# A hybrid geotechnical and geological data-based framework for multiscale regional liquefaction hazard mapping

C. WANG* and Q. CHEN†

Regional assessment of liquefaction hazard necessitates not only models for evaluating liquefaction potentials and their spatial extent, but also the means to account for heterogeneous sources of information. In this work, a framework is developed that integrates geotechnical and geological data for the multiscale mapping of regional liquefaction hazard. The geotechnical data, such as the cone penetration test (CPT) record, are used to obtain primary liquefaction potential data through a classical CPT-based liquefaction model. A simplified method is proposed to derive secondary liquefaction potential data from geological information that essentially enforces geological constraints on the generated liquefaction hazard maps. Both primary and secondary liquefaction potential data are integrated into multiscale random field models through a conditional simulation algorithm to generate liquefaction hazard maps over the region of interest. The proposed framework is applied to an earthquake-prone region for the multiscale mapping and probabilistic assessment of liquefaction hazard. The results are validated with existing knowledge and liquefaction observations, and show improved predictions when both geotechnical and geological data are appropriately accounted for in the liquefaction hazard maps.

KEYWORDS: earthquakes; geology; liquefaction; seismicity; site investigation

## INTRODUCTION

Infrastructure such as bridges, buildings and underground utility lines are often threatened by soil liquefaction during earthquakes. Assessment of liquefaction hazard involves ground shaking hazard analysis and estimation of liquefaction resistance of soil deposits, where various in situ tests, such as the cone penetration test (CPT), the standard penetration test (SPT) and the shear wave velocity test, are commonly utilised to estimate liquefaction resistance or to classify levels of liquefaction severity at individual locations – for example, Seed & Idriss (1971); Robertson & Wride (1998); Andrus & Stokoe II (2000); Youd *et al.* (2001); Cetin *et al.* (2004); Moss *et al.* (2006); Idriss & Boulanger (2008); Shen *et al.* (2016).

When mapping liquefaction hazard over an extended region, classical approaches often rely on surficial geological data and knowledge of past liquefactions (Youd & Perkins, 1987; Knudsen *et al.*, 2000; Witter *et al.*, 2006). The geological data are sometimes supplemented with hydrological data and geotechnical data (Holzer *et al.*, 2006; Hayati & Andrus, 2008; Heidari & Andrus, 2010) or geomorphological information (Papathanassiou *et al.*, 2017). In these approaches, liquefaction hazard level is assumed to be constant within each surficial geological unit. Such an assumption ignores the inherent spatial variability of soil properties and therefore limits the accuracy and the applicability of the generated liquefaction hazard maps. To improve accuracy and to take advantage of the

ever-expanding high-quality geotechnical database, geostatistical methods have received increased attention in recent years. Geostatistical methods allow the incorporation of geotechnical data and allow the explicit consideration of the soil property spatial variability. For instance, Lenz & Baise (2007) analysed the spatial structure of liquefaction potential index (LPI) calculated from both CPT and SPT data and mapped the calculated LPI values over the Alameda County site of California using a kriging method. Chung & Rogers (2010) treated the groundwater table as a spatially varying property of interest. They used a kriging method to evaluate the regional variation of depth to groundwater, which was then converted to a map of LPI over the St Louis Area. The works of Vivek & Raychowdhury (2014), Liu & Chen (2006) and Juang *et al.* (2017) all explicitly considered the spatial variations of soil indices from CPT soundings when evaluating liquefaction potential. It was found that the probability of liquefaction could be significantly underestimated if the spatial dependence of soil indices is not considered. Liu *et al.* (2016) also treated soil indices from CPT soundings as spatially correlated random variables when mapping regional liquefaction-induced lateral spread. In a recent effort by Chen and co-workers (Chen *et al.*, 2016a, 2016b; Liu *et al.*, 2017), a multiscale random field model was developed to account consistently for soil spatial variability across multiple length scales and was applied to mapping soil properties and liquefaction potentials across a region.

The aforementioned geostatistical methods consider the spatial variability of the mapped properties (e.g. soil property indices, LPIs) derived from geotechnical data. However, the spatial structure of the mapped properties could vary within and across different geological units. The geological data, to some extent, provide information on large-scale (regional scale) material heterogeneity. Such information is not yet accounted for in current geostatistical method-based regional liquefaction hazard mapping studies. In this paper, a novel hybrid geotechnical and geological data-based framework will be developed to address this challenge. The framework consistently accounts for spatial variability of soil properties

* Glenn Department of Civil Engineering, Clemson University, Clemson, SC, USA.
† Glenn Department of Civil Engineering, Clemson University, Clemson, SC, USA (Orcid:0000-0002-0394-6710).

across scales, while simultaneously preserving constraints imposed by geological boundaries. Both geotechnical and geological data are integrated into a multiscale random field model through a conditional sequential simulation technique. The framework will be applied to an earthquake-prone region to demonstrate its capabilities and advantages.

## DATA

The project area is a region prone to earthquakes – the Alameda County site of California shown in Fig. 1. The availability of engineering data and the extensive past liquefaction studies in this region make it an ideal site to test and validate the proposed methodology. Information on engineering geology in the Alameda County site has been compiled by Helley & Graymer (1997) and briefly summarised in Holzer et al. (2006). This site is divided into three broad northwest–southeast trending regions, parallel to the Hayward Fault – the most important seismic source in this area. Bedrock is exposed at the surface of the northeast land. The central area, immediately southwest of the bedrock, consists of the Holocene and Pleistocene alluvial fan deposits. The area next to the central area – southwest of the original natural shoreline – is primarily underlain by the artificial fill that rests on younger San Francisco Bay mud.

A total of 210 CPT sounding profiles are collected from the U.S. Geological Survey (USGS) online CPT database (USGS, 2015). Water table information is directly obtained from each CPT sounding record. For CPT soundings without water table information, simple interpolation is used. For the unit weight of soil, constant values of $\gamma_m = 15 \cdot 0 \ \text{kN/m}^3$ for moist soil above the water table and $\gamma_{sat} = 19 \cdot 4 \ \text{kN/m}^3$ for saturated soil below the water table are assumed. However, when data are available, it would be more accurate to use varying unit weights along soil layers. A hypothetical magnitude $M_w = 7 \cdot 1$ earthquake is assumed with a constant peak horizontal ground acceleration, $a_{max} = 0 \cdot 5\boldsymbol{g}$. This combination of $a_{max}$ and $M_w$ has been assumed in a previous liquefaction mapping study (Holzer et al., 2006) for a hypothetical earthquake event on the nearby Hayward Fault. The assumption of a constant $a_{max}$ was justified on the basis that the outcrop area of the surficial geological unit generally parallels and is close to the Hayward Fault (Holzer et al., 2006). Alternatively, varying peak ground acceleration (PGA) values obtained from the latest USGS seismic hazard map could be adopted for the estimation of liquefaction hazard. The analysis region, its surficial geology and locations of CPT soundings are shown in Fig. 1.
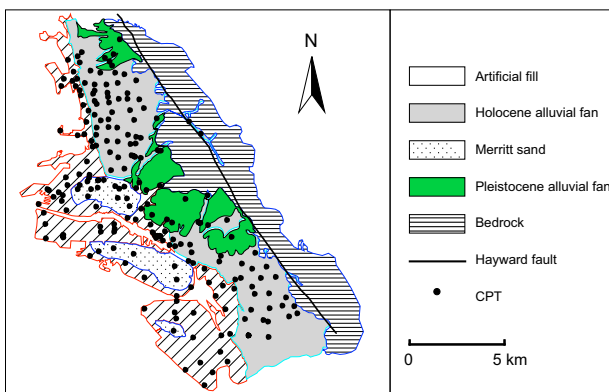


**Fig. 1. Map of the Alameda County site, surficial geology and locations of 210 CPT soundings (black dots). The surficial geology map is generated based on information compiled by Holzer et al. (2006)**

## OVERVIEW OF THE METHODOLOGY

A schematic illustration of the proposed framework is shown in Fig. 2. In this work, liquefaction hazard is quantified and mapped in terms of an averaged index called the LPI, initially proposed by Iwasaki et al. (1978, 1982) and utilised in many liquefaction studies – for example, Lee et al. (2004); Juang et al. (2008); Papathanassiou (2008); Maurer et al. (2014, 2015). The proposed methodology, however, is general and can be applied to map other measures of liquefaction hazard, such as liquefaction-induced ground settlement and lateral spread.

As shown in Fig. 2, the liquefaction hazard map is generated taking into account two types of LPI data. The first type, termed the 'primary data', is evaluated using a geotechnical-based LPI model – for example, the Robertson & Wride (1998) CPT-based liquefaction model. As shown in Fig. 2(a), CPT data are collected within the study area upon which the primary LPI are calculated. The second type, termed the 'secondary data', is obtained based on secondary information such as a surficial geological map (Fig. 1). As shown in Fig. 2(b), the distribution of primary LPIs within each geological unit is characterised. Then, as shown in Fig. 2(c), the secondary LPI data within each geological unit are generated based on the characterised distributions of primary LPIs. The role of the secondary data is to constrain and improve the primary data-based LPI map such that the final map conforms to the large-scale geological boundaries in the analysis region.

Based on the primary CPT-based LPI data and the secondary geology-based LPI data, multiscale random field models are developed to generate realisations of LPIs across the region of interest, as shown in Fig. 2(d). Coupled with Monte Carlo simulations, uncertainties associated with the generated liquefaction hazard maps can also be obtained. Various quantities of interest related to liquefaction hazard can be obtained and results will be presented and discussed in the results section.

## CPT-BASED PRIMARY LPI DATA

In this work, the classical CPT-based liquefaction model by Robertson & Wride (1998) is adopted to calculate the liquefaction potential of a soil layer, where two variables – namely, the cyclic stress ratio (CSR) and the cyclic resistance ratio (CRR) – are evaluated. Details of this classical liquefaction model have been summarised in a previous study (Wang et al., 2017) and are not repeated here. Once CSR and CRR have been obtained, the factor of safety (FS) against liquefaction at a particular depth $z$ can be calculated as

$$\text{FS} = \frac{\text{CRR}}{\text{CSR}} \tag{1}$$

To quantify liquefaction hazard at a particular location, the FS is integrated over the top 20 m of soil to obtain an averaged index termed the LPI (Iwasaki et al., 1978, 1982)

$$\text{LPI} = \int_0^{20} F_L w(z) \, \mathrm{d}z \tag{2}$$

where $z$ denotes the depth (m) and $w(z) = 10 - 0 \cdot 5z$; $F_L$ is defined as Sonmez (2003)

$$F_L = \begin{cases} 0 & \text{FS} \geq 1 \cdot 2 \\ 1 - \text{FS} & \text{FS} \leq 0 \cdot 95 \\ 2 \times 10^6 e^{-18 \cdot 427 \text{FS}} & 0 \cdot 95 < \text{FS} < 1 \cdot 2 \end{cases} \tag{3}$$

where FS is the factor of safety defined in equation (1). It should be noted that there are alternative liquefaction
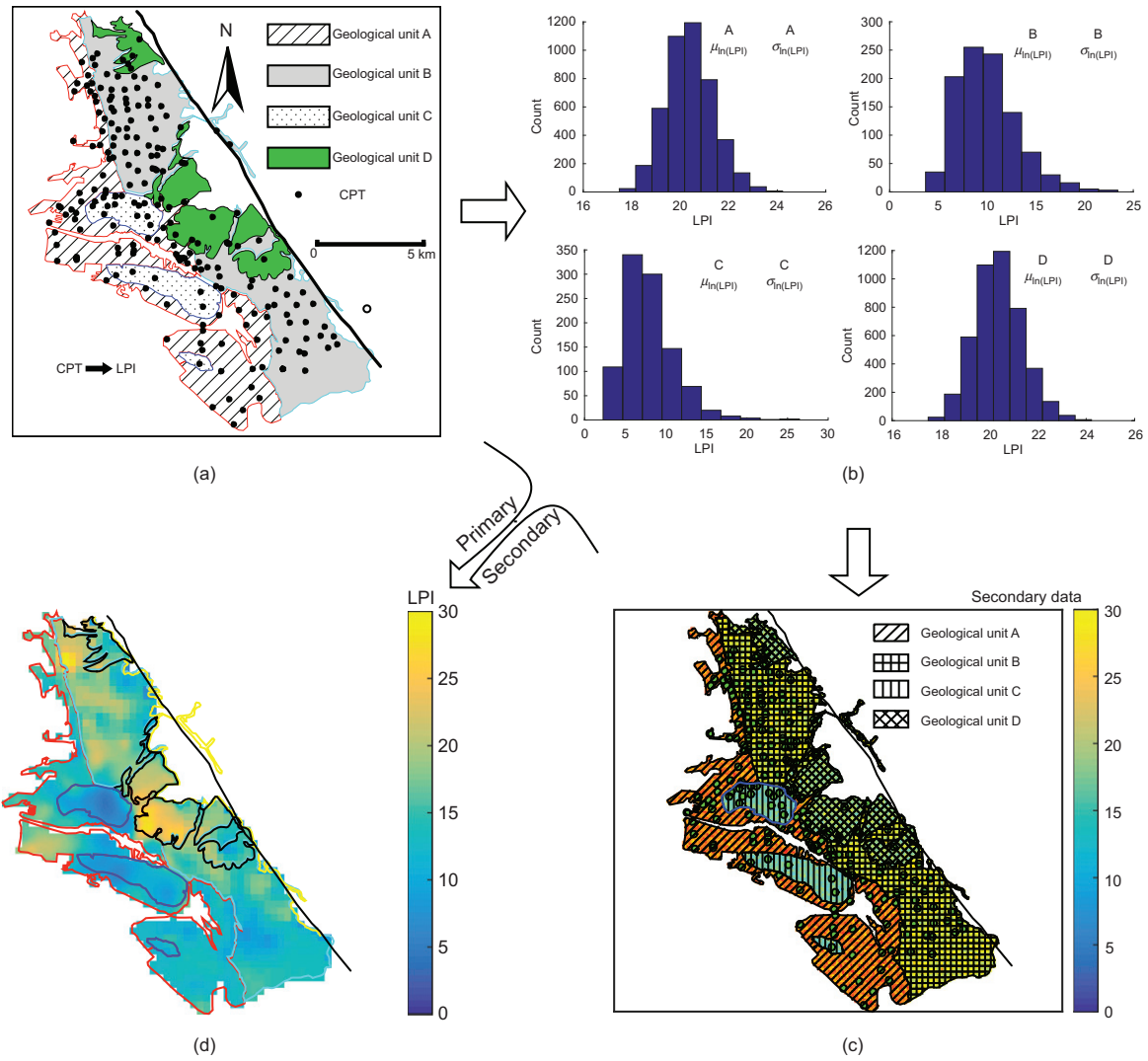
**Fig. 2. Regional mapping of LPI with geological constraints: (a) collect CPT data and calculate primary LPI data at CPT soundings; (b) characterise the distribution of LPIs within each geologic unit; (c) generate secondary LPI data within each geological unit based on the characterised parameters of distribution; (d) random field realisation of LPI over the whole region conditioned upon both primary and secondary data**

evaluation procedures used to assess liquefaction hazard in an LPI framework (Maurer *et al*., 2015), which could also be integrated with the framework developed in this study. The LPI will be used as the primary variable to be mapped over the project region through multiscale random field models.

For the hypothetical earthquake event with $M_w = 7 \cdot 1$ and $a_{max} = 0 \cdot 5g$, the primary LPI values are calculated at all 210 CPT soundings. The histogram of the calculated 210 LPI values is plotted in Fig. 3(a). To assess the spatial correlations of the primary LPI data, the empirical semivariogram $\hat{\gamma}(\boldsymbol{h})$ is calculated as (Goovaerts, 1997)

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})} \sum_{\alpha=1}^{N(\boldsymbol{h})} [z(\boldsymbol{u}_\alpha) - z(\boldsymbol{u}_\alpha + \boldsymbol{h})]^2 \qquad (4)$$

where $N(\boldsymbol{h})$ is the number of pairs of data $z$ located a vector $\boldsymbol{h}$ apart (i.e. a lag bin $\boldsymbol{h}$). In the actual computation, a small tolerance (e.g. $10 \sim 20\%$ of the distance $\boldsymbol{h}$) is usually added to lag bins to accommodate unevenly spaced sample points. Also, it is often convenient to use a scalar distance measure $h$ for the calculation of a semivariogram (Wang *et al*., 2017). Fig. 3(b) shows the calculated empirical semivariogram based on LPIs at 210 CPT soundings. Given the sample semivariogram, a weighted least-square method by Cressie

(1985) is implemented to fit an analytical semivariogram model, shown as a solid line in the plot.

## GEOLOGY-BASED SECONDARY LPI DATA

Knowledge of surficial geology is an important piece of information for regional liquefaction mapping as it typically provides broader area coverage and information on large-scale material heterogeneity. Potentially, any surficial geology-dependent liquefaction information (e.g. previous liquefaction observations, regional geology-based liquefaction hazard map) could be used to derive secondary LPI data that essentially enforce geological constraint to the generated liquefaction hazard map. In this section, a simplified procedure is presented to utilise the calculated primary LPI values and the surficial geological map to derive secondary LPI data. The method consists of the following steps.

(*a*) Identify the boundaries of each geological unit within the study region.
(*b*) Group the primary LPI values by geological units and characterise their statistical distributions.
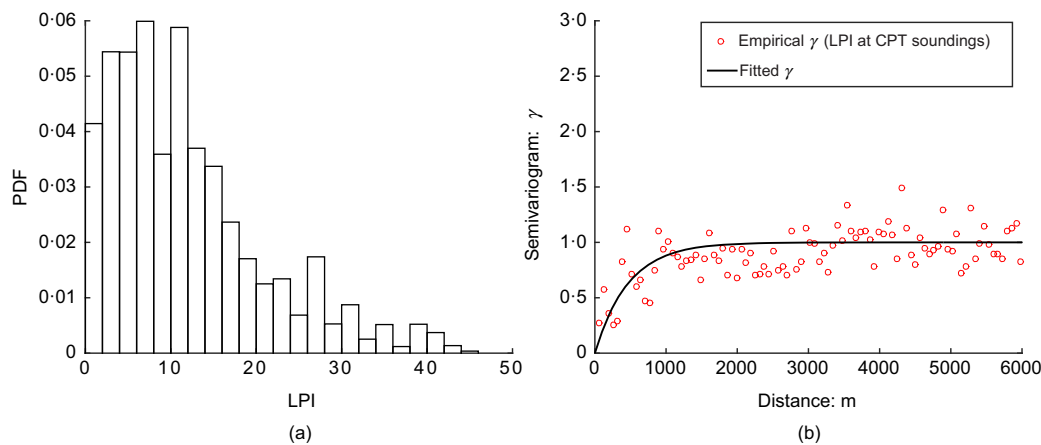
**Fig. 3. Primary LPI data at 210 CPT soundings in the Alameda County. The fitted model semivariogram is an exponential model $\gamma = 1 - e^{(-h/470)}$ with $h$ being the scalar distance: (a) histogram in probability density function (PDF); (b) semivariogram**

(c) Generate random variable realisations of secondary LPI values within each geological unit according to the characterised or assumed statistical distributions. Once generated, the secondary data will be kept constant for the following random field realisations.

(d) Assign the generated secondary LPI values to a predefined grid, which will be integrated into the conditional sequential simulation algorithm as secondary data.

For the project area, the surficial geology map of the Alameda County has been shown in Fig. 1, from which the boundaries of the geological units are identified. As summarised above, the next step is to group primary LPI

values by geological units and characterise their statistical distribution. Fig. 4 shows the histograms of primary LPIs at 210 CPT soundings grouped by the four main geological units in the study region – namely, the artificial fill, the Pleistocene fan, the Holocene fan and the Merritt sand. It can be seen from Fig. 4 that the distributions of LPIs within the artificial fill, the Holocene fan and the Merritt sand geological units can be fitted with a lognormal distribution, whereas the primary data are too sparse to conclude a basic distribution for the Pleistocene fan unit. In the following, a lognormal distribution will be assumed for generating secondary LPI data.

In addition to histograms, box plots and cumulative frequency curves are also used to understand the primary
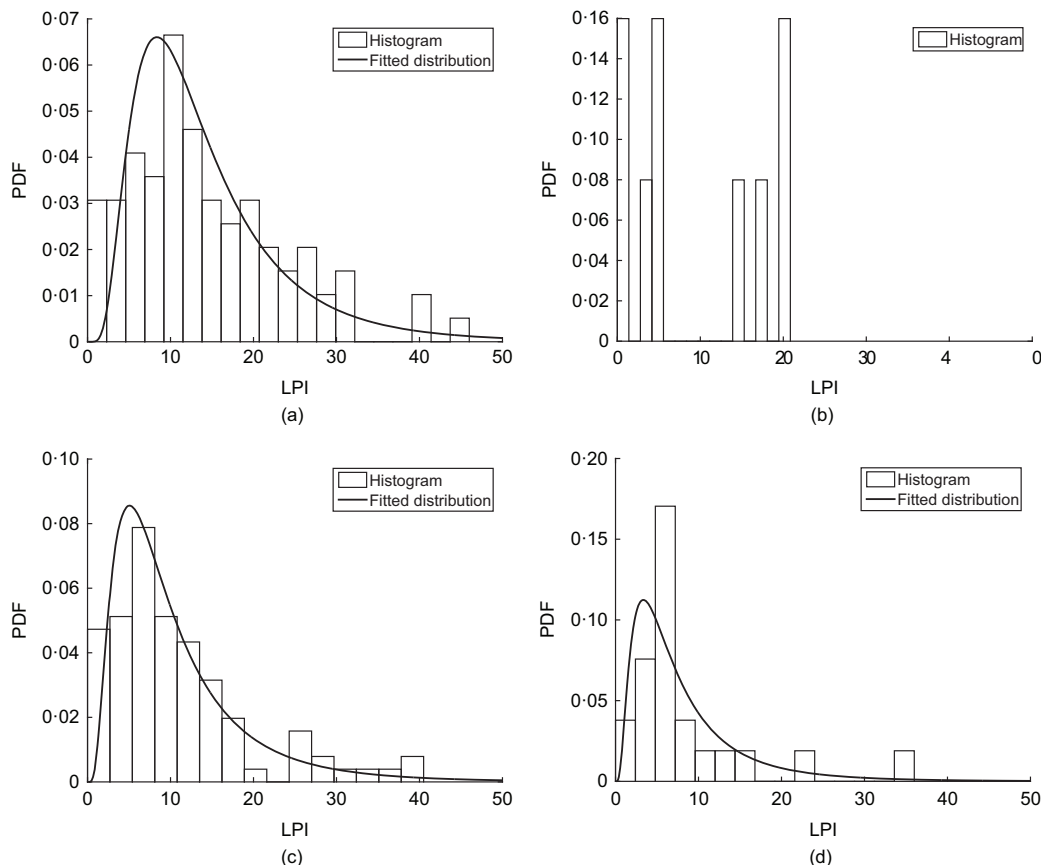


**Fig. 4. Histograms of primary LPIs for each geological unit: (a) artificial fill; (b) Pleistocene fan; (c) Holocene fan; (d) Merritt sand**
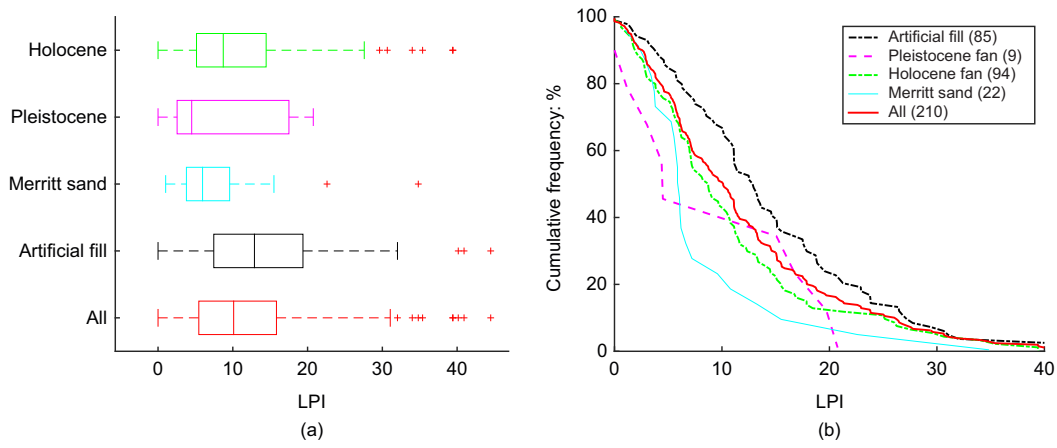
**Fig. 5.** (a) Box plots and (b) cumulative frequency of primary LPI data for each geological unit. The number in the bracket is the number of CPT soundings in that geological unit

LPI values grouped by geological unit and are shown in Fig. 5. It can be seen from the box plots that the median LPI values for each geological unit are ordered from high to low in this manner: artificial fill, Holocene fan, Merritt sand, Pleistocene fan. The cumulative frequency plots (Fig. 5(b)) show that, for a given LPI value, the artificial fill unit has the highest cumulative frequency values, while the Merritt sand unit has the lowest frequency values. The curve for the Pleistocene fan unit shows some abrupt changes due to a very limited amount of data points. Previously, such cumulative frequency curves have been used to assign a constant probability of liquefaction value to the corresponding geological unit (Holzer et al., 2006).

Basic statistical parameters of the primary LPIs for each geological unit are summarised in Table 1. Consistent with the above analysis, artificial fill has the highest mean value of LPI, followed by the Holocene fan. The parameters for Pleistocene fan and Merritt sand are very close to each other.

Once the distribution type and statistical parameters are determined, secondary LPI values for each geological unit are obtained from random field realisations and are assigned to a predefined grid. The results are shown in Fig. 6 for all geological units in the study region. Such secondary LPI values will be incorporated into the multiscale random field models described in the next section. It is worth noting that the above analysis groups geotechnical data based on surficial geological units. Geological units, however, are three-dimensional features. If the subsurface information is available, it may be more appropriate to group geotechnical data according to three-dimensional features.

## MULTISCALE RANDOM FIELD MODELS
### Conditional sequential simulation algorithm

A conditional sequential simulation algorithm is implemented in this work to generate random field realisations of LPIs. This algorithm integrates and preserves multiple sources of known data (e.g. primary and secondary

**Table 1. Statistics of primary LPIs by surficial geological units**

| Geological unit | Mean, $\mu$ | Standard deviation, $\sigma$ |
|---|---|---|
| Artificial fill | 14·5775 | 9·4749 |
| Pleistocene fan | 9·4846 | 8·4259 |
| Holocene fan | 10·8707 | 8·8719 |
| Merritt sand | 8·3618 | 7·6527 |

LPI data). In this algorithm, the realisation of a random variable $Z_n$ is represented by a joint distribution as follows

$$\begin{bmatrix} Z_n \\ \boldsymbol{Z}_p \\ \boldsymbol{Z}_s \end{bmatrix} \sim N\left( \begin{bmatrix} \mu_n \\ \boldsymbol{\mu}_p \\ \boldsymbol{\mu}_s \end{bmatrix}, \begin{bmatrix} \sigma_n^2 & \Sigma_{np} & \Sigma_{ns} \\ \Sigma_{pn} & \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sn} & \Sigma_{sp} & \Sigma_{ss} \end{bmatrix} \right) \quad (5)$$

where $\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the vector of random variables following a joint normal distribution with the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$; $Z_n$ is the random variable to be generated with the expected value $\mu_n$; $\boldsymbol{Z}_p$ is the vector of previously generated or known primary random variables with the vector of expected values $\boldsymbol{\mu}_p$; $\boldsymbol{Z}_s$ is a vector of secondary random variables with the vector of expected values $\boldsymbol{\mu}_s$; $\sigma_n$ is the standard deviation of $Z_n$; $\boldsymbol{\Sigma}$ is the covariance matrix with subscripts 'n', 'p' and 's' denoting 'next', 'previous primary' and 'secondary', respectively. The individual terms in the covariance matrix is defined as

$$\text{COV}\left[ Z_i, Z_j \right] = \rho_{Z_i,Z_j} \sigma_{Z_i} \sigma_{Z_j} \quad (6)$$

where $\rho_{Z_i,Z_j}$ is the correlation between two elements $Z_i$ and $Z_j$ within the random field at any scale with a standard deviation of $\sigma_{Z_i}$ and $\sigma_{Z_j}$, respectively.

Given the joint distribution equation (5), the distribution of the random variable $Z_n$, conditional upon all previously simulated and known primary and secondary data, is given by a univariate normal distribution with the updated mean and variance as

$$(Z_n|(\boldsymbol{Z}_p, \boldsymbol{Z}_s)) \sim N(\tilde{\mu}_n, \tilde{\sigma}_n) \quad (7)$$

with

$$\tilde{\mu}_n = \begin{bmatrix} \Sigma_{np} & \Sigma_{ns} \end{bmatrix} \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sp} & \Sigma_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{Z}_p \\ \boldsymbol{Z}_s \end{bmatrix} \quad (8)$$

$$\tilde{\sigma}_n^2 = \sigma_n^2 - \begin{bmatrix} \Sigma_{np} & \Sigma_{ns} \end{bmatrix} \begin{bmatrix} \Sigma_{pp} & \Sigma_{ps} \\ \Sigma_{sp} & \Sigma_{ss} \end{bmatrix}^{-1} \begin{bmatrix} \Sigma_{pn} \\ \Sigma_{sn} \end{bmatrix} \quad (9)$$

where symbols in equations (8) and (9) have all been defined after equation (5). The value of a random variable $Z_n$ at an unsampled location is drawn from the above joint distribution. Once generated, $Z_n$ becomes a data point in the vector $\boldsymbol{Z}_p$ to be conditioned upon by all subsequent simulations. This process is repeated by following a random path to each unknown location until all the values in the field have been simulated – that is, a map of the primary variable for the region of interest is generated.
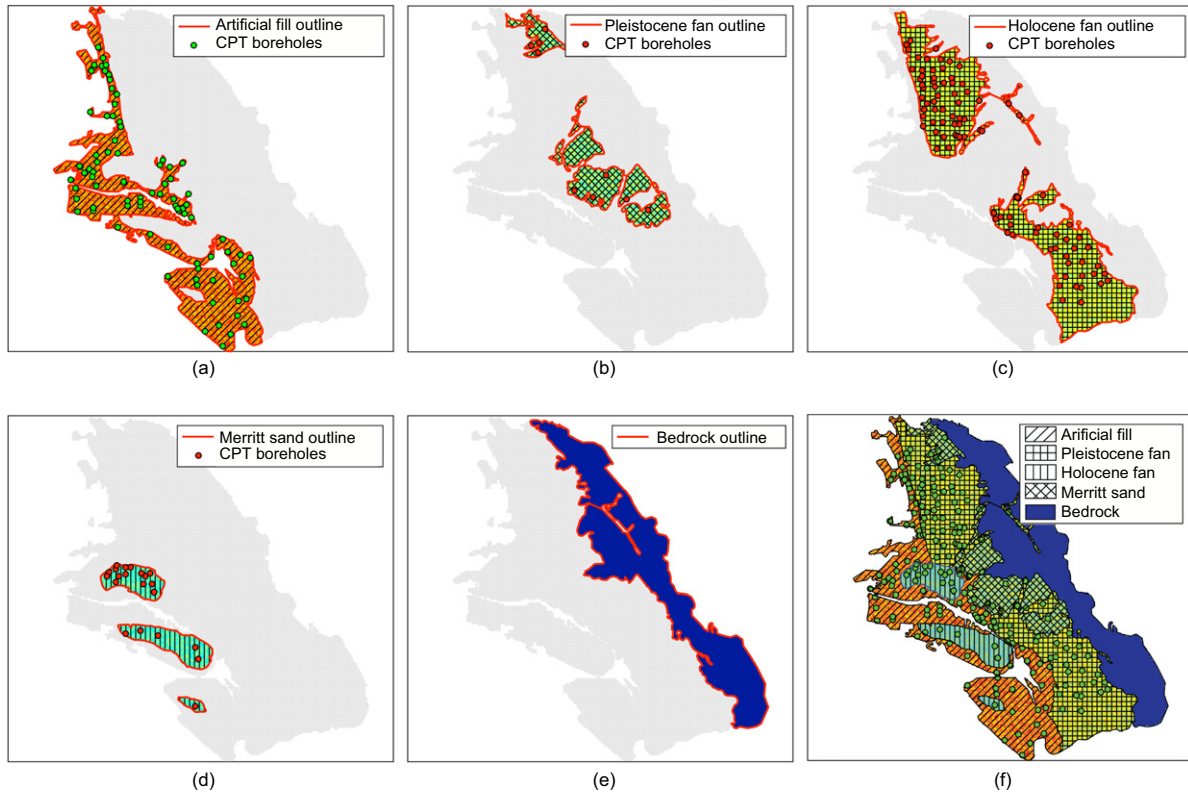
**Fig. 6. Generation of secondary data within each geological unit: (a) artificial fill; (b) Pleistocene fan; (c) Holocene fan; (d) Merritt sand; (e) bedrock; (f) all**

*Characterisation of spatial correlations across scales*

The derivation of spatial correlation across scales is based on the notion that material properties at the coarser scales are the arithmetically averaged values of the properties over corresponding areas at finer scales (Baker *et al.*, 2011; Chen *et al.*, 2012)

$$Z_I^c = \frac{1}{N} \sum_{i=1, i \in I}^{N} Z_i^f \tag{10}$$

where the superscripts 'c' and 'f' refer to the coarse and fine scales, respectively; $N$ is the count of fine-scale elements within a coarse-scale region. Defining the variable of interest at the fine scale, the mean of the variable at the coarse scale, denoted as $\mu_{Z^c}$, can be derived by taking the expectation of equation (10) as

$$\mu_{Z^c} = E[Z_I^c] = \frac{1}{N} \sum_{i=1}^{N} \mu_{Z_{i(I)}^f} = 0 \tag{11}$$

where $\mu_{Z_{i(I)}^f}$ is the mean of the fine-scale elements within a coarse element $I$, which equals zero for variables following the standard Gaussian distribution. Accordingly, if the variance of the fine-scale value is unity, the coarse-scale variance, denoted as $\sigma_{Z^c}^2$, can be computed as

$$\sigma_{Z^c}^2 = E[(Z_I^c)^2] - 0 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{Z_i^f, Z_j^f} \sigma_{Z_i^f} \sigma_{Z_j^f} \tag{12}$$

where $\rho_{Z_i^f, Z_j^f}$ is the correlation between two fine-scale elements $Z_i^f$ and $Z_j^f$ with standard deviations of $\sigma_{Z_i^f}$ and $\sigma_{Z_j^f}$, respectively.

The correlations between all considered scales can be calculated by rearranging the definition of the covariance equation (6) such that

$$\rho_{Z_i, Z_j} = \frac{\text{COV}[Z_i, Z_j]}{\sigma_{Z_i} \sigma_{Z_j}} \tag{13}$$

By making appropriate substitutions at each scale using equations (6) and (13), the correlation between elements at different scales can be obtained as (Chen *et al.*, 2015)

$$\rho_{Z_I^c, Z_{II}^c} = \frac{\sum_{i=1}^{N} \sum_{k=1}^{N} \rho_{Z_{i(I)}^f, Z_{k(II)}^f}}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{Z_{i(I)}^f, Z_{j(I)}^f}} \sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{Z_{i(II)}^f, Z_{j(II)}^f}}} \tag{14}$$

$$\rho_{Z^f, Z_I^c} = \frac{\sum_{i=1}^{N} \rho_{Z^f, Z_{i(I)}^f}}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} \rho_{Z_{i(I)}^f, Z_{j(I)}^f}}} \tag{15}$$

where $\rho_{Z_I^c, Z_{II}^c}$ is the correlation between two coarse-scale elements $I$ and $II$; $\rho_{Z^f, Z_I^c}$ is the correlation between a fine-scale element and a coarse-scale element $I$; $\rho_{Z_{i(I)}^f, Z_{k(II)}^f}$ is the correlation between a fine-scale element $i$ and a fine-scale element $k$, which belong to two different coarse-scale elements $I$ and $II$, respectively. Given the multiscale spatial dependence specified by equations (14) and (15) and an inferred or assumed probability distribution of the random variable, the conditional sequential simulation algorithm is employed to generate random field realisations of variables of interest. In the geostatistics community, the semivariogram $\gamma(\boldsymbol{h})$ measuring the dissimilarity of variate separated by a vector distance $\boldsymbol{h}$ is typically used and can be easily converted to the correlation function $\rho(h)$ (Remy *et al.*, 2009).

*Covariances across data types*

To perform the conditional sequential simulation in equation (7), three covariances must be determined: one for the primary variable, one for the secondary variable and a cross-covariance describing the relationship between these

variables. With relatively sufficient amount of primary LPI data calculated from CPT soundings, the covariance of the primary variable can be easily obtained from the inferred spatial correlation between primary data. However, direct calculation of the secondary and cross-covariances can be challenging. In this work, one simplified approach is adopted based on the Markov–Bayes hypothesis described by Goovaerts (1997) to derive the secondary and cross-covariances by calibrating them to the primary covariance as (Goovaerts, 1997; Moysey *et al.*, 2003)

$$\text{COV}_s(\boldsymbol{h}) = \begin{cases} |B| \cdot \text{COV}_p(\boldsymbol{h}) & \text{for} \quad \boldsymbol{h} = 0 \\ B^2 \cdot \text{COV}_p(\boldsymbol{h}) & \text{for} \quad \boldsymbol{h} > 0 \end{cases} \quad (16)$$

$$\text{COV}_{ps}(\boldsymbol{h}) = B \cdot \text{COV}_p(\boldsymbol{h}) \quad (17)$$

where $B$ is the Markov–Bayes coefficient; $\text{COV}_p$ is the covariance for the primary variable, $\text{COV}_s$ is the covariance for the secondary variable and $\text{COV}_{ps}$ is the covariance between the primary and the secondary variables; $\boldsymbol{h}$ is the distance vector separating two random variables. The Markov–Bayes coefficient $B$ generally varies between 0 and 1 when primary and secondary variables are positively correlated. Its value affects the relative importance of primary data and secondary data and this effect will be illustrated in the results section.

The coefficient $B$ can be chosen based on a calibration procedure recommended by Deutsch & Journel (1998) such that the value $B$ is determined as the difference between two conditional expectations as follows

$$B = E_1 - E_0 \quad (18)$$

where the two conditional expectations are defined as

$$E_1 = \text{E}(\text{Prob}\{Z_s \le z\} | Z_p \le z) \quad (19)$$

$$E_0 = \text{E}(\text{Prob}\{Z_s \le z\} Z_p > z) \quad (20)$$

where E is the expectation operator; $E$ is the expectation; $Z_s$ is the secondary variable (e.g. the geological data-based LPI value) and $\text{Prob}\{Z_s \le z\}$ is the probability of $Z_s$ less than or equal to a threshold value $z$ (e.g. a given LPI threshold value); $Z_p$ is the primary variable (e.g. the geotechnical data-based LPI value). The conditional expectation $E_1$ will be close to 1 if the primary and secondary data support each other – that is, the two data predict similar liquefaction hazard levels. The conditional expectation $E_0$ will be close to 1 if the primary and secondary data contradict each other – that is,

the two data predictions of liquefaction hazard contradict each other.

## RESULTS AND DISCUSSION
### Multiscale random field realisations of LPI
Given both primary and secondary LPI data obtained from CPT and geological information, random field realisations of LPIs across the Alameda County site are performed. The region is first discretised by a relatively coarse grid and an example realisation of LPIs on the coarse grid is shown in Fig. 7(a). Upon the coarse-scale realisation, the grid can be adaptively refined in subregions of higher interest while maintaining consistent spatial correlations across scales. As an example, a subregion marked by the box lines in Fig. 7 is selected for refinement. The corresponding multiscale realisation is shown in Fig. 7(b). This subregion is identified as an area of high liquefaction potential by Kayen *et al.* (1998) and is evidenced by the concentration of higher LPI values. This refinement process can be continued down to multiple finer scales. Since the refinement is only performed where necessary, it is computationally efficient to generate multiscale random field realisations of LPIs for a very large region. The benefit and application of multiscale LPI realisations will be illustrated in the following section.

### Monte Carlo analysis
For regional liquefaction hazard mapping, the multiscale random field models are coupled with Monte Carlo simulations to obtain the expected liquefaction hazard across the region and to perform probabilistic analysis on quantities of interest. To investigate the influence of the Markov–Bayes coefficient $B$ introduced in equation (16), six $B$ values are used – namely, 0, 0·1, 0·4, 0·5, 0·73 and 0·9. For each $B$ value, 1000 Monte Carlo simulations are performed. A hypothetical earthquake event ($M_w = 7·1$, $a_{max} = 0·5$ $g$) is assumed for all simulations.

Figure 8 shows maps of expected LPI values for all six cases of Markov–Bayes coefficient $B$. Each of the six maps is obtained by averaging results from 1000 Monte Carlo simulations. As shown in equation (16), the Markov–Bayes coefficient $B$ is essentially a 'scaling' factor between the primary covariance and the secondary covariance matrices. The larger the coefficient $B$, the stronger influence the secondary data has on the generated LPI maps. In this work, the secondary LPI data come from geological
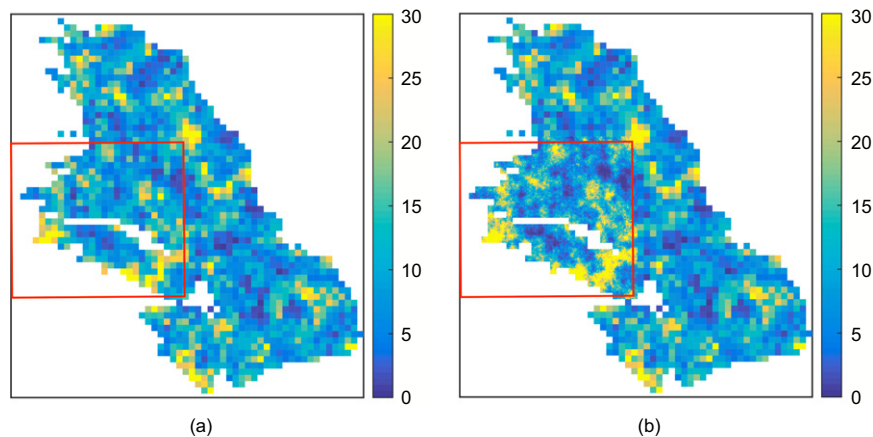


**Fig. 7. Example realisations of LPIs across the Alameda County site for a hypothetical earthquake event ($M_w = 7·1$, $a_{max} = 0·5g$) and the Markov–Bayes coefficient $B = 0$: (a) single scale; (b) multiscale**
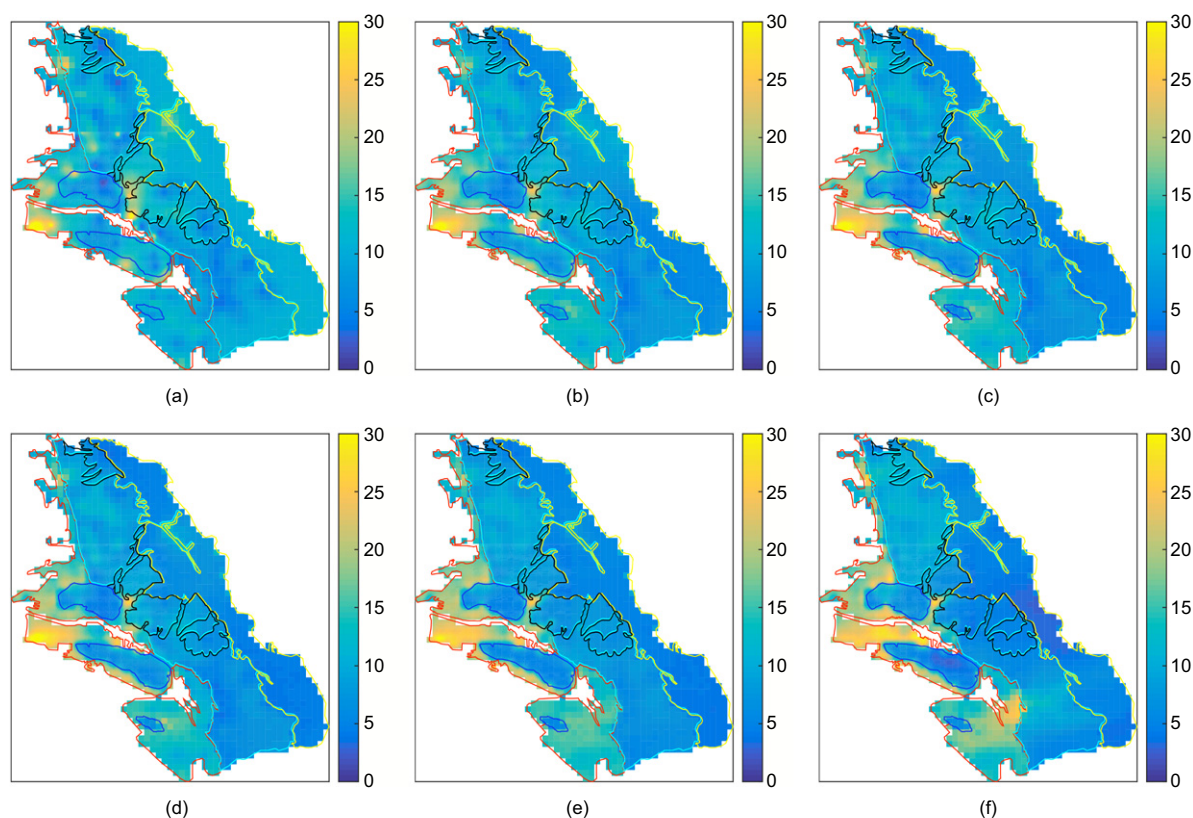
**Fig. 8.** Maps of expected LPI values for all six cases of Markov–Bayes coefficient $B$. $B^* = 0$ is the case without secondary LPI data. $B^* = 0.73$ is the case where the coefficient $B$ is calibrated. Each map is obtained by averaging results from 1000 Monte Carlo simulations: (a) $B^* = 0$; (b) $B = 0.1$; (c) $B = 0.4$; (d) $B = 0.5$; (e) $B^* = 0.73$; (f) $B = 0.9$

information. Therefore, as the value of $B$ increases, the geological boundaries become more distinguishable in the resulting LPI maps, as evidenced in Figs 8(b)–8(f). When no secondary data are incorporated – that is, $B = 0$ – no geological constraint is applied to the LPI map. Such a case is shown in Fig. 8(a) and the resulting LPI map might be inaccurate. For instance, lacking geological constraint, the bedrock unit (refer to Fig. 1 for the geological unit map of the study region) is predicted to have LPI values around 10, which is usually classified as a high liquefaction severity class (Sonmez, 2003). This contradicts the common knowledge that the bedrock is not prone to liquefaction. By applying secondary LPI data, this type of incorrect classification could be mitigated.

The case of $B = 0.73$ is obtained following the calibration procedure described in equation (18) based on an LPI threshold of 5, which is the value separating the moderate and the high liquefaction severity class (Sonmez, 2003). It should be noted that the selection of LPI threshold affects the calibrated $B$ value. For the Alameda County site, the threshold value of 5 is chosen as it appropriately separates liquefaction-prone and non-liquefied units, as evidenced by Fig. 5. As shown in Fig. 8(e), the LPI map with the calibrated $B = 0.73$ manifests reasonable spatial variations of LPI that comply with the known data while preserving boundaries of different geological units. It should be noted that the calibrated $B$ value can be used as a starting point for further adjustments. For instance, if there is a relatively high confidence in the quality of secondary data, or if the secondary data are known to have greater impact on liquefaction, the coefficient $B$ could be increased. If there is very little knowledge or low confidence about the secondary data, a lower value of $B$ could be adopted.

The results obtained and shown in Fig. 8 can be utilised to quantify liquefaction hazard and to calculate various

quantities of interest. As an example, following the procedure proposed by Holzer *et al.* (2006), cumulative frequency (CF) distributions of LPIs are calculated to assess the liquefaction potentials of each geological unit. The CF distributions obtained with different $B$ values are plotted in Fig. 9. For a given geological unit, Holzer *et al.* (2006) related the percentage area predicted to undergo liquefaction during a given earthquake shaking scenario to the value of CF corresponding to LPI = 5. Fig. 8 shows that the artificial fill geological unit has the highest CF values for a given LPI and is therefore most likely to liquefy. The Merritt sand unit has the lowest CF values for a given LPI and is predicted to be the most resistant to liquefaction. The impact of coefficient $B$ on the CF distributions is relatively small.

*Comparison with prior knowledge and observations*

As a modest validation of the framework, maps of the expected liquefaction potential hazards are compared with observations of liquefaction following the 1989 Loma Prieta earthquake (Kayen *et al.*, 1998). As shown in Fig. 10, most of the observed liquefied areas (along the coastline and in the artificial fill unit) are also predicted to have high LPI values by the proposed framework. As previously mentioned, the LPI map with the calibrated Markov–Bayes coefficient $B = 0.73$ predicts the bedrock unit as non-liquefiable (LPI $\approx 0$), which is consistent with the fact that no liquefaction phenomenon was observed during past earthquakes.

*Small-scale analysis*

To gain more insights into the improved predictions when the calibrated coefficient $B = 0.73$ is used, two small areas, denoted as 'area A1' and 'area A2' in Fig. 10, are investigated in more detail. Area A1 is underlain by the Merritt sand
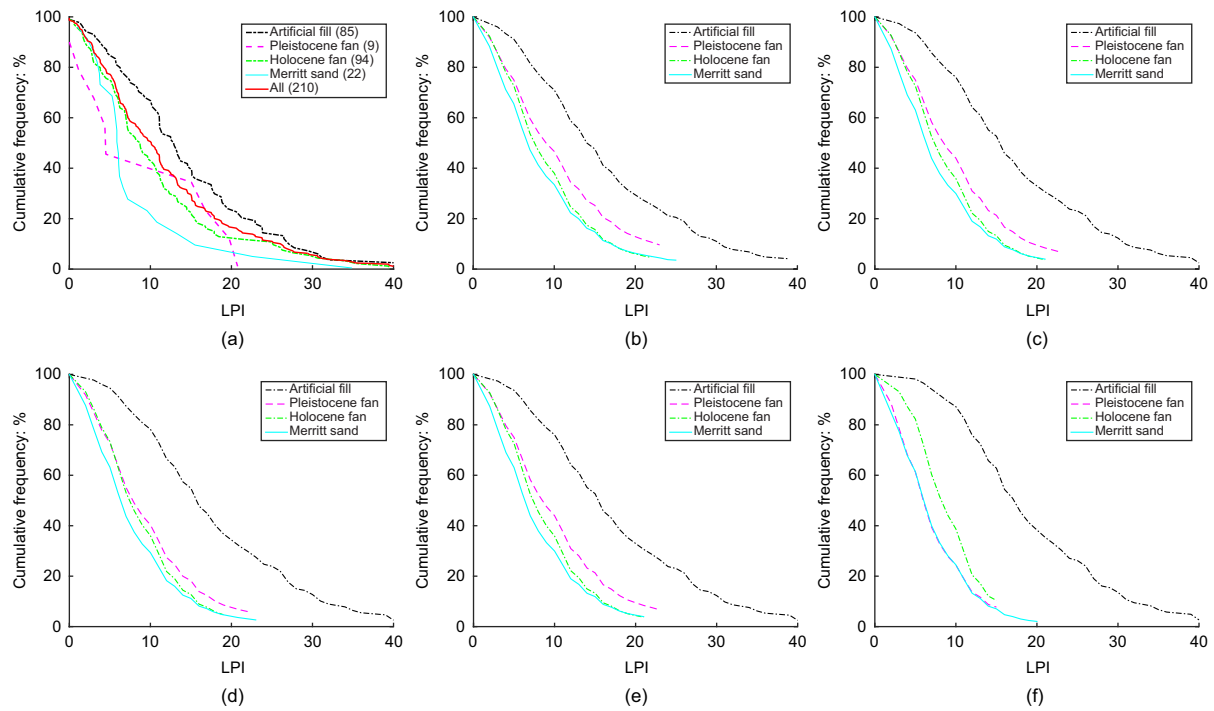
**Fig. 9. Cumulative frequency distributions of LPI: (a) known data; (b) $B = 0.1$; (c) $B = 0.4$; (d) $B = 0.5$; (e) $B = 0.73$; (f) $B = 0.9$**
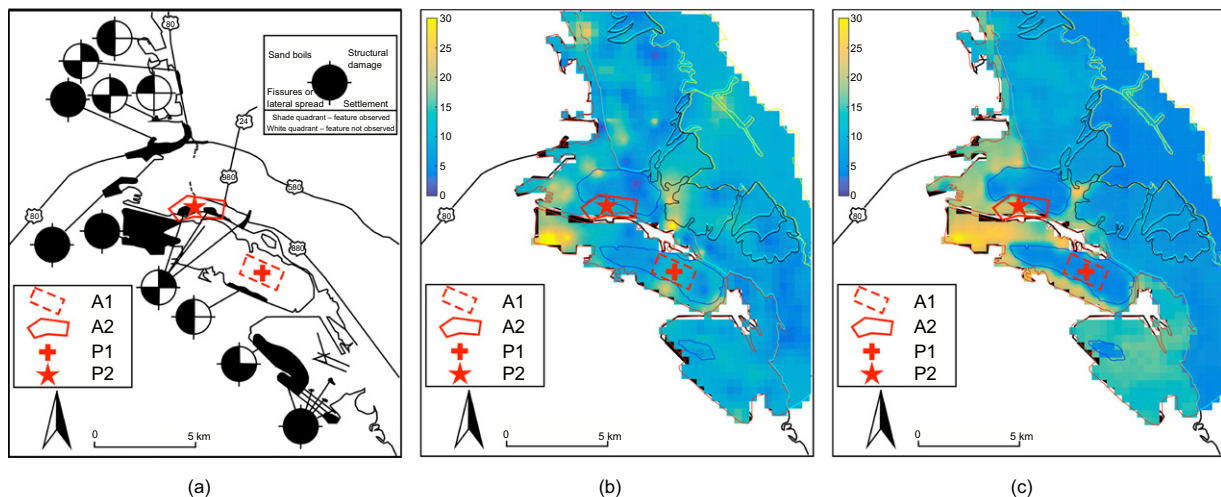


**Fig. 10. Comparison of liquefaction observations from the 1989 Loma Prieta earthquake with LPI maps generated by the proposed framework: (a) liquefaction observations (modified from Kayen *et al.* (1998)); (b) LPI map with $B = 0$; (c) LPI map with $B = 0.73$**

geological unit (refer to Fig. 1), where low liquefaction hazard is expected as evidenced by the fact that there was no liquefaction observation in past earthquakes and confirmed by results of the statistical analysis (summarised in Table 1 and Fig. 5). When no secondary data are used, as shown in Fig. 10(b), relatively high LPI values ($\approx 10$) are predicted in area A1, which are classified as high liquefaction severity by Sonmez (2003) and contradict the knowledge of liquefaction hazard in this area. In contrast, the LPI map with $B = 0.73$ (Fig. 10(c)) correctly predicts low liquefaction potential in area A1.

The second small area, denoted as 'area A2', is expected to suffer high liquefaction hazard as evidenced by the observed liquefaction damage and confirmed by the statistical analysis. The LPI map without secondary data (Fig. 10(b)) incorrectly predicts low liquefaction potential values, while the LPI map with calibrated coefficient $B = 0.73$ (Fig. 10(c))

again correctly predicts the liquefaction potential values, which shows the improved accuracy, especially at small scale, when geological data are appropriately accounted for in the liquefaction mapping.

A more quantitative comparison is presented in Fig. 11 in terms of cumulative frequency distributions of LPI. As shown in Fig. 11(a), the LPI map generated without secondary data predicts that areas A1 and A2 are expected to have similar liquefaction potential hazards, which clearly contradicts knowledge and observations. In contrast, Fig. 11(b) shows that, when the secondary geological-based LPI data are appropriately accounted for, the predicted liquefaction hazard in the two areas are consistent with knowledge and observations.

To further illustrate the advantage of multiscale random field models, the LPI maps are refined to such a high resolution that it is possible to investigate liquefaction hazard
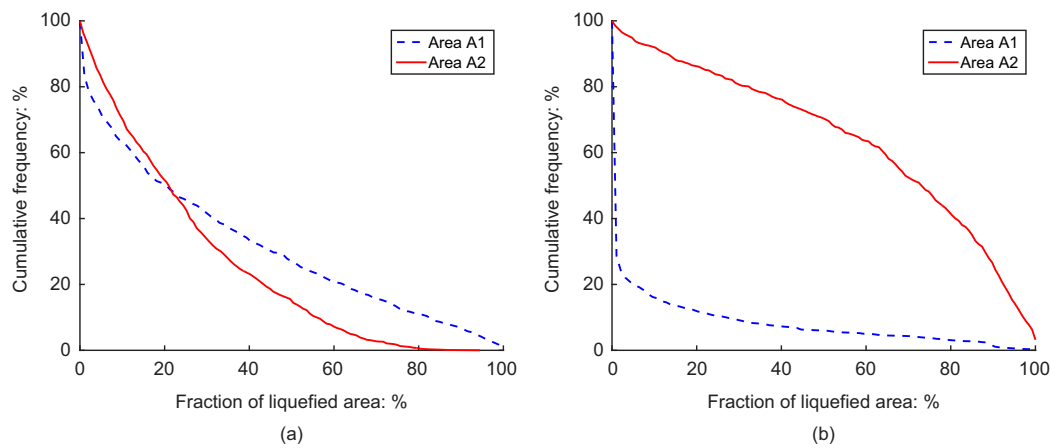
**Fig. 11. Cumulative frequency distributions of LPIs within area A1 and A2 for cases without ($B = 0$) and with ($B = 0 \cdot 73$) secondary data: (a) $B = 0$; (b) $B = 0 \cdot 73$**
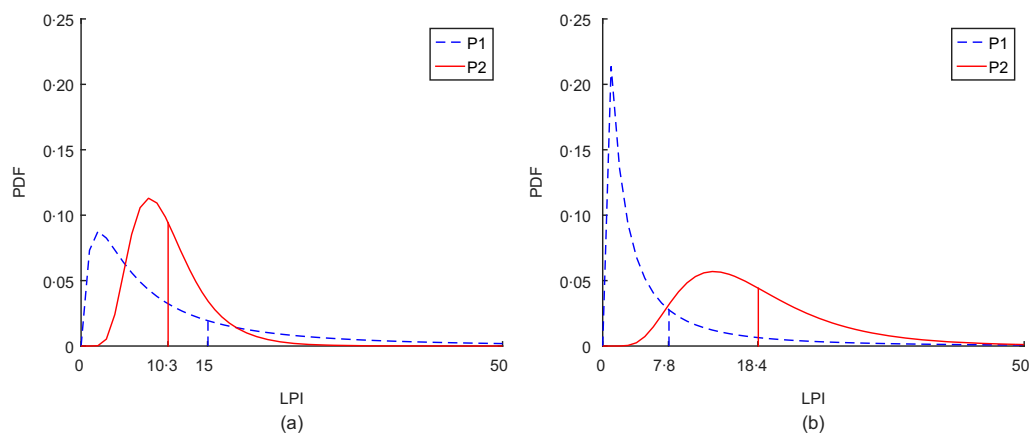


**Fig. 12. Monte Carlo analysis at two local sites P1 and P2 (10·3, 15, 7·8 and 18·4 are mean values of the corresponding distributions): (a) $B = 0$; (b) $B = 0 \cdot 73$**

at particular local sites. For instance, as shown in Fig. 10(a), two local sites denoted as P1 (37°46′8·0″ N 122°15′12·8″ W) and P2 (37°47′50·2″ N 122°17′24·9″ W) are selected where a high-resolution random field is available. Predicted distributions of LPIs at these two sites can be obtained using previously performed Monte Carlo simulations and the distributions are shown in Fig. 12. As has been previously analysed, the results with geological constraint ($B = 0 \cdot 73$) are more accurate. Such small-scale site-specific analysis can be useful information to assist engineering design at these locations.

## CONCLUSIONS

In this work, a hybrid geotechnical and geological data-based framework is developed for the multiscale random field-based regional liquefaction hazard mapping. The geotechnical data such as the CPT data are used to calculate primary LPI values, which show clear spatial correlation, as evidenced by the calculated empirical semivariogram. A simplified method is proposed to derive secondary LPI values from geological information which essentially enforces geological constraints to the generated liquefaction hazard maps. A Markov–Bayes coefficient is introduced to derive covariances related to secondary data. Both primary and secondary LPI data are then integrated into multiscale random field models through a conditional

sequential simulation technique. The proposed framework is applied to map liquefaction hazard at the Alameda County site in California. It is found that the influence of geological constraints on the generated liquefaction hazard maps is significant. Without geological constraints and with only geotechnical data, the generated liquefaction hazard map incorrectly predicts high liquefaction hazard in the bedrock geological unit. As the weight of the geological data increases, the geological boundaries become more distinguishable in the generated map. With an appropriately calibrated Markov–Bayes coefficient ($B = 0 \cdot 73$ for the current study), the accuracy of the liquefaction map is improved when validated against prior knowledge and liquefaction observations of the region. Moreover, the advantage of the multiscale liquefaction hazard map is demonstrated through a small site-specific liquefaction hazard analysis.

## NOTATION

| | |
|---|---|
| $a_{max}$ | peak horizontal ground acceleration |
| $B$ | Markov–Bayes coefficient |
| E | expectation operator |
| $E_1, E_0$ | expectation |
| $F_L$ | function of the factor of safety |
| $g$ | gravitational acceleration |
| $h$ | distance vector |
| $M_w$ | moment magnitude of an earthquake |
| $N(h)$ | number of pairs of data located $h$ apart |
| $\sim N(\mu, \Sigma)$ | joint normal distribution |
| $u$ | spatial location vector |
| $z(u)$ | value of data at location $u$ |
| $Z_n$ | random variable to be generated |
| $Z_p$ | vector of generated primary random variable |
| $Z_s$ | vector of generated secondary random variable |
| $z$ | depth (m) |
| $\hat{\gamma}$ | empirical semivariogram |
| $\gamma_m$ | moist unit weight |
| $\gamma_{sat}$ | saturated unit weight |
| $\mu, \mu$ | mean, mean vector |
| $\rho$ | correlation |
| $\Sigma$ | covariance matrix |
| $\sigma$ | standard deviation |

## REFERENCES

Andrus, R. D. & Stokoe, K. H. II (2000). Liquefaction resistance of soils from shear-wave velocity. *J. Geotech. Geoenviron. Engng* **126**, No. 11, 1015–1025.

Baker, J. W., Seifried, A., Andrade, J. E. & Chen, Q. (2011). Characterization of random fields at multiple scales: an efficient conditional simulation procedure and applications in geomechanics. In *Applications of statistics and probability in civil engineering* (eds M. Faber, J. Koehler and K. Nishijima), pp. 347–348. Leiden, the Netherlands: CRC Press/Balkema.

Cetin, K. O., Seed, R. B., Der Kiureghian, A., Tokimatsu, K., Harder Jr, L. F., Kayen, R. E. & Moss, R. E. (2004). Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential. *J. Geotech. Geoenviron. Engng* **130**, No. 12, 1314–1340.

Chen, Q., Seifried, A., Andrade, J. E. & Baker, J. W. (2012). Characterization of random fields and their impact on the mechanics of geosystems at multiple scales. *Int. J. Numer. Analyt. Methods Geomech.* **36**, No. 2, 140–165.

Chen, Q., Wang, C. & Juang, C. H. (2016a). CPT-based evaluation of liquefaction potential accounting for soil spatial variability at multiple scales. *J. Geotech. Geoenviron. Engng* **142**, No. 2, 04015077.

Chen, Q., Wang, C. & Juang, C. H. (2016b). Probabilistic and spatial assessment of liquefaction-induced settlements through multiscale random field models. *Engng Geol.* **211**, 135–149.

Chung, J. & Rogers, J. D. (2010). Simplified method for spatial evaluation of liquefaction potential in the St. Louis area. *J. Geotech. Geoenviron. Engng* **137**, No. 5, 505–515.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *J. Int. Assoc. Math. Geol.* **17**, No. 5, 563–586.

Deutsch, C. V. & Journel, A. G. (1998). *Geostatistical software library and user's guide.* New York, NY, USA: Oxford University Press.

Goovaerts, P. (1997). *Geostatistics for natural resources evaluation.* New York, NY, USA: Oxford University Press.

Hayati, H. & Andrus, R. D. (2008). Liquefaction potential map of Charleston, South Carolina based on the 1886 earthquake. *J. Geotech. Geoenviron. Engng* **134**, No. 6, 815–828.

Heidari, T. & Andrus, R. D. (2010). Mapping liquefaction potential of aged soil deposits in Mount Pleasant, South Carolina. *Engng Geol.* **112**, No. 1, 1–12.

Helley, E. J. & Graymer, R. W. (1997). *Quaternary geology of Alameda County, and parts of Contra Costa, Santa Clara, San Mateo, San Francisco, Stanislaus, and San Joaquin counties, California: A digital database,* Technical Report USGS Open-File Report 97-97. Reston, VA, USA: United States Geological Survey.

Holzer, T. L., Bennett, M. J., Noce, T. E., Padovani, A. C. & Tinsley, J. C. III (2006). Liquefaction hazard mapping with LPI in the greater Oakland, California, area. *Earthquake Spectra* **22**, No. 3, 693–708.

Idriss, I. M. & Boulanger, R. W. (2008). *Soil liquefaction during earthquakes.* Oakland, CA, USA: Earthquake Engineering Research Institute.

Iwasaki, T., Tatsuoka, F., Tokida, K. & Yasuda, S. (1978). A practical method for assessing soil liquefaction potential based on case studies at various sites in Japan. *Proceedings of the 2nd international conference on microzonation, San Francisco, CA, USA,* pp. 885–896.

Iwasaki, T., Tokida, K., Tatsuoka, F., Watanabe, S., Yasuda, S. & Sato, H. (1982). Microzonation for soil liquefaction potential using simplified methods. *Proceedings of the 3rd international conference on microzonation,* Seattle, WA, USA, vol. 3, pp. 1310–1330.

Juang, C. H., Liu, C. N., Chen, C. H., Hwang, J. H. & Lu, C. C. (2008). Calibration of liquefaction potential index: a re-visit focusing on a new CPTU model. *Engng Geol.* **102**, No. 1, 19–30.

Juang, C. H., Shen, M., Wang, C. & Chen, Q. (2017). Random field-based regional liquefaction hazard mapping data inference and model verification using a synthetic digital soil field. *Bull. Engng Geol. Environ.,* https://doi.org/10.1007/s10064-017-1071-y.

Kayen, R. E., Mitchell, J. K., Seed, R. B. & Nishio, S. (1998). Soil liquefaction in the east bay during the earthquake. In *The Loma Prieta, California earthquake – liquefaction* (ed. T. L. Holzer), Professional Paper No. 1551-B, pp. 61–86. Denver, CO, USA: United States Geological Survey.

Knudsen, K. L., Sowers, J. M., Witter, R. C., Wentworth, C. M., Helley, E. J., Nicholson, R. S., Wright, H. M. & Brown, K. M. (2000). *Preliminary maps of Quaternary deposits and liquefaction susceptibility, nine-county San Francisco Bay region, California,* Technical Report USGS Open-File Report 00-444. Menlo Park, CA, USA: United States Geological Survey.

Lee, D. H., Ku, C. S. & Yuan, H. (2004). A study of the liquefaction risk potential at Yuanlin, Taiwan. *Engng Geol.* **71**, No. 1, 97–117.

Lenz, J. A. & Baise, L. G. (2007). Spatial variability of liquefaction potential in regional mapping using CPT and SPT data. *Soil Dynamics Earthquake Engng* **27**, No. 7, 690–702.

Liu, C. N. & Chen, C. H. (2006). Mapping liquefaction potential considering spatial correlations of CPT measurements. *J. Geotech. Geoenviron. Engng* **132**, No. 9, 1178–1187.

Liu, F., Li, Z., Jiang, M., Frattini, P. & Crosta, G. (2016). Quantitative liquefaction-induced lateral spread hazard mapping. *Engng Geol.* **207**, 36–47.

Liu, W., Chen, Q., Wang, C. & Juang, C. H. (2017). Spatially correlated multiscale $V_{s30}$ mapping and a case study of the suzhou site. *Engng Geol.* **220**, 110–122.

Maurer, B. W., Green, R. A., Cubrinovski, M. & Bradley, B. A. (2014). Evaluation of the liquefaction potential index for assessing liquefaction hazard in Christchurch, New Zealand. *J. Geotech. Geoenviron. Engng* **140**, No. 7, 04014032.

Maurer, B. W., Green, R. A., Cubrinovski, M. & Bradley, B. A. (2015). Assessment of CPT-based methods for liquefaction evaluation in a liquefaction potential index framework. *Géotechnique* **65**, No. 5, 328–336, http://dx.doi.org/10.1680/geot.SIP.15.P.007.

Moss, R. E., Seed, R. B., Kayen, R. E., Stewart, J. P., Der Kiureghian, A. & Cetin, K. O. (2006). CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential. *J. Geotech. Geoenviron. Engng* **132**, No. 8, 1032–1051.

Moysey, S., Caers, J., Knight, R. & Allen-King, R. M. (2003). Stochastic estimation of facies using ground penetrating radar data. *Stochastic Environ. Res. Risk Assessment* **17**, No. 5, 306–318.

Papathanassiou, G. (2008). LPI-based approach for calibrating the severity of liquefaction-induced failures and for assessing the probability of liquefaction surface evidence. *Engng Geol.* **96**, No. 1, 94–104.

Papathanassiou, G., Valkaniotis, S., Chaztipetros, A. & Pavlides, S. (2017). Liquefaction susceptibility map of greece. *Bull. Geol. Soc. Greece* **43**, No. 3, 1383–1392.

Remy, N., Boucher, A. & Wu, J. (2009). *Applied geostatistics with SGeMS: a user's guide.* Cambridge, UK: Cambridge University Press.

Robertson, P. K. & Wride, C. E. (1998). Evaluating cyclic liquefaction potential using the cone penetration test. *Can. Geotech. J.* **35**, No. 3, 442–459.

Seed, H. B. & Idriss, I. M. (1971). Simplified procedure for evaluating soil liquefaction potential. *J. Soil Mech. Found. Div.* **97**, No. 9, 1249–1273.

Shen, M., Chen, Q., Zhang, J., Gong, W. & Juang, C. H. (2016). Predicting liquefaction probability based on shear wave velocity: an update. *Bull. Engng Geol. Environ.* **75**, No. 3, 1199–1214.

Sonmez, H. (2003). Modification of the liquefaction potential index and liquefaction susceptibility mapping for a liquefaction-prone area (Inegol, Turkey). *Environ. Geol.* **44**, No. 7, 862–871.

USGS (U.S. Geological Survey) (2015). *U.S. geological survey hazards program CPT database.* Reston, VA, USA: United States Geological Survey. See http://earthquake.usgs.gov/research/cpt/ (accessed 19/10/2017).

Vivek, B. & Raychowdhury, P. (2014). Probabilistic and spatial liquefaction analysis using CPT data: a case study for Alameda County site. *Nat. Hazards* **71**, No. 3, 1715–1732.

Wang, C., Chen, Q., Shen, M. & Juang, C. H. (2017). On the spatial variability of CPT-based geotechnical parameters for liquefaction potential evaluation. *Soil Dynamics Earthquake Engng* **95**, 153–166.

Witter, R. C., Knudsen, K. L., Sowers, J. M., Wentworth, C. M., Koehler, R. D., Randolph, C. E., Brooks, S. K. & Gans, K. D. (2006). *Maps of Quaternary deposits and liquefaction susceptibility in the central San Francisco Bay region, California.* Technical Report USGS Open-File Report 06-1037. Reston, VA, USA: United States Geological Survey.

Youd, T. L. & Perkins, D. M. (1987). Mapping of liquefaction severity index. *J. Geotech. Engng* **113**, No. 11, 1374–1392.

Youd, T. L., Idriss, I. M., Andrus, R. D., Arango, I., Castro, G., Christian, J. T., Dobry, R., Finn, W. L., Harder, L. F. Jr, Hynes, M. E., Ishihara, K., Koester, J. P., Liao, S. S. C., Marcuson, W. F. III, Martin, G. R., Mitchell, J. K., Moriwaki, Y., Power, M. S., Robertson, P. K., Seed, R. B. and Stokoe, K. H. II (2001). Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *J. Geotech. Geoenviron. Engng* **127**, No. 10, 817–833.