

# BEA: An efficient Bayesian emulation-based approach for probabilistic seismic response

Stylianos Minas<sup>a,\*</sup>, Richard E. Chandler<sup>b</sup>, Tiziana Rossetto<sup>a</sup>

<sup>a</sup> Department of Civil Environmental & Geomatic Engineering, University College London, United Kingdom

<sup>b</sup> Department of Statistical Science, University College London, United Kingdom

## ABSTRACT

This paper presents an advanced Bayesian emulation-based approach (hereafter BEA) that allows a reduced number of analyses to be carried out to compute the probabilistic seismic response and fragility of buildings. The BEA, which is a version of kriging, uses a mean function as a first approximation of the expected Engineering Demand Parameter given Intensity Measure (EDP|IM) and then models the approximation errors as a Gaussian Process (GP). A main advantage of the BEA is its flexibility, as it does not impose a fixed mathematical form on the EDP|IM relationship (unlike other approaches such as the standard cloud method). In addition, BEA makes fewer assumptions than standard methods, and provides improved characterization of uncertainty. This paper first presents the BEA approach and then assesses its computational efficiency as compared to the standard cloud method. This is done through the creation of EDP|IM relationships and fragility functions using the outputs of nonlinear dynamic and nonlinear static analyses for two case-study buildings representing Pre- and Special-Code seismic vulnerability classes. The nonlinear dynamic and static analysis methods represent different levels of accuracy i.e. are of high and low fidelity, respectively. The BEA and standard cloud methods are compared in their ability to recreate three “pseudo-realities”, each represented by an artificially generated EDP|IM relationship derived from a large set of analysis runs. Several input configurations are tested, including, reduced sets of training inputs (analysis runs), training inputs of high and low fidelity, two sampling processes for these inputs (i.e. random and stratified sampling) and two different IM representations. The results demonstrate that BEA yields both an improved accuracy in terms of mean estimates, as well as smaller uncertainty bounds compared to the cloud method. The improved performance of the BEA is maintained for all “pseudo-realities” tested regardless of whether it is trained with high or low fidelity analysis data, with the improvement particularly pronounced in cases when the advanced IM  $I_{Np}$  is used. It is demonstrated that good accuracy can be achieved with BEA even with reduced samples, yielding a saving in 25% in number of analyses required to generate the EDP|IM relationship. Finally, the use of BEA drastically improves both the accuracy and efficiency of the resultant seismic fragility functions.

*Keywords:* probabilistic seismic demand; Bayesian emulation; kriging; fragility curves;

## 1. Introduction

An essential component of the Performance-based earthquake engineering (PBEE) framework is the fragility function [1]. Fragility curves are continuous relationships between the ground motion intensity and the probability that a specified asset will reach or exceed predefined damage states. In the case of buildings and analytical fragility functions, structural models are analyzed under earthquakes of increasing intensity and their response, in terms of Engineering Demand Parameters (EDPs), is recorded. Statistical methods are employed to characterize this response probabilistically as a function

---

\* Corresponding author

Email address: [s.minas@ucl.ac.uk](mailto:s.minas@ucl.ac.uk) (S. Minas)

of seismic hazard, and to build relationships relating Engineering Demand Parameters to Intensity Measures. The ground motion hazard that affects structural response is often represented by a scalar parameter, or low-dimensional vector of parameters, known as the intensity measure (IM). In this paper, IM is a scalar ground motion parameter that is considered to be representative of the earthquake damage potential with respect to the specific structure: using standard notation for conditional probability distributions, the relationship with EDP is denoted by  $EDP|IM$ . The fragility function is then determined by comparing the  $EDP|IM$  relationship to properly calibrated EDP thresholds associated with predefined damage states or performance levels. State-of-the-art fragility functions are derived either analytically, empirically or based on experts' opinion (also addressed in literature as judgment-based or heuristic functions). According to Erberik et al. [2], heuristic fragility curves lack credibility comparing to the other two types. On the other hand, several limitations are associated with empirical approaches, including the difficulty in properly characterizing the levels of seismic intensity and a non-subjective damage state allocation; but more importantly, the lack of sufficient seismic damage data in most areas of the world.

Several approaches exist for the analytical fragility assessment of buildings, offering the user the flexibility to choose from analysis methodologies, various structural parameters, characteristics of input ground motions etc. Large numbers of analysis runs are required in order to account for uncertainties in the earthquake ground motions, hence analytical approaches can be computationally very demanding, especially when nonlinear dynamic analyses are used. To reduce this computational time, past studies have often either made simplifications to the structural model, used a reduced number of ground motions or adopted a simplified structural analysis approach. The present study aims to contribute to the fragility literature by proposing an approach for both reducing the number of computations required to define a fragility curves and for better characterizing the uncertainty associated with the curves.

This paper introduces an advanced Bayesian emulation-based approach (henceforth BEA) which allows a reduced number of analyses to be carried out to compute the probabilistic seismic response and fragility of buildings. The BEA, which is a version of kriging, uses a mean function as a first approximation of the expected  $EDP|IM$  and then models the approximation errors as a Gaussian Process (GP). It builds upon the statistical emulation approach proposed by the authors in [3], but improves on it considerably by offering a full assessment of uncertainty, including uncertainties in model parameter estimation. The main advantage of BEA is that fewer analyses are needed to compute the probabilistic seismic response of buildings. In addition, BEA is more flexible as it does not impose a fixed mathematical form on the  $EDP|IM$  relationship unlike standard approaches for fragility estimation (e.g. the cloud method described below). It is highlighted that past risk models were only able to incorporate parametric expressions of  $EDP|IM$ , however, modern risk codes are able to accommodate non-parametric relationships.

The computational efficiency of BEA is assessed and compared to the standard cloud method. This is done through the creation of  $EDP|IM$  relationships and fragility functions using the outputs of nonlinear dynamic and nonlinear static analyses for two case-study buildings representing Pre- and Special-Code seismic vulnerability classes. Three different "pseudo-reality" scenarios are implemented to investigate the prediction capability of the tested methods under favorable or less favorable conditions. Each "pseudo-reality" is represented by an artificially generated  $EDP|IM$  relationship derived from a large set of analysis runs. A variety of input configurations are tested, including, reduced sets of training inputs (analysis runs), training inputs of different levels of accuracy (high and low fidelity), two sampling processes for these inputs (i.e. random and stratified sampling) and two different IM representations. The results demonstrate that BEA yields both an improved accuracy in terms of mean estimates, as well as smaller uncertainty bounds compared to the cloud method. The improved performance of the BEA is maintained for all "pseudo-realities" tested regardless of whether it is trained with high or low fidelity analysis data. Good results are obtained for standard IMs but strong improvement is particularly evident in cases when an advanced IM is used. It is demonstrated that good

accuracy can be achieved with BEA even with reduced samples, yielding a saving of 25% in number of analyses required to generate the EDP|IM relationship. Finally, the use of BEA drastically improves both the accuracy and efficiency of the resultant seismic fragility functions.

The paper starts with a brief description of the standard cloud method and then provides an overview of statistical emulation and BEA. Next, the procedure followed in the paper for the evaluation of BEA performance versus that of cloud is described, including how the pseudo-reality experiment is designed and sensitivity tests carried out on BEA input. The results are presented and discussed, and conclusions are made as to the efficiency of BEA for fragility curve derivation.

## 2. Cloud analysis

The cloud method is one of the most commonly used approaches for characterizing the relationship between EDP and IM [4–6]. Within this method, computer simulations are used to determine the seismic response of a structure to a series of ground motion (GM) time-histories, each of which has an IM value. The simulations are usually based on nonlinear dynamic analysis (e.g. nonlinear time-history analysis - NLTHA), with the building response expressed in terms of EDP. It is noted that the incapability of some nonlinear static procedures (NSPs) to capture record-to-record variability precludes their use within the cloud analysis framework. However, the recently proposed variant of the capacity spectrum method FRACAS [7], can be implemented in cloud analysis as it utilizes spectra from GM inputs to perform the structural assessment.

The resultant peak values of EDP for given IM levels form a scatter of points, the so-called “cloud”. Least-squares regression is then used to fit a simple model to the cloud of data points. Typically, the EDP is considered to vary as a power-law of the form  $aIM^b$  such that, after taking logarithms, the relationship can be expressed as in equation (1):

$$\ln(EDP) = \ln(a) + b \ln(IM) + e \quad (1)$$

where  $EDP$  is the conditional median of the demand given the  $IM$ ,  $a$ ,  $b$  are the parameters of the regression, and  $e$  is a zero mean random variable representing the variability of  $\ln(EDP)$  given the  $IM$ . However, some situations exhibit substantial heteroscedasticity (i.e. non-constant variance), which needs to be modelled explicitly within the fragility analysis framework [8]; for example by performing linear regressions locally in a region of IM values of interest. The use of logarithmic transformation indicates that the EDPs are assumed to be lognormally distributed conditional upon the values of the IMs. This is a common assumption that has been confirmed as reasonable in many past studies [9,10].

The cloud approach has the advantages of simplicity and rapidity over alternative fragility assessment methods. However, it also has some restrictions. Firstly, an assumption is made that the relationship between IM and EDP is represented by a linear model in the log space. This assumption may be valid for a short range of IM and EDP combinations but not for the entire cloud response space. Additionally, the cloud method requires a large number of earthquake records to be used as an input, and the accuracy of the approach is highly dependent on the record selection process followed [11].

## 3. Developing EDP versus IM relationships through Bayesian emulation

Complex mathematical models of real-world processes exist in all scientific and engineering areas. These mathematical models are usually translated into computer codes and may require significant computational resources. The mathematical functions and the related computer codes may be referred to as simulators [12]. A simulator, as a function  $f(\cdot)$ , takes vectors of inputs  $\mathbf{x}$  and generates unique outputs,  $y = f(\mathbf{x})$  for each given input. In the context of analytical fragility estimation, the input  $\mathbf{x}$  represents a ground motion sequence and the output  $f(\mathbf{x})$  represents the simulated EDP for that

sequence. A simulator will produce outputs that have an associated uncertainty, for example, due to uncertainties in the inputs and in the simulator construction itself [13]. In analytical fragility functions this uncertainty is due to differences in ground motions associated with any IM value, limitations in numerically reproducing physical phenomena of the structure response and simplified assessment. To assess the effect of these uncertainties, one option is to use uncertainty and sensitivity analysis methods requiring many simulator runs, as discussed in [14]. However, this is impractical for complex simulators that are computationally demanding to run [12]. To overcome this, an approximation  $\hat{f}(\mathbf{x})$  of the simulator  $f(\mathbf{x})$ , known as a statistical emulator, may be introduced to act as a surrogate. If  $\hat{f}(\mathbf{x})$  is a good proxy of the simulator, it can be used to carry out uncertainty and sensitivity analysis but with significantly less effort. Additionally,  $\hat{f}(\mathbf{x})$  can be also used to estimate the values of  $f(\cdot)$ , at input configurations where the simulator has not been run.

A small number of data points, obtained by running the simulator at carefully chosen configurations of the inputs, is required to train an emulator (and provide values of standard deviations to quantify the approximation error). Following common practice in this type of problem (e.g. [13]), the present study adopts a Bayesian approach with an emulator based on a Gaussian process (GP) model.

In the standard GP approach to the modelling of simulator outputs, the output function  $f(\mathbf{x})$  is regarded as a realized value of a random process such that the outputs at distinct values of  $\mathbf{x}$  jointly follow a normal (or Gaussian) distribution. In most applications,  $f(\mathbf{x})$  is a ‘smooth’ function in the sense that a small variation in the input  $\mathbf{x}$  will result in a small perturbation in the output  $f(\mathbf{x})$ . Smoothness of a Gaussian process is ensured by specifying an appropriate structure for the covariance between process values at distinct values of  $\mathbf{x}$ . For example, the use of a Gaussian covariance model (not to be confused with the ‘Gaussian’ in the GP, which relates to the distribution, rather than the covariance structure), results in smooth realizations that are infinitely differentiable [15].

As noted above, in the context of fragility estimation, the input  $\mathbf{x}$  represents a ground motion sequence. However, fragility estimates are rarely presented as functions of an entire ground motion sequence: rather, they are presented as a function of an intensity measure  $t = t(\mathbf{x})$ , say, that is used as a proxy for the complete sequence. In practice, the relationship between any such scalar proxy measure and the EDP is imperfect so that the simulator outputs, regarded as a function of  $t$  rather than  $\mathbf{x}$ , are no longer smooth. As an example of this, two different ground motion records characterized by the same value of  $t$  (i.e. having the same IM) will not in general generate identical outputs. Consequently, the standard emulation approach must be modified slightly in order for it to be applied in the context of fragility estimation.

The required modification is still to regard  $f(\mathbf{x})$  as the realized value of a random process, but now such that the conditional distributions of  $f(\mathbf{x})$  given  $t(\mathbf{x})$  themselves form a Gaussian process. Specifically, write  $f(\mathbf{x})$  as:

$$f(\mathbf{x}) = \mu(t) + \varepsilon(\mathbf{x}) \quad (2)$$

where  $\mu(t)$  represents the systematic variation of the output with the IM and  $\varepsilon(\mathbf{x})$  is a discrepancy term with  $E[\varepsilon(\mathbf{x})] = 0$ , representing uncertainty due to the fact that  $t$  does not fully capture the relevant information in  $\mathbf{x}$ . In the first instance it is convenient to assume that  $\varepsilon(\mathbf{x})$  is normally distributed with a constant variance,  $\tau^2$  (this assumption can be checked when applying the methodology). It is also assumed that  $\varepsilon(\mathbf{x})$  is independent of  $\mu(t)$ , and that  $\varepsilon(\mathbf{x}_1)$  is independent of  $\varepsilon(\mathbf{x}_2)$  when  $\mathbf{x}_1 \neq \mathbf{x}_2$ . Note that in equation (2), the term  $\varepsilon(\mathbf{x})$  is deliberately used instead of  $\varepsilon(t)$ . The reason for doing this is to allow two simulator outputs with different  $\mathbf{x}$  but same value of  $t$ , to be considered as separate, hence,

allowing for scatter about the overall mean curve in a plot of  $f(\mathbf{x})$  against  $t(\mathbf{x})$ . Equation (2) then immediately yields the conditional distribution of EDP given IM as normal, with expected value  $\mu(IM)$  and variance  $\tau^2$ .

To exploit GP emulation methodology, a final additional assumption is made, whereby  $\mu(t)$  varies smoothly with  $t$  and hence can itself be considered as a realization of a GP. Specifically, we denote the expected value of  $\mu(t)$  by  $m(t)$ , its variance by  $\sigma^2$  and we specify a correlation function (such as the Gaussian) that ensures smooth variation with  $t$ . Under the Gaussian correlation function, the covariance between  $\mu(t_1)$  and  $\mu(t_2)$  is:

$$\sigma^2 \exp\left[-\frac{(t_1 - t_2)^2}{2\varphi^2}\right] \quad (3)$$

where  $\varphi$  is a parameter controlling the rate of decay of the correlation between function values at increasingly separated values of  $t$ . As an illustrative example, the mean function  $m(t)$  could be specified as linear, i.e.  $m(t) = \beta_0 + \beta_1 t$ : this represents a first approximation to the function  $\mu(t)$  in Eq.(2), with the correlation structure allowing for smooth variation in the ‘‘approximation error’’  $\mu(t) - m(t) = z(t)$ , say.

Combining all of the above, in the case where the mean function  $m(t)$  is linear we find that values of  $f(\mathbf{x})$  at distinct values of  $\mathbf{x}$  are jointly normally distributed such that:

- The expected value of  $f(\mathbf{x})$  is:  $\beta_0 + \beta_1 t$
- The variance of  $f(\mathbf{x})$  is equal to:  $\text{Var}[\mu(t)] + \text{Var}[\varepsilon(\mathbf{x})] = \sigma^2 + \tau^2$
- For  $\mathbf{x}_1 \neq \mathbf{x}_2$ , the covariance between  $f(\mathbf{x}_1)$  and  $f(\mathbf{x}_2)$  is:

$$\begin{aligned} \text{Cov}[f(\mathbf{x}_1), f(\mathbf{x}_2)] &= \text{Cov}[\mu(t_1) + \varepsilon(\mathbf{x}_1), \mu(t_2) + \varepsilon(\mathbf{x}_2)] \\ &= \text{Cov}[\mu(t_1), \mu(t_2)] = \sigma^2 \exp\left[-\frac{(t_1 - t_2)^2}{2\varphi^2}\right] \end{aligned} \quad (4)$$

where  $t_1 = t(\mathbf{x}_1)$  and  $t_2 = t(\mathbf{x}_2)$ .

These expressions depend on the full input vector  $\mathbf{x}$  only through the IM  $t(\mathbf{x})$ . As a result, the simulator outputs can be analyzed as though they are functions of  $t$  alone and, specifically, subjected to a standard GP analysis to estimate the systematic variation  $\mu(t)$ . The only difference between this formulation and a standard GP emulation problem is that when  $t_1 = t_2$ , the covariance between the simulator outputs is  $\sigma^2 + \tau^2$  rather than just  $\sigma^2$ . The additional  $\tau^2$  term allows for additional variance associated with the discrepancy term  $\varepsilon(\mathbf{x})$  in equation (2).

Having formulated the problem in terms of Gaussian processes, the simulator can be run a few times at different values of  $\mathbf{x}$ , and the resulting outputs can be used to estimate the parameters in the GP model which, as set out above, are the regression coefficients  $\beta_0$  and  $\beta_1$ , the variances  $\sigma^2$  and  $\tau^2$ , and the correlation decay parameter  $\varphi$ . A simple way to do this is to use ordinary least-squares regression to estimate the regression coefficients and then to carry out a least-squares fit on the variogram of the

residuals to estimate the remaining parameters as is standard in geostatistics [16]; The resulted  $\beta_0$  and  $\beta_1$  estimates are then plugged into the relevant formulae in order to carry out the interpolation and assess its uncertainty. However, this approach fails to account properly for the uncertainty in the parameter estimates. From this perspective a Bayesian approach is preferable: among other things, for any IM value this provides samples from the posterior predictive distribution of the associated EDP. This predictive distribution automatically accounts for parameter uncertainty: percentiles of the distribution can be used to provide prediction intervals (we use 95% intervals below), and the posterior predictive probability of exceeding any threshold of interest (e.g. corresponding to a given damage state so as to obtain a fragility curve) can be calculated directly. For an introduction to the calculation of predictive distributions in Bayesian analyses, see [17]. In our work, the Bayesian approach is implemented using the `krige.bayes()` routine from the `geoR` package in R [18].

Any Bayesian approach requires the specification of prior distributions for each parameter estimated: following the implementation in `krige.bayes()`, we adopt standard non-informative priors where possible for all parameters in order to ensure that the results reflect the information content of the simulator runs to the maximum extent possible. The model parameters and their prior distribution used in our BEA model are listed below:

- Regression coefficients,  $\beta_0$  and  $\beta_1$  (intercept and slope). Here, non-informative “flat” prior distributions are chosen.
- Partial sill,  $\sigma^2$ . A prior distribution proportional to  $1/\sigma^2$  (“reciprocal”) is used here.
- Range,  $\phi$ . An improper “uniform” prior distribution over the range  $(0, \infty)$  is selected.

It is noteworthy that the `krige.bayes()` function does not allow a Bayesian treatment of the variance  $\tau^2$  so this is fixed at the value obtained from a conventional kriging analysis as defined above. The effect will be that uncertainty in the predictions will still be underestimated, albeit much less than in most standard analyses. The results of the pseudo-reality experiment reported below will help to determine whether this is likely to be a problem in practice.

Having estimated the parameters, the results can be used to interpolate optimally between the existing simulator runs so as to construct an estimate of the predicted EDP|IM curve  $\mu(t)$  and associated standard deviation. The optimal interpolation relies on standard results for calculating conditional distributions in a Gaussian framework. The approach is used extensively in geostatistics to interpolate between observations made at different spatial locations, where it is often referred to as “kriging” [19]. In geostatistics, the term  $\tau^2$  is referred to as a ‘nugget’ and accounts for local-scale variation or measurement error. Nonetheless, in the present context the above argument shows that it can be derived purely by considering the structure of the probabilistic seismic demand problem.

It is noteworthy to mention that similarly to the cloud method, the BEA analysis is also carried in terms of  $\ln(EDP)$ , due to the normality and heteroscedasticity assumptions, which are better justified on the log scale.

#### 4. Methodology

The adequacy and the performance of the proposed Bayesian approach (BEA) for seismic response is assessed through its comparison with the cloud based approach. A roadmap of the methodology followed herein is shown in Figure 1. As a first step, two case study buildings, representing distinct vulnerability classes, are analyzed at two levels of analysis fidelity (i.e. using nonlinear dynamic and static analysis methods, representing high and low fidelity respectively), in order to construct the associated EDP|IM relationships (steps 1-3, Section 4.1). Each of the resultant data sets of IM and EDP is used to generate three “pseudo-realities” based on different assumptions, which are each

characterized by a mean and standard deviation value. The reason for implementing three different “pseudo-reality” scenarios is to investigate the prediction capability of the tested methods under favorable or less favorable conditions. A large number of data sets of IM and EDP are then simulated, using the same mean and standard deviation values as calculated for each “pseudo-reality” (step 4, discussed in Section 4.2). Different sampling configurations are employed, including two sampling types and three sampling sizes to construct the data sets (step 5, Section 4.3), which are used to train the tested methods (step 6). The effect of selecting different covariance structures within BEA is also investigated. Finally, the results of the tested methods are compared and evaluated against the “pseudo-realities” with the aim of three different metrics (step 7, Section 5). A computer code implementing the proposed framework is scripted in R[20]. A code supplement to this study can be accessed online through EPICentre website (<http://www.ucl.ac.uk/epicentre/resources/software>). Note that although the scripted BEA procedure is more complex than the standard method, it does not practically increase the computational expense.

This methodology process aims to shed light on the sensitivity of the BEA to several input assumptions, and sampling configurations, in the estimation of the mean of the EDP|IM distribution and the uncertainty characterization. The comparisons presented here are carried out in terms of EDP|IM predictions, however, a single case-study example is presented in Section 5.7 where the two methods are compared at the fragility level.

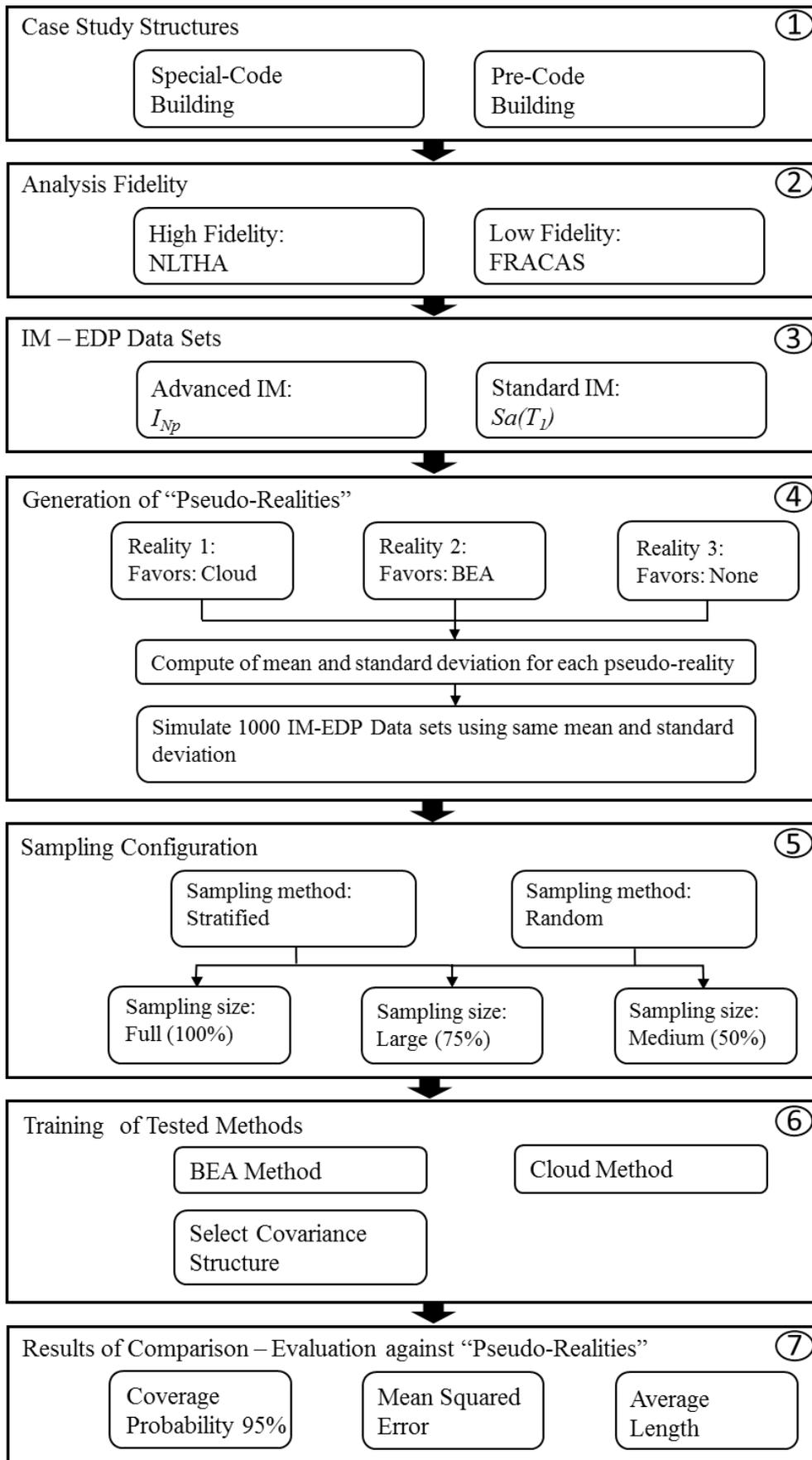


Figure 1 - Schematics of the methodology roadmap.

#### 4.1 Case study structures and analysis setup

Two regular reinforced concrete (RC) moment resisting frames (MRF) are modelled and implemented as case studies for the current research work. These structures, which are regular in plan and elevation and share the same geometry, represent distinct vulnerability classes, as they are characterized by different material properties and reinforcement detailing. The first frame is designed to only sustain gravity loads following the Italian Royal Decree n. 2239 of 1939 [21] that regulated the design of RC buildings in Italy up to 1971, hereafter called the Pre-Code building. The Pre-Code building's first-mode period is  $T_1 = 0.9s$  with a modal mass participation factor of 95.4%. The second frame is designed according to the latest Italian seismic code (or NIBC08; [22]), fully consistent with Eurocode 8 (EC8; [23]), following the High Ductility Class (DCH) rules, hereafter called the Special-Code building. The Special-Code building's first-mode period is  $T_1 = 0.5s$  with a modal mass participation factor of 92.8%. Both building models have a total mass of 173tn. Interstorey heights, the spans of each bay and cross-sections dimensions for each case-study building are reported in Figure 2. Details regarding the design and the modelling of the buildings are available in [24] and [25].

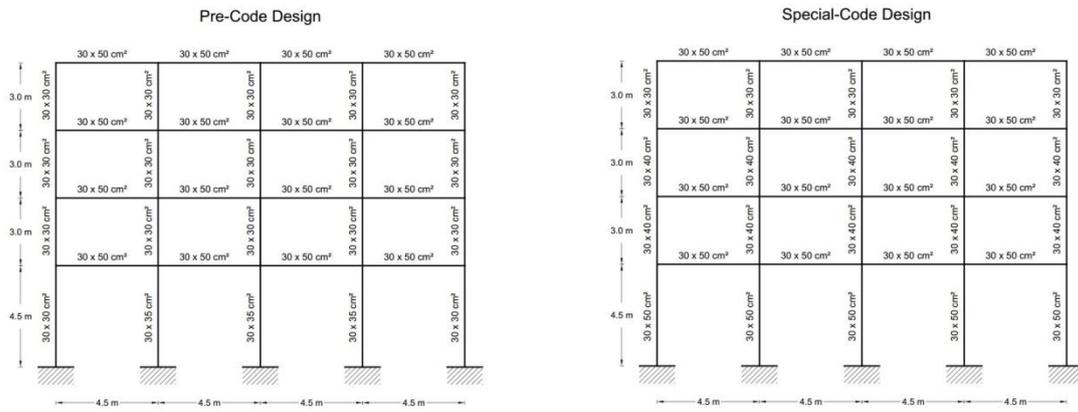


Figure 2 - Elevation dimensions and member cross-sections of the Pre-Code (left) and Special-Code (right) RC frames.

NSP and NLTHA are utilized to estimate the seismic response of the studied frames, representing two different levels of fidelity analysis, namely low fidelity and high fidelity respectively. Unscaled ground motion records from the SIMBAD database (Selected Input Motions for displacement-Based Assessment and Design; [26]), are used here as input to the nonlinear time history of the buildings (high fidelity). SIMBAD includes a total of 467 tri-axial accelerograms, consisting of two horizontal (X-Y) and one vertical (Z) components, generated by 130 worldwide seismic events (including main shocks and aftershocks). In particular, the database includes shallow crustal earthquakes with moment magnitudes ( $M_w$ ) ranging from 5 to 7.3 and epicentral distances  $R \leq 35$  km. A subset of 150 records is considered here to provide a significant number of strong-motion records of engineering relevance for the applications presented in this paper. These records are selected by first ranking the 467 records in terms of their PGA values (by using the geometric mean of the two horizontal components) and then keeping the component with the largest PGA value (for the 150 stations with highest mean PGA). Regarding low fidelity analysis, the simplified approach FRAGility through Capacity spectrum ASsessment (FRACAS) is used [7]. FRACAS is a variant of the capacity spectrum method, which uses ground motion time histories and pushover (PO) curves to compute the elastic/inelastic spectra and capacity curves respectively, in order to estimate the performance points [27]. Here, the same 150 records used for the NLTHA are used for the assessment in FRACAS. Static pushover (PO) analysis is carried out by applying increments of lateral loads to the side nodes of the structure. These lateral loads are proportionally distributed with respect to the interstorey heights (triangular distribution). The PO analysis is conducted until a predefined target displacement is reached, corresponding to the expected collapse state.

Figure 3 (left panel) shows the static PO curves associated with the two case-study buildings. The curves are reported in terms of top center-of-mass displacement divided by the total height of the structure (i.e., the roof drift ratio, RDR) along the horizontal axis of the diagram, and base shear divided by the building's seismic weight along the vertical axis (i.e., base shear coefficient). Figure 3 (right panel) shows the performance points of the two case-study buildings within an acceleration-displacement response spectrum (ADRS) space computed by FRACAS using the GMs records described above, and assuming elastic-perfectly plastic response of the structure.

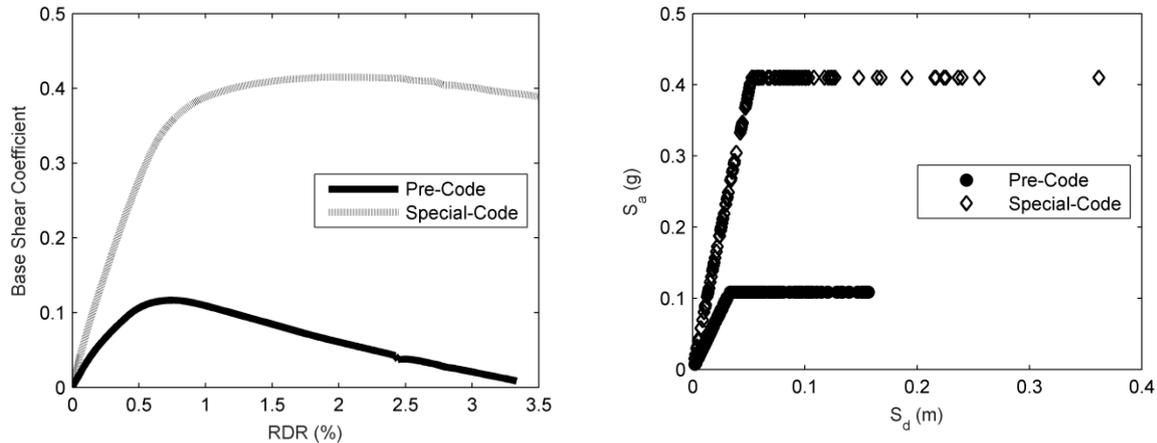


Figure 3 - Static PO curves for the Pre-Code and Special-Code buildings (left panel), and performance points generated by FRACAS using Elastic Perfectly Plastic (EPP) idealization model in ADRS space (right panel).

The results obtained from the various analyses are then expressed in data sets of IM and EDP. In each data set, the ground motion records that did not manage to push the structures to the nonlinear range are discarded. This is because FRACAS, as a capacity spectrum based method, draws a direct correlation between the seismic intensity and the estimated demand parameters within the elastic regime; therefore, bias is introduced [25]. For consistency, the same records are also discarded from the NLTHA analysis. After this amendment, the actual number of GM that pushed the frame into the nonlinear range is relatively small but still significant, corresponding to 15% and 23% of the total number of records used, for the case of Special- and Pre-Code buildings respectively (i.e. 68 and 108 GMs). Table 1 summarizes the number of ground motion records used for analysis of the case-study buildings based on three sample sizes, namely full-, large- and medium-sample.

Table 1 - Number of ground motions used for analysis of the case-study buildings based on sample size.

Building Type	Ground motion sample size		
	Full sample (100%)	Large sample (75%)	Medium sample (50%)
Pre-Code	108	81	54
Special-Code	68	51	34

In this paper, the deformation-based EDP maximum (over all storeys) peak interstorey drift ratio (denoted as MIDR) is adopted. With regard to IM inputs, two scalar IMs are used herein, namely the spectral ordinate  $S_a(T_1)$  and the advanced IM  $I_{N_p}$  [28]. The selection of the adopted IMs was based on the study conducted by [25], on the selection of optimal IMs for the fragility assessment of mid-rise RC buildings.  $I_{N_p}$  is based on the spectral ordinate  $S_a(T_1)$  and the parameter  $N_p$ , and is defined as:

$$I_{N_p} = S_a(T_1)N_p^\alpha \quad (5)$$

where  $\alpha$  parameter is assumed to be  $\alpha = 0.4$  based on the tests conducted by the authors and  $N_p$  is defined as:

$$N_p = \frac{S_{a,avg}(T_1, \dots, T_N)}{S_a(T_1)} = \frac{\left[ \prod_i^N S_a(T_i) \right]^{1/N}}{S_a(T_1)} \quad (6)$$

$T_N$  corresponds to the maximum period of interest and lies within a range of 2 and  $2.5T_1$ , as suggested by [25].

#### 4.2 Defining pseudo-realities

Three different scenarios are considered herein, each corresponding to a different “reality” represented by an artificially generated EDP|IM relationship, constructed as described below. This pseudo-reality approach is used to provide an experimental setup that mimics real applications as closely as possible. In each relationship, the expected value of some function of the EDP is given by a mean function  $\mu(t)$  as in equation (2); where  $\mu(t)$  is the sum of the deterministic mean function  $m(t)$  and an approximation error  $z(t)$ :

$$\mu(t) = m(t) + z(t) \quad (7)$$

The deterministic mean function  $m(t)$  is represented by a simple regression model, which is trained using the sets of EDP|IM data generated as described in Section 4.1. Standard cloud analysis corresponds to a situation in which  $\mu(t)$  is the expected value of  $\ln(EDP)$ , the mean function  $m(t)$  is a linear function of  $\ln(t)$  and the approximation error  $z(t)$  is zero. However, the framework considered here allows exploration of the capability of the tested methods to predict EDP|IM relationship under favorable or less favorable situations.

Our three chosen “pseudo-realities” are as follows:

- a) Reality 1: the expected value of  $\ln(EDP)$  is an exact linear function of  $\ln(IM)$  with no approximation error (i.e. with  $z(t) = 0$  in equation (7)). This “pseudo-reality” is exactly the situation for which the cloud method is designed. The coefficients in the linear function are derived from a sample of real analysis data, as described in Section 4.1: this ensures that as a test case, this represents the kind of scenario that might conceivably be encountered in a real earthquake engineering situation.
- b) Reality 2: this is as Reality 1 except that a non-zero approximation error  $z(t)$  is included in equation (7) so that  $\mu(t)$  is no longer an exact linear function of  $\ln(t)$ . The approximation error is generated as a Gaussian random field (but making sure that the resultant “true” function is monotonically increasing) – again, derived from the real analyses of Section 4.1. The choice of this “reality” favors the BEA methodology.
- c) Reality 3: the function  $\mu(t)$  is constructed in the same way as for Reality 2, but now it represents the expected value of  $EDP$  rather than  $\ln(EDP)$ . This case can be used as an example where both models (cloud and BEA) are wrong, and will shed light on the capability of these two models to estimate the EDP|IM relationship.

As an example, the three pseudo-realities generated for the Special-Code building analyzed at high fidelity is illustrated in Figure 4.

For each “pseudo-reality”, an artificial analysis dataset can be simulated by adding random “residuals” to the fitted mean curve, at the given IM locations of the original analysis set. In each simulation experiment reported below, 1000 such datasets were generated. A subset of the generated EDPs is utilized to calibrate both the cloud and BEA methods. These approaches are then assessed for their ability to predict the target “pseudo-reality”. The predictive performance of the tested methods is not only assessed in terms of estimation of the mean “pseudo-reality”, but also in terms of evaluating the uncertainty of the entire distribution, as both parameters are required for the fragility analysis. As well as predicting the mean “pseudo-reality” function itself, 95% prediction intervals are computed as a way of quantifying the uncertainty in the predictions.

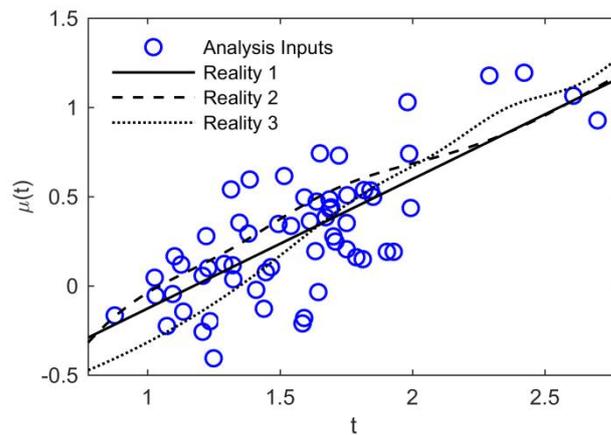


Figure 4 - Three artificially generated “pseudo-realities” from the real high-fidelity analyses (analysis inputs) of the Special-Code case study.

The coverage probability of the 95% prediction intervals (hereafter called coverage), the average length of these intervals and the mean squared error (MSE) are the metrics used to assess the predictive performance of the BEA and cloud approaches. The coverage measures the accuracy of the uncertainty assessment, by calculating the proportion of intervals containing the actual EDP. If the uncertainty assessment is accurate, the coverages from a method should be equal to their nominal value of 0.95. The average interval length assesses the amount of uncertainty in the prediction intervals and, for any collection of  $n$  prediction intervals, is computed as:

$$\text{Average Length} = \sum_{i=1}^n \left( \frac{UL95_i - LL95_i}{n} \right) \quad (8)$$

where  $UL95_i$  and  $LL95_i$  are respectively the upper and lower limits of the  $i$ th interval. Ideally, prediction intervals should be as small as possible subject to the correct coverage.

The MSE assesses empirically the quality of the vector of  $n$  predictions,  $\hat{Y}$ , with respect to a vector of simulated EDPs,  $Y$ , and is defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (9)$$

#### 4.3 Sampling configuration – Training of the BEA and the cloud method

The predictive ability of the BEA and cloud approaches are checked when different sampling configurations are adopted for the selection of EDP|IM training data sets.

Two sampling configurations are considered here, one focusing on the sample size and one on the sampling method. With regard to the sample size, three different cases are investigated, one using the full training input (N and M samples respectively for the Pre-Code and Special Code buildings, as described above), and two utilizing large and medium training samples, which correspond to 75% and 50% of the full samples respectively. The use of smaller sample sizes results in a reduction of the computational time. This is particularly true for high fidelity analyses, which each have a runtime of between a few minutes to two hours on an average personal computer, depending on the complexity of the structural model and the duration of the GM.

Two sampling methods are considered in determining the reduced training data sets from the full sample: random sampling, and stratified sampling. For the stratified sampling, the full input sample is divided into 5 strata (bins) of equal intensity measure width, and then random sampling is applied within each stratum to select the user-specified number of inputs.

Following the sampling procedure, the selected inputs are used to train the cloud and BEA. In the case of BEA, as stated in Section 3, the setup of the emulation model allows one to implement a regression model to describe the mean function  $\mu(t)$  and is suitable for the nature of the problem of interest. In this study, a power-law regression model is adopted to enable a comparison with the cloud method. However, alternative models, such as linear or higher order polynomials may also be used.

The results of this exercise shed light on the capability of the BEA to compute the probabilistic seismic response of buildings using a reduced number of analyses. This is a first step towards the development of an optimal experimental design for BEA, which will be a subject of future work.

## 5. Results of comparison – Evaluation against pseudo-realities

All the steps discussed in the Section 3 are applied here in order to build the BEA. BEA accounts for the uncertainty of the model's parameters, and provides estimates, and the associated variances, for the conditional distribution of the EDP|IM relationship. The outcomes of the BEA approach are here compared to the cloud method outcomes. A series of tests, which are described in detail in Section 4, are carried out to facilitate the evaluation process, highlighting the sensitivity of the BEA to several input assumptions and revealing the range of applicability of the approach proposed herein. In addition to these tests, the sensitivity of BEA predictions to the choice of the covariance structure is explored. Last, a section on fragility analysis using BEA is presented.

### 5.1 Performance assessment for the 2 case study buildings analyzed using NLTHA

In the first test-case investigated, the high fidelity analysis data sets, expressed in terms of MIDR and  $I_{Np}$ , are used to train both the BEA and the cloud method models. Following the procedure described in the methodology section, the two data sets, corresponding to Pre- and Special-Code buildings respectively, are sampled to three sample subsets following the stratified sampling process. Based on these test samples, the BEA and the cloud processes are used to predict the “three assumed realities”. The performance of the emulation and cloud approach is then assessed using the coverage probability, MSE and average interval length metrics. A detailed summary of results for this test-case is presented in Table A.1 in the appendix. The first metric to check is the coverage, which in both the cloud and the BEA case closely matches the nominal coverage probability of 95% (never more than +1.1% difference). Having this condition satisfied, next step is to assess the performance of the two tested approaches in terms of MSE and average length, as shown in Figure 5.

The MSE estimates for the BEA are smaller than those for the Cloud approach for “pseudo-realities 2 and 3”, whilst they are comparable for “pseudo-reality 1”. This observation highlights the superiority of the BEA method over the Cloud method in predicting the mean EDP|IM function, under different “pseudo-reality” scenarios. The same observation also applies to the average length estimates, although the improvement in the BEA's performance is not as significant as in the case of the MSE metric.

With regard to the training subsets, it is observed that the quality of predictions of the mean depends, as expected, on the size of the training subsets. In particular, the performance of the BEA and the cloud models becomes less accurate as the training subset size reduces. However, the same observations discussed above are also present in all instances of reduced training subsets, i.e. the BEA shows improved performance with respect to cloud.

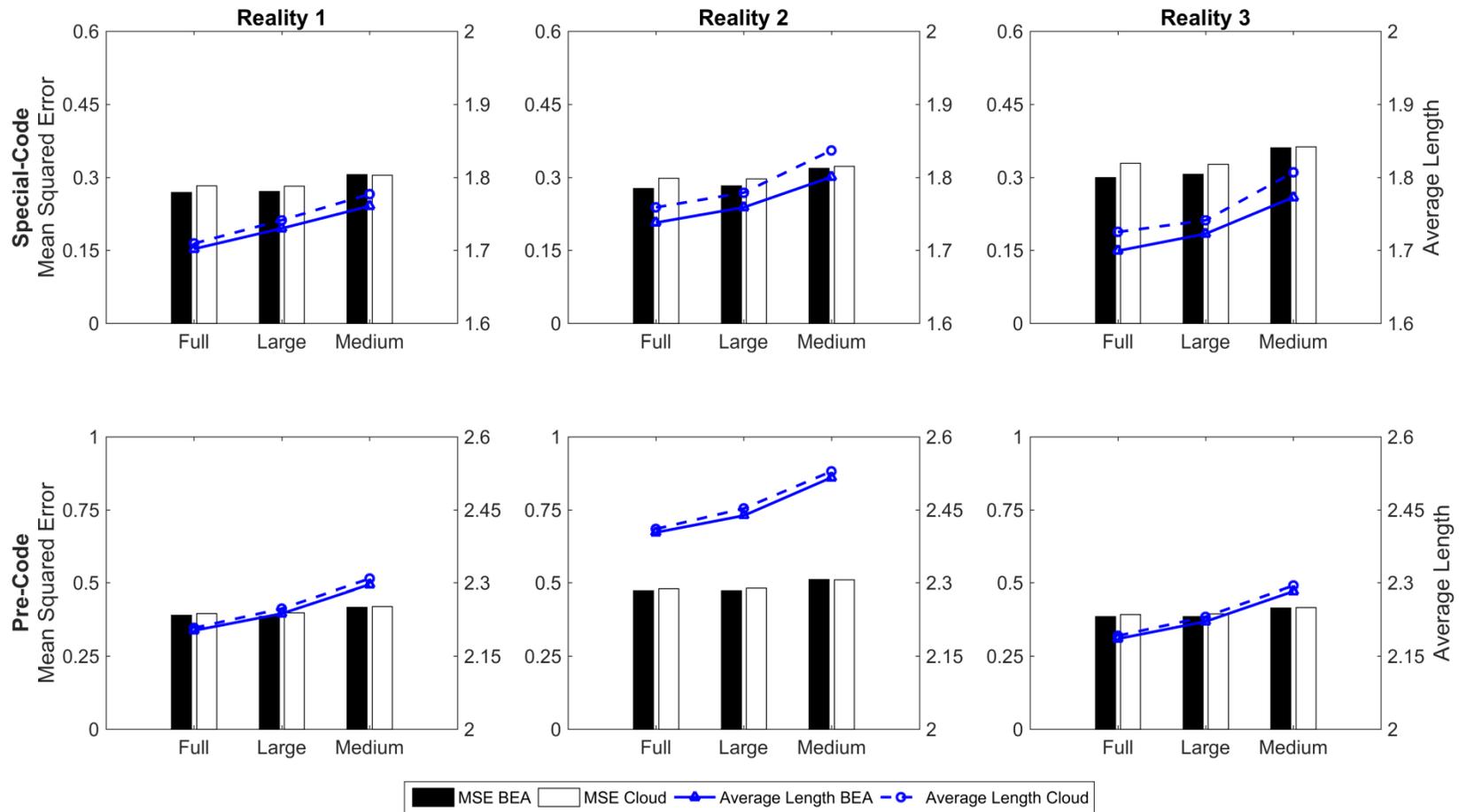
In some cases, it is observed that the BEA can generate predictions that better capture the mean and maintain the same level of accuracy as the cloud method, but requiring a smaller training subset size. For example, though to a lesser extent, when BEA is trained with a large subset of the Special-Code data set, it provides better MSE estimates, while matching the level of accuracy resulting from using the cloud method with the full training set (e.g. Figure 5 top panels, Table A.1). This reduction in 25% of the required NLTHAs translates into significant savings in computational time.

Similar benefits of BEA over the cloud method are also seen in the case of the Pre-Code building, but are not presented here for brevity.

## 5.2 *Sensitivity to analysis fidelity level*

The tests presented in Section 5.1 show that the BEA performs well for high fidelity data, at times outperforming the cloud approach. In this section, the BEA is tested for its performance when low fidelity analysis data are used. The use of low fidelity data for fragility analysis is appealing as it significantly reduces the analysis time. However, this is at the expense of accuracy when compared to the high fidelity data set. Figure 6 and Table A.2 are generated following the same procedures discussed in the Section 5.1.

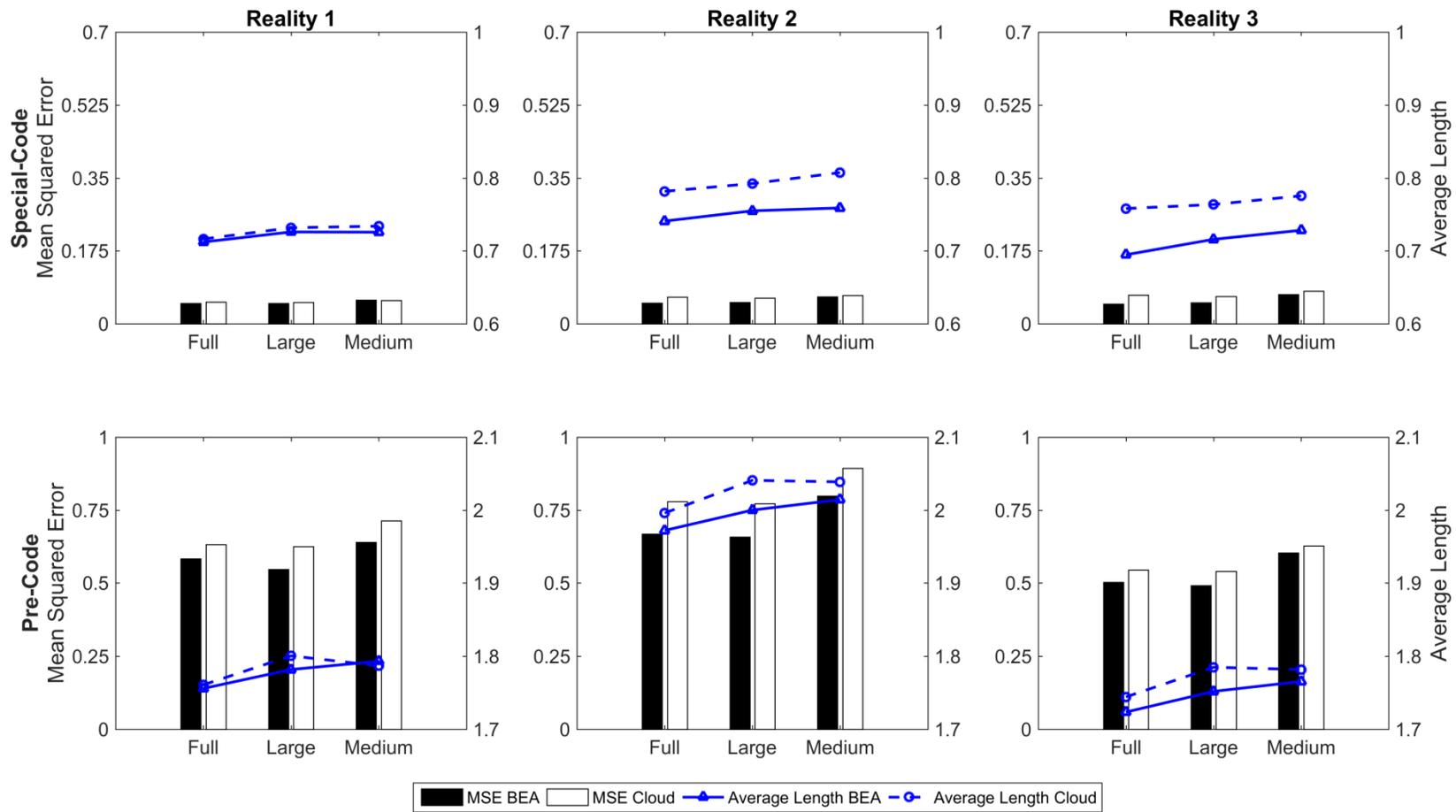
Similar observations to those made for the high fidelity analysis case are made. The coverage closely matches the 95% threshold, ( $\pm 1.9\%$ , see Table A.2). Improvement of MSE estimates is observed when the BEA is employed compared to cloud, especially for “pseudo-realities 2 and 3”. In the case of Special-Code building, this improvement can be translated to BEA MSE estimates reduced up to 21.65% and 30.68% for “pseudo-realities 2 and 3”, respectively. A substantial reduction of average length is also observed when comparing BEA and cloud estimates. This reduction ranges between 4.73% to 8.37%, for “pseudo-realities 2 and 3” for the case of Special-Code building, as shown in the Figure 6 top panels. The aforementioned cases also show the capability of the BEA, under certain conditions, to provide predictions that significantly reduce MSE and average length estimates, even when utilizing smaller training sample sizes. With regard to “pseudo-reality 1”, BEA again produces similar estimates to the cloud, which is the optimal method for this scenario. Improvements can be also observed in the case of the Pre-Code building, however, as in the case of the high fidelity sample, the benefits are less. It is noteworthy, that BEA always outperforms cloud in terms of the MSE and average length estimates when using medium training subsets. However, it is known from previous studies, that cloud method is not recommended to be used with small data sets.



1

2 Figure 5 - Comparison of BEA and Cloud method predictions in terms of Mean Squared Error (left y-axis of each graph) and Average Length (right y-axis of each graph) using  
 3 different sample sizes. Each column represents the different “pseudo-realities”, and each row the tested building types: Special-Code (top panels) and Pre-Code (bottom panels).  
 4 In this test-case high fidelity analysis is used alongside with advanced IM  $I_{Np}$  and stratified sampling process.

5



1

2 Figure 6 - Comparison of BEA and Cloud method predictions in terms of Mean Squared Error (left y-axis of each graph) and Average Length (right y-axis of each graph) using  
 3 different sample sizes. Each column represents the different “pseudo-realities”, and each row the tested building types: Special-Code (top panels) and Pre-Code (bottom panels).  
 4 In this test-case low fidelity analysis is used alongside with advanced IM  $I_{Np}$  and stratified sampling process.

### 5.3 Sensitivity to selection of IM

Previous studies have shown that advanced IMs (such as  $I_{Np}$ ) appear to be better correlated to displacement-based EDPs [25,29] resulting in EDP|IM data sets that have less scatter. In this test-case, the performance of the BEA is investigated when trained with a more scattered input data set, due to the use of the conventional IM,  $Sa(T_I)$ , for both high and low fidelity analysis data.

The outcomes of this test can be visually inspected in Figure 7, while a detailed summary of results is provided in Table A.3. For the sake of brevity, only the results for the Special-Code case-building are shown here. However, the same process is applied to the Pre-Code building and similar trends are observed but are seen to be less significant than for the Special-code.

As before, both approaches satisfy the coverage check. A close observation of the results reveals that the superiority of the BEA over cloud method in capturing the mean (i.e. considerable reduction of the BEA MSE estimates as shown in Table A.3) for both high and low fidelity inputs, even when  $Sa(T_I)$ , is used. The only exceptions are seen to be the cases of the medium sized high and low fidelity samples for “pseudo-reality 2” (Figure 7, middle panels), where cloud MSE predictions are better than those of BEA. The average length obtained from the BEA is always narrower, excluding the case mentioned above. Furthermore, BEA appears to perform significantly better when trained with low fidelity inputs, for “pseudo-realities 2 and 3”.

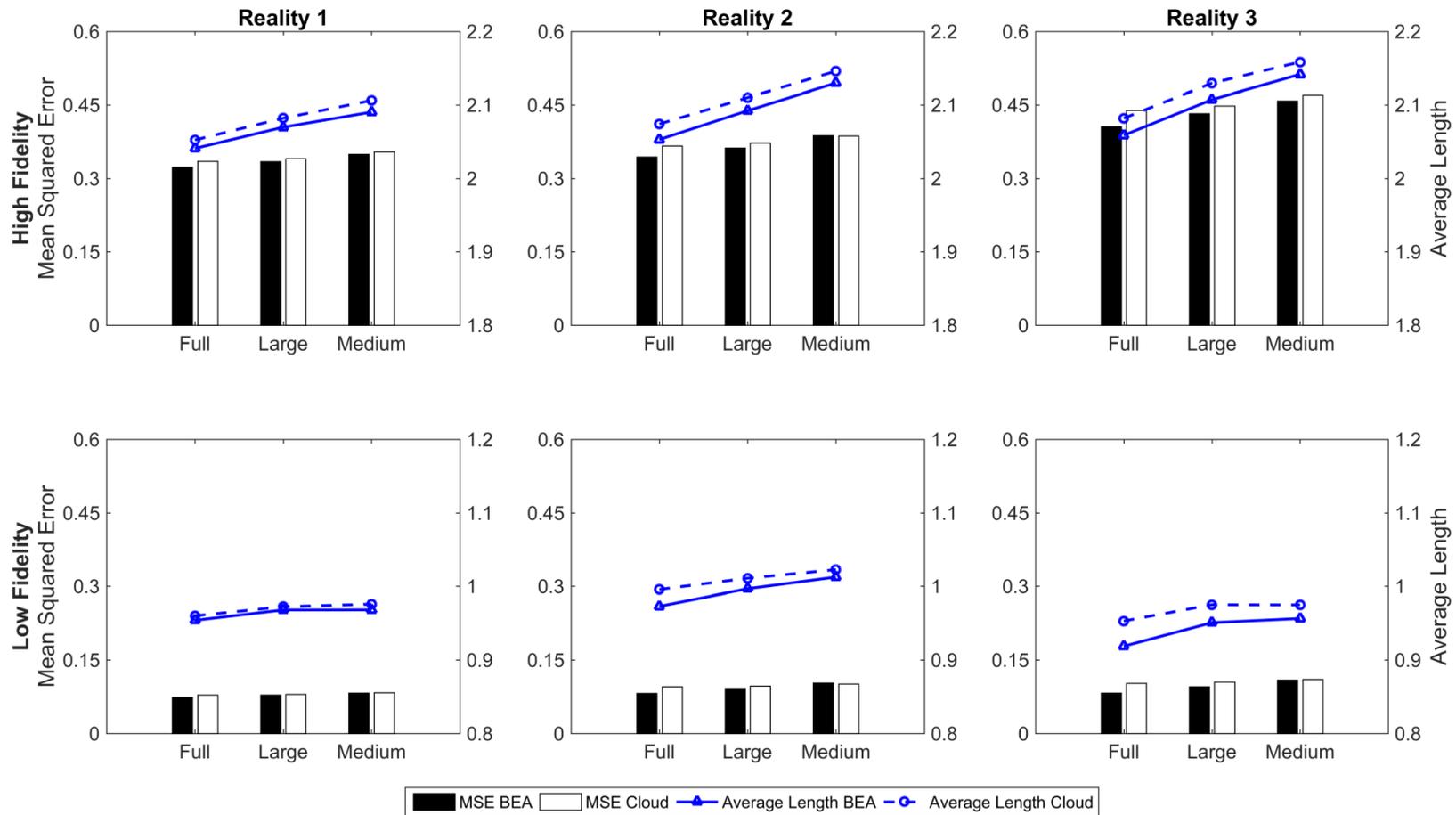
### 5.4 Sensitivity to the sampling approach

In this section a random sampling procedure is used to train both BEA and cloud models. The motivation for doing this is to explore whether the choice of a less sophisticated sampling approach has any effect in the predictions of BEA. This is a preliminary step towards the development of an optimal experimental design for the BEA method.

The same case study arrangement as in Section 5.3 is used here in order to investigate the effect the random sampling procedure. Following a visual inspection of Figure 8, it can be concluded that the observations between the BEA and cloud estimates are essentially consistent with what it is shown in Section 5.3. Nevertheless, it is interesting to evaluate directly the sensitivity of predictions when using random sampling instead of stratified sampling.

The results presented in Table A.3 and Table A.4 reveal that BEA predictions (as well as cloud predictions) show some improvement in both MSE and average length when random sampling is used instead of stratified sampling. Regarding the comparison of BEA and cloud performance metrics, it is observed that the choice of sampling does not have significant effects on capturing the mean function, resulting in similar differences for all fidelity levels and “pseudo-realities”. The BEA with random sampling, maintains its superiority over cloud method even for the low fidelity cases using medium sample sizes. Moreover, random sampling shows slight improvement in terms of average length metric, resulting narrower lengths comparing to ones calculated using stratified sampling.

In theory, one would expect stratified sampling to perform better than random sampling, as the former approach ensures that all parts of the IM range are sampled reasonably well and hence to capture local-scale features in the predicted EDP|IM curves if present. In our pseudo-realities however, there are no such prominent local features so, at first sight, it appears that stratified sampling has few advantages. However, stratified sampling does have an additional advantage that is not quantified by the metrics described above. Specifically, we find that stratified sampling ensures monotonicity of the predicted EDP|IM curve in the vast majority of the cases. In contrast, random sampling is more likely to provide a non-monotonically increasing predicted EDP|IM curve, especially for the cases where reduced training samples are utilized. The proportion of non-monotonically increasing predicted EDP|IM curves obtained from random sampling can reach up to 15%, for the case of reduced samples (i.e. large and medium) expressed in terms of  $I_{Np}$ . This is high compared to the proportion for stratified sampling, which is less than 4%.

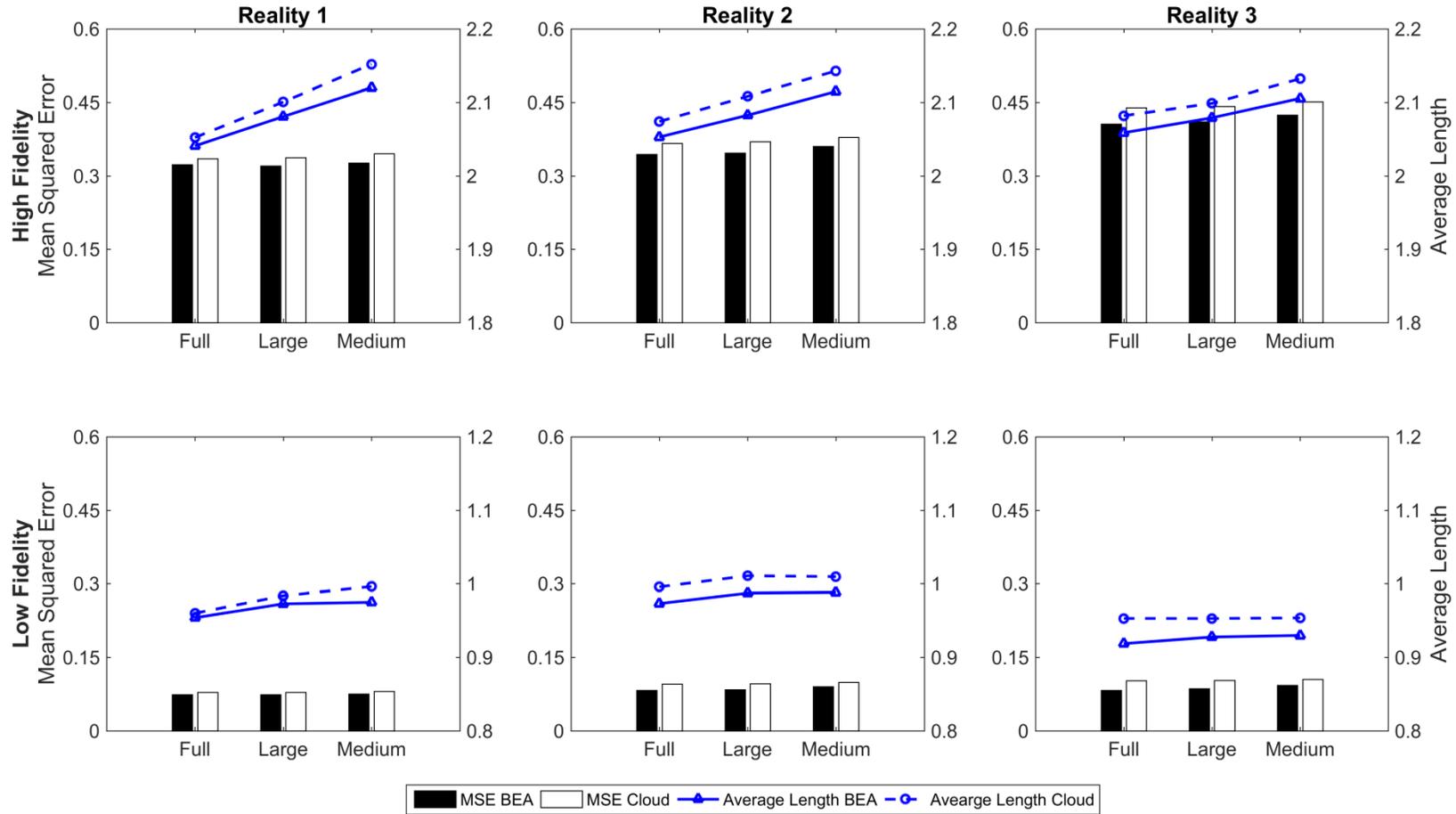


1

2 Figure 7 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error (left y-axis of each graph) and Average Length (right y-axis of each graph) using  
 3 different sample sizes. Each column represents the different “pseudo-realities”, and each row the level of analysis fidelity: High Fidelity (top panels) and Low Fidelity (bottom  
 4 panels). In this test-case Special-Code building is used alongside with the IM  $Sa(T_I)$  and stratified sampling process.

5

1



2

3 Figure 8 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error (left y-axis of each graph) and Average Length (right y-axis of each graph) using  
 4 different sample sizes. Each column represents the different “pseudo-realities”, and each row the level of analysis fidelity: High Fidelity (top panels) and Low Fidelity (bottom  
 5 panels). In this test-case Special-Code building is used alongside with the IM  $Sa(T_I)$  and random sampling process.

6

1 5.5 Sensitivity of BEA to the selection of covariance model

2 The last test conducted as part of the BEA evaluation process is to investigate the sensitivity of BEA to  
3 the selection of the covariance model. It is reminded that the Gaussian covariance model is used to  
4 produce the “pseudo-realities 2 and 3”, however, the user does not know this in practice when choosing  
5 the covariance model for BEA. Thus, this test is partly designed to explore the effect of choosing a  
6 wrong covariance model. For the sake of consistency, the Special-Code building is used here alongside  
7 with the IM  $I_{Np}$ , high fidelity analysis and stratified sampling approach.

8 The Gaussian model is the default choice ensuring the EDP|IM curves are very smooth (infinitely  
9 differentiable). A number of commonly used parametric models are studied here to explore the  
10 performance of the BEA when the wrong covariance model is used. Specifically, the 9 covariance  
11 models tested herein are listed below:

- Gaussian
- Matérn
- Spherical
- Cubic
- Exponential
- Cauchy
- Gneiting
- Circular
- Pure-nugget

12 The above mentioned models, which are available within the cov.spatial function of the geoR package,  
13 are described in detail in its documentation [18].

14 Figure 9 shows as a bar chart the BEA’s predictions when using the various covariance models assessed  
15 in terms of the MSE and the average length metrics (left and right panel respectively) for different  
16 sample subsets. The associated predictions for the cloud method, illustrated as dashed lines, are also  
17 provided as a point reference. A detailed summary of the results is presented in Table A.5. The coverage  
18 probability for all tested covariance structures exceeds matches or slightly exceed the nominal coverage  
19 probability.

20 Close inspection of Figure 9 reveals that irrespective of the choice of the covariance model, the BEA  
21 always outperforms the cloud in both MSE (significant improvements) and average length (slight  
22 improvements). Regarding the average length predictions (Figure 9, right panel), it is observed that  
23 Matérn and Exponential models perform better than remaining models, especially for the full and large  
24 sample cases. Gaussian, Spherical, Cubic, Gneiting, Circular and Pure nugget models perform  
25 effectively in a consistent manner, regardless of the subset size. The worst performance is recorded for  
26 the Cauchy model, where the resultant predictions are similar to the cloud’s predictions.

27 The most favorable models for reducing the average length, namely Matérn and Exponential, are also  
28 seen to be the two best options for estimating the MSE (Figure 9, left panel) for all training subsets.  
29 Circular, Cauchy and Gneiting models using medium subset are characterized by poor performance.  
30 The rest of the tested models perform well in reducing the MSE under all sampling sizes, showing small  
31 differences between them.

32 This test shows that the BEA predictions are sensitive to the selection of different covariance models.  
33 Matérn and Exponential models result in significant reductions in the estimates of both MSE and  
34 average length. Gaussian and the remaining models are a good compromise in improving both MSE  
35 and average length estimates. Cauchy and Pure Nugget covariance models are not recommended.

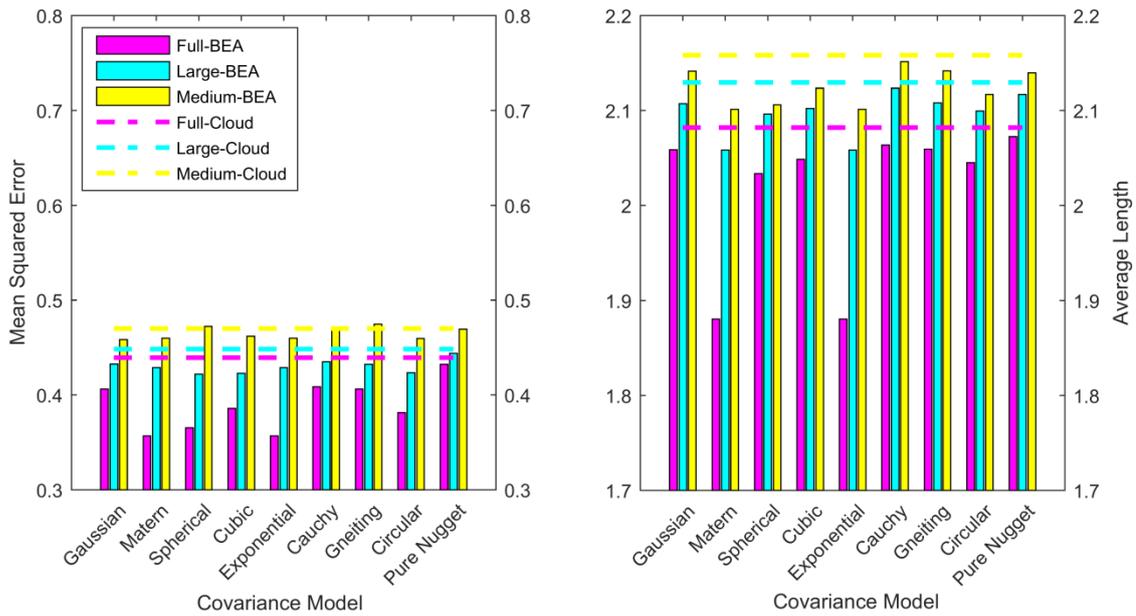
36 However, another consideration when choosing a covariance model is that the resultant mean estimate  
37 curve should be monotonically increasing in order to have a physical meaning. To this aim, an  
38 illustrative example is presented in Figure 10, where BEA mean estimates are obtained using three  
39 different covariance structures, namely Gaussian, Matérn and Cubic models. For this example, each  
40 version BEA is trained using a large training subset as obtained from the Special-Code building  
41 analyzed at low fidelity level.

42 The results obtained from this example are in a good agreement with the tests presented in Figure 9,  
43 where Matérn and Cubic covariance structures contribute to a significant reduction of both the MSE

1 and the average length estimates comparing to the cloud method. However, Figure 10 also reveals the  
 2 oversensitivity of the latter covariance structures at locations where training data are sparse, i.e. middle  
 3 range of IMs.

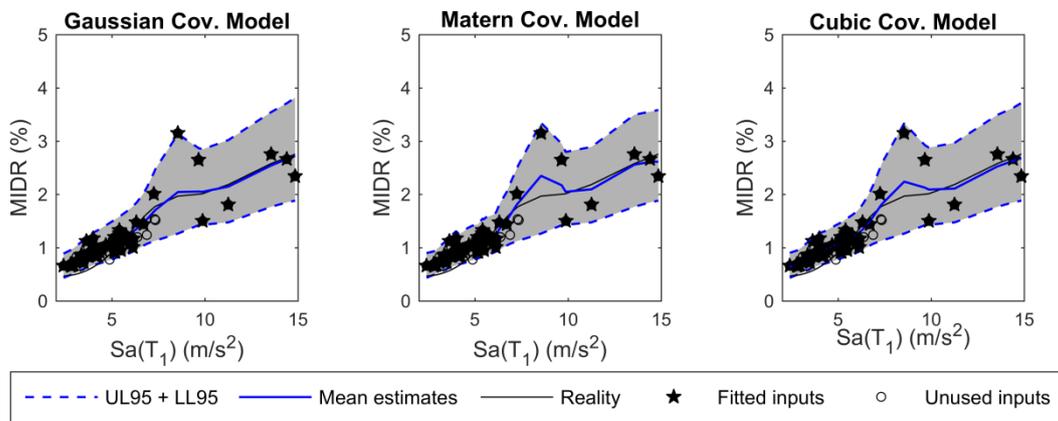
4 This oversensitivity is translated into a ‘jump’ of the curves (also present in the case of Gaussian model)  
 5 followed by a substantial ‘drop’ (Figure 10, middle and right panels), resulting in mean estimates that  
 6 are no longer monotonically increasing. This limitation suggests that the Gaussian model is the most  
 7 appropriate to describe the covariance structures for problems of this nature. Further investigation is  
 8 required to explore the applicability of the different covariance structures.

9 Table 2 illustrates the sensitivity of BEA predictions to the choice of the covariance structure, and how  
 10 these compare to the cloud predictions.



11  
 12 Figure 9 - Sensitivity of BEA predictions when using different covariance models and comparison with cloud  
 13 method predictions, in terms of mean squared error (left panel) and average length (right panel).

14



15  
 16 Figure 10 - BEA mean estimates using Gaussian (left panel), Matérn (middle panel) and Cubic covariance models  
 17 (right panel)

1 The results obtained from this example are in a good agreement with the tests presented in Figure 9,  
 2 where Matérn and Cubic covariance structures contribute to a significant reduction of both the MSE  
 3 and the average length estimates comparing to the cloud method. However, Figure 10 also reveals the  
 4 oversensitivity of the latter covariance structures at locations where training data are sparse, i.e. middle  
 5 range of IMs.

6 This oversensitivity is translated into a ‘jump’ of the curves (also present in the case of Gaussian model)  
 7 followed by a substantial ‘drop’ (Figure 10, middle and right panels), resulting in mean estimates that  
 8 are no longer monotonically increasing. This limitation suggests that the Gaussian model is the most  
 9 appropriate to describe the covariance structures for problems of this nature. Further investigation is  
 10 required to explore the applicability of the different covariance structures.

11 Table 2 - Predictions of BEA for different covariance structures and comparison with cloud method predictions  
 12 in terms of MSE, Average Length and Coverage probability, for the case of a single case-study building. In this  
 13 test-case Special-Code building is used alongside with the IM  $I_{Np}$  and stratified sampling process.

<b>Covariance Model</b>	<b>MSE BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Average Length BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Coverage (%) BEA</b>	<b>Coverage (%) Cloud</b>
Gaussian	0.039 (0.044)	10.42	0.814 (0.828)	1.71	98.41	98.41
Matérn	0.042 (0.044)	4.17	0.756 (0.828)	8.70	100	98.41
Cubic	0.040 (0.044)	8.81	0.783 (0.828)	5.41	100	98.41

14

15 *5.6 Discussion - Case-study example utilizing original analysis data set*

16 The results reported in Sections 5.1 to 5.5 were averaged over 1000 simulation data sets, however, it is  
 17 important to illustrate what those average performance measures mean in practice for a single data set.  
 18 To this aim, the case study introduced in Section 5.5 is studied in detail to benchmark the capability of  
 19 the BEA to estimate the mean and the associated variance, for the conditional distribution of EDP|IM  
 20 relationship.

21 The Gaussian model is employed to describe the covariance structure of the BEA. The resultant data  
 22 outputs are then used to train both the BEA and cloud approach, in order to predict the “pseudo-reality  
 23 3”. For the sake of brevity, only one sampling-case is shown per sample size. The reason for choosing  
 24 this case-study example is because it depicts some of the capabilities of the BEA, and its superiority  
 25 over the cloud method, following the observations discussed in Sections 5.1-5.5. However, consistent  
 26 observations arise when testing alternative case studies.

27 Figure 11 shows the mean estimates and the 95% confidence intervals for both the BEA and the cloud,  
 28 when employing three different sampling sizes. Table 3 shows the sample size, the metrics used to  
 29 assess the performance of the BEA and the cloud approach as well as the respective differences, for the  
 30 three different sample sizes investigated. The coverage probability for both the BEA and the cloud  
 31 approach is always exceeding the nominal coverage probability.

32 A visual inspection of Figure 11 reveals the capability of the BEA to identify the ‘jump’ occurring in  
 33 the mid-range of the “pseudo-reality”. This “jump” cannot be captured by the cloud method due to the  
 34 fixed mathematical model used by the latter approach. The flexibility of the BEA is also highlighted by  
 35 the reduction in MSE, which ranges between 5.7-10.4%, depending on the training subset used.  
 36 Regarding, the average length metric, small improvements are also observed when using the BEA, and  
 37 can be translated to 0.7-2.7% difference comparing to cloud’s average length.

38 Table 3 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error, Average Length  
 39 and Coverage probability, for the case of a single case-study building. In this test-case Special-Code building is  
 40 used alongside with the IM  $I_{Np}$  and stratified sampling process.

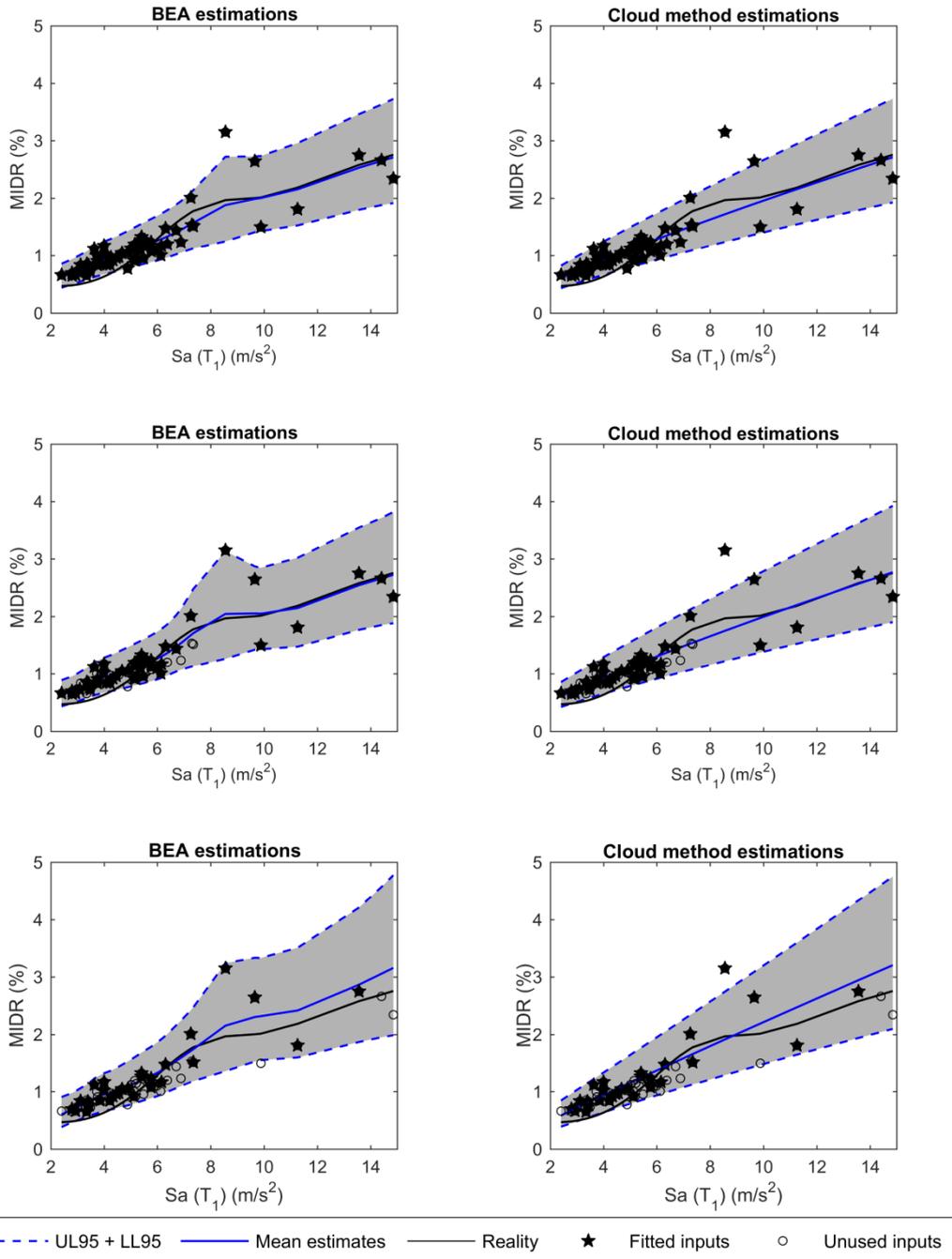
<b>Sample Size</b>	<b>MSE BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Average Length BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Coverage (%) BEA</b>	<b>Coverage (%) Cloud</b>
--------------------	------------------------	-----------------------	-----------------------------------	-----------------------	-------------------------	---------------------------

Full	0.039 (0.042)	7.14	0.743 (0.748)	0.67	98.41	95.24
Large	0.039 (0.044)	10.42	0.814 (0.828)	1.71	98.41	98.41
Medium	0.051 (0.055)	5.74	0.922 (0.948)	2.74	98.41	98.41

1  
2  
3  
4  
5  
6

The observations shown in this section are in most cases in agreement with the conclusion drawn before. Nevertheless, it cannot be ignored that at this section only one individual case is presented. As a result, one should look at previous sections to get a complete picture of the overall performance of the BEA, when the tests are carried out for 1000 sets of artificially generated data.

7  
8



9  
10  
11

Figure 11 - Mean estimations and associated 95% confidence interval of BEA approach (left panels) and the cloud method for full (top row), large (mid row) and medium sample size (bottom row). Case study building: Special-

1 Code; Analysis fidelity: Low; Reality: 3; IM:  $I_{np}$ ; Sampling method: Stratified sampling; Covariance model:  
2 Gaussian.

### 3 5.7 Fragility analysis using BEA

4 In this section, the process for deriving fragility functions using BEA is presented. A fragility function  
5 represents the probability that a *EDP* will exceed a certain level of capacity damage state  $DS_i$  given  
6 an earthquake intensity  $IM$ . For a traditional cloud analysis this probability is usually derived under  
7 the assumption that *EDP* is a lognormal variable, and the parameters of its distribution are fixed at the  
8 best estimates obtained from the predicted  $EDP|IM$  curves (e.g. [30]):

$$P(EDP > DS_i | IM) = 1 - \Phi \left( \frac{\ln DS_i - \ln \mu_{EDP|IM}}{\sigma_{\ln EDP|IM}} \right) \quad (8)$$

9 where  $\Phi$  is the standardized Gaussian CDF,  $\mu_{EDP|IM}$  and  $\sigma_{\ln EDP|IM}$  are the estimated median and the  
10 standard deviation of  $\ln(EDP)$  given  $IM$ . The BEA can improve upon this, however, both in its more  
11 flexible representation of the predicted  $EDP|IM$  curve itself and in its ability to account for parameter  
12 uncertainty. As described in Section 4, in the Bayesian setting the fragility for a given  $IM$  is just the  
13 posterior predictive probability of exceeding  $DS_i$ , and can be obtained directly from the output of the  
14 `krige.bayes()` function using the `geoR` package.

15 The case-study introduced in Section 5.5 is used as an illustrative example for the comparison of BEA  
16 and cloud method at the fragility level. Table 4 shows the descriptions and the thresholds associated  
17 with each damage state, which are used for the derivation of the fragility curves. This damage scale is  
18 based on the re-interpretation of the Homogenized Reinforced Concrete (HRC) damage scale proposed  
19 by [31] and that in [32]. More details on the adjustments of this damage scale to the case-study buildings  
20 can be found in [7].

21 Table 4 - Description of damage states and damage state thresholds used in this study.

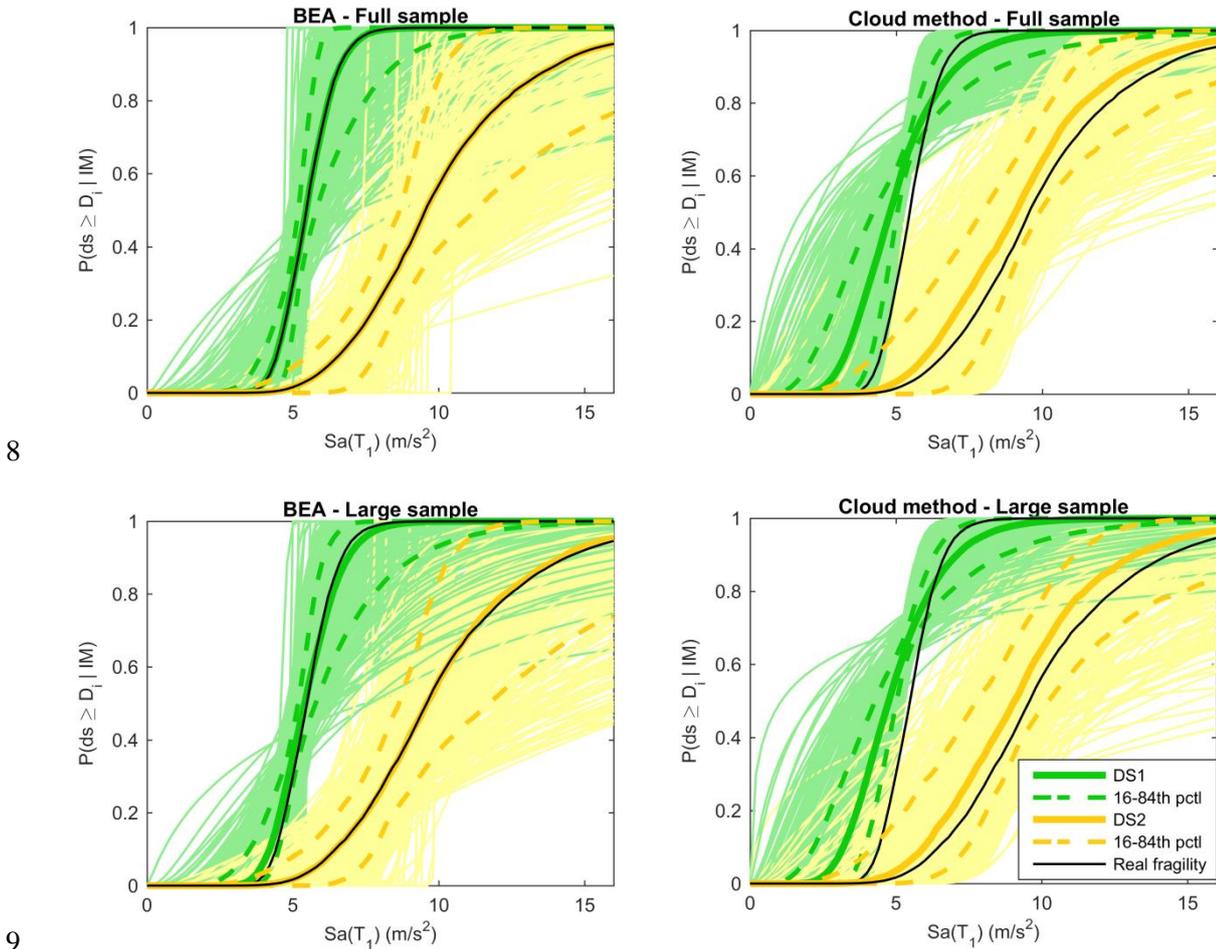
HRC Damage State	DS1 Moderate	DS2 Extensive	DS3 Partial Collapse
Observed Damage	Cracking in most beams and columns. Some yielding in a limited number. Limited concrete spalling	Ultimate strength is reached in some elements	Failure of some columns or impending soft-story failure
Response Characteristics (Threshold defined by the first occurrence of any of these)	Global yield displacement, as obtained by the idealized curve.	Maximum moment capacity of a supporting column is reached.	- There is a drop in strength to 80% of the maximum global capacity; - Shear failure of one element; - The rotation capacity of a critical column is reached.
MIDR Threshold Pre-Code structure (%)	0.49	1.53	3.00
MIDR Threshold Special-Code structure (%)	0.95	2.11	5.62

22

23 In consistency with the process described in Section 4, a large number of EDPs is simulated (1000  
24 simulations) at the given locations of  $IM$ , using the same mean and standard deviation values as  
25 computed for each “pseudo-reality”. The suite of the 1000 fragility functions is then generated for both

1 BEA and cloud method (median and 16-84<sup>th</sup> percentiles), and compared to the median “real fragility”  
 2 as shown in Figure 12. Note that for this example the large sample is used to generate BEA fragility  
 3 curves, while full sample is utilized for the case of cloud.

4 In Figure 12 an apparent superiority of BEA over the cloud method in predicting the fragility is  
 5 observed. Specifically, the median predictions of BEA match almost perfectly the “real fragility” for  
 6 both damage states, even when 25% fewer analyses are used to train BEA (large and full samples are  
 7 used for BEA and cloud respectively). This practically means that BEA not only improves the accuracy



8  
 9  
 10 Figure 12 - A suite of 1000 fragility curves (median and 16-84<sup>th</sup> percentile) generated from sets of simulated  
 11 analyses, for BEA (left panels) and cloud method (right panels), and comparison with median “real fragility”,  
 12 utilizing: full sample size (top panels) and large sample size (bottom panels).

13 captures the variability of the fragility curves, showing varied widths in the 16<sup>th</sup> and 84<sup>th</sup> percentile  
 14 bounds across the IM range, as illustrated in Figure 12. It is highlighted that the third damage state is  
 15 not reached for the Special-Code building using the chosen suite of ground motions.

16 Similar observations are essentially obtained for “pseudo-realities 2 and 3” of all tested case-studies,  
 17 while both BEA and cloud manage to predict well the “real fragility” in all the scenarios related to  
 18 “pseudo-reality 1”.

19 **6. Conclusions**

20 This paper presents a new Bayesian emulation approach, which is a version of kriging, for estimating  
 21 the mean and the associated variance of the conditional distribution of EDP|IM relationship, accounting  
 22 for the uncertainty of the BEA model’s input parameters. The capabilities of this new approach, which

1 overcomes some of the limitations of existing approaches, are investigated when trying to predict  
2 different “pseudo-realities” and the results are compared against standard cloud analysis.

3 Two RC mid-rise building representing the Special- and Pre-Code vulnerability classes of the Italian  
4 building stock are used as case-study structures. These structures are analyzed at two levels of fidelity,  
5 namely high and low fidelity, and their seismic performance is expressed as a function of standard and  
6 advanced IMs, i.e.  $Sa(T_I)$  and  $I_{Np}$  respectively. Three “pseudo-realities” (based on different  
7 mathematical models) are fitted to each of the resultant 8 data sets, and are then used as a target to  
8 evaluate the performance of new emulation-based approach. Next, 1000 artificial sets of EDPs are  
9 generated at given IM levels utilizing each “pseudo-reality” as a mean function and the standard  
10 deviation as calculated from the sample of real analysis data. MSE, average length and coverage  
11 probability are the metrics used to assess the performance of the BEA.

12 A number of tests are conducted to highlight the sensitivity of the proposed emulation-based method to  
13 several input assumptions and show how this method compares to the cloud approach. The main  
14 conclusions of this study can be summarized as follows:

15 The BEA has this ability to better capture the mean function of each “pseudo-reality” for all tested case  
16 studies. This observation is clearly reflected in the MSE predictions of the BEA, which are smaller than  
17 those of the cloud analysis for all tested cases. A reduction of the average length also results when the  
18 BEA is used compared to the cloud method. This minor reduction, can be translated as a reduction of  
19 the uncertainty of the resultant predictions. BEA outperforms cloud for both metrics in the vast majority  
20 of cases. However, more significant improvements are obtained for the “pseudo-realities 2 and 3”.  
21 Regarding, “pseudo-reality 1”, the cloud method is known to be optimal, and therefore the results  
22 presented here suggest that BEA is indistinguishable from the optimal method in this situation.

23 The BEA can essentially result in more accurate mean predictions than cloud method with less  
24 uncertainty. The best estimates of the BEA are attained when combined with Special-Code vulnerability  
25 class, low fidelity analysis and advanced IM  $I_{Np}$ .

26 Under certain conditions, the BEA can reduce the overall computation time by decreasing the analysis  
27 inputs required for training. In particular, the BEA can generate predictions that better capture the mean  
28 function, closely matching the level of accuracy as in the cloud method, but requiring a smaller number  
29 of training inputs. This capability of the BEA is mainly observed when a large training subset is  
30 combined with the favorable inputs mentioned above (namely Special-Code vulnerability class, low  
31 fidelity analysis and advanced IM  $I_{Np}$ ).

32 The estimates of both the BEA and the cloud approach are, as expected sensitive to the size of the  
33 training subsets. Smaller training subsets, result in worse estimates for both MSE and average length  
34 metrics. The use of medium subset resulted in some cases non reliable outputs, therefore it is not  
35 recommended to be used for the training of the BEA.

36 The choice of sampling approach does not yield significant differences in the resultant MSE and average  
37 length metrics, however, stratified sampling is a more suitable sampling method, comparing to random  
38 sampling, for the experimental design of the BEA. This is because stratified sampling is better able to  
39 capture local features in the predicted EDP|IM curve, and is also better able to preserve important global  
40 features such as monotonicity.

41 Some sensitivity is witnessed when different parametric models are used to describe the covariance  
42 structure of BEA. Gaussian model constitutes a good all-around choice for improving both MSE and  
43 average length estimates but also it ensures the smoothness of the GP realizations, which is essential  
44 for the nature of this problem. Finally, BEA drastically improves the prediction of the median fragility  
45 for all damage states, matching perfectly the “real fragility” in most cases, and also capturing  
46 realistically the variability of the fragility curves at a smaller computational expense.

1 Overall, although BEA is more complex as compared to the cloud method, it presents significant  
2 advantages over the latter for predicting nonlinear realities and for representing uncertainty. It is  
3 observed to perform well when high fidelity analysis and larger data samples are used. However, it also  
4 outperforms the cloud method when predicting nonlinear realities with smaller data samples. The cloud  
5 method remains a suitable option when predicting a linear EDP|IM relationship (linear reality).

## 6 **Acknowledgements**

7 The first author would like to acknowledge the Engineering and Physical Sciences Research Council  
8 (EPSRC) in the UK and the AIR Worldwide Ltd through the Urban Sustainability and Resilience  
9 program at University College London for funding for his research work. The authors are very grateful  
10 to Dr. Carmine Galasso and Dr. Joakim Beck for their constructive comments on an earlier version of  
11 this work.

## References

- [1] Deierlein GG, Krawinkler H, Cornell CA. A framework for performance-based earthquake engineering. *Pacific Conf Earthq Eng* 2003;273:1–8. doi:10.1061/9780784412121.173.
- [2] Erberik MA. Seismic Fragility Analysis. In: Beer M, Kougioumtzoglou IA, Patelli E, Au IS-K, editors. *Encycl. Earthq. Eng.*, Berlin, Heidelberg: Springer Berlin Heidelberg; 2014, p. 1–10. doi:10.1007/978-3-642-36197-5\_387-1.
- [3] Minas S, Chandler RE, Rossetto T. New emulation-based approach for probabilistic seismic demand. *16th World Conf. Earthq. Eng. 16WCEE 2017*, Santiago, Chile: 2017.
- [4] Bazzurro P, Cornell CA, Shome N, Carballo JE. Three proposals for characterizing MDOF nonlinear seismic response. *J Struct Eng* 1998;124:1281–9. doi:10.1061/(ASCE)0733-9445(1998)124:11(1281).
- [5] Luco N, Cornell CA. Seismic drift demands for two SMRF structures with brittle connections. *Structural Engineering World Wide. Struct Eng World Wide; Elsevier Sci Ltd, Oxford, Engl* 1998;paper T158-3.
- [6] Jalayer F. *Direct probabilistic seismic analysis: implementing non-linear dynamic assessments*. Stanford University, 2003.
- [7] Rossetto T, Gehl P, Minas S, Galasso C, Duffour P, Douglas J, et al. FRACAS: A capacity spectrum approach for seismic fragility assessment including record-to-record variability. *Eng Struct* 2016;125:337–48. doi:10.1016/j.engstruct.2016.06.043.
- [8] Modica A, Stafford PJ. Vector fragility surfaces for reinforced concrete frames in Europe. *Bull Earthq Eng* 2014;12:1725–53. doi:10.1007/s10518-013-9571-z.
- [9] Ibarra LF, Krawinkler H. *Global Collapse of Frame Structures under Seismic Excitations Global Collapse of Frame Structures under Seismic Excitations*. Berkeley, California: 2005.
- [10] Cornell CA, Jalayer F, Hamburger RO, Foutch DA. Probabilistic Basis for 2000 SAC Federal Emergency Management Agency Steel Moment Frame Guidelines. *J Struct Eng* 2002;128:526–33. doi:10.1061/(ASCE)0733-9445(2002)128:4(526).
- [11] Jalayer F, De Risi R, Manfredi G. Bayesian Cloud Analysis: efficient structural fragility assessment using linear regression. *Bull Earthq Eng* 2014;13:1183–203. doi:10.1007/s10518-014-9692-z.
- [12] O’Hagan A. Bayesian analysis of computer code outputs: A tutorial. *Reliab Eng Syst Saf* 2006;91:1290–300. doi:10.1016/j.res.2005.11.025.
- [13] Kennedy MC, O’Hagan A. Bayesian Calibration of Computer Models. *J R Stat Soc Ser B (Statistical Methodol)* 2001;63:425–64. doi:10.1111/1467-9868.00294.
- [14] Saltelli A, Chan K, Scott E. *Sensitivity analysis. Wiley series in probability and statistics*. Chichester: Wiley; 2000.
- [15] Tarantola A. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia: SIAM; 2005.
- [16] David M. *Geostatistical ore reserve estimation*. Amsterdam: Elsevier Ltd; 1977.
- [17] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis, Third Edition*. CRC Press; 2013.
- [18] Ribeiro Jr P, Diggle P. geoR: A package for geostatistical analysis. *R-NEWS* 2001;1:15–8.
- [19] Sacks J, Welch WJ, Mitchell JSB, Henry PW. Design and Experiments of Computer Experiments. *Stat Sci* 1989;4:409–23.

- 1 [20] R Core Team. R: A language and environment for statistical computing., Vienna, Austria: R  
2 Foundation for Statistical Computing; 2014.
- 3 [21] Regio Decreto 16/11/1939 n. 2229. Norme per la esecuzione delle opere in conglomerato  
4 cementizio semplice e armato. G.U. n. 92. 1940.
- 5 [22] Decreto Ministeriale del 14/01/2008. Norme Tecniche per le Costruzioni. Rome: Gazzetta  
6 Ufficiale della Repubblica Italiana, 29.; 2008.
- 7 [23] EN 1998-1. Eurocode 8: Design of structures for earthquake resistance – Part 1: General rules,  
8 seismic actions and rules for buildings. The European Union Per Regulation 305/2011, Directive  
9 98/34/EC, Directive2004/18/EC; 2004.
- 10 [24] De Luca F, Elefante L, Iervolino I, Verderame GM. Strutture esistenti e di nuova progettazione :  
11 comportamento sismico a confronto. Anidis 2009 XIII Convegno - L' Ing. Sismica Ital.,  
12 Bologna: 2009.
- 13 [25] Minas S, Galasso C, Rossetto T. Spectral Shape Proxies and Simplified Fragility Analysis of  
14 Mid- Rise Reinforced Concrete Buildings. 12th Int. Conf. Appl. Stat. Probab. Civ. Eng.  
15 ICASP12, Vancouver, Canada: 2015, p. 1–8.
- 16 [26] Smerzini C, Galasso C, Iervolino I, Paolucci R. Ground motion record selection based on  
17 broadband spectral compatibility. Earthq Spectra 2013;30:1427–48.  
18 doi:10.1193/052312EQS197M.
- 19 [27] Rossetto T, Gehl P, Minas S, Nassirpour A, Macabuag J, Duffour P. Sensitivity analysis of  
20 different capacity approaches to assumptions in the modeling, capacity and demand  
21 representations. 2nd Int. Conf. Vulnerability Risk Anal. Manag. 6th Int. Symp. Uncertain.  
22 Model. Anal. (ISUMA2014). 13-16 July, Liverpool, UK: 2014.
- 23 [28] Bojórquez E, Iervolino I. Spectral shape proxies and nonlinear structural response. Soil Dyn  
24 Earthq Eng 2011;31:996–1008. doi:10.1016/j.soildyn.2011.03.006.
- 25 [29] Ebrahimian H, Jalayer F, Lucchini A, Mollaioli F, Manfredi G. Preliminary ranking of  
26 alternative scalar and vector intensity measures of ground shaking. Bull Earthq Eng  
27 2015;13:2805–40. doi:10.1007/s10518-015-9755-9.
- 28 [30] Jalayer F, Cornell CA. Alternative non-linear demand estimation methods for probability-based  
29 seismic assessments. Earthq Eng Struct Dyn 2009;38:951–72. doi:10.1002/eqe.876.
- 30 [31] Rossetto T, Elnashai AS. Derivation of vulnerability functions for European-type RC structures  
31 based on observational data. Eng Struct 2003;25:1241–63. doi:10.1016/S0141-0296(03)00060-  
32 9.
- 33 [32] Dolšek M, Fajfar P. The effect of masonry infills on the seismic response of a four-storey  
34 reinforced concrete frame — a deterministic assessment. Eng Struct 2008;30:1991–2001.  
35 doi:10.1016/j.engstruct.2008.01.001.

36

1 **Appendix**

2

3 Table A.1 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error, Average Length  
 4 and Coverage probability. In this test-case high fidelity analysis is used alongside with advanced IM  $I_{Np}$  and  
 5 stratified sampling process.

Case Study Building	Reality	Sample Size	MSE BEA (Cloud)	Difference (%)	Average Length BEA (Cloud)	Difference (%)	Coverage (%) BEA	Coverage (%) Cloud
Special-Code	1	Full	0.269 (0.283)	4.72	1.702 (1.710)	0.43	95.84	95.60
		Large	0.271 (0.282)	3.87	1.730 (1.741)	0.65	95.75	95.60
		Medium	0.306 (0.304)	-0.57	1.760 (1.777)	0.92	95.52	95.68
	2	Full	0.277 (0.299)	7.08	1.738 (1.759)	1.19	95.84	95.43
		Large	0.283 (0.297)	4.99	1.758 (1.779)	1.13	95.90	95.71
		Medium	0.319 (0.323)	1.19	1.801 (1.837)	1.98	95.59	95.56
	3	Full	0.299 (0.329)	9.01	1.699 (1.725)	1.50	96.06	95.49
		Large	0.307 (0.327)	6.15	1.723 (1.741)	1.06	95.81	95.54
		Medium	0.361 (0.363)	0.60	1.773 (1.807)	1.89	95.54	95.57
Pre-Code	1	Full	0.390 (0.396)	1.54	2.203 (2.208)	0.24	95.87	95.86
		Large	0.387 (0.398)	2.74	2.237 (2.248)	0.46	96.10	96.07
		Medium	0.418 (0.420)	0.46	2.297 (2.309)	0.52	95.82	95.87
	2	Full	0.473 (0.479)	1.33	2.403 (2.411)	0.30	95.98	95.96
		Large	0.473 (0.483)	1.98	2.438 (2.453)	0.60	96.05	96.13
		Medium	0.511 (0.510)	-0.2	2.517 (2.529)	0.50	95.88	95.89
	3	Full	0.386 (0.392)	1.61	2.186 (2.192)	0.28	95.89	95.88
		Large	0.386 (0.395)	2.31	2.221 (2.231)	0.44	96.08	96.08
		Medium	0.415 (0.416)	0.19	2.283 (2.294)	0.50	95.86	95.87

6

7

1 Table A.2 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error, Average Length  
 2 and Coverage probability. In this test-case low fidelity analysis is used alongside with advanced IM  $I_{Np}$  and  
 3 stratified sampling process.

Case Study Building	Reality	Sample Size	MSE BEA (Cloud)	Difference (%)	Average Length BEA (Cloud)	Difference (%)	Coverage (%) BEA	Coverage (%) Cloud
Special-Code	1	Full	0.049 (0.052)	5.96	0.712 (0.716)	0.58	95.48	95.02
		Large	0.049 (0.052)	5.48	0.726 (0.732)	0.79	95.48	95.25
		Medium	0.057 (0.056)	-0.79	0.726 (0.734)	1.13	94.98	95.00
	2	Full	0.050 (0.064)	21.65	0.741 (0.782)	5.22	96.21	95.25
		Large	0.052 (0.062)	16.23	0.755 (0.792)	4.73	96.03	95.51
		Medium	0.065 (0.069)	5.37	0.759 (0.807)	6.00	94.97	95.14
	3	Full	0.048 (0.069)	30.68	0.695 (0.758)	8.37	96.90	95.54
		Large	0.051 (0.066)	23.28	0.716 (0.764)	6.26	96.41	95.68
		Medium	0.070 (0.078)	10.16	0.728 (0.776)	6.09	95.21	95.11
Pre-Code	1	Full	0.583 (0.632)	7.80	1.756 (1.761)	0.28	95.06	94.77
		Large	0.547 (0.625)	12.58	1.782 (1.801)	1.04	94.91	94.80
		Medium	0.640 (0.714)	10.38	1.793 (1.787)	-0.35	94.15	94.10
	2	Full	0.668 (0.779)	14.28	1.972 (1.996)	1.20	95.05	94.90
		Large	0.658 (0.772)	14.83	2.000 (2.041)	2.00	94.98	94.93
		Medium	0.799 (0.894)	10.64	2.015 (2.039)	1.19	94.00	94.08
	3	Full	0.503 (0.545)	7.69	1.723 (1.744)	1.19	95.07	94.9
		Large	0.492 (0.541)	9.04	1.752 (1.785)	1.85	95.12	94.98
		Medium	0.603 (0.627)	3.85	1.765 (1.782)	0.92	94.11	94.02

4

5

1 Table A.3 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error, Average Length  
 2 and Coverage probability. In this test-case Special-Code building is used alongside with the IM  $Sa(T_1)$  and  
 3 stratified sampling process.

Fidelity	Reality	Sample Size	MSE BEA (Cloud)	Difference (%)	Average Length BEA (Cloud)	Difference (%)	Coverage (%) BEA	Coverage (%) Cloud
High fidelity	1	Full	0.323 (0.336)	3.76	2.041 (2.052)	0.56	95.90	95.58
		Large	0.335 (0.341)	1.83	2.070 (2.082)	0.59	95.80	95.57
		Medium	0.349 (0.354)	1.41	2.090 (2.106)	0.75	95.47	95.55
	2	Full	0.344 (0.367)	6.24	2.053 (2.074)	1.02	95.97	95.67
		Large	0.363 (0.372)	2.64	2.092 (2.110)	0.84	95.67	95.47
		Medium	0.388 (0.387)	-0.26	2.130 (2.146)	0.75	95.42	95.75
	3	Full	0.406 (0.439)	7.51	2.059 (2.082)	1.11	96.03	95.67
		Large	0.433 (0.448)	3.49	2.107 (2.129)	1.05	95.75	95.58
		Medium	0.458 (0.470)	2.44	2.141 (2.158)	0.78	95.44	95.60
Low fidelity	1	Full	0.074 (0.078)	5.66	0.954 (0.960)	0.64	95.58	95.30
		Large	0.078 (0.080)	2.23	0.968 (0.973)	0.46	95.58	95.25
		Medium	0.082 (0.083)	1.00	0.968 (0.976)	0.80	94.97	95.13
	2	Full	0.082 (0.095)	13.89	0.973 (0.996)	2.29	95.92	95.28
		Large	0.092 (0.097)	5.53	0.997 (1.011)	1.41	95.62	95.28
		Medium	0.103 (0.101)	-1.94	1.013 (1.023)	0.98	95.07	95.15
	3	Full	0.083 (0.102)	19.20	0.919 (0.953)	3.60	96.30	95.42
		Large	0.096 (0.105)	9.27	0.951 (0.975)	2.50	95.80	95.32
		Medium	0.109 (0.111)	1.16	0.956 (0.975)	1.90	95.00	95.02

4

5

1 Table A.4 - Comparison of BEA and cloud method predictions in terms of Mean Squared Error, Average Length  
 2 and Coverage probability. In this test-case Special-Code building is used alongside with the IM  $Sa(T_1)$  and random  
 3 sampling process. Full sample cases are not shown here as they coincide with the corresponding entries in Table  
 4 4.

<b>Fidelity</b>	<b>Reality</b>	<b>Sample Size</b>	<b>MSE BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Average Length BEA (Cloud)</b>	<b>Difference (%)</b>	<b>Coverage (%) BEA</b>	<b>Coverage (%) Cloud</b>
High fidelity	1	Large	0.320 (0.337)	5.07	2.081 (2.101)	0.95	96.12	96.05
		Medium	0.326 (0.345)	5.51	2.120 (2.152)	1.48	95.70	95.62
	2	Large	0.347 (0.370)	6.18	2.083 (2.108)	1.22	96.07	95.87
		Medium	0.360 (0.379)	4.92	2.115 (2.143)	1.29	95.70	95.58
	3	Large	0.411 (0.442)	7.07	2.079 (2.099)	0.94	96.12	95.82
		Medium	0.424 (0.452)	6.00	2.105 (2.132)	1.24	95.65	95.48
Low fidelity	1	Large	0.074 (0.079)	6.41	0.973 (0.984)	1.11	96.05	95.85
		Medium	0.075 (0.081)	6.91	0.975 (0.996)	2.13	95.50	95.27
	2	Large	0.084 (0.096)	12.51	0.987 (1.011)	2.32	96.07	95.50
		Medium	0.090 (0.099)	8.80	0.988 (1.010)	2.11	95.28	94.82
	3	Large	0.086 (0.103)	16.52	0.928 (0.953)	2.65	96.10	95.32
		Medium	0.093 (0.105)	11.85	0.930 (0.954)	2.50	95.32	94.77

5

6

1 Table A.5 - Sensitivity of BEA predictions when using different covariance models and estimation of the percent  
 2 differences to the associated cloud results. Case study building: Special-Code; Analysis fidelity: High; Reality: 3;  
 3 IM: Sa(T1); Sampling method: Stratified sampling.

Sample Size	Covariance Model	MSE	Difference (%)	Average Length	Difference (%)	Coverage (%)
Full	Gaussian	0.406	7.51	2.059	1.11	96.03
	Matérn	0.357	18.78	1.880	9.69	96.48
	Spherical	0.365	16.82	2.034	2.32	96.62
	Cubic	0.386	12.15	2.049	1.60	96.37
	Exponential	0.357	18.78	1.880	9.69	96.48
	Cauchy	0.409	6.98	2.064	0.87	95.95
	Gneiting	0.406	7.51	2.059	1.08	96.03
	Circular	0.381	13.17	2.045	1.77	96.55
Large	P. Nugget	0.432	1.58	2.073	0.45	95.83
	Gaussian	0.433	3.49	2.107	1.05	95.75
	Matérn	0.429	4.31	2.058	3.34	95.79
	Spherical	0.422	5.85	2.096	1.56	95.95
	Cubic	0.423	5.68	2.102	1.28	95.97
	Exponential	0.429	4.31	2.058	3.34	95.79
	Cauchy	0.435	2.96	2.124	0.27	95.71
	Gneiting	0.432	3.56	2.108	1.00	95.77
Medium	Circular	0.423	5.54	2.099	1.41	96.18
	P. Nugget	0.444	0.94	2.117	0.58	95.63
	Gaussian	0.458	2.44	2.141	0.78	95.44
	Matérn	0.460	2.16	2.101	2.63	95.03
	Spherical	0.472	-0.53	2.106	2.41	95.30
	Cubic	0.462	1.71	2.124	1.59	95.50
	Exponential	0.460	2.16	2.101	2.63	95.03
	Cauchy	0.471	-0.23	2.151	0.31	95.43
Medium	Gneiting	0.475	-1.04	2.142	0.75	95.50
	Circular	0.459	2.21	2.117	1.91	95.17
	P. Nugget	0.469	0.11	2.140	0.86	95.48

4