Soil Dynamics and Earthquake Engineering 69 (2015) 83–92

Contents lists available at ScienceDirect

Soil Dynamics and Earthquake Engineering

journal homepage: www.elsevier.com/locate/soildyn

Assessing the impact of ground-motion variability and uncertainty on empirical fragility curves

Ioanna Ioannou [a,*], John Douglas [b], Tiziana Rossetto [a]

[a] EPICentre, Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom
[b] BRGM – DRP/RSV, 3 avenue C. Guillemin, BP 36009, 45060 Orleans Cedex 2, France

ARTICLE INFO

Article history:
Received 27 June 2014
Received in revised form
10 October 2014
Accepted 25 October 2014
Available online 22 November 2014

Keywords:
Earthquake risk evaluation
Ground-motion fields
Empirical fragility curves
Spatial correlation
Kriging
Monte Carlo

ABSTRACT

Empirical fragility curves, constructed from databases of thousands of building-damage observations, are commonly used for earthquake risk assessments, particularly in Europe and Japan, where building stocks are often difficult to model analytically (e.g. old masonry structures or timber dwellings). Curves from different studies, however, display considerable differences, which lead to high uncertainty in the assessed seismic risk. One potential reason for this dispersion is the almost universal neglect of the spatial variability in ground motions and the epistemic uncertainty in ground-motion prediction. In this paper, databases of building damage are simulated using ground-motion fields that take account of spatial variability and a known fragility curve. These databases are then inverted, applying a standard approach for the derivation of empirical fragility curves, and the difference with the known curve is studied. A parametric analysis is conducted to investigate the impact of various assumptions on the results. By this approach, it is concluded that ground-motion variability leads to flatter fragility curves and that the epistemic uncertainty in the ground-motion prediction equation used can have a dramatic impact on the derived curves. Without dense ground-motion recording networks in the epicentral area empirical curves will remain highly uncertain. Moreover, the use of aggregated damage observations appears to substantially increase uncertainty in the empirical fragility assessment. In contrast, the use of limited randomly-chosen un-aggregated samples in the affected area can result in good predictions of fragility.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

1. Introduction

Fragility curves of buildings exposed to earthquakes express the likelihood of damage to these assets from future seismic events. Empirical fragility curves are based on the statistical analysis of post-earthquake observations of the damage sustained by the exposed buildings and the corresponding ground-motion intensity level at the building locations. Currently at least 119 empirical fragility curves have been published [1]. These curves have generally been constructed assuming that the measurement error in the intensity-measure levels (IMLs) is negligible. However, given the general lack of a dense strong-motion network in the areas of damaging earthquakes, the intensity levels are typically estimated though ground motion prediction equations (GMPEs) or, more recently, ShakeMaps. Hence, the IMLs are associated with high measurement error. In recent years, a handful of studies have proposed undertaking a Bayesian regression analysis to explicitly model this error [2–4]. Nonetheless, the impact of this measurement error on empirical fragility curves is not well understood.

This study aims to examine the impact of the measurement error in the IMLs on empirical fragility curves. A simulation study is undertaken to investigate this issue, following a similar philosophy to Gehl et al. [5], who studied the influence of the number of dynamic runs on the accuracy of fragility curves. In the next section the method of simulation is introduced. This approach is applied in the subsequent section to undertake a parametric analysis to study the influence of different assumptions on the empirical fragility curves. The paper finishes with some discussion of the results, the limitations of existing empirical fragility curves, implications for the development of future empirical fragility functions as well as possible ways forward.

2. Method

The impact of ground-motion variability and uncertainty on empirical fragility curves is studied here by undertaking a series of experiments. In these, an earthquake with specified characteristics

* Corresponding author.
E-mail address: ioanna.ioannou@ucl.ac.uk (I. Ioannou).

http://dx.doi.org/10.1016/j.soildyn.2014.10.024
0267-7261/© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/3.0/).

(i.e. magnitude, location and faulting mechanism) affects a number of buildings ($N_{Buildings}$) located in a number of towns (N_{Towns}).

The construction of empirical fragility curves requires observations of two variables, namely: the damage sustained by the considered buildings and their corresponding IMLs. IMLs are generated assuming the absence or the presence of ground-motion observations.

2.1. Seismic damage

In this study, the damage experienced by each building in the affected area is considered random due to the uncertainty in its IML as well as the uncertainty in its structural performance given this IML. Therefore, seismic damage for each building is determined here by modelling these two uncertainties through a Monte Carlo analysis. The procedure adopted is an extension of the procedure used by Douglas [6] in order to study the density of seismic networks required to monitor ground motions from induced seismicity. According to this analysis, a large number, $N_{Realisations}$, of IMLs and subsequent damage states are generated.

According to the procedure proposed by Douglas [6], each realisation of IMLs for the considered buildings occurs from the generation of a ground-motion field using a given GMPE coupled with models of spatial variability. To simulate the spatially-correlated ground-motion fields the procedure of Strasser and Bommer ([7], pp. 2625–2626) is used. The package geoR [8] of the statistical software R allows such fields to be generated quickly and then manipulated. The between-event and within-event ground-motion variabilities are included within the fields. The deterministic ground-motion field produced by evaluating the considered IMLs for all building locations in the region is perturbed by the addition of a random field derived from a multivariate normal distribution based on a standard deviation equal to the within-event variability of the selected GMPE and an exponential correlation function, $G(h)$, which is found to fit the observed spatial correlation of earthquake ground motions [9,10]:

$$G(h) = \exp\left(-\frac{h}{h_0}\right) \tag{1}$$

where h is the separation distance between locations of interest and h_0 is the correlation range. Because one ground-motion field differs greatly from another, this procedure is repeated many times so that robust conclusions can be drawn from the combined results. The sensitivity of the results on the chosen GMPE, the value of h_0 and other input parameters (e.g. size of the region, density of ground-motion measurements and aggregation level) are investigated in this paper.

In order to simulate earthquake-damage fields, a known fragility curve expressing the fragility of hypothetical buildings in a region is applied. This curve takes as input the simulated ground-motion fields and yields the building damage observations used as the empirical dataset for the study. Consequently, the impact of sparse or uncertain observations on fragility curves can be evaluated by comparing the resulting empirical fragility curves derived from different sampling and assumptions, with the curve used as input in the simulations. The advantage of this approach is that the 'true' fragility of the structures is known and can be compared with the empirical fragility curves resulting from the experiments.

In particular, for a realisation k, resulting in $iml_{realisationk}$, the damage sustained by each building is randomly generated as follows. In order to simplify the analysis, the determination of the exact damage state of each building is not required. Instead, we concentrate on whether the building has reached or not a given damage state, ds_i, assuming an appropriate fragility curve from the literature. In particular, for a realisation k, the building, j, is assigned an indicator, Y_{jk}, where:

$$Y_{jk} = \begin{cases} 1 & DS \geq ds_i \\ 0 & DS < ds_i \end{cases} \tag{2}$$

The indicator is randomly assigned to the building j, by assuming that it follows a special case of the binomial distribution, termed the Bernoulli distribution:

$$Y_{jk}|IM = iml_{realisation_k} \sim \binom{n}{y_{jk}} {\mu_{jk}}^{y_{jk}} \left[1-\mu_{jk}\right]^{n-y_{jk}}$$

$$\text{where } \mu_{jk} = P(DS \geq ds_i | im_{realisation_k}) = \Phi\left(\frac{\ln(iml_{realisation_k}) - \lambda}{\zeta}\right) \tag{3}$$

where n is the number of buildings for a given intensity measure level, $iml_{realisationk}$, and in this case, $n=1$; μ_j equals the probability that the building is in damage state ds_i or above given $iml_{realisationk}$; μ_{jk} is the mean of the Bernoulli distribution, which is typically expressed in the literature in terms of a cumulative lognormal distribution; Φ is the cumulative standard normal distribution; λ is the lognormal mean; and ζ is the lognormal standard deviation.

2.2. Ground-motion intensity

The determination of the IML at the location of each building is necessary for the construction of empirical fragility curves. These levels are considered known and measured without uncertainty, an assumption commonly made when deriving such curves. The determination of these 'true' IMLs depends on the absence or presence of ground-motion observations.

2.2.1. The absence of ground motion recording stations

In the absence of ground-motion records, empirical fragility curves are derived here by following the common assumption that the IML at the location of each building is equal to the median values obtained from a pre-selected GMPE. It should be noted that the selected equation is not necessarily the same as the one used to generate the damage levels because in practice the appropriate GMPE for an earthquake is not known.

2.2.2. The presence of ground motion recording stations

The random fields of peak ground accelerations (PGAs) are recovered assuming the presence of ground motion recording stations located at a varying number of buildings. This consideration suggests that the IMLs for the buildings at which records are available are known and equal to the corresponding values provided by the random field. The IMLs for the remaining buildings are estimated from these records using a procedure known as kriging. In this study, kriging uses the same correlation model as the one used for the generation of the random fields.

2.3. Empirical fragility curves

The empirical fragility curve is then constructed for the k realisations of IMLs by considering the Y_{jk} indicators generated for all considered buildings, according to the procedure described in Section 2.1, and the corresponding 'true' IMLs as determined in Section 2.2. Their construction follows the procedure proposed in the Global Earthquake Model empirical fragility assessment guidelines [3]:

$$Y_{jk}|IM_{true} = iml_{'true',k} \sim \binom{n}{y_{jk}} {\mu_{jk}}^{y_{jk}} \left[1-\mu_{jk}\right]^{n-y_{jk}}$$

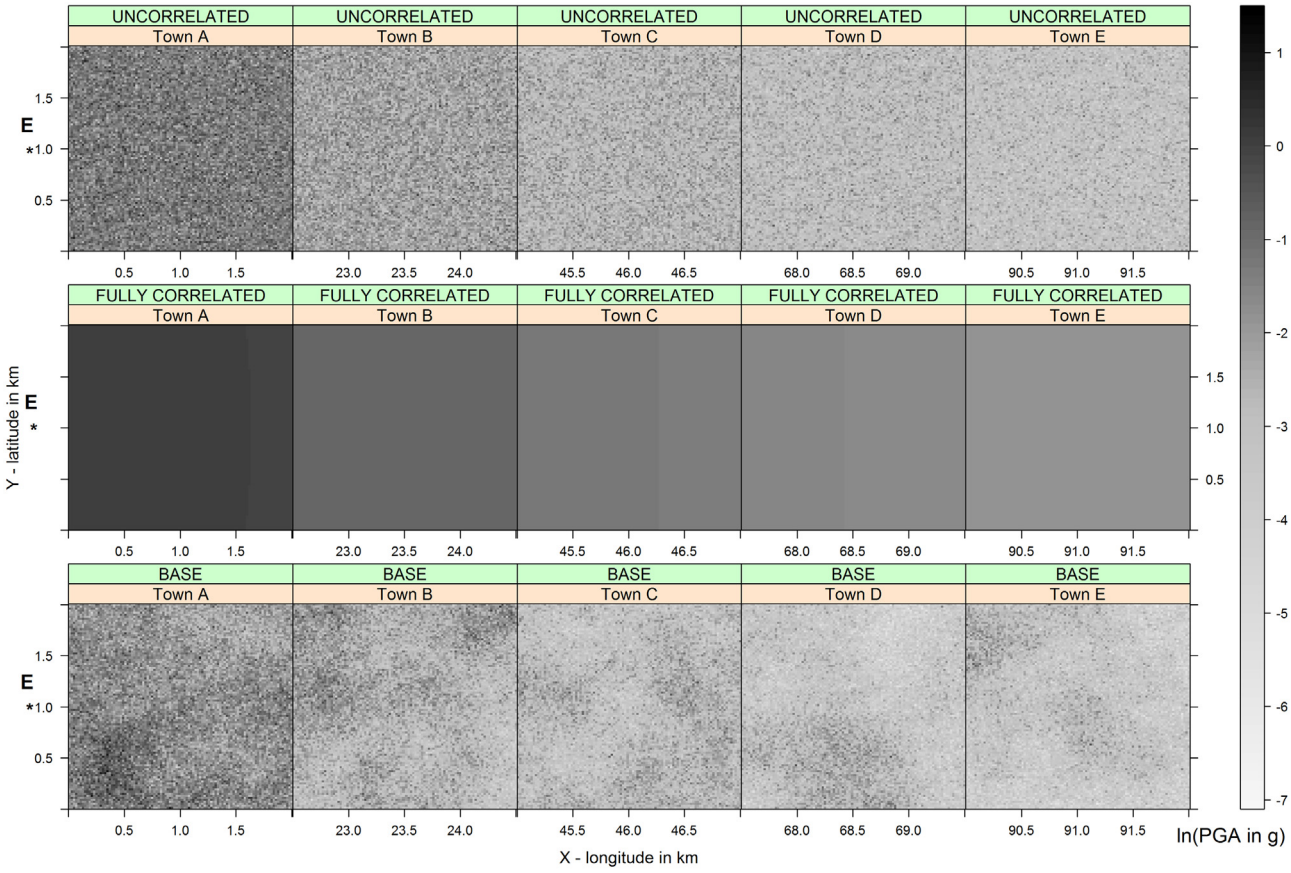

Fig. 1. Location of the epicentre (star), the five towns and realisations of PGA fields for 'UNCORRELATED' (top row), 'FULLY CORRELATED' (middle row), and 'BASE' (bottom row) scenarios. How the PGA (in g) fields shown here were generated is discussed later on in the paper.

$$\text{where } \mu_{jk} = P\left(DS \geq ds_i | iml_{true',k}\right) = \Phi\left(\frac{\ln\left(iml_{true',k}\right) - \lambda_k}{\zeta_k}\right) \quad (4)$$

where λ_k is the lognormal mean and ζ_k is the lognormal standard deviation for realisation k estimated by Eq. (4). Both of these parameters fully describe the empirical fragility curve for realisation k.

3. Results

A simulation study is undertaken to gain insight into the influence on empirical fragility curves of the measurement error in IMLs. Within this study, five hypothetical Turkish towns (termed A to E), distributed at equal intervals of 22.5 km as presented in Fig. 1, are considered. For simplicity, all ground motions are predicted assuming uniform stiff soil (e.g., $V_{s30}=400$ m/s) conditions throughout the region. Each town includes $N=10{,}000$ buildings, spatially distributed on the nodes of a grid with spacing of 20 m. To reduce the complexity of this study, masonry buildings are assumed with fragility expressed by the analytical fragility curves of Erberik [11], which adopt PGA as the intensity measure and use three damage states. Here, only the curve corresponding to moderate damage state is used for simplicity. The buildings in the five towns (i.e. 5 towns $\times$ 10,000 buildings $=$ 50,000 buildings in total) are assumed to be affected by a normal-faulting earthquake with known location and magnitude. The effect on the empirical fragility curves of the uncertainty in PGA estimates is examined by assuming the absence as well as the presence of strong-motion stations in the epicentral region. Table 1 lists the characteristics of the different parametric analyses conducted.

Common to all sensitivity analyses is the reference scenario, termed 'BASE', which assumes that the aforementioned 50,000 buildings are affected by an earthquake, with moment magnitude $M_w=7.2$, whose epicentre is located 10 km west of the centre of town A, as depicted in Fig. 1. In addition, the PGAs experienced by each building are estimated by the GMPE (termed GMPE1) derived by Akkar and Bommer [12], which models a large intra-event as well as inter-event variability. Following the procedure outlined in Section 2.1, this variability as well as the spatial correlation is taken into account by generating 1000 realisations of the PGA levels experienced by the 50,000 buildings. It should be noted that we assume $h_0=10$ km, which is a typical estimate of this parameter in recent studies [9,10], although it should be noted that this parameter appears to vary with the structural period, geographical location and earthquake and consequently there is much uncertainty over the appropriate value. For each realisation of the PGAs, the damage experienced by each building is generated according to the procedure outlined in Section 2.1. The construction of empirical curves requires the determination of the 'true' PGAs. For this scenario, the absence of strong-motion stations in the vicinity is considered. Therefore, the 'true' values are considered equal to the median GMPE1 values. The empirical fragility curves for moderate damage for each realisation, depicted in Fig. 2a, are substantially flatter than the fragility curves of Erberik [11], which express the actual fragility of the masonry buildings. This suggests that the ground-motion variability leads to a considerable increase in the uncertainty of empirical fragility curves.

The impact of the variability in PGA on the mean empirical fragility curve as well as its 90% confidence intervals is examined here by considering an alternative scenario, termed 'CHECK', which sets the variability in GMPE1 to zero. According to this

Table 1
Scenario simulations and characteristics of the inverted fragility curves.

Name	M_w	h_0 (km)	Grid (m^2)	Aggregation of buildings	N	PGA_{median}	$PGA_{realisation}$		PGA = 0.07 g			PGA = 0.20 g		
							σ	τ	95%	Mean	5%	95%	Mean	5%
Sensitivity to uncertainty in GMPE														
CHECK	7.2	10	20 × 20	No	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	9.0	9.2	9.5	67.1	67.7	68.3
BASE	7.2	10	20 × 20	No	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	11.0	23.0	38.0	37.6	60.7	80.6
Sensitivity to GMPE														
GMPE2	7.2	10	20 × 20	No	5 × 10,000	GMPE2	σ_{GMPE1}	τ_{GMPE1}	4.3	14.3	27.9	24.7	42.5	60.7
GMPE3	7.2	10	20 × 20	No	5 × 10,000	GMPE3	σ_{GMPE1}	τ_{GMPE1}	3.7	13.1	26.1	23.8	41.5	59.4
Sensitivity in h_0														
Uncorrelated	7.2	Uncorrelated	20 × 20	No	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	8.4	18.7	32.2	44.9	62.4	78.4
Fully correlated	7.2	Fully correlated	20 × 20	No	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	0.07	19.8	71.0	7.6	63.2	99.0
Sensitivity to buildings' density														
Density = 50 m × 50 m	7.2	10	50 × 50	No	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	12.0	23.2	36.5	42.7	60.2	76.5
Sensitivity to the number of buildings														
N = 5 × 100 (Ordered 1)	7.2	10	20 × 20	No (Ordered 1)	5 × 100	GMPE1	σ_{GMPE1}	τ_{GMPE1}	7.4	22.9	43.8	29.9	60.4	87.9
N = 5 × 100 (Ordered 2)	7.2	10	20 × 20	No (Ordered 2)	5 × 100	GMPE1	σ_{GMPE1}	τ_{GMPE1}	8.9	22.6	39.4	35.6	61.0	83.1
N = 5 × 100 (Random)	7.2	10	20 × 20	No (Random)	5 × 100	GMPE1	σ_{GMPE1}	τ_{GMPE1}	9.7	22.8	38.5	36.9	60.6	81.5
N = 5 × 1000 (Ordered 1)	7.2	10	20 × 20	No (Ordered 1)	5 × 1000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	9.1	22.8	40.1	34.5	60.5	84.1
Sensitivity to buildings aggregation														
Aggregated = 1 point per town	7.2	10	20 × 20	1 × 1 per town	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	1.7	17.7	44.1	18.8	63.4	97.7
Aggregated = 100 points per town	7.2	10	20 × 20	10 × 10 per town	5 × 10,000	GMPE1	σ_{GMPE1}	τ_{GMPE1}	2.1	18.0	40.8	22.5	63.6	96.5
Name	M_w	h_0 (km)	Grid (m^2)	Aggregation of buildings	N	PGA_{median}	$PGA_{realisation}$		PGA = 0.07 g			PGA = 0.20 g		
									95%	Mean	5%	95%	Mean	5%
Kriging														
5 × 1 st, COR.	7.2	10	20 × 20	10 × 10 p town	5 × 100	GMPE1	1 station p town		12.1	25.2	39.5	28.6	48.7	73.9
5 × 10 st, COR.	7.2	10	20 × 20	10 × 10 p town	5 × 100	GMPE1	10 stations p town		15.5	21.5	28.4	42.1	53.9	65.2
5 × 50 st, COR.	7.2	10	20 × 20	10 × 10 p town	5 × 100	GMPE1	50 stations p town		12.6	16.8	21.0	53.0	59.8	66.2
5 × 100 st, COR.	7.2	10	20 × 20	10 × 10 p town	5 × 100	GMPE1	100 stations p town		5.8	9.1	12.6	61.6	68.1	75.0
5 × 1 st, UNCOR.	7.2	Uncorrelated	20 × 20	10 × 10 p town	5 × 100	GMPE1	1 station p town		0.0	21.1	59.8	20.8	73.6	100.0
5 × 10 st, UNCOR.	7.2	Uncorrelated	20 × 20	10 × 10 p town	5 × 100	GMPE1	10 stations p town		7.1	20.5	31.3	57.5	73.3	89.0
5 × 50 st, UNCOR.	7.2	Uncorrelated	20 × 20	10 × 10 p town	5 × 100	GMPE1	50 stations p town		9.6	17.5	25.0	62.6	71.3	79.4
5 × 100 st, UNCOR.	7.2	Uncorrelated	20 × 20	10 × 10 p town	5 × 100	GMPE1	100 stations p town		5.8	9.1	12.6	61.7	67.9	74.6
5 × 1 st, FCOR.	7.2	Fully correlated	20 × 20	10 × 10 p town	5 × 100	GMPE1	1 station p town		11.0	25.0	40.6	27.3	48.9	73.7
5 × 10 st, FCOR.	7.2	Fully correlated	20 × 20	10 × 10 p town	5 × 100	GMPE1	10 stations p town		15.3	21.5	27.9	44.5	54.1	63.5
5 × 50 st, FCOR.	7.2	Fully correlated	20 × 20	10 × 10 p town	5 × 100	GMPE1	50 stations p town		12.6	16.7	21.2	52.7	59.8	66.8
5 × 100 st, FCOR.	7.2	Fully correlated	20 × 20	10 × 10 p town	5 × 100	GMPE1	100 stations p town		5.8	9.0	12.6	61.7	67.9	74.6

scenario, the 'true' PGAs are considered known and equal to their corresponding values obtained from the 1000 random fields. The procedure described above to construct empirical fragility curves is again used and the results obtained are presented in Fig. 2a. The mean fragility curve for this scenario is almost identical to the 'true' fragility curve, as expected. The very narrow width of the 90% confidence intervals can be attributed to the large sample size, i.e. 50,000 buildings, used to produce each curve.

The differences in the empirical fragility curves constructed by the two scenarios, i.e. 'BASE' and 'CHECK', are quantified in Table 1, which lists the mean and 90% confidence intervals of these curves for two example PGAs (0.07 g and 0.20 g). The flatter mean curve for 'BASE' leads to a $P(DS \geq$ moderate damage/PGA = 0.07 g) that is 150% higher than its 'CHECK' counterpart. The difference is reduced to 10% for 0.20 g. With regard to the width of the 90% confidence intervals, this appears to be 54 and 36 times wider for 'BASE' than 'CHECK' given PGA = 0.07 g and 0.20 g, respectively.

Overall, the variability in the PGAs leads to a substantial increase in uncertainty in the empirical fragility curves, which is manifested both in terms of a flatter curve as well as wider confidence intervals.

In the construction of empirical fragility curves for 'BASE', the 'true' IMLs have been considered equal to the median values of GMPE1. Nonetheless, the selection of a GMPE to express the 'true' PGA levels depends on the analyst. What is the impact of using median values of an alternative GMPE? To address this question, the GMPEs proposed by Cauzzi and Faccioli [13] and Zhao et al. [14], identified by Delavaud et al. [15] as suitable for Europe and the Middle East, are used to estimate the 'true' PGAs. The modified 'BASE' scenario is re-run and the estimated mean and 90% confidence intervals for the two scenarios, termed 'GMPE2' and 'GMPE3', respectively, are depicted in Fig. 2b. The 'true' intensity measure levels for 'GMPE2' and 'GMPE3' are higher than their 'BASE' counterparts as presented in Fig. 3. This leads to the mean

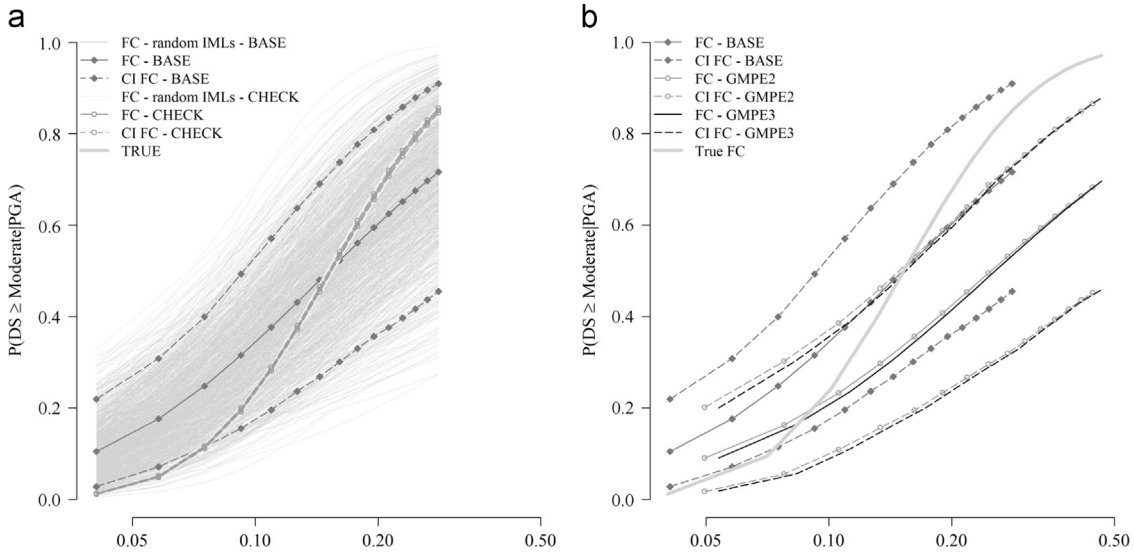

Fig. 2. Sensitivity of fragility curves (FC) and their 90% confidence intervals (CI FC) to (a) uncertainty in GMPE and (b) selection of GMPE.

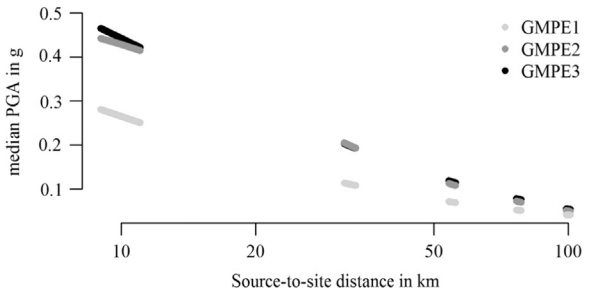

Fig. 3. Median PGAs (in g) against the source-to-site distance for GMPE1, GMPE2 and GMPE3.

fragility curves for the two scenarios to be shifted to the right indicating an improved seismic performance than their 'BASE' counterparts (see Fig. 2b). It should be noted, however, that improved performance does not make the curves closer to the 'true' fragility curve [11]. In particular, for PGA=0.07 g, the difference between the mean fragility curve for 'GMPE2' and the 'true' fragility curve is reduced to 55% and 37% for 'GMPE2' and 'GMPE3', respectively. By contrast, the width of the 90% confidence intervals is the same as for 'BASE'. This is expected given that the width is related to the ground-motion variability, which remains the same for all three scenarios (the standard deviations associated to GMPE2 and GMPE3 were assumed equal to that for GMPE1). The true total standard deviations of GMPE2 (0.344 in terms of $\log_{10}$ PGA) and GMPE3 (0.314) are similar to that of GMPE1 (0.279) and hence using these values instead would not have a large impact on the results.

Having established that the event's characteristics mainly affect the range of IMLs, the influence on the empirical fragility curves of the spatial correlation parameters, namely h_0 and the building density, are examined next.

The impact on the empirical fragility curves of the correlation introduced by the exponential model, used to model the Gaussian random fields in Section 2.1, is examined here by re-running the 'BASE' scenario assuming that the intra-event residuals in GMPE1 are uncorrelated or fully correlated. According to the former scenario, the intensity measure level experienced by each building does not depend on the intensity at adjacent buildings. Fig. 4a shows that this assumption leads to a steeper mean fragility curve, which appears to be closer to its 'true' counterpart, and notably smaller confidence intervals than 'BASE'. This agrees with the findings of Crowley et al. [9]. In particular for PGA=0.20 g, the confidence intervals around the empirical fragility curves are 22% narrower than for 'BASE'. By contrast, if the intra-event residuals are considered fully correlated, the confidence intervals round the fragility curve are significantly wider. In particular, they are 2.6 times wider than their 'BASE' counterparts (see Table 1).

The impact of the density of the buildings is studied by re-running the 'BASE' scenario assuming that, in each town, the 10,000 buildings are located on the nodes of a wider 50×50 m^2 grid. By repeating the procedure outlined above, the mean and 90% confidence intervals of the empirical fragility curves constructed from this scenario are plotted against their 'BASE' counterparts in Fig. 4b. A wider grid means that the buildings cover a larger area. This leads to an IML range 11% larger than for the 'BASE'. In addition, the confidence intervals appear to be narrower, suggesting that surveying buildings far away from each other reduces the impact of the spatial correlation. For example, for PGA=0.20 g the width of the confidence limits is 20% smaller than for 'BASE'.

The sensitivity analyses so far used the entire building inventories in the five examined towns. In reality, surveying 50,000 buildings would be a time-consuming and expensive task, which leads to the questions: can the accuracy of a large sample size be reached using fewer samples? And, if so, does the adopted sampling technique matter?

The impact of the sample size on empirical fragility curves is examined by considering two scenarios. The first scenario, termed '$N=5 \times 100$ (Ordered 1)', considers 100 buildings uniformly distributed on the 20×20 m^2 grid around the centre of each town, as presented in Fig. 5a. The second scenario, termed '$N=5 \times 1000$ (Ordered 1)', considers 1000 buildings uniformly distributed on the 20×20 m^2 grid around the centre of each town, as presented in Fig. 5b. For each building, the corresponding 1000 indicators generated for 'BASE' are assigned. For each realisation of indicators, empirical fragility curves are then constructed by fitting the probit model (Eq. (3)) to the 500 or 5000 indicators and their associated median PGA values obtained from GMPE1. Fig. 6 shows that the mean fragility curves for both scenarios are identical to their 'BASE' counterpart. However, the width of the confidence intervals varies considerably with the sample size. From Table 1, it can be inferred that the width of the 90% confidence intervals for '$N=5 \times 100$ (Ordered 1)' is 35% larger than for 'BASE'. This difference is reduced to 15% for '$N=5 \times 1000$ (Ordered 1)'.

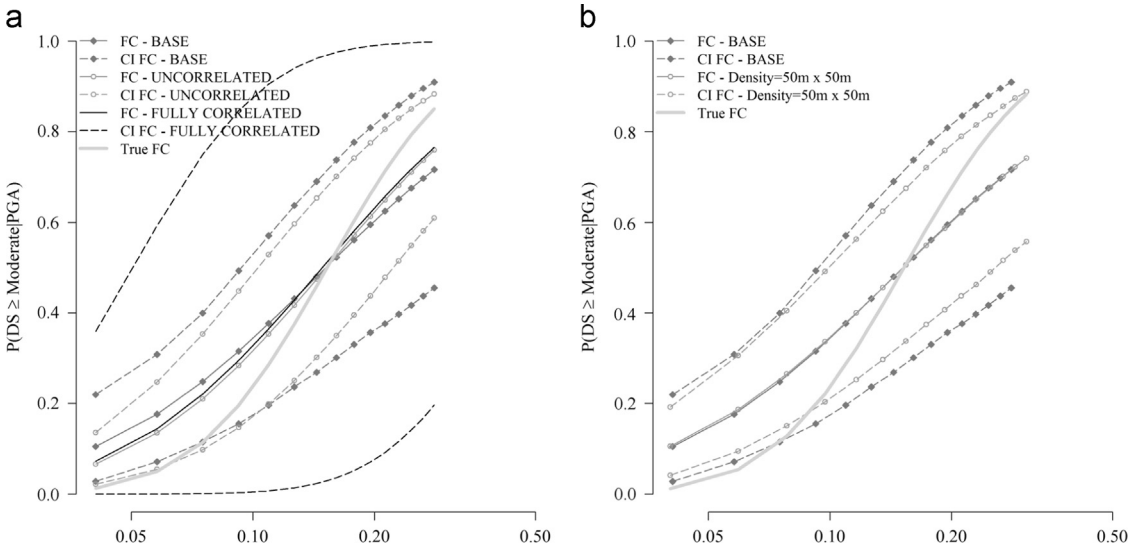

Fig. 4. Sensitivity of fragility curves (FC) and their 90% confidence intervals (CI FC) to (a) h_0 and (b) spatial distribution of buildings.

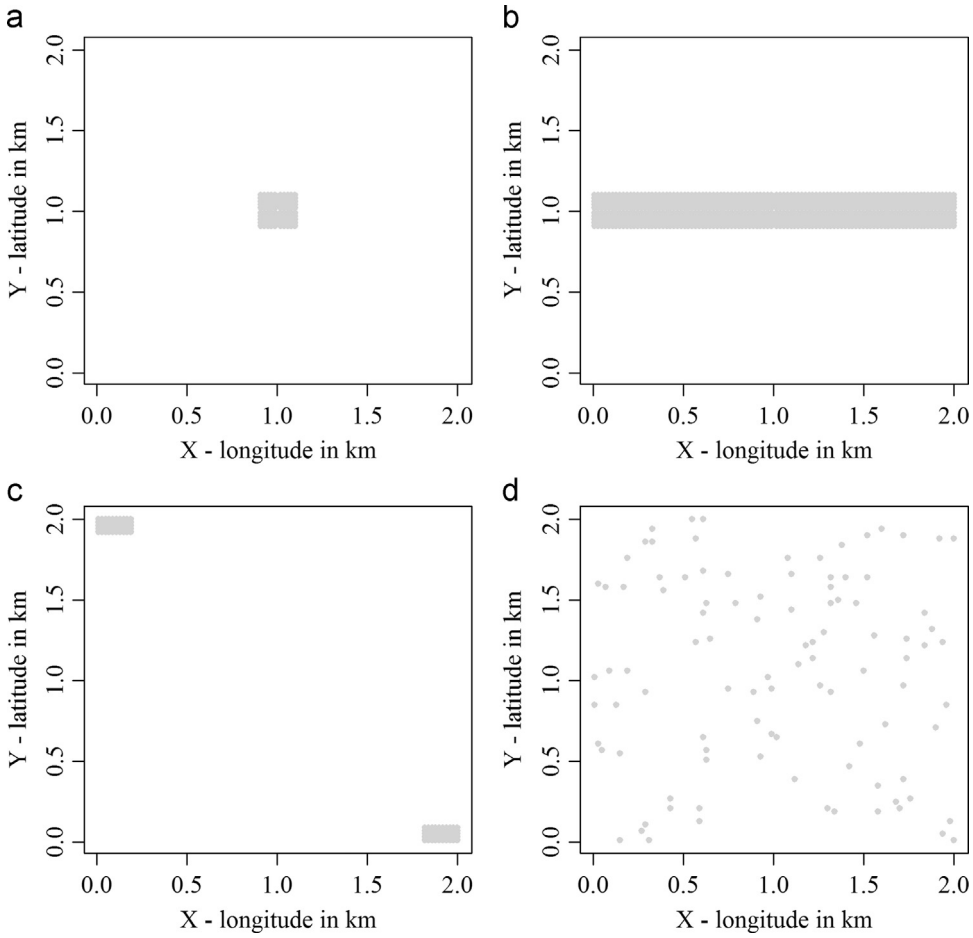

Fig. 5. Distribution of 100 and 1000 selected buildings in town A for the four considered scenarios. (a) N=5x100(Order 1), (b) N=5x1000(Order 1), (c) N=5x100(Order 2), and (d) N=5x100(Random).

The influence of the sampling technique adopted is examined next by examining whether the large uncertainty around the mean empirical fragility curve for scenario '$N=5\times100$ (Ordered 1)' is reduced by changing the sampling technique. This investigation is conducted by considering two further scenarios. The scenario, termed '$N=5\times100$ (Ordered 2)' assumes that 50 buildings uniformly distributed on the $20\times20\ m^2$ grid are obtained from two different areas of each town, as presented in Fig. 5c. The scenario termed '$N=5\times100$ (Random)' randomly selects 100 buildings from each town (see Fig. 5d). Empirical fragility curves are constructed using the procedure adopted for the two aforementioned scenarios. Fig. 6 shows that the mean fragility curves for both scenarios are also identical to 'BASE'. In this case, the width of the confidence intervals appears to be closer to 'BASE'. From Table 1, it can be

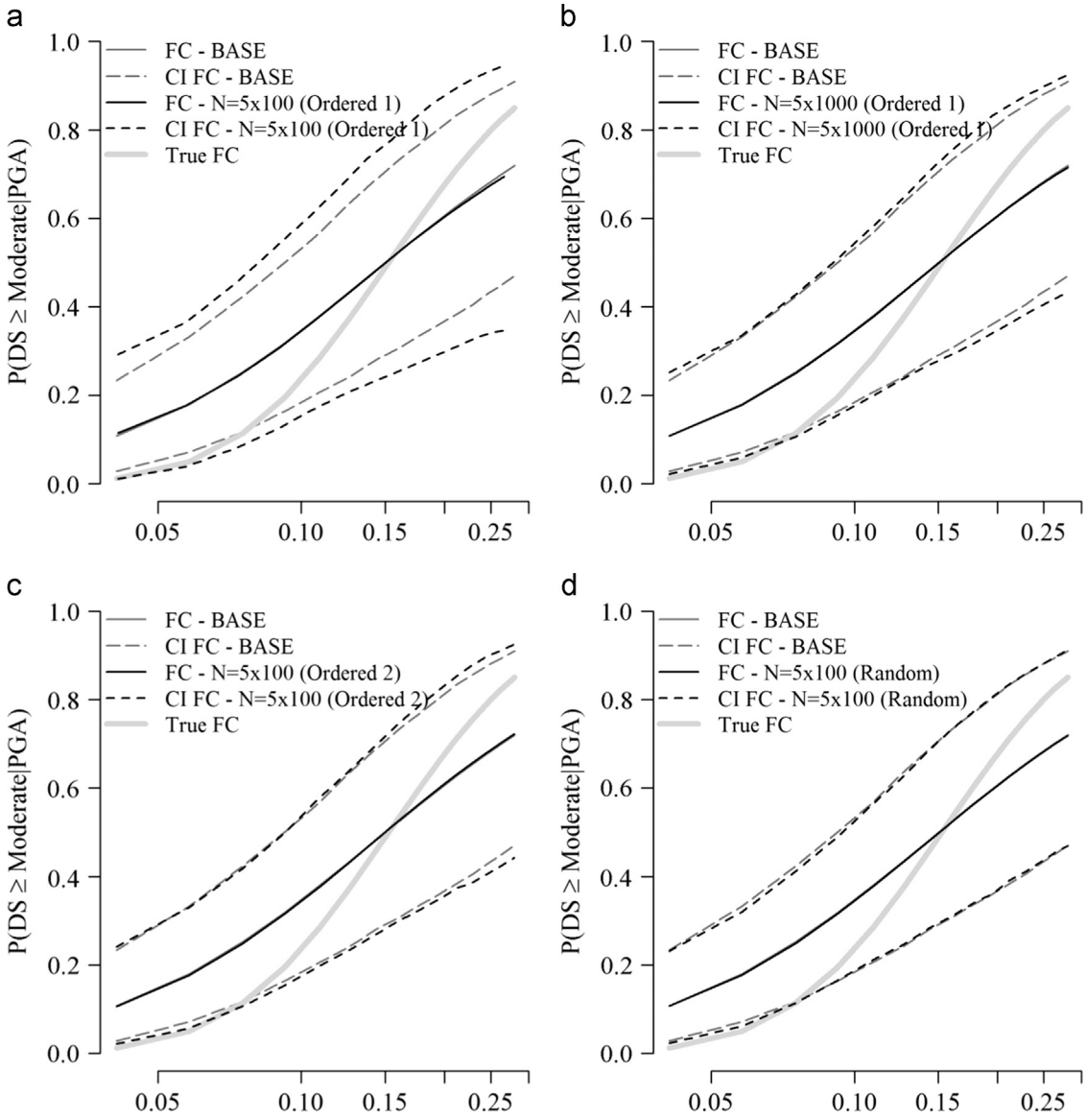

Fig. 6. Sensitivity of fragility curves (FC) and their 90% confidence intervals (CI FC) to the number of buildings and their location aggregation. ((a)-(d), PGA in g).

inferred that the width of the 90% confidence intervals for '$N=5\times100$ (Ordered 2)' and '$N=5\times100$ (Random)' for PGA=0.07 g is 13% and less than 10% the width of 'BASE', respectively. This suggests that the sampling technique adopted affects the results substantially. More importantly, relatively small carefully selected samples can yield results close to the curves obtained if every single affected building is surveyed.

Post-earthquake damage observations are often available in aggregated form. What is the effect of aggregation on the empirical fragility curves in the absence of strong-motion stations? This question is addressed by considering two scenarios with different levels of aggregation. The scenario termed 'aggregated=1 per town' aggregates the 10,000 buildings in each town to a single bin. The PGA for each of the five bins is estimated at the corresponding town's centre. This is a common assumption found in the literature. Similarly, the scenario termed 'aggregated=10×10 per town' aggregates the 10,000 buildings in each town to 100 equally sized sub-areas. The intensity measure level for each subarea is estimated at its centre. The mean and 90% confidence intervals of the empirical fragility curves for the two scenarios are presented in Fig. 7. The mean curve is identical for the two scenarios and they appear to be steeper than the 'BASE' mean curve, although the differences, especially at the lower end of the curves, remain significant. For example, for PGA=0.07 g the probability of exceedance for 'Aggregated=1 per town' is reduced, compared to its 'BASE' counterpart, to 92%. The confidence intervals appear to be significantly wider than for 'BASE'. Fig. 7b shows that the aggregated results are closer to the fully aggregated results, which is expected given that aggregation of buildings assumes that they are all subjected to the same ground motion. From Table 1, larger differences in the width of the confidence intervals can be observed for PGA=0.20 g. In particular, the 90% confidence intervals for 'Aggregated=1 per town' and 'Aggregated=10×10 per town' are 83% and 72% wider than for 'BASE'. This indicates that data aggregation leads to a significant loss of information, which appears to substantially increase the uncertainty in the empirical fragility curves. This observation seems to contradict the conclusions of Bal et al. [16] that the uncertainty introduced by the geographical resolution of observations of the mean repair-to-replacement ratio (MDR) is too small to justify a systematic effort to reduce it. It should be noted that Bal et al. [16] generated a large number of loss scenarios and they did not perform empirical fragility analysis. In

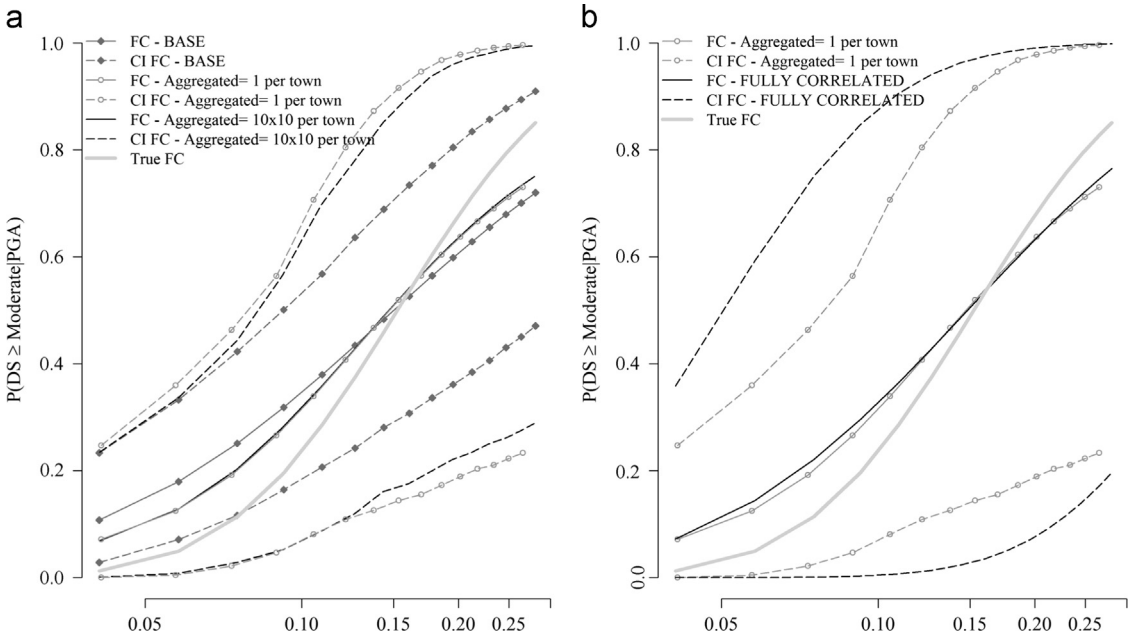

Fig. 7. Sensitivity of fragility curves (FC) and their 90% confidence intervals (CI FC) to the level of aggregation ((a), and (b), PGA in g).

particular, they generated random fields of ground motion intensity for a hypothetical town affected by a given earthquake and for each IML, the response of each building was simulated. A damage state was then assigned and finally the MDR was estimated and aggregated over the total number of buildings in the town. This aggregation might be the reason for the relatively small uncertainty around MDRs for each event. More research is needed in order to examine the effect of empirical fragility curves accounting for ground-motion uncertainty on the scenario loss.

So far, the considerable influence of ground-motion variability on empirical fragility curves in the absence of ground motion recording stations has been highlighted. Nonetheless, the presence of ground motion recording stations (accelerographs) is expected to reduce this uncertainty. Does this mean that the presence of a relatively small number of stations can lead to more accurate empirical curves?

In order to address this question, the influence of the presence of a varying number of recording stations in each town is examined following the procedure described in Section 2.2.2. It should be noted that this procedure is computationally intensive. In order to improve its efficiency, the sensitivity analyses are based on the modification of the scenario '$N=5\times100$ (Random)'. In particular, each town has 100 randomly-distributed buildings. The damage in each building is characterised by the 1000 indicators obtained for the '$N=5\times100$ (Random)' case. For this analysis, however, the 'true' PGAs account for the presence of 1, 10, 50 and 100 stations per town located in some of the 100 selected buildings following the procedure outlined in Section 2.2.2. The procedure is repeated assuming that the correlation model required for kriging varies. In particular, the case where the spatial correlation is ignored is considered by setting $h_0=0$ km and the case where the 'true' intensity experienced by the 100 buildings is considered fully correlated is also simulated by assuming $h_0=100{,}000$ km.

The mean and 90% confidence intervals of the empirical fragility curves obtained assuming that the PGA values have been recorded (i.e. '5 × 100 st, COR.', '5 × 100 st, UNCOR.' or '5 × 100 st, FCOR.') in all 100 buildings for the three correlation models are presented in Fig. 8 together with their counterparts obtained for 'BASE' and the 'true' fragility curve. As expected, the results for the former three scenarios are identical. Their mean fragility curves are identical to the 'true' fragility curve corresponding to the moderate damage state. It should be noted that the confidence intervals for the three considered scenarios appear to be wider than their 'CHECK' counterparts due to the smaller number of buildings used (i.e. 500 instead of 50,000).

On the other hand, the presence of one station per town produces empirical fragility curves, whose mean and confidence intervals depend on the correlation model adopted. With regard to the mean fragility curves, the curve for scenario '5 × 1 st, UNCOR' appears to be closer to the 'true' fragility curve than that for '5 × 1 st, FCOR', Nonetheless, the differences between the mean curves and the 'true' remain significant in the lower tails for all cases. For example, for PGA = 0.07 g, the probability of exceedance for '5 × 1 st, UNCOR' and '5 × 1 st, FCOR' is 129% and 174% higher than for its 'true' counterpart, respectively. The curve for scenario '5 × 1 st, COR' is included in the envelope formed by the two extreme correlation models and it is almost identical to its counterpart for the fully correlated scenario. The mean curve for 'BASE' is also included in this envelope. With regard to the confidence intervals, their width appears to be significantly wider for '5 × 1 st, UNCOR' and reduced for '5 × 1 st, FCOR'. In this case, the 90% confidence intervals for 'BASE' appear to be very close to their counterparts for '5 × 1 st, FCOR'. The aforementioned observations suggest that the presence of a very small number of stations, distributed in the affected area, does not improve the accuracy of the empirical fragility curves. To explore the reason behind this, the actual IMLs are plotted against their corresponding values estimated by kriging in Fig. 8. The closer the values to the 45° line, the better the prediction provided by a given number of stations in the area. From this figure, it can be seen that the presence of only one station per town leads to poor prediction of the actual IMLs and this affects the accuracy of the empirical fragility curves.

The presence of larger number of stations, i.e. '5 × 10 st, COR' and '5 × 50 st, COR', improves the prediction of the actual values (see Fig. 9) and this results in increasingly steeper mean fragility curves with increasingly narrower confidence intervals. In particular, for PGA = 0.07 g, the width of the confidence intervals is 52% and 69%, for the two aforementioned scenarios, smaller than its 'BASE' counterpart. This indicates that the presence of ground-motion measurements at 10% or 50% of the surveyed buildings greatly improves the confidence around the mean empirical

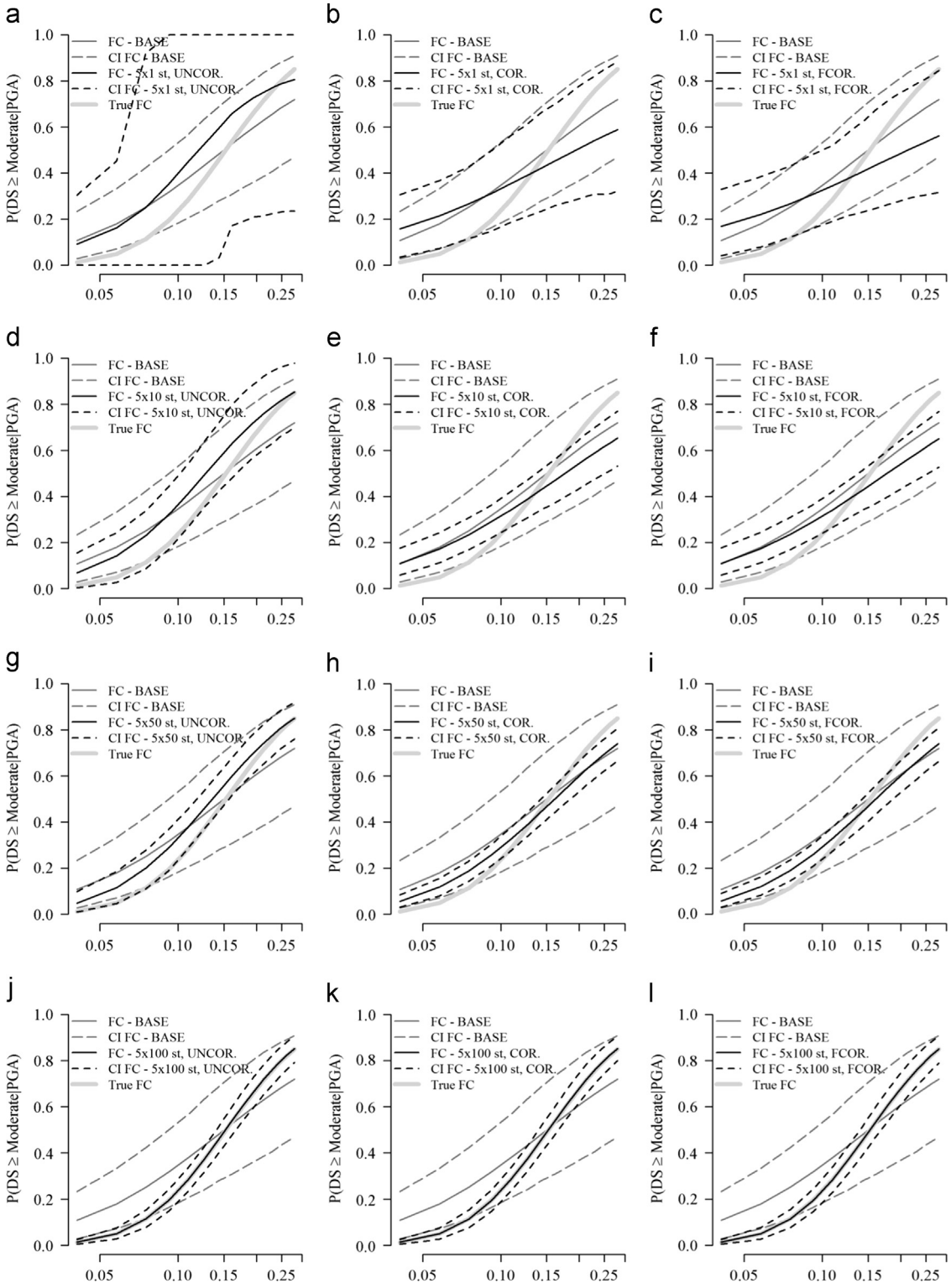

Fig. 8. Fragility curves (FC) and their 90% confidence intervals (CI FC) corresponding to the moderate damage state assuming the presence of a varying number of ground-motion stations assuming three different correlation models ((a)-(l), PGA in g).

fragility curve. However, despite a decrease from 134% to 83% for the two aforementioned scenarios, the difference between the mean fragility curves and the 'true' fragility curve remains high. Similar observations can be noted for the other two correlation models. This lack of significant improvement in the mean empirical fragility curves provided by the extrapolation of PGAs from stations located in half of the total number of buildings (which is equivalent to 25 stations per km^2) considered indicates that a very dense network of ground motion recording stations is required in order to construct reliable empirical fragility curves.

4. Conclusions

In this study, a series of experiments were conducted in order to assess the importance of modelling the variability and

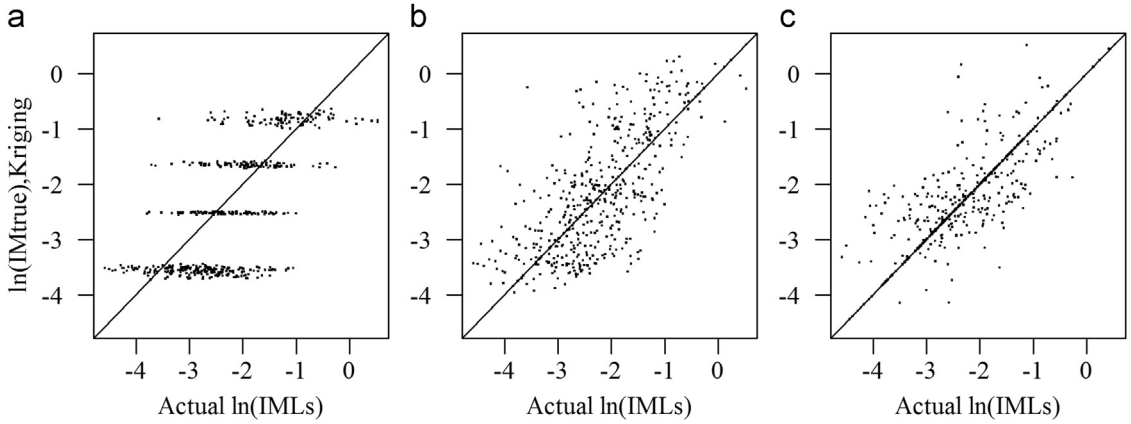

Fig. 9. Actual ln(IMLs), for the scenario '$N=5 \times 100$ (random)', against the ln(IMLs) estimated by kriging (scenarios '5×1, 10 and 50st, COR').

uncertainty in earthquake ground motions on empirical fragility curves. The following four main conclusions can be drawn from this study.

- The impact of variability in ground-motion levels is significant, and leads to considerably flatter fragility curves and wider confidence intervals. Further research is required in order to examine the impact of the uncertainty in the intensity measure levels in the seismic risk assessment's main products. These include, e.g., the estimation of the economic loss suffered by a group of buildings for a given event or the annual failure probability in a given location, estimated using the corresponding empirical fragility curves.
- There is a need for a very dense network of ground motion recording stations in order for the recorded data to reduce uncertainty in empirical fragility functions. This observation is in line with the main findings of Crowley et al. [9].
- The use of aggregated damage data, which is typical of existing empirical fragility curves, is found to increase substantially the uncertainty in the empirical fragility assessment. This raises questions about the accuracy of existing empirical fragility curves, but requires further research as only five towns and a single earthquake have been considered in this study.
- Finally, the sampling technique adopted in the collection of data can improve the accuracy of the empirical fragility curves. It was found that the use of a randomly selected, relatively small sample of buildings (e.g. 100 buildings per town) can potentially lead to improved fragility curves compared to using buildings from a single neighbourhood.

Overall, the findings of this study highlight, in line with similar appeals in the literature, the need for denser networks of ground motion recording stations in urban areas, without which the reliability of the empirical fragility curves is questionable. Low-cost Internet-enabled sensors, such as those in the Quake–Catcher Network [17], provide a possible cost-effective method of dramatically increasing the density of available ground-motion observations of damaging earthquakes. The results also highlight the important role that reconnaissance teams can play in the collection of small samples of high-resolution damage data for the construction of reliable empirical fragility curves.

Acknowledgements

The authors would like to acknowledge the French Embassy in the London (UK) for supporting the Collaborative Science and Technology Workshop on "Seismic Fragility of Urban Buildings and Infrastructure" that led to this work. John Douglas's contributions to this study were funded by internal BRGM Projects, whereas those of Ioanna Ioannou and Tiziana Rossetto were funded by the EPSRC Project "Challenging RISK" (EP/K022377/1). We thank two anonymous reviewers for their detailed comments on a previous version of this paper.

References

[1] Rossetto T, Ioannou I, Grant DN. Existing empirical fragility and vulnerability relationships: compendium and guide for selection. Pavia, Italy: GEM Foundation; 2013.
[2] Straub D, Kiureghian Der A. Improved seismic fragility modeling from empirical data. Struct Saf 2008;30:320–36.
[3] Rossetto T, Ioannou I, Grant DN. Guidelines for empirical vulnerability assessment. Pavia, Italy: GEM Foundation; 2014.
[4] YazcanU. Empirical vulnerability modeling considering geospatial ground motion variability. In: Proceedings of the 11th international conference on structural safety and reliability. New York, USA; 2013.
[5] Gehl, P, Douglas, J, Seyedi, D., Influence of the number of dynamic analyses on the accuracy of structural response estimates, Earthqu Spectra, 2014, http://dx.doi.org/10.1193/102912EQS320M.
[6] Douglas J. Seismic network design to detect felt ground motions from induced seismicity. Soil Dyn Earthqu Eng 2013;48:193–7.
[7] Strasser FO, Bommer JJ. Review: strong ground motions—have we seen the worst? Bull Seismol Soc Am 2009;99:2613–37.
[8] Ribeiro Jr. PJ, Diggle PJ. geoR: a package for geostatistical analysis. R-News 2001, 15–18;1.
[9] Crowley H, Stafford PJ, Bommer J. Can earthquake loss models be validated using field observations? J Earthq Eng 2008;12:1078–104.
[10] Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struct Dyn 2009;38:1687–708.
[11] Erberik MA. Generation of fragility curves for Turkish masonry buildings considering in-plane failure modes. Earthq Eng Struct Dyn 2008;37:387–405.
[12] Akkar S, Bommer JJ. Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean Region, and the Middle East. Seismol Res Lett 2010;81:195–206.
[13] Cauzzi C, Faccioli E. Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records. J Seismol 2008;12:453–75.
[14] Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, et al. Attenuation relations of strong ground motion in Japan using site classification based on predominant period. Bull Seismol Soc Am 2006;96:898–913.
[15] Devalaud E, Cotton F, Akkar S, Scherbaum F, Danciu L, Beauval C, et al. Towards a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. J Seismol 2012;16:451–73.
[16] Bal IE, Bommer JJ, Stafford PJ, Crowley H, Pinho R. The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. Earthq Spectra 2010;26:619–34.
[17] Cochran ES, Lawrence JF, Christensen C, Jakka R. The Quake–Catcher Network: citizen science expanding seismic horizons. Seismol Res Lett 2009;80:26–30.