ASCE

Accelerogram-Based Method for Quick Assessment of Liquefaction Occurrence

Weiwei Zhan, S.M.ASCE[1]; and Qiushi Chen, A.M.ASCE[2]

Abstract: An accelerogram-based method is developed and validated for the quick assessment of liquefaction occurrence based on ground motion records. In this method, two frequency-related ground motion indices, termed RL and MIFr, are defined and extracted from accelerograms using signal-processing techniques. RL and MIFr indicate the richness of the low-frequency components and the temporal variation rate of the mean instantaneous frequency in the ground motion records, respectively. A new liquefaction database consisting of ground motion stations with both ground motion records and the corresponding liquefaction observations is compiled. Logistic regression is used to develop a new liquefaction classification model that takes RL and MIFr as inputs and calculates a liquefaction indicator (LQI) that can be used to assess liquefaction occurrence. The performance of the proposed method is evaluated and compared with existing accelerogram-based liquefaction assessment methods using a common database, and the method is further validated using a new liquefaction data set. The proposed method demonstrated superior performance, with an overall accuracy of 92.8% for the common data set. The proposed method has promising potentials for applications in real-time disaster mitigation systems and rapid postearthquake loss estimations. **DOI: 10.1061/(ASCE)GT.1943-5606.0002577.**

Author keywords: Liquefaction; Accelerogram; Signal processing; Logistic regression; Database.

Introduction

Earthquake-induced liquefaction is one of the leading causes of earthquake damage worldwide. Common approaches of liquefaction triggering assessment include the simplified stress-based procedure, regional liquefaction hazard mapping approaches, laboratory and physical model testing, and computational mechanics–based approaches. Among them, the simplified stress-based procedure (Seed and Idriss 1971; Seed et al. 1985) is the most commonly used method in practice to predict liquefaction triggering. In this procedure, liquefaction triggering is assessed through the comparison of two variables (Youd et al. 2001): (1) cyclic stress ratio (CSR), representing the seismic demand on a soil layer, and (2) cyclic resistance ratio (CRR), representing the capacity of the soil's resistance to liquefaction. CRR is commonly evaluated using field tests such as the standard penetration test (SPT) (Seed and Idriss 1971; Seed et al. 1985; Cetin et al. 2004; Idriss and Boulanger 2008; Boulanger and Idriss 2014; Juang et al. 2012), cone penetration test (CPT) (Robertson and Wride 1998; Moss et al. 2006; Boulanger and Idriss 2014; Ku et al. 2012; Chen et al. 2016a, b), and shear-wave velocity measurements (V_s) (Andrus and Stokoe 2000; Kayen et al. 2013).

Regional liquefaction hazard mapping approaches often rely on surficial geological data and knowledge of past occurrence of liquefaction (Youd and Perkins 1978), or supplement geological data with geomorphological data (Matsuoka et al. 2015) and geotechnical data (Baise et al. 2006; Holzer et al. 2011; Chen et al. 2016a, b; Wang and Chen 2018). Liquefaction assessment methods using broadly available geospatial proxies have also been developed (Zhu et al. 2015, 2017). Regional mapping approaches provide the assessment of liquefaction triggering on a large scale and can be used for rapid response and loss estimation. Laboratory testing (e.g., cyclic triaxial and cyclic direct simple shear tests) and physical model testing (e.g., shaking-table tests and centrifuge tests) allow for the study of the liquefaction triggering mechanism in a controlled environment, with direct measurement of pore-water pressure generation and shear strain development. They can provide valuable insights into the liquefaction.

However, the applications of laboratory and physical model testing are limited due to challenges such as obtaining undisturbed samples and relatively high costs. Computational mechanics–based approaches can predict pore pressure generation within an effective-stress site response framework and have been used to evaluate liquefaction triggering (Matasovic and Hashash 2012). However, continued efforts in terms of model calibration, validation, and result interpretation are needed for this type of approach (NASEM 2016).

Recently, accelerogram-based approaches that determine the occurrence of liquefaction based on only ground motion records have been developed (Suzuki et al. 1998; Miyajima 1998; Ozaki 1999; Kostadinov and Yamazaki 2001; Yuan et al. 2010; Özener et al. 2020). Accelerogram-based methods' real-time application is promising as inexpensive and broad networks of ground motion monitors are developed in the fields of both earthquake engineering and seismology, such as low-cost accelerometers for structural health monitoring, the community seismic network (Clayton et al. 2015), and the smartphone seismic network MyShake (Kong et al. 2016). The SUPREME system in the Tokyo urban area (Shimizu et al. 2000) shows a good example of applications of accelerogram-based methods in quick response to liquefaction hazard for large and wide pipeline networks.

Basically, given ground motion records (which have been increasingly available), one can use an accelerogram-based method to quickly (in real-time) determine if liquefaction has occurred or

[1]Graduate Research Assistant, Glenn Dept. of Civil Engineering, Clemson Univ., Clemson, SC 29631.

[2]Associate Professor, Glenn Dept. of Civil Engineering, Clemson Univ., Clemson, SC 29631 (corresponding author). ORCID: https://orcid.org/0000-0002-0394-6710. Email: qiushi@clemson.edu

Note. This manuscript was submitted on June 22, 2020; approved on April 5, 2021; published online on May 25, 2021. Discussion period open until October 25, 2021; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Geotechnical and Geoenvironmental Engineering***, © ASCE, ISSN 1090-0241.

not at discrete locations. Accelerogram-based approaches typically extract some characteristics of ground motion record (termed as ground motion indices), and use some classification models (e.g., threshold values), to determine if liquefaction has occurred. Suzuki et al. (1998) were among the first to propose an accelerogram-based approach for liquefaction occurrence detection. Suzuki et al. (1998) used four ground motion indices, i.e., the horizontal peak ground acceleration (PGA), maximum spectrum intensity, maximum horizontal ground displacement, and zero-crossing period. Liquefaction is considered to have occurred if all four indices exceed certain threshold values. Miyajima (1998) classified the possibility of liquefaction occurrence to low, high, and very high using a point system and four ground motion indices derived from the horizontal and vertical component accelerograms, i.e., the maximum ratio of vertical to horizontal acceleration amplitude, the ratio of low-frequency portion to the whole area of the Fourier amplitude spectrum, the averaged predominant frequency, and the maximum decrease rate of the predominant frequency.

Ozaki (1999) used two horizontal accelerograms and defined a ground motion index, i.e., the ratio of Arias intensity of filtered to nonfiltered acceleration time history, to classify occurrence of liquefaction into no liquefaction, possible liquefaction, and liquefaction. Kostadinov and Yamazaki (2001) processed the three-component accelerograms and used the mean instantaneous frequency and the peak ground velocity to classify sites into no liquefaction, liquefaction suspicion, and liquefaction. Yuan et al. (2010) used the time-frequency decreasing ratio of surface acceleration calculated by the zero-crossing method to identify the occurrence of liquefaction. Özener et al. (2020) compared the modal frequency corresponding to the maximum power at each time step with its fitted step function to determine the liquefaction triggering time.

The effectiveness of an accelerogram-based approach for liquefaction detection depends on both the indices extracted from ground motion records and the classification model that classifies liquefaction occurrence based on the extracted indices. It is important to extract ground motion indices that best characterize the effect of liquefaction on ground motion records. Commonly, ground motion indices are divided into four categories, i.e., amplitude, frequency, energy, and duration. Kostadinov and Yamazaki (2001) evaluated the performance of the amplitude-, frequency-, and energy-related ground motion indices of different accelerogram-based approaches in liquefaction occurrence assessment and found that frequency-related ground motion indices are the most effective ones.

For the classification model, all the existing accelerogram-based approaches use simple threshold values of selected ground motion indices to determine liquefaction occurrence. These threshold values are usually determined from a relatively limited number of ground motion records and liquefaction observations of several earthquakes, which may limit their applicability and effectiveness to new cases. For instance, Miyajima (1998) recommended that the threshold value of the proposed ground motion indices be calibrated when applying the method to different earthquakes.

In this work, a new accelerogram-based method is proposed for the quick assessment of liquefaction occurrence based on observed liquefaction modification effects to the surface accelerograms. A rigorous regression analysis–based classification model is developed using two representative frequency-related ground motion indices as explanatory variables. In order to train the classification model, a comprehensive liquefaction database consisting of 135 ground motion stations with both ground motion records and liquefaction observations is compiled. Receiver operating characteristics (ROC) analysis is used to determine an optimal threshold value when a binary classification is desired. The developed model is then evaluated and compared with existing accelerogram-based methods using a common data set, and is further validated using a new data set to demonstrate its predictive capability and potential.

Characteristics of Ground Motions at Liquefied Sites

The ground motions of liquefied sites have shown some distinct characteristics when compared with those of nonliquefied sites (Kramer et al. 2015). The development of an accelerogram-based liquefaction detection method hinges upon identifying and extracting these characteristics. In this work, two frequency-related characteristics of ground motions of liquefied sites are discussed. As will be shown in subsequent sections, these two characteristics motivate the definition of two ground motion indices used for the proposed liquefaction model development. The north–south (NS) direction accelerogram recorded at the ground surface of the Wildlife Liquefaction Array (WLA) is used to illustrate these two characteristics (Fig. 1). The WLA site was confirmed liquefied during the 1987 M_w 6.6 Superstition Hills earthquake based on the evidence of extensive sand boils in the ground surface and high pore-water pressure ratios measured by installed piezometers (Kramer et al. 2018).

Rapid Frequency Drop

Analyses of accelerogram records have revealed that there is often a rapid drop in ground motion frequency when liquefaction is triggered, and there is an increasing prevalence of low-frequency motions in the latter part of a ground motion (Kramer et al. 2015). Depending on the timing of liquefaction occurrence, this rapid liquefaction-induced frequency drop can start at a different time within an accelerogram. Taking the accelerogram of the WLA site as an example (Fig. 1), a relatively sudden change in frequency contents can be observed at approximately 13–15 s. Afterward, the accelerogram is dominated by low-frequency ground motions.

Kramer et al. (2015) explained the mechanism of this liquefaction-induced rapid frequency drop. When the seismic response of a soil profile and surface ground motion is strongly affected by the stiffness of the underlying soil, the high pore-water pressure reduces the effective stress and thus the stiffness of the underlying soil. This softening effect of liquefaction allows the site to transmit more low-frequency ground motions, an extreme case of which is where fully liquefied soils behave like a fluid that cannot

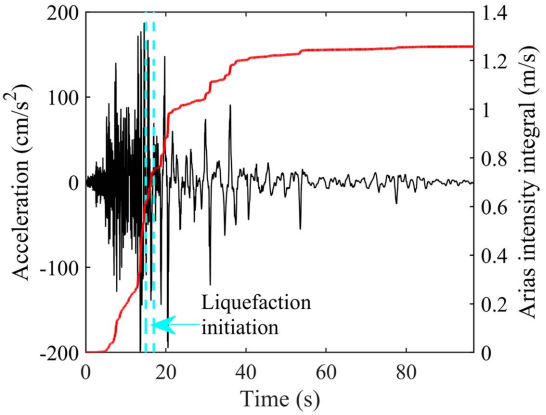

Fig. 1. Example ground motion of a liquefied site, the Wildlife Liquefaction Array, during the 1987 Superstition Hills earthquake. The solid line is the integral to calculate the Arias intensity.

transmit S-waves. The intrinsic frequency drop also occurs when the incoming seismic waves transit from P-waves to S-waves to surface waves. However, their effects on the frequency drop are believed to be less significant than the ones caused by liquefaction. In addition, there are cases where changes in frequency contents may be difficult to detect, in particular, for sites where liquefiable soil is underlain by a soft nonliquefiable soil and the ground surface motion is dominated by low frequencies even before liquefaction occurs (Kramer et al. 2015).

High-Frequency Acceleration Spikes

The second characteristic of accelerogram records of liquefied sites is the acceleration spikes in the accelerogram. These acceleration spikes begin to appear after the initiation of liquefaction, and they are evident at approximately 15–54 s for the WLA site (Fig. 1). These acceleration spikes are called high-frequency acceleration spikes because of the short half-period of their waveform. The mechanism of this type of acceleration spike is the quick soil stiffness recovery and dissipation of pore-water pressure along with soil behavior transitioning from contraction to dilation (Kramer et al. 2015). Hence, these high-frequency acceleration pulses are also called dilatation pulses (Kramer et al. 2015) and cyclic mobility spikes (Bradley and Cubrinovski 2011) because of their mechanical origin.

The amplitude of these high-frequency acceleration spikes may become strong enough to be the PGA of an accelerogram. This amplitude amplification mechanism is interesting and not yet well understood because it contradicts the common perception that the effect of liquefaction is to deamplify the ground acceleration (Bouckovalas et al. 2016). In terms of energy measures of ground motion, these liquefaction-induced high-frequency acceleration spikes can account for a considerable portion of the corresponding Arias intensity. Take the WLA site as an example (Fig. 1), the high-frequency acceleration spikes prevail from approximately 15 to 54 s. During this period, the integral to calculate the corresponding Arias intensity value (represented by the solid line in the figure) increases from 0.6 to 1.2 m/s (or an increase of 0.6 m/s), which accounts for approximately 50% of the final integral value, i.e., the corresponding Arias intensity value.

The aforementioned two characteristics of liquefied accelerograms can be challenging to identify in the ground motion records of some liquefied sites. First, the effects of liquefaction on accelerograms are influenced by many site-specific conditions that affect the liquefaction triggering time and postliquefaction pore pressure dissipation (Kramer et al. 2016). Second, there are intrinsic changes in the frequency contents of typical earthquake motion. For instance, P-waves that arrive first have higher frequencies than S-waves that arrive shortly thereafter; the surface waves that often follow the S-waves have lower frequencies than the S-waves (Kramer et al. 2015). These changes in the frequency contents may obscure the liquefaction-induced effects on ground motions. An effective accelerogram-based method requires careful analyses of frequency characteristics and the use of appropriate indices to quantify them, as will be discussed in the "Proposed Methodology" section.

Proposed Methodology

Flow of the Proposed Procedure

The flow of the proposed procedure to assess liquefaction occurrence using only accelerograms is illustrated in Fig. 2. The horizontal accelerograms are collected to be the input accelerograms

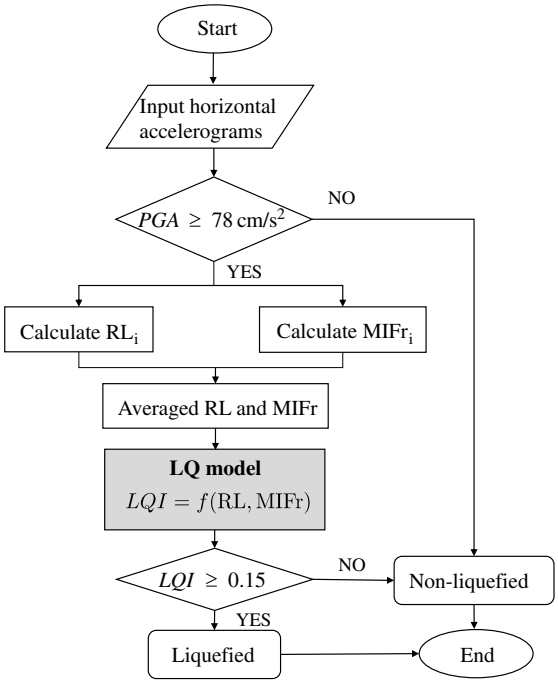

Fig. 2. Flow of the proposed procedure to assess liquefaction occurrence using accelerograms.

because they are more likely to be influenced by liquefaction than vertical components, and the interaction mechanism between vertical and horizontal ground motions are not yet well understood (Miyajima 1998; Kostadinov and Yamazaki 2001). As shown in Fig. 2, the first step in the procedure is to filter out low-intensity ground motions that are not likely going to trigger liquefaction. In this work, a simple PGA threshold value is applied for this purpose. Then, two frequency-related ground motion indices, the ratio of low-frequency portion to the whole area of the Fourier amplitude spectrum (termed RL), and the mean instantaneous frequency decrease rate (termed MIFr), are extracted from each horizontal accelerogram. They are shown as RL_i and $MIFr_i$ in Fig. 2. For each site, a single RL and single MIFr value can be obtained by averaging their corresponding values from the two horizontal accelerograms.

At the core of the procedure is a new liquefaction classification model, which will be detailed in a subsequent section. The liquefaction classification model takes RL and MIFr as inputs and calculates a liquefaction indicator (LQI) that can be used to assess liquefaction occurrence. If a binary classification is desired (i.e., liquefied versus nonliquefied), a threshold value of LQI can be employed. In subsequent sections, key components of the procedure, as well as choices of PGA and LQI threshold values shown in the flowchart, will be presented in detail.

PGA Threshold Value

In the proposed procedure, a PGA threshold value is used to filter out low-intensity ground motions. This is necessary because these ground motions are not likely to trigger liquefaction. Moreover, some low-intensity ground motions or noises may possess similar frequency characteristics of liquefaction-affected ground motions. In seismic codes and recommendations (de Magistris et al. 2013), a PGA threshold value is often used as the liquefaction exclusion criterion. Suzuki et al. (1998) used 100 cm/s^2 as the threshold PGA value because the liquefaction occurrence was suggested likely when

the PGA exceeds this value based on Seed and Idriss's (1971) liquefaction case history database.

According to a recent database of 201 liquefaction case histories (de Magistris et al. 2013), the minimum peak ground acceleration to trigger liquefaction is found to be $0.08g$ (78 cm/s^2). This PGA threshold value is adopted in this work. It is worth mentioning that other indices, such as the peak ground velocity (PGV), have also been used for similar purposes. For instance, Kostadinov and Yamazaki (2001) employed a PGV of 10 cm/s as the threshold value based on practice in Japan.

RL and MIFr Calculations

Commonly, ground motion parameters or indices are divided into four categories: amplitude-, frequency-, energy-, and duration-related ones. Kostadinov and Yamazaki (2001) evaluated the performance of the amplitude-, frequency-, and energy-related ground motion parameters of different accelerogram-based approaches in liquefaction occurrence assessment and found the frequency-related parameters are the most effective ones. In this work, two frequency-related ground motion indices, extracted using the frequency and time-frequency analyses of the original accelerograms, are selected for the purpose of assessing liquefaction triggering. The first index is the ratio of the low-frequency portion to the whole area of the Fourier amplitude spectrum, termed RL. The second index is the mean instantaneous frequency decrease rate, termed MIFr. The calculation of these two indices will be illustrated using the two example accelerograms shown in Fig. 3, one from a liquefied site [the NS component of Station 57 in the compiled database), and the second one from a nonliquefied site [the east-west (EW) component of Station 98 in the compiled database]. Table S1 summarizes all ground motion records in the compiled database.

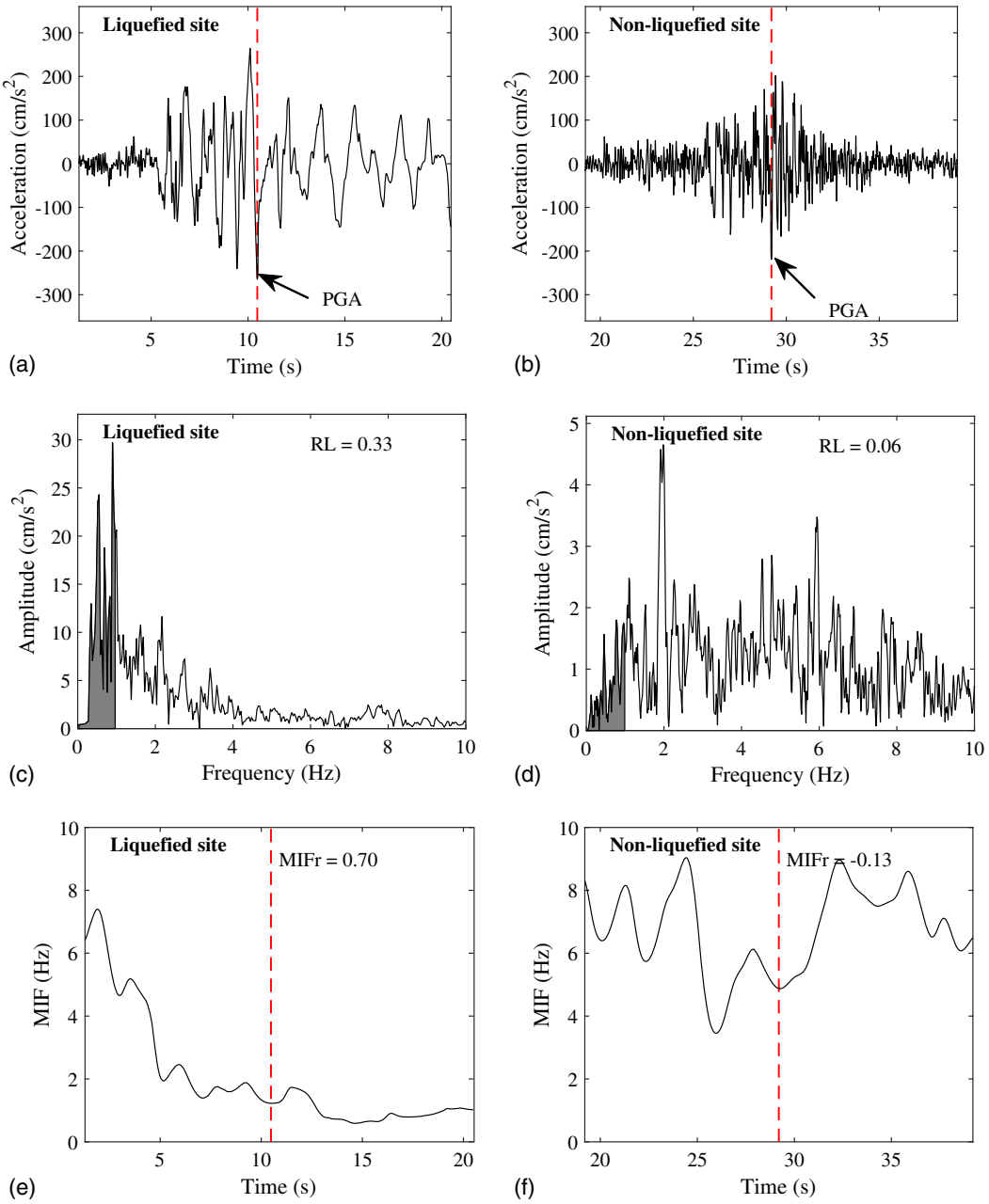

Fig. 3. Acceleration time history, RL, and MIFr plots of a liquefied and a nonliquefied site. The dashed line in plots (a), (b), (e), and (f) indicates the time of PGA. The gray area in plots (c) and (d) represents the portion of the low-frequency components.

Calculation of RL

Accelerograms at liquefied sites have more low-frequency components because the softening effect of liquefaction tends to allow more low-frequency ground motion components to be transmitted to the ground. This low-frequency-rich characteristic of ground motions of liquefied sites was first characterized as the ratio of low-frequency portion to the total area of the Fourier amplitude spectrum (RL) by Miyajima (1998). In order to calculate RL, the accelerogram is firstly decomposed to its frequency domain through the Fourier transform (FT) as follows:

$$p(f) = \int_{-\infty}^{+\infty} a(t) e^{-i2\pi ft} dt \tag{1}$$

where $p(f)$ = Fourier amplitude spectrum of the original acceleration time history $a(t)$; t = vector of time; f = vector of frequency; and $i = \sqrt{-1}$.

Then, RL is calculated using the following equation:

$$\mathrm{RL} = \frac{\int_0^{f_1} p(f) df}{\int_0^{f_2} p(f) df} \tag{2}$$

where f_1 = upper frequency limit of the low-frequency components; and f_2 = upper frequency limit of the whole-frequency components. Rather than using the highest frequency of FT (i.e., half of the sampling frequency of each ground motion record), f_2 is set to 10 Hz to simplify the calculation of RL; f_1 is selected as 1 Hz instead of 2 Hz suggested by Miyajima's (1998) method because many case studies suggest the ground motion frequency after liquefaction initiation is typically less than 1 Hz (Smyrou et al. 2016).

Physically, RL is a ratio of energy measures that indicates the richness of low-frequency components in a ground motion record. RL ranges from 0 to 1, and a large RL means a large portion of low-frequency components. Results of RL calculation for the liquefied and nonliquefied accelerogram are illustrated in Figs. 3(c and d), respectively. It can be easily seen that the liquefied accelerogram has a larger RL than the nonliquefied accelerogram. However, some ground motion records recorded at soft soil sites may have high RL values, which will be discussed in the "Results and Analyses" section.

Calculation of MIFr

The frequency of liquefaction-affected ground motions can drop rapidly after liquefaction initiation due to the significant liquefaction-induced softening effects (Kramer et al. 2015). To characterize the frequency variation of the nonstationary ground motion signals over time, it is necessary to conduct joint time-frequency analysis. Short-time Fourier transform (STFT) is applied here to decompose the acceleration time series into time-frequency space (Rioul and Vetterli 1991) as follows:

$$P(\tau, f) = \int_{-\infty}^{+\infty} a(t) w(t-\tau) e^{-i2\pi ft} dt \tag{3}$$

where $P(t, f)$ = power spectral density or power spectrum obtained through STFT; $a(t)$ = acceleration time history; τ = vector of time corresponding to the midpoint of each time window of STFT; f = vector of frequencies at which the STFT is evaluated; and $w(t-\tau)$ = window function where it is nonzero only in a finite region around time τ, where here a Hamming window of time length of 2.56 s (Kostadinov and Yamazaki 2001) is used.

As shown in Eq. (3), the basic idea of STFT is to apply a moving time window to a signal, e.g., the accelerogram, so that the Fourier transform can be performed on individual intervals of the signal. The mean instantaneous frequency (MIF) is defined as the weighted average of all frequencies present at a given time using the time-frequency representation of the acceleration time history obtained from STFT (Kostadinov and Yamazaki 2001)

$$\mathrm{MIF}(\tau) = \frac{\int_0^{\frac{f_s}{2}} f P(\tau, f) df}{\int_0^{\frac{f_s}{2}} P(\tau, f) df} \tag{4}$$

where f_s = sampling frequency of each ground motion record; and $\mathrm{MIF}(\tau)$ = measure of frequency variation of the ground motion records over time. Then, the mean instantaneous frequency decrease rate (MIFr) can be defined

$$\mathrm{MIFr} = \frac{\mathrm{MIF}_b - \mathrm{MIF}_a}{\mathrm{MIF}_b} \tag{5}$$

where MIF_b and MIF_a = averaged mean instantaneous frequency in the time interval of $[(t_{\mathrm{PGA}} - t_w), t_{\mathrm{PGA}}]$ and $[t_{\mathrm{PGA}}, (t_{\mathrm{PGA}} + t_w)]$, respectively, where t_{PGA} here denotes the time when PGA occurs in the original accelerogram, and t_w is the length of the time interval and is chosen as 10 s in this work to keep the time interval for MIFr calculation representative for as many ground motion records as possible. MIFr ranges from negative values to 1. A large MIFr means the frequency of the ground motions drops rapidly around the time of its amplitude peak. A small or even negative MIFr means the frequency of the ground motions drops slowly or even increases around the time of its amplitude peak, respectively.

For the liquefied and nonliquefied accelerograms shown in Figs. 3(a and b), the MIF variations within the time interval $[(t_{\mathrm{PGA}} - t_w), (t_{\mathrm{PGA}} + t_w)]$ and corresponding MIFr values are shown in Figs. 3(e and f), respectively. The liquefied accelerogram shows a rapid frequency decrease rate (MIFr = 0.70), whereas the nonliquefied accelerogram shows a small negative frequency decrease rate (MIFr = −0.13).

Logistic Regression Analysis

With values of RL and MIFr extracted from ground motion records, logistic regression analysis is applied to assess the liquefaction occurrences. In the logistic regression framework, a liquefaction probability function (P_L) can be defined in terms of the vector of independent variables $\boldsymbol{X}$ as follows (assuming a sigmoid-type function):

$$P_L(\boldsymbol{X}) = \frac{1}{1 + e^{-\boldsymbol{\theta}^T \boldsymbol{X}}} \tag{6}$$

where $\boldsymbol{X}$ = vector of independent variables, e.g., $\boldsymbol{X} = [\mathrm{RL}, \mathrm{MIFr}]$ in this work; and $\boldsymbol{\theta}$ = vector of regression coefficients. The Logit transformation of $P_L(\boldsymbol{X})$ can be expressed

$$\mathrm{logit}[P_L(\boldsymbol{X})] = \ln\left[\frac{P_L(\boldsymbol{X})}{1 - P_L(\boldsymbol{X})}\right] = \boldsymbol{\theta}^T \boldsymbol{X} \tag{7}$$

The form of decision boundary is defined by $\boldsymbol{\theta}^T \boldsymbol{X}$ term in Eqs. (6) and (7). For the case with two independent variables, the simple linear and second-order polynomial functions are investigated in this work [Eqs. (8) and (9), respectively]. Other functional forms can also be easily adopted if needed

$$\boldsymbol{\theta}^T \boldsymbol{X} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \tag{8}$$

$$\boldsymbol{\theta}^T \boldsymbol{X} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 \times x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 \tag{9}$$

The maximum likelihood principle is applied to estimate the regression coefficients $\boldsymbol{\theta}$. The likelihood function that correlates the

Table 1. Summary of earthquake data

ID	Earthquake	Latitude (degrees)	Longitude (degrees)	Depth (km)	M_w	Tect env[a]	Fault type[b]	GM[c]	LQ[d]	References[e]
1	1964 Niigata, Japan	38.38	139.26	10.0	7.6	2	T	1	1	Abe (1995)
2	1968 Tokachi-Oki, Japan	39.84	143.20	13.2	8.2	0	T	3	1	Abe (1995)
3	1978 Miyagiken-Oki, Japan	38.21	142.02	48.5	7.6	0	R	1	0	Abe (1995)
4	1983 Nihonkai-Chibu, Japan	40.38	139.17	16.5	7.7	2	R	2	1	ISC (2020)
5	1985 Michoacan, Mexico	18.14	−102.54	15.6	8.0	0	R	5	0	ISC (2020)
6	1987 Superstition Hills, US	33.02	−115.83	9.0	6.6	2	SS	2	1	VDC (2020)
7	1987 Chibaken-Toho-Oki, Japan	35.45	140.38	45.1	6.5	1	SS	5	0	ISC (2020)
8	1989 Loma Prieta, US	37.04	−121.88	17.5	7.0	2	RO	12	1	VDC (2020)
9	1993 Kushiro-Oki, Japan	42.98	144.34	92.3	7.6	1	N	8	1	Abe (1995)
10	1993 Hokkaido-Nansei-Oki, Japan	42.83	139.23	12.0	7.7	0	T	4	0	Abe (1995)
11	1994 Northridge, US	34.21	−118.55	17.5	6.7	2	R	10	0	VDC (2020)
12	1994 Hokkaido-Toho-Oki, Japan	43.75	147.29	29.2	8.2	1	T	4	0	Abe (1995)
13	1994 Sanriku-Haruka-Oki, Japan	40.52	143.42	27.0	7.7	0	T	1	0	Abe (1995)
14	1995 Hyogoken-Nanbu, Japan	34.59	135.01	17.9	6.9	2	SS	21	11	VDC (2020)
15	1997 Kagoshimaken-Hokuseibu, Japan	31.95	130.34	7.7	5.9	2	SS	4	0	Miyake et al. (2003)
16	2010 Darfield, New Zealand	−43.53	172.17	11.0	7.1	2	R	33	2	GeoNet (2020)
17	2011 Christchurh, New Zealand	−43.58	172.68	5.0	6.2	2	R	21	9	GeoNet (2020)
18	2001 Nisqually, US	47.17	−122.73	52.4	6.8	1	N	32	4	VDC (2020)

[a]Tectonic environment: 0 = interface; 1 = intraslab; and 2 = shallow crustal.
[b]T = thrust; R = reverse; SS = strike slip; N = normal; and RO = reverse oblique.
[c]Total number of ground motion records collected from corresponding earthquake in this work.
[d]Number of liquefied-site records collected from corresponding earthquake in this work.
[e]ISC = International Seismological Centre; and VDC = Strong Motion Virtual Data Center.

independent variables $\boldsymbol{X}$ with the regression coefficients $\boldsymbol{\theta}$ is defined

$$L(\boldsymbol{\theta}|\boldsymbol{X}) = \prod_{i=1}^{m}[[P_L(\boldsymbol{X})]^{y^{(i)}}[1 - P_L(\boldsymbol{X})]^{(1-y^{(i)})}] \tag{10}$$

where $L(\boldsymbol{\theta}|\boldsymbol{X})$ = likelihood function; $y^{(i)}$ = dependent variable for case i; and m = total number of cases. If a binary classification is desired, $y^{(i)} = 1$ for a liquefied case and $y^{(i)} = 0$ for a nonliquefied case. The regression coefficients $\boldsymbol{\theta}$ can be easily solved through maximizing the log-likelihood function

$$\max_{\boldsymbol{\theta}} \ln[L(\boldsymbol{\theta}|\boldsymbol{X})] = \max_{\boldsymbol{\theta}} \sum_{i=1}^{m}[y^{(i)} \ln P_L(\boldsymbol{X}) + (1 - y^{(i)}) \ln[1 - P_L(\boldsymbol{X})]] \tag{11}$$

To develop and validate the logistic regression classification model, a new liquefaction case history database will be compiled and will be detailed in the next section.

Database of Liquefaction and Ground Motion Records

In this section, a specifically designed liquefaction case-history database consisting of ground motion stations with both ground motion records and the corresponding liquefaction observations is compiled. Whether or not a ground motion record is from a liquefied site is inferred from postearthquake reconnaissance reports and published papers. Surface manifestations such as sand boils, ground fissures filled with sand, large permanent displacements or vertical settlements of the soil, uplifting of pipelines or tanks, tilting of buildings, and some foundation failures are used as indicators of liquefaction occurrence. Depending on the soil profile, however, some liquefaction evidence such as sand boils may not be observed on the surface, when liquefaction occurs at a relatively large distance below the surface (Kostadinov and Yamazaki 2001).

Kostadinov and Yamazaki (2001) compiled such a database covering 15 major earthquakes that occurred between 1964 and 1997. The ground motion records were classified into the three groups based on the following criteria:

- Liquefied sites (LQ): there was clear evidence of liquefaction manifestation at the recording site.
- Liquefaction-suspicious sites (SUS): there was no direct evidence of surface manifestation of liquefaction at the location of the recording site. However, liquefaction was observed at close vicinity (within 50 m) of the recording site or cyclic mobility at the site was confirmed by an analytical study.
- Nonliquefied sites (NonLQ): there was no evidence of surface manifestation of liquefaction at the close vicinity (up to 50 m) of the recording site, nor was there any confirmation of cyclic mobility.

The database of Kostadinov and Yamazaki (2001) (termed the K-Y database) includes 11 liquefied sites, 6 liquefaction-suspicious sites, and 66 nonliquefied sites. In this work, a new liquefaction database is developed that extends the K-Y database to include four more recent earthquakes with 109 more ground motion records and liquefaction observations: the 2002 M_w 6.8 Nisqually earthquake, the 2010 M_w 7.1 Darfield earthquake, and the 2011 M_w 6.2 Christchurch. Table 1 summarizes the earthquake information included in the new database. The total number of ground motion records is 167, and the number of LQ records is 32.

In the new database, only ground motion stations with liquefaction evidence confirmed by literature are labeled as liquefied (LQ). Nonliquefied sites (labeled as NonLQ) are either confirmed by literature or interpreted from remote-sensing images. Because the proposed method aims at a binary classification, the six liquefaction-suspicious sites (SUS) in the K-Y database are relabeled as liquefied (LQ) in this database because there is strong evidence of liquefaction for most SUS sites confirmed by other literature. For instance, the record from the Treasure Island site during the 1989 Loma Prieta earthquake was classified as SUS in the K-Y database because the surface manifestation of liquefaction were found within 100 m of the seismic station, whereas many studies had shown this site was

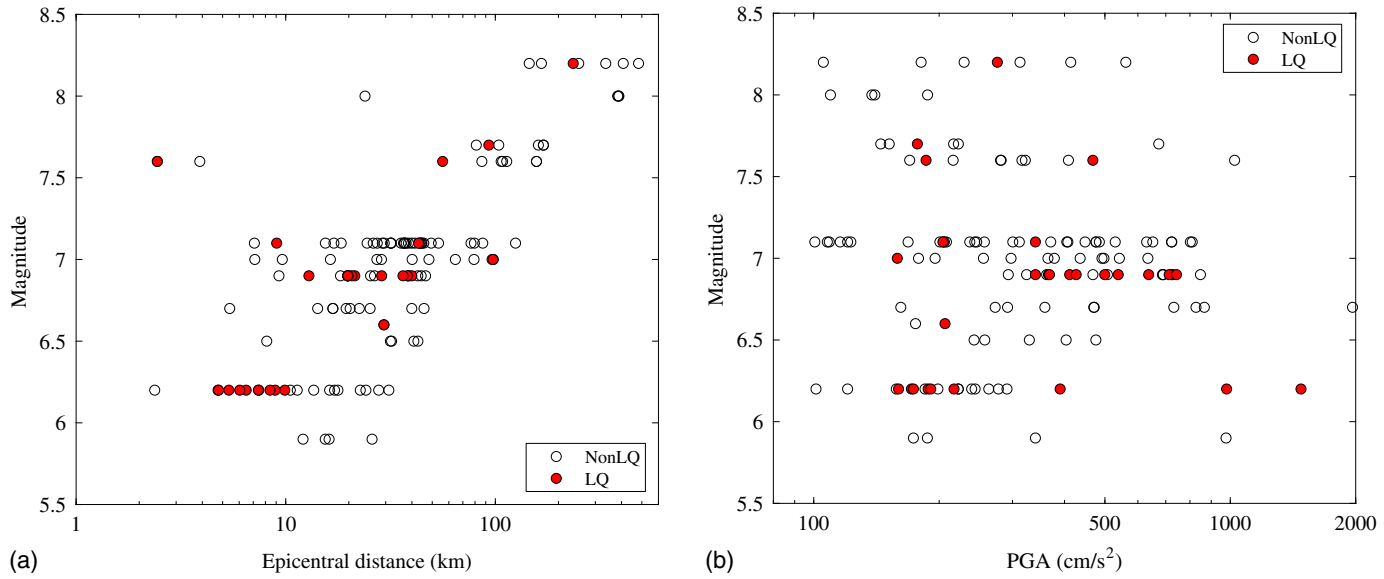

Fig. 4. Magnitude, epicentral distance, and PGA distribution of ground motion records in the training data set.

liquefied. Besides, two weak ground motion records (Nos. 11 and 41) from the K-Y database were removed because they have peak ground acceleration less than 78 cm/s^2.

To evaluate the predictive capability of the proposed accelerogram-based liquefaction assessment method, the new database is split into training and validation datas ets. Data from Nos. 1–17 earthquakes are used as the training data set. Data from the 2001 M_w 6.8 Nisqually earthquake (No. 18 in Table 1) are used as the validation data set because detailed liquefaction observations and ground motions recorded by a dense ground motion network in the Seattle Basin are available for this earthquake (Frankel et al. 2002). In total, the training data set includes 135 sites, and among them, 28 are liquefied sites. Complete details of the data in the training data set are summarized in Table S1.

The source earthquake magnitude of the ground motions in the training database ranges from M_w 5.9 to 8.2. The epicentral distance (R_{epi}) ranges from 2 to 410 km. The PGA ranges from 101 to 1,965 cm/s^2. The relationships between epicentral distance and PGA with earthquake magnitude are shown in Figs. 4(a and b), respectively. The validation data set includes 32 sites in total, and among them, four are liquefied sites. The validation data set will be used for evaluating the performance of the proposed method.

Results and Analyses

In this section, the methodology and the database described in previous sections are used to develop and validate a liquefaction classification model. The final form of the developed model is presented first. The performance of the proposed method is evaluated and compared with existing accelerogram-based methods using a common data set. Finally, the model is validated using a new data set, the 2001 M_w 6.8 Nisqually earthquake, which is not included in the training data, to demonstrate its potential and effectiveness.

Developed Liquefaction Classification Model

The training database compiled in the "Database of Liquefaction and Ground Motion Records" section consists of ground motion records from 28 liquefied sites (labeled as LQ) and 107 nonliquefied sites (labeled as NonLQ). Values of the two ground motion indices, RL and MIFr, are calculated for all ground motion records in the training database. The histograms of RL and MIFr for LQ and NonLQ cases are shown in Fig. 5, and the relative frequency is defined as number of occurrences over the total number in that category (i.e., LQ and NonLQ). Values of RL range from 0.026 to 0.816 with a mean of 0.175. Values of MIFr range from −0.168 to 0.924 with a mean of 0.386. It can be seen from Fig. 5 that liquefied sites tend to have larger values of RL and MIFr, as discussed previously. The complete results for all sites in the training data set are included in Table S1.

The scatter plot of RL versus MIFr is shown in Fig. 6. It can be seen from Fig. 6 that there is a clear clustering of LQ and NonLQ cases on this RL-MIFr plot, which means that it is possible to develop a liquefaction classification model using these two indices. Instead of manually drawing a boundary line to separate the LQ and NonLQ data in Fig. 6, this work applies the logistic regression classification method described in the "Proposed Methodology" section to make the decision boundary more robust. As mentioned in the "Proposed Methodology" section, linear, and simple nonlinear (second-order polynomials) decision boundaries are tested in this work. The optimal decision boundary is selected based on both the model accuracy and its physical implications. The result of the logistic regression analysis is boiled down to the following equation:

$$\mathrm{LQI} = \frac{1}{1 + e^{6.44 - 47.61 \times \mathrm{RL} \times \mathrm{MIFr}}} \tag{12}$$

where LQI is calculated based on the two indices (RL and MIFr) extracted from accelerograms. LQI is analogous to the probability of liquefaction (P_L) used in many simplified stress-based approaches for liquefaction assessment (Cetin et al. 2004; Juang et al. 2012). Eq. (12) is a continuous function that gives an LQI value between 0 and 1. Its three-dimensional (3D) visualization is shown in Fig. 7. If a binary classification is desired (i.e., liquefied versus nonliquefied), a threshold value of LQI is needed. In this work, ROC analysis (Fawcett 2006) is used to select the optimal LQI threshold, which yields an optimal LQI threshold value of 0.15, as shown in Fig. 6. Details of the ROC analysis will be presented in the next section.

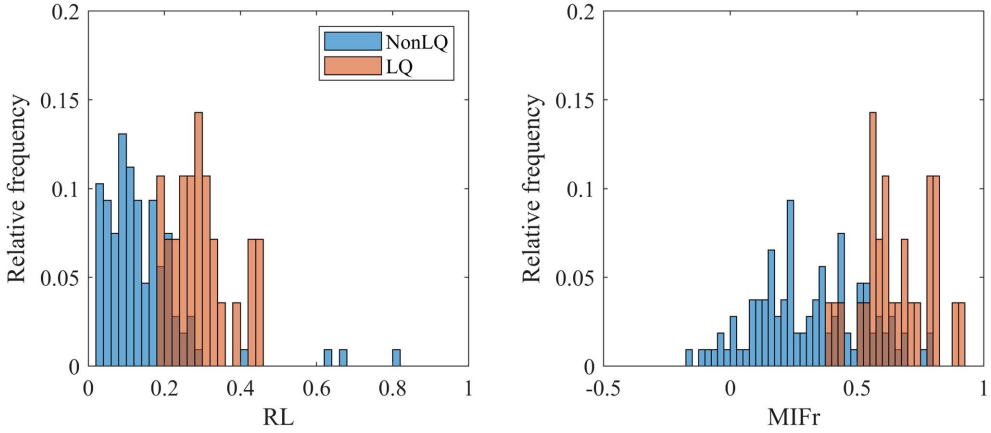

Fig. 5. Histograms of RL and MIFr values of ground motion records at liquefied and nonliquefied stations in the training data set.

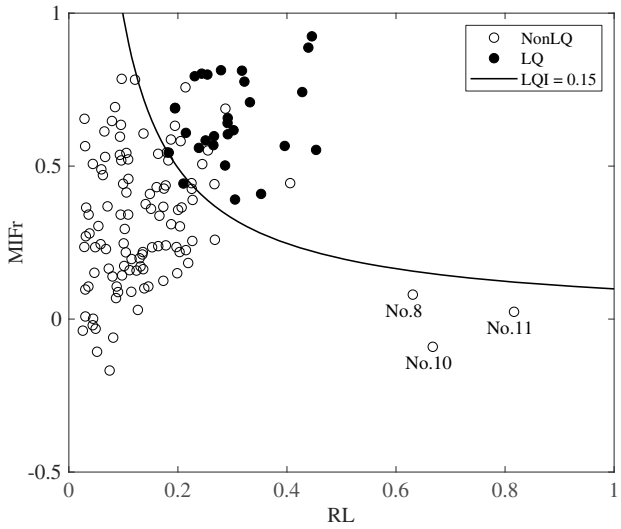

Fig. 6. Curve recommended for quick assessment of liquefaction occurrence based on RL and MIFr derived from accelerograms along with liquefaction case history data.

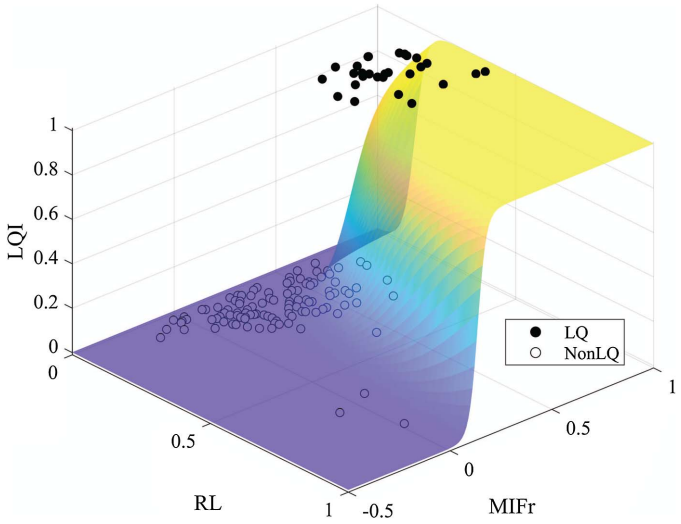

Fig. 7. Three-dimensional visualization of the proposed logistic regression model [Eq. (12)] for liquefaction classification.

ROC Analysis and Choice of LQI Threshold

ROC analysis can be used to evaluate the model performance and select the optimal decision threshold value (Fawcett 2006). As shown in Fig. 8(a), in the ROC analysis, a 2×2 confusion matrix (also called a contingency table) can be constructed representing the four prediction outcomes of the classification model, namely, true positive (TP, i.e., liquefaction is predicted and observed), false positive (FP, i.e., liquefaction is predicted but not observed), true negative (TN, i.e., liquefaction is not predicted and not observed), and false negative (FN, i.e., liquefaction is not predicted but is observed). This confusion matrix forms the basis for other performance metrics, the examples of which are also given in Fig. 8(a).

The ROC curve in Fig. 8(b) plots the true positive rate (TPR) and false positive rate (FPR) for varying LQI threshold values. The diagonal of the ROC graph is interpreted as random guessing, and classification models that fall below the diagonal are considered as worse than random guessing. A perfect classifier plots as a point at (0, 1). The area under the ROC curve (AUC) is equivalent to the probability that liquefied sites have higher computed LQI than nonliquefied sites. The range of AUC is from 0 to 1.0, and a lager AUC value indicates a better model performance (Maurer et al. 2015). The AUC of the proposed logistic regression model [Fig. 8(b)] is 0.977, indicating the model's good performance. The optimum decision threshold is defined herein as the threshold LQI value that minimizes the rate of misprediction [i.e., FPR + (1 − TPR)] (Maurer et al. 2015). Using this criterion, a threshold LQI of 0.15 is selected as the optimum decision threshold for a binary classification against the training data set [Fig. 8(c)]. When LQI = 0.15 is used as the decision threshold, the proposed logistic regression classification model has a value of TPR = 0.9643 and FPR = 0.0935.

The histogram of the classification results using LQI = 0.15 is shown in Fig. 8(d). The success rate of the proposed model is 96.4% (or 27 out of 28) for LQ sites, and 90.7% (or 97 out of 107) for NonLQ sites, respectively. The only incorrect prediction of LQ sites (i.e., the FN case) is from the CBGS station in the 2011 Christchurch earthquake (Station 115 in Table S1).

Overall, the model is demonstrated to be effective and shows promising potential for broader applications because the liquefaction cases included in the database are from different regions worldwide, covering a wide earthquake magnitude ranges (from M_w 5.9 to 8.2) and different fault mechanisms (Table 1). There is a data

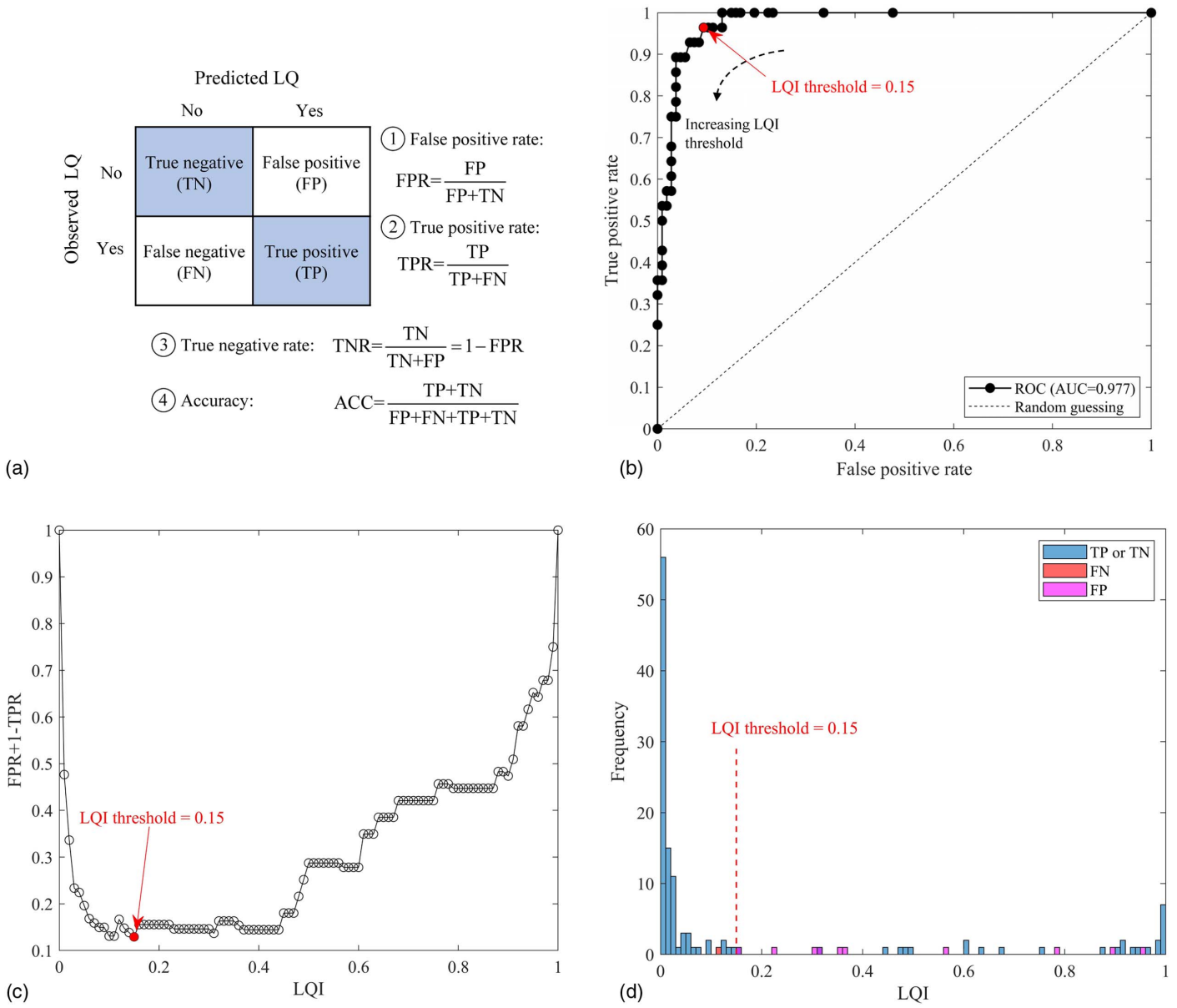

Fig. 8. ROC analysis: (a) illustration of the confusion matrix and definitions of four performance metrics; (b) ROC curve; (c) LQI value versus misprediction rate; and (d) histogram of four model prediction outcomes.

scarcity in the top-right and bottom zones of Fig. 6. For example, only three ground motion records (Nos. 8, 10, and 11 in Table S1) are plotted in the bottom zone of Fig. 6, and they were recorded at soft old lake bed in Mexico City, Mexico, during the 1985 Michoacán earthquake and indicated the mechanism of resonant excitation of the shallow sedimentary structure (Anderson et al. 1986). More ground motion records will be needed to fill in the data scarcity in these zones in the future to further validate the proposed model.

Model Evaluation: K-Y Data Set

In this section, the performance of the proposed model is evaluated and compared with four existing accelerogram-based liquefaction assessment methods using a common data set compiled by Kostadinov and Yamazaki (2001), termed the K-Y data set. In the K-Y data set, there are 11 LQ sites, 66 NonLQ sites, and 6 SUS sites, bringing the total number of sites to 83. The four existing methods are the Suzuki method (Suzuki et al. 1998), the Miyajima method (Miyajima 1998), the Ozaki method (Ozaki 1999), and the Kostadinov-Yamazaki (K-Y) method (Kostadinov and Yamazaki 2001). The proposed method and Suzuki method both use a two-level classification (LQ or NonLQ), whereas the other three methods use a three-level classification (LQ, SUS, or NonLQ). Because the liquefaction observations in Kostadinov and Yamazaki (2001) were grouped into three classes, the prediction accuracy of the Miyajima, Ozaki, and K-Y methods are quantified through directly comparing the prediction results with the corresponding liquefaction observations. For the proposed and the Suzuki methods, the SUS cases are considered as liquefied observations (LQ) because there was relatively strong evidence of liquefaction confirmed by other studies. In other words, if the liquefaction-suspicious sites are predicted as liquefied in the proposed and the Suzuki methods, they are counted as correct predictions.

The performance of the proposed method and the four existing accelerogram-based methods is summarized in Table 2 in terms of

J. Geotech. Geoenviron. Eng., 2021, 147(8): 04021060

Table 2. Performance evaluation of the proposed and four existing accelerogram-based methods using the K-Y data set

Performance metric	Suzuki	Miyajima	Ozaki	K-Y	Proposed method
FPR (%)	9.1	28.8	24.2	9.1	9.1
TPR (%)	76.5	100	100	94.1	100
TNR (%)	90.9	71.2	75.8	90.9	90.9
ACC (%)	88.0	77.1	80.7	91.6	92.8

Note: Definitions of the performance metrics are given in Fig. 8(a).

the performance metrics defined in Fig. 8(a). The proposed method yielded the highest overall accuracy, followed by the Suzuki method, K-Y method, Ozaki method, and Miyajima method. The TPR and TNR of the proposed method is 100% (17 out of 17), and 90.9% (60 out of 66), respectively. The results of the existing four methods are reported in Table 2 in Kostadinov and Yamazaki (2001), and the correct NonLQ predictions by the proposed method include the two low-intensity sites predicted correctly because of the applied PGA threshold as discussed previously in the "Proposed Methodology" section.

Model Validation: 2001 M_w 6.8 Nisqually Earthquake Data Set

In this section, the proposed model is validated using a new data set not included in the training data set. This new data set consists of 32 ground motion records recorded at the Seattle Basin during the M_w 6.8 Nisqually earthquake on February 28, 2001. Information of all 32 ground motion records is summarized in Table S2, and their locations are marked on the map in Fig. 9. Locations of ground motion stations are from Frankel et al. (2002), and locations of liquefaction observations are extracted from Geotechnical Extreme Events Reconnaissance (GEER) Association renaissance reports (Bray et al. 2001). A total of four stations were reported by Frankel et al. (2002) to have liquefied, i.e., Stations SDS, SDW, HAR, and BOE. Among them, there was liquefaction just a few meters from Station SDS and liquefaction beneath the basement where SDW

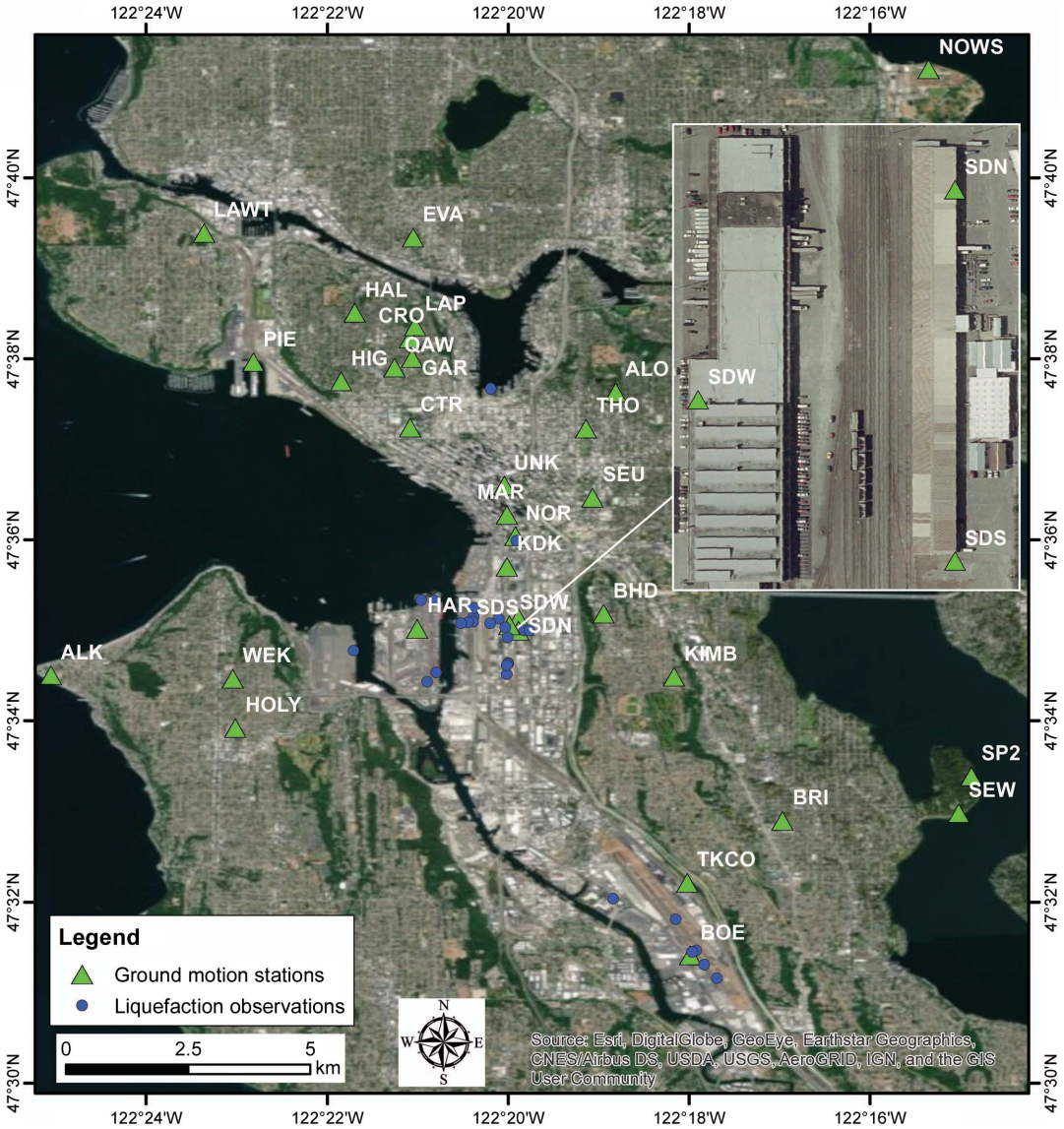

Fig. 9. Map showing the location of ground motion stations and liquefaction observationsin the Seattle Basin (zoom-in view of the SD array shown in the inset). (Map sources: Esri, DigitalGlobe, GeoEye, Earthstar Geographics, CNES/Airbus DS, USDA, USGS, AeroGRID, IGN, and the GIS User Community.)

J. Geotech. Geoenviron. Eng., 2021, 147(8): 04021060

was installed. A sand boil erupted adjacent to HAR. For Station BOE at Boeing Field, extensive liquefaction from the Nisqually mainshock was reported for this area.

The epicentral distance of these ground motion stations ranges from 51 to 68 km with an average value of 57 km, and the PGA values range from 32.6 to 408.0 cm/s^2 with an average value of 193.9 cm/s^2. The large distance from the recording sites to the hypocenter of the mainshock and the small distance between different recording sites imply that the variation of these ground motions is more likely caused by local site effects and basin effects than the rupture directivity and focal-mechanism effects (Frankel et al. 2002). This Nisqually earthquake data set makes a great validation case for the proposed classification model because the ground motion records are recorded by a rare dense array of ground motion stations located close to each other, and the variation is more likely caused by local site effects.

RL and MIFr of all 32 ground motion records are calculated and plotted in Fig. 10 along with the proposed model as the boundary between liquefied and nonliquefied cases. The proposed model successfully predicted two out of four liquefied cases, and 27 out of 28 nonliquefied cases, respectively. The three sites with incorrect predictions are the HAR, EVA, and SDS sites, and all are marked in Fig. 10. These three incorrect predictions are worthy of a closer investigation.

First, site HAR is a false negative case (i.e., liquefaction is not predicted but is observed). Frankel et al. (2002) reported that a sand boil erupted adjacent to the HAR site shortly after the earthquake, which implies that the liquefaction effect on the ground motion at this location could be delayed. Moreover, the HAR can be considered a marginal case in the model prediction because it is located right on the model boundary line, as shown in Fig. 10.

Site EVA is a false positive case (i.e., liquefaction is predicted but is not observed). EVA is a stiff-soil site that had been affected by the basin effect that induced longer-period ground motions, as analyzed by Frankel et al. (2002). The temporal overlap of the low-frequency waves induced by basin effect and the high-frequency S-phase waves caused the rapid drop of mean instantaneous frequency at EVA site (MIFr = 0.584), leading to its misclassification.

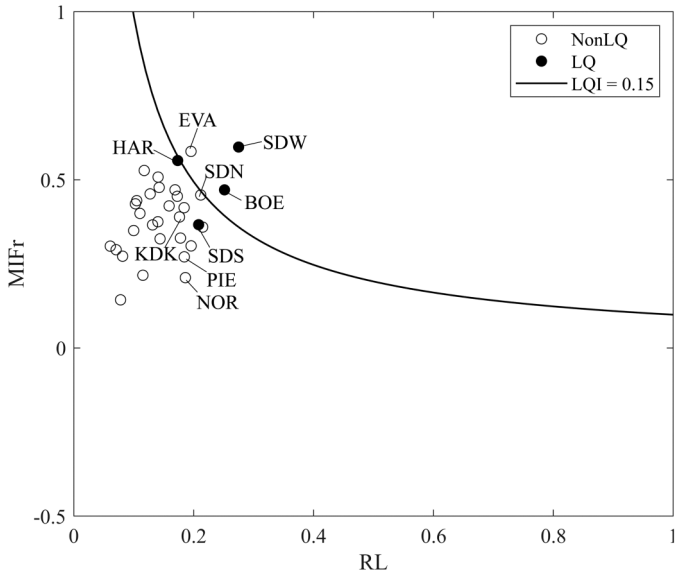

Fig. 10. Classification results of the 2001 Nisqually earthquake data set.

The third incorrect prediction, the SDS site, is a false negative (i.e., liquefaction is not predicted but is observed). This false prediction is mainly caused by the abnormal north–south accelerogram component (Fig. 11 shows both the north–south and east-west accelerogram components at the SDS site). This one-sided cusped waveform may be induced by the dynamic soil–structure interaction because the overlaid structure is 200 m long in the north–south direction.

Furthermore, SDN, SDS, and SDW were deployed in a triangular array (denoted as the SD Array and shown in the inset of Fig. 9), with a distance of 200 m apart, all located on the artificial fill south of the Seattle, Washington, downtown area. The recorded PGAs are 243, 356, and 408 cm/s^2 for SDN, SDS, and SDW, respectively (Table S2). This large variation of ground motion amplitudes among such a small area is an indication of different subsurface soils and potential soil–structure interaction effects.

Finally, the proposed method also demonstrated its robustness when applied to soft-soil sites. In the 2001 Nisqually earthquake validation data set, there are eight soft-soil sites: HAR, NOR, BOE, KDK, and PIE, and the three sites of the SD Array (SDN, SDS, and SDW). Among them, four were observed to be liquefied (HAR, BOE, SDS, and SDW), and four were nonliquefied (NOR, KDK, PIE, and SDN). These soft-soil sites are located on the artificial fill and young alluvium, and most of them have a V_{s30} value of less than 180 m/s [i.e., a National Earthquake Hazards Reduction Program (NEHRP) Class E]. Among them, the two misclassified sites (HAR and SDS) have been discussed. It is worth pointing out that the model successfully predicted all four nonliquefied sites. In other words, none of these soft-soil nonliquefied sites are misclassified as liquefied sites, although there are significant low-frequency-wave components in their ground motions.

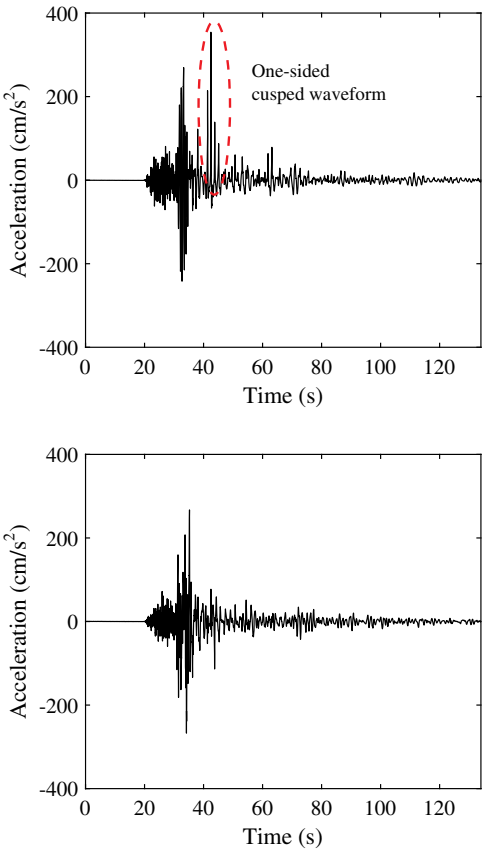

Fig. 11. Acceleration time histories recorded by the SDS station during the 2001 Nisqually earthquake.

Discussion

The accelerogram-based method developed in this study has demonstrated superior performance for quick assessment of liquefaction occurrence. There remain, however, challenges with respect to class imbalance, liquefaction observation uncertainty, and potential data biases. The effects of these challenges on model performance are analyzed and discussed in this section, and the supporting figures and tables are included in the Supplemental Materials.

Class Imbalance Effect

Class imbalance problem refers to the problem where the total number of data in the positive class (e.g., liquefied) is far less than the total number of data in the negative class (e.g., nonliquefied). Class imbalance has a strong influence on the probabilities of a logistic regression model (Zhu et al. 2017). The class imbalance of the training data set is clearly shown in Fig. 6, where the ratio of LQ to NonLQ records is 28:107 (or about 1:4). To investigate the effect of the class imbalance problem, a downsampling strategy is implemented. In the process, all 28 LQ sites in the training data set are kept and a random sampling of 28 NonLQ sites from the 107 NonLQ sites is conducted. Then, a logistic regression model is trained, and the corresponding LQI threshold value is determined following the same procedure described in the previous methodology sections. This process is repeated 10 times, i.e., 10 random samplings of 28 NonLQ sites are taken. Coefficients of the logistic regression and the performance metrics of the corresponding classification models are summarized in Table S3. It can be seen that the reported AUC and accuracy are preserved for all 10 downsampling realizations. Then, the averaged regression coefficients and LQI threshold are used as the parameters of a new classification model. The classification results of this new model are shown in Fig. S1 (LQI threshold is rounded to 0.5 in the figure).

Comparing Fig. S1 (results of the new model) and Fig. 6 (results of the original model), it can be observed that the decision boundary lines of the original and new classification models are quite similar. Their main difference is the LQI threshold. Because the LQI threshold is determined from the ROC analysis (described in the "ROC Analysis and Choice of LQI Threshold" section), it is expected that the LQI threshold can be influenced by the model coefficients and the training data set. The new model with a LQI threshold of 0.5 leads to one more false negative prediction and two fewer false positive predictions.

The performance of the original model and the model trained using the downsampling method is further analyzed using performance metrics including TPR, FPR, and accuracy (ACC). Table S4 summarizes the comparison of the two models evaluated against the training and validation data sets. For the training data set, ACC of the new model increases slightly (from 0.919 to 0.926), whereas TPR decreases (from 0.964 to 0.929). TPR, FPR, and ACC are the same when the new and original models are evaluated against the validation data set. For this work, a higher TPR value is desired because a false negative prediction may lead to severe consequence (i.e., failure to alarm the liquefaction may result in more severe loss than to falsely alarm the liquefaction in a real-time liquefaction hazard response system). Therefore, the original model [Eq. (12)] is recommended over the one built on the downsampling strategy.

Liquefaction Observation Uncertainty

A liquefaction database may contain false negatives and false positives due to liquefaction observation uncertainty (Boulanger and Idriss 2016). To investigate the potential effects of this uncertainty on the proposed model, the six liquefaction-suspicious sites (SUS), which were labeled as liquefied (LQ) in the model development, are relabeled as nonliquefied (NonLQ) in this section. A new model is then trained using the same method (but with these six SUS sites labeled as NonLQ). The performance metrics of the original and new models are compared in Table S5 and visualized in Fig. S2 for both the training and validation data sets. The results suggest the liquefaction observation uncertainty of the six SUS sites slightly affect the classification boundary line, and the new performance metrics are comparable to the original model.

Potential Sources and Effects of Data Biases

The frequency-related ground motion parameters (RL and MIFr) and the performance of the proposed model [i.e., Eq. (12)] are subject to the influence of data bias of various factors such as nearby structures, earthquake source, and tectonic environment, which are investigated and discussed in this section.

Nearby Structures

Structures and foundations could affect the liquefaction triggering and liquefaction-related behavior in the soil (Bray and Macedo 2017; Bullock et al. 2019; Karimi et al. 2018). The observation data in the compiled database are grouped into the free-field observations (total of 78) and observations near structures (total of 57). The type of observations is determined based on information reported by Kostadinov and Yamazaki (2001) (for the first 81 records in Table S1) and by Kaiser et al. (2012) (for the New Zealand earthquake data). Analyses are then performed on each data type.

Fig. S3 plots the histograms of RL and MIFrs for free-field and near-structure ground motions. Both free-field and near-structure histograms suggest higher RL and MIFr values for liquefied ground motion records, which is consistent with the observations when grouping all liquefaction observations together. There is no clear trend separating the free-field histograms from the near-structure histograms. Fig. S4 shows the liquefaction classification results for the training data set separated by types of liquefaction observation (free-field versus near-structure) and instrument. The performance metrics are summarized in Table S6.

These analyses suggest that the calculated RL and MIFr values, and the proposed model, is not significantly impacted by the type of ground motions, i.e., free-field and near-structure ground motions (Fig. S3). In other words, these results suggest that the structures do not severely modify the two frequency-related ground motion parameters. On the other hand, however, a nearby structure is known to significantly affect ground deformations of liquefiable sites (Bray and Macedo 2017; Bullock et al. 2019; Karimi et al. 2018). The effect of nearby structure on other ground motions parameters would warrant future studies in this area.

Earthquake Event

To make sure the proposed model is not biased toward a specific earthquake, the AUCs and performance metrics (TPR, FPR, and ACC) of the model when applied to individual earthquake events are summarized in Table S7. Events with only liquefied or with only nonliquefied observations are excluded because the calculation of AUC requires both liquefaction and nonliquefaction. Events with total observation number under 10 are also excluded in the table because the limited sample size will introduce large variability of classification model performance metrics. The model performs well on the four earthquakes (out of the 17 earthquakes in the training data set) with AUC values greater than 0.6.

Fault Type

The effect of fault type on the model performance is evaluated and summarized in Table S8. Most of the ground motions in the training data set are from the earthquakes occurred on reverse faulta (71/135), followed by strike slip (32/135). According to Table S8, the developed model performs well on these two fault types with AUC values greater than 0.97. The model has better performance when applied to ground motion records induced by normal and thrust fault (AUCs are 1) than those induced by a reverse oblique fault. However, because the number of ground motions of earthquakes occurred on normal, thrust, and reverse oblique faults are relatively small in the training data set (8, 12, and 12, respectively) (Fig. S5), additional data will be needed to further evaluate the fault type effects on the model performance.

Tectonic Environment

The tectonic environment of the earthquakes in Table 1 has been grouped by three classes: interface, intraslab, and shallow crustal, according to Bozorgnia et al. (2020) and Gingery (2014). The exceptions are the 1994 Northridge and 1997 Kagoshimaken earthquakes (Nos. 11 and 15 events in Table 1), which were not reported in the literature, and are classified to be in shallow crustal tectonic environment in this work following the procedure described by Bozorgnia et al. (2020). The effect of tectonic environment is then evaluated and summarized in Table S9. Most of the ground motions in the training data set are from the shallow crustal environment (106/135). But according to Table S9, the developed model has equal or even better performance when applied to ground motion records from the interface and intraslab earthquakes (AUCs are 1). However, because the number of ground motions of earthquakes in the interface and intraslab environments are relatively small in training data set (12 and 17, respectively) (Fig. S6), additional data will be needed to further evaluate the performance and applicability of the proposed method for different tectonic environments.

Conclusions

In this work, an accelerogram-based method has been developed and validated for the quick assessment of liquefaction occurrence based on ground motion records. Two frequency-related ground motion indices, termed RL and MIFr, are extracted from accelerograms to quantify the effects of liquefaction on ground motions. A new liquefaction database is compiled, consisting of 135 companion ground motion records and liquefaction observations from 17 major earthquakes worldwide. Logistic regression is used to develop a new liquefaction classification model that takes RL and MIFr as inputs and calculates a LQI.

Main findings from the study include (1) an optimal LQI value of 0.15, obtained through ROC analysis, is found to best classify liquefied and nonliquefied sites in the database; (2) the proposed method demonstrated superior performance when compared with four existing accelerogram-based methods, and the overall accuracy is 92.8%, evaluated using the common K-Y liquefaction data set; and (3) the model is validated with a new data set, i.e., the 2001 Nisqually earthquake, not included in the original training data set. Overall, 29 out of the 32 sites (i.e., 2 correct LQ and 27 correct NonLQ sites) were correctly predicted by the model. In addition, the effects of the class imbalance, liquefaction observation uncertainty, and various sources of data bias, have been analyzed and discussed in detail.

The developed liquefaction model is given in a simple form that can be used by engineers and practitioners to quickly assess the liquefaction occurrence using ground motion records. The results of this study show that this model has a promising potential for applications in real-time disaster mitigation systems, and for the rapid response and loss estimation. Future work will continue expanding the liquefaction database and apply the proposed model to other earthquakes worldwide. Moreover, it would be interesting to compare the proposed method with other geotechnical and geospatial data-based approaches. Although there is limited availability of geotechnical data for most of the sites in the present data set to calculate indices such as the liquefaction potential index (LPI) or the liquefaction severity number (LSN), there are some sites with highly instrumented geotechnical arrays (e.g., the Wildlife Liquefaction Array), where such a comparison may be possible. The code of the developed model has been made publicly available (Data Availability Statement) so that the proposed accelerogram-based liquefaction assessment method could be used, further tested, and improved by the community.

Data Availability Statement

Some or all data, models, or code generated or used during the study are available in a repository or online in accordance with funder data retention policies. The Github repository hosting the developed code can be accessed through https://github.com/qschen/liquefaction-detection.git.

Acknowledgments

The authors would like to express their sincere gratitude to Professor C. Hsein Juang, who gave valuable advice on logistic regression analysis, to Professor Fumio Yamazaki, who kindly shared the 83 ground motion records from Kostadinov and Yamazaki (2001) for model development and evaluation, and to Professor Arthur D. Frankel, who shared the ground motion records of the 2001 Nisqually earthquake for the model validation. The authors would like to acknowledge the US Geological Survey, GeoNet (New Zealand), NIED (Japan), and PARI (Japan) for making their ground motion databases publicly accessible.

Supplemental Materials

Figs. S1–S6 and Tables S1–S9 are available online in the ASCE Library (www.ascelibrary.org).

References

Abe, K. 1995. "Recent great earthquakes and tectonics in Japan." *J. Phys. Earth* 43 (4): 395–405. https://doi.org/10.4294/jpe1952.43.395.

Anderson, J., P. Bodin, J. Brune, J. Prince, S. Singh, R. Quaas, and M. Onate. 1986. "Strong ground motion from the Michoacan, Mexico, earthquake." *Science* 233 (4768): 1043–1049. https://doi.org/10.1126/science.233.4768.1043.

Andrus, R. D., and K. H. Stokoe II. 2000. "Liquefaction resistance of soils from shear-wave velocity." *J. Geotech. Geoenviron. Eng.* 126 (11): 1015–1025. https://doi.org/10.1061/(ASCE)1090-0241(2000)126:11(1015).

Baise, L. G., R. B. Higgins, and C. M. Brankman. 2006. "Liquefaction hazard mapping–statistical and spatial characterization of susceptible units." *J. Geotech. Geoenviron. Eng.* 132 (6): 705–715. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:6(705).

Bouckovalas, G. D., Y. Z. Tsiapas, A. I. Theocharis, and Y. K. Chaloulos. 2016. "Ground response at liquefied sites: Seismic isolation or amplification?" *Soil Dyn. Earthquake Eng.* 91 (Dec): 329–339. https://doi.org/10.1016/j.soildyn.2016.09.028.

Boulanger, R. W., and I. Idriss. 2014. *CPT and SPT based liquefaction triggering procedures*. Rep. No. UCD/CGM-14/01. Davis, CA: Center for Geotechnical Modeling, Dept. of Civil and Environmental Engineering, Univ. of California.

Boulanger, R. W., and I. Idriss. 2016. "CPT-based liquefaction triggering procedure." *J. Geotech. Geoenviron. Eng.* 142 (2): 04015065. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001388.

Bozorgnia, Y., J. Stewart, and N. Abrahamson. 2020. *Data resources for NGA-subduction project*. PEER Rep. No. 2020/02. Berkeley, CA: Pacific Earthquake Engineering Research Center.

Bradley, B. A., and M. Cubrinovski. 2011. "Near-source strong ground motions observed in the 22 February 2011 Christchurch earthquake." *Seismol. Res. Lett.* 82 (6): 853–865. https://doi.org/10.1785/gssrl.82.6.853.

Bray, J. D., et al. 2001. *Some observations of geotechnical aspects of the February 28, 2001 Nisqually earthquake in Olympia, South Seattle, and Tacoma, Washington*. Berkeley, CA: Univ. of California, Berkeley.

Bray, J. D., and J. Macedo. 2017. "6th Ishihara lecture: Simplified procedure for estimating liquefaction-induced building settlement." *Soil Dyn. Earthquake Eng.* 102 (Nov): 215–231. https://doi.org/10.1016/j.soildyn.2017.08.026.

Bullock, Z., Z. Karimi, S. Dashti, K. Porter, A. Liel, and K. Franke. 2019. "A physics-informed semi-empirical probabilistic model for the settlement of shallow-founded structures on liquefiable ground." *Géotechnique* 69 (5): 406–419. https://doi.org/10.1680/jgeot.17.P.174.

Cetin, K. O., R. B. Seed, A. Der Kiureghian, K. Tokimatsu, L. F. Harder Jr., R. E. Kayen, and R. E. Moss. 2004. "Standard penetration test-based probabilistic and deterministic assessment of seismic soil liquefaction potential." *J. Geotech. Geoenviron. Eng.* 130 (12): 1314–1340. https://doi.org/10.1061/(ASCE)1090-0241(2004)130:12(1314).

Chen, Q., C. Wang, and C. H. Juang. 2016a. "CPT-based evaluation of liquefaction potential accounting for soil spatial variability at multiple scales." *J. Geotech. Geoenviron. Eng.* 142 (2): 04015077. https://doi.org/10.1061/(ASCE)GT.1943-5606.0001402.

Chen, Q., C. Wang, and C. H. Juang. 2016b. "Probabilistic and spatial assessment of liquefaction-induced settlements through multiscale random field models." *Eng. Geol.* 211 (Aug): 135–149. https://doi.org/10.1016/j.enggeo.2016.07.002.

Clayton, R. W., T. Heaton, M. Kohler, M. Chandy, R. Guy, and J. Bunn. 2015. "Community seismic network: A dense array to sense earthquake strong motion." *Seismol. Res. Lett.* 86 (5): 1354–1363. https://doi.org/10.1785/0220150094.

de Magistris, F. S., G. Lanzano, G. Forte, and G. Fabbrocino. 2013. "A database for PGA threshold in liquefaction occurrence." *Soil Dyn. Earthquake Eng.* 54 (Nov): 17–19. https://doi.org/10.1016/j.soildyn.2013.07.011.

Fawcett, T. 2006. "An introduction to ROC analysis." *Pattern Recognit. Lett.* 27 (8): 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Frankel, A. D., D. L. Carver, and R. A. Williams. 2002. "Nonlinear and linear site response and basin effects in Seattle for the M 6.8 Nisqually, Washington, earthquake." *Bull. Seismol. Soc. Am.* 92 (6): 2090–2109. https://doi.org/10.1785/0120010254.

GeoNet. 2020. "Earthquake catalogue." Accessed March 30, 2020. https://www.geonet.org.nz/.

Gingery, J. R. 2014. "Effects of liquefaction on earthquake ground motions." Ph.D. thesis, Dept. of Structural Engineering, Univ. of California, San Diego.

Holzer, T. L., T. E. Noce, and M. J. Bennett. 2011. "Liquefaction probability curves for surficial geologic deposits." *Environ. Eng. Geosci.* 17 (1): 1–21. https://doi.org/10.2113/gseegeosci.17.1.1.

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

ISC (International Seismological Centre). 2020. "ISC-GEM earthquake catalogue." Accessed March 30, 2020. http://www.isc.ac.uk/iscgem/.

Juang, C. H., J. Ching, Z. Luo, and C.-S. Ku. 2012. "New models for probability of liquefaction using standard penetration tests based on an updated database of case histories." *Eng. Geol.* 133 (Apr): 85–93. https://doi.org/10.1016/j.enggeo.2012.02.015.

Kaiser, A., et al. 2012. "The Mw 6.2 Christchurch earthquake of February 2011: Preliminary report." *N. Z. J. Geol. Geophys.* 55 (1): 67–90. https://doi.org/10.1080/00288306.2011.641182.

Karimi, Z., S. Dashti, Z. Bullock, K. Porter, and A. Liel. 2018. "Key predictors of structure settlement on liquefiable ground: A numerical parametric study." *Soil Dyn. Earthquake Eng.* 113 (Oct): 286–308. https://doi.org/10.1016/j.soildyn.2018.03.001.

Kayen, R., R. Moss, E. Thompson, R. Seed, K. Cetin, A. D. Kiureghian, Y. Tanaka, and K. Tokimatsu. 2013. "Shear-wave velocity–based probabilistic and deterministic assessment of seismic soil liquefaction potential." *J. Geotech. Geoenviron. Eng.* 139 (3): 407–419. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000743.

Kong, Q., R. M. Allen, L. Schreier, and Y.-W. Kwon. 2016. "MyShake: A smartphone seismic network for earthquake early warning and beyond." *Sci Adv* 2 (2): e1501055. https://doi.org/10.1126/sciadv.1501055.

Kostadinov, M. V., and F. Yamazaki. 2001. "Detection of soil liquefaction from strong motion records." *Earthquake Eng. Struct. Dyn.* 30 (2): 173–193. https://doi.org/10.1002/1096-9845(200102)30:2<173::AID-EQE3>3.0.CO;2-7.

Kramer, S., S. Sideras, and M. Greenfield. 2016. "The timing of liquefaction and its utility in liquefaction hazard evaluation." *Soil Dyn. Earthquake Eng.* 91 (Dec): 133–146. https://doi.org/10.1016/j.soildyn.2016.07.025.

Kramer, S. L., B. A. Asl, P. Ozener, and S. S. Sideras. 2015. "Effects of liquefaction on ground surface motions." In *Perspectives on earthquake geotechnical engineering*, 285–309. New York: Springer.

Kramer, S. L., S. S. Sideras, M. W. Greenfield, and B. Hushmand. 2018. "Liquefaction, ground motions, and pore pressures at the Wildlife Liquefaction Array in the 1987 Superstition Hills earthquake." In *Proc., Geotechnical Earthquake Engineering and Soil Dynamics V: Liquefaction Triggering, Consequences, and Mitigation*. Reston, VA: ASCE. https://doi.org/10.1061/9780784481455.037.

Ku, C.-S., C. H. Juang, C.-W. Chang, and J. Ching. 2012. "Probabilistic version of the Robertson and Wride method for liquefaction evaluation: Development and application." *Can. Geotech. J.* 49 (1): 27–44. https://doi.org/10.1139/t11-085.

Matasovic, N., and Y. Hashash. 2012. *Practices and procedures for site-specific evaluations of earthquake ground motions: A synthesis of highway practice*. Washington, DC: National Cooperative Highway Research Program of the Transportation Research Board.

Matsuoka, M., K. Wakamatsu, M. Hashimoto, S. Senna, and S. Midorikawa. 2015. "Evaluation of liquefaction potential for large areas based on geomorphologic classification." *Earthquake Spectra* 31 (4): 2375–2395. https://doi.org/10.1193/072313EQS211M.

Maurer, B., R. Green, M. Cubrinovski, and B. Bradley. 2015. "Assessment of CPT-based methods for liquefaction evaluation in a liquefaction potential index framework." *Géotechnique* 65 (5): 328–336. https://doi.org/10.1680/geot.SIP.15.P.007.

Miyajima, M. 1998. "Detective method of liquefaction using strong ground motion records." In *Proc., 3rd China-Japan-US Trilateral Symp. on Lifeline Earthquake Engineering, Kunming, China, August 1998*, 133–140. Beijing: China Earthquake Administration.

Miyake, H., T. Iwata, and K. Irikura. 2003. "Source characterization for broadband ground-motion simulation: Kinematic heterogeneous source model and strong motion generation area." *Bull. Seismol. Soc. Am.* 93 (6): 2531–2545. https://doi.org/10.1785/0120020183.

Moss, R., R. B. Seed, R. E. Kayen, J. P. Stewart, A. Der Kiureghian, and K. O. Cetin. 2006. "CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential." *J. Geotech. Geoenviron. Eng.* 132 (8): 1032–1051. https://doi.org/10.1061/(ASCE)1090-0241(2006)132:8(1032).

NASEM (National Academies of Sciences, Engineering, and Medicine). 2016. *State of the art and practice in the assessment of earthquake-induced soil liquefaction and its consequences*. Washington, DC: National Academies Press. https://doi.org/10.17226/23474.

Ozaki, R. 1999. "Study on real-time earthquake mitigation liquefaction monitoring and earthquake countermeasures." Ph.D. thesis, Graduate School of Engineering, Kobe Univ.

Özener, P. T., M. W. Greenfield, S. S. Sideras, and S. L. Kramer. 2020. "Identification of time of liquefaction triggering." *Soil Dyn. Earthquake Eng.* 128 (Jan): 105895. https://doi.org/10.1016/j.soildyn.2019.105895.

Rioul, O., and M. Vetterli. 1991. "Wavelets and signal processing." *IEEE Signal Process Mag.* 8 (4): 14–38. https://doi.org/10.1109/79.91217.

Robertson, P. K., and C. Wride. 1998. "Evaluating cyclic liquefaction potential using the cone penetration test." *Can. Geotech. J.* 35 (3): 442–459. https://doi.org/10.1139/t98-017.

Seed, H. B., and I. M. Idriss. 1971. "Simplified procedure for evaluating soil liquefaction potential." *J. Soil Mech. Found. Div.* 97 (9): 1249–1273. https://doi.org/10.1061/JSFEAQ.0001662.

Seed, H. B., K. Tokimatsu, L. F. Harder, and R. M. Chung. 1985. "Influence of SPT procedures in soil liquefaction resistance evaluations." *J. Geotech. Eng.* 111 (12): 1425–1445. https://doi.org/10.1061/(ASCE)0733-9410(1985)111:12(1425).

Shimizu, Y., A. Watanabe, K. Koganemaru, W. Nakayama, and F. Yamazaki. 2000. "Super high-density realtime disaster mitigation system." In *Proc., 12th World Conf. on Earthquake Engineering*. Auckland, New Zealand: New Zealand Society for Earthquake Engineering.

Smyrou, E., İ. E. Bal, P. Tasiopoulou, and G. Gazetas. 2016. "Wavelet analysis for relating soil amplification and liquefaction effects with seismic performance of precast structures." *Earthquake Eng. Struct. Dyn.* 45 (7): 1169–1183. https://doi.org/10.1002/eqe.2701.

Suzuki, T., Y. Shimizu, and W. Nakayama. 1998. "Characteristics of strong motion records at the liquefied sites and judgment for liquefaction." In *Proc., 11th European Conf. on Earthquake Engineering*. Rotterdam, Netherlands: A.A. Balkema.

VDC (COSMOS Strong-Motion Virtual Data Center). 2020. "Worldwide strong ground motion data." Accessed March 30, 2020. https://www.strongmotioncenter.org/vdc/scripts/default.plx.

Wang, C., and Q. Chen. 2018. "A hybrid geotechnical and geological data-based framework for multiscale regional liquefaction hazard mapping." *Géotechnique* 68 (7): 614–625. https://doi.org/10.1680/jgeot.17.P.074.

Youd, T. L., et al. 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. Eng.* 127 (10): 817–833. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:10(817).

Youd, T. L., and D. M. Perkins. 1978. "Mapping liquefaction-induced ground failure potential." *J. Soil Mech. Found. Div.* 104 (4): 433–446. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:10(817).

Yuan, X., R. Sun, L. Chen, and F. Tang. 2010. "A method for detecting site liquefaction by seismic records." *Soil Dyn. Earthquake Eng.* 30 (4): 270–279. https://doi.org/10.1016/j.soildyn.2009.12.003.

Zhu, J., L. G. Baise, and E. M. Thompson. 2017. "An updated geospatial liquefaction model for global application." *Bull. Seismol. Soc. Am.* 107 (3): 1365–1385. https://doi.org/10.1785/0120160198.

Zhu, J., D. Daley, L. G. Baise, E. M. Thompson, D. J. Wald, and K. L. Knudsen. 2015. "A geospatial liquefaction model for rapid response and loss estimation." *Earthquake Spectra* 31 (3): 1813–1837. https://doi.org/10.1193/121912EQS353M.