Seismic Fragility Evaluation with Incomplete Structural Appraisal Data: An Iterative Statistical Approach

Vincent Z. Wang[1]; Michael Mallett[2]; and Andrew Priory[3]

Abstract: This paper presents an iterative statistical approach to evaluating seismic structural safety using incomplete appraisal data. Despite the continuous improvement to traditional structural assessment procedures and the recent progress in structural health monitoring methodologies, practically acquired structural appraisal data may often be incomplete. The occurrence of the appraisal data missingness could be ascribed to the malfunction of data acquisition systems, the abnormality during data transfer, and the inaccessibility of critical quantities, among other reasons. The study begins with a quantitative investigation into the sensitivity of the seismic fragility evaluation with respect to the structural appraisal data missingness through the defined additional information loss and probability of noninformativeness. Subsequently, a remedy for the missingness of the structural appraisal data, instead of a precaution against it, is formulated by employing the expectation-maximization (EM) algorithm. With synthetic or real seismic ground accelerations involved, the efficacy of the EM algorithm embedded remedy is demonstrated by examples of typical linear or nonlinear hysteretic systems in the framework of statistical hypothesis testing. Resorting to the bootstrap technique, the influence of the related correlations and missingness probability is also examined. **DOI: 10.1061/(ASCE)ST.1943-541X.0000804.** *© 2013 American Society of Civil Engineers.*

Author keywords: Seismic fragility; Statistics; Incomplete data; Missing data; Uncertainty quantification; EM algorithm; Bootstrap; Linear or hysteretic systems; Analysis and computation.

Introduction

Seismic activities have long been identified as one of the major natural hazards against which civil structures need to be designed. Among various measures of the safety performance of a structure pertaining to seismic activities, the fragility is to quantify the likelihood that a predefined event occurs. This event is typically undesirable and can usually be formally described as such: Corresponding to a given intensity of seismic ground motions, the response of interest of the structure goes beyond a prescribed bound. Practically, there exist scores of options for the response of interest in accordance with the vast variety in engineering structures. As a purpose-related issue, the final choice is contingent upon the structure being investigated and the specific structural aspects with which a study is concerned. For instance, for a given multistory reinforced concrete frame, the bending moments at the beam-column joints and the story drifts can respectively serve as the response of interest when its load-carrying capacity and deformation performance are of primary concern. Examples of the response of interest in the literature include displacement (Sasani and Der Kiureghian 2001; Kafali and Grigoriu 2007), drift (Kim and Shinozuka 2004; Lupoi et al. 2006; Choe et al. 2008; Park et al. 2009; Celik and Ellingwood 2010), stress resultants (Lupoi et al. 2006; Casciati et al. 2008; Choe et al. 2008), and curvature (Lupoi et al. 2006), along with other derived quantities (Oller and Barbat 2006) and combinations of multiple quantities (Cimellaro et al. 2009; Cimellaro and Reinhorn 2011).

Once the response of interest has been selected, it is also indispensable to gather enough information to evaluate its value. The information required comprises the structural condition, loading configurations, modeling considerations, etc. Of particular interest is the structural condition. For structures to be constructed or recently constructed, it may be obtained from relevant regulatory guidelines on structural design, such as the code requirements for concrete buildings [American Concrete Institute (ACI) 2011; British Standards Institution (BSI) 2004; Standards Australia 2009] and steel structures (American Institute of Steel Construction (AISC) 2010; BSI 2005; Standards Australia 1998]. Quality control procedures could also shed some light on the condition of a newly constructed structure. When ready mixed concrete or precast/prefabricated structural components and subassemblies are involved, it is often the case that standard quality control practice, e.g., as per ASTM E122 (ASTM 2009), and various in-house quality control measures may provide some information useful for the evaluation of the structural condition.

By contrast, if the condition of a structure having been in service for a considerable amount of time needs to be assessed, those data directly from the aforementioned structural design requirements and quality control procedures may not be accurate enough for subsequent fragility analysis as they generally do not allow for possible damage or deterioration to which the structure may have been subjected. Thus structural appraisal techniques with the capability of inferring the up-to-date condition of the structure come into play. ASTM C805/C805M (ASTM 2008) and BS EN 13791 (BSI 2007) are among the standards that direct the industry practice in this regard. Diverse structural appraisal techniques are often applied simultaneously for improved accuracy. Sullivan (1991) evaluated the in situ strengths of the concrete in an office building using nondestructive tests and core extraction. Hassan et al. (1995) investigated the modulus of elasticity of concrete in in-service

[1]Lecturer in Civil Engineering, School of Engineering and Physical Sciences, James Cook Univ., Townsville, Queensland 4811, Australia (corresponding author). E-mail: vincent.wang@jcu.edu.au

[2]Former Honours Project Student, School of Engineering and Physical Sciences, James Cook Univ., Townsville, Queensland 4811, Australia.

[3]Former Honours Project Student, School of Engineering and Physical Sciences, James Cook Univ., Townsville, Queensland 4811, Australia.

Note. This manuscript was submitted on March 24, 2012; approved on January 14, 2013; published online on January 16, 2013. Discussion period open until March 7, 2014; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Structural Engineering***, © ASCE, ISSN 0733-9445/04013048(13)/$25.00.

bridges obtained by ultrasonic measurements and extracted cores, and compared the results with those from load tests. A series of nondestructive assessment methods was employed by Pascale et al. (2003) to study the properties of high-strength concrete. It is also noted that statistical techniques are frequently invoked in a structural-appraisal scenario to account for the uncertainty involved. Leshchinsky (1992) reviewed several statistical criteria relevant to the nondestructive assessment of concrete strength. Geyskens et al. (1998) proposed a Bayesian approach to relate the modulus of elasticity of concrete to the corresponding compressive strength. Maes (2002) formulated an updating scheme in the empirical Bayes framework to assess the reliability of a structure subject to demanding environmental conditions. Recently, a multidisciplinary research field, structural health monitoring (SHM), has been growing very fast (Chang 2011). SHM aims to detect the presence of damage in a structure, pinpoint the damage, quantify its extent, and prognosticate its potential development as well as its influence on the service life of the structure. The advances in SHM have been complementing conventional structural appraisal techniques, giving rise to enhanced capability for the acquisition of the data on the current condition of in-service structures.

With continuing research effort put into the area of structural appraisal and health monitoring, there is no doubt that more and more strategies will be constructed to further the state of the art of obtaining the latest data on the condition of the structure being investigated. An accompanying issue arises that: in practical situations the structural appraisal data collected can be incomplete. The missingness of the structural appraisal data occurs when the data acquisition system does not function as it has been originally designed. Nowadays it is not uncommon that a data acquisition system consists of numerous sensors, and such factors as installation errors, mechanical actions, electrical surges, and power outage in any of them may lead to the data missingness to some extent. For the sensors embedded into a structure and expected to work constantly during the service life of the structure, problems pertaining to the lifetimes of the sensors and the maintenance and replacement plans could result in the data missingness as well. Another potential source that the appraisal data missingness can be ascribed to is the reliability, robustness, and working environment of the data transfer mechanism employed. High levels of undesirable noise, interference, and distortion may considerably affect the amount of the useful information that can be extracted from the signals. The appraisal data missingness happens if the disturbance cannot be allowed for by the data transfer mechanism. More specifically, this includes the occasions when the disturbance is of such a kind that the data transfer mechanism has not been designed and prepared to resist its influence, or when some redundancy does exist in the data transfer mechanism but is unfortunately not enough to compensate for the detrimental effect of the disturbance. Besides, the missingness of the structural appraisal data could also be attributed to the fact that the quantities critical to the performance of a civil structure are not always readily convenient to be directly observed or measured. For example, in a typical frame building the nonstructural elements such as ceilings and finishes may prevent the frame joints from direct measurement access, and thus the relevant appraisal data can be missing. Upon identifying representative scenarios in which the structural appraisal data may become incomplete, appropriate measures that can be taken in these scenarios need to be sorted out. Effort has reasonably been focused on coming up with comprehensive precautions against the missingness of the appraisal data, as exhibited by the rapid progress of a large number of structural appraisal and health monitoring techniques. However, the other side of the coin is that the likelihood of the occurrence of the appraisal data missingness is still far from negligible. If a data missingness event does occur, the consequence is that the structural appraisal data turn out to be incomplete. Note that even when the probability of the data missingness event is small, the missingness probability associated with an individual appraisal data point can nevertheless be fairly high if the event occurs. Therefore, it would be interesting and worthwhile to formulate some remedies, in contrast to precautions, for the structural appraisal data missingness and examine their effectiveness in seismic safety performance evaluation.

In the next section, a pilot study is presented in which the sensitivity of the seismic fragility evaluation with respect to the missingness of structural appraisal data is demonstrated. Subsequently, an iterative statistical algorithm is employed to formulate an efficacious approach to deal with incomplete structural appraisal data with application to both linear systems and nonlinear hysteretic systems. The influence of some parameters involved is also discussed.

Sensitivity of the Seismic Fragility Evaluation to the Appraisal Data Missingness: A Pilot Study

Despite the relatively large number of extensively studied data missingness cases in other areas such as life sciences, clinical research, etc., to the authors' knowledge, few results on the seismic safety performance of civil structures are available. Accordingly, it becomes helpful to illustrate by some straightforward examples the structural appraisal data missingness and its effect on the fragility evaluation. More details along this line can be found in Wang et al. (2011).

Consider, for instance, a three-story shear frame modeled as a three-degree-of-freedom (DOF) linear dynamic system. Suppose that the story stiffness has a trivariate normal distribution. Practically, the parameters in this probability distribution, i.e., the mean vector and the covariance matrix, can be estimated from pertinent structural appraisal data, which may be incomplete. Throughout the study presented in this paper, the missing-completely-at-random (MCAR) assumption (Heitjan and Basu 1996) is used to model the incomplete structural appraisal data. MCAR refers to the situation in which the conditional probability of an observed missingness pattern given the observed and missing data equals the probability of the observed missingness pattern. That is, it essentially assumes that the observed missingness pattern is statistically independent of the observed and missing data. Letting the mean and the coefficient of variation of each of the three random variables for the story stiffness be 1.5×10^4 kN/m and 0.3 respectively, and assuming that the correlation between any two of the three random variables is 0.5, Fig. 1 illustrates the corresponding complete and incomplete structural appraisal data with a sample size m of 15. Notice that the missing data are indicated by blanks. As one can see in Fig. 1(b) where the appraisal data for each story are subjected to a missingness probability p of 0.3, a total of six, five, and five data points are missing for the first, second, and third stories, respectively.

To estimate the mean vector, the incomplete appraisal data can be treated as three individual univariate samples. For the estimation of the covariance matrix, two immediate options exist, i.e., pairwise manner and listwise manner. In the pairwise manner, the covariance between two random variables is estimated based on the missingness pattern exhibited by the realizations of these two random variables only. The ostensible advantage of the pairwise manner is that it appears to make the best of the situation of the data missingness, while when the bigger picture of estimating the full covariance matrix for the story stiffness is considered, it turns out to be dubious as the full covariance matrix thus obtained may not

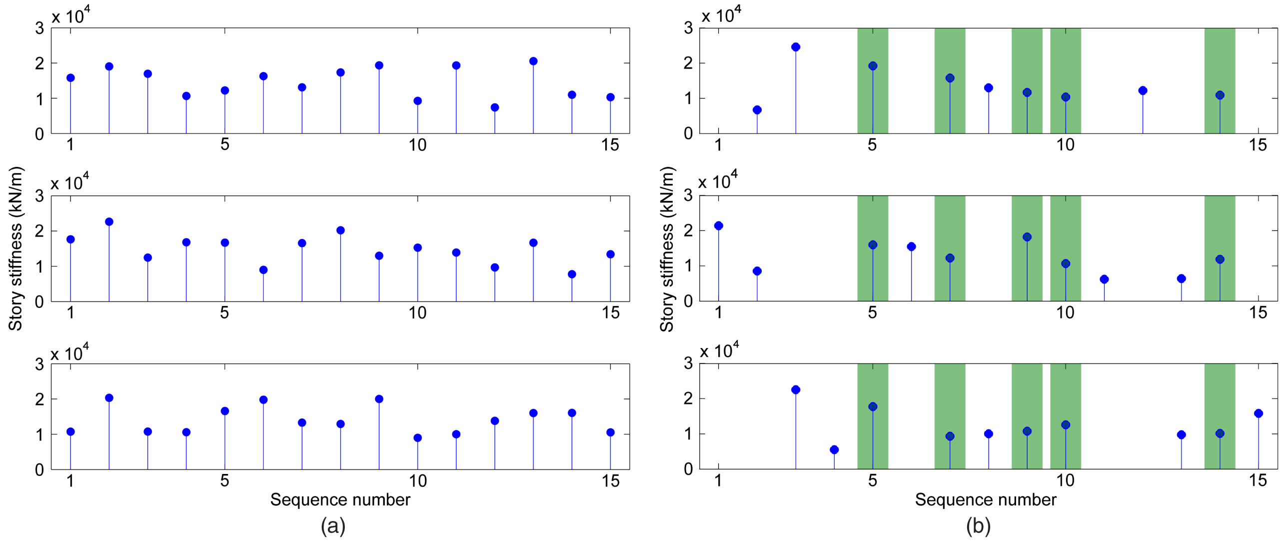

Fig. 1. An illustration of the structural appraisal data missingness: (a) complete appraisal data; and (b) incomplete appraisal data for the story stiffness of a three-story shear frame

always be positive semidefinite. Instead, the listwise manner can be used where only the observed units without any missing data [i.e., the five shaded units in Fig. 1(b)] are employed to estimate the covariance matrix. Obviously, the listwise manner imposes a second round of information loss on top of the data missingness having occurred. For the incomplete appraisal data illustrated in Fig. 1(b), besides the 16 missing data points, 14 observed data points do not essentially contribute to the covariance matrix estimation if the listwise manner is applied. This corresponds to a 48.3% additional information loss. Here the additional information loss is quantified by the percentage of the number of noncontributory observed data points with respect to the total number of observed data points. More generally, Fig. 2(a) plots the average additional information loss for different values of the missingness probability p and the number of stories. It is apparent that on average a considerable portion of the observed data points is noncontributory when the number of stories is large. Another concern is that the probability of the event that each observed unit contains at least one missing data point, which is defined as the probability of noninformativeness, becomes higher and higher if either the missingness probability or the number of stories increases. Several examples of the probability of noninformativeness corresponding to selected values of the sample size m and the missingness probability p are presented in Fig. 2(b). For those cases with notably high probabilities of noninformativeness, very often the listwise estimation procedures cannot even be implemented, not to mention the level of accuracy achieved. With some examples of the seismic fragility evaluation in the listwise manner documented by Wang et al. (2011), exploring alternative procedures to cope with the incomplete appraisal data would be preferable.

Fragility Evaluation of Linear Systems with Incomplete Appraisal Data

The fragilities of typical linear systems under the circumstances of the appraisal data missingness are dealt with in this section.

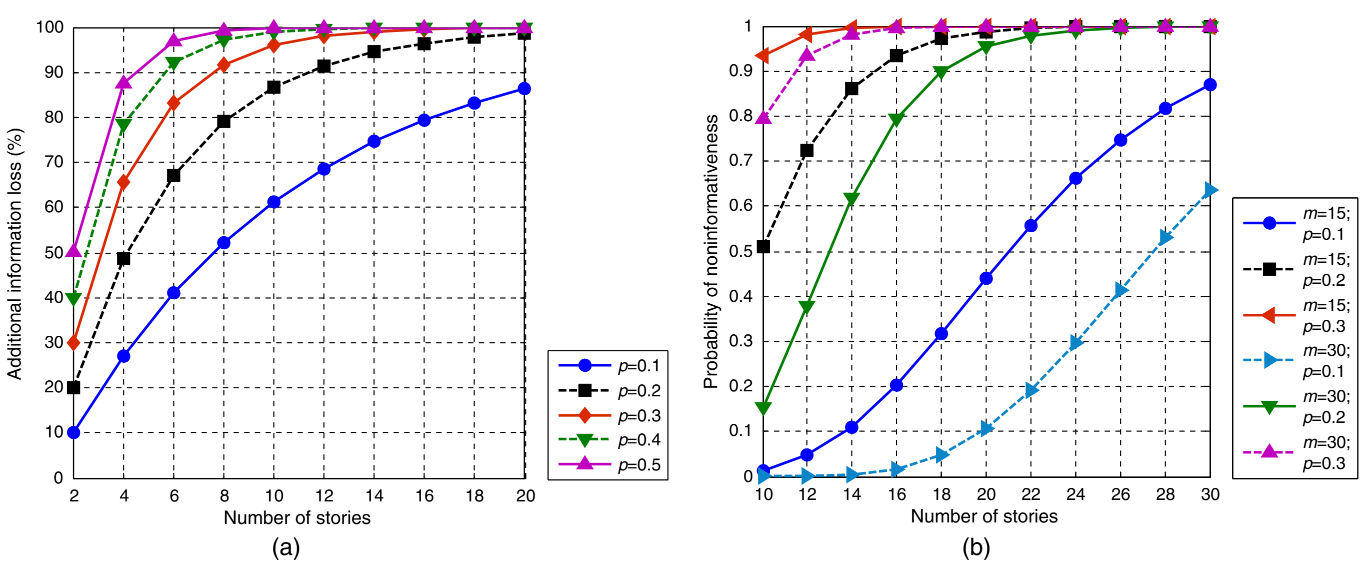

Fig. 2. Concerns about the listwise estimation procedures: (a) additional information loss; and (b) probability of noninformativeness

The equation of motion of a linear structural system subject to a seismic-ground-acceleration time history can be written as

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \mathbf{K}\mathbf{U}(t) = -\mathbf{M}\mathbf{X}(t) \tag{1}$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ denote the mass, damping, and stiffness matrices, respectively; t is the time; $\mathbf{U}(t)$ is the displacement-time-history random vector of the structural system; and $\mathbf{X}(t)$ is the seismic-ground-acceleration random vector. Assume that only the horizontal vibrations of the system are looked at in Eq. (1), as is the case for the vast majority of relevant structural engineering practice. Accordingly, $\mathbf{U}(t)$ only contains the displacements of the horizontal DOFs of the structural system, and $\mathbf{X}(t)$ is simply based on the seismic ground acceleration in the horizontal direction $X(t)$:

$$\mathbf{U}(t) = (U_1(t), U_2(t), \ldots, U_n(t))^{\mathrm{T}} \tag{2}$$

$$\mathbf{X}(t) = (X(t), X(t), \ldots, X(t))^{\mathrm{T}} \tag{3}$$

Specifically in this section, linear multistory shear frames are considered for the conciseness in terms of the stiffness matrix $\mathbf{K}$, and the resulting formulation could analogously be extended to other types of linear or nonlinear systems, as to be illustrated in the next section. The stiffness matrix $\mathbf{K}$ of a linear n-story shear frame can be derived as

$$\mathbf{K} = \begin{pmatrix} K_1 + K_2 & -K_2 & 0 & \cdots & 0 \\ -K_2 & K_2 + K_3 & -K_3 & \ddots & \vdots \\ 0 & -K_3 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & K_{n-1} + K_n & -K_n \\ 0 & \cdots & 0 & -K_n & K_n \end{pmatrix} \tag{4}$$

where K_i is a random variable, and denotes the story stiffness for the ith story ($i = 1, 2, \ldots, n$). Hence, the stiffness matrix $\mathbf{K}$ is characterized by the random vector $\mathbf{K}'$ defined as

$$\mathbf{K}' = (K_1, K_2, \ldots, K_n)^{\mathrm{T}} \tag{5}$$

Let $f(\mathbf{K}'|\boldsymbol{\theta})$ be the joint probability density function for $\mathbf{K}'$, where $\boldsymbol{\theta}$ contains the n_0 distribution parameters, as in Eq. (6):

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{n_0})^{\mathrm{T}} \tag{6}$$

Following a common interpretation, the seismic fragility of the shear frame, denoted by q, is shown in Eq. (7):

$$q = P\left\{\max_{0\leq t\leq T}(|U_1(t)|, |U_{i+1}(t) - U_i(t)| \text{ for } i = 1, 2, \ldots, n-1) > u_{\mathrm{m}}\right\} \tag{7}$$

where T is the maximum time considered; u_{m} is a preselected threshold displacement; and P is the probability measure in the probability space (S, $\mathbb{S}$, P) associated with the issue. Obviously, the value of the probability q is contingent upon the probability distribution of $\mathbf{K}'$, among other contributing factors. In many engineering applications, Eq. (7) is often evaluated by resorting to some numerical procedures (e.g., the Monte Carlo simulation).

To estimate the distribution parameters θ_i ($i = 1, 2, \ldots, n_0$), the realizations of the random vector $\mathbf{K}'$ can be obtained from the structural appraisal and health monitoring results, in combination with some relevant analytical or empirical relationships where appropriate. Ideally, this leads to the random sample $\mathbf{K}_{\mathrm{r}}'$ below:

$$\mathbf{K}_{\mathrm{r}}' = (\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_n) \tag{8}$$

in which

$$\mathbf{k}_i = (k_{1i}, k_{2i}, \ldots, k_{mi})^{\mathrm{T}} \tag{9}$$

where $i = 1, 2, \ldots, n$; and $\mathbf{k}_i$ is the realizations of K_i with the sample size of m. Invoking the canonical maximum likelihood routine, one could get an estimate/estimator

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{n_0})^{\mathrm{T}} \tag{10}$$

for $\boldsymbol{\theta}$, and proceed to evaluate Eq. (7) for the seismic fragility q. However, as described in the first section, practically a data missingness event may occur, meaning that in some cases not all the entries in the matrix $\mathbf{K}_{\mathrm{r}}'$ are available. Further, as demonstrated in the pilot study, if only incomplete data for $\mathbf{K}_{\mathrm{r}}'$ are collected, unfortunately neither the pairwise estimation procedures nor the listwise ones seem to be appealing. It is thereby propounded in this study to employ an iterative statistical approach known as the expectation-maximization (EM) algorithm (Dempster et al. 1977; Wu 1983; Meng and van Dyk 1997) for improved performance in the estimation of the parameter $\boldsymbol{\theta}$ and in turn the computation of the probability q. There are two steps (i.e., E-step and M-step) in each iteration of the EM algorithm. Heuristically, in a typical E-step the mean vector of the jointly sufficient statistics based on the complete data is estimated, and the M-step then aims to achieve in the maximum likelihood framework a better estimate of the parameter $\boldsymbol{\theta}$ than that in the last iteration. The two-step iterative process continues until some predefined convergence criteria are met, yielding the final estimate/estimator $\hat{\boldsymbol{\theta}}$. Examples of the convergence criteria include that in terms of a kind of vector norm the absolute or relative error between two consecutive iterates for the estimate of the parameter $\boldsymbol{\theta}$ does not exceed a given small positive number. Rigorous and complete expositions of the EM algorithm can be found by referring to the previously mentioned statistical literature. Starting with a simple case where the appraisal data missingness is restricted to a single story, the subsections below examine in detail the efficacy of this iterative statistical approach in the context of linear system fragility evaluation, and the application to a nonlinear hysteretic system is deferred to the next section.

Occurrence of the Appraisal Data Missingness at a Single Story

A 10-story ($n = 10$) shear frame is considered. Suppose that the corresponding story stiffness random vector $\mathbf{K}'$ has a 10-variate normal distribution. The distribution parameter $\boldsymbol{\theta}$ thus consists of the 65 independent entries ($n_0 = 65$) in the mean vector and the covariance matrix. Further, for each K_i ($i = 1, 2, \ldots, 10$) assume that its mean and coefficient of variation are 1.5×10^7 kN/m and 0.3, respectively; for each pair of K_i and K_j (i, $j = 1, 2, \ldots, 10$; $i \neq j$) the correlation is taken to be 0.5. The 10-by-10 lumped mass matrix with each diagonal entry equal to 1.5×10^6 kg is used for $\mathbf{M}$. The damping effect is simulated through the Rayleigh damping with the coefficients for the matrices $\mathbf{M}$ and $\mathbf{K}$ computed from a modal damping ratio of 0.01 for the first two modes. The horizontal seismic ground acceleration $x(t)$, i.e., the realization of $X(t)$, is based on the classical, widely used Clough-Penzien one-sided power spectral density (Clough and Penzien 1993). As this section is only concerned with linear systems, the displacement time histories $u_i(t)$ ($i = 1, 2, \ldots, 10$), i.e., the realizations of $U_i(t)$ ($i = 1, 2, \ldots, 10$), are obtained by the mode superposition in conjunction with the state space

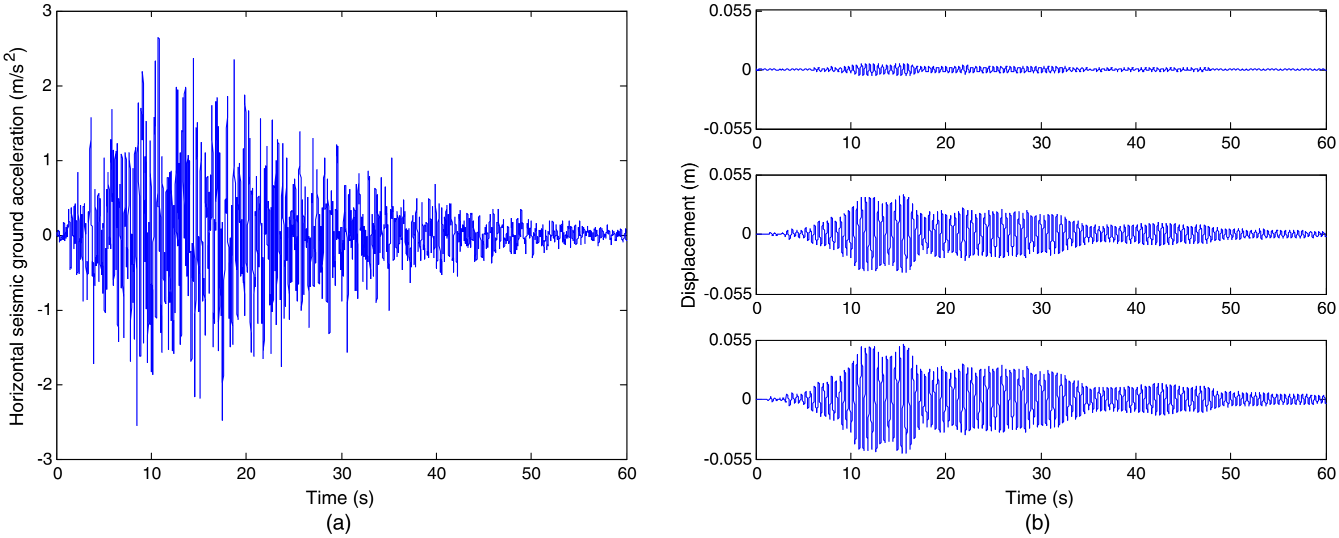

Fig. 3. For the 10-story shear frame: (a) an example of the horizontal seismic ground acceleration; and (b) the corresponding displacement time histories at the first (top), fifth (middle), and ninth (bottom) stories

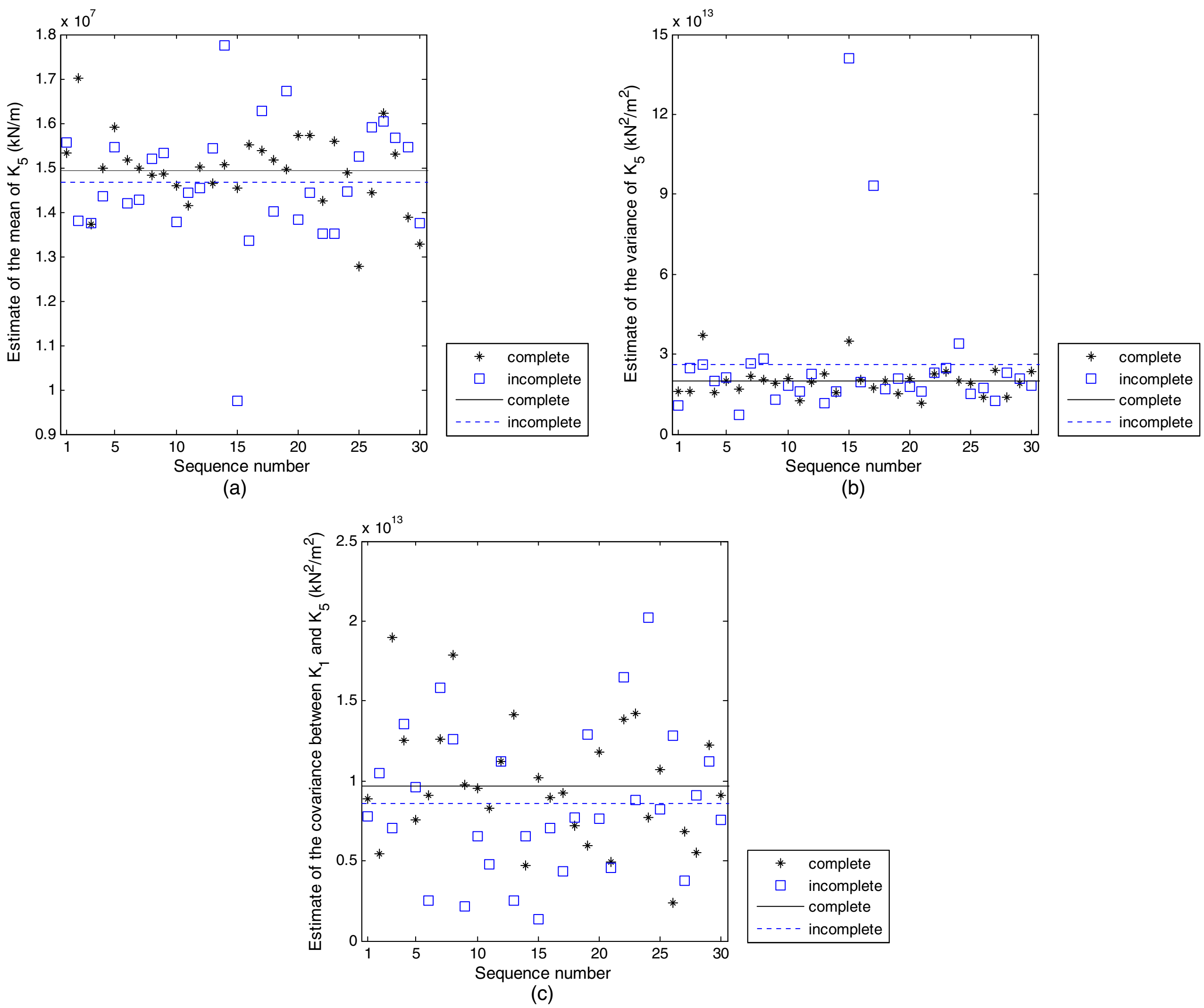

Fig. 4. For the case that the appraisal data missingness can only occur at the fifth story of the 10-story shear frame the estimates of: (a) the mean; (b) variance of K_5; (c) the covariance between K_1 and K_5 (The horizontal solid and dashed lines indicate the sample means of the estimates corresponding to the complete and incomplete appraisal data respectively, as also in Figs. 5 and 7)

Table 1. Comparison between the Seismic Fragilities in the Complete-Data Scenario and Those in the Incomplete Data Scenario Where the Data Missingness Can Occur at and Only at the Fifth Story of the Linear Shear Frame

Trial ID	Sample and hypothesis testing result for the seismic fragility when $\xi = 0.981$ m/s^2: Sample	Result	Sample and hypothesis testing result for the seismic fragility when $\xi = 3.924$ m/s^2: Sample	Result
L1A	0.010; 0.005; 0.080; 0.055; 0.026; 0.023; 0.046; 0.051; 0.041; 0.046	T	0.792; 0.812; 0.857; 0.871; 0.878; 0.845; 0.871; 0.867; 0.877; 0.853	T
	0.013; 0.053; 0.070; 0.060; 0.027; 0.012; 0.024; 0.028; 0.014; 0.014		0.838; 0.927; 0.908; 0.861; 0.873; 0.844; 0.888; 0.855; 0.869; 0.912	
L1B	0.009; 0.009; 0.081; 0.061; 0.031; 0.014; 0.041; 0.061; 0.030; 0.034	T	0.827; 0.789; 0.873; 0.849; 0.878; 0.840; 0.868; 0.881; 0.890; 0.855	T
	0.008; 0.049; 0.074; 0.054; 0.029; 0.012; 0.029; 0.032; 0.018; 0.016		0.820; 0.916; 0.896; 0.863; 0.877; 0.869; 0.874; 0.881; 0.858; 0.905	
L2A	0.050; 0.036; 0.065; 0.016; 0.064; 0.046; 0.009; 0.021; 0.031; 0.031	T	0.898; 0.819; 0.862; 0.867; 0.883; 0.841; 0.846; 0.872; 0.875; 0.887	T
	0.018; 0.055; 0.008; 0.036; 0.343; 0.051; 0.121; 0.004; 0.015; 0.028		0.878; 0.890; 0.909; 0.873; 0.920; 0.883; 0.908; 0.840; 0.803; 0.900	
L2B	0.049; 0.027; 0.063; 0.014; 0.070; 0.043; 0.013; 0.017; 0.030; 0.050	T	0.878; 0.816; 0.869; 0.853; 0.875; 0.822; 0.861; 0.876; 0.864; 0.884	T
	0.016; 0.051; 0.014; 0.029; 0.330; 0.043; 0.120; 0.008; 0.014; 0.034		0.876; 0.871; 0.895; 0.897; 0.909; 0.906; 0.868; 0.834; 0.807; 0.879	
L3A	0.023; 0.049; 0.034; 0.020; 0.041; 0.036; 0.033; 0.016; 0.033; 0.055	T	0.897; 0.866; 0.814; 0.840; 0.913; 0.862; 0.841; 0.824; 0.873; 0.892	T
	0.031; 0.099; 0.035; 0.044; 0.013; 0.024; 0.029; 0.048; 0.009; 0.034		0.919; 0.878; 0.871; 0.827; 0.827; 0.839; 0.812; 0.854; 0.783; 0.877	
L3B	0.020; 0.072; 0.032; 0.019; 0.040; 0.056; 0.027; 0.017; 0.043; 0.051	T	0.913; 0.842; 0.808; 0.857; 0.870; 0.922; 0.831; 0.818; 0.859; 0.898	T
	0.044; 0.096; 0.030; 0.050; 0.010; 0.020; 0.018; 0.043; 0.016; 0.046		0.909; 0.887; 0.868; 0.867; 0.825; 0.829; 0.820; 0.840; 0.776; 0.866	

Note: The first and second rows in each trial respectively show the samples obtained from the complete-data and incomplete-data scenarios, and the hypothesis testing results are indicated by the logical values with "T" and "F" respectively denoting "failing to reject H_0" and "rejecting H_0", as is also the case in Tables 2 and 5.

representation. The sampling frequency for $x(t)$ is 200 Hz, and the time step size for $u_i(t)$ $(i = 1, 2, \ldots, 10)$ is 0.005 s. Fig. 3 shows an example of $x(t)$ and the corresponding $u_1(t)$, $u_5(t)$, and $u_9(t)$.

In order to investigate the performance of the approach, two parallel scenarios are constructed independently. In the first scenario, no data missingness event occurs, and the 10-variate normal random vector $\mathbf{K}'$ is fully observed 30 times ($m = 30$). That is, the random sample $\mathbf{K}_r'$ is a 30-by-10 matrix with all its entries obtained, and these complete appraisal data are utilized to evaluate the seismic fragilities of the shear frame thereafter. On the contrary, it is assumed in the second scenario that the appraisal data missingness can occur only at the fifth story, and a missingness probability p equal to 0.3 is introduced such that for each of the 30 entries in $\mathbf{k}_5$ the probability of the event that this entry is missing is 0.3. This is achieved in the simulation through a binomial distribution with the parameters of 30 and 0.7. In the first scenario, namely the complete-data scenario, $\hat{\boldsymbol{\theta}}$ is straightforwardly computed from the sample mean and the sample covariance matrix; and in the second scenario, or the incomplete-data scenario, the EM algorithm is applied to get it. Throughout this study, the EM algorithm is implemented through the *R* environment (R Development Core Team 2011) and the package *norm*, a contributed package by Novo and Schafer (2010), and the convergence criteria are introduced such that the two-step iterative process stops if and only if the maximum relative difference between all the corresponding entries of the estimates of the parameter $\boldsymbol{\theta}$ in two consecutive iterations is not more than 0.0001 (Novo and Schafer 2010). To allow for the uncertainty involved and characterize the stochastic properties of the estimator $\hat{\boldsymbol{\theta}}$, the entire procedures in each of the two scenarios are independently carried out 30 times. Consequently, one obtains a random sample of size 30 for each individual θ_i $(i = 1, 2, \ldots, 65)$. Indeed, this in general yields in the complete-data scenario

$$^{c}\hat{\boldsymbol{\theta}}_i = (^{c}\hat{\theta}_{1i}, {}^{c}\hat{\theta}_{2i}, \ldots, {}^{c}\hat{\theta}_{m_0 i})^{T} \tag{11}$$

and in the incomplete-data scenario

$$^{i}\hat{\boldsymbol{\theta}}_i = (^{i}\hat{\theta}_{1i}, {}^{i}\hat{\theta}_{2i}, \ldots, {}^{i}\hat{\theta}_{m_0 i})^{T} \tag{12}$$

where $i = 1, 2, \ldots, n_0$; $^{c}\hat{\boldsymbol{\theta}}_i$ and $^{i}\hat{\boldsymbol{\theta}}_i$ contain the estimates of θ_i in the complete-data and incomplete-data scenarios, respectively; and m_0 is the sample size. As noted above, $n_0 = 65$ and $m_0 = 30$ in this section, resulting in a total of 3,900 estimates. Take the mean and variance of K_5 and the covariance between K_1 and K_5, which are three out of the 65 entries in $\boldsymbol{\theta}$, as an example. Their estimates in either the complete-data or incomplete-data scenarios are illustrated in Fig. 4, where those in the incomplete-data scenario appear to exhibit greater amounts of variability. Indeed, the sample standard deviations of the estimates in the complete- and incomplete-data scenarios are respectively 8.59×10^5 kN/m and 1.42×10^6 kN/m for Fig. 4(a), 5.53×10^{12} kN^2/m^2 and 2.62×10^{13} kN^2/m^2 for Fig. 4(b), and 3.81×10^{12} kN^2/m^2 and 4.55×10^{12} kN^2/m^2 for Fig. 4(c).

Now with all the necessary modules ready, the seismic fragilities can be sought. For a given set of $\hat{\theta}_i$ $(i = 1, 2, \ldots, n_0)$ and a specified seismic intensity, Monte Carlo simulation is used to compute the corresponding seismic fragility. Again, taking account of the uncertainty involved, at a specified seismic intensity two samples for the seismic fragility are constructed:

$$^{c}\mathbf{q} = (^{c}q_1, {}^{c}q_2, \ldots, {}^{c}q_{m_1})^{T} \tag{13}$$

in the complete-data scenario, and

$$^{i}\mathbf{q} = (^{i}q_1, {}^{i}q_2, \ldots, {}^{i}q_{m_1})^{T} \tag{14}$$

in the incomplete-data scenario. Let ${}^{c}Q$ and ${}^{i}Q$ be the underlying populations from which the samples ${}^{c}\mathbf{q}$ and ${}^{i}\mathbf{q}$ are respectively taken, and denote the cumulative distribution functions (CDFs) of ${}^{c}Q$ and ${}^{i}Q$ by $F_{{}^{c}Q}(z)$ and $F_{{}^{i}Q}(z)$, respectively. The following statistical hypotheses can be constructed:

$$H_0\text{: } \forall\ z, \qquad F_{{}^{c}Q}(z) = F_{{}^{i}Q}(z)$$
$$H_1\text{: } \exists\ z, \qquad \ni F_{{}^{c}Q}(z) \neq F_{{}^{i}Q}(z) \tag{15}$$

The objective of setting up this pair of hypotheses is to provide a decision-making procedure so that the efficacy of the formulated remedy for the structural appraisal data missingness can be assessed. In this study the two-sample Kolmogorov-Smirnov test is performed for the hypothesis testing task. In particular, the 3,900 estimates in Eqs. (11) and (12) are divided into three trials. Trial L1 comprises the estimates with the first subscript ranging from 1 to 10; Trial L2 contains those having a first subscript between 11 and 20; and all the other estimates belong to Trial L3. Thus corresponding to each trial, ${}^{c}\mathbf{q}$ and ${}^{i}\mathbf{q}$ both have a sample size of 10 ($m_1 = 10$). Further, to help reduce the error introduced by the Monte Carlo simulation, each of the above three trials is run twice independently, and this is indicated by appending to the trial number a letter A or B. For example, the first and second independent runs of Trial L1 are denoted by Trials L1A and L1B, respectively. With $u_{\mathrm{m}} = 0.013$ m and a significance level $\alpha = 0.05$, the samples and hypothesis testing results for the seismic fragilities are presented in Table 1, where the fragilities are with respect to a peak ground acceleration $\xi = 0.981\ \mathrm{m/s^2}$ or $3.924\ \mathrm{m/s^2}$. As one can see from Table 1, for any of the 12 pairs of the hypotheses no significant difference is exhibited between the fragilities in the complete-data scenario and those in the incomplete-data scenario. An appealing implication pops up: When a data missingness event occurs, it appears to be useful to try the formulated remedy as it may yield fragility evaluation results that are not significantly different from those based on the complete structural appraisal data.

Occurrence of the Appraisal Data Missingness at Multiple Stories

More complicated circumstances can now be looked at where the appraisal data missingness may occur at more than one story of the shear frame. As an example, assume that each entry in $\mathbf{K}_{\mathrm{r}}'$ has a missingness probability p of 0.15. As in the above subsection,

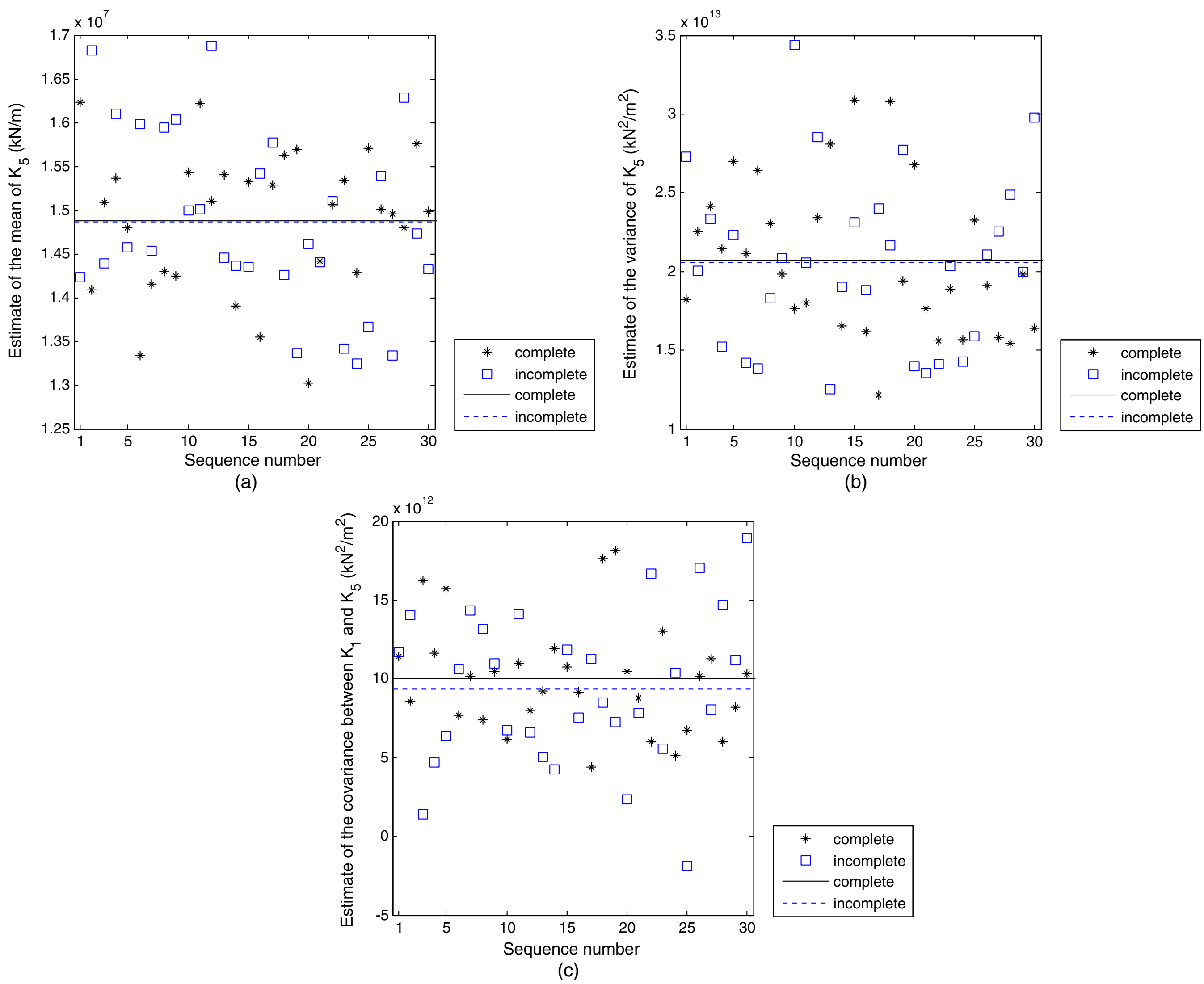

Fig. 5. For the case that the appraisal data missingness can occur at more than one story of the 10-story shear frame, the estimates of: (a) the mean; (b) variance of K_5; (c) the covariance between K_1 and K_5

Table 2. Comparison between the Seismic Fragilities in the Complete-Data Scenario and Those in the Incomplete-Data Scenario Where the Data Missingness Can Occur at Multiple Stories of the Linear Shear Frame

Trial ID	Sample and hypothesis testing result for the seismic fragility when $\xi = 0.981\ \mathrm{m/s^2}$: Sample	Result	Sample and hypothesis testing result for the seismic fragility when $\xi = 3.924\ \mathrm{m/s^2}$: Sample	Result
L4A	0.025; 0.059; 0.022; 0.039; 0.031; 0.058; 0.049; 0.024; 0.055; 0.008 0.050; 0.028; 0.053; 0.030; 0.021; 0.045; 0.049; 0.040; 0.031; 0.045	T	0.785; 0.892; 0.781; 0.878; 0.836; 0.876; 0.884; 0.871; 0.873; 0.870 0.867; 0.853; 0.871; 0.852; 0.862; 0.821; 0.860; 0.844; 0.844; 0.890	T
L4B	0.034; 0.061; 0.020; 0.030; 0.033; 0.069; 0.041; 0.028; 0.046; 0.009 0.039; 0.032; 0.050; 0.017; 0.022; 0.042; 0.045; 0.034; 0.032; 0.033	T	0.785; 0.890; 0.804; 0.850; 0.855; 0.881; 0.891; 0.870; 0.870; 0.875 0.857; 0.837; 0.896; 0.866; 0.855; 0.812; 0.872; 0.861; 0.828; 0.863	T
L5A	0.033; 0.044; 0.037; 0.075; 0.044; 0.039; 0.006; 0.036; 0.030; 0.036 0.057; 0.014; 0.079; 0.040; 0.041; 0.028; 0.036; 0.079; 0.054; 0.014	T	0.839; 0.887; 0.831; 0.885; 0.851; 0.865; 0.826; 0.802; 0.848; 0.889 0.864; 0.817; 0.873; 0.903; 0.881; 0.851; 0.835; 0.871; 0.934; 0.872	T
L5B	0.033; 0.065; 0.040; 0.075; 0.035; 0.058; 0.010; 0.052; 0.037; 0.053 0.048; 0.018; 0.082; 0.046; 0.040; 0.017; 0.028; 0.072; 0.068; 0.018	T	0.839; 0.882; 0.788; 0.896; 0.894; 0.876; 0.840; 0.823; 0.850; 0.886 0.888; 0.853; 0.859; 0.897; 0.869; 0.840; 0.838; 0.883; 0.915; 0.858	T
L6A	0.013; 0.036; 0.037; 0.041; 0.031; 0.057; 0.023; 0.041; 0.021; 0.024 0.056; 0.059; 0.046; 0.039; 0.021; 0.021; 0.053; 0.041; 0.025; 0.098	T	0.866; 0.896; 0.845; 0.922; 0.885; 0.837; 0.892; 0.850; 0.830; 0.856 0.907; 0.874; 0.890; 0.921; 0.860; 0.831; 0.845; 0.843; 0.852; 0.894	T
L6B	0.019; 0.030; 0.024; 0.044; 0.020; 0.053; 0.027; 0.035; 0.022; 0.017 0.072; 0.057; 0.045; 0.041; 0.021; 0.032; 0.047; 0.036; 0.021; 0.098	T	0.869; 0.889; 0.835; 0.925; 0.850; 0.853; 0.883; 0.843; 0.867; 0.817 0.893; 0.856; 0.869; 0.908; 0.864; 0.836; 0.866; 0.844; 0.869; 0.899	T

two parallel scenarios (i.e., complete-data and incomplete-data scenarios) are analogously introduced with Eqs. (11)–(15) constructed accordingly. To summarize, $n = 10$, $n_0 = 65$, $m = 30$, $m_0 = 30$, and $m_1 = 10$. It is worth noting that, if the listwise estimation procedures are used, this set of the parameter values is expected to give rise to an additional information loss of 76.8% according to the pilot study. Some representative estimates based on the EM algorithm, together with those obtained in the complete-data scenario, are plotted in Fig. 5. With the sample means of the estimates in the complete-data scenario indicated by the horizontal solid lines and those in the incomplete-data scenario by the horizontal dashed lines, reasonably good agreement can be observed in all the three subfigures and especially in Fig. 5(a) where the sample mean of the estimates of the mean of K_5 in the incomplete-data scenario differs from its complete-data counterpart by a relative error of only 0.13%. The final fragility evaluation results are presented in Table 2, in which the trial IDs are designated in the same manner as in the single-story case. For all the trials investigated, the EM algorithm embedded remedy can in the incomplete-data scenario produce the results comparable to those in the complete-data scenario in the sense that the null hypothesis H_0 in Eq. (15) cannot be rejected at the significance level $\alpha = 0.05$.

Influence of the Correlations and the Missingness Probability

To study the influence of the correlations between the random variables in $\mathbf{K}'$ on the performance of the formulated remedy in evaluating the fragilities of the 10-story shear frame, two trials, i.e., Trials L7 and L8, are designed. In Trial L7 the correlation between K_i and K_j ($i, j = 1, 2, \ldots, 10$; $i \neq j$) is assumed to be 0.1, while 0.9 is used in Trial L8. In each trial the bootstrap technique see, e.g., (Efron and Tibshirani 1993) is applied to improve the simulation efficiency. Particularly, M bootstrap samples are constructed from ${}^{\mathrm{c}}\mathbf{q}$ for the complete-data scenario, and another M ones are based on ${}^{\mathrm{i}}\mathbf{q}$ for the incomplete-data scenario:

$$ {}^{\mathrm{c}}\mathbf{q}_i^* = ({}^{\mathrm{c}}q_{1i}^*, {}^{\mathrm{c}}q_{2i}^*, \ldots, {}^{\mathrm{c}}q_{m_2 i}^*)^{\mathrm{T}} \tag{16} $$

$$ {}^{\mathrm{i}}\mathbf{q}_i^* = ({}^{\mathrm{i}}q_{1i}^*, {}^{\mathrm{i}}q_{2i}^*, \ldots, {}^{\mathrm{i}}q_{m_2 i}^*)^{\mathrm{T}} \tag{17} $$

where $i = 1, 2, \ldots, M$; and m_2 is the size of the bootstrap samples. For each pair of ${}^{\mathrm{c}}q_i^*$ and ${}^{\mathrm{i}}q_i^*$ ($i = 1, 2, \ldots, M$), one can with analogy to Eq. (15) construct the statistical hypotheses shown below:

$$ \begin{aligned} &H_0\colon \forall\ z, \qquad F_{{}^{\mathrm{c}}Q_i^*}(z) = F_{{}^{\mathrm{i}}Q_i^*}(z) \\ &H_1\colon \exists\ z, \qquad \ni F_{{}^{\mathrm{c}}Q_i^*}(z) \neq F_{{}^{\mathrm{i}}Q_i^*}(z) \end{aligned} \tag{18} $$

where ${}^{\mathrm{c}}Q_i^*$ and ${}^{\mathrm{i}}Q_i^*$ are the underlying populations for ${}^{\mathrm{c}}\mathbf{q}_i^*$ and ${}^{\mathrm{i}}\mathbf{q}_i^*$, respectively; and $F_{{}^{\mathrm{c}}Q_i^*}(z)$ and $F_{{}^{\mathrm{i}}Q_i^*}(z)$ are the corresponding CDFs. Again, for a given significance level α, the two-sample Kolmogorov-Smirnov test is carried out. The results of all these M hypothesis tests are represented by an acceptance rate R_{a}, which is defined as the ratio of M_0 (i.e., the number of the tests in which H_0 cannot be rejected at the significance level α) to M (i.e., the total number of the tests):

$$ R_{\mathrm{a}} = M_0/M \tag{19} $$

Thereby, R_{a} varies between 0 and 100%, and a higher value of R_{a} may indicate superior performance of the formulated remedy. Note that the implication here is that there may exist a situation in which the null hypothesis has to be rejected. For $\xi = 2.453\ \mathrm{m/s^2}$, $m = 30$, $p = 0.3$ for each entry in $\mathbf{K}_{\mathrm{r}}'$, $m_0 = 30$, $m_1 = 30$, $M = 1{,}000$, and $m_2 = 5$, 10, or 15, Table 3 lists the results of

Table 3. Seismic Fragilities in the Complete-Data and Incomplete-Data Scenarios Corresponding to Different Values of the Correlations

Trial ID	Sample	Acceptance rate R_a (%) $m_2 = 5$	$m_2 = 10$	$m_2 = 15$
L7	0.490; 0.482; 0.585; 0.535; 0.559; 0.565; 0.606; 0.573; 0.471; 0.572; 0.470; 0.560; 0.562; 0.544; 0.559; 0.500; 0.467; 0.587; 0.565; 0.530; 0.402; 0.421; 0.483; 0.538; 0.488; 0.599; 0.571; 0.470; 0.613; 0.551	90.8	91.5	93.1
	0.492; 0.656; 0.502; 0.549; 0.519; 0.472; 0.578; 0.635; 0.543; 0.514; 0.545; 0.465; 0.527; 0.559; 0.647; 0.682; 0.568; 0.470; 0.596; 0.649; 0.555; 0.534; 0.578; 0.518; 0.522; 0.517; 0.542; 0.561; 0.591; 0.520			
L8	0.232; 0.296; 0.219; 0.255; 0.362; 0.339; 0.273; 0.295; 0.242; 0.244; 0.242; 0.160; 0.221; 0.240; 0.237; 0.222; 0.293; 0.333; 0.313; 0.238; 0.186; 0.276; 0.245; 0.342; 0.266; 0.354; 0.288; 0.146; 0.253; 0.273	83.4	79.8	82.5
	0.394; 0.206; 0.158; 0.260; 0.184; 0.175; 0.233; 0.256; 0.218; 0.212; 0.226; 0.186; 0.252; 0.307; 0.310; 0.238; 0.293; 0.257; 0.227; 0.266; 0.231; 0.210; 0.272; 0.319; 0.229; 0.283; 0.284; 0.228; 0.169; 0.160			

Note: The first and second three rows in each trial are the samples obtained from the complete-data and incomplete-data scenarios, respectively.

Table 4. Seismic Fragilities in the Complete-Data and Incomplete-Data Scenarios Corresponding to Different Values of the Missingness Probability

Trial ID	Sample	Acceptance rate R_a (%) $m_2 = 5$	$m_2 = 10$	$m_2 = 15$
	0.419; 0.464; 0.407; 0.440; 0.345; 0.385; 0.542; 0.407; 0.317; 0.377; 0.430; 0.436; 0.423; 0.295; 0.476; 0.449; 0.435; 0.443; 0.392; 0.390; 0.466; 0.593; 0.466; 0.440; 0.424; 0.344; 0.512; 0.471; 0.427; 0.459			
L9	0.394; 0.579; 0.377; 0.464; 0.478; 0.466; 0.403; 0.356; 0.430; 0.427; 0.510; 0.389; 0.351; 0.391; 0.393; 0.463; 0.327; 0.430; 0.406; 0.415; 0.336; 0.372; 0.425; 0.494; 0.422; 0.364; 0.376; 0.376; 0.397; 0.582	86.9	85.2	81.4
L10	0.420; 0.469; 0.374; 0.325; 0.559; 0.329; 0.445; 0.394; 0.506; 0.544; 0.412; 0.498; 0.455; 0.372; 0.492; 0.455; 0.432; 0.403; 0.467; 0.480; 0.473; 0.496; 0.479; 0.391; 0.367; 0.377; 0.458; 0.447; 0.471; 0.533	85.4	85.5	86.9

Note: The first three rows show the sample obtained from the complete-data scenario, and the second and third three rows are the two samples corresponding to their respective incomplete-data scenarios.

$^c\mathbf{q}$, $^i\mathbf{q}$, and R_a. It is clear that the R_a values are fairly high for both the trials. It can also be observed from this table that, with a difference of 7.4, 11.7, or 10.6 percentage points in R_a respectively for an m_2 value of 5, 10, or 15, the formulated remedy seems to perform better for the correlation configuration in Trial L7 than it does for that in Trial L8.

Concerning the influence of the missingness probability, two additional trials, i.e., Trials L9 and L10, are introduced. In Trials L9 and L10 the missingness probability p for each entry in $\mathbf{K}_r'$ is respectively taken as 0.25 and 0.35, whereas in both the trials 0.5 is used as the correlation between K_i and K_j ($i, j = 1, 2, \ldots, 10$; $i \neq j$). Keeping the other parameters the same as those in studying the influence of the correlations and carrying out the procedures directly analogous to Eqs. (16)–(19), the obtained $^c\mathbf{q}$, $^i\mathbf{q}$, and R_a results are summarized in Table 4. Specifically, when each entry in $\mathbf{K}_r'$ has the missingness probability p of 0.35, an R_a value of 85.4%, 85.5%, or 86.9% is achieved for $m_2 = 5$, 10, or 15, respectively. By contrast, if the listwise estimation procedures are turned to in this situation, an addition information loss of 97.9% would be expected to occur, and the probability of noninformativeness amounts to 0.939 as well.

Fragility Evaluation of Hysteretic Systems with Incomplete Appraisal Data

Hysteresis is a nonlinear phenomenon frequently observed and recorded by structural engineers in a seismic event. To further examine the effectiveness of the EM algorithm embedded remedy, this section is devoted to the fragility evaluation of a typical hysteretic system (Mostaghel 1999; Mostaghel and Byrd 2000), and particularly the seismic ground accelerations from an actual earthquake are used. On February 22, 2011, an M 6.3 earthquake hit New Zealand's southern city of Christchurch, leading to catastrophic life and property losses. New Zealand's Earthquake Commission (EQC), GNS Science, and Land Information New Zealand (LINZ) captured scores of the seismic ground accelerations during the earthquake and made them publicly accessible through the companion website GeoNet (http://www.geonet.org.nz/earthquake/historic-earthquakes/top-nz/quake-14.html). Utilizing this resource, the fragility evaluation in this section is based on a selected real seismic-ground-acceleration time history, which is depicted in Fig. 6(a). The equations of motion of the two-story bilinear hysteretic shear frame being considered are shown in Eqs. (20) and (21), and more details can be found in Mostaghel (1999) and Mostaghel and Byrd (2000):

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \gamma\mathbf{K}\mathbf{U}(t) + (1-\gamma)\tilde{\mathbf{K}}\mathbf{V}(t) = -\mathbf{M}\mathbf{X}(t) \quad (20)$$

$$\dot{\mathbf{V}}(t) = \mathbf{F}(\mathbf{U}(t), \mathbf{V}(t), \dot{\mathbf{U}}(t), \dot{\mathbf{V}}(t))\dot{\mathbf{U}}(t) \quad (21)$$

in which

$$\mathbf{V}(t) = (V_1(t), V_2(t))^{\mathrm{T}} \quad (22)$$

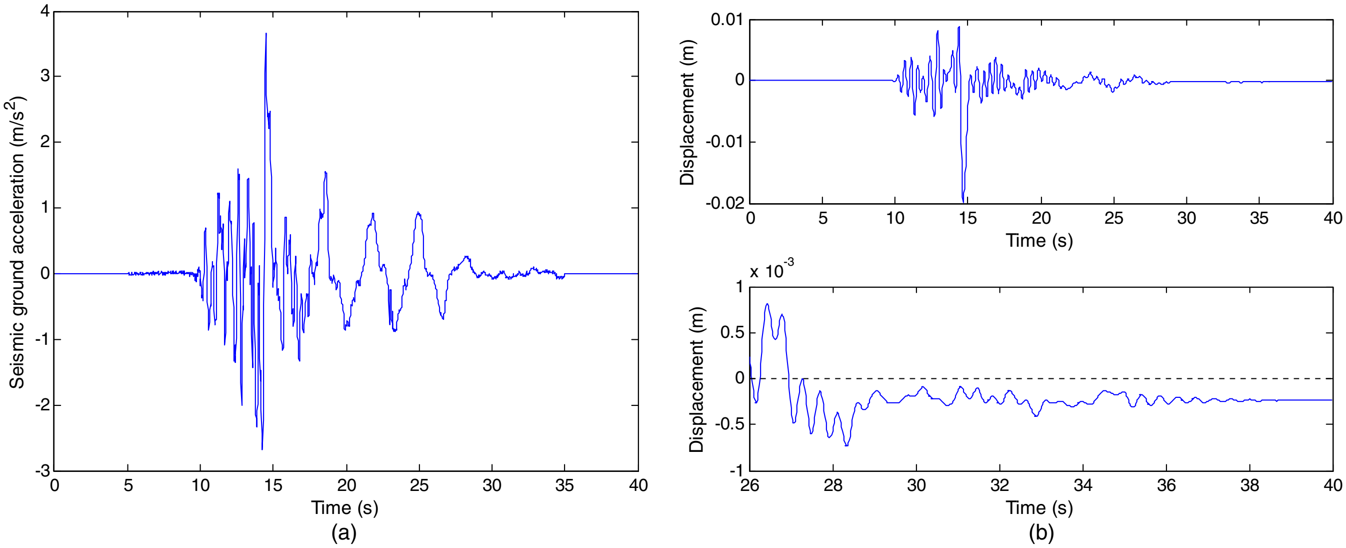

Fig. 6. (a) A horizontal seismic ground acceleration in the February 2011 Christchurch earthquake; and (b) an example of the corresponding displacement time history at the first story of the bilinear hysteretic shear frame when $\eta = 3.924\ \mathrm{m/s^2}$

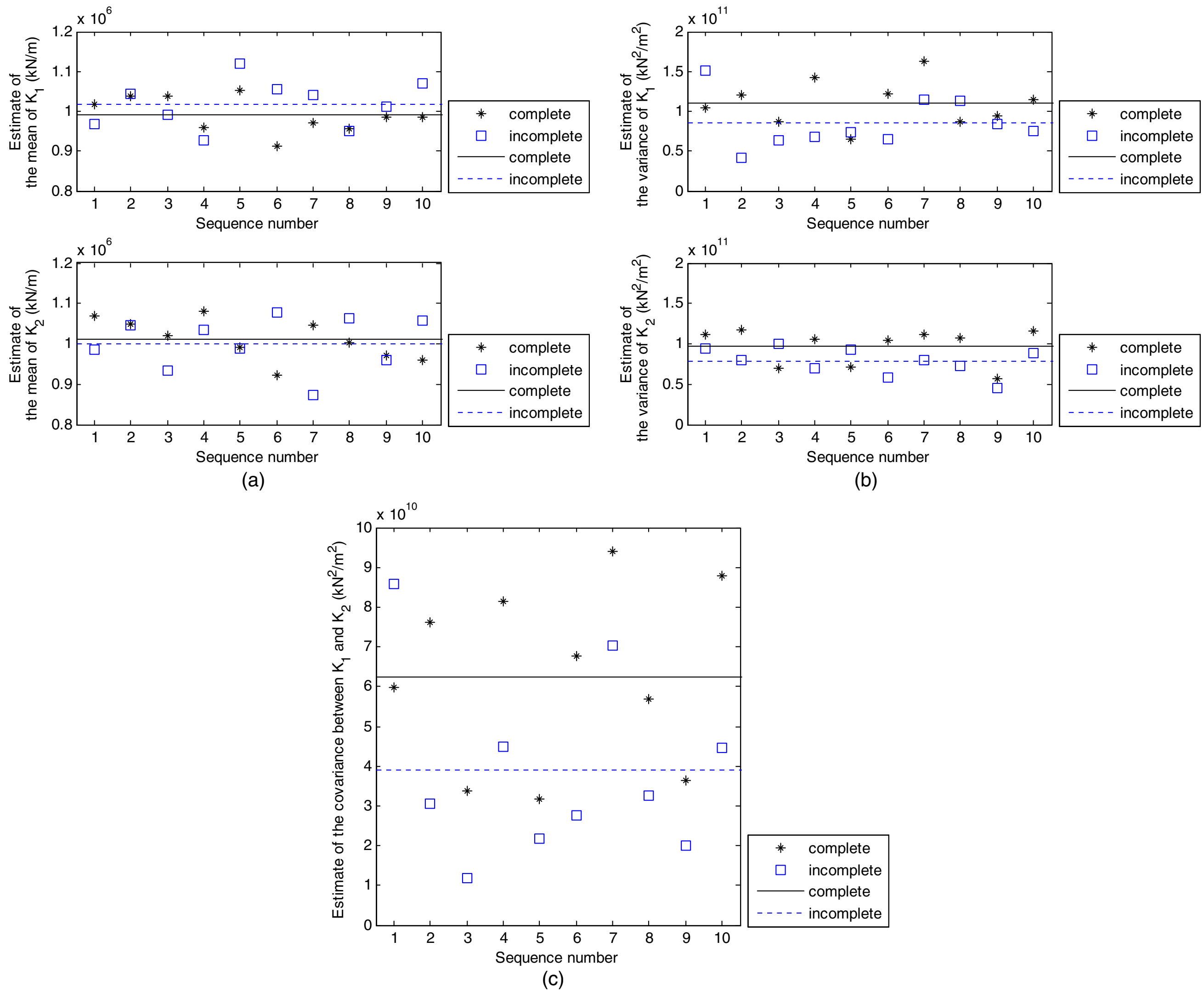

Fig. 7. For the case of the bilinear hysteretic shear frame, the estimates of: (a) the means; (b) variances of K_1 and K_2; and (c) the covariance between K_1 and K_2

Table 5. Comparison between the Seismic Fragilities in the Complete-Data Scenario and Those in the Incomplete-Data Scenario for the Hysteretic System

Trial ID	Sample and hypothesis testing result for the seismic fragility when $\eta = 3.924\ \text{m/s}^2$: Sample	Result	Sample and hypothesis testing result for the seismic fragility when $\eta = 7.848\ \text{m/s}^2$: Sample	Result
H11A	0.127; 0.115; 0.089; 0.181; 0.048; 0.191; 0.180; 0.122; 0.116; 0.142 0.182; 0.023; 0.114; 0.126; 0.055; 0.043; 0.128; 0.147; 0.082; 0.053	T	0.715; 0.683; 0.692; 0.710; 0.724; 0.807; 0.720; 0.792; 0.745; 0.725 0.724; 0.756; 0.804; 0.833; 0.684; 0.695; 0.689; 0.776; 0.756; 0.672	T
H11B	0.098; 0.117; 0.065; 0.165; 0.038; 0.215; 0.203; 0.130; 0.118; 0.148 0.177; 0.031; 0.101; 0.121; 0.049; 0.040; 0.122; 0.157; 0.085; 0.063	T	0.693; 0.647; 0.688; 0.745; 0.706; 0.795; 0.703; 0.777; 0.748; 0.747 0.730; 0.768; 0.816; 0.833; 0.636; 0.722; 0.695; 0.780; 0.742; 0.679	T

$$\tilde{\mathbf{K}} = \begin{pmatrix} K_1 & -K_2 \\ 0 & K_2 \end{pmatrix} \tag{23}$$

$$\mathbf{F}(\mathbf{U}(t),\ \mathbf{V}(t),\ \dot{\mathbf{U}}(t),\ \dot{\mathbf{V}}(t)) = \begin{pmatrix} f_0 & 0 \\ -f_1 & f_1 \end{pmatrix} \tag{24}$$

$$\begin{aligned} f_0 = {} & h_3(\dot{U}_1(t))(h_4(V_1(t) - \lambda\delta)h_4(U_1(t)) \\ & + h_4(V_1(t) - \delta)h_3(U_1(t))) \\ & + h_2(\dot{U}_1(t))(h_1(V_1(t) + \lambda\delta)h_1(U_1(t)) \\ & + h_1(V_1(t) + \delta)h_2(U_1(t))) \end{aligned} \tag{25}$$

$$\begin{aligned} f_1 = {} & h_3(\dot{U}_2(t) - \dot{U}_1(t))(h_4(V_2(t) - \lambda\delta)h_4(U_2(t) - U_1(t)) \\ & + h_4(V_2(t) - \delta)h_3(U_2(t) - U_1(t)) \\ & + h_2(\dot{U}_2(t) - \dot{U}_1(t))(h_1(V_2(t) + \lambda\delta)h_1(U_2(t) - U_1(t)) \\ & + h_1(V_2(t) + \delta)h_2(U_2(t) - U_1(t)) \end{aligned} \tag{26}$$

$$h_1(z) = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}; \qquad h_2(z) = \begin{cases} 0, & z \geq 0 \\ 1, & z < 0 \end{cases};$$
$$h_3(z) = \begin{cases} 1, & z > 0 \\ 0, & z \leq 0 \end{cases}; \qquad h_4(z) = \begin{cases} 0, & z > 0 \\ 1, & z \leq 0 \end{cases} \tag{27}$$

where $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$, t, $\mathbf{U}(t)$, and $\mathbf{X}(t)$ are analogously defined as in the prior section; $\mathbf{V}(t)$ contains the auxiliary unknown functions related to the nonlinear mechanism; $\tilde{\mathbf{K}}$ is the auxiliary stiffness matrix; γ is the post-yield-to-pre-yield stiffness ratio; λ is to quantify the distinction between the tension and compression strengths involved; and δ prescribes the deformation when yielding occurs. Besides, $\mathbf{K}'$, $\boldsymbol{\theta}$, $f(\mathbf{K}'|\boldsymbol{\theta})$, q, $\mathbf{K}_r'$, and $\hat{\boldsymbol{\theta}}$ are introduced in analogy with the linear system circumstances. Suppose that K_1 and K_2 have a bivariate normal distribution, and that each of them has a mean of 1×10^6 kN/m and a coefficient of variation of 0.3, and the correlation between them is 0.5. Both of the diagonal entries in the lumped-mass matrix $\mathbf{M}$ are taken as 1.25×10^6 kg. The Rayleigh damping is used for $\mathbf{C}$ where the coefficients involved are computed from the linear system corresponding to Eqs. (20) and (21) (i.e., $\gamma = 1$) with a modal damping ratio of 0.05 for both the modes. The fourth order Runge–Kutta method is applied to perform the time marching, and the time step size is chosen as 0.02 s. For the parameters γ, λ, and δ equal to 0.15, 0.1, and 0.005 m respectively, Fig. 6(b) illustrates a corresponding first-story displacement time history obtained when the seismic ground acceleration in Fig. 6(a) is scaled such that η, the pseudo-spectral acceleration with a natural frequency of 17.48 rad/s and a damping ratio of 0.05, becomes $3.924\ \text{m/s}^2$. The residual displacement associated with the hysteresis is clearly demonstrated in Fig. 6(b), where the original equilibrium position is marked by the horizontal dashed line. A complete-data scenario and an incomplete-data scenario are set up in the same way as in the preceding section, resulting in Eqs. (11)–(15). The parameters used are summarized below: $n = 2$, $n_0 = 5$, $m = 30$, $m_0 = 10$, $m_1 = 10$, $u_{\text{m}} = 0.02$ m, $p = 0.3$, and $\alpha = 0.05$. The means and variances of K_1 and K_2 and the covariance between K_1 and K_2 are estimated in the complete-data scenario by their respective sample statistics and in the in-complete data scenario by the EM algorithm, and the results are plotted in Fig. 7. For the data being considered, the EM algorithm used can be observed to produce in the incomplete-data scenario the estimates in closer agreement with those from the complete-data scenario for the means than for the variances and the covariance, given that the agreement is measured by the difference between the sample means of the estimates as respectively described in Fig. 7 by the horizontal solid and dashed lines for the complete-data and incomplete-data scenarios. Table 5 lists the results in Trials H11A and H11B, where the suffixes A and B denote two independent runs of the entire Monte Carlo simulation process. In each trial, the seismic-fragility samples as in Eqs. (13) and (14) are reported at $\eta = 3.924\ m/s^2$ or $7.848\ m/s^2$. Besides the cases in the previous section regarding typical linear systems, the EM algorithm embedded remedy appears to be also effectual to the appraisal data missingness for the hysteretic shear frame in that, in terms of any of the four pairs of the hypotheses as in Eq. (15), the differences between the fragilities in the complete- and incomplete-data scenarios are not statistically significant enough to be discerned by the two-sample Kolmogorov-Smirnov test with $\alpha = 0.05$.

Concluding Remarks

With a data missingness event accepted as a fait accompli in relevant structural appraisal and health monitoring practice, seismic fragility evaluation is cast as an issue involving incomplete data. The formulation of the EM algorithm embedded remedy for the appraisal data missingness strives to provide a thinking-beyond-the-box alternative in addition to the substantial research effort in some well recognized, fast progressing fields germane to the advancement of the precautions. The major findings of the study, along with some recommendations for future research, are summarized in as below:

1. Straightforward as the pairwise and listwise parameter-estimation manners are in dealing with incomplete structural appraisal data, they may create problems for the subsequent seismic fragility evaluation. To be more specific, the pairwise manner could have difficulty in satisfying the relevant positive-semidefiniteness criterion while the listwise manner seems to be incompetent in regard to the additional information loss and the probability of noninformativeness defined.
2. On the other hand, as demonstrated by the examples of the linear and nonlinear hysteretic systems incorporating both numerically synthesized and practically recorded seismic ground accelerations, the proposed EM algorithm embedded remedy appears to be able to exhibit superior performance when only incomplete structural appraisal data are available for the evaluation of the seismic fragilities. Further, the meticulously devised parameter study illustrates, to some extent, the robustness of the remedy against the values of the pertinent correlations and missingness probability.
3. Implementation-wise, considerably more amount of effort was exerted to compute a fragility value for the nonlinear hysteretic systems than for the linear systems. Hence, it is recommended for future research that some surrogate model be established to alleviate the otherwise cumbersome computational burden especially when the performance of the proposed remedy in the fragility evaluation of the nonlinear hysteretic systems is to be investigated in a more comprehensive way.
4. Besides the applications to the linear or nonlinear hysteretic systems included in this study, the formulated remedy may also have the potential to contribute to the seismic safety design of sustainable civil structures. Indeed, such cutting-edge techniques as online SHM and wireless sensor network assisted structural assessment can help acquire with decent accuracy the latest information on the condition of a structure prone to seismic risk. These techniques form the first-level strategy toward sustainability: So long as they work appropriately and effectively, the health condition and seismic safety of the structure are being attended to arguably well throughout its service life. Imposed upon the first-level strategy and essentially introducing a second-level sustainability strategy, the formulated remedy could hopefully build up in the seismic safety design the preparedness to cope with the situations in which only incomplete structural appraisal data are practically available.

Notation

The following symbols are used in this paper:

$\mathbf{C}$ = damping matrix;
$\mathbf{K}$ = stiffness matrix;
$\tilde{\mathbf{K}}$ = auxiliary stiffness matrix for the hysteretic system;
$\mathbf{K}'$ = story-stiffness random vector;
$\mathbf{K}'_\mathrm{r}$ = realizations of $\mathbf{K}'$;
M = number of the bootstrap samples constructed from ${}^\mathrm{c}\mathbf{q}$, aka number of the bootstrap samples constructed from ${}^\mathrm{i}\mathbf{q}$;
$\mathbf{M}$ = mass matrix;
M_0 = out of the M hypothesis tests, number of the tests in which the null hypothesis cannot be rejected at the specified significance level;
m = sample size of the structural appraisal data $\mathbf{K}'_\mathrm{r}$;
m_0 = sample size of ${}^\mathrm{c}\hat{\theta}_i$ and ${}^\mathrm{i}\hat{\theta}_i$;
m_1 = sample size of ${}^\mathrm{c}\mathbf{q}$ and ${}^\mathrm{i}\mathbf{q}$;
m_2 = sample size of ${}^\mathrm{c}\mathbf{q}_i^*$ and ${}^\mathrm{i}\mathbf{q}_i^*$;
n = dimension of $\mathbf{K}'$;
n_0 = number of the distribution parameters for $\mathbf{K}'$;
P = probability measure for the seismic fragility definition;
p = missingness probability;
${}^\mathrm{c}Q$ = underlying population corresponding to the sample ${}^\mathrm{c}\mathbf{q}$;
${}^\mathrm{i}Q$ = underlying population corresponding to the sample ${}^\mathrm{i}\mathbf{q}$;
${}^\mathrm{c}Q_i^*$ = underlying population corresponding to the sample ${}^\mathrm{c}\mathbf{q}_i^*$;
${}^\mathrm{i}Q_i^*$ = underlying population corresponding to the sample ${}^\mathrm{i}\mathbf{q}_i^*$;
q = seismic fragility;
${}^\mathrm{c}\mathbf{q}$ = seismic fragility sample in the complete-data scenario;
${}^\mathrm{i}\mathbf{q}$ = seismic fragility sample in the incomplete-data scenario;
${}^\mathrm{c}\mathbf{q}_i^*$ = ith bootstrap sample constructed from ${}^\mathrm{c}\mathbf{q}$;
${}^\mathrm{i}\mathbf{q}_i^*$ = ith bootstrap sample constructed from ${}^\mathrm{i}\mathbf{q}$;
R_a = acceptance rate;
S = sample space for the seismic fragility definition;
$\mathbb{S}$ = σ-algebra on S for the seismic fragility definition;
T = maximum time considered;
t = time;
$\mathbf{U}(t)$ = displacement-time-history random vector;
$u_i(t)$ = realization of the ith entry in $\mathbf{U}(t)$;
u_m = preselected threshold displacement;
$\mathbf{V}(t)$ = auxiliary unknown functions for the hysteretic system;
$\mathbf{X}(t)$ = seismic-ground-acceleration random vector;
$x(t)$ = realization of an entry in $\mathbf{X}(t)$;
α = significance level;
γ = post-yield-to-pre-yield stiffness ratio;
δ = deformation when yielding occurs;
η = pseudo-spectral acceleration;
$\boldsymbol{\theta}$ = distribution parameters for $\mathbf{K}'$;
$\hat{\boldsymbol{\theta}}$ = estimate/estimator for $\boldsymbol{\theta}$;
${}^\mathrm{c}\hat{\boldsymbol{\theta}}_i$ = estimates of the ith entry in $\boldsymbol{\theta}$ in the complete-data scenario;
${}^\mathrm{i}\hat{\boldsymbol{\theta}}_i$ = estimates of the ith entry in $\boldsymbol{\theta}$ in the incomplete-data scenario;

λ = parameter quantifying the distinction between the tensile and compressive strengths involved; and
ξ = peak ground acceleration.

References

American Institute of Steel Construction (AISC). (2010). "Specification for structural steel buildings." *ANSI/AISC 360-10*, Chicago.
American Concrete Institute (ACI). (2011). "Building code requirements for structural concrete and commentary." *ACI 318-11*, Farmington Hills, MI.
ASTM. (2008). "Standard test method for rebound number of hardened concrete." *ASTM C805/C805M-08*, West Conshohocken, PA.
ASTM. (2009). "Standard practice for calculating sample size to estimate, with specified precision, the average for a characteristic of a lot or process." *ASTM E122-09*, West Conshohocken, PA.
British Standards Institution (BSI). (2004). "Eurocode 2: Design of concrete structures—Part 1-1: General rules and rules for buildings." *BS EN 1992-1-1:2004*, London.
British Standards Institution (BSI). (2005). "Eurocode 3: Design of steel structures—Part 1-1: General rules and rules for buildings." *BS EN 1993-1-1:2005*, London.
British Standards Institution (BSI). (2007). "Assessment of in-situ compressive strength in structures and pre-cast concrete components." *BS EN 13791:2007*, London.
Casciati, F., Cimellaro, G. P., and Domaneschi, M. (2008). "Seismic reliability of a cable-stayed bridge retrofitted with hysteretic devices." *Comput. Struct.*, 86(17–18), 1769–1781.
Celik, O. C., and Ellingwood, B. R. (2010). "Seismic fragilities for non-ductile reinforced concrete frames—Role of aleatoric and epistemic uncertainties." *Struct. Saf.*, 32(1), 1–12.
Chang, F.-K., ed. (2011). *Proc., 8th Int. Workshop on Structural Health Monitoring*, DEStech Publications, Lancaster, PA.
Choe, D.-E., Gardoni, P., Rosowsky, D., and Haukaas, T. (2008). "Probabilistic capacity models and seismic fragility estimates for RC columns subject to corrosion." *Reliab. Eng. Syst. Saf.*, 93(3), 383–393.
Cimellaro, G. P., and Reinhorn, A. M. (2011). "Multidimensional performance limit state for hazard fragility functions." *J. Eng. Mech.*, 47–60.
Cimellaro, G. P., Roh, H., and De Stefano, A. (2009). "Spectral and fragility evaluations of retrofitted structures through strength reduction and enhanced damping." *Earthquake Eng. Vib.*, 8(1), 115–125.
Clough, R. W., and Penzien, J. (1993). *Dynamics of structures*, 2nd Ed., McGraw-Hill, New York.
Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum likelihood from incomplete data via the EM algorithm." *J. R. Stat. Soc. Ser. B Methodol.*, 39(1), 1–38.
Efron, B., and Tibshirani, R. J. (1993). *An introduction to the bootstrap*, Chapman & Hall, New York.
Geyskens, P., Der Kiureghian, A., and Monteiro, P. (1998). "Bayesian prediction of elastic modulus of concrete." *J. Struct. Eng.*, 89–95.
Hassan, M., Burdet, O., and Favre, R. (1995). "Ultrasonic measurements and static load tests in bridge evaluation." *NDT and E Int.*, 28(6), 331–337.
Heitjan, D. F., and Basu, S. (1996). "Distinguishing 'missing at random' and 'missing completely at random'." *Am. Stat.*, 50(3), 207–213.
Kafali, C., and Grigoriu, M. (2007). "Seismic fragility analysis: Application to simple linear and nonlinear systems." *Earthquake Eng. Struct. Dyn.*, 36(13), 1885–1900.
Kim, S.-H., and Shinozuka, M. (2004). "Development of fragility curves of bridges retrofitted by column jacketing." *Probab. Eng. Mech.*, 19(1–2), 105–112.
Leshchinsky, A. M. (1992). "Non-destructive testing of concrete strength: Statistical control." *Mater. Struct.*, 25(2), 70–78.
Lupoi, G., Franchin, P., Lupoi, A., and Pinto, P. E. (2006). "Seismic fragility analysis of structural systems." *J. Eng. Mech.*, 385–395.
Maes, M. A. (2002). "Updating performance and reliability of concrete structures using discrete empirical Bayes methods." *J. Offshore Mech. Arct. Eng.*, 124(4), 239–244.
Meng, X.-L., and van Dyk, D. (1997). "The EM algorithm—An old folk-song sung to a fast new tune." *J. R. Stat. Soc. Ser. B - Methodol.*, 59(3), 511–567.
Mostaghel, N. (1999). "Analytical description of pinching, degrading hysteretic systems." *J. Eng. Mech.*, 125(2), 216–224.
Mostaghel, N., and Byrd, R. A. (2000). "Analytical description of multi-degree bilinear hysteretic system." *J. Eng. Mech.*, 126(6), 588–598.
norm: Analysis of Multivariate Normal Datasets with Missing Values [Computer software]. A. A. Novo and J. L. Schafer, University Park, PA.
Oller, S., and Barbat, A. H. (2006). "Moment-curvature damage model for bridges subjected to seismic loads." *Comput. Methods Appl. Mech. Eng.*, 195(33–36), 4490–4511.
Park, J., Towashiraporn, P., Craig, J. I., and Goodno, B. J. (2009). "Seismic fragility analysis of low-rise unreinforced masonry structures." *Eng. Struct.*, 31(1), 125–137.
Pascale, G., Di Leo, A., and Bonora, V. (2003). "Nondestructive assessment of the actual compressive strength of high-strength concrete." *J. Mater. Civ. Eng.*, 15(5), 452–459.
R: A Language and Environment for Statistical Computing. [Computer software]. R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria.
Sasani, M., and Der Kiureghian, A. (2001). "Seismic fragility of RC structural walls: Displacement approach." *J. Struct. Eng.*, 127(2), 219–228.
Standards Australia. (1998). "Steel structures." *AS 4100-1998*, Sydney, New South Wales, Australia.
Standards Australia. (2009). "Concrete structures." *AS 3600-2009*, Sydney, New South Wales, Australia.
Sullivan, P. J. E. (1991). "Testing and evaluation of concrete strength in structures." *ACI Mater. J.*, 88(5), 530–535.
Wang, V. Z., Mallett, M., and Priory, A. (2011). "Sensitivity of the seismic fragility evaluation of a multi-story shear frame to the appraisal data missingness." *Proc., 2011 World Congress on Advances in Structural Engineering and Mechanics*, Techno-Press, Daejeon, South Korea, 188–195.
Wu, C. F. J. (1983). "On the convergence properties of the EM algorithm." *Ann. Stat.*, 11(1), 95–103.