



HAL
open science

Using Bayesian model averaging to improve ground motion predictions

M Bertin, S Marin, C Millet, C Berge-Thierry

► **To cite this version:**

M Bertin, S Marin, C Millet, C Berge-Thierry. Using Bayesian model averaging to improve ground motion predictions. *Geophysical Journal International*, 2020, 220 (2), pp.1368-1378. 10.1093/gji/ggz486 . hal-04552014

HAL Id: hal-04552014

<https://hal.science/hal-04552014>

Submitted on 6 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Using Bayesian model averaging to improve ground motion predictions

M. Bertin¹, S. Marin², C. Millet^{1,2} and C. Berge-Thierry³

¹CMLA, ENS Paris-Saclay, 94230 Cachan, France. E-mail: michael.bertin@cmla.ens-cachan.fr

²CEA, DAM, DIF, F-91297 Arpajon Cedex, France

³CEA Saclay, 91191 Gif-sur-Yvette Cedex, France

Accepted 2019 November 26. Received 2019 October 16; in original form 2018 November 30

SUMMARY

In low-seismicity areas such as Europe, seismic records do not cover the whole range of variable configurations required for seismic hazard analysis. Usually, a set of empirical models established in such context (the Mediterranean Basin, northeast U.S.A., Japan, etc.) is considered through a logic-tree-based selection process. This approach is mainly based on the scientist's expertise and ignores the uncertainty in model selection. One important and potential consequence of neglecting model uncertainty is that we assign more precision to our inference than what is warranted by the data, and this leads to overly confident decisions and precision. In this paper, we investigate the Bayesian model averaging (BMA) approach, using nine ground-motion prediction equations (GMPEs) issued from several databases. The BMA method has become an important tool to deal with model uncertainty, especially in empirical settings with large number of potential models and relatively limited number of observations. Two numerical techniques, based on the Markov chain Monte Carlo method and the maximum likelihood estimation approach, for implementing BMA are presented and applied together with around 1000 records issued from the RESORCE-2013 database. In the example considered, it is shown that BMA provides both a hierarchy of GMPEs and an improved out-of-sample predictive performance.

Key words: Numerical approximations and analysis; Probabilistic forecasting; Earthquake ground motions.

1 INTRODUCTION

Ground-motion prediction equations (GMPEs) are empirical models used in seismic hazard assessment (SHA). These equations are usually determined from regression techniques using ground-motion data recorded by seismic networks. Ground motion can also be inferred from physically based models. Since the advent of probabilistic seismic hazard analysis (PSHA; Cornell 1968; McGuire 1976), three physical quantities are classically considered as random variables: the location of earthquakes, the magnitude of earthquakes and the expected ground motion [such as peak ground acceleration (PGA) or pseudospectral acceleration (PSA)] at a given site. Despite the sophistication of models, some discrepancies between the predicted and the observed ground motions do exist. These discrepancies are due to incomplete and simplified representation of physical processes, model structural inadequacies, measurement errors, conversion laws between magnitudes scales, uncertainties in the fault mechanism and soil conditions (and more generally uncertainties in the metadata assessment). In low-seismic regions such as Europe, ground-motion data are not significantly meaningful and predictions are often obtained from models calibrated with stronger

ground motions. This is referred to as a data sparsity problem, which is, with the epistemic uncertainties of empirical models, one of the most important challenges in PSHA.

Predictions obtained with a single GMPE can lead to statistical bias or underestimation of the output variability. The multimodel approach consisting in the selection and combination of a number of models using a logic-tree (Scherbaum & Kuehn 2011; Bommer 2012; Marzocchi *et al.* 2015) has emerged as a standard SHA practice. Further, this approach has become a regulatory requirement in PSHA (Budnitz *et al.* 1997). The estimate of branch weights is a matter of experts' judgement, which raises discussions and difficulties. Different frameworks were proposed for model selection and ranking for particular data sets (Scherbaum 2009; Delavaud *et al.* 2012; Kale & Akkar 2012). The logic-tree-based approach, however, ignores the uncertainty in model selection (or model uncertainty). Although there may be strong arguments (the physicist's expertise) leading to the selection of models, there is often no such strong argument for absolutely disqualifying the unselected models. If at least one of them could lead to radically different conclusions, then ignoring it underestimates uncertainty. Other approaches were proposed using probabilistic frameworks (Arroyo *et al.* 2014; Kowsari

et al. 2019) to address model uncertainty. Model averaging is a natural response to model uncertainty, especially when there is a large amount of uncertainty about model specifications that is not resolved by universal accepted theory.

This paper focuses on the so-called Bayesian model averaging (hereafter denoted as BMA), which is an extension of the standard Bayesian inference approach used in the social and health sciences (Raftery 1995; Hoeting *et al.* 1999; Culka 2014). The objective is to explore how this approach can improve accuracy and reliability of ground-motion prediction in PSHA or deterministic seismic hazard assessment (DSHA) when several GMPEs are considered. The combination of GMPEs into a mixture model has already been investigated with backbone models (Atkinson & Adams 2013; Haendel 2014; Bommer *et al.* 2015) and this paper deals with the contribution of the Bayes theory to that approach. With the BMA approach, the overall prediction probability density function (PDF) is an average of the posterior distributions under each of the models considered, weighted by their posterior model probability. The BMA weight reflects the relative quality of prediction of a given model regarding the set of training data considered. The performance of BMA is tested against the Reference database for Seismic ground-motion pRediction in Europe (RESORCE) database (Akkar *et al.* 2013), using nine GMPEs. This approach provides a hierarchy of GMPEs that avoids arbitrariness. Then it also provides a new model in the form of a linear combination of the GMPEs, with better predictions. This is due to the BMA formalism that allows the variance to be decomposed into within-model and between-model variances. Because the BMA approach accounts for the effect of model combination, it has the potential to improve the statistical predictive quality compared to the best model. The goal of this study is to present the BMA method's feasibility and potential for the use of GMPEs, without any PSHA obligation, yet. Therefore, the weights and adjustments presented in this study are for discussion of a BMA analysis and are not appropriate for direct use in a PSHA that might depend on the application.

The outline of the paper is as follows: Section 2 describes the BMA background and how the BMA model can be obtained. Section 3 reports the numerical results using nine GMPEs along with the RESORCE-2013 data set (Akkar *et al.* 2013). Finally conclusions and further ideas of research are reported in Section 4.

2 BAYESIAN MODEL AVERAGING

2.1 Principle of the BMA method

BMA (Kass & Raftery 1995; Baran & Moller 2015; Schöniger *et al.* 2015; Zhu *et al.* 2016) provides a coherent mechanism for accounting the model uncertainty. If y is the quantity of interest, such as the PGA or the PSA, then its posterior distribution given data D is

$$p(y|D) = \sum_{k=1}^K p(y|M_k, D)p(M_k|D), \quad (1)$$

where M_1, M_2, \dots, M_K are the GMPEs considered. The implicit assumption made here is that the observations D are reachable by the space formed with the K models considered. The posterior probability assigned to M_k is given by the classical Bayes' formula

$$p(M_k|D) = \frac{L(M_k|D)p(M_k)}{\sum_{j=1}^K L(M_j|D)p(M_j)}, \quad (2)$$

where $p(M_k)$ is the prior probability and $L(M_k|D) = p(D|M_k)$ is the so-called marginal likelihood of model M_k , which is a key quantity for model comparison. In line with probability theory, the standard Bayesian response to dealing with parameter uncertainty is the average. This involves averaging over parameter values with the posterior distribution of that parameter in order to get the predictive distribution. Analogously, eq. (1) shows that model uncertainty is also resolved through averaging, but this time averaging over models with the (discrete) posterior model distribution. Since the posterior probabilities $p(M_k|D)$ add up to one, they can be considered as posterior model weights, or simply BMA weights. It is worth noting that an alternative approach is given by the Akaike information criterion (AIC). Burnham & Anderson (2002) provide a Bayesian justification for AIC and suggest the use of AIC-based weights as posterior model probabilities.

The posterior mean of y , given the observation data D , is as follows:

$$E(y|D) = \sum_{k=1}^K p(M_k|D)E(y|M_k, D), \quad (3)$$

where $E(y|M_k, D)$ is the expectation of the posterior distribution of y conditioned with the data D using the model M_k alone. The associated variance $\text{Var}(y|D)$ can be split up into two distinct terms V_1 and V_2 . The first term V_1 is given by

$$V_1 = \sum_{k=1}^K p(M_k|D)\text{Var}(y|M_k, D), \quad (4)$$

and refers to variations caused by differences within individual models. The second term, which is given by

$$V_2 = \sum_{k=1}^K p(M_k|D)[E(y|M_k, D) - E(y|D)]^2, \quad (5)$$

is due to the interaction between the models. These two terms V_1 and V_2 are classically referred to as the *within-model variance* and the *between-model variance*, respectively.

In eq. (2), $p(M_k)$ is the prior probability that M_k is the true model (given that one of the models considered is true). Using experts' knowledge appears difficult, though not arbitrary, since there is no universal formula for converting knowledge into prior distributions [nonetheless, some approaches derived from experimental design theory are investigated in Runge *et al.* (2013)]. Thus, when there is little prior information about the relative plausibility of the models considered, the assumption that all models are equally likely, that is $p(M_k) = 1/K$, seems a reasonable choice, the uniform distribution being usually considered as one of the less informative ones. That choice may not appear entirely satisfactory since no prior distribution is really neutral, but we do not address that specific issue here and only illustrate the method using the most common approach. In this case, the BMA weights as given by eq. (2) are directly proportional to the marginal likelihood of models.

2.2 Models likelihood

Over the last decade, there has been a rapidly growing awareness of the importance of dealing with both model error and measurement error. In practice, however, these two terms cannot be conveniently distinguished and properly interpreted. Hence, following Park & Grandhi (2014) and Brynjarsdottir & O'Hagan (2014), we adopt here a formulation that involves a single error term through the

additive constant ϵ_k that summarizes the model-observation discrepancy and that represents each GMPE by

$$y(x) = f_k(x) + \epsilon_k, \quad (6)$$

where y is the logarithm of the 5 per cent damped horizontal PSA at a given frequency, f_k is the deterministic prediction of the GMPE M_k , and x is a set of earthquake characteristics used as inputs in the GMPEs (moment magnitude, distance, focal depth, V_{S30} , fault type, etc.).

The use of the additive term in eq. (6) for modelling the prediction error immediately follows from the general structure of GMPEs. In particular, each GMPE is designed using a log-normal error term to describe the PSA. As a result, it appears natural to assume that ϵ_k follows a normal distribution, that is, $\epsilon_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, where μ_k and σ_k are the expectation and the standard deviation, respectively. This assumption is further discussed and justified in Section 3. The choice of statistical distributions for μ_k and σ_k is somewhat arbitrary in the absence of detailed knowledge about ϵ_k . Here we use classical non-informative priors and assume that $\mu_k \sim \mathcal{U}(\mu_a, \mu_b)$ and $\sigma_k \sim \mathcal{U}(\sigma_a, \sigma_b)$. The four parameters $\sigma_a, \sigma_b, \mu_a$ and μ_b govern the domain over which appreciable values of the likelihood function occur and thus they have to be fixed so as to avoid introducing biases in $L(M_k|D)$.

Following the Bayesian formalism, the marginal likelihood of model M_k is computed as the integral of the likelihood function $l(M_k, \mu_k, \sigma_k|D)$ over the hyperparameters: hyperparameters are *marginalized* by the integration; hence

$$L(M_k|D) = \int_{\mu_a}^{\mu_b} \int_{\sigma_a}^{\sigma_b} l(M_k, \mu_k, \sigma_k|D) p(\mu_k) p(\sigma_k) d\mu_k d\sigma_k, \quad (7)$$

where $p(\mu_k)$ and $p(\sigma_k)$ are the prior distributions of the hyperparameters. Assuming that D consists of N components d_1, d_2, \dots, d_N , we obtain

$$l(M_k, \mu_k, \sigma_k|D) = \prod_{n=1}^N g(d_n|M_k, \mu_k, \sigma_k), \quad (8)$$

where $g(d_n|M_k, \mu_k, \sigma_k)$ is the prediction PDF for y using the GMPE M_k and the hyperparameters μ_k and σ_k . In eq. (8), the independence of observations d_1, d_2, \dots, d_N is assumed, meaning that they result from N independent events. While for the sake of simplicity here, we use the independence assumption, one should bear in mind that multiple detections from a single event are not necessarily independent. Using the assumption $\epsilon_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, we obtain

$$l(M_k, \mu_k, \sigma_k|D) = \prod_{n=1}^N \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(d_n - f_k(x_n) - \mu_k)^2}{2\sigma_k^2}\right), \quad (9)$$

where x_n is the set of model inputs corresponding to the n th seismic record. Clearly, eq. (7) shows that priors matter for the computation of the likelihood function. It should be emphasized that it is preferable to be explicit about the prior assumptions. Recent researches in prior sensitivity (Fernandez *et al.* 2001) serve to highlight which aspects of the prior are particularly critical for the results and how we can make our prior choices more robust.

BMA as described above is the formal probabilistic way of obtaining predictive inference, and is, more generally, the approach to any inference problem involving quantities of interest that are not model-specific. The predictive in eq. (1) involves averaging at two levels: over (continuous) parameter values, given each possible model, and discrete averaging over all possible models. Once the likelihood function is estimated, the posterior PDF of y can be

Table 1. Periods for which the 5 per cent damped horizontal PSA is used for the BMA calibration.

	1	2	3	4	5	6	7	8	9	10
Period [s]	0.02	0.05	0.1	0.15	0.2	0.3	0.5	1	1.5	2

obtained by integration over μ_k and σ_k :

$$p(y|M_k, D) = \int_{\mu_a}^{\mu_b} \int_{\sigma_a}^{\sigma_b} g(y|M_k, \mu_k, \sigma_k) p(\mu_k, \sigma_k|D, M_k) \times d\mu_k d\sigma_k, \quad (10)$$

where the posterior PDF of (μ_k, σ_k) given M_k is computed as

$$p(\mu_k, \sigma_k|D, M_k) = \frac{l(M_k, \mu_k, \sigma_k|D) p(\mu_k) p(\sigma_k)}{L(M_k|D)}. \quad (11)$$

Generally, the main computational challenge is constituted by the typically large parameter space in eq. (10). There are many methods and algorithms to address this problem (Hoeting *et al.* 1999; Vrugt *et al.* 2008; Edeling *et al.* 2014; Park & Grandhi 2014; Chitsazan & Tsai 2015). We briefly present two different strategies in the next section: the maximum likelihood estimation (MLE) method and the Markov chain Monte Carlo (MCMC) method.

2.3 Estimation methods

Here, and in the following, $\theta_k \in \Theta_k$ is the vector of parameters of model M_k , for example, $\theta_k = (\mu_k, \sigma_k)$ or $\theta_k = \epsilon_k$. Since θ_k groups the (unknown) parameters, it can also contain some GMPE's coefficients as additional random parameters. If all parameters are jointly uniformly distributed then the prior distribution of these parameters is constant over the entire space Θ_k and the Bayes theorem can be written as

$$p(\theta_k|D, M_k) = \frac{p(D|\theta_k, M_k) p(\theta_k|M_k)}{p(D|M_k)} = c p(D|\theta_k, M_k), \quad (12)$$

where c is a constant. The marginal likelihood is then given by

$$p(D|M_k) = \int_{\Theta_k} p(D|\theta_k, M_k) p(\theta_k|M_k) d\theta_k, \quad (13)$$

where we use the conventional representation of weak prior information through the density $p(\theta_k|M_k) \propto 1$. We can think of calibration of θ_k as a preliminary to computing the BMA weights $p(M_k|D)$, following rule (2).

The main computational issues concern the estimate of eq. (13) and the fact that we need a large number of model evaluations for each realization of θ_k . For peak shape likelihood functions, however, an excellent approximation of eq. (13) is given by the MLE and, indeed, the problem can readily be solved analytically for the GMPEs, provided that we impose a normal distribution for ϵ_k . The details of calculations are given in the Appendix. When the posterior is an intractable expression, a fairly general approach is to use an MCMC method (Metropolis *et al.* 1953; Hastings 1970), which constructs a Markov chain whose stationary distribution is the posterior distribution. A simple MCMC implementation uses the Metropolis algorithm that, for a current realization θ^m , generates θ^* from a symmetric distribution, then computes the Metropolis acceptance probability

$$\alpha = \min\left\{1, \frac{l(M_k, \theta_k^*|D)}{l(M_k, \theta_k^m|D)}\right\}, \quad (14)$$

and finally sets $\theta^{m+1} = \theta^*$ with probability α and $\theta^{m+1} = \theta^m$ with probability $1 - \alpha$. The width of the proposal distribution (e.g.

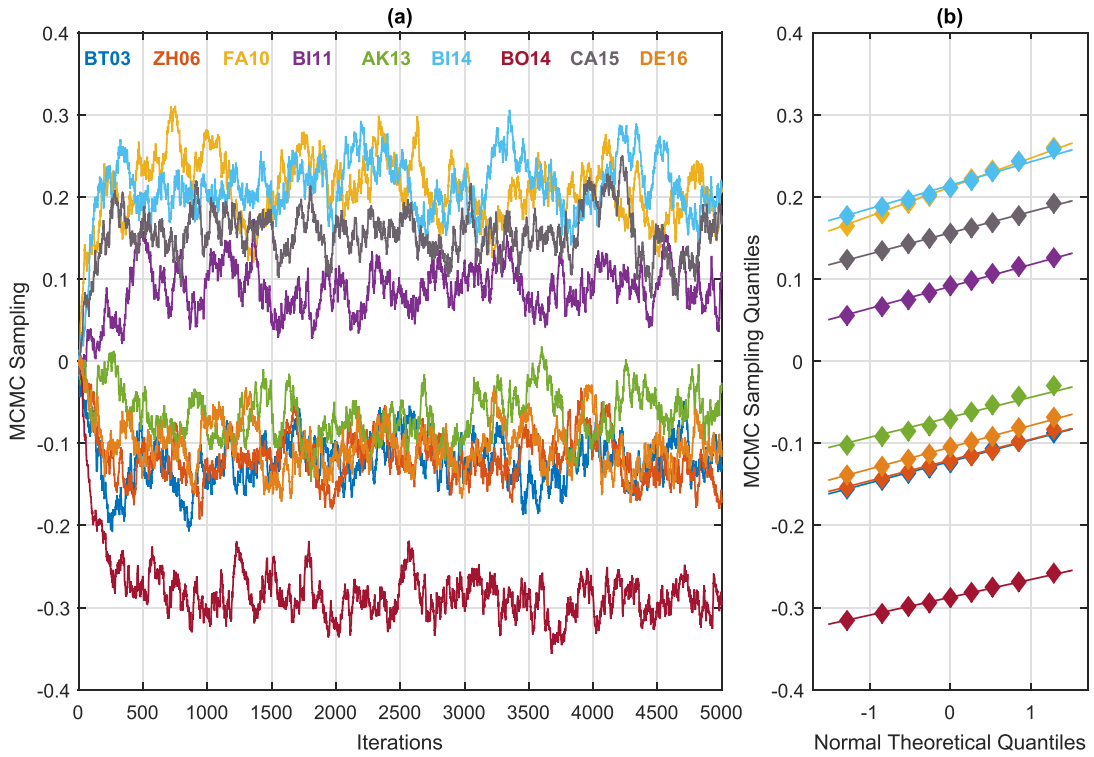


Figure 1. (a) Metropolis–Hastings sampling of ϵ_k for $T = 1$ s. (b) Empirical deciles versus theoretical deciles of the normal distribution. The models proposed by Berge-Thierry *et al.* (2003), Zhao *et al.* (2006), Faccioli *et al.* (2010), Bindi *et al.* (2011), Akkar *et al.* (2014), Bindi *et al.* (2014), Boore *et al.* (2014), Cauzzi *et al.* (2015) and Derras *et al.* (2016) are designated by BT03, ZH06, FA10, BI11, AK13, BI14, BO14, CA15 and DE16, respectively.

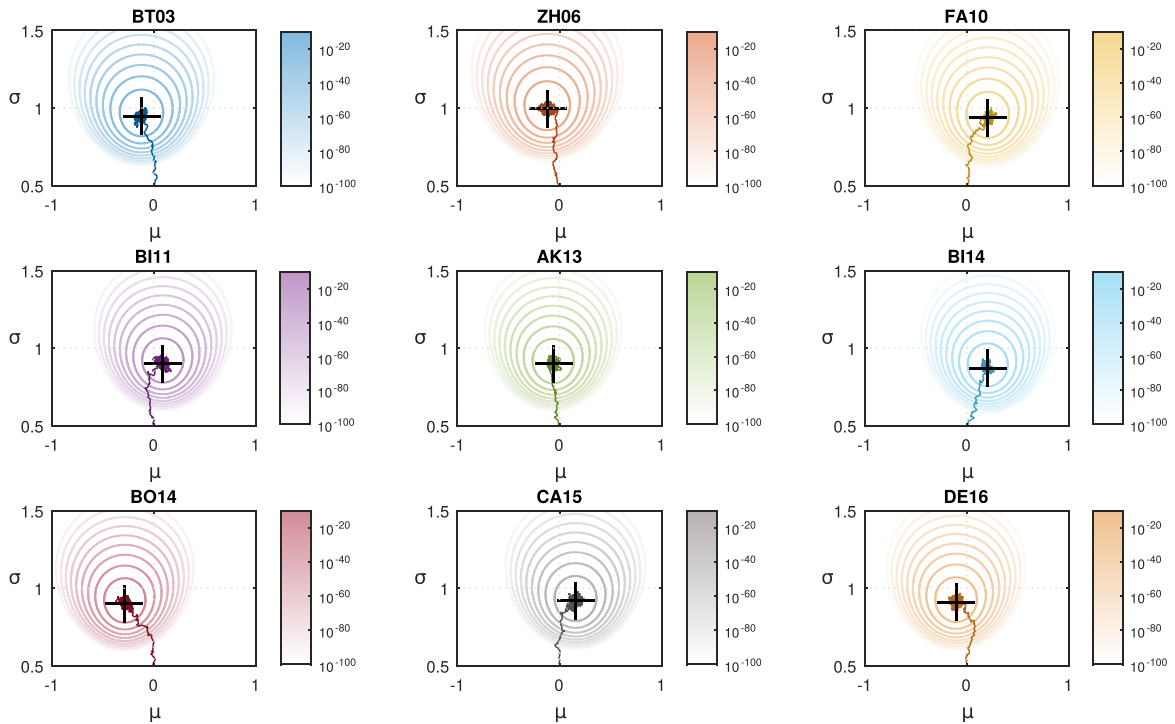


Figure 2. Equal log-likelihood contours of μ_k and σ_k for $T = 1$ s. The dark lines represent the MCMC Markov chains starting from the point $\mu_k = 0$ and $\sigma_k = 0.5$ and the black crosses indicate the maxima as obtained using the MLE method. Each panel corresponds to a GMPE (see the caption of Fig. 1).

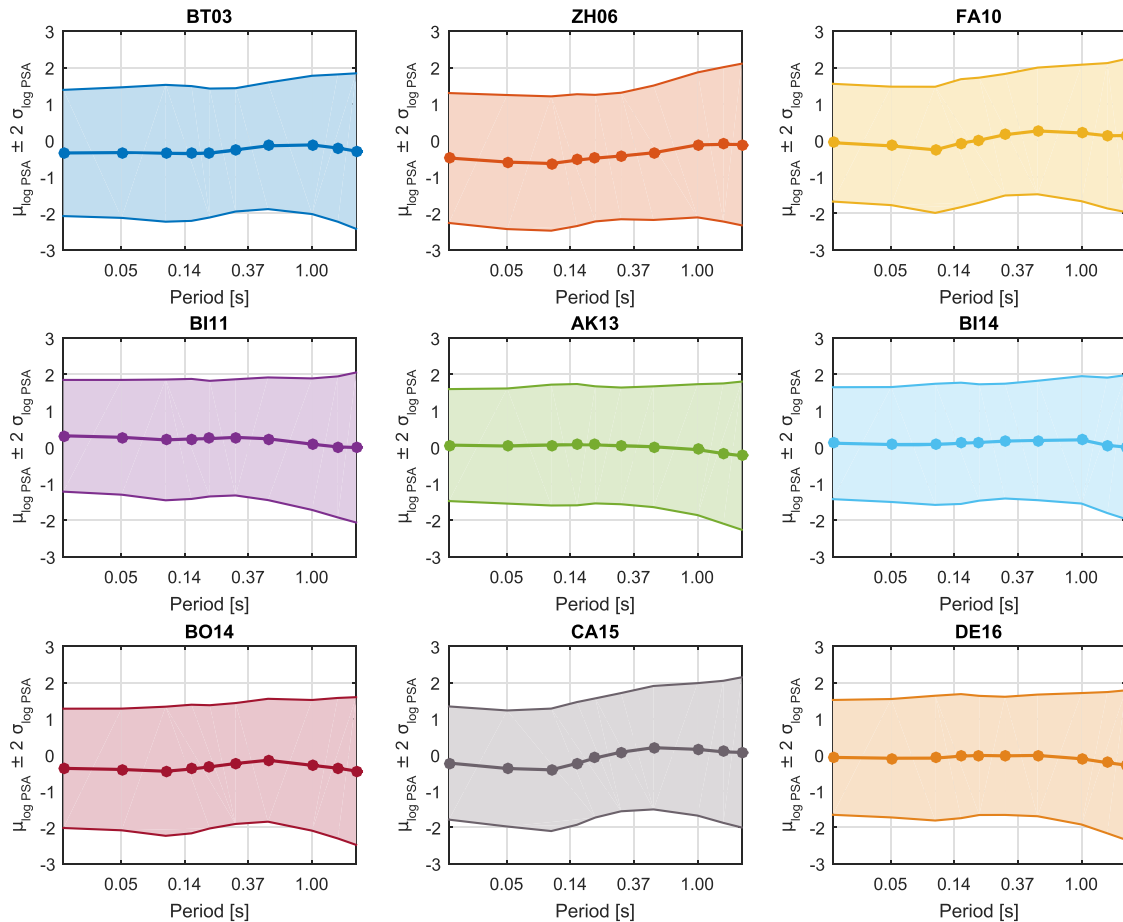


Figure 3. Optimal values μ_k^* and 95 per cent confident interval $\mu_k^* \pm 2\sigma_k^*$ as function of the period (see Table 1). The horizontal PSA is computed in units of g . Each panel corresponds to a GMPE (see the caption of Fig. 1).

the standard deviation for a normal distribution) for computing θ^* is a tuning parameter and clearly a limitation of the standard sampling algorithm, although there are various methods that remedy the problem [see Roberts & Rosenthal (2009) for an overview of ‘auto-tuning’ algorithms].

The multiple central processing units (CPUs) of recent computers can be used to speed up computation through multiple independent MCMC chains. Markov chains construct a progressive picture of the target distribution, proceeding by local exploration of the state space Θ_k until all the regions of interest, meaning areas of maximum likelihood, have been covered. Hence, eq. (13) can be estimated with a standard Monte Carlo estimate. Further, in this study, we diagnose the successful convergence and the proper mixing of the Markov chains through computing the Gelman–Rubin’s potential scale reduction factor (Gelman & Rubin 1992; Cowles & Carlin 1996).

3 APPLICATION

3.1 The selection of GMPEs

The selection of GMPEs, which is one of the fundamental modelling tasks in SHA studies (either probabilistic or deterministic), may impact strongly the SHA results in terms of level and dispersion predictions. In a first stage, we consider GMPEs obtained from local,

regional and worldwide data that satisfy the *a priori* requirements for shallow crustal tectonic settings (Cotton *et al.* 2006; Bommer *et al.* 2010). A parsimony principle is adopted to avoid adding more uncertainties in the estimation of unknown parameters required in complex models (due to the lack of information in the European database RESORCE). Some of the models considered here have been previously analysed by Delavaud *et al.* (2012) in the Seismic Hazard Harmonization in Europe (SHARE) project, using the likelihood-based ranking method (Scherbaum 2009), and selected as the candidate GMPEs for seismic hazard in Europe. Using the same method, Beauval *et al.* (2012) investigated GMPEs with the French accelerometric database (RAP). Instead, in Global Earthquakes Model (GEM) a Euclidean-distance-based ranking method (Kale & Akkar 2012) has been proposed and applied to the Middle East. These works suggest that GMPEs introduced by Berge-Thierry *et al.* (2003) and further developed by Zhao *et al.* (2006), Cauzzi & Faccioli (2008), Akkar & Bommer (2010) and Bindi *et al.* (2010) are appropriate for Europe. In this list, the simplest model (Berge-Thierry *et al.* 2003), originally developed for the French nuclear safety regulation, involves the surface wave magnitude to describe the size of earthquakes, which implies to use a formula for converting it into the largely used moment magnitude. We choose to use (Scordilis 2006) conversion relations that are published to help building homogenous global seismic catalogues.

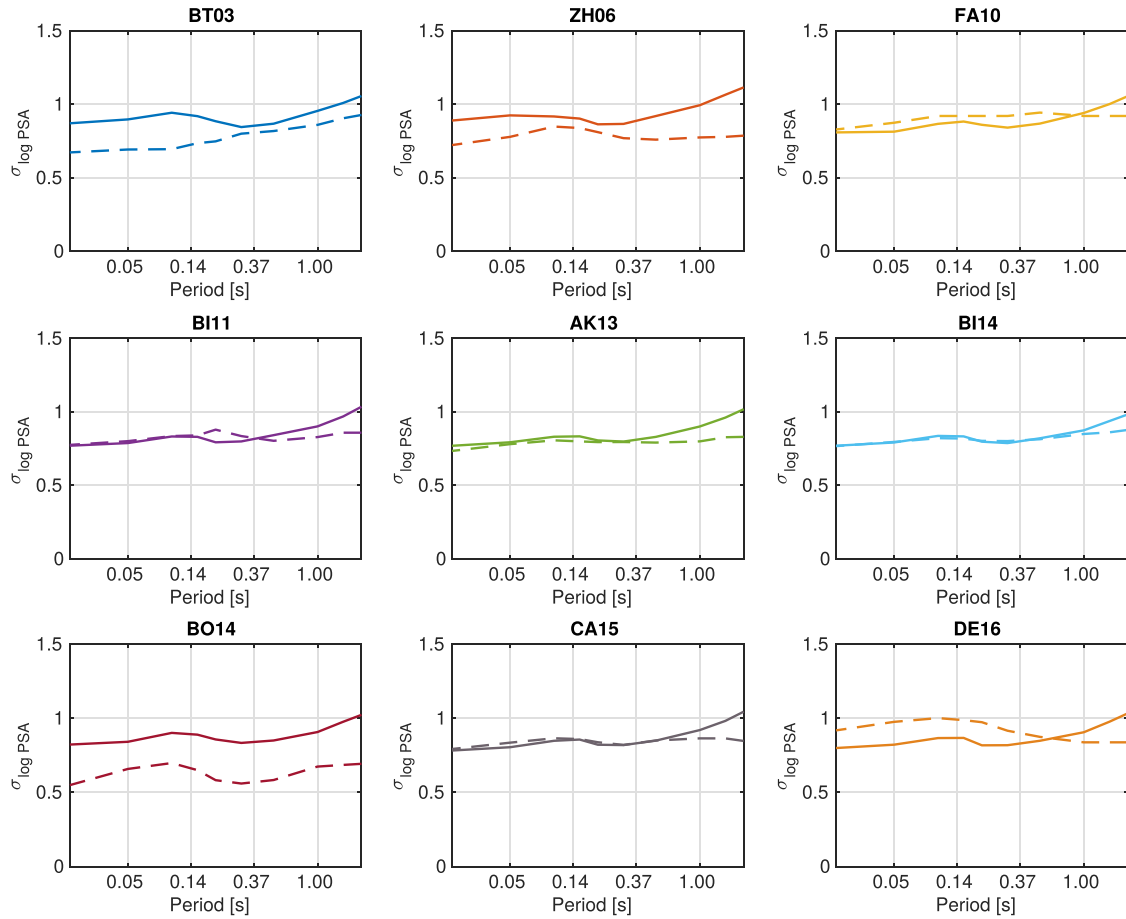


Figure 4. Optimal standard deviations σ_k^* obtained with the MLE method (solid lines) and original values obtained from the initial GMPE estimation on their own data set (dashed lines). The horizontal PSA is computed in units of g . Each panel corresponds to a GMPE (see the caption of Fig. 1).

In a second stage, we take into account models that are often used as standards in SHA. These include the U.S. Next Generation Attenuation (NGA) models that were developed for active regions from strong ground motions. According to the comparisons of the NGA ground-motion relations (Abrahamson *et al.* 2008) and their applicability in the Euro-Mediterranean region (Stafford *et al.* 2008), the model proposed by Boore & Atkinson (2007) is chosen here as an example of NGA model, along with the model proposed by Derras *et al.* (2016). These models are chosen following a parsimony principle (they use a limited number of parameters compared to the other ones). In fact, since the NGA models may really be different in some ranges, the selection of a single one would be difficult.

Finally, to investigate the performance of several selected GMPEs in face with the earthquakes from Europe and the Middle East of the RESORCE database, two models developed using the database itself by Akkar *et al.* (2014) and Bindi *et al.* (2014) are considered. We therefore end up with nine GMPEs for which, when updates of the models are available, we use the most recent version (Berge-Thierry *et al.* 2003; Zhao *et al.* 2006; Faccioli *et al.* 2010; Bindi *et al.* 2011, 2014; Akkar *et al.* 2014; Boore *et al.* 2014; Cauzzi *et al.* 2015; Derras *et al.* 2016). The BMA method requires a large amount of data, thus we choose to ignore regional differences by excluding non-ergodic models (Kotha *et al.* 2016; Landwehr *et al.* 2016; Sedaghati & Pezeshk 2017) that could be difficult to interpret. The set of GMPEs considered in this paper is not supposed to be

exhaustive neither fitted for an SHA study, since the main purpose is to test and demonstrate the performance of the BMA method. For a site-specific PSHA study, the regional differences should be taken into account, for instance, using non-ergodic models and in this case, there is no guarantee that the BMA would be the best approach. In fact, it would mainly depend on the amount of regional data available.

3.2 The RESORCE database

In order to avoid solving our problem on the basis of a small number of observations, we use a recent version of the RESORCE database, RESORCE-2013 (Akkar *et al.* 2013), which contains ground-motion data from pan-European earthquakes. This freely accessible database gathers 5882 seismic records from 1814 events, and the corresponding seismological and geotechnical parameters.

According to the selected GMPEs, the relevant input variables represented by x in eq. (6) are the distance (epicentral distance, hypocentral distance and the Joyner–Boore distance), the V_{S30} velocity, the focal depth, the fault mechanism and the event magnitude. Hence, to make sure that the data lie in the common application domain covered by the selected GMPEs, we only retain the elements for which the moment magnitude, the distance and the V_{S30} velocity are in the intervals [5; 7.3], [4; 150] (km) and [300; 1200] ($m s^{-1}$), respectively. Regarding the diversity of magnitude measures in the database, and given that the most commonly used measure is the

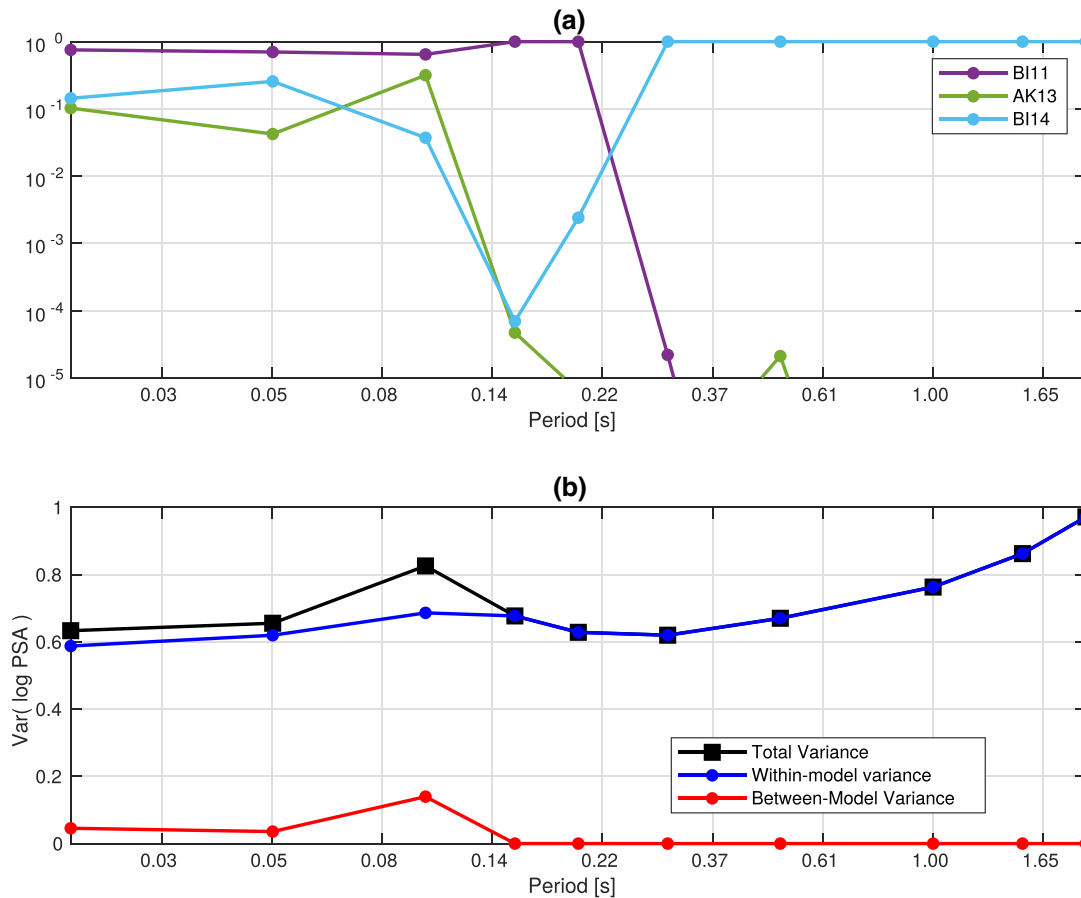


Figure 5. (a) Posterior BMA weights $P(M_k|D)$ as functions of the period. Only the three dominant models are represented. (b) Total variance, *within-model* variance and *between-model* variance of the BMA combination. The horizontal PSA is computed with GMPEs in units of g .

moment magnitude M_w (in 79 per cent of the records), we restrict our approach to events for which the magnitude M_w is provided. Similarly, we only select events for which the fault mechanism is indicated as *Normal*, *Strike-Slip* and *Reverse*. These restrictions result in a subset of $N = 939$ records, which ensures the convergence of our statistical study.

As outlined above, the GMPE's output is the logarithm of the horizontal 5 per cent damped pseudospectral acceleration [y in eq. (6)]. This variable is expressed in multiples of g , where g is the acceleration due to Earth's gravity, and computed for 10 periods common to the 9 GMPEs (see Table 1), using the exact coefficient values provided in the literature. The corresponding observation is obtained using the geometric mean of the two horizontal PSA components provided in the RESORCE database.

3.3 Application of the BMA approach

According to most Bayesian applications, we try to make as few assumptions as possible in order to ensure the analysis is essentially data driven. Even though the normal distribution assumption for ϵ_k sounds reasonable, the MCMC Metropolis–Hastings algorithm provides a way to test this hypothesis against observations. A classical way of doing this is to compute the posterior distribution $p(\epsilon_k|D, M_k)$ for each GMPE, and to assign a uniform prior for ϵ_k over the interval $[-2, 2]$. The resulting MCMC sampling is given in Fig. 1(a) for 5000 iterations (including a warm-up period of 500 iterations

for initialization of the algorithm), a standard deviation of 0.005 for the Metropolis algorithm, and using $\epsilon_k = 0$ as the initial guess for the MCMC. For the sake of simplicity, only the results obtained for the period $T = 1$ s are given here. Other periods exhibit similar behaviours. Fig. 1(b) shows a quantile–quantile plot (Jobson 1991) that enlightens the nature of the posterior distribution. The linear relation between the empirical deciles of the Markov chains and the theoretical deciles of the normal distribution is a validation of the Gaussian assumption. This result is consistent with our assumption that the prediction error ϵ_k in eq. (6) can be modelled by a normal distribution, that is, $\epsilon_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$.

The MLE estimation of optimal hyperparameters μ_k^* and σ_k^* can be computed from eqs (A2) and (A3) of the Appendix and compared to the MCMC-based approach. Fig. 2 shows the results for $T = 1$ s, using $\mu_k = 0$ and $\sigma_k = 0.5$ as starting points, and a standard deviation of 0.01 for the Metropolis algorithm. The likelihood shape appears strongly peaked in the hyperparameter space (logarithmic contours in Fig. 2) and therefore the MLE approximation seems to be a reasonable choice. The calibration results for the whole range of periods (Table 1) are summarized in Fig. 3. While the priors for μ_k and σ_k are uniform distributions over the intervals $[-1, 1]$ and $[0.5, 5]$, respectively, it is important to point out the fact that the final distributions are not centred (i.e. $\mu_k^* \neq 0$). The final distribution should be interpreted as a measure of the statistical discrepancy between the GMPE's predictions and the training database. Also, as discussed in Section 3.1 two GMPEs (Akkar

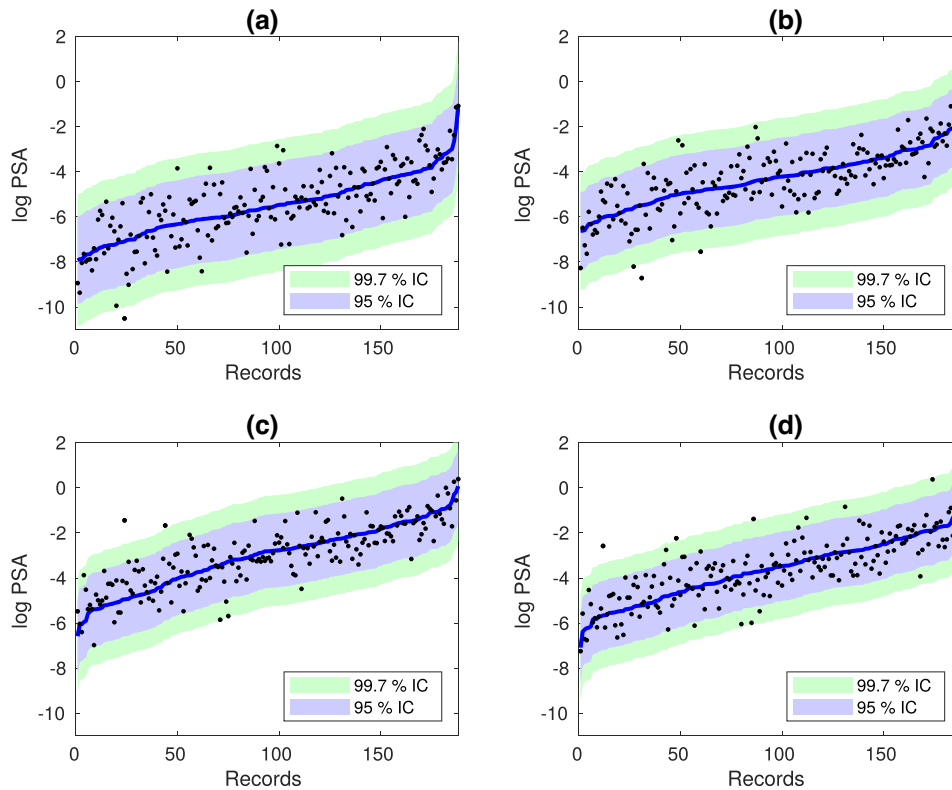


Figure 6. 99.7 per cent and 95 per cent confidence intervals of BMA predictions (light green and light blue areas, respectively) compared to 200 events that were not used for the BMA calibration (black dots). Events are ordered to have an increasing BMA expectation (blue line). The horizontal PSA is computed with GMPEs in *g*. Results are shown for periods (a) 2 s, (b) 1 s, (c) 0.15 s and (d) 0.05 s.

et al. 2014; Bindi *et al.* 2014) were designed from the RESORCE database, which means that the statistical discrepancy, as computed here with a subset of RESORCE-2013, should be close to that obtained with the database originally considered. This is illustrated in Fig. 4, where the corresponding standard deviations are plotted as functions of the period and compared with the original values obtained from the initial GMPE estimation on their own data set. The overall agreement is deemed very good, especially for the GMPEs proposed by Akkar *et al.* (2014) and Bindi *et al.* (2014). For large periods, we observe differences that might be the result of differences between the selected subsets. This lack of data robustness is not surprising, and there is evidence in the literature (Feldkircher 2012) that this effect can be mitigated through adding prior information.

Fig. 5 shows the BMA weights, as well as the within-model and between-model variances. While the models proposed by Bindi *et al.* (2011), Akkar *et al.* (2014) and Bindi *et al.* (2014) dominate over the set of periods considered here (Fig. 5a), a variance-based examination of results (Fig. 5b) shows two distinct regions. For large periods (low frequencies), the variance is only due to the within-model variance, which means that the BMA predictive PDF is dominated by the model having the highest BMA weight (Bindi *et al.* 2014). For lower periods (high frequencies), a low between-model variance is observed, reflecting the fact that three models among the nine considered contribute to the BMA PDF. Fig. 5(a) shows that most of other models have low weights throughout, suggesting that they are less useful relative to the three dominant models. However, this does not mean that the corresponding GMPEs are not suited to make predictions on European seismicity. The weights reflect the models' overall performance over the training

data set, relative to the other models. It is not surprising, then, that the GMPEs established with the RESORCE database have higher BMA weights.

Fig. 6 illustrates the way in which BMA can yield reasonable confidence intervals. The BMA PDF is computed using 739 records that are randomly selected among those of the initial RESORCE-2013 training data set (with 939 observations). Then the predictions of the BMA PDF are compared to the 200 remaining observations. On average, 94.4 per cent of the observations lay within the 95 per cent confidence interval. An alternative way of dealing with the problem of prediction capability is to resort to the so-called *leave-one-out* procedure (Geisser 1993), which uses a sample of observations that is not itself used to calibrate the model. The measure of the fit of the resulting model to the sample of observations can be summarized through the Predicted Residual Error Sum of Squares (PRESS) statistic:

$$\text{PRESS} = \frac{1}{N} \sum_{n=1}^N [E(y_{-n}(x_n)) - d_n]^2, \quad (15)$$

where d_n is the n th observation and $y_{-n}(x_n)$ is the BMA prediction based on the calibrated GMPEs using the whole data set except d_n . Given this procedure, the PRESS statistic can be calculated for the nine GMPEs for the same data set, with the lowest values of PRESS indicating the best models. Figs 7(a) and (b) show the GMPE's performance before the Bayesian calibration of prediction error and after being calibrated with observations. Clearly, there are substantial gains in using both Bayesian calibration and the BMA approach for improving the predictive overall performance of GMPEs. Further, in addition to the calibration and the ranking of the

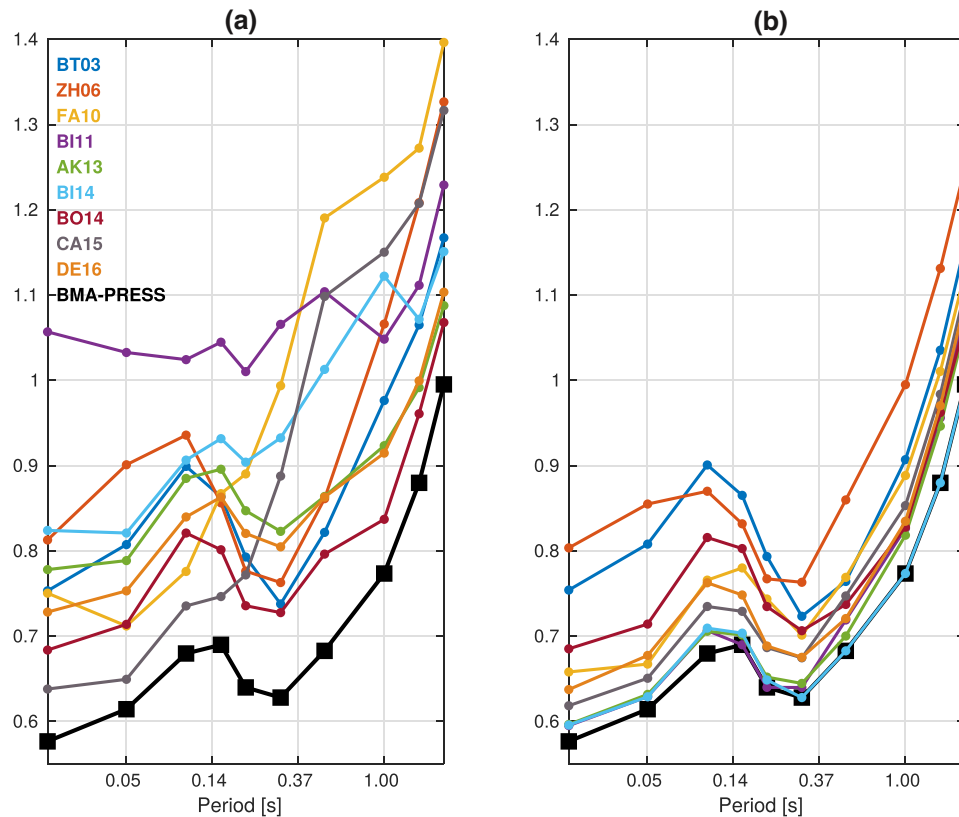


Figure 7. Quadratic mean of prediction error. (a) Prior predictions; (b) posterior predictions. The black line corresponds to the predicted residual error sum of squares (PRESS) of the BMA combination of calibrated GMPEs (*leave-one-out* strategy).

GMPEs, the BMA approach offers a way to combine the calibrated models together in order to get much better performances for low periods, as shown in Fig. 7(b).

4 CONCLUSION

In this paper, we have proposed a new method for statistical post-processing GMPEs' output to produce calibrated and predictive PDFs. Calibrating the bias along with the standard deviation of the log-normal prediction error allows a significant improvement of GMPEs' performance. The BMA method offers then a way to rank, select and combine together the calibrated GMPEs in a weighted sum requiring neither additional hypothesis nor experts' judgement. In our case study, based on a subset of the RESORCE-2013 database, the BMA has a lower PRESS statistic than any of the individual GMPEs. Results suggest the use of a single model (Bindi *et al.* 2014) for low frequencies and the combination of three GMPEs (Bindi *et al.* 2011, 2014; Akkar *et al.* 2014) for high frequencies. The goal of this study is to present the BMA method's feasibility and potential for the use of multiple GMPEs, without any PSHA obligation, yet. Therefore, results presented here can provide useful information but should be taken cautiously. As the BMA weights and adjustment deeply depend on the observations and on the model set considered, a BMA analysis requires a careful and relevant discussion, depending on the application, to lead to good predictions in a PSHA.

As indicated in Section 2.1, the BMA procedure is based on the implicit assumption that the whole set of observations is reachable

by the space formed with the models considered. This is a strong assumption that leads to a well-known limitation of BMA applications (Clarke 2018) that are best suited to M-closed problems, which is indeed not the case with a limited set of GMPEs. This results in the overconfidence of the method that tends to present weights close to zero or one. However, despite the limitations of the model set used, our application shows a performing combination of three GMPEs for high frequencies. This is an encouraging result that leaves room for improvement. Clearly, adding more distinct GMPEs to the study would ensure a larger and more performant BMA combination. However, adding too many models in a single probabilistic framework may discard data from the common observational set considered, which may produce much lower confidence intervals. Indeed, the range of metadata (magnitude, distance, V_{S30} , etc.) of the training data set has to be available and compatible with the domains on which every GMPEs were established.

While much of the emphasis in this paper was to show that the Bayesian learning theory approach can be applied to the task of selection and combination of GMPEs, other BMA-inspired techniques could also be proposed to achieve much better results. The BMA itself could be modified to integrate over combinations of models rather than over individual learners. Another approach could be to learn the optimal model combination, given a fixed set of models. The model space could be enriched by extending the calibration to GMPEs' coefficients and combining several functional forms of GMPEs. The GMPEs' coefficients are usually estimated with standard regression techniques, and recent works have focused on estimating them with a Bayesian inference using the MCMC

method (Kowsari *et al.* 2019). The resulting model, that can be seen as the Bayesian theory transduction of the backbone approach (Douglas 2018), should be updated each time new ground-motion observations are available (Stafford 2019). Alternatively, strategies could be developed to allow calibrations and combinations to be inferred simultaneously.

ACKNOWLEDGEMENTS

The work carried out under the SINAPS@ project receives French funding managed by the French National Research Agency under the program 'Future Investments' (SINAPS@ reference No. ANR-11-RSNR-0022). SINAPS@ is a SEISM Institute project (<http://www.institut-seism.fr/en/projects/sinaps/>). Readers can access the RESORCE-2013 database used in this study on the freely accessible platform <http://www.resorce-portal.eu/> (DOI: 10.1007/s10518-013-9506-8).

REFERENCES

- Abrahamson, N. *et al.*, 2008. Comparison of the NGA Ground Motion Relations, *Earthq. Spectra*, **24**, 45–66.
- Akkar, S. & Bommer, J.J., 2010. Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East, *Seismol. Res. Lett.*, **81**(2), 195–206.
- Akkar, S. *et al.*, 2013. Reference database for seismic ground-motion in Europe (RESORCE), *Bull. Earthq. Eng.*, **12**(1), 311–339.
- Akkar, S., Sandikkaya, M.A. & Bommer, J.J., 2014. Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in Europe and the Middle East, *Bull. Earthq. Eng.*, **12**(1), 359–387.
- Arroyo, D., Ordaz, M. & Rueda, R., 2014. On the selection of ground-motion prediction equations for probabilistic seismic-hazard analysis, *Bull. seism. Soc. Am.*, **104**(4), 1860–1875.
- Atkinson, G.M. & Adams, J., 2013. Ground motion prediction equations for application to the 2015 Canadian national seismic hazard maps, *Can. J. Civ. Eng.*, **40**(10), 988–998.
- Baran, S. & Moller, A., 2015. Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging, *Environmetrics*, **26**(2), 120–132.
- Beauval, C., Tasan, H., Laurendeau, A., Delavaud, E., Cotton, F., Guéguen, P. & Kuehn, N., 2012. On the testing of ground-motion prediction equations against small-magnitude data, *Bull. seism. Soc. Am.*, **102**(5), 1994–2007.
- Berge-Thierry, C., Cotton, F., Scotti, O., Griot-Pommeroy, D.-A. & Fukushima, Y., 2003. New empirical response spectral attenuation laws for moderate European earthquakes, *J. Earthq. Eng.*, **7**(2), 193–222.
- Bindi, D., Luzi, L., Massa, M. & Pacor, F., 2010. Horizontal and vertical ground motion prediction equations derived from the Italian Accelerometric Archive (ITACA), *Bull. Earthq. Eng.*, **8**(5), 1209–1230.
- Bindi, D., Pacor, F., Luzi, L., Puglia, R., Massa, M., Ameri, G. & Paolucci, R., 2011. Ground motion prediction equations derived from the Italian strong motion database, *Bull. Earthq. Eng.*, **9**(6), 1899–1920.
- Bindi, D., Massa, M., Luzi, L., Ameri, G., Pacor, F., Puglia, R. & Augliera, P., 2014. Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0s using the RESORCE dataset, *Bull. Earthq. Eng.*, **12**(1), 391–430.
- Bommer, J.J., 2012. Challenges of building logic trees for probabilistic seismic hazard analysis, *Earthq. Spectra*, **28**(4), 1723–1735.
- Bommer, J.J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H. & Fäh, D., 2010. On the selection of ground-motion prediction equations for seismic hazard analysis, *Seismol. Res. Lett.*, **81**(5), 783–793.
- Bommer, J.J. *et al.*, 2015. A SSHAC level 3 probabilistic seismic hazard analysis for a new-build nuclear site in South Africa, *Earthq. Spectra*, **31**(2), 661–698.
- Boore, D.M. & Atkinson, G.M., 2007. NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters, PEER Report 2007/01, Pacific Earthquake Engineering Research Center.
- Boore, D.M., Stewart, J.P., Seyhan, E. & Atkinson, G.M., 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra*, **30**(3), 1057–1085.
- Brynjarsdottir, J. & O'Hagan, A., 2014. Learning about physical parameters: the importance of model discrepancy, *Inverse Probl.*, **30**(11), 114007.
- Budnitz, R.J., Apostolakis, G., Boore, D.M., Cluff, L.S., Coppersmith, K.J., Cornell, C.A. & Morris, P.A., 1997. Recommendations for probabilistic seismic hazard analysis: guidance on uncertainty and use of experts, United States Nuclear Regulatory Commission, NUREG/CR-6372.
- Burnham, K. & Anderson, D., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edn, Springer.
- Cauzzi, C. & Faccioli, E., 2008. Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records, *J. Seismol.*, **12**(4), 453–475.
- Cauzzi, C., Faccioli, E., Vanini, M. & Bianchini, A., 2015. Updated predictive equations for broadband (0.01–10 s) horizontal response spectra and peak ground motions, based on a global dataset of digital acceleration records, *Bull. Earthq. Eng.*, **13**(6), 1587–1612.
- Chitsazan, N. & Tsai, F.T.-C., 2015. A hierarchical Bayesian model averaging framework for groundwater prediction under uncertainty, *Groundwater*, **53**(2), 305–316.
- Clarke, B.S. & Clarke, J., 2018. *Predictive Statistics: Analysis and Inference beyond Models*, ch. 11.5.3, Cambridge Univ. Press.
- Cornell, A.C., 1968. Engineering seismic risk analysis, *Bull. seism. Soc. Am.*, **58**(5), 1583–1606.
- Cotton, F., Scherbaum, F., Bommer, J.J. & Bungum, H., 2006. Criteria for selecting and adjusting ground-motion models for specific target regions: application to central Europe and rock sites, *J. Seismol.*, **10**(2), 137–156.
- Cowles, M.K. & Carlin, B.P., 1996. Markov chain Monte Carlo convergence diagnostics: a comparative review, *J. Am. Stat. Assoc.*, **91**(434), 883–904.
- Culka, M., 2014. Applying Bayesian model averaging for uncertainty estimation of input data in energy modelling, *Energy Sustainability Soc.*, **4**(1), 1–17.
- Delavaud, E. *et al.*, 2012. Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe, *J. Seismol.*, **16**(3), 451–473.
- Derras, B., Bard, P.-Y. & Cotton, F., 2016. Site-condition proxies, ground motion variability, and data-driven GMPEs: insights from the NGA-West2 and RESORCE data sets, *Earthq. Spectra*, **32**(4), 2027–2056.
- Douglas, J., 2018. Calibrating the backbone approach for the development of earthquake ground motion models, in *Best Practice in Physics-based Fault Rupture Models for Seismic Hazard Assessment of Nuclear Installations: Issues and Challenges Towards Full Seismic Risk Analysis*, Cadarache, France, May 14–16.
- Edeling, W.N., Cinnela, P. & Dwight, R.P., 2014. Predictive RANS simulation via Bayesian model-scenario averaging, *J. Comput. Phys.*, **275**, 65–91.
- Faccioli, E., Bianchini, A. & Villani, M., 2010. New ground motion prediction equations for $t > 1$ s and their influence on seismic hazard assessment, in *Proceedings of the University of Tokyo Symposium on Long-Period Ground Motion and Urban Disaster Mitigation*, Tokyo, Japan, March 17–18.
- Feldkircher, M. & Zeugner, S., 2012. The impact of data revisions on the robustness of growth determinants: a note on 'determinants of economic growth. Will data tell?', *J. Appl. Econometrics*, **27**, 686–694.
- Fernández, C., Ley, E. & Steel, M., 2001. Benchmark priors for Bayesian model averaging, *J. Econometrics*, **100**, 381–427.
- Geisser, S., 1993. *Predictive Inference: An Introduction*, Chapman and Hall.
- Gelman, A. & Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences, *Stat. Sci.*, **7**, 457–511.
- Haendel, A., Specht, S., Kuehn, N.M. & Scherbaum, F., 2014. Mixture of ground-motion prediction equation as backbone models for a logic tree: an application to the subduction zone in Northern Chile, *Bull. Earthq. Eng.*, **13**, 483–501.

- Hastings, W., 1970. Monte Carlo sampling methods using Markov chains and their application, *Biometrika*, **57**, 97–109.
- Hoeting, J.A., Madigan, D., Raftery, A.E. & Volinsky, C.T., 1999. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors), *Stat. Sci.*, **14**(4), 382–417.
- Jobson, J.D., 1991. *Applied Multivariate Data Analysis*, (p. 63), Springer Texts in Statistics, ISBN 978-0-387-97660-0, doi:10.1002/9781118887486.
- Kale, Ö. & Akkar, S., 2012. A method to determine the appropriate GMPEs for a selected seismic prone region, in *15th World Conf. Earthquake Engineering*, Lisbon, Portugal.
- Kass, R.E. & Raftery, A.E., 1995. Bayes factors, *J. Am. Stat. Assoc.*, **90**(430), 773–795.
- Kotha, S.R., Bindi, D. & Cotton, F., 2016. Partially non-ergodic region specific GMPE for Europe and Middle-East, *Bull. Earthq. Eng.*, **14**(4), 1245–1253.
- Kowsari, M., Halldorsson, B., Hrafinkelsson, B., Snæbjörnsson, J. & Jónsson, S., 2019. Calibration of ground motion models to Icelandic peak ground acceleration data using Bayesian Markov chain Monte Carlo simulation, *Bull. Earthq. Eng.*, **17**(6), 1–30.
- Landwehr, N., Kuehn, N.M., Scheffer, T. & Abrahamson, N., 2016. A non-ergodic ground-motion model for California with spatially varying coefficients, *Bull. seism. Soc. Am.*, **106**(6), 2574–2583.
- Marzocchi, W., Taroni, M. & Selva, J., 2015. Accounting for epistemic uncertainty in PSHA: logic tree and ensemble modeling, *Bull. seism. Soc. Am.*, **105**(4), 2151–2159.
- McGuire, R.K., 1976. Methodology for incorporating parameter uncertainties in seismic hazard analyses for loss risk design intensities, in *Int. Symp. Earthquake Structural Engineering*, St. Louis, U.S.A.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E., 1953. Equations of state calculations by fast computing machines, *J. Chem. Phys.*, **21**(6), 1087–1092.
- Park, I. & Grandhi, R.V., 2014. A Bayesian statistical method for quantifying model form uncertainty and two model combination methods, *Reliab. Eng. Syst. Saf.*, **129**, 46–56.
- Raftery, A., 1995. Bayesian model selection in social research, *Sociol. Methodol.*, **25**, 111–163.
- Roberts, G.O. & Rosenthal, J.S., 2009. Examples of adaptive MCMC, *J. Comput. Graph. Stat.*, **18**, 359–367.
- Runge, A.K., Scherbaum, F., Curtis, A. & Riggelsen, C., 2013. An interactive tool for the elicitation of subjective probabilities in probabilistic seismic-hazard analysis, *Bull. seism. Soc. Am.*, **103**(5), 2862–2874.
- Scherbaum, F., Delavaud, E. & Riggelsen, C., 2009. Model selection in seismic hazard analysis: an information-theoretic perspective, *Bull. seism. Soc. Am.*, **99**(6), 3234.
- Scherbaum, F. & Kuehn, N.M., 2011. Logic tree branch weights and probabilities: summing up to one is not enough, *Earthq. Spectra*, **27**(4), 1237–1251.
- Schöniger, A., Wöhling, T. & Nowak, W., 2015. A statistical concept to assess the uncertainty in Bayesian model weights and its impact on model ranking, *Water Resour. Res.*, **51**(9), 7524–7546.
- Scordilis, E.M., 2006. Empirical global relations converting M_S and m_b to moment magnitude, *J. Seismol.*, **10**(2), 225–236.
- Sedaghati, F. & Pezeshk, S., 2017. Partially nonergodic empirical ground-motion models for predicting horizontal and vertical PGV, PGA, and 5% damped linear acceleration response spectra using data from the Iranian plateau, *Bull. seism. Soc. Am.*, **107**(2), 934–948.
- Stafford, P.J., 2019. Continuous integration of data into ground-motion models using Bayesian updating, *J. Seismol.*, **23**, 39–57.
- Stafford, P.J., Strasser, F.O. & Bommer, J.J., 2008. An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region, *Bull. Earthq. Eng.*, **6**(2), 149–177.
- Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M. & Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, **44**(12), doi:10.1029/2007WR006720.
- Zhao, J.X. et al., 2006. Attenuation relations of strong ground motion in Japan using site classification based on predominant period, *Bull. seism. Soc. Am.*, **96**(3), 898–913.
- Zhu, G. et al., 2016. Multi-model ensemble prediction of terrestrial evapotranspiration across north China using Bayesian model averaging, *Hydrol. Process.*, **30**(16), 2861–2879.

APPENDIX: MAXIMUM LIKELIHOOD ESTIMATION

As a general rule, the MLE approximation of (13) can be solved with classical optimization techniques, such as Newton–Raphson or Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithms. When the distribution of the uncertain parameter is known and simple, however, the problem can be entirely solved analytically. In our case, the fact that we consider a normal distribution renders the computation straightforward. Starting from the system

$$\begin{cases} \frac{\partial}{\partial \mu_k} \sum_{n=1}^N \log g(d_n | M_k, \mu_k, \sigma_k) = 0 \\ \frac{\partial}{\partial \sigma_k} \sum_{n=1}^N \log g(d_n | M_k, \mu_k, \sigma_k) = 0, \end{cases} \quad (\text{A1})$$

we get the following analytical expressions for the optimal hyper-parameters

$$\mu_k^* = \frac{1}{N} \sum_{n=1}^N d_n - f_k(x_n), \quad (\text{A2})$$

and

$$\sigma_k^* = \left[\frac{1}{N} \sum_{n=1}^N (d_n - f_k(x_n) - \mu_k^*)^2 \right]^{1/2}. \quad (\text{A3})$$

The MLE estimation of the marginal likelihood is then obtained by substituting these expressions in eq. (9) so as to obtain

$$\log L(M_k | D) \approx -N \left(\frac{\log 2\pi}{2} + \log \sigma_k^* \right) - \frac{N}{2} + c_p, \quad (\text{A4})$$

where c_p is a constant term corresponding to the uniform priors: $c_p = -\log(\mu_b - \mu_a) - \log(\sigma_b - \sigma_a)$. Finally, let's note that once the optimum is established, one has a convenient and straightforward definition of the posterior prediction PDF (10):

$$y(x) | M_k, D \sim \mathcal{N}(f_k(x) + \mu_k^*, \sigma_k^*). \quad (\text{A5})$$