THE BAYESIAN MODEL AVERAGING OF GROUND MOTION PREDICTION EQUATIONS

M. BERTIN[1], S. MARIN[2], C. MILLET[1,2], C. BERGE-THIERRY[3] and H. HEBERT[2]

[1]CMLA, ENS Paris-Saclay, 94230 Cachan, France
[2]CEA, DAM, DIF, F-91297 Arpajon Cedex, France
[3]CEA Saclay, 91191 Gif-sur-Yvette Cedex, France

E-mail contact of main author: mikl.bertin@gmail.com

Abstract. In France's tectonic context, where seismicity is moderate, records do not cover the whole range of variable configurations useful for the evaluation of seismic hazard. Usually, a set of empirical models established in similar context (Italy, the Mediterranean Basin, U.S.A., Japan, etc) is considered through a model selection process and with the help of a logic tree. Ultimately, this approach is mainly based on the scientist's expertise. There is a data-driven method able to address the issue of model predictive quality evaluation and of model selection without any additional hypothesis: the Bayesian Model Averaging approach (BMA). This method is an extension of classical Bayesian calibration techniques and allows to take into account a set of several models. By making these models encounter a dataset of observations through a statistical framework, BMA approach provides an unbiased evaluation of each model likelihood and moreover, it produces a weighted average formula using each model considered to get the best predictive result. The BMA method is processed with nine GMPEs (Ground-Motion Prediction Equations) issued from several databases including NGA2 (Next Generation Attenuation) and RESORCE (Reference database for Seismic grOund-motion pRediction in Europe). We use the Markov Chain Monte-Carlo method (MCMC) and the Maximum Likelihood Estimation approach (MLE) with around 1000 records issued from the RESORCE-2013 database. Results tend to show that BMA algorithm has a promising potential: it provides not only a hierarchy of GMPEs based upon their statistical accordance to seismic records, but also a new predictive model – combination of the GMPEs – with an enhanced statistical predictive quality compared to the most efficient models used separately.

Article submitted in *Journal of Geophysical Research*, May 2018.

Acknowledgements: The work carried out under the SINAPS@ project receives French funding managed by the 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/).

Key Words: Bayesian Model Averaging, Seismic Hazard Assessment, Ground Motion Prediction Equations.

1 Introduction

Ground Motion Prediction Equations (GMPEs) are empirical models used in Seismic Hazard Assessment (SHA) usually determined by adapted linear regression techniques from ground-motion data recorded by seismic networks. Ground-motion can also be inferred from more theoretical physical-based models. Since the proposal of the probabilistic approach (PSHA) by Cornell and McGuire [19; 36], in the aim of representing the random characteristic of the seismic processes, three variables are considered as random variables in the methodology: location and magnitude of earthquakes and expected ground-motion at a given site. In this study, we

will focus on the epistemic uncertainties included in the simulation of the ground-motion at a given site due to the lack of knowledge and data. Despite the sophistication of models, some discrepancies between predicted and observed seismic intensity measures (such as PGA, PSA, etc) do exist in certain conditions and situations due to: incomplete and simplified representation of the physical processes, model structural inadequacies, measurement errors, conversion between magnitudes scales, uncertainties in fault mechanism, soil conditions (and more generally uncertainties in the metadata assessment). In low-seismic regions, the sparsity of data obliges to accommodate models developed using strong motions recorded in other regions. Prediction of a single model can lead to bias of the forecast or under-estimation of its variance. Quantification of uncertainties in ground-motion prediction remains a challenge in PSHA and is partly arbitrary solved by fixing values or threshold (like the number of standard deviation considered in the integral computation). To overcome the limitations of the single-model approach and improve the estimation of uncertainties, the multi-model simulations have come out as a convenient way and have become a regulatory requirement in PSHA (IAEA [30], NRC [38]).

With the multi-model approach, published models have to be selected and results have to be combined. The standard methodology in PSHA consists in the use of a logic-tree weighting each model. Epistemic uncertainties of weighted GMPEs predictions are then associated to a Monte-Carlo framework in order to study the uncertainty of aleatory variables. Weighting the GMPEs is generally a matter of experts' judgement which could raise some discussions and difficulties. To address this issue, weights could be determined in a quantitative way reducing subjectivity by comparing the models performance in front of observed data. Some earlier studies proposed different frameworks for model selection and ranking for particular datasets [42; 23; 33]. In a multi-model ensemble perspective, an alternative approach named Bayesian Model Averaging (BMA) is proposed.

This is an extension of standard Bayesian inference approach in the presence of multiple competing models and has been widely used in the social and health sciences [41; 29]. With the BMA approach, the overall prediction PDF (*Probability Density Functions*) is a weighted average of predicted PDFs based on each of the individual models; the weights are the estimated posterior model probabilities and reflect the models prediction quality regarding a specific set of training data, relatively to the other models. The BMA forecast variance decomposes into two components, corresponding to within-model variance and between-model variance. While a standard ensemble spread only captures the first component, the BMA approach accounts for the effect of models combination and offers a way to predict unknown responses more reliably than each individual model in a set. The objective of the paper is to explore how BMA can improve accuracy and reliability of ground-motion prediction in either PSHA or DSHA (Deterministic Seismic Hazard Assessment) when multi-models are used. The performance of BMA including nine GMPEs will be investigated for the RESORCE database events ground-motion forecasting. The prediction uncertainty quantification can be very helpful for decision makers in seismic codes for design purposes and reduction of damages from earthquakes. This approach provides not only a hierarchy of GMPEs based upon their statistical accordance to seismic records, avoiding any arbitrary choices, but also a new predictive model, combination of the GMPEs, with an enhanced statistical predictive quality compared to the most efficient GMPE used separately.

In section 2 we present the BMA main ideas and how the BMA model can be estimated. In section 3 we give BMA results using nine GMPEs along with the RESORCE-2013 [4] database records of events and in section 4 we propose some concluding remarks.

2 Bayesian model averaging

2.1 Principle of the BMA method

The standard procedure, i.e. selecting a set of empirical models through a logic tree to make predictions of seismic hazard, omits a source of uncertainty. These models have been selected among several competing models, many of which are possible candidates for describing a given seismic situation. Other plausible models could provide different answers and this is a source of uncertainty in drawing conclusions. Put another way, 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 other models. And if at least one of them could lead to radically different conclusions, ignoring it underestimates uncertainty and is a risky approach.

Bayesian model averaging [41; 34; 29] adress this problem by considering the entire ensemble of models first considered. If one wants to produce a forecast of a given quantity of interest y, i.e the Peak Ground Acceleration or the Pseudo Spectral Acceleration, based on the training data D and with a set of K models $M_1, M_2, ..., M_K$ (the GMPEs), we have an expression for the forecast PDF $p(y|D)$ given by the law of total probability:

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

where $p(y|M_k, D)$ is the prediction PDF of y conditioned with the data D using the model M_k alone, and $p(M_k|D)$ is the posterior probability of model M_k being the best suited to make predictions given the training data D. Prediction PDF $p(y|M_k, D)$ in (1) refers to the most classical Bayesian approach where posterior probability of one or several unknown model parameter(s) are infered from the data D through the evaluation of the model likelihood, approach detailed in the next section. What represents the specificity of the BMA approach in (1) is the posterior probability of model $p(M_k|D)$. Since we have stated that the *suitable* choice of model, i.e. the selection of the best one by a user considering the observations D, is a source of uncertainty that cannot be ignored, we can consider it as a new unknown parameter, a K-dimensional, discrete, random variable, for which one can process a classical Bayesian inference as well. These posterior probabilities add up to one, $\sum_{k=1}^{K} p(M_k|D) = 1$, so they can be considered as weights, and the BMA prediction PDF is therefore a weighted average.

BMA weights are obtained from prior probabilities of models $p(M_k)$ following the Bayes' theorem:

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

where $L(M_k|D)$ is the marginal likelihood of the model M_k, i.e. the probability to predict the data D when using the model M_k: $L(M_k|D) = p(D|M_k)$. Priors $p(M_k)$ correspond to the prior trust in models, before testing them on the training data D. Using the expert judgment to quantify these prior probabilities can then appear tempting but with the BMA approach, we often consider these priors as a uniform distribution $p(M_k) = 1/K$. There are two main raisons for that. First the Bayesian perspective usually consists in making the less informative hypothesis possible ensuring that the inference mainly results from the good fitting of a model to data, and not from additionnal, restraining assumptions. Then, using expert knowledge to quantify the prior distribution appears difficult, if not arbitrary, since there is not any known, logical relation allowing to convert the knowledge about a model set into prior model probabilities.

From the composite distribution (1), we can express the expectation and the variance of the BMA prediction of y conditioned with the observation data D:

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

and

$$\mathrm{Var}(y|D) = \sum_{k=1}^{K} p(M_k|D)\mathrm{Var}(y|M_k, D) + \sum_{k=1}^{K} p(M_k|D)\left(E(y|M_k, D) - E(y|D)\right)^2, \tag{4}$$

where $E(y|M_k, D)$ and $\mathrm{Var}(y|M_k, D)$ are respectively the expectation and variance of the posterior distribution of y conditioned with the data D using the model M_k alone. The variance of the composite BMA model (4) shows two distinct terms. The first term, called *within-model variance*, accounts for the average degree of uncertainty in each model prediction. The second term, called *between-model variance*, corresponds to the degree of uncertainty in response prediction resulting from model selection uncertainty.

With a uniformally distributed prior, the BMA weights are directly proportional to the marginal likelihood of models. Thus, for the weights evaluation and for the posterior prediction PDF in (1) as well, the BMA approach requires the computation of the models likelihood, i.e. their posterior statistical accordance to observations D.

2.2 Models likelihood

Like any other regression technique, evaluating a model likelihood implies that there are several or at least one degree of liberty attached to the model and/or to the data on which one can act to seek the best fitting between the model and the data. These uncertainties can play different roles: they can be unknown physical parameters, empirical coefficients one wants to adjust, model input errors, model prediction errors (aleatory or epistemic) or experimental errors for instance. In the case of GMPEs here, we will limit to a single prediction error term to present the method. What follows can be easily extended to a random vector of uncertainties instead of a unique scalar random quantity.

Since the origin of uncertainties such as the model prediction error and the measurement error cannot be conveniently distinguished and properly interpreted in practice, we adopt here a common formulation of the problem involving an unknown error term that is supposed to account for model discrepancy and for the measurement errors [39; 15]. Thus, each GMPE is represented by:

$$y(x) = f_k(x) + \epsilon_k, \tag{5}$$

where y is the unknown quantity of interest for which we want a prediction, here the logarithm of the 5 %-damped PSA at a given frequency, f_k is the deterministic prediction of the GMPE M_k, x is a set of earthquake characteristics used as inputs of the GMPEs (moment magnitude, distance, focal depth, VS30, fault type, etc), and ϵ_k is the random prediction error attached to the model M_k. The additive formulation is appropriate in that case since we consider the logarithm of the ground acceleration and because each GMPE is established considering a log-normal error term: $PSA \sim \exp(y + \epsilon)$.

The only assumption we have to make here is the distribution form of the prediction error. Because GMPEs are built following a log-normal discrepancy between predictions and data, it appears natural to assume for ϵ_k a normal distribution:

$$\epsilon_k \sim \mathcal{N}\left(\mu_k, \sigma_k^2\right), \tag{6}$$

where μ_k and σ_k are hyperparameters for the normal distribution: the expectation and the standard deviation, respectively. This assumption is further discussed and justified in section 3. These hyperparameters are considered unknown and the aim of the Bayesian inference is to find their posterior distribution. The reason to do that instead of directly using ϵ_k is to make less assumptions: a direct inference on ϵ_k would require to define precisely its prior distribution, quantity for which we do not have any specific knowledge. One could argue that it is not entirely true: the GMPEs could provide a convenient prior for ϵ_k with their centered log-normal error term but that quantity only accounts for the discrepancy with the data used to establish the model. Here we want to consider any dataset of observation.

Following that configuration, the marginal likelihood can be defined as the integral of the likelihood function over the hyperparameters (hyperparameters are *marginalized* by the integration):

$$L(M_k|D) = \int_{\sigma_k}\int_{\mu_k} l(M_k,\mu_k,\sigma_k|D)\, p(\mu_k)p(\sigma_k)\, \mathrm{d}\mu_k\mathrm{d}\sigma_k, \tag{7}$$

where $p(\mu_k)$ and $p(\sigma_k)$ are the prior distributions of the hyperparameters, and $l(M_k,\mu_k,\sigma_k|D)$ is the likelihood function (of μ_k and σ_k) that is the probability to predict the observation D with the model M_k when using μ_k and σ_k values in the random prediction error term ϵ_k. To get a specific expression for the likelihood function, let's consider that D is a vector of N single observations d_n of the quantity y, i.e. a set of N 5 % damped horizontal PSA at a given frequency issued from N distinct seismic records, we can then write

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

where $g(\cdot|M_k,\mu_k,\sigma_k)$ is the prediction PDF for y using the GMPE M_k and the hyperparameters μ_k and σ_k. Using the assumptions (6) and (7), it follows:

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

where x_n is the set of model inputs corresponding to the n^{th} seismic record.

As regards the hyperparameters priors, it is often assumed in Bayesian approaches that they must follow a non-informative distribution. The invariant Jeffreys' prior [31] can be chosen for instance, or the more common uniform prior: $\mu_k \sim \mathcal{U}(\mu_a,\mu_b)$ and $\sigma_k \sim \mathcal{U}(\sigma_a,\sigma_b)$. Let's note that these assumptions are not absolutely neutral. Indeed, the inference process can be highly dependant to where the hyperparameters bounds are located. In our case, the domain $[\mu_a,\mu_b]\times[\sigma_a,\sigma_b]$ must be large enough and must contain the essential part of the likelihood function otherwise a bias is clearly induced in the computation.

The proper evaluation of the likelihood function and above all of the marginal likelihood integral appears crucial here: once they can be correctly estimated, we then have acces to the joint posterior probability of hyperparameters:

$$p(\mu_k,\sigma_k|D) = \frac{l(M_k,\mu_k,\sigma_k|D)\, p(\mu_k)p(\sigma_k)}{L(M_k|D)}, \tag{10}$$

then to the posterior PDF of the model prediction error:

$$p(\epsilon_k|D) = \int_{\sigma_k}\int_{\mu_k} g(\epsilon_k|\mu_k,\sigma_k)\, p(\mu_k,\sigma_k|D)\, \mathrm{d}\mu_k\mathrm{d}\sigma_k, \tag{11}$$

and directly to the prediction PDF of y used in the weighted average (1):

$$p(y|M_k,D) = \int_{\sigma_k}\int_{\mu_k} g(y|M_k,\mu_k,\sigma_k)\, p(\mu_k,\sigma_k|D)\, \mathrm{d}\mu_k \mathrm{d}\sigma_k. \tag{12}$$

The main difficulty is that as a general rule, the likelihood function is not straightforward to compute, its evaluation can be costly. The estimation of its integral is then a non-trivial challenge, all the more difficult as there may be a large number of uncertain parameters, and therefore a hyperparameter space of large dimension. There is a wide class of methods and algorithms able to adress this problem [49; 39; 29; 18; 47]. We briefly present two of them in the next subsection: the MLE method (*Maximum Likelihood Estimation*), specifically suitable for cases where the parameter distribution is known and simple, and the MCMC algorithm (*Markov Chain Monte Carlo*), useful for more general, large dimension problems.

2.3 Estimation methods

For each model M_k, the Bayesian formulation can be used to determine the model parameters θ_k. It can be a single scalar $\theta_k = \epsilon_k$ or a vector $\theta_k = [\mu_k,\sigma_k]^\top$, a vector that could also contain some GMPE's coeficients as additionnal random parameters for instance. We consider that every admissible θ_k belongs to a parameter space noted Θ_k. If all parameters are joinly uniformly distributed then the prior distribution of these parameters is constant over the entire space Θ_k and Bayes theorem can be written as

$$P(\theta_k|D,M_k) = \frac{P(D|\theta_k,M_k)P(\theta_k|M_k)}{P(D|M_k)} = c\; P(D|\theta_k,M_k), \tag{13}$$

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

$$P(D|M_k) = \int_{\Theta_k} P(D|\theta_k,M_k)P(\theta_k|M_k)\, \mathrm{d}\theta_k, \tag{14}$$

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 the rule (2).

The main computational issues concern the estimate of (14), 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 (14) is given by the Maximum Likelihood Estimation (MLE) and indeed, the problem can readily be solved analytically for the GMPEs, provided that we impose the condition (6). The details of calculations are given in the appendix. When the posterior is an intractable expression, a fairly general approach is to use a MCMC method [37; 28], 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

$$\alpha = \min\left\{1, \frac{l(M_k,\theta_k^*|D)}{l(M_k,\theta_k^m|D)}\right\}, \tag{15}$$

and finally, sets $\theta^{m+1} = \theta^*$ with probability α and $\theta^{m+1} = \theta^m$, with probability $1-\alpha$. 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, (14) can be estimated with a standard Monte-Carlo estimate. In this study, we diagnose the successful convergence and the proper mixing of the Markov chains when reaching a Gelman-Rubin's potential scale reduction factor [27; 21] close to one.

3 Application

3.1 The selection of GMPEs

The selection of GMPEs is one of the most important elements of any SHA study (be it Probabilistic or Deterministic), impacting strongly the SHA results in terms of level and dispersion predictions. In this study, GMPEs have been obtained from local, regional and worldwide data all of which satisfy the a priori requirements for shallow crustal tectonic settings [20; 12]. A parsimony principle has been 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). Many models had been analyzed by Delavaud et al. [23] in the SHARE (Seismic Hazard hArmonization in Europe) project as the candidate GMPEs for seismic hazard in Europe with the likelihood-based ranking method [42]. In another study, Beauval et al. (2012 [7]) also used this method to investigate GMPEs with the french accelerometric database (RAP). In GEM (Global Earthquakes Model) an Euclidean distance-based ranking method [33] has been proposed and applied to the Middle East. According to these studies combined results using the following GMPEs seem appropriate for Europe: Berge-Thierry et al. (2003 [8]), Zhao et al. (2006 [50]), Cauzzi & Faccioli (2008 [16]), Akkar & Bommer (2010 [2]), Bindi et al. (2010 [9]). Many models have been updated by the authors since then and will be used in their most recent versions: Faccioli (2010 [25]), Akkar & Bommer (2013 [3]), Cauzzi (2015 [17]), Bindi et al. (2011 [10]) for Italy and Bindi (2013 [11]) for Europe.

Furthermore, the U.S. NGA (Next Generation Attenuation) models developed for more active regions from strong ground motion must have to be considered due to their extensive and standard use in seismic hazard analysis. According to the NGA comparison done by Abrahamson et al. (2008 [1]) and their applicability in the Euro-Mediterranean region by Stafford et al. (2008 [45]), Boore and Atkinson (2007 [13]) has been chosen as a template among all NGA models for two reasons: first, in general, the median ground motions are similar to one another, therefore, a single model is sufficient; then it is simple (a minimum of parameters is used in comparison with the other models). The latest model updates have been used for this study: Boore and Atkinson (2014 [14]). Berge-Thierry et al. (2003 [8]) is the oldest and simplest model used. This model has been developed for the French nuclear safety regulation (Règle Fondamentale de Sûreté, RFS 2001-01) and is still cited as a reference by the Nuclear Safety Authority (reference Guide 22, ASN, 2017 [5]). As this last model used surface magnitude instead of moment magnitude, a relation to convert magnitudes is necessary. Scordilis (2006 [44]) published such relations for different magnitude scales to help building homogenous global seismic catalogs in the most reliable and used scale, the moment magnitude. To investigate the performance of several selected GMPEs in face with the earthquakes from Europe and the Middle East of the RESORCE database, three models developed using the database itself by Akkar et al. (2013 [3]), Bindi et al. (2013 [11]), and Derras et al. (2016 [24]) have been considered. In the results presented here, the nine selected GMPEs Berge-Thierry et al. (2003 [8]), Zhao et al. (2006 [50]), Bindi et al. (2011 [10]), Akkar et al. (2013 [3]), Bindi et al. (2013 [11]), Boore and Atkinson (2014 [14]), Cauzzi (2015 [17]) and Derras et al. (2016 [24]) will be designated with the keywords BT03, ZH06, FA10, BI11, AK13, BI13, BO14, CA15, DE16, respectively.

3.2 The RESORCE database

As source of training data for the Bayesian calibration of empirical models, we use seismic records and associated metadata gathered in the RESORCE database (Reference database for Seismic grOund-motion pRediction in Europe). RESORCE is a freely accessible platform for accessing and retrieving reliable ground-motion data from pan-European earthquakes and associated seismological and geotechnical parameters. For this study, we use the state of the

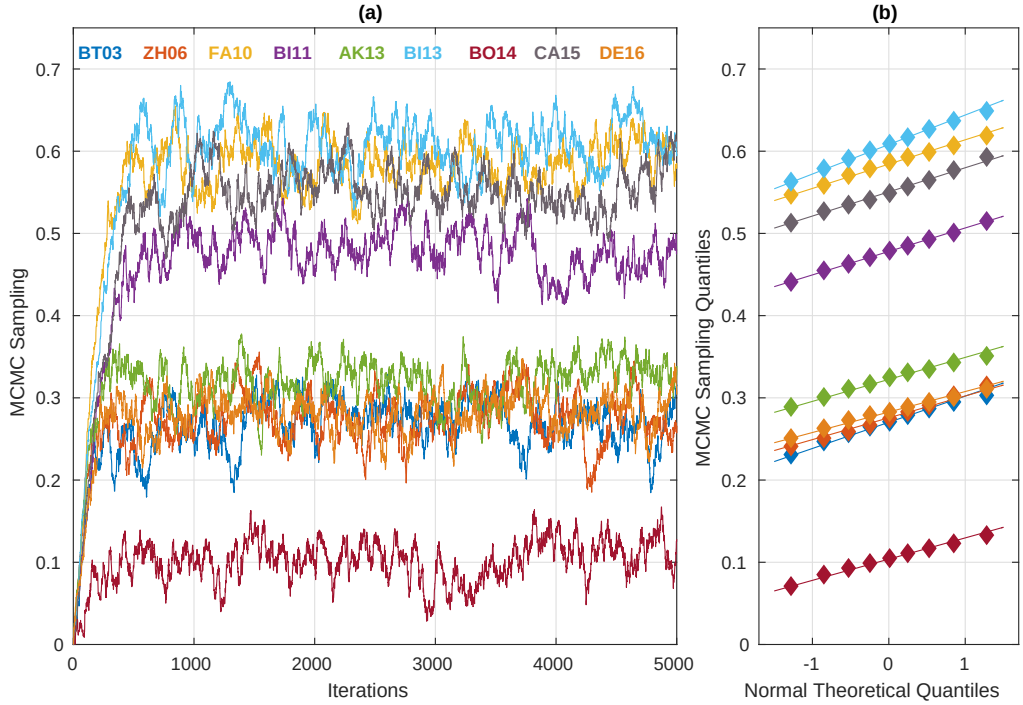

Figure 1: *(a) MCMC Metropolis-Hastings sampling of each prediction error ϵ_k attached to the nine considered GMPEs for the period $T = 1$ s. (b) Quantile-quantile plot: the nine empirical deciles of each sample versus the nine theoretical deciles of the normal distribution.*

RESORCE-2013 database version [4] as it was when accessed on January 2018. That represents 5 882 seismic records from 1 814 events and 1 540 strong-motion stations.

Some selection steps must be carried out on these data in order to be able to proceed the Bayesian study with the selected GMPEs. First, we have to make sure the records used present enough information regarding the GMPE's inputs and outputs, saying the distance (epicentral distance, hypocentral distance, Joyner-Boore distance), the V_{S30} speed, the focal depth, the fault mechanism, the event magnitude, and the horizontal pseudo-spectral ground acceleration (two components) must all be avalaible. We select specific metadata ranges to ensure these records are in the valid application domain for the nine selected GMPEs, meaning the moment magnitude must be in the range $[5, 7.3]$, the distance in the range $[4, 150]$ km and the V_{S30} speed must be in the range $[300, 1200]$ m.s^{-1}. Then, several magnitudes types are present in the database (M_w, M_s, M_b, M_d, M_l) but the moment magnitude M_w remains the majority (for 79 % of records), we therefore only select records of events for which the magnitude is provided in M_w in order to avoid any issue regarding the magnitude conversion. Likewise, we only select events for which the fault mechanism is indicated as *Normal*, *Strike-Slip* and *Reverse*. This selection procedure provides a subset with $N = 939$ records to use for the BMA analysis.

The GMPE's output used in the statistical comparison with RESORCE observations is the logarithm of the horizontal 5 % damped pseudo-spectral acceleration (PSA), expressed in g, computed for a specific set of periods (see Table 1) for which we can use exact coefficient values provided in the litterature for the nine selected GMPEs. The corresponding observations from the RESORCE database are obtained with the geometric mean of the two horizontal PSA components.

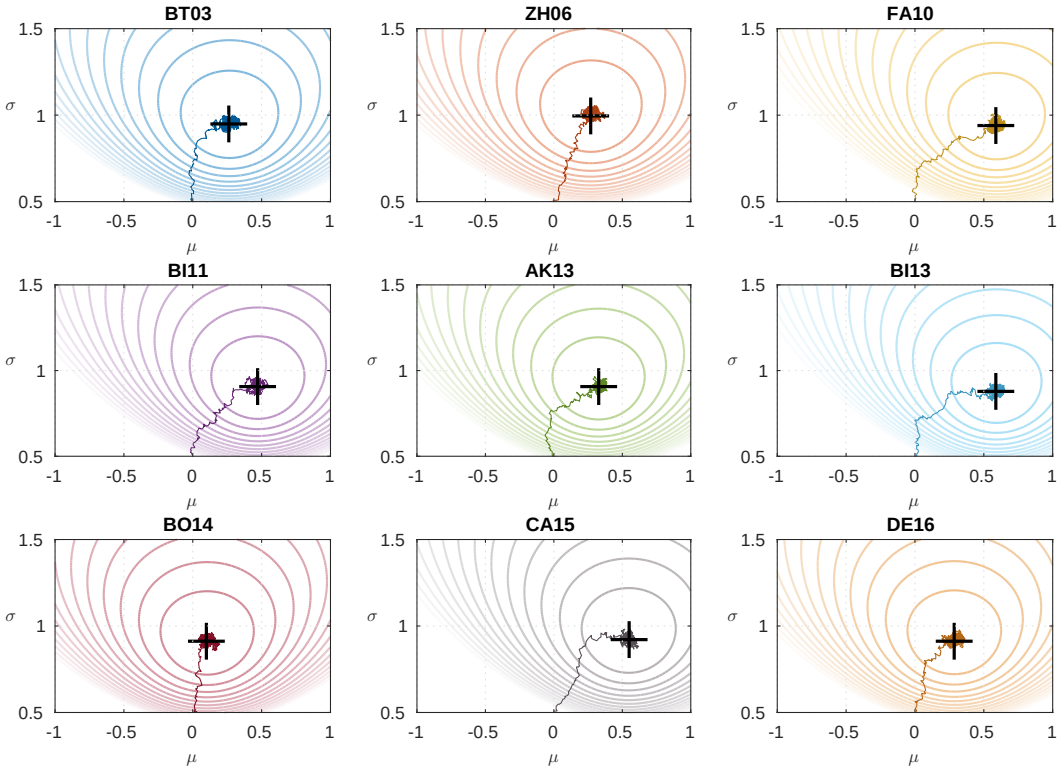

Figure 2: *For each one of the nine considered GMPEs and for the period $T = 1$ s: contours indicate the log-likelihood function of the hyperparameters μ_k and σ_k, the dark line reprensents the MCMC markov chain for the hyperparameters $(\mu_k^{(m)}, \sigma_k^{(m)})_{m \leq M}$ starting from the point $\mu_k^{(0)} = 0$ and $\sigma_k^{(0)} = 0.5$, and the black cross shows the analytical maximum of likelihood (μ_k^*, σ_k^*) obtained with the MLE method.*

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

Table 1: *Set of periods for which the 5 %-damped horizontal PSA is used for the BMA calibration*

3.3 Application of the BMA approach

As a general rule, a proper Bayesian inference requires to make as few assumptions as possible. Even if the normal distribution hypothesis (6) for the prediction error ϵ_k appears reasonable, the MCMC Metropolis-Hastings algorithm provides a formal way to validate that hypothesis. As a matter of fact, before working with the hyperparameters μ_k and σ_k, one could wonder what the posterior distribution of ϵ_k would look like if we did not assume a normal distribution. So, in the first place, let's assume the prediction error have a non-informative prior: a uniform distribution. The Metropolis-Hastings algorithm can then be carried out directly on ϵ_k for each GMPE for the period 1 s (results are similar for the other periods). The uniform prior is $\epsilon_k \sim \mathcal{U}(-2,2)$, the starting point is $\epsilon_k^{(0)} = 0$, and the chain's standard deviation is $c = 0.005$. Figure 1-(a) shows for the resulting MCMC sampling for the nine GMPEs. The Markov chain's length is M=5 000, including 500 iterations of warm-up. Figure 1-(b) shows a quantile-quantile plot that enlights

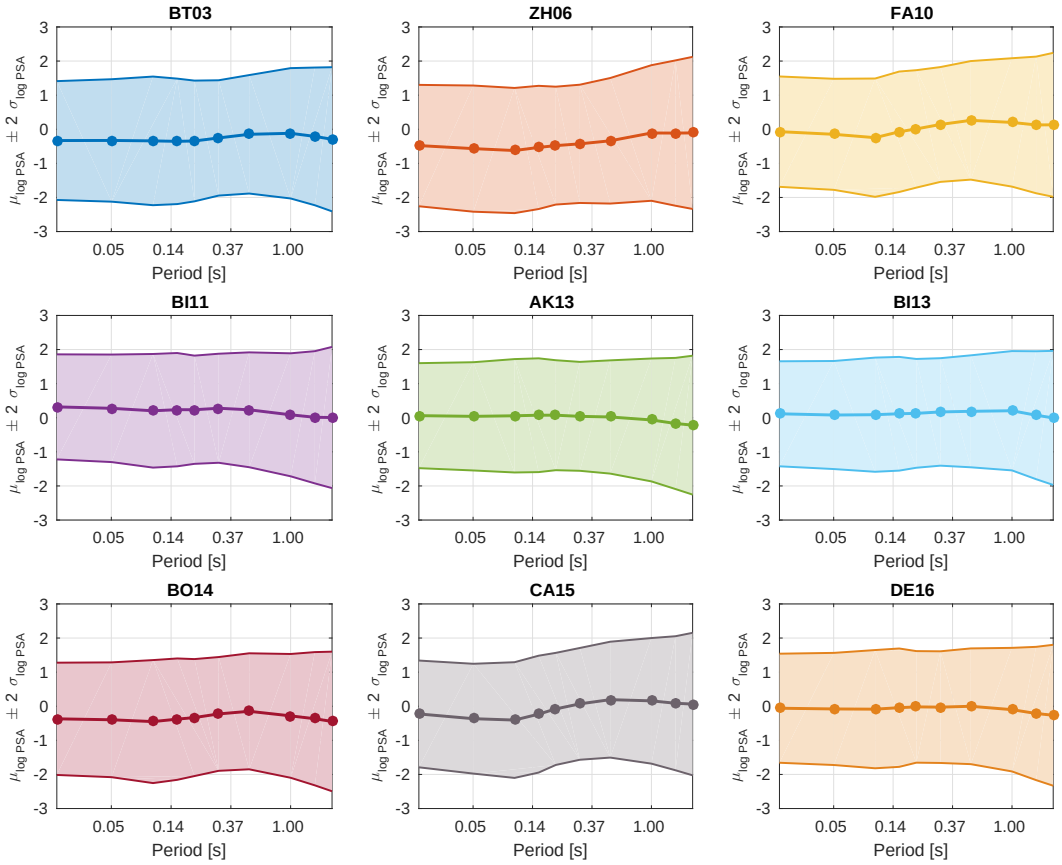

Figure 3: *Summary of calibration results. For each one of the nine considered GMPEs and for periods from $T = 0.02$ s to $T = 2$ s (see Table 1): the dark line shows the optimal biais values μ_k^* and the colored surface indicates the intervals $\mu_k^* \pm 2\sigma_k^*$ where μ_k^* and σ_k^* are issued from the Bayesian calibration with RESORCE-2013 data. The horizontal PSA is computed with GMPEs in g.*

the nature of posterior ϵ_k's distribution [32]. The linear relation between the empirical deciles of the Markov chains and the theoretical deciles of the normal distribution is a validation of the normality assumption. Therefore, from now on we work with the assumption (6) and adjustment variables are the hyperparameters μ_k and σ_k, the prediction bias and the prediction error's standard deviation, respectively.

The MLE estimation of optimal hyperparameters (18,19) is then computed and compared to a MCMC algorithm processed in the hyperparameter space. Figure 2 shows the result for the nine GMPEs and for the period 1 s. The hyperparameters priors are $\mu_k \sim \mathcal{U}(-1,1)$ and $\sigma_k \sim \mathcal{U}(0.5,5)$, the chosen starting point is $\mu_k^{(0)} = 0$ and $\sigma_k^{(0)} = 0.5$, and the MCMC standard deviation is 0.01 for both hyperparameters. Let's notice that the calibration ends up with non-zeros corrective bias μ_k^* for each GMPE. That was expectable and does not represent any individual, predictive bias from GMPEs: values of μ_k^* only account for the statistical discrepancy between GMPE's predictions and the training database used here (that is not the original observation database used to establish the GMPEs). Calibration results for the whole range of periods (Table 1) are summarized in figure 3.

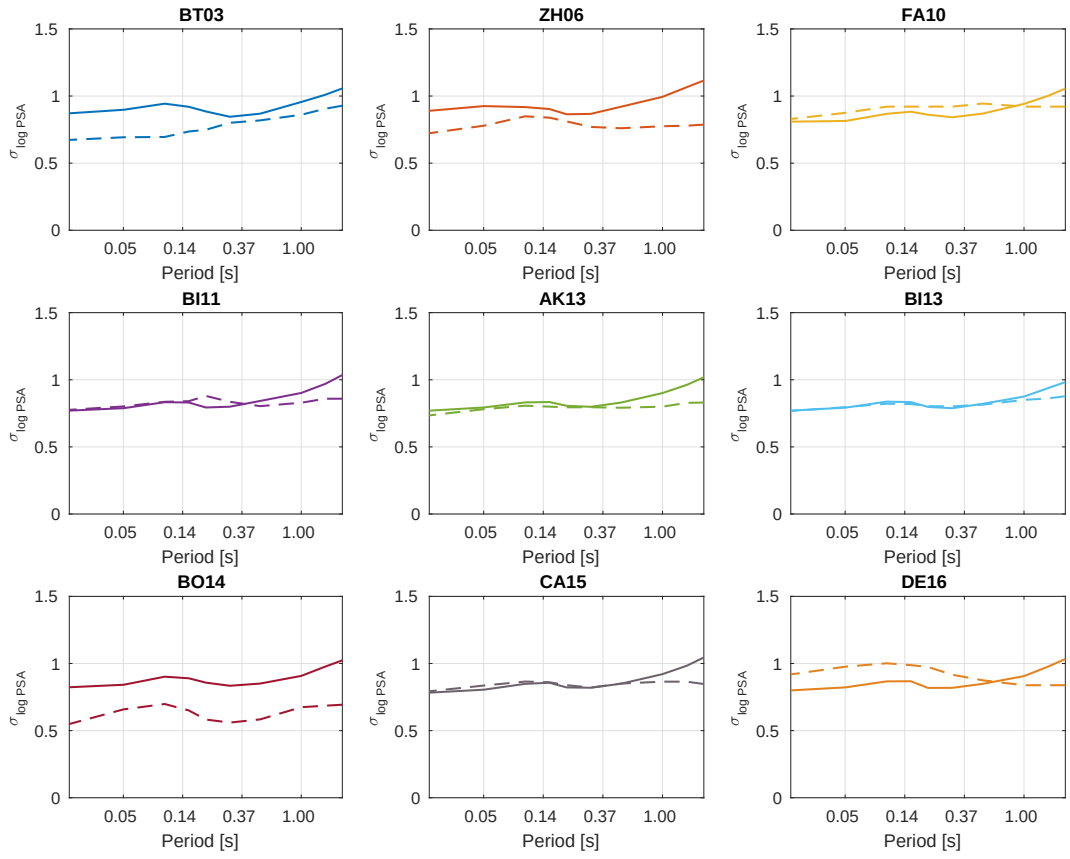

Figure 4: *For each one of the nine considered GMPEs and for periods from $T = 0.02$ s to $T = 2$ s (see Table 1): dashed lines show the original σ_{GMPE} estimated from the GMPE regression, plain lines show the optimal σ_k^* values issued from the Bayesian calibration with RESORCE-2013 data. The horizontal PSA is computed with GMPEs in g.*

More specifically, in figure 4 the posterior standard deviation of prediction errors σ_k^* are compared with the original σ_{GMPE} values (obtained from the initial GMPE estimation on their own dataset). Despite the fact these GMPEs are issued from different regions and datasets, the statistical distance with RESORCE-2013 data (σ_k^*) does not differ much from the discrepancy relative to original data (σ_{GMPE}): same orders are obtained. The fact that the prediction bias μ_k is calibrated along with the standard deviation σ_k is certainly the reason of that: if the inference was performed on the standard deviation alone, larger values would have been obtained. Besides, values appears quite close for GMPEs established using data from RESORCE such as Akkar et al. (2013 [3]) and Bindi et al. (2013 [11]). For such models, the prediction bias μ_k^* appears close to 0 as well (figure 3). The only differences result from the fact we did not select the exact same subsets of records from RESORCE database.

The BMA weights and the BMA variance components (4) are shown in figure 5. The fact that most of computed BMA weights appear very close to 0 does not mean that corresponding GMPEs are absolutely not suited to make predictions on European seismicity: these weights are *relative* values obtained with (2) and should be interpreted only compared to a specific set of models and to a specific set of observed data. With our set of nine GMPEs and our subset of 939 records issued from RESORCE-2013, three GMPEs appear to share the highest values

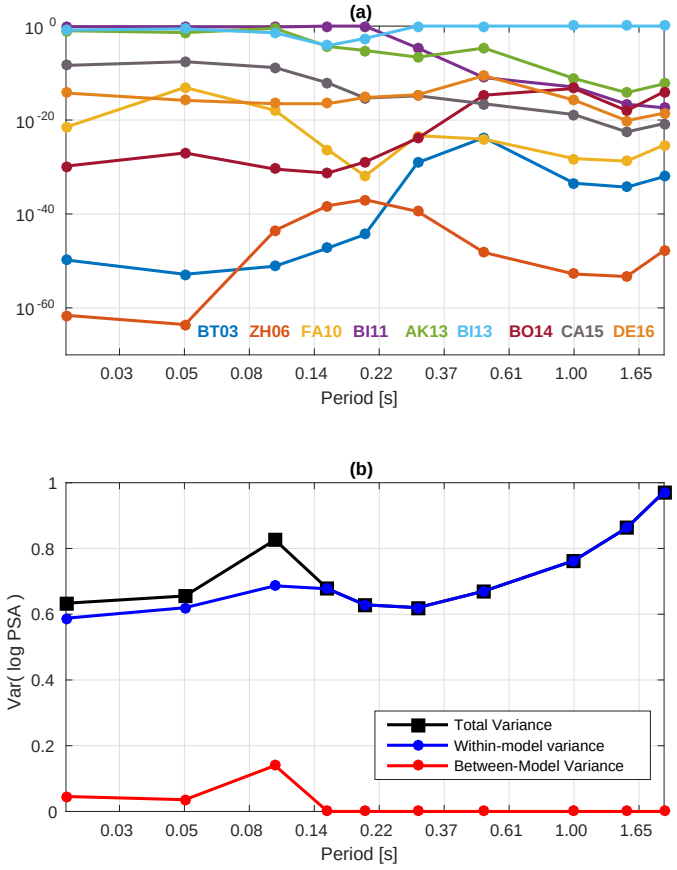

Figure 5: *(a) Posterior BMA weights $P(M_k|D)$ for each one of the nine considered GMPEs and for periods from $T = 0.02$ s to $T = 2$ s (see Table 1). (b) Resulting variance of the BMA combination of the nine considered GMPEs and its 2 components: the within-model variance and the the between-model variance. The horizontal PSA is computed with GMPEs in g.*

of BMA weight: Bindi et al. (2011 [10]), for Italy and Akkar et al. (2013 [3]) and Bindi et al. (2013 [11]) for Europe. Looking at the BMA variance components (figure 5-(b)) offers a clear insight of what could be the resulting Bayesian strategy. As a matter of fact, for large periods (low frequencies), the total variance of the BMA combination is only the result of within-model variance: this means only the model with highest weight does contribute (Bindi et al., 2013) and could then be selected alone. But for lower periods (high frequencies), the between-model component of variance is not neglectable anymore, meaning more than one model have a significant contribution to BMA predictions and should therefore be taken into account together: Bindi et al., 2011, Akkar et al., 2013 and Bindi et al., 2013.

The validation of that BMA strategy is shown in figure 6. 739 records of events are randomly selected among the subset of 939 considered observations and are used to calibrate the BMA combination of models, and predictions of the resulting model are then compared to the 200 remaining observations. On average, results show that 94.4 % of these observations are in the 95 % confidence interval and 98.8 % of them are in the 99.7 % confidence interval. The fact that the BMA combination based upon the 3 most performing GMPEs provides correct predictions regarding the RESORCE observations may be reassuring but it is not very surprising given

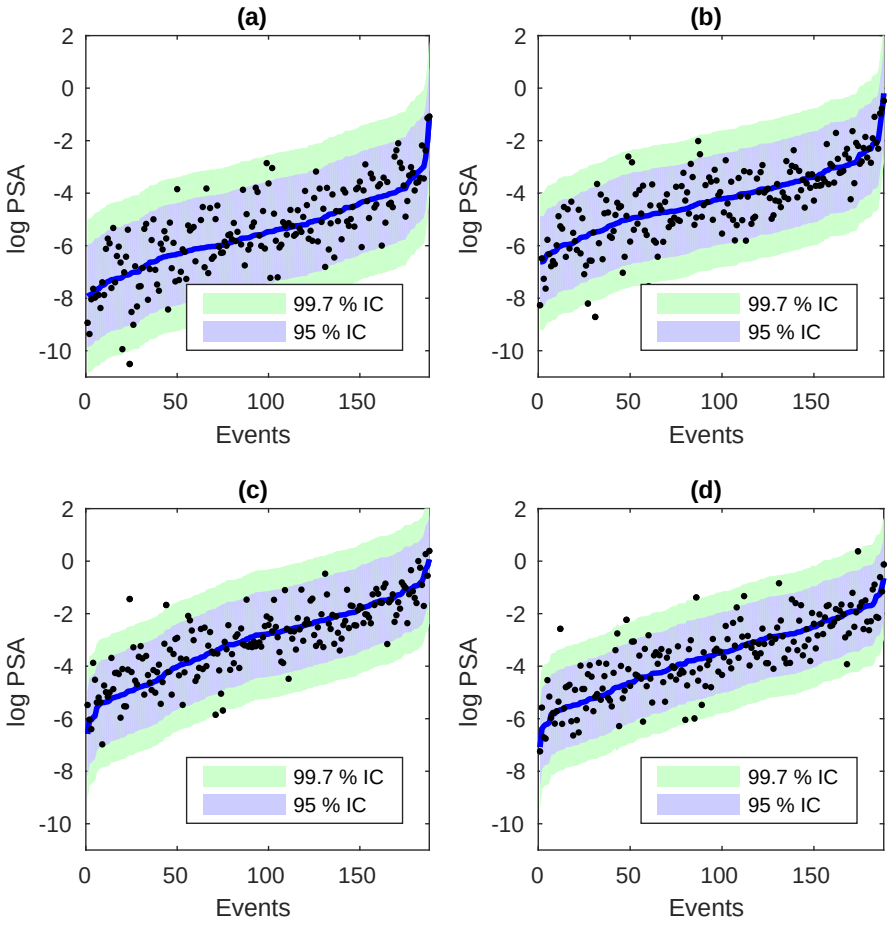

Figure 6: *99.7 % and 95 % confidence intervals of BMA predictions (light green and light blue areas, respectively) compared to 200 events that were not used for the BMA calibration (black dots). Events are ordered to have an increasing BMA expectation (blue line). The horizontal PSA is computed with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

that some of the selected GMPEs were established with RESORCE data. In fact, the strength of the BMA approach can be brought out when compared with the models taken individually.

Figure 7 shows the comparison between the predicted residual error sum of squares (PRESS) of the BMA model combination and the quadratic mean of prediction error of each individual GMPE. More specifically, a *leave-one-out* strategy [26] is used to estimate the PRESS statistic:

$$\text{PRESS} = \frac{1}{N}\sum_{n=1}^{N}\left(E[F_{\text{BMA},n}(x_n)] - d_n\right)^2, \tag{16}$$

where d_n is the n^{th} observation, $E[.]$ is the expectation and $F_{\text{BMA},n}(x_n)$ is the random BMA prediction for the n^{th} observation based upon the calibration of GMPEs using the whole dataset except d_n. Figure 7-(a) shows the comparison whith the GMPEs before the Bayesian calibration of prediction error and 7-(b) shows the GMPE's performance after being calibrated with observations: the calibration brings a clear improvement of preditive performance for each GMPE. Moreover, when the most performing models are then combined, the resulting BMA combina-

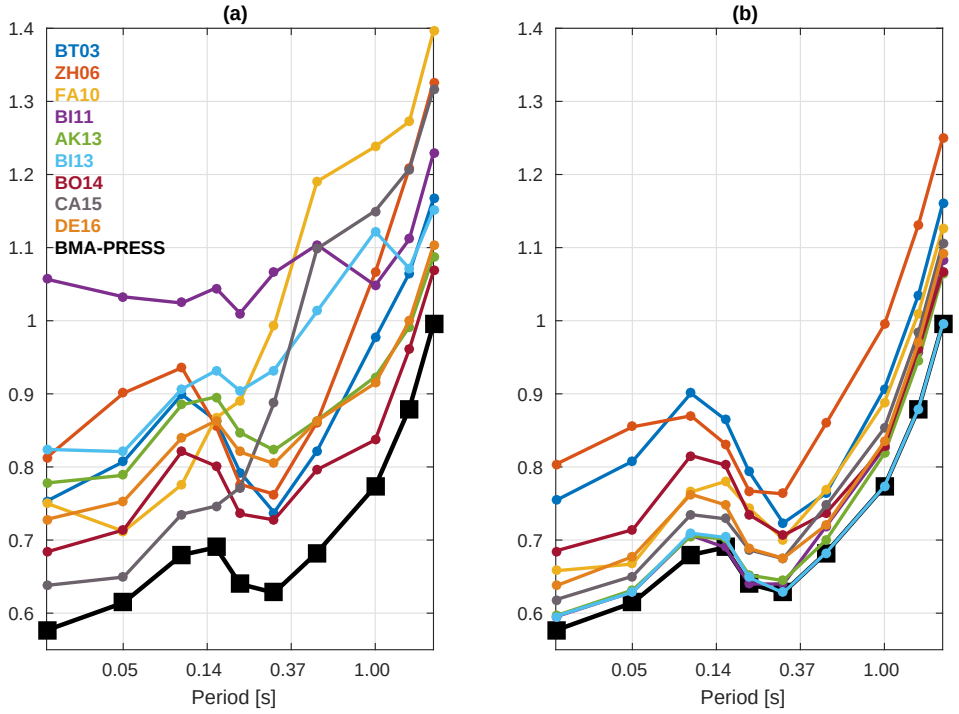

Figure 7: *Coloured lines: quadratic mean of prediction error for each GMPE according to RESORCE data, prior predictions in figure (a), and posterior predictions in figure (b), i.e. after Bayesian calibration. Black line: predicted residual error sum of squares (PRESS) of the BMA combination of calibrated GMPEs (leave-one-out strategy for the calibration (16)).*

tion shows even better performances (the PRESS statistic is lower for the BMA combination rather than for each GMPE for low periods in figure 7-(b)). That result shows the specificity of the BMA method and what one can expect from it: in addition to the calibration and the ranking of the considered models, that approach offers a way to combine the calibrated models together in order to reach a better statistical accordance to observations.

4 Conclusion

The BMA approach has been carried out with nine GMPEs and around 1000 records of events issued from the RESORCE-2013 database. By using a classical approach, i.e. considering a random prediction error term ϵ_k attached to the GMPEs, but without ignoring the prediction bias μ_k, the BMA procedure has already proven a promising potential as a data-driven alternative to logic-trees in a SHA perspective. Making as few assumption as possible, each empirical model is calibrated according to observations, then models are ranked according to their respective likelihood and finaly a new model, combination of the most performing ones, is proposed and that combination provides better predictions compared to the most efficient GMPE used separately. More specifically, by studying the horizontal 5 %-damped PSA for different frequencies, it appeared that distinct strategies should be carried out according to the frequency range: the selection of a unique model for low frequencies (Bindi et al. 2013 [11]) and the combination of three GMPEs for higher frequencies (Bindi et al. 2011 [10], Akkar et al. 2013 [3], Bindi et al. 2013 [11]).

We must recall that the goal of this study is to present the BMA method feasibility and potential for the use of GMPEs, without any PSHA obligation, yet. Results presented here should therefore be taken cautiously. Indeed, interpretation and generalization of the BMA results deeply depend on the observations used and the model set considered. The more observations (and associated meta-data) we have to represent the studied seismicity, the more performant is the BMA combination we get. Likewise, the more models we use, the more general can be considered the resulting BMA process. The latter point is crucial: the implicit assumption behind the BMA approach is that the whole set of observations is reachable by the space formed with the models considered. Put in another way, all probabilities are implicitely conditional on the set of models under consideration. The larger is the set of models, the larger is then the scope of the Bayesian inference.

Hence, several approaches can be considered in order to extend the method presented here. First, more GMPEs could be added to the study. Despite being promising, that kind of approach may end up with a limitation: adding too much models could make decrease the size of observation data eligible for the study. Indeed, the range of metadata (magnitude, distance, V_{S30}, etc) of the training dataset has to be available and compatible with the domains on which every GMPEs were established. The risk is to make decrease these metadata ranges. Another way to make the BMA approach deliver its full potential is to increase the dimension of the space of considered models by adding degrees of freedom. For instance, in addition to the prediction error term, the Bayesian calibration can be performed on the GMPE's coefficients as well. Several works already adopt that kind of approach: GMPE's coefficients can be deduced from data using an Artificial Neural Network (Derras et al., 2016 [24]), or with a Bayesian inference using the MCMC method (Kowsari et al., 2017 [35]). Several functional forms of GMPEs can be under consideration, their respective coefficients and prediction errors can then be calibrated using a MCMC algorithm, and these calibrated models can finally be combined with a BMA strategy. Furthermore, such an approach may allow to use observations whose features are not necessarily covered by GMPEs' validity domains. The overall idea is that the BMA method should be integrated in a dynamical process by adding any new observation, any new GMPE or functional form of GMPE when they become available so that the calibration, the ranking, the selection and combination of models are constantly updated in a machine learning process ensuring the best predictive performances.

5 Appendix

As a general rule, the MLE approximation of (14) can be solved with classical optimization techniques, such as Newton-Raphson or BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithms. But in some specific cases, where the distribution of the uncertain parameter is known and simple, the problem can be entirely solved analytically. In our case, a normal distribution is involved (6), that makes the computation straightforward. The solving of the following system:

$$\begin{cases} \dfrac{\partial}{\partial \mu_k} \displaystyle\sum_{n=1}^{N} \log g(d_n | M_k, \mu_k, \sigma_k) = 0 \\ \dfrac{\partial}{\partial \sigma_k} \displaystyle\sum_{n=1}^{N} \log g(d_n | M_k, \mu_k, \sigma_k) = 0, \end{cases} \tag{17}$$

provides the following analytical expressions for the optimal hyperparameters:

$$\mu_k^* = \frac{1}{N} \sum_{n=1}^{N} d_n - f_k(x_n), \tag{18}$$

and

$$\sigma_k^* = \left(\frac{1}{N} \sum_{n=1}^{N} (d_n - f_k(x_n) - \mu_k^*)^2 \right)^{1/2} . \tag{19}$$

The MLE estimation of the marginal likelihood is then obtained by using these expressions in (9):

$$\log L(M_k|D) \approx -N \left(\frac{\log 2\pi}{2} + \log \sigma_k^* \right) - \frac{N}{2} + c_p, \tag{20}$$

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 (12):

$$y(x)|M_k, D \sim \mathcal{N}\left(f_k(x) + \mu_k^*, \sigma_k^*\right) . \tag{21}$$

References

[1] Abrahamson, N., Atkinson, G., Boore, D., Bozorgnia, Y., Campbell, K., Chiou, B., Idriss, I.M., Silva, W., and Youngs, R. Comparison of the NGA Ground Motion Relations. *Earthq. Spectra*, 24, 45–66, 2008.

[2] S. Akkar and J. J Bommer. Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean region and the Middle East. *Seismological Research Letters*, 81(2):195–206, 2010.

[3] S. Akkar, M. A. Sandıkkaya, and J. J Bommer. Empirical ground-motion models for point- and extended-source crustal earthquake scenarios in europe and the middle east. *Bulletin of earthquake engineering*, 12(1):359–387, 2014.

[4] Akkar S., Sandıkkaya M.A., Åđenyurt M., Azari Sisi A., Ay B.Ö., Traversa P., Douglas J., Cotton F., Luzi L., Hernandez B., Godey S. Reference database for seismic ground-motion in Europe (RESORCE), Bulletin of Earthquake Engineering *Bulletin of Earthquake Engineering*, 12(1): 311–339, 2013. http://www.resorce-portal.eu/

[5] Conception des réacteurs à eau sous pression, published by the French Nuclear and Safety Authority (ASN) and the Nuclear and Radioprotection Institut (IRSN), July 18th, 2017, https://www.asn.fr/Professionnels/Installations-nucleaires/Guides-de-l-ASN/Guide-de-l-ASN-n-22-Conception-des-reacteurs-a-eau-sous-pression

[6] S. Baran and A. Moller. Joint probabilistic forecasting of wind speed and temperature using bayesian model averaging. *Environmetrics*, 26(2):120–132, 2015.

[7] Beauval C., Tasan H., Laurendeau A., Delavaud E., Cotton F., GuÃl'guen P. and Kuehn N. On the testing of ground-motion prediction equations against small-magnitude data. *Bulletin of the Seismological Society of America*, 102(5):1994–2007, 2012.

[8] Catherine Berge-Thierry, Fabrice Cotton, Oona Scotti, Daphné-Anne Griot-Pommera, and Yoshimitsu Fukushima. New empirical response spectral attenuation laws for moderate european earthquakes. *Journal of Earthquake Engineering*, 7(02):193–222, 2003.

[9] D. Bindi, L. Luzi, M. Massa, and F. Pacor. Horizontal and vertical ground motion prediction equations derived from the Italian Accelerometric Archive (ITACA). *Bulletin of Earthquake Engineering*, 8(5):1209–1230, 2010.

[10] D. Bindi, F. Pacor, L. Luzi, R. Puglia, M. Massa, G. Ameri, and R. Paolucci. Ground motion prediction equations derived from the italian strong motion database. *Bulletin of Earthquake Engineering*, 9(6):1899–1920, 2011.

[11] D. Bindi, M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, and P. Augliera. 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. *Bulletin of*

Earthquake Engineering, 12(1): 391–430, 2013.

[12] Bommer, J.J., Douglas, J., Scherbaum, F., Cotton, F., Bungum, H., FÃd'h, D. On the selection of ground-motion prediction equations for seismic hazard analysis *Seismological Research Letters*, 81:5, 783–793, 2010.

[13] D.M. Boore and G.M. Atkinson. 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*, 2007.

[14] David M Boore, Jonathan P Stewart, Emel Seyhan, and Gail M Atkinson. Nga-west2 equations for predicting pga, pgv, and 5% damped psa for shallow crustal earthquakes. *Earthquake Spectra*, 30(3):1057–1085, 2014.

[15] Jenno Brynjarsdottir and Anthony O'Hagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.

[16] C. Cauzzi and E. Faccioli. Broadband (0.05 to 20 s) prediction of displacement response spectra based on worldwide digital records. *Journal of Seismology*, 12(4): 453–475, 2008.

[17] Carlo Cauzzi, Ezio Faccioli, Manuela Vanini, and Aldo Bianchini. 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. *Bulletin of Earthquake Engineering*, 13(6):1587–1612, 2015.

[18] Nima Chitsazan and Frank T.-C. Tsai. A hierarchical bayesian model averaging framework for groundwater prediction under uncertainty. *Groundwater*, 53(2):305–316, 2015.

[19] A. C. Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58, 1968.

[20] Cotton, F., Scherbaum, F., Bommer, J. J., Bungum, H. Criteria for selecting and adjusting ground-motion models for specific target regions: Application to central Europe and rock sites. *Journal of Seismology*, 10:2, 137–156, 2006.

[21] Mary Kathryn Cowles and Bradley P. Carlin. Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, 91(434):883–904, 1996.

[22] Monika Culka. Applying bayesian model averaging for uncertainty estimation of input data in energy modelling. *Energy, Sustainability and Society*, 4(1):1–17, 2014.

[23] Elise Delavaud, Fabrice Cotton, Sinan Akkar, Frank Scherbaum, Laurentiu Danciu, Céline Beauval, Stéphane Drouet, John Douglas, Roberto Basili, M. Abdullah Sandikkaya, Margaret Segou, Ezio Faccioli, and Nikos Theodoulidis. Toward a ground-motion logic tree for probabilistic seismic hazard assessment in europe. *Journal of Seismology*, 16(3):451–473, Jul 2012.

[24] Boumédiène Derras, Pierre-Yves Bard, and Fabrice Cotton. Site-condition proxies, ground motion variability, and data-driven gmpes: Insights from the nga-west2 and resorce data sets. *Earthquake Spectra*, 32(4):2027–2056, 2016.

[25] E Faccioli, A Bianchini, and M Villani. 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, March*, 17–18, 2010.

[26] Geisser, S. Predictive Inference: An Introduction. *NY: Chapman and Hall*, ISBN 0-412-03471-9, 1993.

[27] Gelman, A., and D. B. Rubin. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7: 457–511, 1992.

[28] Hastings, W. Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57: 97–109, 1970.

[29] Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors). *Statist. Sci.*, 14(4):382–417, 1999.

[30] Specific Safety Guide SSG-9, Seismic Hazards in Site Evaluation for Nuclear Installations. http://www-pub.iaea.org/MTCD/publications/PDF/Pub1448_web.pdf, *International Atomic Energy Agency*, 2010.

[31] Jaynes, E. T. Prior Probabilities. *IEEE Trans. on Systems Science and Cybernetics*, SSC-4, 227, 1968.

[32] J. D. Jobson Applied Multivariate Data Analysis, p.63, *Springer Texts in Statistics*, 1991.

[33] Ö. Kale and S. Akkar. A method to determine the appropriate gmpes for a selected seismic prone region. In *15th World Conference on Earthquake Engineering*, 2012.

[34] R.E. Kass and A.E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[35] Milad Kowsari, Benedikt Halldorsson, and Jónas Snæbjörnsson. On the Manifestation of Ground Motion Model Differences on Seismic Hazard in North Iceland. In *International Conference on Earthquake engineering and Structural Dynamics*, 2017.

[36] R. K. McGuire. Methodology for incorporating parameter uncertainties in seismic hazard analyses for loss risk design intensities. In *Internat. Symposium on Earthquake Structural Engineering*, 1976.

[37] Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. Equations of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, 1953.

[38] NUREG/CR-6372 Prepared by Senior Seismic Hazard Analysis Committee (SSHAC) R. J. Budnitz (Chairman), G. Apostolakis, D. M. Boore, L. S. Cluff, K. J. Coppersmith, C. A. Cornell, P. A. Morris. Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts, *US NRC*, 1997.

[39] Inseok Park and Ramana V. Grandhi. A bayesian statistical method for quantifying model form uncertainty and two model combination methods. *Reliability Engineering & System Safety*, 129: 46–56, 2014.

[40] E. Wang W. Zhao R. Zhu, H. Zheng. Multi-model ensemble simulation of flood events using bayesian model averaging. In *20th International Congress on Modelling and Simulation, Adelaide, Australia*, 2013.

[41] Adrian Raftery. Bayesian model selection in social research (with discussion by andrew gelman & donald b. rubin, and robert m. hauser, and a rejoinder), 1995.

[42] Frank Scherbaum, Elise Delavaud, and Carsten Riggelsen. Model selection in seismic hazard analysis: An information-theoretic perspectivemodel selection in seismic hazard analysis: An information-theoretic perspective. *Bulletin of the Seismological Society of America*, 99(6):3234, 2009.

[43] Anneli Schoniger, Thomas Wohling, and Wolfgang Nowak. A statistical concept to assess the uncertainty in bayesian model weights and its impact on model ranking. *Water Resources Research*, 51(9):7524–7546, 2015.

[44] E.M. Scordilis. Empirical Global Relations Converting MS and mb to Moment Magnitude. *Journal of Seismology*, 10(2): 225–236, 2006.

[45] P.J. Stafford, F.O. Strasser and J.J. Bommer. An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region. *Bulletin of Earthquake Engineering*, 6(2): 149–177, 2008.

[46] Jonathan P. Stewart, John Douglas, Mohammad Javanbarg, Yousef Bozorgnia, Norman A. Abrahamson, David M. Boore, Kenneth W. Campbell, Elise Delavaud, Mustafa Erdik, and Peter J. Stafford. Selection of ground motion prediction equations for the global earthquake model. *Earthquake Spectra*, 31(1):19–45, 2015.

[47] Jasper A. Vrugt, Cajo J. F. ter Braak, Martyn P. Clark, James M. Hyman, and Bruce A. Robinson. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with markov chain monte carlo simulation. *Water Resources Research*, 44(12), 2008.

[48] J. Selva W. Marzocchi, M. Taroni. Accounting for epistemic uncertainty in psha: Logic tree and ensemble modeling. *Bulletin of the Seismological Society of America*, 105(4):2151–2159, 2015.

[49] R.P. Dwight W.N. Edeling, P. Cinnela. Predictive rans simulation via bayesian model-scenario averaging. *Journal of Computational Physics*, 275:65–91, 2014.

[50] John X Zhao, Jian Zhang, Akihiro Asano, Yuki Ohno, Taishi Oouchi, Toshimasa Takahashi, Hiroshi Ogawa, Kojiro Irikura, Hong K Thio, Paul G Somerville, et al. Attenuation relations of strong ground motion in japan using site classification based on predominant period. *Bulletin of the Seismological Society of America*, 96(3):898–913, 2006.

[51] Gaofeng Zhu, Xin Li, Kun Zhang, Zhenyu Ding, Tuo Han, Jinzhu Ma, Chunlin Huang, Jianhua He, and Ting Ma. Multi-model ensemble prediction of terrestrial evapotranspiration across north china using bayesian model averaging. *Hydrological Processes*, 30(16):2861–2879, 2016.