Available online at www.sciencedirect.com
ScienceDirect
Structural Safety 30 (2008) 320–336
STRUCTURAL SAFETY
www.elsevier.com/locate/strusafe

Improved seismic fragility modeling from empirical data

Daniel Straub *, Armen Der Kiureghian

Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710, USA

Received 6 February 2007; received in revised form 15 May 2007; accepted 15 May 2007
Available online 19 July 2007

Abstract

The improved empirical fragility model addresses statistical dependence among observations of seismic performances, which arises from common but unknown factors influencing the observations. The proposed model accounts for this dependence by explicitly including common variables in the formulation of the limit state for individual components. Additionally, the fact that observations of the same component during successive earthquakes are correlated is considered in the estimation of the model parameters. As demonstrated by numerical examples considering the fragility of electrical substation equipment, the improved formulation can lead to significantly different fragility estimates than those obtained with the conventional assumption of statistical independence among the empirical observations. Furthermore, the conventional approach underestimates the statistical uncertainty associated with the resulting fragility estimates. The paper concludes with an investigation of the effects of statistical uncertainty and component statistical dependence on the system fragility. Numerical examples demonstrate that these effects are significant and must be addressed in the analysis of redundant systems.

Keywords: Bayesian analysis; Electrical substation equipment; Lifeline systems; Probabilistic models; Seismic fragility; Statistical dependence

1. Introduction

A seismic fragility model describes the performance of an engineering component or system subject to earthquake excitations in probabilistic terms. Fragility models are used to estimate the risk of earthquake hazard acting on the components of lifeline systems such as oil, water and gas pipelines, electrical power distribution systems, transportation networks and building systems. They are obtained either from statistical analysis of observed failures during past earthquakes (the empirical approach) or from structural modeling of the seismic performance of components and systems. Examples of the empirical approach are documented in Basöz et al. [1], Shinozuka et al. [2], O'Rourke and So [3], Der Kiureghian [4] and Osaki and Takada [5]. Examples of the structural modeling approach include Shinozuka et al. [2], Singhal and Kiremidjian [6] and Ellingwood [7]. Additionally, structural models have been combined with empirical fragility estimates, e.g.,

* Corresponding author.
E-mail address: straub@berkeley.edu (D. Straub).

0167-4730/$ - see front matter
doi:10.1016/j.strusafe.2007.05.004

Singhal and Kiremidjian [8] and Kim and Shinozuka [9], or have been extended with model correction factors calibrated to experiments, Gardoni et al. [10]. In this paper, we focus on the empirical modeling approach, but the considerations made are also relevant for fragility models obtained from structural analyses.

The main objective of this paper is the appropriate representation of uncertainties in the estimation of fragility and the investigation of their effects on system reliability. These uncertainties give rise to statistical dependence among the observations of seismic performances, based on which fragility is estimated. As will be shown, neglecting this statistical dependence can lead to large errors in the resulting fragility estimates. In the general statistical literature, the problem of dependence among observations is well known, e.g., Skellam [11], Williams [12] or Poortema [13] and various methods to account for these effects in regression models have been proposed. Straub and Der Kiureghian [14] modify such a regression model to represent seismic fragility, but that model can include only one level of statistical dependence. This paper aims at a more flexible and, thus, realistic model, including various levels of dependence. In addition to the above, statistical dependence among component performances can have a significant influence on the system fragility, particularly for redundant systems, and must be considered when analyzing the seismic performance of infrastructure systems. Finally, realistic estimation of the uncertainties highlights the advantages of collecting further information and improving the fragility models by including additional explanatory variables.

Consider a set of statistically dependent observations. Let the dependence among these observations be modeled through a set of factors, each factor being common among a different disjoint subset of the observations. Fig. 1 illustrates this idea for a set of observations with two distinct dependence factors. Each observation is noted by a symbol × in the figure. Observations that have factor 1 in common are grouped together by gray lines, and observations that have factor 2 in common are grouped together by dashed black lines. Of course one can have any number of factors and, hence, such groupings of the observations. Examples for such dependence structures in real-world observations are aplenty. Below, we describe two specific examples from civil engineering infrastructures subjected to earthquake effects.

Consider observations indicating the states (e.g., failure or survival) of a particular type of bridges (e.g., two-span, reinforced concrete highway bridges) in past earthquakes. Suppose we wish to use these data to construct a fragility model for the particular bridge type. Each observation indicates the state of a specific bridge in a specific earthquake event. Since a given bridge may have been subjected to multiple earthquake events, we may have multiple observations for each bridge. These can be grouped together to represent the dependence between the observations, which arises from the common capacity of the bridge. A second grouping of the observations may be considered in terms of the geographic locations of the bridges. If several bridges are in close proximity of each other, they will be subjected to essentially the same ground motion during each earthquake. Therefore, their states after each earthquake are likely to be dependent. To account for this effect, one would group together observations in each earthquake for bridges that are in close proximity of each other. Provided data is available, one could further refine this model by considering groupings of observations that account for such factors as the age of each bridge, the type of structural system or foundation, the local soil conditions, etc.

The second example relates to data on the performance of electrical substation equipment in past earthquakes, which is the specific application investigated in this paper. Each observation describes the number of failed/survived equipment items of specific type (viz. circuit breakers, transformers, surge arresters, etc.) within a substation affected by an earthquake. In this case, statistical dependence among the observations

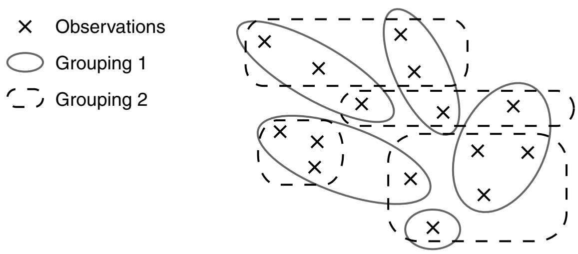

Fig. 1. Illustration of groupings of observations.

may exist due to the same equipment items being subjected to multiple earthquakes (grouping 1), all equipment within a substation being subjected to essentially the same ground motion (grouping 2), and other factors, such as the age or make of the equipment items, load conditions of each substation (i.e., whether the switches were open or closed during the earthquake), etc. Hence, multiple groupings may be used to describe the dependence structure, provided the necessary information is available. In the application in this paper, we consider two levels of grouping, corresponding to the groupings 1 and 2 described above.

2. Parameter estimation

We employ Bayesian statistics [15,16] to estimate the parameters $\boldsymbol{\theta}$ of the fragility model from the set of observations $\mathbf{z}$, which contains data on failures and survivals of equipment items in different electrical substations during past earthquake events. In this method, the posterior distribution of $\boldsymbol{\theta}$ given the observations, denoted $f(\boldsymbol{\theta}|\mathbf{z})$, is obtained by combining the prior distribution of $\boldsymbol{\theta}$, $f(\boldsymbol{\theta})$, and the likelihood of $\boldsymbol{\theta}$, $L(\boldsymbol{\theta}|\mathbf{z})$

$$f(\boldsymbol{\theta}|\mathbf{z}) \propto L(\boldsymbol{\theta}|\mathbf{z})f(\boldsymbol{\theta}). \tag{1}$$

The likelihood is a function that is proportional to the probability of making the observations given the parameters $\boldsymbol{\theta}$. The proportionality constant in (1) is obtained from the condition that integration of $f(\boldsymbol{\theta}|\mathbf{z})$ over the entire domain of $\boldsymbol{\theta}$ must yield one. For the application considered in this paper, no analytical solution is available and the integration over $f(\boldsymbol{\theta}|\mathbf{z})$ must be performed numerically. Here, we employ Markov chain Monte Carlo simulation (MCMC) to obtain the posterior statistics of $\boldsymbol{\theta}$. MCMC allows simulating directly from the posterior distribution with an unknown proportionality constant [17].

As an alternative to the Bayesian analysis, we perform maximum likelihood estimation (MLE) of the parameters [18]. The corresponding estimate, denoted $\boldsymbol{\theta}_{\mathrm{MLE}}$, is obtained by maximizing the likelihood or log-likelihood function with respect to the parameters, i.e.

$$\boldsymbol{\theta}_{\mathrm{MLE}} = \arg\max\{\ln[L(\boldsymbol{\theta}|\mathbf{z})]\}. \tag{2}$$

From a Bayesian viewpoint, the uncertainty in the estimation of $\boldsymbol{\theta}$ can be represented by a probability distribution. For large data, this probability distribution can be approximated by a multinormal with mean $\boldsymbol{\theta}_{\mathrm{MLE}}$ and a covariance matrix equal to the inverse of the Hessian matrix of the log-likelihood function evaluated at $\boldsymbol{\theta}_{\mathrm{MLE}}$ [18].

From the above description it is clear that we need to formulate the likelihood function for both approaches and the prior distribution for the Bayesian estimation. Derivation or selection of these functions for the considered fragility model are described in the following section.

3. Fragility model for electrical substation equipment

The electric power network is a vital lifeline infrastructure. An important subsystem within this network is the electrical substation, which consists of an interconnected system of electrical equipment, such as transformers, circuit breakers, switches, surge arrestors, capacitor banks, bus supports, etc. The reliability of the power system depends on the reliabilities of substations within the network, and the latter depend on the reliabilities of their constituent components, i.e., the electrical equipment in each substation. Our interest here is in estimating the seismic fragility of selected electrical substation equipment based on observations of their performances during past earthquakes. Seismic fragility is defined as the conditional probability of failure, i.e., the complement of reliability, for a given measure of the intensity of ground motion. When the performances of equipment items of a different type are statistically dependent, in addition to the marginal fragilities of the individual equipment types, one needs to determine their joint fragilities.

In this section, after describing the available data, we formulate the fragility model for a typical equipment item and derive the corresponding likelihood function and select an appropriate prior distribution. Results of the analysis are presented in the next section.

3.1. Available data

A large data-set on the performance of electrical substation equipment in past earthquakes has been compiled by Anagnos [19]. For each earthquake and each affected substation, the data provides the number of each equipment type in the substation and the number that failed during the earthquake. The "failure" state for an equipment item is defined as a state that makes the equipment inoperable after the earthquake. Measures of the intensity of the earthquake at the substation, expressed in terms of peak ground acceleration (PGA) and various spectral accelerations, are also provided. These were estimated based on the known magnitude and distance of the earthquake and existing attenuation laws, combined with recordings of ground motions at nearby sites. While we focus on the analysis of the performance of equipment items in electrical substations, the model presented in the following is suitable for any set of failure data to which the same or similar model assumptions apply. In particular, the model may be valid for other geographical groupings of observations and it can be formulated with other seismic intensity measures, such as peak ground velocity.

3.2. Model formulation

The fragility model presented is based on the following assumptions:

(a) We consider the performance history of a specific component type at all substations and earthquakes. Information on the equipment, age, repair history, etc., is not available. In common statistical terminology, this means that the observations for each equipment type are exchangeable.
(b) The measure of intensity of the ground motion (the demand) during each earthquake is selected as the PGA at the substation. As mentioned above, the reported PGA is an estimate based on an attenuation law. Therefore, it contains a measurement error, which must be accounted for in the analysis.
(c) Since the demand on each component is defined by a single variable, the PGA, the component capacity is also expressed in terms of this variable.
(d) Component capacities in different substations are statistically independent and identically distributed. Within each substation, we introduce a measure of dependence between the capacities of different components so as to allow for such common effects as the age of the substation and the functional state of the substation, e.g., whether the components under consideration were connected or disconnected at the time of the earthquake. Information about these effects is not available in the data, but they potentially introduce dependence between the component capacities within a substation, which may affect the parameter estimates.
(e) Obviously there is a modeling error in describing the earthquake demand on each component, since the PGA is only a crude measure of the destructive force of the ground motion. We assume this effect is random and that it is statistically independent for different earthquakes and for different substations during the same earthquake. The fact that the substations are far apart from one another justifies the latter assumption. However, for each earthquake, we assume this random effect is the same for all equipment within a substation.
(f) For substations that experience repeated earthquake events, we assume that a component that survives an earthquake has the same capacity during the next earthquake, while a component that fails is replaced with a new component before the next earthquake.

The dependences described in items (d–f) introduce three levels of groupings. However, as shown below, the groups for (d) and (e) are identical so that they can be combined. Thus, the problem involves two levels of groupings of dependent data.

In civil engineering, the performance of a structural component is typically modeled by a limit state function. Here, the performance of equipment i in substation j during earthquake k is modeled by the limit state function

$$g_{ijk} = R_{ij} Y_{R,jk} - S_{jk} Y_{S,jk}, \tag{3}$$

where R_{ij} is the uncertain intrinsic capacity of equipment i in substation j (expressed in terms of PGA), S_{jk} is the PGA of the ground motion at substation j during earthquake k, $Y_{S,jk}$ is an uncertain factor common to all equipment capacities in substation j during earthquake k (see item (d) above), and $Y_{S,jk}$ is the uncertain model error for the seismic demand in substation j during earthquake k (see item (e) above). Because $g_{ijk} \leqslant 0$ implies failure and $g_{ijk} > 0$ survival of the equipment, the above formulation is equivalent to

$$g_{ijk} = r_{ij} - s_{jk} + y_{jk}, \tag{4}$$

where $r_{ij} = \ln(R_{ij})$, $s_{jk} = \ln(S_{jk})$ and $y_{jk} = \ln(Y_{jk})$, in which $Y_{jk} = Y_{S,jk}/Y_{S,jk}$. Since our data does not contain information on the individual effects $Y_{S,jk}$ and $Y_{S,jk}$, our following analysis can only assess the combined effect Y_{jk}.

It is mathematically convenient and, therefore, common to represent the component capacities R_{ij} as independently and identically distributed lognormal random variables [2,7]. Similarly, we model Y_{jk} for different earthquakes and substations as independently and identically distributed lognormal random variables, which are also independent of R_{ij}. These imply independent normal distributions for r_{ij} and y_{jk}. Let μ_r and μ_y denote their means, and σ_r and σ_y denote their standard deviations. With the objective of making an unbiased estimate of μ_r, we set $\mu_y = 0$.

Let $\hat{s}_{jk}$ denote the estimated value of $\ln(S_{jk})$ based on the attenuation law. In general, this estimate is associated with a measurement (or estimation) error ε_{jk} such that the true value of $\ln(S_{jk})$ is $s_{jk} = \hat{s}_{jk} + \varepsilon_{jk}$. The limit state function for equipment i in substation j during earthquake k, thus, becomes

$$g_{ijk} = r_{ij} - \hat{s}_{jk} - \varepsilon_{jk} + y_{jk}. \tag{5}$$

We assume that ε_{jk} is independently and identically distributed for all substations and earthquakes and is described by a normal distribution with parameters μ_ε and σ_ε. These parameters should be estimated prior to analyzing the failure data, as the measurement uncertainty is related to the observations and not the fragility model. Typically $\mu_\varepsilon = 0$, i.e., the attenuation model is unbiased for the log PGA value. For the sake of simplicity of the notation, in the following we set

$$x_{jk} = \hat{s}_{jk} + \varepsilon_{jk} - y_{jk}, \tag{6}$$

so that

$$g_{ijk} = r_{ij} - x_{jk}. \tag{7}$$

It should be clear that x_{jk}, which can be viewed as an effective demand at substation j during earthquake k, are statistically independent normal random variables with means $\hat{s}_{jk}$ and common variances $\sigma_\varepsilon^2 + \sigma_y^2$ for all j and k.

3.3. Likelihood function

Let $\boldsymbol{\theta} = (\mu_r, \sigma_r, \sigma_y)$ denote the unknown parameters of the above model and let $L(\boldsymbol{\theta}|\mathbf{z})$ denote the likelihood function, where $\mathbf{z}$ represents the data. Given that the states of the equipment in different substations are statistically independent, we have

$$L(\boldsymbol{\theta}|\mathbf{z}) = \prod_j L(\boldsymbol{\theta}|\mathbf{z}_j), \tag{8}$$

where $\mathbf{z}_j$ is the data for substation j. Considering the fact that some equipment have identical capacities in successive earthquakes, and that the x_{jk} are statistically independent for different earthquakes, we can write

$$L(\boldsymbol{\theta}|\mathbf{z}_j) = \int_{\mathbf{x}_j} \prod_i \Pr(\mathbf{z}_{ij}|\mathbf{x}_j) f(\mathbf{x}_j)\,\mathrm{d}\mathbf{x}_j, \tag{9}$$

where $\mathbf{z}_{ij}$ is the data for equipment i in substation j (for all earthquakes), $\mathbf{x}_j$ is the collection of x_{jk} values for all earthquakes that affect substation j, and $f(\mathbf{x}_j)$ is the joint probability density function (PDF) of $\mathbf{x}_j$. Owing to the statistical independence of x_{jk}, $f(\mathbf{x}_j) = \prod_k f(x_{jk})$, where $f(x_{jk})$ is the normal PDF with mean $\hat{s}_{jk}$ and variance $\sigma_\varepsilon^2 + \sigma_y^2$.

To understand what is involved in computation of the above expression, we examine the cases where a substation is affected by one or two earthquakes. If substation j is affected only by earthquake k, then $\mathbf{z}_{ij}$ has only the one element z_{ijk} and $\mathbf{x}_j$ has only the one element x_{jk} and we can write

$$L(\boldsymbol{\theta}|\mathbf{z}_j) = \int_{-\infty}^{+\infty} \prod_i \Pr(z_{ijk}|x_{jk}) f(x_{jk})\,\mathrm{d}x_{jk}, \tag{10}$$

The conditional probability in the preceding expression is given by

$$\Pr(z_{ijk}|x_{jk}) = \Pr(r_{ij} - x_{jk} \leqslant 0|x_{jk}) = \Phi\left[\frac{x_{jk} - \mu_r}{\sigma_r}\right] \tag{11}$$

if equipment i in substation j has failed during earthquake k and

$$\Pr(z_{ijk}|x_{jk}) = \Pr(r_{ij} - x_{jk} > 0|x_{jk}) = \Phi\left[-\frac{x_{jk} - \mu_r}{\sigma_r}\right] \tag{12}$$

if it has survived, in which $\Phi[\,\cdot\,]$ denotes the standard normal cumulative probability function.

If substation j is first affected by earthquake k and then by earthquake l, then

$$L(\boldsymbol{\theta}|\mathbf{z}_j) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \prod_i \Pr(z_{ijk}, z_{ijl}|x_{jk}, x_{jl}) f(x_{jk}) f(x_{jl})\,\mathrm{d}x_{jk}\,\mathrm{d}x_{jl}. \tag{13}$$

The conditional probability for each equipment item now depends on the states of the equipment in the two successive earthquakes. If the equipment item survived the first earthquake, then its capacity was the same during the second event. In that case, if the equipment survived the second earthquake, we have

$$\Pr(z_{ijk}, z_{ijl}|x_{jk}, x_{jl}) = \Phi\left[-\frac{\max(x_{jk}, x_{jl}) - \mu_r}{\sigma_r}\right] \tag{14}$$

and if it failed the second earthquake, we have

$$\begin{aligned}\Pr(z_{ijk}, z_{ijl}|x_{jk}, x_{jl}) &= \Phi\left[\frac{x_{jl} - \mu_r}{\sigma_r}\right] - \Phi\left[\frac{x_{jk} - \mu_r}{\sigma_r}\right] && \text{for } x_{jk} < x_{jl},\\ &= 0 && \text{otherwise,}\end{aligned} \tag{15}$$

On the other hand, if the equipment item failed the first earthquake, then it must have been replaced and the capacity values during the two earthquakes are statistically independent. Thus, if the new equipment item survived the second earthquake, we have

$$\Pr(z_{ijk}, z_{ijl}|x_{jk}, x_{jl}) = \Phi\left[\frac{x_{jk} - \mu_r}{\sigma_r}\right]\Phi\left[-\frac{x_{jl} - \mu_r}{\sigma_r}\right] \tag{16}$$

and if it failed the second earthquake, we have

$$\Pr(z_{ijk}, z_{ijl}|x_{jk}, x_{jl}) = \Phi\left[\frac{x_{jk} - \mu_r}{\sigma_r}\right]\Phi\left[\frac{x_{jl} - \mu_r}{\sigma_r}\right]. \tag{17}$$

The above formulation can be extended to substations affected by 3 or more successive earthquakes, naturally with increasing number of integration folds and combinations of failed and survived equipment cases to be considered. However, not many electrical substations are subjected to repeated damaging earthquakes. For the particular data reported by Anagnos [19], most substations are subjected to only one earthquake, and a few substations are subjected to two earthquake.

To facilitate parameter estimation, it is convenient to introduce the following alternative parameterization:

$$\zeta = \sqrt{\sigma_r^2 + \sigma_y^2}, \tag{18}$$

$$\rho = \frac{\sigma_y^2}{\sigma_r^2 + \sigma_y^2}, \tag{19}$$

ζ is the logarithmic standard deviation representing the uncertainty in the fragility model, which includes the uncertainty in the equipment capacity R_{ij} and the uncertainty arising from the model error Y_{jk} for each substation and earthquake. ρ is the correlation coefficient between the safety margins of any pair of equipment items in a substation during each earthquake. This correlation arises from the common value of the model error Y_{jk} for all equipment in substation j and earthquake k. Together with μ_r, the set of unknown model parameters now are $\boldsymbol{\theta} = (\mu_r, \zeta, \rho)$. Note that μ_r is unbounded, $0 \leqslant \zeta$, and $0 \leqslant \rho \leqslant 1$.

3.4. Prior distribution

If information is available prior to the analysis, e.g., from engineering judgment based on previous experience with similar equipment, it may be used to select a prior distribution for $\boldsymbol{\theta}$. When this is not the case, a non-informative prior according to Box and Tiao [15] is appropriate. In practice, identification of the non-informative prior can be difficult in complex models involving several parameters and dependent observations, as shown by Gelman [20]. In such cases, an alternative is to use a weakly informative prior. A weakly informative prior may also be appropriate if the non-informative prior defies engineering judgment.

We consider the model parameters $\boldsymbol{\theta} = (\mu_r, \zeta, \rho)$ to be independent a-priori. The prior joint distribution is thus expressed as the product of prior marginal distributions. For μ_r, which is the mean of the logarithm of the component capacity, the non-informative prior is a diffuse distribution. This implies the improper distribution $f(m_S) \propto 1/m_S$ for the median m_S of the capacity. Computation with this prior yields a posterior distribution, which has unrealistically large densities for large median capacity values that defy engineering judgment. For this reason, a weakly informative prior is preferred. Thus, the prior distribution for μ_r is selected as normal with mean $\mu_{\mu_r} = -1.5$ and standard deviation $\sigma_{\mu_r} = 1.5$. This is equivalent to assuming that, with 90% probability, the median capacity lies roughly within 0.02–3 g. Later, we investigate the sensitivity of the fragility estimates to this selection.

For ζ, a diffuse prior distribution on $\ln(\zeta)$ is a good choice as a non-informative prior. This leads to the improper prior distribution

$$f(\zeta) \propto \frac{1}{\zeta}, \quad 0 \leqslant \zeta. \tag{20}$$

Since ρ is bounded by 0 and 1, a uniform prior distribution is an appropriate choice

$$f(\rho) = 1, \quad 0 \leqslant \rho \leqslant 1. \tag{21}$$

The selected prior joint probability density function for the parameters $\boldsymbol{\theta} = (\mu_r, \zeta, \rho)$, thus, is

$$\begin{aligned} f(\boldsymbol{\theta}) &\propto \tfrac{1}{\zeta} \exp\left[-\frac{1}{2}\left(\frac{\mu_r - \mu_{\mu_r}}{\sigma_{\mu_r}}\right)^2\right], \quad && 0 \leqslant \zeta,\ 0 \leqslant \rho \leqslant 1, \\ &= 0, && \text{otherwise}. \end{aligned} \tag{22}$$

3.5. Predictive model for equipment fragility

Based on (4), the limit state function for equipment i in substation j for a future earthquake with PGA equal to $\exp(s)$ can be written as $g_{ij} = r_{ij} - s + y_j$, where y_j is the model error term common to all equipment in the substation. Note that the measurement error term ε is not included because s is a given, not measured value. Since the equipment items are exchangeable, the fragility estimates are identical and, hence, in the following we drop the subscripts i and j.

For a given set $\boldsymbol{\theta} = (\mu_r, \zeta, \rho)$ of the model parameters, the conditional fragility of an equipment item during the future earthquake is given by

$$p(s|\boldsymbol{\theta}) = \Phi\left[\frac{s - \mu_r}{\zeta}\right]. \tag{23}$$

In particular, for a given point estimate $\hat{\boldsymbol{\theta}} = (\hat{\mu}_r, \hat{\zeta}, \hat{\rho})$ of the parameters, one obtains the fragility point estimate $p(s|\hat{\boldsymbol{\theta}})$. This estimate, however, does not account for the uncertainty present in the estimation of the model parameters $\boldsymbol{\theta}$. To account for this uncertainty, we introduce the predictive fragility model $\tilde{p}(s)$ defined by the total probability rule

$$\tilde{p}(s) = \int_{\boldsymbol{\theta}} p(s|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{z})\, \mathrm{d}\boldsymbol{\theta}. \tag{24}$$

The fragility for a single equipment item does not depend on the parameter ρ, the correlation coefficient between the safety margins of any two equipment items. However, the dependence between past observations influences the fragility of a single equipment item through the likelihood function and, therefore, the posterior distribution of the parameters $\boldsymbol{\theta}$. It is worth noting that the predictive fragility is the mean of the conditional fragility $p(s|\boldsymbol{\theta})$ over the posterior distribution of $\boldsymbol{\theta}$. It follows that if $\hat{\boldsymbol{\theta}}$ is selected as the posterior mean of $\boldsymbol{\theta}$, then $p(s|\hat{\boldsymbol{\theta}})$ represents a first-order approximation of $\tilde{p}(s)$. The conditional fragility $p(s|\boldsymbol{\theta})$ itself is an uncertain quantity with a distribution indicating the statistical uncertainty in the estimation of the fragility.

The above fragility estimate is valid for an equipment item for which no specific information on performances during past earthquake is available. If the equipment is known to have survived previous earthquakes, the above formulation is conservative, because such an observation has a censoring effect on the distribution of the intrinsic capacity of the equipment, r_{ij}. (Note that the subscripts are required here, because we are dealing with a specific equipment item.) In that case, the exact estimation is based on the posterior distribution of r_{ij}, which can be computed as $f(r_{ij}|\boldsymbol{\theta}, \mathbf{z}_j) \propto f(\mathbf{z}_j|\boldsymbol{\theta}, r_{ij}) f(r_{ij}|\boldsymbol{\theta})$, where $\mathbf{z}_j$ denotes all observations related to substation j. This computation requires an additional numerical integration; here we do not consider this case.

4. Numerical investigation

4.1. Data

For the numerical analysis, we consider the failure data on 1-phase 230 kV transformers (identified by TR1) and on 230 kV live tank GE ATB4-6 circuit breakers (identified by CB9) in electrical substations from Anagnos [19]. The data contains observed numbers of failures in different substations during different earthquakes, together with the estimated PGA at each substation site. The data, sorted by substation number and year of earthquake, are summarized in Tables 1 and 2. An earlier study with this data was reported by Der Kiureghian [4], where dependence of observations within a substation was considered but found to have no effect because

Table 1
Failure data for TR1 (1-phase 230 kV transformers)

j	k	Date	$\hat{S}_{jk}$ (g)	n_{jk}	N_{jk}	n_{jk}/N_{jk}
1	1	1971	0.07	0	13	0.00
1	6	1987	0.09	1	13	0.08
2	6	1987	0.20	0	3	0.00
3	1	1971	0.211	0	7	0.00
3	6	1987	0.22	0	7	0.00
4	3	1983	0.30	0	3	0.00
5	2	1978	0.28	7	12	0.58
6	6	1987	0.19	0	6	0.00
7	1	1971	0.097	0	6	0.00
7	6	1987	0.26	0	6	0.00
8	4	1984	0.06	0	3	0.00
9	8	1989	0.206	0	9	0.00
10	8	1989	0.237	3	10	0.30
11	8	1989	0.281	1	16	0.06
12	8	1989	0.206	0	9	0.00
13	6	1987	0.62	5	10	0.50
14	9	1994	0.367	2	4	0.50
15	8	1989	0.186	4	10	0.40
16	1	1971	0.536	1	6	0.17

j: Substation, k: earthquake, $\hat{S}_{jk}$: estimated PGA, n_{jk}: number of failures, and N_{jk}: number of TR1 equipment in substation j during k.

Table 2
Failure data for CB9 (230 kV live tank GE ATB4-6 circuit breakers)

j	k	Year	$\hat{S}_{jk}$ (g)	n_{jk}	N_{jk}	n_{jk}/N_{jk}
1	6	January 1987	0.086	2	12	0.17
2	6	January 1987	0.198	6	6	1.00
2	7	April 1987	0.110	1	1	1.00
7	1	1971	0.097	1	6	0.17
7	6	January 1987	0.259	3	6	0.50
9	8	1989	0.206	0	6	0.00
12	8	1989	0.206	2	9	0.22
15	8	1989	0.186	3	3	1.00
17	5	1986	0.836	6	6	1.00
18	6	January 1987	0.141	9	30	0.30
18	7	April 1987	0.079	3	21	0.14
19	6	January 1987	0.120	0	6	0.00
20	6	January 1987	0.112	0	23	0.00
21	6	January 1987	0.220	24	24	1.00
22	9	1994	0.570	5	10	0.50

j: Substation, k: earthquake, $\hat{S}_{jk}$: estimated PGA, n_{jk}: number of failures, and N_{jk}: number of CB9 equipment in substation j during k.

the selected prior distribution led to an improper posterior distribution. The present study overcomes this shortcoming through the alternative parameterization in (18) and (19).

The individual equipment items cannot be identified in Anagnos' data. However, for substations that experience repeated earthquakes (in this case a maximum of two), it must be established whether an equipment item that failed had survived previous earthquakes or had been newly installed after the failure of a previous equipment item. This problem is only relevant for the CB9 data. For these data, we assume that an equipment item that failed during a second earthquake had survived the previous earthquake, if the total number of failed equipment during the second earthquake is not greater than the number of equipment that survived the first earthquake. Thus, for example, for CB9 we assume that the three equipment that failed in substation $j = 7$ in the January 1987 earthquake were among the five that survived the 1971 earthquake. The assumption is reasonable, because new equipment is likely to be less vulnerable due to improvements in design and manufacturing. Ideally, these changes would be considered in the model by different values of the parameters describing the equipment capacity. However, because all sub-types are lumped together in the data and the model, this distinction cannot be made.

The measurement error ε_{jk} associated with the estimation of the PGA at each substation is assumed to have standard deviation $\sigma_\varepsilon = 0.3$. This value is based on engineering judgment and considers that the PGA estimates were obtained from attenuation laws modified with recordings at nearby sites.

As mentioned earlier, in addition to the marginal fragility estimates, a joint fragility estimate including the different types of equipment is of interest. However, this requires that a sufficient number of paired observations of both equipment types at the same substation during same earthquakes are available. Unfortunately, this is not the case in the above data-set, as is evident from comparison of Table 1 with Table 2. Therefore, no such analysis is performed here.

4.2. Posterior distribution

The posterior distribution of the parameters is computed using 20,000 MCMC samples. Listed in Table 3 are the posterior means, $\hat{\mathbf{M}}_{\boldsymbol{\theta}}$, standard deviations, $\hat{\mathbf{S}}_{\boldsymbol{\theta}}$, and correlation coefficients, $\hat{\mathbf{R}}_{\boldsymbol{\theta\theta}}$, of $\boldsymbol{\theta}$ for the two equipment types. It is noted that the posterior mean estimates of μ_r indicate a median capacity (in terms of PGA) of approximately $\exp(-0.03) = 0.970$ g for TR1 and $\exp(-1.71) = 0.181$ g for CB9. The variances ζ^2 indicate much more uncertainty in the estimated fragility of CB9 than TR1. Furthermore, the mean estimates of ρ indicate a much higher correlation between the performances of a pair of CB9 equipment in a substation than those of a pair of TR1 equipment. These results suggest that the circuit breakers are more influenced by the details of the ground motion beyond the PGA than are the transformers, i.e., the variable y_j plays a more significant role for the CB9

Table 3
Posterior statistics of model parameters

Equipment	$\boldsymbol{\theta}$	$\hat{\mathbf{M}}_{\boldsymbol{\theta}}$	$\hat{\mathbf{S}}_{\boldsymbol{\theta}}$	$\hat{\mathbf{R}}_{\boldsymbol{\theta\theta}}$
TR1	$\begin{bmatrix}\mu_r\\ \zeta\\ \rho\end{bmatrix}$	$\begin{bmatrix}-0.03\\ 1.39\\ 0.35\end{bmatrix}$	$\begin{bmatrix}0.63\\ 0.68\\ 0.17\end{bmatrix}$	$\begin{bmatrix}1 & 0.86 & 0.16\\ 0.86 & 1 & 0.23\\ 0.16 & 0.23 & 1\end{bmatrix}$
CB9	$\begin{bmatrix}\mu_r\\ \zeta\\ \rho\end{bmatrix}$	$\begin{bmatrix}-1.71\\ 5.06\\ 0.71\end{bmatrix}$	$\begin{bmatrix}0.88\\ 4.24\\ 0.12\end{bmatrix}$	$\begin{bmatrix}1 & -0.12 & -0.09\\ -0.12 & 1 & 0.16\\ -0.09 & 0.16 & 1\end{bmatrix}$

than for the TR1. The fact that the transforms have much higher resonant frequencies (typically 5–10 Hz) than the circuit breakers (typically 1–2 Hz) may explain this result. As is well known, at high frequencies the spectral acceleration converges on the PGA, while spectral ordinates at lower frequencies are influenced by the frequency content of the ground motion and, therefore, its detailed fluctuations for the given PGA. This suggests that, while the PGA may be a good measure of the destructive power of the earthquake for a high-frequency equipment such as the TR1, it may not be a good indicator of damage for CB9.

4.3. Comparison with the conventional approach

For comparison, we also apply the conventional fragility estimation approach, which neglects statistical dependence between the observations (i.e., no grouping of observations is considered). This corresponds to setting $\rho = 0$ and estimating the remaining two parameters from the data by assuming that, for given s_{jk}, the number of failures at a given substation during an earthquake is described by a Binomial distribution. For substation j and earthquake k, this distribution is specified by the parameter $p_{jk} = \Phi[(s_{jk} - \mu_r)/\zeta]$ and the number of equipment items in the substation. The same marginal prior distributions as above are selected for μ_r and ζ. The resulting posterior means, $\hat{\mathbf{M}}_{\boldsymbol{\theta}}$, standard deviations, $\hat{\mathbf{S}}_{\boldsymbol{\theta}}$, and correlation coefficients, $\hat{\mathbf{R}}_{\boldsymbol{\theta\theta}}$, are summarized in Table 4. The striking differences between these estimates and those in Table 3 are solely due to neglecting the statistical dependence between the observations.

One way to examine the validity of the Binomial model employed in the conventional formulation is to compare the dispersion in the predicted number of failed equipment with the dispersion observed in the data [21]. Such an analysis with the present data, which for the sake of brevity is not reported here, clearly shows that the Binomial model strongly under-predicts the dispersion observed in the data for both equipment items. It is concluded, therefore, that the Binomial model and the conventional approach are inappropriate for the present application. The results based on the conventional approach are reported here only for the purpose of demonstrating the importance of accounting for statistical dependence inherent among the observations.

4.4. Fragility estimates

To assess the influence of statistical dependence among the observations and to examine the effect of statistical uncertainty on the estimated fragility, three fragility curves are compared in Fig. 2: (a) the case where

Table 4
Posterior statistics of model parameters based on the conventional approach

Equipment	$\boldsymbol{\theta}$	$\hat{\mathbf{M}}_{\boldsymbol{\theta}}$	$\hat{\mathbf{S}}_{\boldsymbol{\theta}}$	$\hat{\mathbf{R}}_{\boldsymbol{\theta\theta}}$
TR1	$\begin{bmatrix}\mu_r\\ \zeta\end{bmatrix}$	$\begin{bmatrix}-0.48\\ 0.85\end{bmatrix}$	$\begin{bmatrix}0.29\\ 0.29\end{bmatrix}$	$\begin{bmatrix}1 & 0.86\\ 0.86 & 1\end{bmatrix}$
CB9	$\begin{bmatrix}\mu_r\\ \zeta\end{bmatrix}$	$\begin{bmatrix}-1.64\\ 0.48\end{bmatrix}$	$\begin{bmatrix}0.12\\ 0.13\end{bmatrix}$	$\begin{bmatrix}1 & 0.40\\ 0.40 & 1\end{bmatrix}$

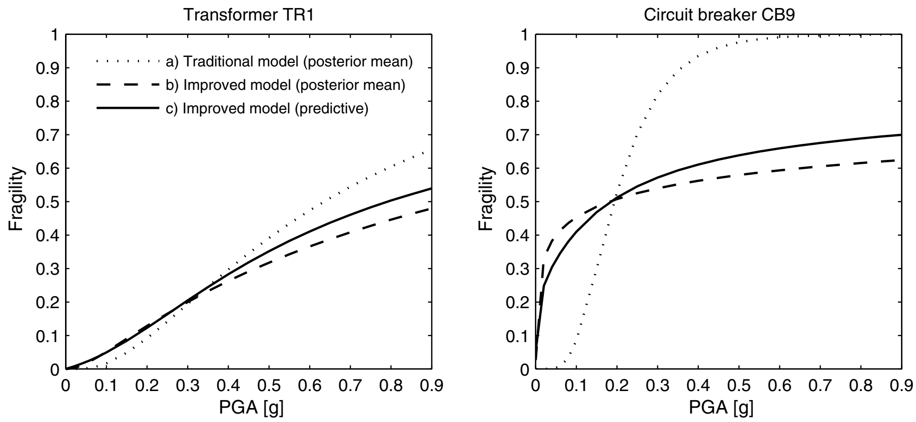

Fig. 2. Fragility curves for a single component.

statistical dependence and statistical uncertainty are both neglected, i.e., $p(s|\hat{\mathbf{M}}_{\boldsymbol{\theta}})$ with $\hat{\mathbf{M}}_{\boldsymbol{\theta}}$ as obtained by the conventional approach given in Table 4; (b) the case where only statistical dependence is considered, i.e., $p(s|\hat{\mathbf{M}}_{\boldsymbol{\theta}})$ with $\hat{\mathbf{M}}_{\boldsymbol{\theta}}$ based on the presented model as in Table 3; and (c) the case where both effects are included, i.e., the predictive model $\tilde{p}(s)$.

The difference between cases (a) and (b) in Fig. 2 demonstrates the influence of accounting for the statistical dependence among observations. The difference is much more pronounced for CB9 than for TR1. The estimates based on the conventional approach suggest a strongly informative fragility curve for CB9 (steeper curves indicate less uncertainty in the model) and a less informative fragility curve for the TR1. This result, however, is deceiving, as the estimates based on the improved model clearly indicate a much flatter fragility curve for CB9. In fact, one may conclude from the results in Fig. 2 that the PGA is not a good indicator of damage for CB9, as the fragility curve for this equipment tends to be nearly flat for PGA values greater than 0.1 g. The fact that the CB9 has a relatively low resonant frequency explains this phenomenon, as described earlier. The comparison between cases (b) and (c) in Fig. 2 shows the effect of the statistical uncertainty in the model parameters. As mentioned earlier, the point estimate $p(s|\hat{\mathbf{M}}_{\boldsymbol{\theta}})$ is a first-order approximation of the predictive fragility $\tilde{p}(s)$, so one does not expect a large difference between the two cases. Significant differences, however, may be observed in the tails of the fragility estimate, i.e., near 0 and 1 probability values. Although not evident in the scale of Fig. 2, the predictive fragility curve has larger values at the lower tail and smaller values at the tail near probability 1.

4.5. *Effect of statistical dependence among observations on the statistical uncertainty*

To further demonstrate the influence of the model assumptions on the statistical uncertainty, Figs. 3 and 4 show the predictive fragility estimates $\tilde{p}(s)$ together with 95% credible intervals (the Bayesian counterpart to confidence intervals) on the conditional fragility $p(s|\boldsymbol{\theta})$, representing the effect of the statistical uncertainty inherent in the parameter estimates. Recall that $\tilde{p}(s)$ is the mean of the distribution of $p(s|\boldsymbol{\theta})$. Additionally, the Figures include the observed failure rates. One should note that these are not realizations of the fragility, but of a discrete distribution describing the probability of n_{jk}/N_{jk} as a function of the fragility and N_{jk}. Thus, they are not directly related to the credible bounds of fragility. Fig. 3 presents the results obtained with the improved model and Fig. 4 those obtained with the conventional approach.

As observed in Figs. 3 and 4, consideration of statistical dependence among observations leads to larger uncertainty on the estimated fragility. This is because dependence among observations reduces the information content of the data. This effect is more pronounced for CB9, for which the dependence among observations is stronger (as evidenced by a larger estimated correlation coefficient). It follows that neglecting these dependences, as in the conventional approach, may lead to a serious underestimation of the statistical uncertainty and overconfidence in the estimated fragility.

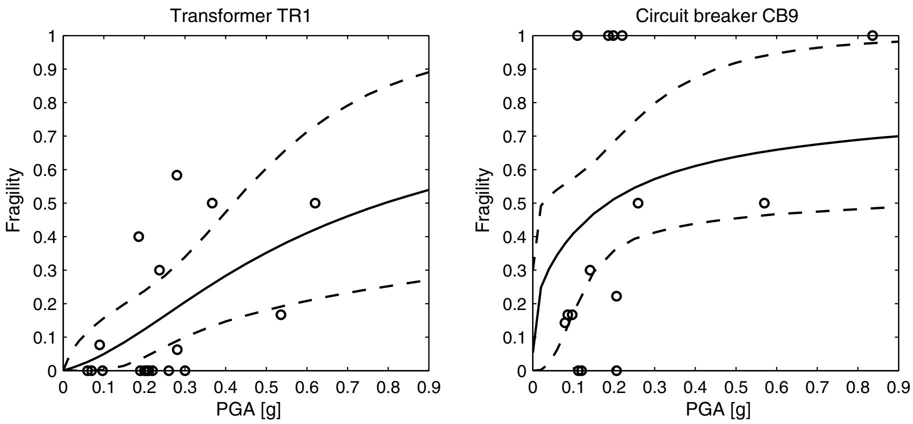

Fig. 3. Predictive fragility models (solid line), 95% credible intervals with respect to uncertainty in $\boldsymbol{\theta}$ (dashed lines) and observed failure rates (circles).

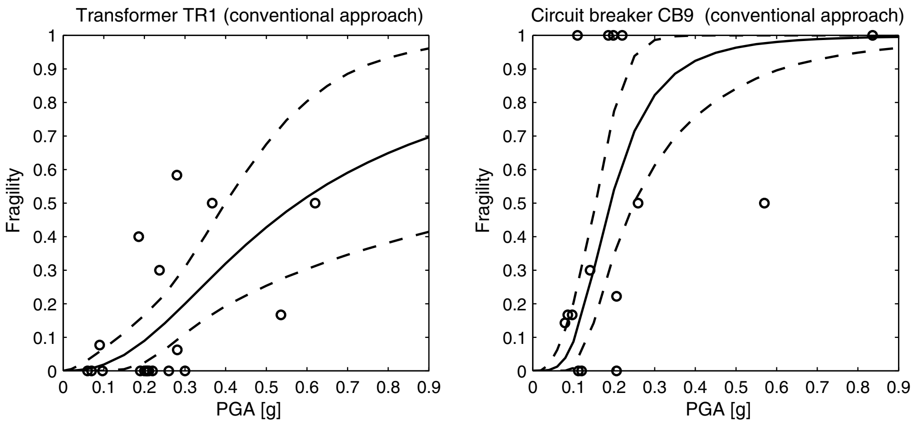

Fig. 4. Predictive fragility models (solid line) and 95% credible intervals with respect to uncertainty in $\boldsymbol{\theta}$ (dashed lines) as obtained with the conventional approach (assuming independence among observations). The observed failure rates are shown as circles.

4.6. Sensitivity of the results to the prior distribution

Validity of Bayesian analysis has been questioned in the past because of the need to select a prior distribution. We strongly believe that the Bayesian approach is the proper framework for the analysis of uncertainties. However, it is important to be aware of the effect of the prior distribution on the results. For this purpose, it is helpful to compare the results of the Bayesian analysis with those obtained by a maximum likelihood estimation (MLE), which does not employ a prior distribution. Fig. 5 shows the results based on the MLE for the three cases investigated earlier. These results exhibit the same trends as observed from the Bayesian analysis shown in Fig. 2. Here, because of the assumed normal distribution of the parameters, the MLE tends to underestimate the statistical uncertainty; therefore, the difference between curves (b) and (c) is somewhat smaller than in Fig. 2. By comparison of the results in Figs. 2 and 5, one concludes that the prior distribution selected in the Bayesian analysis does not have a strong influence on the fragility estimates.

To further examine the influence of the choice of the prior distribution, we perform a sensitivity analysis. To this end, the fragility estimates are compared for three choices of the prior distribution of μ_r: (a) the original selection, which is a normal distribution with mean $\mu_{\mu_r} = -1.5$ and standard deviation $\sigma_{\mu_r} = 1.5$ (equivalent to stating that the median capacity lies between 0.02 g and 3 g with 90% probability), (b) a normal distribution with mean $\mu_{\mu_r} = -1.5$ and standard deviation $\sigma_{\mu_r} = 0.9$ (equivalent to stating that the median

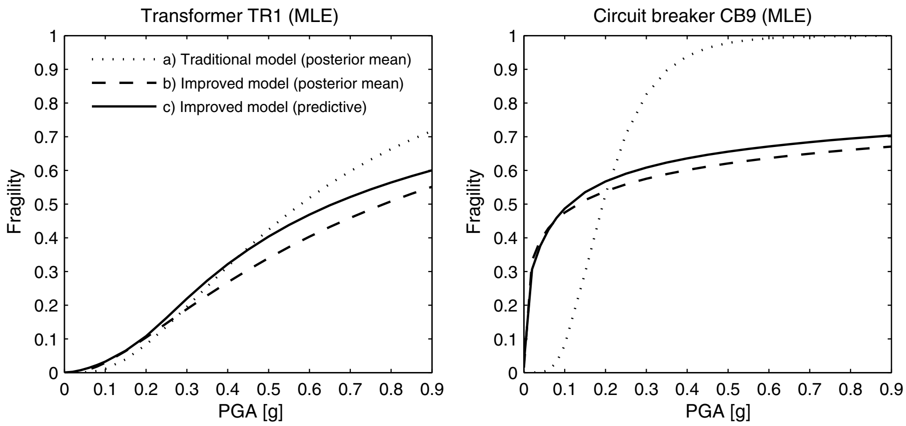

Fig. 5. Fragility curves for a single component, computed with MLE.

capacity lies between 0.05 g and 1 g with 90% probability), and (c) a normal distribution with $\mu_{\mu_r} = -2$ and $\sigma_{\mu_r} = 4$ (equivalent to stating that the median capacity lies between 2×10^{-4} g and 10^3 g with 90% probability). The marginal prior distributions for the other parameters, ζ and ρ, are not altered, as these are considered non-informative. The resulting predictive fragility models for TR1 and CB9 are shown in Fig. 6.

It is seen in Fig. 6 that the choice of the prior distribution has a moderate influence on the predictive fragility estimate. In general, a less informative prior (case b) results in a flatter (less informative) fragility curve. Not surprisingly, the differences are larger for values of the PGA for which few data are available (PGA > 0.3 g), this showing that the choice of the prior distribution is of importance mainly for extrapolation of the model. Ultimately, the choice of the prior distribution is that of the analyst. If it is believed that the results of the Bayesian analysis are too disperse, then the question to be asked is whether the chosen prior distribution does truly reflect the available information and engineering judgment. Only if this is not the case, a more informative prior distribution should be chosen; otherwise, the results of the Bayesian analysis reflect the fact that the observations provide only limited information. For the present case, we believe that the statement that the median capacity lies between 0.02–3 g with 90% probability is not overly confident. In fact, most experts, if consulted, would provide a narrower interval. The choice is therefore appropriate as a weakly informative prior.

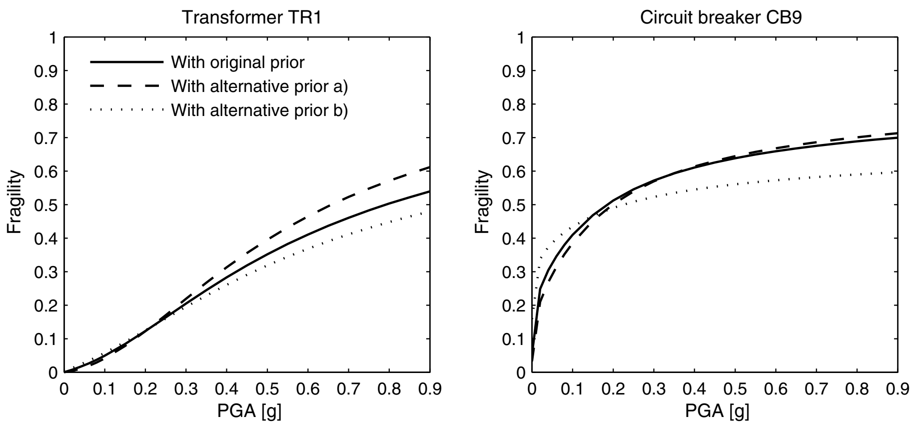

Fig. 6. Predictive fragility estimates obtained with alternative prior distributions of μ_r. (a) $\mu_{\mu_r} = -1.5$ and $\sigma_{\mu_r} = 1.5$; (b) $\mu_{\mu_r} = -1.5$ and $\sigma_{\mu_r} = 0.9$; and (c) $\mu_{\mu_r} = -2$ and $\sigma_{\mu_r} = 4$.

5. System reliability

The reliability of a system composed of a set of components depends not only on the reliabilities of the individual components, but also on statistical dependence that may exist among the component states. Thus when computing the reliability of an electrical substation system or the entire power network for a future earthquake event, it is necessary to consider not only the estimated fragilities of the individual equipment items, but also any statistical dependence that may exist between the estimated states of different equipment within each substation or the entire network. This dependence arises from two sources, as described below.

One source of statistical dependence among the estimated states of equipment items is the statistical uncertainty in the fragility model parameters $\boldsymbol{\theta}$, which are common to all equipment of the same type. This uncertainty, illustrated in Fig. 3, is epistemic in nature, i.e., it is caused by the limited number of observations and can be reduced by collecting and analyzing additional data. The statistical uncertainty introduces dependence among our estimates of the performances of all equipment items of the same type, irrespective of whether they belong to the same substation or not.

The second source of statistical dependence is the common uncertain influencing factors that affect all equipment in a substation during an earthquake. This effect has been represented in the limit state model for each equipment type through the random variable y (previously represented as y_{jk}). This variable is common for all equipment of similar type within a substation and leads to the correlation coefficient ρ between their limit states, as defined in (19). The uncertainty represented by y can be interpreted as aleatory. It primarily represents the variability of the future earthquake ground motion for a given PGA, which may be considered as inherently random. This source of statistical dependence has a physical nature and cannot be eliminated by additional data gathering. However, it may be reduced by refining the fragility model, e.g., using a more informative measure of the ground motion intensity than the PGA.

When computing the reliability of a power network for a future earthquake event, we must account for the above dependences and, therefore, must distinguish between equipment in the same substation and equipment in different substations. Here, we investigate the effect of the statistical dependence by considering substations idealized as K-out-of-N systems. Such a system consists of N components and functions if at least K, $1 \leqslant K \leqslant N$, components work. The extremes $K = 1$ and $K = N$ respectively represent parallel and series systems. Although real lifeline systems are more complex (e.g., Vanzi [22]), K-out-of-N systems allow demonstrating the effect of the statistical dependence for various degrees of redundancy. Here, we assume the system is composed of only one equipment type, either transformers or circuit breakers; we assume other equipment in the substation do not fail.

For a given $s = \ln(S)$, where S is the earthquake intensity in terms of PGA, the limit state function for a typical equipment item i is given by $g_i = r_i - s + y$, where y is the model error term. For given $\boldsymbol{\theta}$ and y, the fragility of the equipment is given by $p(s|\boldsymbol{\theta},y) = \Phi[(s - \mu_r - y)/\sigma_r]$, which is a common value for all equipment items in the substation. Since for the given $\boldsymbol{\theta}$ and y, the component states are statistically independent, the conditional system survival probability is given by the cumulative Binomial distribution with argument $(N - K)$ and parameters N and $p(s|\boldsymbol{\theta},y)$. The predictive system fragility, therefore, is

$$\tilde{P}(s) = 1 - \int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}|\mathbf{z}) \int_{-\infty}^{+\infty} \frac{1}{\sigma_y} \varphi\left(\frac{y}{\sigma_y}\right) \sum_{i=0}^{N-K} \binom{N}{i} [p(s|\boldsymbol{\theta}, y)]^i [1 - p(s|\boldsymbol{\theta}, y)]^{(N-i)} \, \mathrm{d}y \, \mathrm{d}\boldsymbol{\theta}, \tag{25}$$

where $\varphi()$ denotes the standard normal PDF. The integral on $\boldsymbol{\theta}$ in Eq. (25) is evaluated by using samples of $f(\boldsymbol{\theta}|\mathbf{z})$ generated during parameter estimation by MCMC.

To investigate the effects of the two types of statistical dependence described earlier, we consider the following two additional cases. First, the dependence arising from the variable y, which is physical in nature, is removed if we replace the above expression with

$$\tilde{P}_1(s) = 1 - \int_{\boldsymbol{\theta}} \sum_{i=0}^{N-K} \binom{N}{i} [p(s|\boldsymbol{\theta})]^i [1 - p(s|\boldsymbol{\theta})]^{(N-i)} f(\boldsymbol{\theta}|\mathbf{z}) \, \mathrm{d}\boldsymbol{\theta}. \tag{26}$$

If we assume equipment states are statistically independent, i.e., remove the dependences arising from both y and $\boldsymbol{\theta}$, the system fragility expression reduces to

$$\tilde{P}_2(s) = 1 - \sum_{i=0}^{N-K} \binom{N}{i} [\tilde{p}(s)]^i [1 - \tilde{p}(s)]^{(N-i)}. \tag{27}$$

Fig. 7 shows the influence of the statistical dependence between the components of redundant systems by comparing three fragility estimates for a parallel system with five components ($N = 5$, $K = 1$). The three cases considered are as follows: case (a) full account of statistical dependence, Eq. (25), case (b) only considering statistical dependence arising from statistical uncertainty, Eq. (26) and case (c) neglecting statistical dependence, Eq. (27).

For this parallel system, consideration of dependence among component performances has a significant influence on system failure probability, as illustrated by the difference between cases (a) and (b) in Fig. 7. The influence of the dependence arising from statistical uncertainty is reflected in the difference between the fragility curves for cases (b) and (c). In general, for parallel systems, neglecting statistical dependence among components leads to an overestimation of reliability. As observed in Fig. 7, the effect is larger for smaller

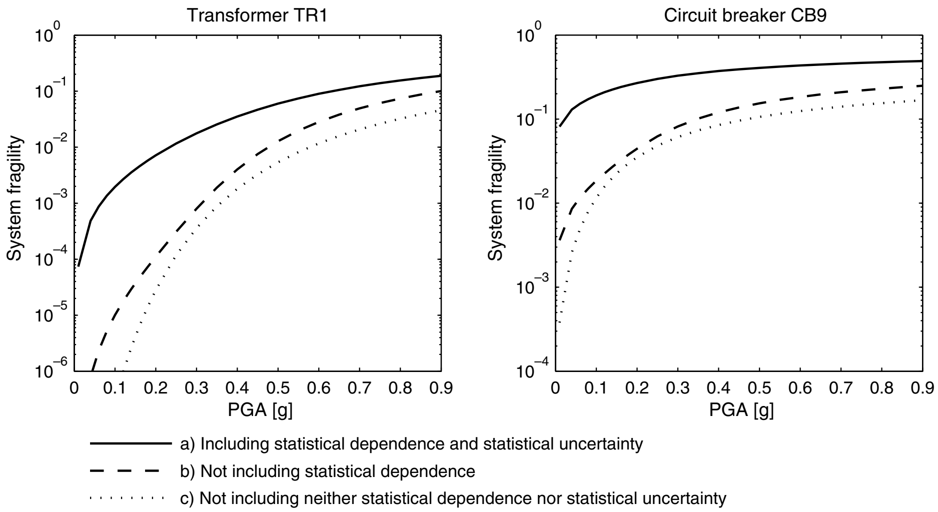

Fig. 7. Fragility curves for a parallel system with five components.

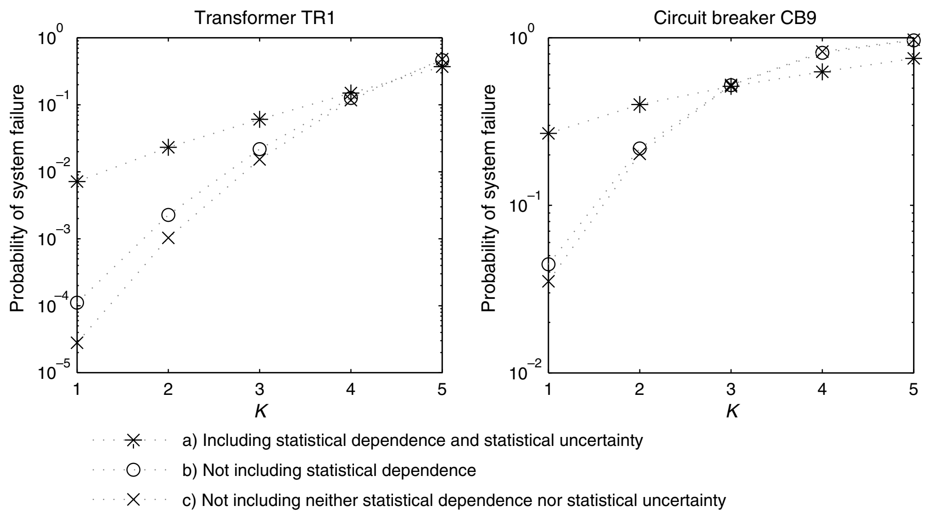

Fig. 8. Fragility for K-out-of-N systems with $N = 5$ (PGA = 0.2 g).

probabilities of failure, i.e., for lower values of the PGA. E.g., for a parallel system consisting of five TR1 equipment subject to an earthquake with PGA $= 0.2$ g, neglecting statistical dependence leads to an underestimation of the probability of system failure by approximately two orders of magnitude.

The effect of statistical dependence is a function of system redundancy. Unlike parallel systems, for series systems increasing dependence among components leads to increasing reliability. In this case, neglecting statistical dependence is conservative. This is illustrated in Fig. 8, which shows the estimates for the three cases as a function of K in a K-out-of-N system with $N = 5$ and PGA=0.2 g. As expected, for systems with little or no redundancy (i.e., large K), consideration of statistical dependence leads to lower estimates of the system failure probability (in the present case for $K = 5$ for TR1 and $K \geqslant 4$ for CB9). However, the relative difference between the estimates obtained under the different assumptions is smaller than for highly redundant systems. Therefore, neglecting statistical dependence in system reliability analysis is critical mainly for redundant systems.

In addition to dependence among component performances within a substation, as presented here, statistical uncertainty also introduces dependence among components at different substations. This can be considered in the analysis, as demonstrated in Straub and Der Kiureghian [14]. However, for the cases investigated there it has been found that this effect is relatively low and can be neglected for most practical purposes.

6. Concluding remarks

In assessing seismic fragility of structural components and systems from empirical data, it is important to properly account for possible statistical dependence among the observations. Such statistical dependence may arise from uncertain factors that are common among subsets of the observations, or from repeated observations of the same component. As demonstrated through an application to seismic performance data of electrical substation equipment, neglecting the statistical dependence among observations, as is conventionally done, can lead to erroneous fragility estimates. Furthermore, statistical dependence among observations in general tends to reduce the information content in the data and, hence, leads to increased statistical uncertainty in the estimation of the model parameters and the fragility. Neglecting the dependence may, therefore, lead to a level of confidence in the fragility estimates that is not supported by the data. Consideration of these problems has motivated the introduction of the fragility model in this paper, which explicitly addresses the effect of statistical dependence among empirical observations. For the specific application, three types of dependence are considered: dependence between the capacities of equipment in a substation due to factors common to the equipment, dependence among the seismic demands on the equipment in a substation due to their proximity, and dependence among observations for equipment items subjected to repeated earthquakes.

Results of the analysis for the two types of equipment items, transformers TR1 and circuit breakers CB9, show that the dependence among the observations is significant and has a pronounced effect on the fragility estimates. The estimated correlation coefficient between the limit states is 0.35 for TR1 and 0.71 for CB9. We believe the higher dependence for CB9 is due to its lower resonant frequency and, thus, higher sensitivity to the details of the ground motion for a given peak ground acceleration (PGA). Furthermore, the large statistical uncertainty in the estimation of the fragility curve for CB9 and its relatively flat form suggest that the PGA is not a good indicator of the damaging potential of the ground motion for this equipment.

When considering the fragility of a system, it is necessary to account for statistical dependence among the system components. This dependence can be due to common effects, such as described in the preceding paragraph, or due to statistical uncertainty present in the model parameters. As demonstrated by application of the model to idealized systems, neglecting these statistical dependences can lead to a significant overestimation of system reliability for redundant systems. Because most infrastructure systems are redundant, it is concluded that these effects are important and should generally be considered for reliability analysis of critical infrastructure systems.

Acknowledgement

The first author acknowledges support of the Swiss National Science Foundation (SNF) through Grant PA002-111428.

References

[1] Basöz NI, Kiremidjian AS, King SA, Law KH. Statistical analysis of bridge damage data from the 1994 Northridge, CA, earthquake. Earthquake Spectra 1999;15(1):25–54.
[2] Shinozuka M, Feng MQ, Lee J, Naganuma T. Statistical analysis of fragility curves. J Eng Mech, Trans ASCE 2000;126(12):1224–31.
[3] O'Rourke MJ, So P. Seismic fragility curves for on-grade steel tanks. Earthquake Spectra 2000;16(4):801–15.
[4] Der Kiureghian A. Bayesian methods for seismic fragility assessment of lifeline components. In: Taylor C, VanMarcke E, editors. Acceptable risk processes: lifelines and natural hazards, ASCE council on disaster reduction and technical council on lifeline earthquake engineering monograph, vol. 21; 2002, p. 61–77.
[5] Osaki T, Takada T. Seismic fragility assessment based on Bayesian updating. In: Applications of statistics and probability in civil engineering: proceedings of the 9th international conference, vol. 1; 2003, p. 185–9.
[6] Singhal A, Kiremidjian AS. Method for probabilistic evaluation of seismic structural damage. J Struct Eng, Trans ASCE 1996;122(12):1459–67.
[7] Ellingwood BR. Earthquake risk assessment of building structures. Reliab Eng Syst Safe 2001;74:251–62.
[8] Singhal A, Kiremidjian AS. Bayesian updating of fragilities with application to RC frames. J Struct Eng, Trans ASCE 1998;124(8):922–9.
[9] Kim SH, Shinozuka M. Development of fragility curves of bridges retrofitted by column jacketing. Probabilist Eng Mech 2004;19:105–12.
[10] Gardoni P, Der Kiureghian A, Mosalam KM. Probabilistic capacity models and fragility estimates for reinforced concrete columns based on experimental observations. J Eng Mech, Trans ASCE 2002;128(10):1024–38.
[11] Skellam JG. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J Roy Stat Soc, Ser B 1948;10(2):257–61.
[12] Williams DA. Extra-binomial variation in logistic linear models. Appl Stat 1982;31(2):144–8.
[13] Poortema K. On modelling overdispersion of counts. Stat Neerl 1999;53(1):5–20.
[14] Straub D, Der Kiureghian A. Seismic reliability assessment of infrastructure systems based on fragility models. In: Proc. ICASP 07, Tokyo, in press.
[15] Box GEP, Tiao GC. Bayesian inference in statistical analysis. New York: Wiley; 1992.
[16] Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. London: Chapman and Hall; 2003.
[17] Gilks WR, Richardson S, Spiegelhalter DJ. Markov chain Monte Carlo in practice. London: Chapman and Hall; 1996.
[18] Lindley DV. Introduction to probability and statistics from a Bayesian viewpoint, vol. 1 & 2. Cambridge University Press; 1965.
[19] Anagnos T. Development of an electrical substation equipment performance database for evaluation of equipment fragilities. Pacific Earthquake Engineering Center, Report no. 2001-06. http://www.engr.sjsu.edu/tanagnos/Substation/index.htm; 2001.
[20] Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Anal 2006;1(3):515–33.
[21] Collett D. Modelling binary data. London: Chapman and Hall; 1991.
[22] Vanzi I. Seismic reliability of electric power networks: methodology and application. Struct Safe 1996;18(4):311–27.