Submitted to ***Canadian Geotechnical Journal***

Incorporating region-variability of model bias into liquefaction-triggering procedures for sandy and gravelly soils through BUS-powered hierarchical Bayesian updating

Mao-Xin Wang[a], Y.F. Leung[a], M.K. Lo[a]

[a] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.

Correspondence to: Y.F. Leung (E-mail: andy.yf.leung@polyu.edu.hk)

Abstract

The accuracy of cyclic stress-based liquefaction-triggering assessment procedures can vary systematically from region to region, but it is challenging to regionalize models due to the lack of region-specific data. This paper presents a hierarchical Bayesian modeling (HBM)-based framework for incorporation of inter-region and intra-region variabilities of the bias factor in liquefaction-triggering procedures. A key feature is that the BUS approach (Bayesian Updating with Structural reliability methods) is combined with subset simulation to efficiently update high-dimensional statistics of bias factors. Another feature is a new four-hyperparameter HBM considering both region-specific means and variances of bias factors. This framework is utilized to develop three sets of region-specific liquefaction probability models for practical applications, covering liquefaction-susceptible sandy and gravelly soils. The results show that the four-hyperparameter HBM generally matches better with liquefaction observations and produces larger total variance, compared to the lumped-region modeling and the HBM with only region-specific means. Meanwhile, the population-level distribution and the weighting factor of liquefaction/non-liquefaction occurrence can considerably affect model performance. Furthermore, a discrete integration-based probabilistic method is suggested for liquefaction-triggering hazard assessment. Illustrative examples indicate that different HBM configurations can yield notably different liquefaction hazard results, while neglecting the region-variability tends to be unconservative.

Keywords: Earthquake-induced soil liquefaction; Liquefaction-triggering procedure; Inter-region variability; Hierarchical Bayesian modeling; Bayesian evidence; Region-specific liquefaction probability

1. Introduction

Earthquake-induced soil liquefaction often causes significant structural damage and disruption to infrastructure systems. The liquefaction potential of soils is commonly assessed by the semi-empirical liquefaction-triggering procedures pioneered by Seed and Idriss (1971). Depending on the types of in-situ tests used to measure liquefaction resistance, these procedures can generally be classified into standard penetration test (SPT)-, cone penetration test (CPT)-, and shear wave velocity (Vs)-based categories (e.g., Robertson and Wride 1998; Youd et al. 2001; Moss et al. 2006; Boulanger and Idriss 2014; Robertson 2022). While the existing procedures have been proven to be reliable, many of them follow a deterministic manner by estimating the nominal factor of safety (F_{S0}). Yet, soil liquefaction may still occur if $F_{S0} > 1$ due to significant uncertainty involved in the liquefaction triggering analysis. The assessment of liquefaction potential through probabilistic methods has attracted increasing attention (e.g., Juang et al. 2010; Boulanger and Idriss 2012; Green et al. 2020; Guan and Wang 2022).

To enable probabilistic applications of the existing liquefaction-triggering procedures, extensive efforts have been made to quantify the model uncertainty. For instance, Juang et al. (2002) developed a generic framework to calibrate the uncertainty of liquefaction-triggering model using the first-order reliability method. This framework has been subsequently used for other types of models (e.g., Juang et al. 2012; Duan et al. 2022). In addition, Ku et al. (2012) evaluated the uncertainty associated with the Robertson and Wride (1998) procedure using the maximum likelihood method. The Bayesian network has also been introduced to describe the uncertainty in liquefaction potential assessment (e.g., Cetin et al. 2002; Hu and Liu 2019). The previous studies generally lumped liquefaction case histories around the world together in

56 analyses, such that the model error (or bias) was implicitly assumed to be consistent across
57 different regions. Yet, the accuracy of liquefaction-triggering assessment can vary
58 systematically from region to region (Green 2022), due to the inter-region differences of
59 ground-motion and geologic/geotechnical characteristics (e.g., Liu and Leung 2018; Baker et
60 al. 2021; Xiao et al. 2021).

61 An ideal way to improve the model accuracy is to calibrate the model bias using data only
62 from the region of interest (e.g., Maurer et al. 2019; Green et al. 2020). In most cases, however,
63 the scarce liquefaction data in a region is statistically not adequate for the calibration. To
64 address this issue, Zhang et al. (2016) developed a two-hyperparameter hierarchical Bayesian
65 modeling (HBM) to describe the inter-region variability of the bias factor for a CPT-based
66 procedure. Subsequently, Zhang et al. (2020) calibrated the inter-region variability for a SPT-
67 based procedure. These works focused on the sandy soil liquefaction. Yet, documented field
68 observations have also indicated the consequence of gravel liquefaction (e.g., Brandenberg et
69 al. 2020; Rollins et al. 2022). Therefore, it is of practical importance to calibrate the region-
70 specific model bias in liquefaction potential assessments for both sandy and gravelly soils.

71 The HBM regards the group (or region)-specific statistics as random variables that follow
72 the same population-level distribution parameterized by the hyperparameters. The previously
73 used two-hyperparameter HBM (termed as HBM-2Hyp hereafter) assumes that only the means
74 of the bias factor are region-specific following a lognormal distribution (Zhang et al. 2016,
75 2020). Considering the growing number of documented liquefaction data (Brandenberg et al.
76 2020; Geyin et al. 2021), it becomes more reasonable to also characterize the inter-region
77 variability of the standard deviation of the bias factor. The inter-group (e.g., inter-region or

inter-site) variability of both the means and the variances of group-level variables has been incorporated in some previous studies (e.g., Herschtal et al. 2016; Tabarroki et al. 2024). For instance, Zhang et al. (2014) proposed a four-hyperparameter HBM to describe the inter-site variability of the mean and the coefficient of variation of the model bias factor for pile foundation design. Considering the inter-site variability of the mean and covariance, Ching et al. (2021b) and Wu et al. (2022) developed a computationally efficient HBM framework to construct quasi-site-specific probability distributions for soil/rock properties. On the other hand, different types of the population-level distribution such as normal and inverse gamma distributions have been adopted in HBM applications (e.g., Ching et al. 2021a; Xiao et al. 2021; Tao et al. 2023). It is thus of interest to examine the effects of the region-specific bias factor variance and the population-level distribution type on liquefaction potential assessment. A potential challenge with the incorporation of more HBM hyperparameters and/or target regions is the high-dimensional Bayesian updating problem, which is difficult to be tackled by classical Markov Chain Monte Carlo simulation (MCMCS) methods. Meanwhile, a quantitative index is essential for effectively comparing the performances of different HBM configurations.

This study aims (1) to present a Bayesian updating and model selection framework for calibration of the inter-region and intra-region variabilities of the liquefaction-triggering model bias; and (2) to develop region-specific liquefaction probability models for both sandy and gravelly soils. A key feature of this framework is the BUS approach (Bayesian Updating with Structural reliability methods) (Straub and Papaioannou 2015) with subset simulation (SuS) (Au and Beck 2001) used for posterior sampling and the Bayesian evidence utilized for model selection. Besides, a new four-hyperparameter HBM (termed as HBM-4Hyp hereafter) is

100 established to describe the bias factor, where three candidate distribution types for the
101 population-level probabilistic modeling are examined. The Bayesian framework is
102 subsequently utilized to incorporate the region-variability of model bias into the SPT-based
103 Boulanger and Idriss (2012) procedure (SPT-BI12), the CPT-based Robertson and Wride (1998)
104 procedure (CPT-RW98), and the Vs-based Rollins et al. (2022) procedure (Vs-R22). Finally, a
105 probabilistic liquefaction-triggering hazard analysis method is suggested, and the implication
106 of region-feature modeling for liquefaction-triggering assessment is investigated.

107 **2. General liquefaction potential assessment**

108 In cyclic stress-based liquefaction-triggering procedures, the cyclic stress ratio (CSR) and
109 cyclic resistance ratio (CRR) represent the seismic loading within a soil profile and the
110 liquefaction resistance of soils, respectively. The relationship between the actual factor of
111 safety (F_S) and the nominal F_{S0} can be expressed as:

112 $$F_S = cF_{S0} = c\frac{\mathrm{CRR}}{\mathrm{CSR}} \tag{1}$$

113 where c denotes the model bias factor due to imperfect prediction and the model error is
114 neglected when using c = 1; CSR is usually estimated as (e.g., Youd et al. 2001):

115 $$\mathrm{CSR} = 0.65\frac{\sigma_v}{\sigma_v'}\frac{a_{\max}}{g}r_d\frac{1}{MSF}\frac{1}{K_\sigma} \tag{2}$$

116 where σ_v and σ_v' are the vertical total and effective stresses at depth z, respectively; $a_{\max}$ is
117 peak ground acceleration; g is gravitational acceleration; r_d is the shear stress reduction factor
118 that accounts for the deformability of soil profile; MSF is the earthquake magnitude scaling
119 factor that accounts for the ground motion duration; K_σ is the overburden correction factor.

120 Detailed mathematical expressions of CRR, r_d, MSF, and K_σ for the adopted SPT-BI12,

CPT-RW98, and Vs-R22 procedures are presented in the Supplementary material. Although developed using post-earthquake case histories across different regions, the existing procedures would be directly applied to specific sites, as demonstrated in Fig. 1.

The factor c in Eq. (1) is usually assumed to follow a lognormal distribution with the mean of μ_c and standard deviation of σ_c (e.g., Juang et al. 2012; Zhang et al. 2016), resulting in the normally distributed $\ln F_S$ as follows:

$$\ln F_S \sim \text{Normal}\left\{\ln \mu_c - 0.5\ln\left(1+\sigma_c^2/\mu_c^2\right) + \ln F_{S0},\ \ln\left(1+\sigma_c^2/\mu_c^2\right)\right\} \tag{3}$$

Therefore, the probability of liquefaction (P_{L0}) can be calculated as:

$$P_{L0} = P\left(F_S < 1\right) = P\left(\ln F_S < 0\right) = \Phi\left[\frac{-\ln \mu_c + 0.5\ln\left(1+\sigma_c^2/\mu_c^2\right) - \ln F_{S0}}{\sqrt{\ln\left(1+\sigma_c^2/\mu_c^2\right)}}\right] \tag{4}$$

where $\Phi[\cdot]$ is the cumulative distribution function (CDF) of the standard normal variable. Based on the liquefaction case histories, μ_c and σ_c can be estimated from Bayesian inference by lumping data across different regions, as illustrated in Fig. 2a.

3. Bayesian framework for development of region-specific liquefaction-triggering models

3.1. Inter- and intra-region variabilities of model bias

The model bias may be similar for different cases in a region (e.g., low intra-region variability) while different regions may result in systematically different biases (e.g., high inter-region variability). Thus, the total variability of the bias factor can be separated into the inter- and intra-region components by defining the region-specific bias factor c_i (i = 1, 2, …, r), where r denotes the number of regions considered. Similar to Eq. (3), c_i is assumed to follow a lognormal distribution with the mean of μ_{ci} and standard deviation of σ_{ci}. These statistics can be estimated from liquefaction data through two Bayesian modeling ideas, which are shown in

Fig. 2b and c. The HBM-2Hyp uses region-specific μ_{ci} and region-invariant $\sigma_c = \sigma_{c1} = \sigma_{c2} = \ldots = \sigma_{cr}$, while the HBM-4Hyp considers both μ_{ci} and σ_{ci} as region-specific variables.

The region-specific means or standard deviations are considered to be similar by assuming a population-level distribution of μ_{ci} (or σ_{ci}) parametrized by two hyperparameters, which are the mean μ_μ (or μ_σ) and standard deviation σ_μ (or σ_σ) of μ_{ci} (or σ_{ci}). As a result, there are two hyperparameters (μ_μ and σ_μ) and four hyperparameters (μ_μ, σ_μ, μ_σ, and σ_σ) in HBM-2Hyp and HBM-4Hyp, respectively (see Fig. 2). The hyperparameter μ_μ represents the "average" model bias across different regions, whereas σ_μ measures the inter-region variability of the average bias. Also, μ_σ represents the "average" intra-region variability and σ_σ reflects the inter-region variability of the bias uncertainty. Suppose that $\boldsymbol{\Theta}$ is the vector of all uncertain parameters required in the probabilistic modeling, resulting in $\boldsymbol{\Theta} = [\mu_\mu, \sigma_\mu, \sigma_c, \mu_{c1}, \mu_{c2}, \ldots, \mu_{cr}]$ and $[\mu_\mu, \sigma_\mu, \mu_\sigma, \sigma_\sigma, \mu_{c1}, \mu_{c2}, \ldots, \mu_{cr}, \sigma_{c1}, \sigma_{c2}, \ldots, \sigma_{cr}]$ for HBM-2Hyp and HBM-4Hyp, respectively.

For given $[\mu_\mu, \sigma_\mu]$ or $[\mu_\sigma, \sigma_\sigma]$, μ_{ci} or σ_{ci} can be described by the population-level distribution. Three candidate distributions are considered in this study, including lognormal distribution, normal distribution truncated to be positive (Trunc-normal), and Gumbel distribution truncated to be positive (Trunc-Gumbel). For instance, the probability distribution function (PDF) of μ_{ci} or σ_{ci} for lognormal distribution can be expressed as (e.g., Wang et al. 2020; Deng et al. 2022):

$$f\left(x_{ci}\middle|\mu_x,\sigma_x,Mdl\right)=\frac{1}{x_{ci}\sqrt{2\pi\ln\left(1+\sigma_x^2/\mu_x^2\right)}}\exp\left\{-\frac{\left[\ln x_{ci}-\ln\mu_x+0.5\ln\left(1+\sigma_x^2/\mu_x^2\right)\right]^2}{2\ln\left(1+\sigma_x^2/\mu_x^2\right)}\right\} \quad (5)$$

where x denotes "μ" for $f\left(\mu_{ci}\middle|\mu_\mu,\sigma_\mu,Mdl\right)$ or "σ" for $f\left(\sigma_{ci}\middle|\mu_\sigma,\sigma_\sigma,Mdl\right)$; Mdl represents the model configuration (e.g., Mdl = [CPT-4Hyp, lognormal]). The PDFs for the other candidate distributions are listed in Table 1, where p and q denote distribution-specific parameters and

are transformed from [μ_μ, σ_μ] or [μ_σ, σ_σ]. The parameters μ_μ, σ_μ, μ_σ, and σ_σ are random variables, and their samples are generated through MCS or MCMCS in Bayesian updating. The statistics of these parameters can be estimated from posterior samples and will be explicitly provided in the next section. Increasing μ_μ and σ_μ would lead to lower overall P_{L0} and larger inter-region difference in P_{L0}, respectively. Additionally, P_{L0} for each region would generally become greater with increasing μ_σ at $F_{S0} > 1$ levels, whereas increasing σ_σ tends to enlarge the inter-region difference of P_{L0}.

3.2. Hierarchical Bayesian modeling of the inter- and intra-region variabilities

Let **Data** = [$\mathbf{Dt}_1$, $\mathbf{Dt}_2$, …, $\mathbf{Dt}_r$] denote case histories from r earthquake influencing regions and $\mathbf{Dt}_i$ denote [Dt_{i1}, Dt_{i2}, …, Dt_{i,n_i}], where Dt_{ij} = 1 for liquefied case and 0 otherwise, and n_i is the number of observations in the i-th region. For given Mdl and **Data**, $\boldsymbol{\Theta}$ can be estimated by Bayesian updating, following the general idea below (e.g., Cao et al. 2019; Lo and Leung 2019; Zhang et al. 2022):

$$f\left(\boldsymbol{\Theta}\middle|\mathbf{Data},Mdl\right)=KL\left(\mathbf{Data}\middle|\boldsymbol{\Theta},Mdl\right)f\left(\boldsymbol{\Theta}\middle|Mdl\right) \tag{6}$$

where $f(\boldsymbol{\Theta}|\mathbf{Data},Mdl)$ denotes the posterior joint PDF of $\boldsymbol{\Theta}$; $K=\left[\int_{\boldsymbol{\Theta}} L\left(\mathbf{Data}\middle|\boldsymbol{\Theta},Mdl\right)f\left(\boldsymbol{\Theta}\middle|Mdl\right)d\boldsymbol{\Theta}\right]^{-1}$ is a normalization constant that is independent of $\boldsymbol{\Theta}$; $f(\boldsymbol{\Theta}|Mdl)$ is the prior joint PDF of $\boldsymbol{\Theta}$ and is expressed as:

$$f\left(\boldsymbol{\Theta}\middle|Mdl\right)=\begin{cases} f\left(\mu_\mu\right)f\left(\sigma_\mu\right)f\left(\sigma_c\right)\prod_{i=1}^{r}f\left(\mu_{ci}\middle|\mu_\mu,\sigma_\mu,Mdl\right) & \text{for 2Hyp} \\ f\left(\mu_\mu\right)f\left(\sigma_\mu\right)f\left(\mu_\sigma\right)f\left(\sigma_\sigma\right)\prod_{i=1}^{r}\left[f\left(\mu_{ci}\middle|\mu_\mu,\sigma_\mu,Mdl\right)f\left(\sigma_{ci}\middle|\mu_\sigma,\sigma_\sigma,Mdl\right)\right] & \text{for 4Hyp} \end{cases} \tag{7}$$

where $f\left(\mu_{ci}\middle|\mu_\mu,\sigma_\mu\right)$ and $f\left(\sigma_{ci}\middle|\mu_\sigma,\sigma_\sigma\right)$ are obtained from Eq. (5) or equations presented in Table 1; $f(\mu_\mu)$, $f(\sigma_\mu)$, $f(\mu_\sigma)$, and $f(\sigma_\sigma)$ are the prior marginal PDFs of μ_μ, σ_μ, μ_σ, and σ_σ,

184 respectively. According to the previous studies (Zhang et al. 2016, 2020), each prior PDF is
185 taken as 1/3 by assuming a uniform distribution within the [0, 3] interval.

186 The $L(\mathbf{Data}|\mathbf{\Theta},Mdl)$ term in Eq. (6) represents the likelihood function for the observed data,
187 and is expressed as:

188 $$L\left(\mathbf{Data}\middle|\mathbf{\Theta},Mdl\right)=\prod_{i=1}^{r}\left[\prod_{j=1}^{n_{Li}}P_{Lij}^{w_{Li}\left(Dt_{ij}\right)}\prod_{j=n_{Li}+1}^{n_i}\left(1-P_{Lij}\right)^{w_{NLi}\left(1-Dt_{ij}\right)}\right] \tag{8}$$

189 where n_{Li} is the number of liquefied cases in the i-th region; Dt_{ij} is taken as 1 for j = 1, 2, …,
190 n_{Li} and 0 for $j \geq n_{Li}+1$; P_{Lij} is the predicted liquefaction probability for the j-th case history in
191 the i-th region, and is estimated by adjusting Eq. (4) as:

192 $$P_{Lij}=\Phi\left(\frac{-\ln\mu_{ci}+0.5\ln\left(1+\sigma_{ci}^2/\mu_{ci}^2\right)-\ln F_{Sij}}{\sqrt{\ln\left(1+\sigma_{ci}^2/\mu_{ci}^2\right)}}\right) \tag{9}$$

193 in which σ_{ci} is simply taken as σ_c in the 2Hyp modeling; F_{Sij} is the calculated safety factor.

194 As parameters used to reduce the sampling bias effects, w_{Li} and w_{NLi} in Eq. (8) are the
195 weighting factors of the liquefaction and non-liquefaction cases in the i-th region, respectively,
196 which are defined as (e.g., Cetin et al. 2002; Zhang et al. 2016):

197 $$w_{Li}=\frac{Q_p}{Q_{si}}=\frac{Q_p\left(n_{Li}+n_{NLi}\right)}{n_{Li}} \tag{10}$$

198 $$w_{NLi}=\frac{1-Q_p}{1-Q_{si}}=\frac{\left(1-Q_p\right)\left(n_{Li}+n_{NLi}\right)}{n_{NLi}} \tag{11}$$

199 where Q_{si} is the sample proportion of liquefied cases in the i-th region; Q_p denotes the true
200 proportion of liquefied cases in nature. Since Q_p is always unknown, a typical way of estimating
201 w_{NLi} and w_{Li} is to specify a reasonable w_{NLi}/w_{Li} ratio. Cetin et al. (2002) consulted with a panel
202 of eight experts, suggesting that the most common estimate of the ratio should be between 1.5
203 and 2. They further conducted Bayesian updating analyses and found that w_{NLi}/w_{Li}=1.5 yielded

the optimal result. This value has been consistently adopted in many subsequent studies (e.g., Boulanger and Idriss 2012), regardless of the specific databases and methods.

Using $w_{NLi}/w_{Li} = 1.5$ and $Q_{si} = Q_{s,total} = \sum_{i=1}^{r} n_{Li} \Big/ \sum_{i=1}^{r} \left(n_{Li} + n_{NLi}\right)$, Q_p can thus be computed as:

$$Q_p = \frac{1}{1.5/Q_{s,total} - 0.5} = \frac{1}{1.5\sum_{i=1}^{r}\left(n_{Li} + n_{NLi}\right) \Big/ \sum_{i=1}^{r} n_{Li} - 0.5} \tag{12}$$

where $Q_{s,total}$ represents the total sample proportion of all liquefaction cases. Given $Q_{s,total}$ = 0.557 (Cetin et al. 2002), Zhang et al. (2016) derived $Q_p = 0.456$ for their model development, and this value was subsequently used for another dataset (Zhang et al. 2020). Yet, Q_p may be considered as dataset-specific as long as the weighting ratio of 1.5 is satisfied. In this study, both the fixed $Q_p = 0.456$ and the calculated dataset-specific Q_p (from Eq. (12)) are taken as candidate modeling choices, and their resulting models will be compared later.

3.3. Bayesian updating through the BUS approach with subset simulation

The BUS approach (Straub and Papaioannou 2015) is utilized herein to evaluate the posterior PDF of $\boldsymbol{\Theta}$ in Eq. (7). This approach is powerful in dealing with high-dimensional Bayesian inference (Betz et al. 2018a, 2018b; Jiang et al. 2018). The basic idea of BUS is to treat the updating problem as a structural reliability problem such that the samples in an equivalent failure domain Ω_Θ are taken as the posterior samples. The observed information described by $L(\mathbf{Data}|\boldsymbol{\Theta},Mdl)$ is used to define Ω_Θ as:

$$\Omega_\Theta = \left\{ H\left(\boldsymbol{\Theta}_+ \middle| Mdl\right) \le 0 \right\} \tag{13}$$

where $\boldsymbol{\Theta}_+$ is the augmented outcome space [$\boldsymbol{\Theta}$, ζ]; ζ is the standard uniform variable in [0, 1]

interval; $H(\boldsymbol{\Theta}_+)$ is the limit state function:

$$H\left(\boldsymbol{\Theta}_+ \middle| Mdl\right) = \ln\zeta - \ln\left[\kappa L\left(\mathbf{Data}\middle|\boldsymbol{\Theta}, Mdl\right)\right] \tag{14}$$

where κ is the likelihood multiplier to ensure $\kappa L(\mathbf{Data}|\boldsymbol{\Theta},Mdl) \leq 1$ and will be determined adaptively as discussed later.

The subset simulation (Au and Beck 2001) is coupled with the BUS approach (termed as BUS-SuS hereafter) to efficiently solve the reliability problem in obtaining the posterior samples. Consequently, the occurrence probability of the event $Z = \{\boldsymbol{\Theta}_+ \in \Omega_\Theta\}$) can be represented as a product of larger conditional probabilities of a set of intermediate events $Z_l = \{H(\boldsymbol{\Theta}_+|Mdl) < h_l\}$ ($l = 1, 2, \ldots, m$), expressed as:

$$P\left\{\boldsymbol{\Theta}_+ \in \Omega_\Theta\right\} = P\left(Z\right) = P\left(Z_1\right)\prod_{l=2}^{m} P\left(Z_l\middle|Z_{l-1}\right) \tag{15}$$

where these intermediate events satisfy $Z_1 \supset Z_2 \supset \ldots \supset Z_{m-1} \supset Z_m$; $h_1 > h_2 > \ldots > h_m = 0$ are a decreasing series of threshold values to be determined automatically. The intermediate probabilities $P(Z_1)$ and $P(Z_l|Z_{l-1})$ are equal to a predetermined value p_0, which is specified as 0.1 (Betz et al. 2018b).

BUS-SuS starts with a direct Monte Carlo simulation (MCS) of N_l sets of $\boldsymbol{\Theta}_+$ samples, based on (i) the [0, 3] uniform distributions for $[\mu_\mu, \sigma_\mu, \sigma_c]$ or $[\mu_\mu, \sigma_\mu, \mu_\sigma, \sigma_\sigma]$, (ii) the population-level distributions for μ_{ci} or $[\mu_{ci}, \sigma_{ci}]$, and (iii) standard uniform distribution for ζ in the first level ($l = 1$). Specifically, the population-level sampling is performed with isoprobabilistic transformation expressed as:

$$\mu_{ci} = F^{-1}\left(\Phi\left(\xi\right)\middle|\mu_\mu, \sigma_\mu, Mdl\right) \tag{16a}$$

$$\sigma_{ci} = F^{-1}\left(\Phi\left(\xi\right)\middle|\mu_\sigma, \sigma_\sigma, Mdl\right) \tag{16b}$$

where ξ is a random realization of the standard normal variable; $F^{-1}\left(\cdot\middle|\mu_\mu,\sigma_\mu,Mdl\right)$ and $F^{-1}\left(\cdot\middle|\mu_\sigma,\sigma_\sigma,Mdl\right)$ represent the inverses of CDFs of μ_{ci} and σ_{ci}, respectively, which are included in Table 1.

The $\boldsymbol{\Theta}_+$ samples are substituted into Eq. (14) to calculate the corresponding $H(\boldsymbol{\Theta}_+|Mdl)$ values, which are ranked in an ascending order and the (p_0N_l+1)-th ranking value is taken as h_1. Consequently, the p_0N_s samples with the lowest $H(\boldsymbol{\Theta}_+|Mdl)$ values are regarded as seeds to generate $(1\text{-}p_0)N_s$ new samples of $\boldsymbol{\Theta}_+$ for the Z_2 event via MCMCS (Papaioannou et al. 2015). Both MCS and MCMCS are performed in the standard normal space. Similarly, the sample ranking and MCMCS are conducted for the next SuS level ($l = 2$), producing N_l samples of $\boldsymbol{\Theta}_+$. This process will be continued until the (p_0N_l+1)-th ranking value of $H(\boldsymbol{\Theta}_+)$ is equal to or less than 0. During the process, κ in Eq. (14) is evaluated adaptively as the reciprocal of the maximum of the $L(\mathbf{Data}|\boldsymbol{\Theta},Mdl)$ values in the l-th SuS level, and is written as (Jiang et al. 2018):

$$\kappa_l = \frac{1}{\max\left\{1/\kappa_{l-1},\left[L_{l,1},L_{l,2},...,L_{l,N_l}\right]\right\}} \tag{17}$$

where $\kappa_{l\text{-}1}$ and κ_l represent the constant values in the (l-1)-th and l-th levels, respectively. Once the SuS process is finished, the failure samples in the final level are taken as seeds to generate N_l samples in the Ω_Θ domain, which are treated as the posterior samples of interest.

The resulting posterior statistics (i.e., mean E(·) and standard deviation Std(·)) of $\boldsymbol{\Theta} = [\mu_\mu, \sigma_\mu, \sigma_c, \mu_{c1}, \mu_{c2}, \ldots, \mu_{cr}]$ or $[\mu_\mu, \sigma_\mu, \mu_\sigma, \sigma_\sigma, \mu_{c1}, \mu_{c2}, \ldots, \mu_{cr}, \sigma_{c1}, \sigma_{c2}, \ldots, \sigma_{cr}]$ are used to estimate E(·) and Std(·) of the region-specific bias factors c_i (i = 1, 2, …, r) through the laws of total expectation and total variance, which can be expressed as:

$$\mathrm{E}\left(c_i\right) = \mathrm{E}\left(\mu_{ci}\right) \tag{18}$$

$$\mathrm{Std}\left(c_i\right)=\begin{cases}\sqrt{\mathrm{E}^2\left(\sigma_c\right)+\mathrm{Std}^2\left(\mu_{ci}\right)+\mathrm{Std}^2\left(\sigma_c\right)} & \text{for 2Hyp}\\ \sqrt{\mathrm{E}^2\left(\sigma_{ci}\right)+\mathrm{Std}^2\left(\mu_{ci}\right)+\mathrm{Std}^2\left(\sigma_{ci}\right)} & \text{for 4Hyp}\end{cases} \tag{19}$$

For other regions outside the dataset, the bias factor is denoted as c_0, of which E(·) and Std(·) can be similarly estimated as:

$$\mathrm{E}\left(c_0\right)=\mathrm{E}\left(\mu_\mu\right) \tag{20}$$

$$\mathrm{Std}\left(c_0\right)=\sqrt{\mathrm{E}^2\left(\sigma_{c0}\right)+\mathrm{Std}^2\left(\mu_{c0}\right)+\mathrm{Std}^2\left(\sigma_{c0}\right)} \tag{21}$$

where $\mathrm{E}(\sigma_{c0})$ is equal to $\mathrm{E}(\sigma_c)$ and $\mathrm{E}(\mu_\sigma)$ for 2Hyp and 4Hyp, respectively; $\mathrm{Std}(\sigma_{c0})$ is equal to $\mathrm{Std}(\sigma_c)$ and $\sqrt{\mathrm{E}^2\left(\sigma_\sigma\right)+\mathrm{Std}^2\left(\mu_\sigma\right)+\mathrm{Std}^2\left(\sigma_\sigma\right)}$ for 2Hyp and 4Hyp, respectively; $\mathrm{Std}(\mu_{c0}) = \sqrt{\mathrm{E}^2\left(\sigma_\mu\right)+\mathrm{Std}^2\left(\mu_\mu\right)+\mathrm{Std}^2\left(\sigma_\mu\right)}$.

3.4. Model selection through Bayesian evidence with Gaussian copula

When N_c candidate HBM configurations Mdl_k (k = 1, 2, …, N_c) are available, it is desirable to select the most probable candidate through Bayesian model comparison techniques, such as residual analysis, leave-some-out cross-validation analysis, likelihood function evaluation, and deviance information criterion (e.g., Bozorgzadeh et al. 2019; Wu et al. 2022; Tao et al. 2023; Tabarroki et al. 2024). One of the most commonly used performance indexes is the evidence derived within Bayesian analysis (e.g., Yuen 2010; Cao and Wang 2014). For instance, Zhang et al. (2016) suggested an evidence calculation method for comparing HBM-based liquefaction-triggering models. In this study, a copula-based method is introduced to derive the evidence for HBM. Therefore, the general case where the posterior PDF may not be approximated by a specific distribution (e.g., normal distribution) can be addressed. Previous studies have shown that copula functions provide a flexible way of establishing the joint distribution in geotechnical applications (e.g., Tian et al. 2016; Tang et al. 2020).

The plausibility of a model class Mdl_k given **Data** can be estimated by applying Bayes' theorem (e.g., Cao and Wang 2014; Zhang et al. 2016):

$$P\left(Mdl_k\left|\mathbf{Data}\right.\right)=\frac{P\left(\mathbf{Data}\left|Mdl_k\right.\right)P\left(Mdl_k\right)}{P\left(\mathbf{Data}\right)} \tag{22}$$

where $P(\mathbf{Data})$ is a normalization constant independent of Mdl_k; $P(Mdl_k)$ is the prior probability of the k-th candidate and is typically set to $1/N_c$ for equal weights. Therefore, the ranking of $P(Mdl_k|\mathbf{Data})$ for different models is only dependent on $P(\mathbf{Data}|Mdl_k)$, which is usually termed as *evidence*, and can be expressed as:

$$P\left(\mathbf{Data}\left|Mdl_k\right.\right)=\int_{\boldsymbol{\Theta}} L\left(\mathbf{Data}\left|\boldsymbol{\Theta},Mdl_k\right.\right)f\left(\boldsymbol{\Theta}\left|Mdl_k\right.\right)d\boldsymbol{\Theta} \tag{23}$$

For a general case without requirement for normality of the posterior distribution, the logarithm of evidence can be derived as (Yuen 2010):

$$\begin{aligned}\ln\left[P\left(\mathbf{Data}\left|Mdl_k\right.\right)\right]=&\int_{\boldsymbol{\Theta}}\ln\left[L\left(\mathbf{Data}\left|\boldsymbol{\Theta},Mdl_k\right.\right)\right]f\left(\boldsymbol{\Theta}\left|\mathbf{Data},Mdl_k\right.\right)d\boldsymbol{\Theta}\\&-\int_{\boldsymbol{\Theta}}\ln\left[\frac{f\left(\boldsymbol{\Theta}\left|\mathbf{Data},Mdl_k\right.\right)}{f\left(\boldsymbol{\Theta}\left|Mdl_k\right.\right)}\right]f\left(\boldsymbol{\Theta}\left|\mathbf{Data},Mdl_k\right.\right)d\boldsymbol{\Theta}\end{aligned} \tag{24}$$

where the first integral represents the average logarithmic goodness-of-fit of a candidate model, and the second integral measures the relative entropy between the prior and posterior PDFs.

Given N_c posterior samples of $\boldsymbol{\Theta}$ derived from the BUS-SuS approach, Eq. (24) can be numerically evaluated through:

$$\ln\left[P\left(\mathbf{Data}\left|Mdl_k\right.\right)\right]=\frac{1}{N_l}\left\{\sum_{j=1}^{N_l}\ln\left[L\left(\mathbf{Data}\left|\hat{\boldsymbol{\Theta}}_j,Mdl_k\right.\right)\right]-\sum_{j=1}^{N_l}\ln\left[\frac{f\left(\hat{\boldsymbol{\Theta}}_j\left|\mathbf{Data},Mdl_k\right.\right)}{f\left(\hat{\boldsymbol{\Theta}}_j\left|Mdl_k\right.\right)}\right]\right\} \tag{25}$$

where $f\left(\hat{\boldsymbol{\Theta}}_j\left|Mdl_k\right.\right)$ and $L\left(\mathbf{Data}\left|\hat{\boldsymbol{\Theta}}_j,Mdl_k\right.\right)$ are calculated from Eqs. (7) and (8) with the j-th posterior sample $\hat{\boldsymbol{\Theta}}_j$, respectively; $f\left(\hat{\boldsymbol{\Theta}}_j\left|\mathbf{Data},Mdl_k\right.\right)$ can be expressed through copula theory (e.g., Tian et al. 2016; Tang et al. 2020), representing the joint posterior PDF of $\boldsymbol{\Theta}$ as the product

of the posterior marginal PDFs of individual parameters and a copula density function. A feature of the copula method is its flexibility in characterizing the posterior PDF.

In this study, the Gaussian copula is selected such that $f\left(\hat{\boldsymbol{\Theta}}_j \middle| \mathbf{Data}, Mdl_k\right)$ is estimated as:

$$f\left(\hat{\boldsymbol{\Theta}} \middle| \mathbf{Data}, Mdl_k\right) = \begin{cases} D\left(\hat{F}\left(\hat{\mu}_\mu\right), \hat{F}\left(\hat{\sigma}_\mu\right), \hat{F}\left(\hat{\sigma}_c\right), \hat{F}\left(\hat{\mu}_{c1}\right), \hat{F}\left(\hat{\mu}_{c2}\right), \ldots, \hat{F}\left(\hat{\mu}_{cr}\right); \boldsymbol{\theta}\right) & \\ \times \hat{f}\left(\hat{\mu}_\mu\right) \hat{f}\left(\hat{\sigma}_\mu\right) \hat{f}\left(\hat{\sigma}_c\right) \prod_{i=1}^{r} \hat{f}\left(\hat{\mu}_{ci}\right) & \text{for 2Hyp} \\ D\begin{pmatrix} \hat{F}\left(\hat{\mu}_\mu\right), \hat{F}\left(\hat{\sigma}_\mu\right), \hat{F}\left(\hat{\mu}_\sigma\right), \hat{F}\left(\hat{\sigma}_\sigma\right), \\ \hat{F}\left(\hat{\mu}_{c1}\right), \hat{F}\left(\hat{\mu}_{c2}\right), \ldots, \hat{F}\left(\hat{\mu}_{cr}\right), \hat{F}\left(\hat{\sigma}_{c1}\right), \hat{F}\left(\hat{\sigma}_{c2}\right), \ldots, \hat{F}\left(\hat{\sigma}_{cr}\right); \boldsymbol{\theta} \end{pmatrix} & \\ \times \hat{f}\left(\hat{\mu}_\mu\right) \hat{f}\left(\hat{\sigma}_\mu\right) \hat{f}\left(\hat{\mu}_\sigma\right) \hat{f}\left(\hat{\sigma}_\sigma\right) \prod_{i=1}^{r} \hat{f}\left(\hat{\mu}_{ci}\right) \prod_{i=1}^{r} \hat{f}\left(\hat{\sigma}_{ci}\right) & \text{for 4Hyp} \end{cases} \tag{26}$$

where $\hat{F}(\cdot)$ represents the empirical marginal CDFs for the posterior samples of individual parameters; $\hat{f}(\cdot)$ denotes the kernel smoothing-based marginal PDFs for the posterior parameter samples; $D(\cdot)$ is the copula density function and is written as:

$$D\left(u_1, u_2, \ldots, u_n; \boldsymbol{\theta}\right) = \det\left(\boldsymbol{\theta}\right)^{-1/2} \exp\left[-\frac{1}{2}\boldsymbol{\eta}\left(\boldsymbol{\theta}^{-1} - \mathbf{I}\right)\boldsymbol{\eta}^T\right] \tag{27}$$

where n is the number of variables considered; $\boldsymbol{\eta}$ is a 1-by-n vector $[\Phi^{-1}(u_1), \Phi^{-1}(u_2), \ldots, \Phi^{-1}(u_n)]$, in which $\Phi^{-1}(\cdot)$ represents the inverse of CDF of the standard normal variable; $\det(\cdot)$ is determinant; T denotes matrix transpose; $\mathbf{I}$ is an n-by-n identity matrix; $\boldsymbol{\theta}$ is a n-by-n dependence parameter matrix. The maximum-likelihood estimation is adopted to derive $\boldsymbol{\theta}$ as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{j=1}^{N_l} \ln D\left(u_{1j}, u_{2j}, \ldots, u_{nj}; \boldsymbol{\theta}\right) \tag{28}$$

For each candidate model configuration, the posterior samples of $\boldsymbol{\Theta}$ generated from the BUS-SuS approach are substituted into Eq. (25) to calculate $\ln\left[P\left(\mathbf{Data}\middle|Mdl_k\right)\right]$. Subsequently, the configuration with the largest value of $\ln\left[P\left(\mathbf{Data}\middle|Mdl_k\right)\right]$ is identified as the best-fit one. The performance of the evidence estimator is dependent on the quality of the fitting of posterior

PDFs, and an inappropriate fitting can result in biased estimates. It is worth noting that Bayesian evidence can also be obtained as a by-product of BUS using the algorithms proposed by Betz et al. (2018a, 2018b). Comparative results showed that there was generally a strong correlation between the evidences derived from the suggested method and the BUS estimator. The suggested estimator can be used regardless of the specific BUS algorithm employed, and thus may serve as an alternative or supplement to the BUS estimator. Fig. 3 summarizes the flowchart of the proposed Bayesian framework.

4. Bayesian updating for three liquefaction-triggering procedures

4.1. Dataset preparation and model configurations

Based on the Bayesian framework, the SPT- and CPT-based case history databases for sandy soils compiled in Boulanger and Idriss (2014) are used to develop the region-specific SPT-BI12 and CPT-RW98 models, respectively, whereas the database documented by Rollins et al. (2022) is adopted to develop the region-specific Vs-R22 models for gravelly soils. After excluding some undesirable data (e.g., regions with inadequate data and case histories with unclear liquefaction classification or unrealistically large F_{S0}), the final datasets for SPT-BI12, CPT-RW98, and Vs-R22 contain 217, 234, and 153 case histories, respectively. Detailed exclusion criteria and dataset information are given in the Supplementary material. For each liquefaction-triggering procedure, 12 candidate model configurations (2 hyperparameter choices × 3 population-level distributions × 2 Q_p values) are considered in this study.

4.2. Bayesian updating results

Fig. 4a-c display the logarithmic evidence (i.e., $\ln[P(\mathbf{Data}|Mdl_k)]$) values of the 12 candidate

configurations for the SPT-BI12, CPT-RW98, and Vs-R22 models, respectively. It is observed that the conventional lognormal distribution produces lower evidence than the other distributions in many cases, and is therefore generally not the best-fit population-level distribution. In addition, the evidence for HBM-4Hyp is generally higher than that for HBM-2Hyp, due to the more complete statistical representation by HBM-4Hyp. Considering the best-fit population-level distribution, the posterior E(·) and Std(·) of the parameters [μ_μ, σ_μ, σ_c, c_1, c_2, …, c_r] or [μ_μ, σ_μ, μ_σ, σ_σ, c_1, c_2, …, c_r] for the updated SPT-BI12, CPT-RW98, and Vs-R22 models are listed in Tables 2-4, respectively. Several observations are made as follows:

(1) The observed E(μ_μ) > 1 indicates that the deterministic model tends to, on average, underestimate F_S. Also, the positive E(σ_μ) and E(σ_σ) for different models illustrate the existence of the inter-region variability. Specifically, the inter-region variability for the Vs-R22 models (with E(σ_μ)=0.95~0.99 and E(σ_σ)=0.71~0.72) is generally larger than that for the other models, whereas the SPT-BI12 models correspond to the smallest inter-region variability (with E(σ_μ)=0.11~0.15 and E(σ_σ)=0.54~0.57). Meanwhile, E(μ_σ) reflects the intra-region variability, which appears to be the largest for the CPT-RW98 models (with E(μ_σ)=0.81~0.87).

(2) E(c_i) and Std(c_i) exhibit notable region-to-region differences, which are in accordance with the positive E(σ_μ) and E(σ_σ), respectively. For example, the influenced region by the 1989 Loma Prieta earthquake corresponds to remarkably larger E(c_i) than the other regions for the Vs-R22 models. This is because the calculated F_{S0} values are consistently smaller than 1, which is opposite to the observed nonexistence of liquefaction phenomena in this region.

(3) The E(c_i) values for HBM-2Hyp and HBM-4Hyp are generally comparable. Yet, HBM-4Hyp produces larger Std(c_i); the reason is that the inter-region variability for HBM-4Hyp is

represented by the inter-region differences of both μ_{ci} and σ_{ci}, enlarging the total uncertainty of c_i. This is consistent with the larger evidence for HBM-4Hyp (see Fig. 4).

(4) The dataset-specific Q_p calculated from Eq. (12) generally yields smaller E(μ_μ) and E(c_i) compared to Q_p = 0.456. This is because the calculated Q_p is consistently larger than 0.456 and is related to higher population proportion of liquefaction occurrence.

To display the posterior distributions of model parameters, Fig. 5 presents the histograms of [μ_μ, σ_μ, σ_c] or [μ_μ, σ_μ, μ_σ] in the updated CPT-RW98 and SPT-BI12 models considering different combinations of Q_p and hyperparameters. The uniformly distributed prior relative frequency is also shown for comparison. It is observed that the model parameters after Bayesian updating are distributed in considerably narrower ranges. This is more evident for the SPT-BI12 models due to their weaker inter- and intra-region variabilities. Note that the best-fit population-level distribution will be mainly used in subsequent analyses, with focus on evaluating the effects of hyperparameter and Q_p selections.

4.3. Relationship between liquefaction probability and nominal factor of safety

Based on the updating results in Tables 2-4, the liquefaction probability P_{L0} can be calculated for a region of interest by substituting F_{S0} (from Eq. (1)) as well as E(c_i) and Std(c_i) into Eq. (4), which is expressed as:

$$P_{L0} = \Phi\left[\frac{-\ln \mathrm{E}\left(c_i\right) + 0.5\ln\left(1+\mathrm{Std}^2\left(c_i\right)/\mathrm{E}^2\left(c_i\right)\right) - \ln F_{S0}}{\sqrt{\ln\left(1+\mathrm{Std}^2\left(c_i\right)/\mathrm{E}^2\left(c_i\right)\right)}}\right] \tag{29}$$

Eq. (29) represents an approximate simplified expression for the liquefaction probability accounting for the region-specific variability. Therefore, the E(c_i) and Std(c_i) values provided in Tables 2-4, derived using the laws of total expectation and variance, can be directly employed

without the requirement for posterior samples. An alternative MCS-based approach of averaging the liquefaction probability values obtained for individual posterior samples of μ_{ci} and σ_{ci} may provide more consistent estimates.

To illustrate the performance of the region-specific P_{L0} models, the probability mass function (PMF) of the number of predicted liquefied cases in each region is obtained and compared with the actual number of liquefied cases. Specifically, the sum of independent indicators of liquefaction (= 1) or non-liquefaction (= 0) for individual case histories in each region is assumed to follow a Poisson binomial distribution, where the occurrence probability of each indicator is specified as a case-history-specific P_{L0} value. In this study, PMF is obtained by the 'poibin' package in the **R** software (Hong 2013).

Fig. 6a and b display the PMF results for the region-specific SPT-BI12 and CPT-RW98 models, respectively. It is seen that the actual number of liquefied cases is generally within the range of relatively large PMF. The HBM-4Hyp, which tends to predict more liquefied cases than HBM-2Hyp, is observed to better match the actual liquefaction observations. Meanwhile, using the dataset-specific Q_p calculated from Eq. (12) appears to partially improve the fitting performance of the CPT-RW98 models (e.g., in Loma Prieta and Canterbury regions), for which the dataset-specific $Q_p = 0.649$ is considerably larger than the conventionally used $Q_p = 0.456$. For the SPT-BI12 models, the PMFs for the two Q_p choices (0.461 and 0.456) are not significantly different. These graphical results are consistent with the quantitative ranking of model evidence in Fig. 5. Note that the PMF-based comparison between the actual number of liquefied cases and the HBM predictions does not consider the sample imbalance, although the model-to-model difference can be illustrated. One may assign a reasonable weighting factor

409 for a fairer comparison.

410 **4.4. Comparison of region-specific models for different HBM configurations**

411 Using Q_p = 0.456, Fig. 7a and b present the P_{L0} versus F_{S0} trends for the region-specific SPT-
412 BI12 and CPT-RW98 models, respectively. Although not shown herein, the predicted P_{L0}
413 would generally become larger with increasing Q_p. Several findings can be obtained. First, P_{L0}
414 decreases with increasing F_{S0} (or safety level), and $P_{L0} > 0.5$ can still be observed for $F_{S0} > 1$
415 due to the uncertainty involved. Second, there are evident region-to-region differences of P_{L0},
416 though they are relatively small for the SPT-based models. It is thus necessary to account for
417 the inter-region variability of model bias. Third, HBM-4Hyp results in more notable inter-
418 region difference of P_{L0} than HBM-2Hyp, which is mainly attributed to the larger inter-region
419 difference of Std(c_i) in HBM-4Hyp. In contrast, the lumped-region modeling that neglects
420 region-specific features (see Fig. 2a) would lead to inaccurate estimations.

421 Furthermore, Fig. 7c compares the P_{L0}-F_{S0} relationships derived from different HBM
422 configurations. The comparison is based on the CPT-RW98 and Vs-R22 models for regions
423 not included in the calibration databases. The 4Hyp-based models generally yield greater P_{L0}
424 than the 2Hyp-based models at a given F_{S0}, especially for CPT-RW98. Additionally, the
425 predicted P_{L0} becomes greater with increasing Q_p, and is also notably influenced by the
426 selection of population-level distribution.

427 A more detailed comparison of different HBM configurations is conducted. The influence
428 of population-level distribution is illustrated in Fig. 8a using the fixed [2Hyp, Q_p=0.456].
429 Specifically, the ratios of P_{L0} for the three candidate distributions to P_{L0} for the best-fit
430 distribution at multiple F_{S0} levels are displayed, where the boxplot represents the distribution

of ratios for all regions. It is observed that the conventional lognormal distribution tends to produce smaller P_{L0} compared to the generally identified best-fit Trunc-Gumbel distribution, and the difference is likely to be enlarged with increasing F_{S0}. Therefore, calibrating the population-level distribution is necessary to avoid biased liquefaction potential assessment.

Using the best-fit population-level distribution, Fig. 8b presents the ratios of P_{L0} for three parameter configurations to the conventional configuration [2Hyp, Q_p=0.456]. The ratio values are observed to be as large as 10 at $F_{S0} \geq 2$. Thus, the influence of parameter configuration is more significant than that of the population-level distribution. The HBM-4Hyp generally results in larger P_{L0} than HBM-2Hyp, and adopting dataset-specific Q_p tends to yield slightly larger P_{L0} compared to Q_p=0.456. Thus, the [2Hyp, Q_p=0.456] configuration can be unconservative for statistical description of the inter- and intra-region variabilities.

5. Implication of region-variability for liquefaction-triggering hazard assessment

The region-specific SPT-IB12 and CPT-RW98 models are applied to Yuanlin (YL-B-3 profile) and Zhangbin Industrial Park (LW-A10 profile), respectively, which are located in Taiwan and were influenced by the 1999 Chi-Chi earthquake. It should be noted that the two profiles are not included in our HBM-based model updating. The in-situ geotechnical investigations are obtained from the database compiled by Dr. C. H. Juang (https://cecas.clemson.edu/chichi/TW-LIQ/In-situ-Test.htm). For the YL-B-3 profile, the corrected SPT blow counts $(N_1)_{60}$ were obtained from the raw SPT blow counts using the model of Juang et al. (2005). The fines content (FC) and $(N_1)_{60}$ of this profile are displayed in Fig. 9a. Also, the measured cone tip penetration resistance (q_c) and sleeve friction (f_s) of the LW-A10 profile are shown in Fig. 9b.

5.1. Scenario-based liquefaction potential assessment

The deterministic SPT-BI12 and CPT-RW98 procedures are adopted to estimate F_{S0} considering earthquake magnitude (M) of 7.6, where the recorded $a_{\max}$ values are 0.18 g and 0.12 g for the YL-B-3 and LW-A10 profiles, respectively (Juang et al. 2005; Ku et al. 2012). The updated region-specific models are then used to estimate P_{L0} using Eq. (29). The results of F_{S0} and P_{L0} are included in Fig. 9. The largest liquefaction probabilities in the YL-B-3 and LW-A10 profiles are observed at the depths of about 5 m and 10.5 m, respectively, and are both larger than 0.5. The results are in accordance with the actual occurrence of liquefaction at the two sites. Specifically, HBM-4Hyp predicts higher P_{L0} than HBM-2Hyp.

5.2. Probabilistic method for liquefaction-triggering hazard assessment

The performance-based earthquake engineering design (e.g., Juang et al. 2010; Rathje et al. 2014) usually requires an estimate of the total probability of liquefaction subjected to all possible ground motions in an exposure period (e.g., 50 years). Therefore, a discrete integration-based probabilistic method is suggested herein to assess liquefaction-triggering hazard accounting for multiple ground motion levels.

This method convolves ground motion hazard with a liquefaction-triggering procedure, following the idea of probabilistic seismic demand analysis (e.g., Wang et al. 2021). The mean annual rate of liquefaction (λ_L) is calculated as:

$$\lambda_L = \sum_i \sum_j \sum_k \sum_l P\left(F_S < 1 \middle| M_i, a_{\max,j}, r_{d,k}, MSF_l\right) P\left(MSF_l \middle| M_i, a_{\max,j}\right) P\left(r_{d,k} \middle| M_i, a_{\max,j}\right) P\left(M_i, a_{\max,j}\right) \quad (30)$$

where i, j, k, and l identify the discretized bins of M, $a_{\max}$, r_d, and MSF, respectively; $P(F_S < 1|\cdot)$ is calculated from P_{L0} model through Eq. (29); $P(M_i, a_{\max,j})$ is the joint annual probability

of occurrence of $M = M_i$ and $a_{\max} = a_{\max,j}$, and is obtained as:

$$P\left(M_i, a_{\max,j}\right) = \sum_m P\left(M_i, R_m \left| a_{\max,j}\right.\right) P\left(a_{\max,j}\right) \tag{31}$$

where R_m represents the m-th bin of source-to-site distance R; $P(M_i, R_m|a_{\max,j})$ is the joint annual probability of $M = M_i$ and $R = R_m$ given $a_{\max} = a_{\max,j}$, and can be computed from the disaggregation of seismic hazard (Baker et al. 2021); $P(a_{\max,j})$ represents the annual probability of $a_{\max} = a_{\max,j}$, which can be derived as (Rathje et al. 2014):

$$P\left(a_{\max,j}\right) = \frac{\lambda\left(a_{\max,j-1}\right) - \lambda\left(a_{\max,j+1}\right)}{2} \tag{32}$$

where $\lambda(a_{\max,j\text{-}1})$ and $\lambda(a_{\max,j+1})$ denote the mean annual rates of $a_{\max} > a_{\max,j\text{-}1}$ and $a_{\max} > a_{\max,j+1}$, respectively.

The $P(r_{d,k}|M_i, a_{\max,j})$ term in Eq. (30) represents the occurrence probability of $r_d = r_{d,k}$ conditioned on $M = M_i$ and $a_{\max} = a_{\max,j}$, and is statistically described to account for the uncertainty in r_d estimation. In this study, r_d is assumed to follow a normal distribution with the mean of μ_{rd} and standard deviation of σ_{rd} (Cetin and Seed 2004), which are often provided by empirical models (e.g., Moss et al. 2006; Lasley et al. 2016). Given the uncertainty in *MSF* estimation as well (e.g., Green et al. 2020), ln*MSF* is also assumed to be a normal variable with the mean of $\mu_{\ln MSF}$ and standard deviation of $\sigma_{\ln MSF}$. Specifically, μ_{rd} and $\mu_{\ln MSF}$ denote the deterministic estimates of r_d and logarithmic *MSF*, respectively (see Supplementary material). Both σ_{rd} and $\sigma_{\ln MSF}$ are assumed to be 0.15 (Lasley et al. 2016; Green et al. 2020), unless stated otherwise. Sensitivity analyses of the two parameters will be separately discussed later.

Under the assumption of Poisson's process (e.g., Kramer and Mayfield 2007; Wang et al. 2022), λ_L can be used to estimate the liquefaction probability ($P_{L,T}$) in an exposure period T:

$$P_{L,T} = 1 - \exp\left(-\lambda_L T\right) \tag{33}$$

In other words, $P_{L,T}$ represents the total probability of liquefaction accounting for the uncertain occurrence of earthquake during a target period.

5.3. Liquefaction-triggering hazard analysis results

For the purpose of illustrating the proposed models, the soil profiles are assumed to be located at a distance of 10 km from the center of a 200 km-long strike-slip linear fault. The seismicity of the hypothetical source is described by a bounded Gutenberg-Richter relationship:

$$\log_{10}\left(\lambda_M\right) = 4.3 - M \tag{34}$$

where λ_M is the mean annual rate that earthquake magnitude exceeds M. The central values of discretized M bins range from 5 to 8 with increments of 0.1. For each bin of M, earthquake rupture occurs uniformly along the fault with the rupture length L as a function of M (Wells and Coppersmith 1994). The source-site relationship is schematically illustrated in Fig. 10.

The seismic hazard for $a_{\max}$ is derived using the ground-motion model proposed by Campbell and Bozorgnia (2014). Since the shear-wave velocities of the YL-B-3 and LW-A10 profiles are not available, the average upper 30 m shear-wave velocity (Vs_{30}) required by the ground-motion model is estimated as follows: (1) the raw SPT-N values are transformed into Vs values using the empirical model developed by Xiao et al. (2021); (2) the depth-dependent Vs values are used to estimate the average upper 15 m shear-wave velocity (Vs_{15}), which is then used in the Vs_{30}-Vs_{15} relationship established by Boore et al. (2011).

First, the profiles are assumed to be in regions influenced by the 1999 Chi-Chi earthquake. Fig. 11a displays λ_L at different depths and the variation of $P_{L,T}$ at the depth of 5.8 m with exposure period for the YL-B-3 profile. Similarly, the results for the LW-A10 profile are presented in Fig. 11b. Regarding the [2Hyp, Q_p=0.456] configuration, largest λ_L values of

around 0.0055 and 0.0186 are observed at the 5.8 m and 10.5 m depths in the YL-B-3 and LW-A10 profiles, respectively. In contrast, the 4Hyp modeling results in considerably larger λ_L and $P_{L,T}$ because of the larger uncertainty considered in the model. Meanwhile, λ_L and $P_{L,T}$ are higher for the larger calculated Q_p, though the influence on the SPT-BI12 models (for which the calculated Q_p is comparable with 0.456) is small. As also shown in the figures, $P_{L,T}$ increases with increasing exposure period; thus, a soil deposit with $P_{L,T} < 0.5$ for relatively short T can still be likely to liquefy.

To compare the effects of choosing region-specific models, the profiles are then assumed to be located in different regions, subjected to the same earthquake source, path, and site conditions. This ensures that the difference of results only originates from the liquefaction-triggering models. Fig. 12a and b display the results of λ_L from different region-specific SPT-BI12 models (for the depth of 5.8 m in YL-B-3 profile) and CPT-RW98 models (for the depth of 10.5 m in LW-A10 profile), respectively. Besides, the lumped-region modeling is also included for comparison. It is seen that the λ_L values for different regions are notably different, and can considerably differ from the results for the lumped-region modeling.

Furthermore, the effects of region features on the liquefaction-triggering hazard analysis is studied by quantifying the ratio of $P_{L,T}$ between region-specific and lumped-region models, namely [$P_{L,T}$(region,para)/$P_{L,T}$(lumped,Q_p=0.456)]. Fig. 13a and b present the boxplots of the $P_{L,T}$ ratios in different regions for the SPT- and CPT-based models, respectively, in which three exposure periods and four parameter configurations are considered. Except for the [CPT-2Hyp, Q_p=0.456] modeling, the $P_{L,T}$ ratio values are found to be generally larger than 1, and exceed 1.3 in a few cases. Thus, the lumped-region modeling could underestimate the liquefaction

probability by more than 30%. The distribution of the $P_{L,T}$ ratio for the SPT-BI12 models appears to be insensitive to the exposure period, while the relative difference between the region-specific and lumped-region models is larger for shorter exposure period. In particular, the relative difference is more evident for HBM-4Hyp as compared to HBM-2Hyp. The above results thus indicate the importance of refined region-feature representation in liquefaction-triggering hazard assessment.

5.4. Effects of σ_{rd} and $\sigma_{\ln MSF}$ on liquefaction-triggering hazard analysis

As mentioned previously, σ_{rd} and $\sigma_{\ln MSF}$ are greater than 0, due to the uncertainty in predictions of r_d and MSF (e.g., from ground-motion variability and parametric fitting error). The σ_{rd} versus depth trends reported in previous studies (Cetin and Seed 2004; Moss et al. 2006; Lasley et al. 2016; Green et al. 2020) are plotted in Fig. 14a. It is seen that σ_{rd} increases with depth, reaching around 0.15-0.2 at depth larger than 15 m. In contrast, $\sigma_{\ln MSF}$ has been reported to be about 0.15 (Green et al. 2020). To investigate the effects of σ_{rd} and $\sigma_{\ln MSF}$, the $P_{L,50}$ values along the LW-A10 profile for different σ_{rd} and $\sigma_{\ln MSF}$ values (0, 0.1, 0.2, and 0.5) are displayed in Fig. 14b and c, respectively. The results are derived by the [CPT-2Hyp, Q_p=0.456] model for the 1999 Chi-Chi earthquake. It is observed that the effects of standard deviations on $P_{L,T}$ are negligibly small when σ_{rd} and $\sigma_{\ln MSF}$ are not larger than 0.2, yet become relatively large when increasing σ_{rd} and $\sigma_{\ln MSF}$ to 0.5.

6. Discussion

To examine the adequacy of the data size used to calibrate models, an additional set of CPT-4Hyp models is developed by increasing the current minimum data size for each region from

561 4 to 8. After removing two regions (i.e., 1983 Borah Peak and 2011 Tohoku) with less than 8
562 case histories, 8 regions are retained for the model re-development. The resulting P_{L0}-F_{S0}
563 relationships are compared with those derived from the original database in Fig. 15a. The
564 results for the two minimum data sizes (i.e., two datasets) are observed to be comparable,
565 suggesting that the data size considered is generally adequate for the HBM-4Hyp calibration.
566 Regarding the SPT-BI12 procedure, Fig. 15b compares Std(c) versus R_{FT} trends for four
567 Bayesian modeling choices, where R_{FT} is defined as the ratio of false to true predictions (i.e.,
568 (FP+FN)/(TP+TN) in Confusion Matrix) from a deterministic liquefaction-triggering
569 procedure and data for each region. It can be seen that the Std(c) values generally increase with
570 increasing R_{FT}, implying that the intra-region variability of bias factor can be roughly measured
571 by R_{FT}. Additionally, HBM-4Hyp results in much smaller Std(c) than the independent-regions
572 modeling. Thus, HBM-4Hyp differs significantly from the independent-regions modeling that
573 fully neglects information borrowing across regions. In contrast, HBM-2Hyp leads to a
574 negligible regional variation of Std(c), which is similar to the lumped-regions modeling with a
575 common Std(c). The comparative results demonstrate the capability of HBM-4Hyp to capture
576 the inter-region variability of bias variance.
577 In addition to the method suggested, another major contribution of this study is to
578 demonstrate that there can be large region-to-region differences in the standard deviations of
579 model bias factors and HBM-2Hyp with a common variance could underestimate liquefaction
580 hazard. In this case, HBM-4Hyp is advantageous over HBM-2Hyp for characterizing the
581 region-specific statistics. Yet, a notable challenge in implementing HBM-4Hyp is the increased
582 data size requirement for calibration. Therefore, it is recommended that practitioners should

583 assess the plausibility of region-invariant variance and the adequacy of data size before making
584 a choice between HBM-4Hyp and HBM-2Hyp, both of which are considered in this study. This
585 work represents a step towards refining the representation of the inter-region variability of
586 model bias. However, further efforts are necessary to continue advancing research in the
587 regionalization of liquefaction-triggering models (e.g., Zhang et al. 2016, 2020; Green 2022),
588 taking into account the continuously evolving global and regional liquefaction databases (e.g.,
589 Brandenberg et al. 2020; Geyin et al. 2021).

590 **7. Conclusions**

591 This paper presented a Bayesian framework to incorporate the inter-region and intra-region
592 variabilities of model bias into liquefaction-triggering procedures. The BUS approach with SuS
593 is utilized for posterior sampling and the Gaussian copula-based Bayesian evidence is
594 quantified for model selection. A new four-hyperparameter hierarchical Bayesian model
595 (HBM-4Hyp) was suggested, which assumes the means (μ_{ci}) or standard deviations (σ_{ci}) of the
596 bias factor (c_i) are region-specific, but are controlled by a population-level distribution. This
597 model was systematically compared with HBM-2Hyp (with region-invariant σ_{ci}). Meanwhile,
598 three candidate distributions were adopted for the population-level statistical modeling.
599 Scenario-based and probabilistic liquefaction-triggering hazard analyses were conducted to
600 investigate the implication of region-feature modeling. The main findings drawn from this
601 study are summarized as follows:

602 (1) The HBM-4Hyp requires around twice the number of parameters for HBM-2Hyp,
603 leading to a high-dimensional Bayesian updating problem. This can be efficiently addressed by
604 the BUS approach with SuS, and the benefits of such would be more notable with increasing

the number of regions compiled in future global databases. Meanwhile, the Bayesian evidence is shown to be effective for model comparison, and a HBM configuration with higher evidence is generally observed to better match liquefaction observations.

(2) The proposed Bayesian framework is utilized for incorporating the region-features into three liquefaction-triggering procedures (SPT-BI12, CPT-RW98, and Vs-R22), covering liquefaction-susceptible sandy and gravelly soil types. The coefficients of the updated region-specific probabilistic liquefaction models are provided for potential applications.

(3) There exist notable inter-region variabilities of μ_{ci} and σ_{ci} especially for the CPT- and Vs-based models. Compared to HBM-2Hyp, HBM-4Hyp tends to yield higher Bayesian evidence and larger total standard deviation of c_i. When the calculated Q_p is considerably different from the conventional value of 0.456, this dataset-specific Q_p is likely to enlarge Bayesian evidence, and the predicted liquefaction probabilities are found to increase with increasing Q_p.

(4) The lognormal distribution may not be optimal for the population-level statistical descriptions in HBM, as illustrated by the lower evidence compared to the other candidate distributions, and could considerably underestimate liquefaction potential in some cases.

(5) A discrete integration-based probabilistic method is suggested to implement the developed models for liquefaction-triggering hazard assessment considering multiple ground-motion levels. Neglecting the region-variability is likely to underestimate liquefaction hazard, and HBM-4Hyp generally yields higher liquefaction hazard than HBM-2Hyp.

Acknowledgment

The work presented in this paper is financially supported by the Research Grants Council of Hong Kong Special Administrative Region (Project No. 15222021).

Supplementary material

Supplementary data related to the mathematical expressions and liquefaction datasets for the three liquefaction-triggering procedures are available with the article. Other data generated or analyzed during this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare there are no competing interests.

References

Au, S. K., Beck, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics, 16(4), 263-277.

Baker, J., Bradley, B., Stafford, P. (2021). Seismic hazard and risk analysis. Cambridge University Press.

Betz, W., Beck, J. L., Papaioannou, I., Straub, D. (2018a). Bayesian inference with reliability methods without knowing the maximum of the likelihood function. Probabilistic Engineering Mechanics, 53, 14-22.

Betz, W., Papaioannou, I., Beck, J. L., Straub, D. (2018b). Bayesian inference with subset simulation: strategies and improvements. Computer Methods in Applied Mechanics and Engineering, 331, 72-93.

Boore, D. M., Thompson, E. M., Cadet, H. (2011). Regional correlations of VS 30 and velocities averaged over depths less than and greater than 30 meters. Bulletin of the Seismological Society of America, 101(6), 3046-3059.

Boulanger, R. W., Idriss, I. M. (2012). Probabilistic standard penetration test–based liquefaction-triggering procedure. Journal of Geotechnical and Geoenvironmental

649 Engineering, 138(10), 1185-1195.
650 Boulanger, R. W., Idriss, I. M. (2014). CPT and SPT based liquefaction triggering
651 procedures. Report No. UCD/CGM.-14, 1.
652 Bozorgzadeh, N., Harrison, J. P., Escobar, M. D. (2019). Hierarchical Bayesian modelling of
653 geotechnical data: application to rock strength. Géotechnique, 69(12), 1056-1070.
654 Brandenberg, S. J., Zimmaro, P., Stewart, J. P., et al. (2020). Next-generation liquefaction
655 database. Earthquake Spectra, 36(2), 939-959.
656 Campbell, K. W., Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average
657 horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra.
658 Earthquake Spectra, 30(3), 1087-1115.
659 Cao, Z., Wang, Y. (2014). Bayesian model comparison and selection of spatial correlation
660 functions for soil parameters. Structural Safety, 49, 10-17.
661 Cao, Z. J., Zheng, S., Li, D. Q., Phoon, K. K. (2019). Bayesian identification of soil stratigraphy
662 based on soil behaviour type index. Canadian Geotechnical Journal, 56(4), 570-586.
663 Cetin, K. O., Der Kiureghian, A., Seed, R. B. (2002). Probabilistic models for the initiation of
664 seismic soil liquefaction. Structural safety, 24(1), 67-82.
665 Cetin, K. O., Seed, R. B. (2004). Nonlinear shear mass participation factor (rd) for cyclic shear
666 stress ratio evaluation. Soil Dynamics and Earthquake Engineering, 24(2), 103-113.
667 Ching, J., Phoon, K. K., Ho, Y. H., Weng, M. C. (2021a). Quasi-site-specific prediction for
668 deformation modulus of rock mass. Canadian Geotechnical Journal, 58(7), 936-951.
669 Ching, J., Wu, S., Phoon, K. K. (2021b). Constructing quasi-site-specific multivariate
670 probability distribution using hierarchical Bayesian model. Journal of Engineering

671 Mechanics, 147(10), 04021069.
672 Deng, Z. P., Pan, M., Niu, J. T., Jiang, S. H. (2022). Full probability design of soil slopes
673 considering both stratigraphic uncertainty and spatial variability of soil
674 properties. Bulletin of Engineering Geology and the Environment, 81(5), 195.
675 Duan, W., Zhao, Z., Cai, G., Pu, S., Liu, S., Dong, X. (2022). Evaluating model uncertainty of
676 an in situ state parameter-based simplified method for reliability analysis of liquefaction
677 potential. Computers and Geotechnics, 151, 104957.
678 Geyin, M., Maurer, B. W., Bradley, B. A., Green, R. A., van Ballegooy, S. (2021). CPT-based
679 liquefaction case histories compiled from three earthquakes in Canterbury, New Zealand.
680 Earthquake Spectra, 37(4), 2920-2945.
681 Green, R. A. (2022). Regionalization of Liquefaction Triggering Models. In Conference on
682 Performance-based Design in Earthquake. Geotechnical Engineering (pp. 437-451).
683 Cham: Springer International Publishing.
684 Green, R. A., Bommer, J. J., Stafford, P. J., et al. (2020). Liquefaction hazard in the Groningen
685 region of the Netherlands due to induced seismicity. Journal of Geotechnical and
686 Geoenvironmental Engineering, 146(8), 04020068.
687 Guan, Z., Wang, Y. (2022). Quantifying reliability of liquefaction severity map developed from
688 sparse cone penetration tests. Canadian Geotechnical Journal, 60(5), 623-641.
689 Hong, Y. (2013). On computing the distribution function for the Poisson binomial distribution.
690 Computational Statistics & Data Analysis, 59, 41-51.
691 Hu, J., Liu, H. (2019). Bayesian network models for probabilistic evaluation of earthquake-
692 induced liquefaction based on CPT and Vs databases. Engineering Geology, 254, 76-88.

693 Jiang, S. H., Papaioannou, I., Straub, D. (2018). Bayesian updating of slope reliability in
694 spatially variable soils with in-situ measurements. Engineering Geology, 239, 310-320.
695 Juang, C. H., Jiang, T., Andrus, R. D. (2002). Assessing probability-based methods for
696 liquefaction potential evaluation. Journal of Geotechnical and Geoenvironmental
697 Engineering, 128(7), 580-589.
698 Juang, C. H., Ching, J., Luo, Z., Ku, C. S. (2012). New models for probability of liquefaction
699 using standard penetration tests based on an updated database of case histories.
700 Engineering Geology, 133, 85-93.
701 Juang, C. H., Ou, C. Y., Lu, C. C., Luo, Z. (2010). Probabilistic framework for assessing
702 liquefaction hazard at a given site in a specified exposure time using standard penetration
703 testing. Canadian Geotechnical Journal, 47(6), 674-687.
704 Juang, C. H., Yang, S. H., Yuan, H., Fang, S. Y. (2005). Liquefaction in the Chi-Chi earthquake-
705 effect of fines and capping non-liquefiable layers. Soils and foundations, 45(6), 89-101.
706 Kramer, S. L., Mayfield, R. T. (2007). Return period of soil liquefaction. Journal of
707 Geotechnical and Geoenvironmental Engineering, 133(7), 802-813.
708 Ku, C. S., Juang, C. H., Chang, C. W., Ching, J. (2012). Probabilistic version of the Robertson
709 and Wride method for liquefaction evaluation: development and application. Canadian
710 Geotechnical Journal, 49(1), 27-44.
711 Lasley, S. J., Green, R. A., Rodriguez-Marek, A. (2016). New stress reduction coefficient
712 relationship for liquefaction triggering analyses. Journal of Geotechnical and
713 Geoenvironmental Engineering, 142(11), 06016013.
714 Liu, W. F., Leung, Y. F. (2018). Spatial variability of saprolitic soil properties and relationship

715 with joint set orientation of parent rock: Insights from cases in Hong Kong. Engineering
716 Geology, 246, 36-44.
717 Lo, M. K., Leung, Y. F. (2019). Bayesian updating of subsurface spatial variability for improved
718 prediction of braced excavation response. Canadian Geotechnical Journal, 56(8), 1169-
719 1183.
720 Maurer, B. W., Green, R. A., Van Ballegooy, S., Wotherspoon, L. (2019). Development of
721 region-specific soil behavior type index correlations for evaluating liquefaction hazard in
722 Christchurch, New Zealand. Soil Dynamics and Earthquake Engineering, 117, 96-105.
723 Moss, R. E., Seed, R. B., Kayen, R. E., Stewart, J. P., Der Kiureghian, A., Cetin, K. O. (2006).
724 CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction
725 potential. Journal of Geotechnical and Geoenvironmental Engineering, 132(8), 1032-1051.
726 Papaioannou, I., Betz, W., Zwirglmaier, K., Straub, D. (2015). MCMC algorithms for subset
727 simulation. Probabilistic Engineering Mechanics, 41, 89-103.
728 Rathje, E. M., Wang, Y., Stafford, P. J., Antonakos, G., Saygili, G. (2014). Probabilistic
729 assessment of the seismic performance of earth slopes. Bulletin of Earthquake
730 Engineering, 12, 1071-1090.
731 Robertson, P. K. (2022). Evaluation of flow liquefaction and liquefied strength using the cone
732 penetration test: An update. Canadian Geotechnical Journal, 59(4), 620-624.
733 Robertson, P. K., Wride, C. E. (1998). Evaluating cyclic liquefaction potential using the cone
734 penetration test. Canadian Geotechnical Journal, 35(3), 442-459.
735 Rollins, K. M., Roy, J., Athanasopoulos-Zekkos, A., et al. (2022). A new Vs-based liquefaction-
736 triggering procedure for gravelly soils. Journal of Geotechnical and Geoenvironmental

737 Engineering, 148(6), 04022040.
738 Seed, H. B., Idriss, I. M. (1971). Simplified procedure for evaluating soil liquefaction potential.
739 Journal of the Soil Mechanics and Foundations Division, 97(9), 1249-1273.
740 Straub, D., Papaioannou, I. (2015). Bayesian updating with structural reliability methods.
741 Journal of Engineering Mechanics, 141(3), 04014134.
742 Tabarroki, M., Ching, J., Yuan, S. H., Phoon, K. K., Teng, F. (2024). Data-driven hierarchical
743 Bayesian model for predicting wall deflections in deep excavations in clay. Computers
744 and Geotechnics, 168, 106135.
745 Tang, X. S., Wang, M. X., Li, D. Q. (2020). Modeling multivariate cross-correlated
746 geotechnical random fields using vine copulas for slope reliability analysis. Computers
747 and Geotechnics, 127, 103784.
748 Tao, Y., Phoon, K. K., Sun, H., Cai, Y. (2023). Hierarchical Bayesian model for predicting
749 small-strain stiffness of sand. Canadian Geotechnical Journal. https://doi.org/10.1139/cgj-
750 2022-0598.
751 Tian, M., Li, D. Q., Cao, Z. J., Phoon, K. K., Wang, Y. (2016). Bayesian identification of
752 random field model using indirect test data. Engineering Geology, 210, 197-211.
753 Wang, M. X., Tang, X. S., Li, D. Q., Qi, X. H. (2020). Subset simulation for efficient slope
754 reliability analysis involving copula-based cross-correlated random fields. Computers and
755 Geotechnics, 118, 103326.
756 Wang, M. X., Li, D. Q., Du, W. (2021). Probabilistic seismic displacement hazard assessment
757 of earth slopes incorporating spatially random soil parameters. Journal of Geotechnical
758 and Geoenvironmental Engineering, 147(11), 04021119.

759 Wang, M. X., Li, D. Q., Liu, Y., Du, W. (2022). Probabilistic decoupled approach to estimate
760 seismic rotational displacements of flexible slopes considering depth-dependent soil
761 variability. Acta Geotechnica, 17(4), 1551-1567.
762 Wells, D. L., Coppersmith, K. J. (1994). New empirical relationships among magnitude,
763 rupture length, rupture width, rupture area, and surface displacement. Bulletin of the
764 seismological Society of America, 84(4), 974-1002.
765 Wu, S., Ching, J., Phoon, K. K. (2022). Quasi-site-specific soil property prediction using a
766 cluster-based hierarchical Bayesian model. Structural Safety, 99, 102253.
767 Xiao, S., Zhang, J., Ye, J., Zheng, J. (2021). Establishing region-specific N-Vs relationships
768 through hierarchical Bayesian modeling. Engineering Geology, 287, 106105.
769 Youd, T. L., Idriss, I. M., Andrus, R. D., et al. (2001). Liquefaction resistance of soils: summary
770 report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of
771 liquefaction resistance of soils. Journal of Geotechnical and Geoenvironmental
772 Engineering, 127(4), 297-313.
773 Yuen, K.V. (2010). Bayesian Methods for Structural Dynamics and Civil Engineering. John
774 Wiley & Sons (Asia) Pte Ltd, Singapore.
775 Zhang, J., Li, J. P., Zhang, L. M., Huang, H. W. (2014). Calibrating cross-site variability for
776 reliability-based design of pile foundations. Computers and Geotechnics, 62, 154-163.
777 Zhang, J., Juang, C. H., Martin, J. R., Huang, H. W. (2016). Inter-region variability of
778 Robertson and Wride method for liquefaction hazard analysis. Engineering Geology, 203,
779 191-203.
780 Zhang, J., Xiao, S., Huang, H., Zhou, J. (2020). Calibrating a standard penetration test based

method for region-specific liquefaction potential assessment. Bulletin of Engineering Geology and the Environment, 79, 5185-5204.

Zhang, J., Yang, S., Zhang, L. L., Zhou, M. L. (2022). Bayesian estimation of soil-water characteristic curves. Canadian Geotechnical Journal, 59(4), 569-582.

List of Figures

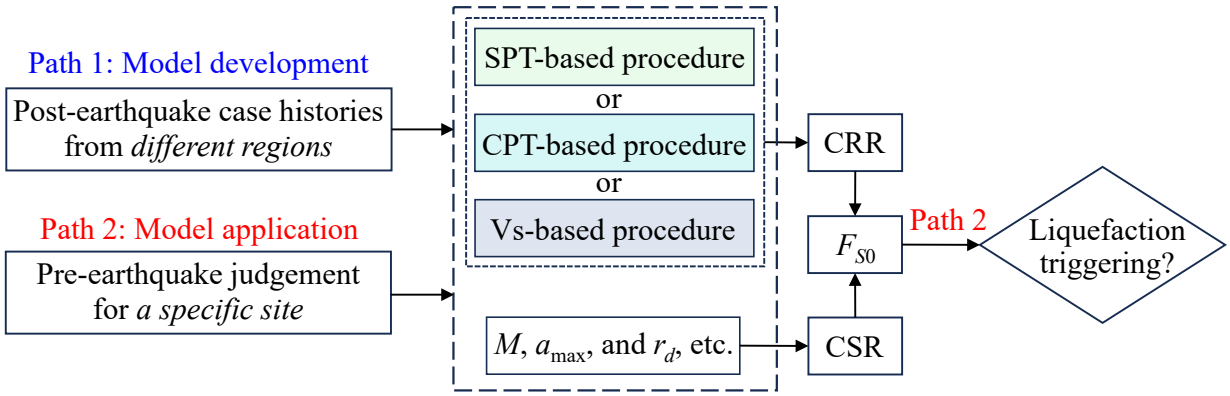


Fig. 1. Introduction to the development and application of conventional liquefaction-triggering assessment procedures.

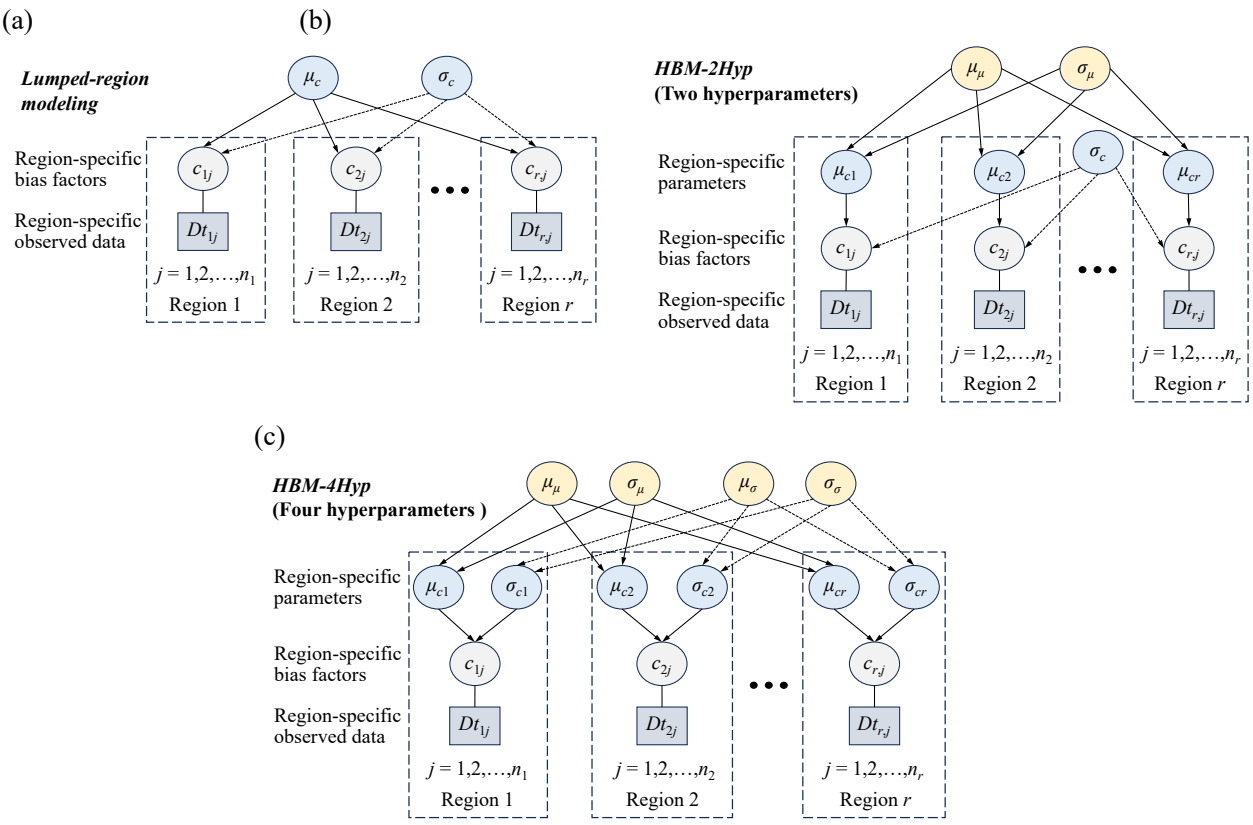


Fig. 2. Schematic illustrations of (a) lumped-region modeling, (b) HBM-2Hyp, and (c) HBM-4Hyp for updating liquefaction-triggering procedures.

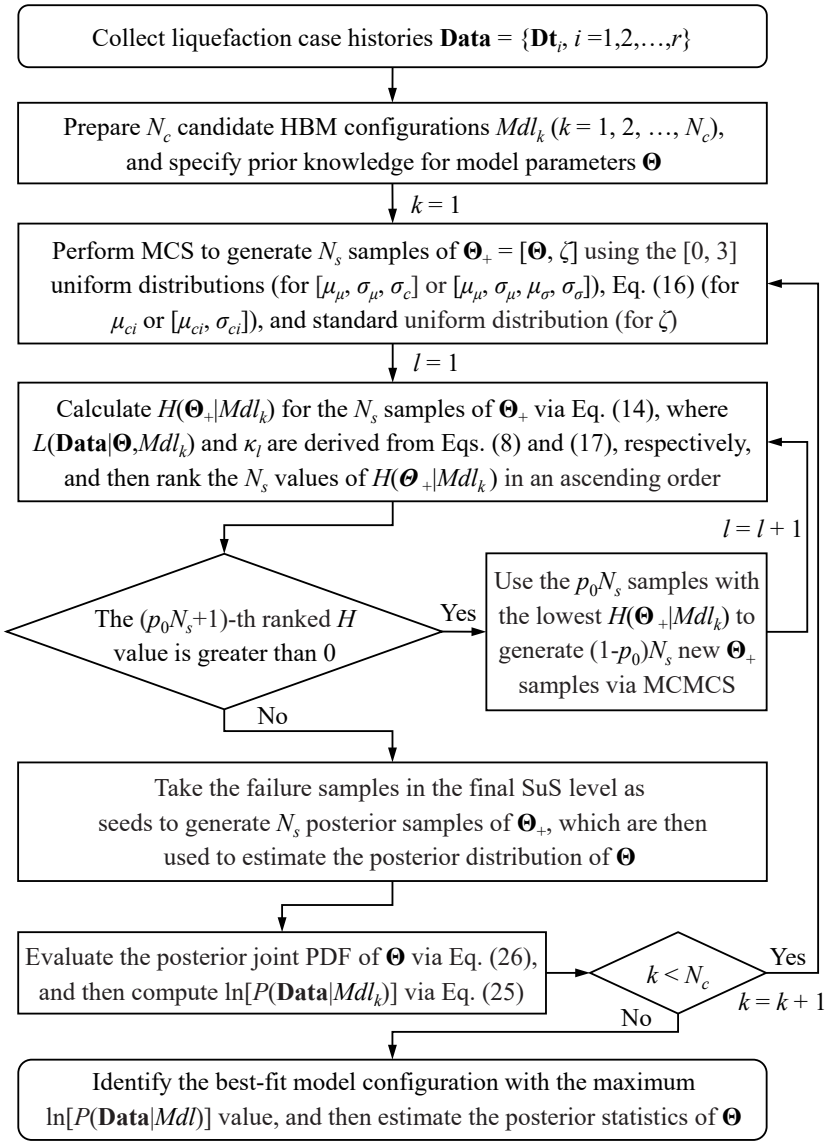


Fig. 3. Flowchart of the BUS-powered hierarchical Bayesian updating framework.

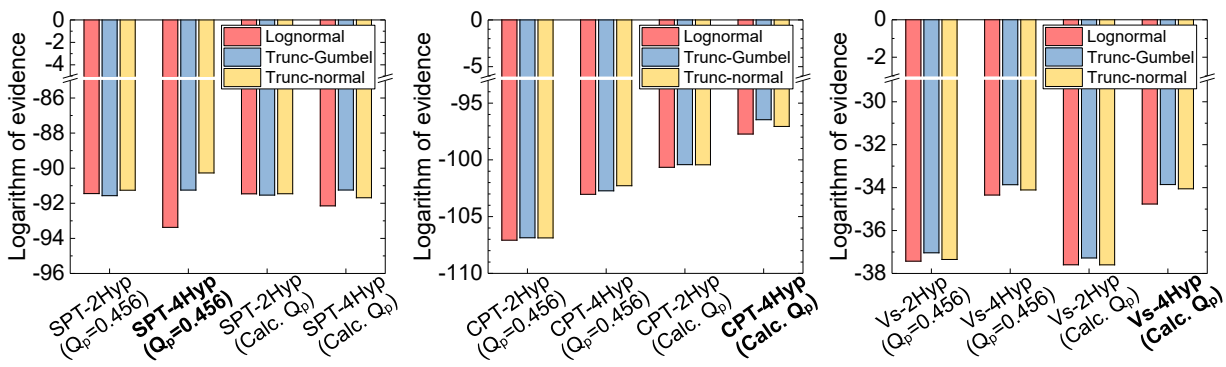


Fig. 4. Model evidence values for different HBM configurations of updated (a) SPT-BI12, (b) CPT-RW98, and (c) Vs-R22 models. The best-fit configuration is indicated as bold on the x-axis.

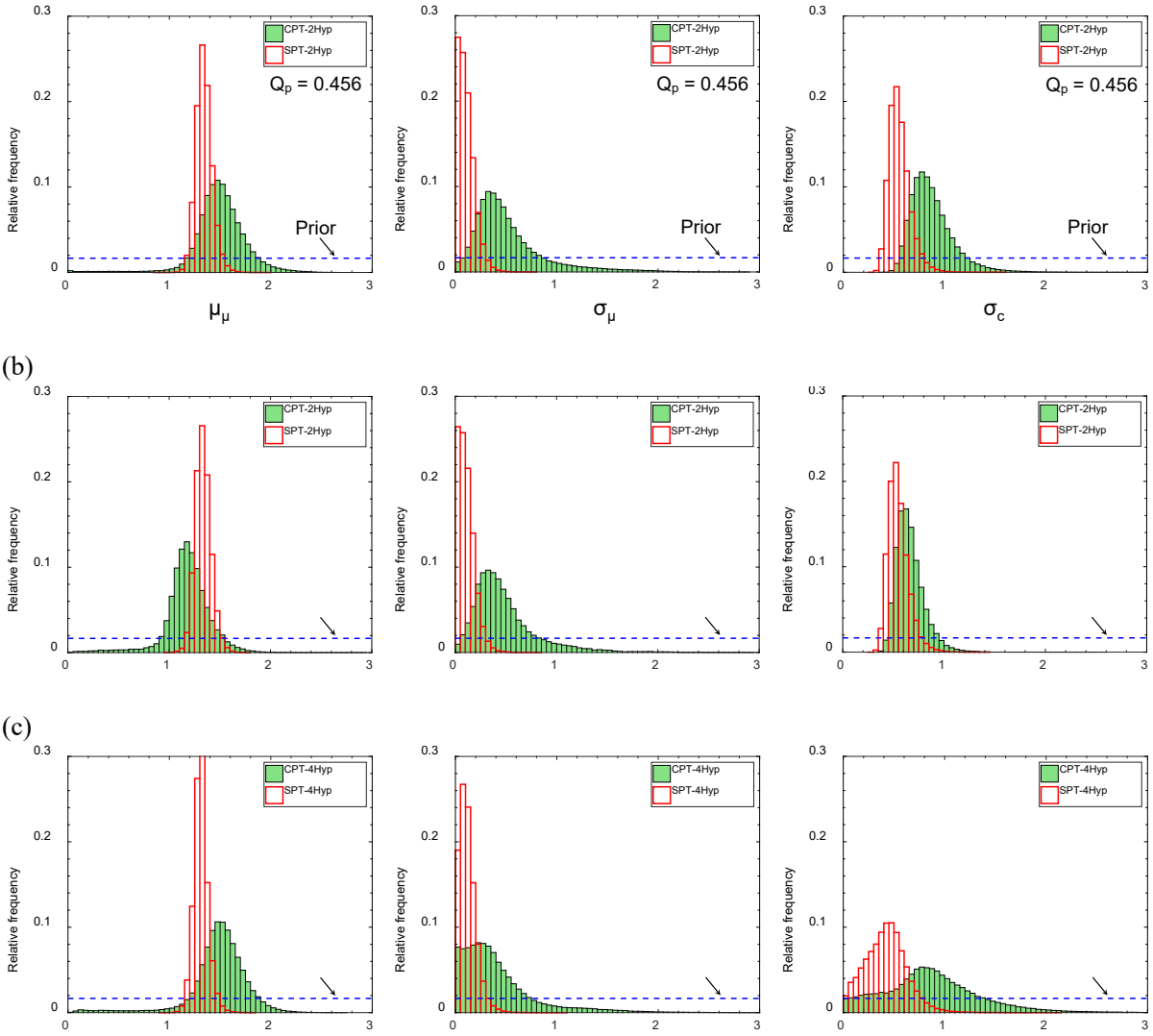
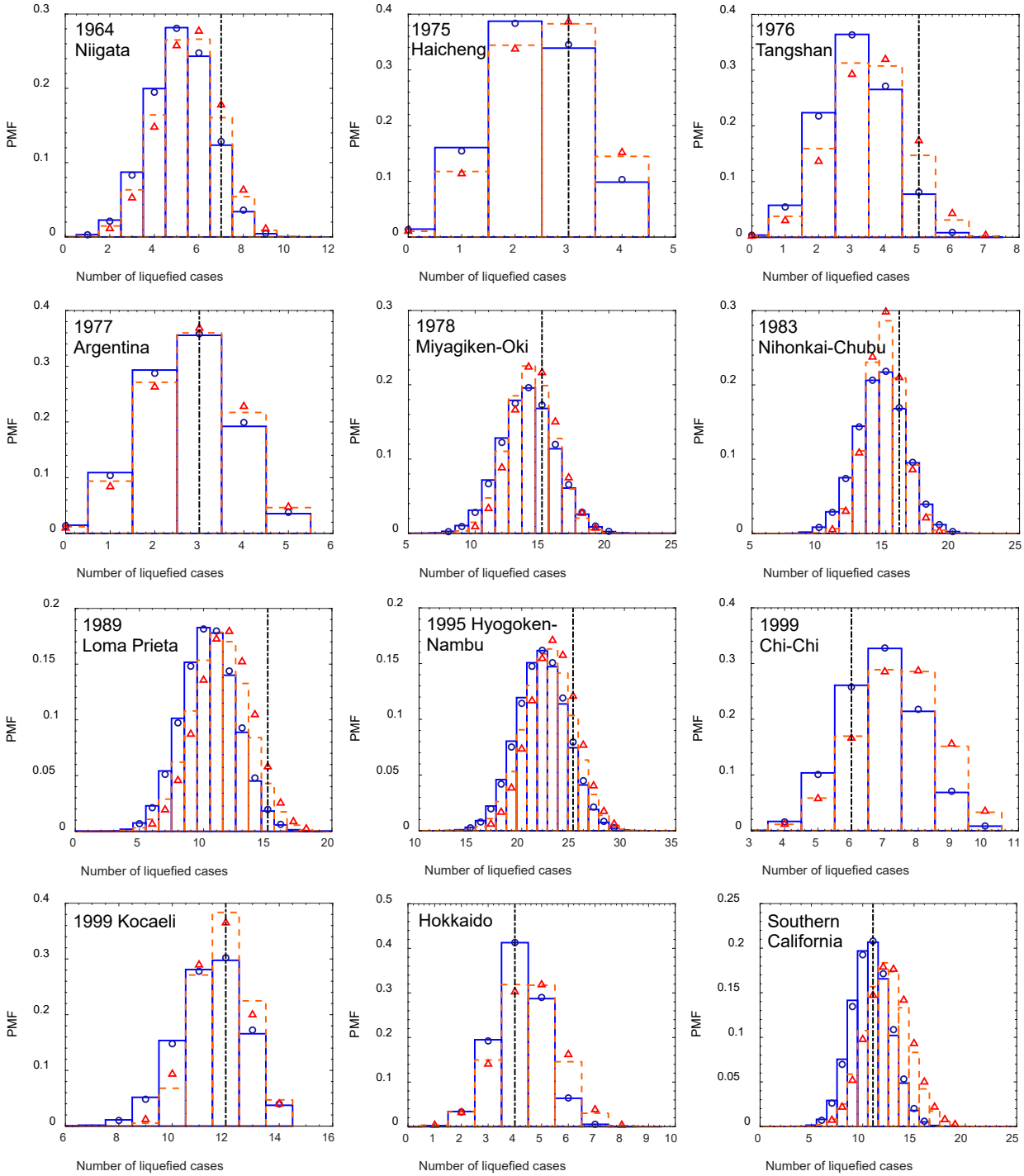


Fig. 5. Posterior sample distributions for (a) [μ_μ, σ_μ, σ_c] in CPT-2Hyp and SPT-2Hyp models (Q_p=0.456), (b) [μ_μ, σ_μ, σ_c] in CPT-2Hyp and SPT-2Hyp models (calculated Q_p), and (c) [μ_μ, σ_μ, μ_σ] in CPT-4Hyp and SPT-4Hyp models (Q_p=0.456). The prior relative frequency is derived considering the bin width of histogram.

815 (a)


1964
Niigata
1975
Haicheng
1976
Tangshan
1977
Argentina
1978
Miyagiken-Oki
1983
Nihonkai-Chubu
1989
Loma Prieta
1995 Hyogoken-
Nambu
1999
Chi-Chi
1999 Kocaeli
Hokkaido
Southern
California
PMF
Number of liquefied cases
SPT-2Hyp, Qp = 0.456
SPT-4Hyp, Qp = 0.456 (best-fit)
SPT-2Hyp, calculated Qp
SPT-4Hyp, calculated Qp
Actual number of liquefied cases

816

(b)

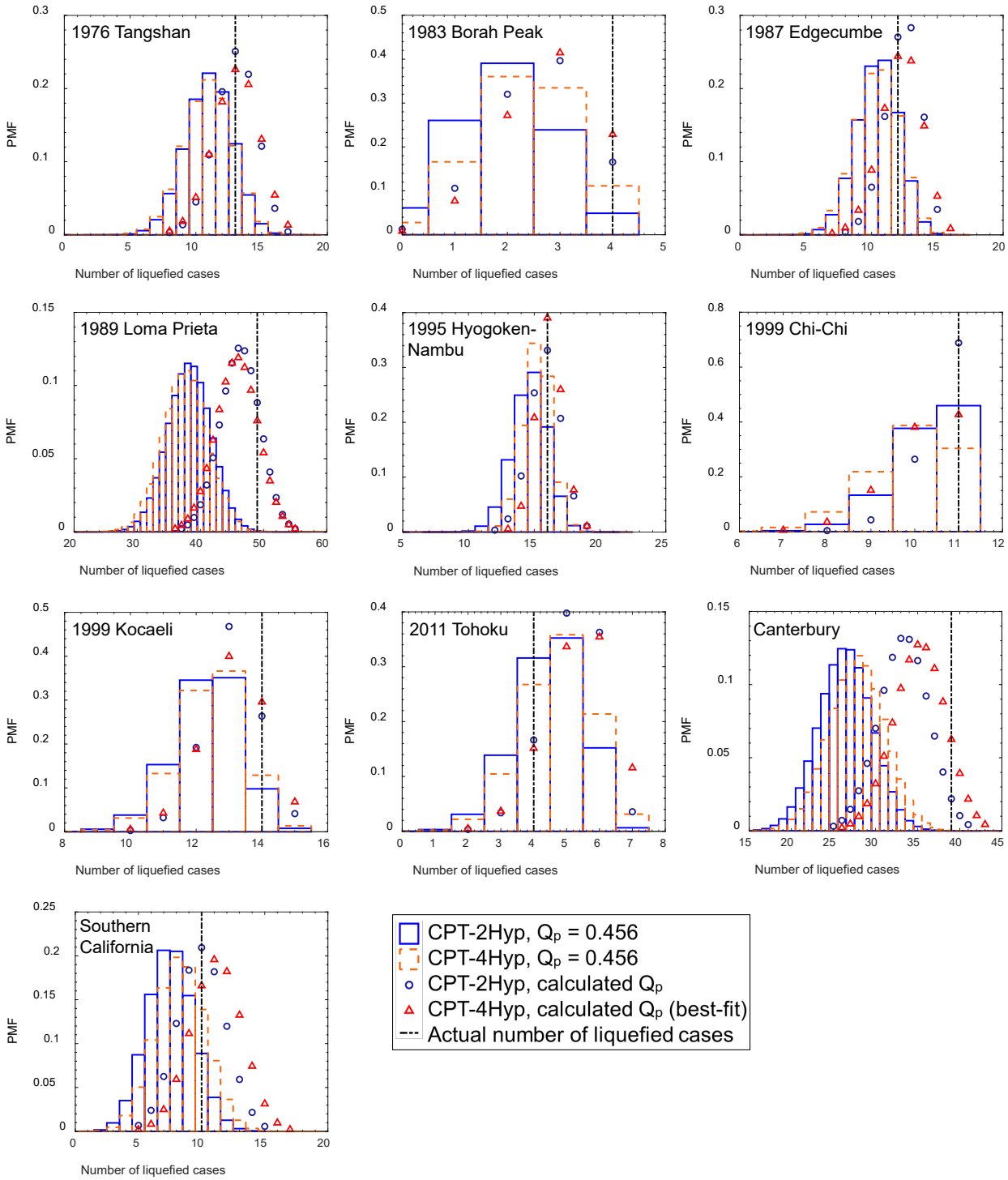


Fig. 6. Probability mass functions of predicted number of liquefied cases for region-specific (a) SPT-BI12 models and (b) CPT-RW98 models with four HBM configurations. The actual number of liquefied cases is also shown.

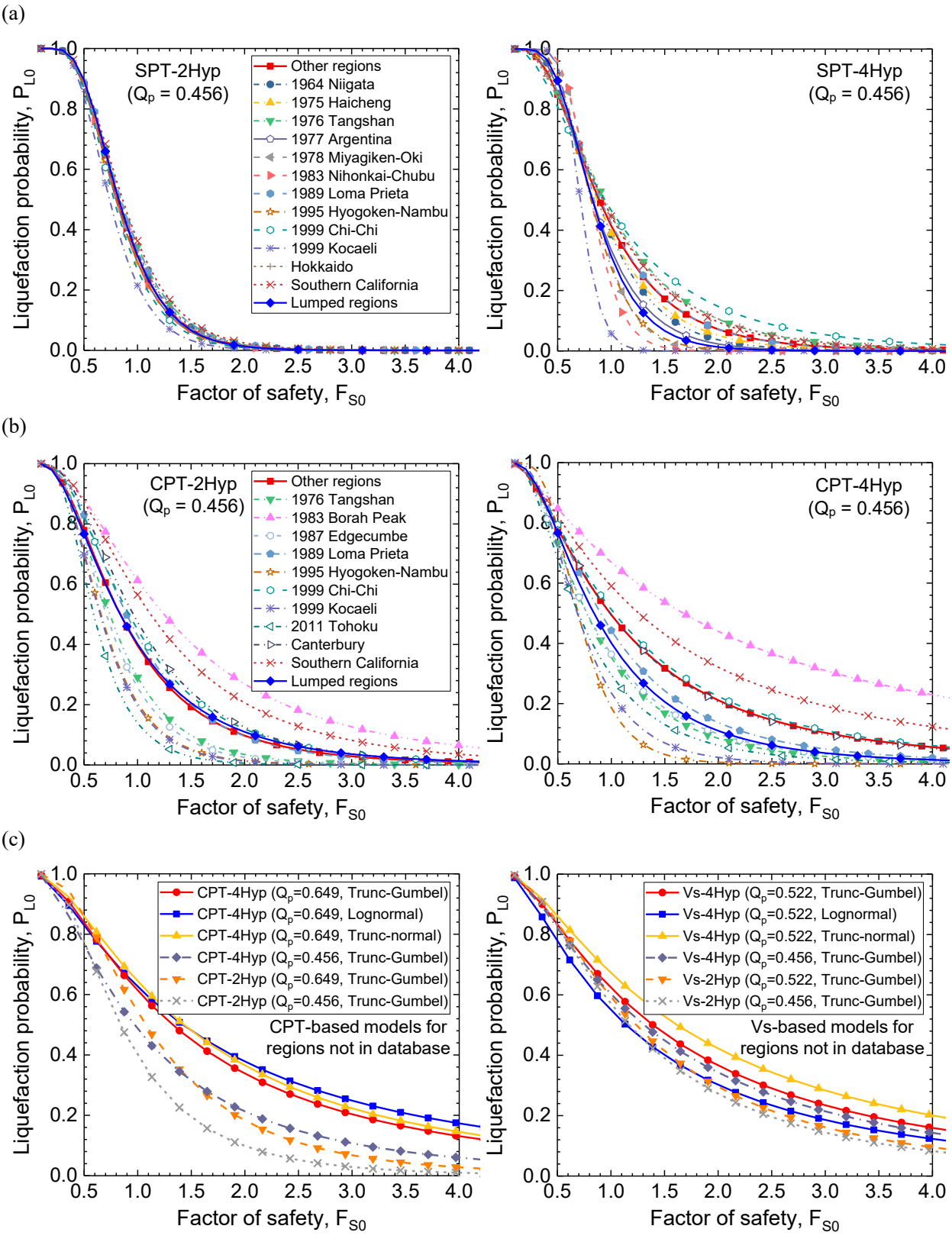
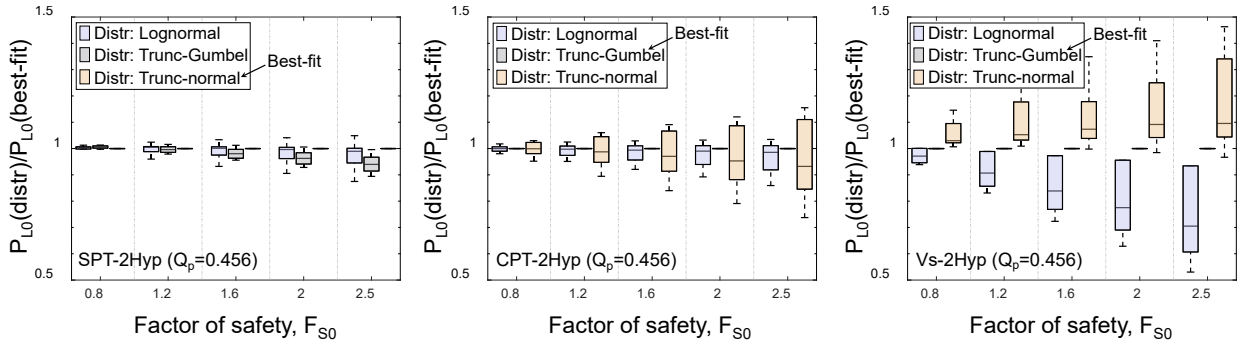


Fig. 7. Relationships between P_{L0} and F_{S0} for (a) region-specific SPT-BI12 models with Q_p=0.456 and best-fit distribution, (b) region-specific CPT-RW98 models with Q_p=0.456 and best-fit distribution, and (c) CPT-RW98 and Vs-R22 models with different HBM configurations for regions not in the calibration databases.

(a)



(b)

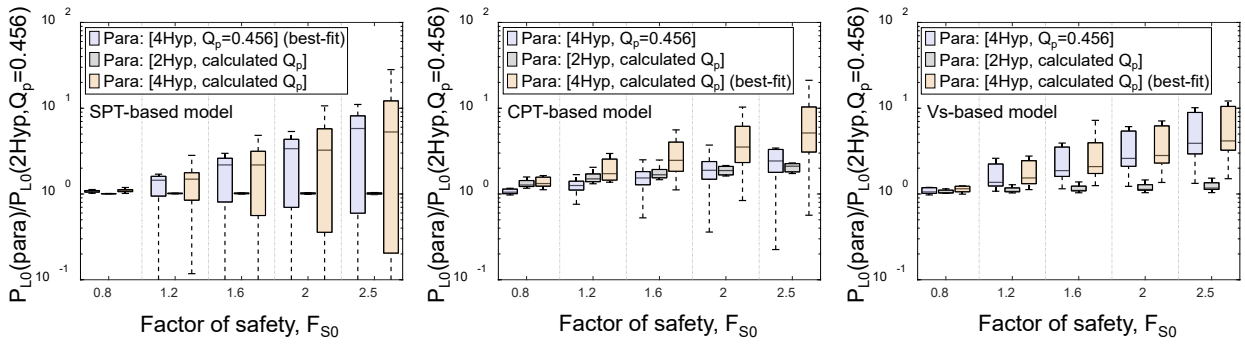
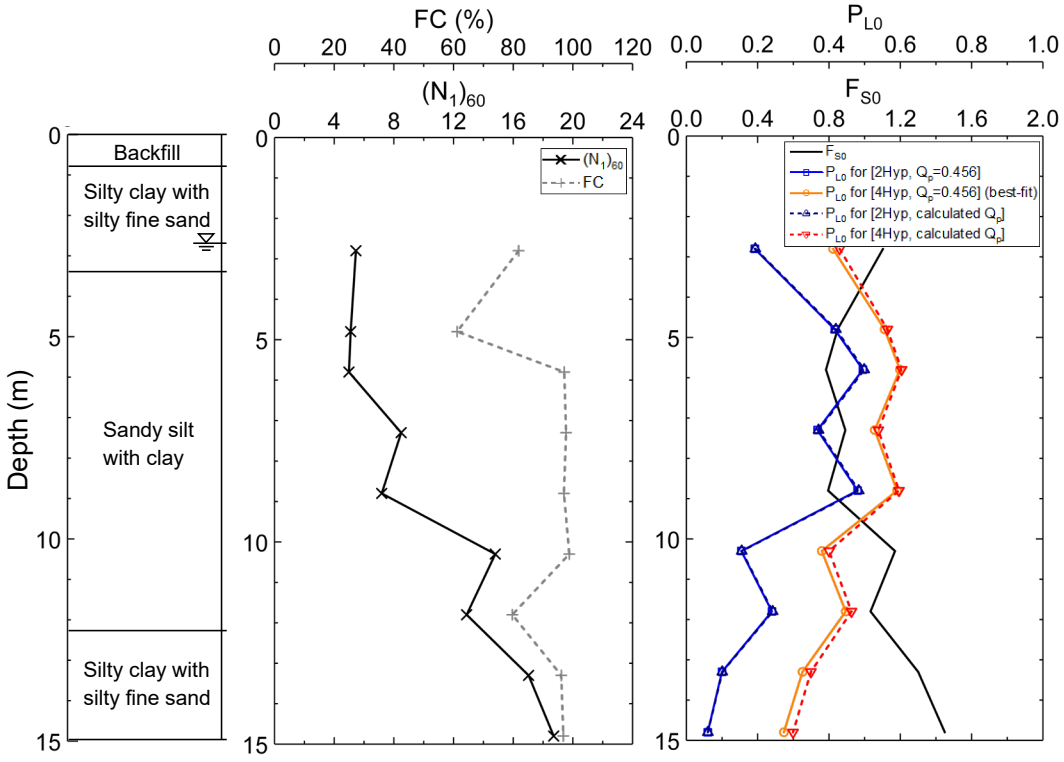


Fig. 8. (a) Ratios of P_{L0} for three candidate distributions to P_{L0} for best-fit distribution using [2Hyp, Q_p=0.456]; (b) ratios of P_{L0} for three parameter configurations to P_{L0} for [2Hyp, Q_p=0.456] using the best-fit distribution. Each boxplot displays the distribution of ratio values for all region-specific models, in which the central line marks the 50th percentile, the edges of box mark the 25th and 75th percentiles, and the ends of whiskers represent the 0.35th and 99.65th percentiles.

(a)



(b)

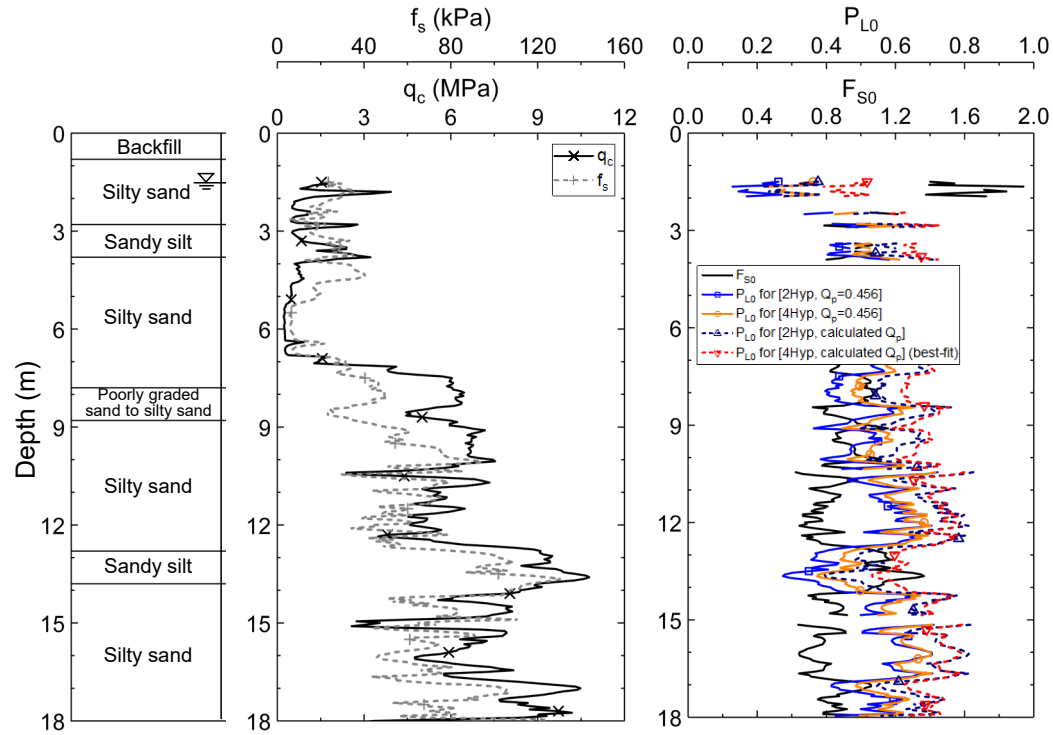
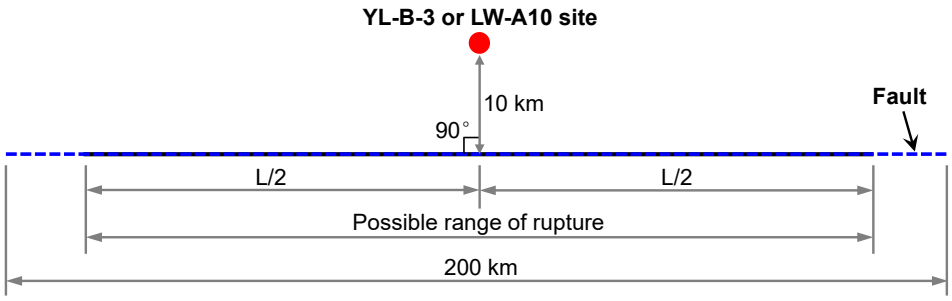


Fig. 9. Scenario-based liquefaction potential assessments for (a) YL-B-3 profile and (b) LW-A10 profile using updated SPT-BI12 models and CPT-RW98 models, respectively. The 1999 Chi-Chi earthquake region is considered.

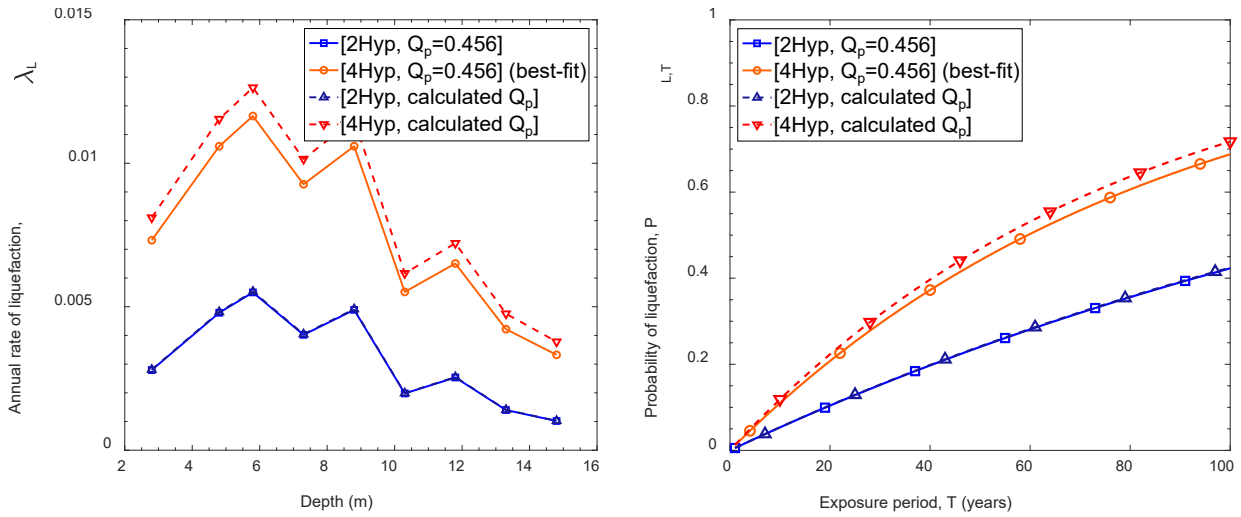


848

849 **Fig. 10.** Source-site relationship in the illustrative example.

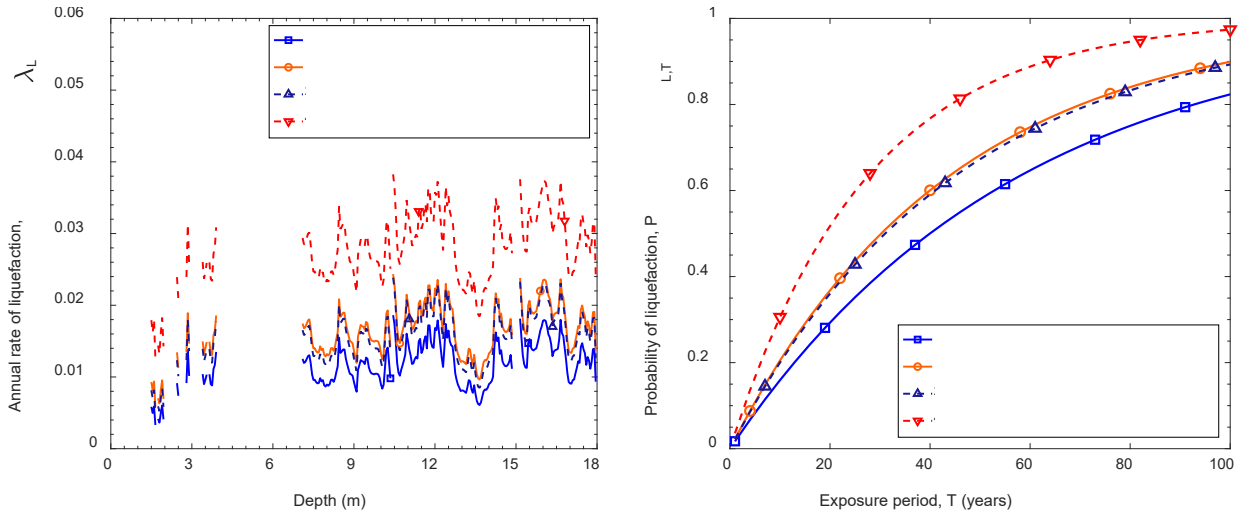
850

851 (a)



852

853 (b)



854

855 **Fig. 11.** Probabilistic liquefaction-triggering hazard results: (a) λ_L versus depth and $P_{L,T}$ at 5.8 m depth versus

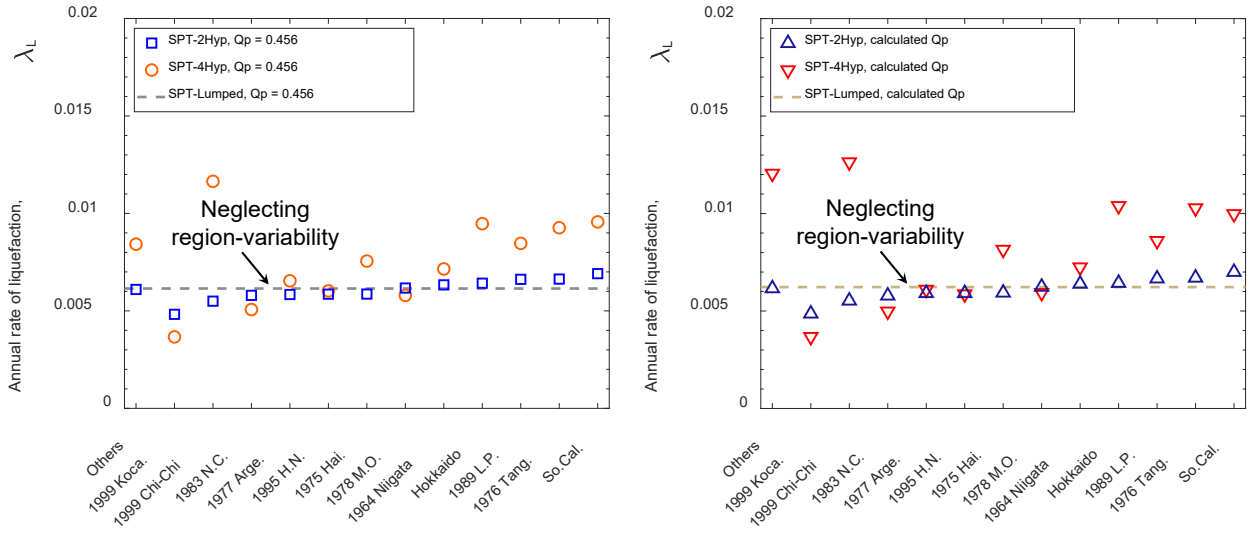
856 exposure period for YL-B-3 profile; (b) λ_L versus depth and $P_{L,T}$ at 10.5 m depth versus exposure period for

857 LW-A10 profile. The profiles are assumed to be located in regions influenced by the 1999 Chi-Chi

858 earthquake.

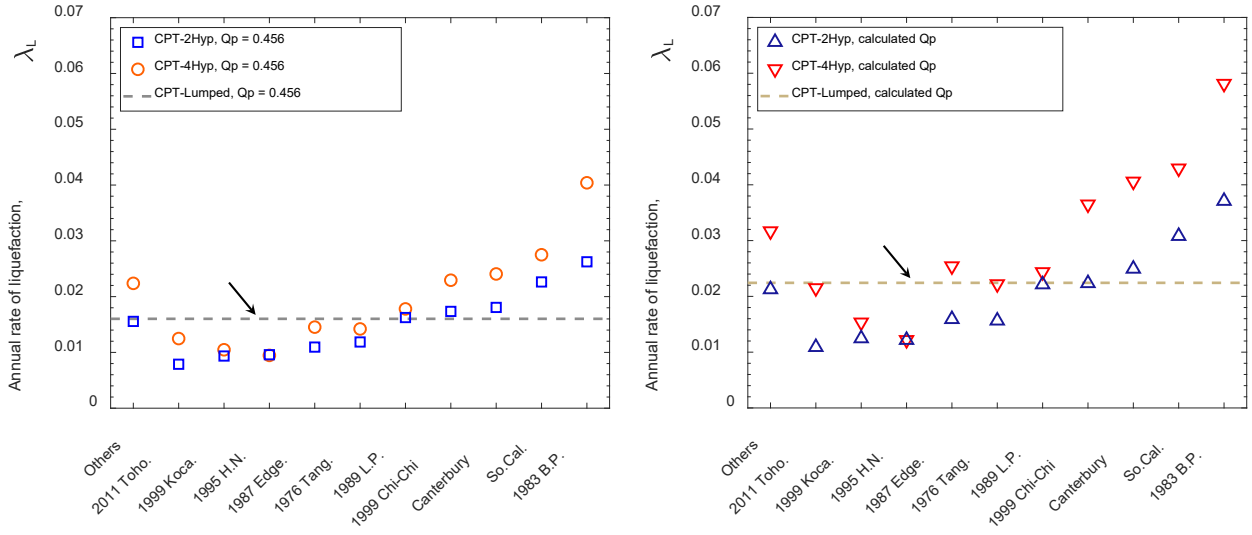
859

860 (a)



861

862 (b)



863

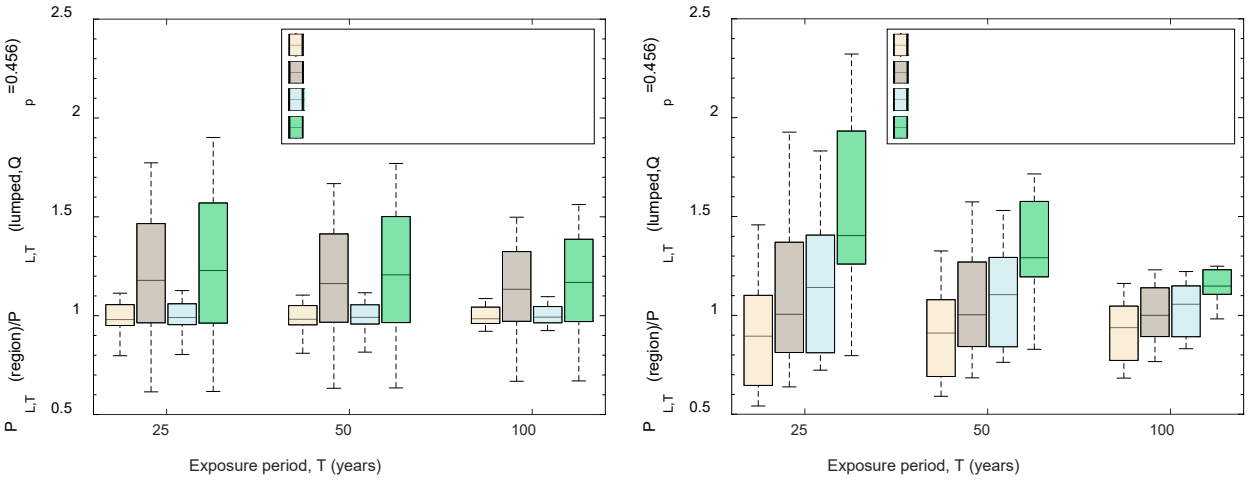
864 **Fig. 12.** Comparison of annual liquefaction rates from region-specific (a) SPT-BI12 models for the 5.8 m

865 depth in YL-B-3 profile and (b) CPT-RW98 models for the 0.5 m depth in LW-A10 profile. The profiles are

866 assumed to be located in different regions.

867

868 (a) (b)

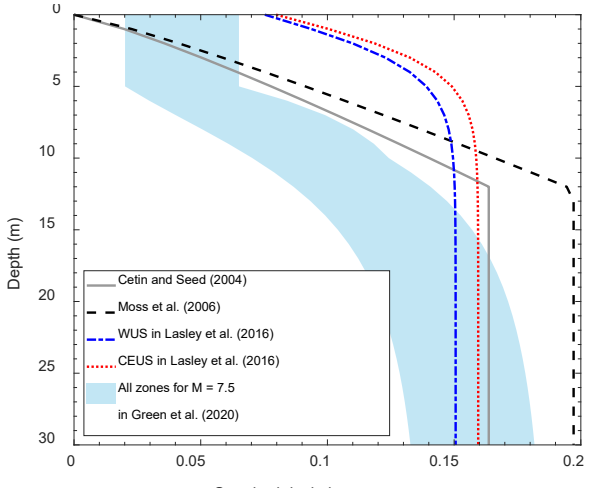


869

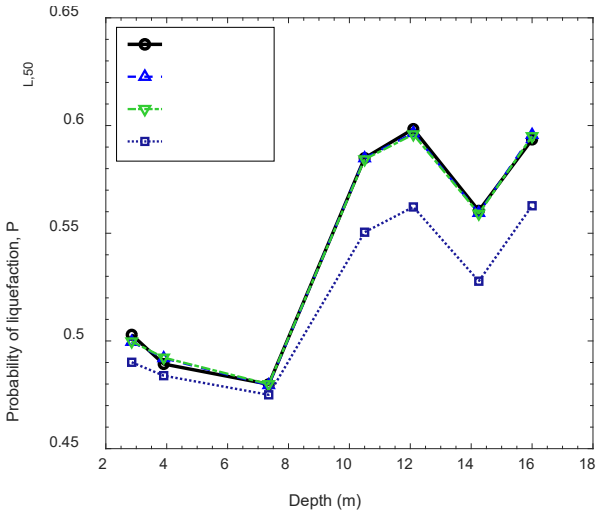
870 **Fig. 13.** Ratios of exposure period-consistent liquefaction probabilities (a) between region-specific and
871 lumped-region SPT-BI12 models and (b) between region-specific and lumped-region CPT-RW98 models.
872 The 5.8 m depth in YL-B-3 profile and 10.5 m depth in LW-A10 profile are considered for the SPT- and
873 CPT-based models, respectively.

874

875 (a) (b)

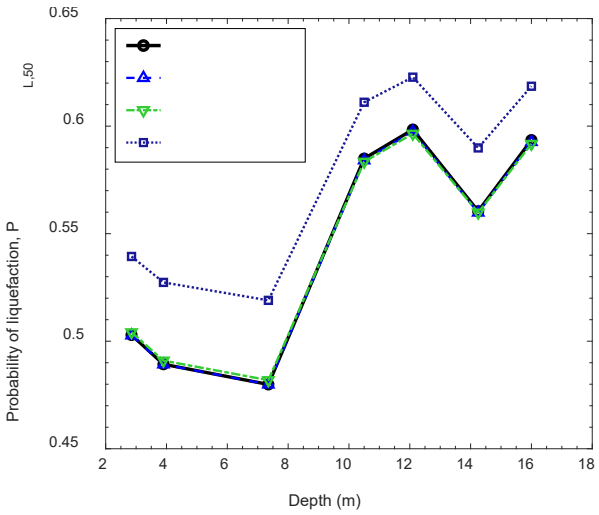


876



877

878 (c)

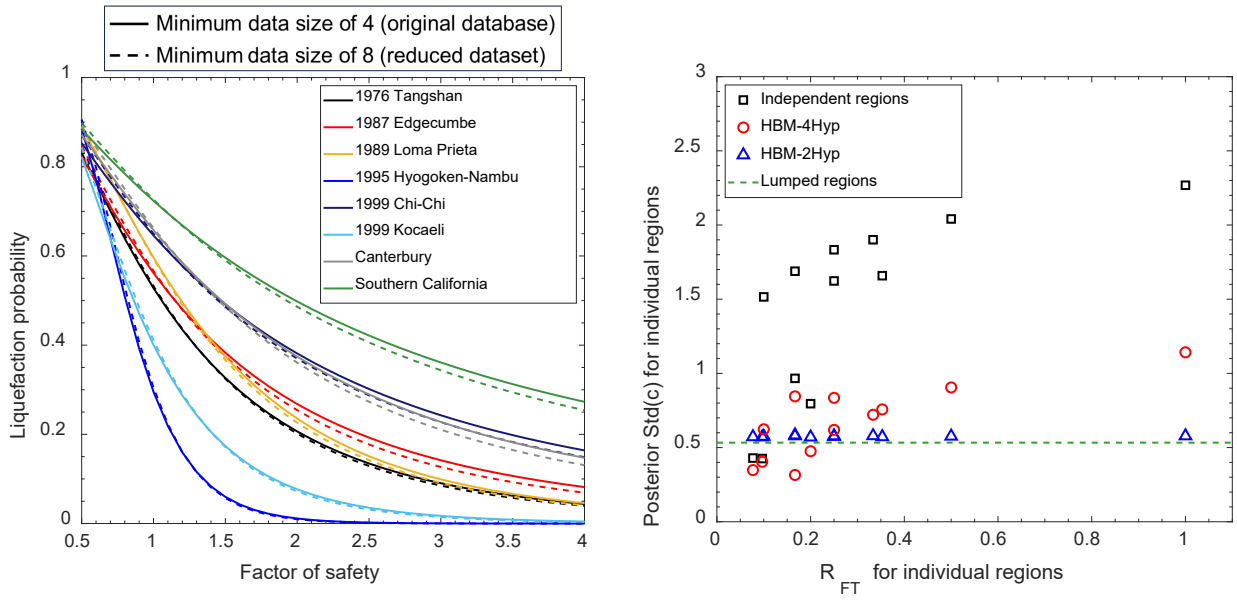


879

880 **Fig. 14.** (a) Variation of reported standard deviation of r_d with depth; (b) effect of σ_{rd} on $P_{L,50}$ using [CPT-

881 2Hyp, Q_p=0.456]; (c) effect of $\sigma_{\ln MSF}$ on $P_{L,50}$ using [CPT-2Hyp, Q_p=0.456].

882 (a) (b)



883

884 **Fig. 15.** (a) Comparison between P_{L0}-F_{S0} relationships derived from the reduced dataset and the original

885 database; (b) comparison among Std(c) values derived from different Bayesian models.

886