Inferring earthquake ground motion fields with Bayesian Networks

Pierre Gehl*[1], John Douglas[2], Dina D'Ayala[3]

[1]BRGM/DRP, 3 avenue C. Guillemin, BP 36009, 45060 Orléans Cedex 2, France

*=corresponding author: p.gehl@brgm.fr

[2]University of Strathclyde

[3]University College London

Abstract

Bayesian Networks (BNs) have the ability to perform inference on uncertain variables given evidence on observed quantities, which makes them relevant mathematical tools for the updating of ground-motion fields based on strong-motion records or macroseismic observations. Therefore the present article investigates the use of BN models of spatially correlated Gaussian random fields as an accurate and scalable method for the generation of ground-motion maps. The proposed BN model is based on continuous Gaussian variables, as opposed to discrete variables as in previous formulations, and it is built to account for cross-correlated ground-motion parameters as well as macroseismic observations. This approach is validated with respect to the analytical solution (i.e., conditional multivariate normal distributions) and it is also compared to the USGS ShakeMap method, thus demonstrating a better ability to model jointly the inter- and intra-event error terms of ground-motion models. The scalability of the approach, i.e. its capacity to be applied to large grids, is ensured by a grid sub-division strategy, which appears to be computationally efficient and accurate within an error rate of a fraction of percent. Finally, the BN implementation is demonstrated on a real-world example (the Mw 6.2 Kumamoto, Japan, 2016 foreshock), where vector-valued shake-maps of cross-correlated intensity measures are generated, along with the integration of macroseismic observations.

INTRODUCTION

Over the past decade, rapid loss assessment following earthquakes has emerged as a crucial research topic, with the objective of providing emergency responders and critical facility operators with accurate estimates of intensity levels or probable damage across the affected area (e.g., Wald et al., 2008; Erdik et al., 2011). For instance, at the hazard level, the updating of the spatially-distributed ground-motion field, or ground-motion map, is achieved by combining estimates from ground-motion prediction equations (GMPEs) and field observations (Wald et al., 2005; Worden et al., 2010). A comparison of the most common statistical techniques is provided by Douglas (2007) for the Les Saintes (Guadeloupe, France) 2004 earthquake. Worden et al. (2010) also provide valuable insights into pending issues, namely: the treatment of uncertainties near the observations, the quality of the estimates for poorly-observed events and the computation of joint distributions for correlated intensity measures (IMs). A rigorous probabilistic analysis of the relation between macroseismic intensity and peak ground acceleration (PGA) has been proposed by Ebel and Wald (2003), but without accounting for correlation between spatially-distributed ground motions.

The inference abilities of Bayesian Networks (BNs) appear to be appealing for such a problem because they use observations as evidence in order to update directly the prior distributions of various variables, such as estimates from GMPEs or the damage distribution (Jaiswal et al., 2011). The application of BNs to earthquake engineering has been formalized by Bensi et al. (2011a) for the analysis of infrastructure systems of interdependent elements, which requires the estimation of statistics for joint events over spatially-distributed assets. Besides forward risk analyses (Bensi et al., 2013), BNs may also be used for the backward analysis of a system when a partial knowledge of losses is available immediately after an earthquake (e.g., Pozzi and Der Kiureghian, 2013; Gehl et al., 2017). Most proposed BN formulations are, however, hampered by scalability and computational issues, which complicate their application to real-world systems (Cavalieri et al., 2017).

Therefore, the present paper builds upon the original BN approach by Bensi et al. (2011a), while applying the Bayesian framework to the ground-motion assessment part only. It is expected that the removal of the variables related to damage and system performance estimation will greatly reduce the computational difficulties, mostly by enabling the use of continuous Gaussian BNs, as opposed to the discrete BNs used in previous studies. Moreover, the BN formulation is augmented with additional variables representing secondary cross-correlated IMs and even macroseismic intensities, so that the Bayesian updating can be performed with diverse sources of field observations. The proposed developments pursue multiple objectives: *(i)* to demonstrate the accuracy of the BN approach for the generation of ground-motion maps, which is a pre-requisite before complete BNs enabling loss estimation may be used in a decision support system, *(ii)* to verify the feasibility and scalability of the BN approach for large spatial grids in the case of real-world earthquakes, and *(iii)* to investigate the potential benefits that can be gained from inferring ground-motion fields with a BN, especially in terms of uncertainty treatment (e.g., joint updating of inter- and intra-event error terms) and the generation of maps for vector-valued intensity measures. The underlying equations necessary to model the formulation of the proposed BN are presented in Section "Demonstration of the Bayesian network approach", which contains also a comparative analysis of the BN approach with respect to the well-established ShakeMap algorithm (Worden et al., 2016). The scalability issue is addressed in Section "Computational performance", where a sub-grid division strategy is investigated to ensure the stability of the BN. Finally, Section "Application to the M_w 6.2 Kumamato Earthquake (April 14 2016)" applies the BN approach to a specific event, the M_w 6.2 earthquake on 14 April 2016 near Kumamoto, Japan, thus providing an opportunity to demonstrate the implementation of the BN on an actual earthquake and to analyze the information gain when considering multiple cross-correlated intensity measures.

DEMONSTRATION OF THE BAYESIAN NETWORK APPROACH

This section provides details on the construction of the BN and an investigation of its validity with respect to other methods for the generation ground-motion maps.

Proposed approach for the construction of the Bayesian Network

The distribution of a given strong motion parameter or IM over a given geographical grid is usually estimated from the following variables (Crowley and Bommer, 2006):

- M_w, the moment magnitude of the earthquake event;
- *Epi*, the location of the epicenter of the earthquake, if a point-source event is assumed, or the rupture location and extent for finite-fault scenarios;
- Other parameters such as the faulting mechanism, the fault geometry and the depth to top of rupture, depending on the specific GMPE that is used;
- X_i, the logarithm of the median estimate of the IM at the grid point *i*, as predicted by the selected GMPE (i.e., $X_i = \ln \overline{IM_i}$);
- η, the inter-event (or between-event) error term from the GMPE;

- ζ, the intra-event (or within-event) error term from the GMPE;
- Y_i, the logarithm of the IM distribution at the grid point i, accounting for the aleatory variability generated by the GMPE error terms (i.e., $Y_i = \ln IM_i$);

It should be noted that X_i, η and ζ depend on the GMPE chosen and hence they are a function of its database, functional form and the technique used for its derivation. There can be considerable differences in these variables depending on the GMPE chosen (epistemic uncertainty), particularly at the edges of their applicability (e.g., large magnitudes and close source-to-site distances) (Douglas and Edwards, 2016). When there are few observations these differences would map to large differences in the ground-motion fields estimates. However, when dense observations exist the BN method presented below would lead to these differences being reduced and the choice of the original GMPE would then be less important.

According to Crowley et al. (2008a) and Park et al. (2007), the same inter-event variability should be applied to all grid points within a given earthquake scenario, while the joint distribution of the intra-event term should follow the spatial correlation among grid points. As shown by Bensi et al. (2011b), representing the dependency among grid points is facilitated by a Cholesky factorization of the correlation matrix. Let us assume a grid of n points, where the variability of the intra-event term is represented by a correlated Gaussian random field defined by standard normal variables Z_i at grid points i. The proposed decomposition is then performed as follows:

$$\mathbf{Z} = \mathbf{T} \cdot \mathbf{U} \tag{1}$$

where the $n \times n$ transformation matrix $\mathbf{T}$ is a lower triangular matrix obtained through a Cholesky factorization, so that $\mathbf{R} = \mathbf{T}.\mathbf{T}^{\mathrm{T}}$, with $\mathbf{R}$ being the correlation matrix of each couple of the grid points. The $n \times 1$ vector $\mathbf{U}$ represents the standard normal variables, which are statistically independent from each other and are used to model the variation in the correlation among the grid points. The correlation matrix $\mathbf{R}$ is built thanks to a spatial correlation model, such as the one proposed by Jayaram and Baker (2009), where the correlation coefficient ρ_{ij} between the ground-motion parameters at two sites i and j is expressed as:

$$\rho_{ij} = \exp\left(-\frac{3 \cdot r_{ij}}{b}\right) \tag{2}$$

where r_{ij} is the distance between the sites and b is the correlation distance, assumed here to be equal to 13.5 km for PGA, which is consistent with recent studies on spatial correlation (e.g., Jayaram and Baker, 2009; Esposito and Iervolino, 2011).

When the ground-motion field is generated to estimate losses for various types of assets, such as an infrastructure system, the method may need to provide estimates for more than one IM, depending on the type of fragility models used. Therefore, the cross-correlation between the IMs of interest must be taken into account when computing their joint distribution. When modelling a ground-motion field of n_{IM} cross-correlated IMs over n sites, the corresponding correlation matrix must be of the order of $n_{IM} \times n$, if it is directly used in Equation 1 (Weatherill et al., 2014). Therefore, because this matrix can rapidly become large, Weatherill et al. (2014) advocate the use of a sequential simulation method, which first generates a field of primary IMs, represented by the correlated vector $\mathbf{Z_1}$ of standard normal variables. Then, the field of secondary IMs, represented by the correlated vector $\mathbf{Z_2}$ of standard normal variables, is conditioned upon the distribution of the primary IMs. These variables may then be expressed as follows (Oliver, 2003):

$$\begin{cases} \mathbf{Z}_1 = \mathbf{T}_1 \cdot \mathbf{U}_1 \\ \mathbf{Z}_2 = \mathbf{T}_2 \cdot \left(\rho_{12} \cdot \mathbf{U}_1 + \sqrt{1-\rho_{12}^2} \cdot \mathbf{U}_2\right) \end{cases} \tag{3}$$

where $\mathbf{T_1}$ and $\mathbf{T_2}$ are the $n \times n$ triangular transformation matrices that are factorized from the correlation matrices $\mathbf{R_1}$ and $\mathbf{R_2}$, for the primary and secondary IMs respectively. $\mathbf{U_1}$ and $\mathbf{U_2}$ are $n \times 1$ vectors of independent standard normal variables. Finally, ρ_{12} represents the cross-IM correlation coefficient between the primary and the secondary IMs.

The proposed BN structure corresponding to the above detailed variables is presented in Figure 1. The selected GMPE directly establishes a deterministic relationship between M_w, *Epi* and $\overline{IM_i}$ at site *i*. For this study we assume here that the magnitude and epicentre are known for a given earthquake.

Hence, the BN structure may be greatly simplified with respect to the original BN formulation by Bensi et al. (2011a): only the variables that have a probabilistic dependency between each other are displayed, namely Y_i, W and U_i (representing η and ζ).

[Figure 1 about here]

Since all the BN variables may be expressed as normal distributions (i.e., W and U_i are standard normal variables, and the normal distribution of the parameters Y_i is a very common assumption in ground-motion prediction), it is possible to define the BN in Figure 1 as a Gaussian Bayesian Network (GBN), as introduced by Murphy (2002). In this case, all BN nodes become continuous normal variables with parameters expressed as a linear combination of the values of the parent nodes. In the proposed example, the root nodes U_i and W are defined by a marginal distribution (i.e., normal probability density function represented by N):

$$\begin{cases} p(U_i) = N\,(0,1) \\ p(W) = N\,(0,1) \end{cases} \tag{4}$$

Meanwhile, the conditional distribution of the child nodes Y_i (i.e., $Y_{1,i}$ as primary IM and $Y_{2,i}$ as secondary IM) is expressed as follows:

$$\begin{cases} p(Y_{1,i} | \mathbf{U}_1, W) = N\left(X_{1,i} + \sigma_\varsigma \cdot \sum_{j=1}^{n} t_{ij}^{(1)} \cdot U_j + \sigma_\eta \cdot W, \varepsilon^2 \right) \\ p(Y_{2,i} | \mathbf{U}_1, \mathbf{U}_2, W) = N\left(X_{2,i} + \sigma_\varsigma \cdot \rho_{12} \cdot \sum_{j=1}^{n} t_{ij}^{(2)} \cdot U_j + \sigma_\varsigma \cdot \sqrt{1-\rho_{12}^2} \cdot \sum_{j=1}^{n} t_{ij}^{(2)} \cdot U_{n+j} + \sigma_\eta \cdot W, \varepsilon^2 \right) \end{cases} \tag{5}$$

where σ_ζ and σ_η respectively represent the standard deviations of the intra-event and inter-event error terms, which are provided by the GMPE. The coefficients $t^{(1)}{}_{ij}$ and $t^{(2)}{}_{ij}$ respectively represent the elements of the transformation matrices $\mathbf{T_1}$ and $\mathbf{T_2}$. If it is assumed that $Y_{1,i}$ and $Y_{2,i}$ are completely determined by knowledge of U_i and W, a value close to zero has to be assumed for their standard deviation ε to achieve convergence. A standard deviation corresponding to the record-to-record variability may also be assigned, if it is specified by the GMPE.

In the case that an earthquake event is recorded by a set of accelerometers, the recorded ground motions may be used to update the predicted ground-motion field. Thanks to the proposed Bayesian approach, an inference can be performed through the U_i and W variables, which are used to pass the message to the neighboring sites. To this end, the original BN formulation is augmented with the addition of the nodes representing the observed ground motions (i.e., red nodes and edges in Figure 1), which are then used as evidence for the Bayesian inference. It can be seen, therefore, that the spatial correlation structure between the IMs plays a major role in the propagation of the observations to the grid points in the vicinity. Such a BN has the merit of providing probabilistic distributions of the ground-motion estimates, while ensuring that the joint distribution of the predicted parameters complies with the spatial correlation of the intra-event residuals.

Once the Y_i distributions are obtained at the grid points, they may be interpolated at the locations of the vulnerable sites (e.g., built areas or infrastructure elements), while local amplification factors may also be added to account for site effects. The expression of the problem as a GBN has the merit of manipulating only continuous variables, which do not require a preliminary discretization and the creation of conditional probability tables that grow exponentially with the number of parents.

Single-IM Bayesian inference on a synthetic example

A trivial synthetic example is introduced in order to demonstrate how the ground-motion map is updated with the BN approach. It consists of a 3 x 3 square grid (grid step = 1 km) with a M_w 5.5 earthquake occurring in its vicinity (at coordinates [-3; 5]), while two ground-motion records are assumed to be available (see the spatial configuration in Figure 2a): the two observations (i.e., *Yobs1* and *Yobs2*) are assumed to be 15% smaller and 10% larger than the predictions, respectively. For simplification purposes, only a single IM is considered here, which is the PGA estimated using the GMPE of Chiou and Youngs (2008).

[Figure 2 about here]

The corresponding BN is detailed in Figure 2b, where the link structure between U_i and *Y(i)* variables is characteristic of the triangular transformation matrix **T**, following the Cholesky decomposition. This

BN structure, consisting of a table describing the directed links between the variables and of normal distribution parameters for each variable (see Equations 4 and 5), is then implemented in the Bayes Net toolbox (see Data and Resources). The junction-tree algorithm, which carries out exact inference and thus provides exact probability distributions, is used within the toolbox. This algorithm consists in the following steps:

- Moralization of the BN: all edges are represented as undirected links, and all the parents of a same node are linked by a new undirected edge, if they were not previously linked.
- Variable elimination: each node is successively removed while its adjacent nodes are connected through additional undirected edges (i.e., fill-in edges), if they were not previously linked. Then a clique is formed by the eliminated node and all its adjacent nodes.
- Once all variables have been eliminated, the cliques are assembled into a junction tree (see Figure 3).
- The potential of each clique (i.e., joint probability distribution of the variables within the clique) is computed by multiplying the marginal and conditional Gaussian distributions that are associated with the variables (see Equations 4 and 5).

[Figure 3 about here]

Once the junction tree is built, the BN is considered as initialized and it can be used to perform inference on any scenarios. In the proposed example, the evidence is set on the *Yobs1* and *Yobs2* variables and propagated through the junction tree, as shown in Figure 3. The evidence propagation is carried out in two successive stages:

- Evidence collection: the evidence is collected from the leaves of the junction tree to the root clique. Operations of probability marginalization (i.e., removal of a variable) and multiplication are performed in order to update the potential of the root clique.
- Evidence distribution: the evidence is distributed from the root clique to all cliques along the junction tree. Operations of probability marginalization, division and multiplication in order to update the remaining cliques.

The posterior probability distribution can then be observed for any variable of interest. For instance, the updated distribution of variable *Y(1)* is obtained by marginalizing the potential of the clique [*U1* ; *W* ; *Y(1)*] with respect to *Y(1)*. The prior and posterior distribution parameters of the variables involved in the synthetic example are summarized in Table 1.

[Table 1 about here]

As expected, the ground-motion grid is modified by the field observations, i.e. lower values are found towards the lower left of the grid where the assumed observation *Yobs1* is lower than the initial prediction. An analysis of the distributions of the BN variables after the inference reveals two complementary levels of updating (Figure 4):

- On a global level, the distribution of the *W* variable, which represents the inter-event error η that is common to all grid points, is updated to provide a biased GMPE prediction that balances the general under- or over-estimation of the ground motion when compared with the observations. In the present example, the two hypothetical ground-motion records are globally lower than the initial GMPE estimates with an unbiased inter-event error: as a result, the variable η is updated to account for the observed bias; the standard deviation σ_η is also reduced, even though it does not converge towards zero due to the limited number of observations.
- On a local level, the distribution of the U_i variables, which are used to map the spatially-correlated intra-event errors ζ_i, is updated in order to match the local variations of the ground motion in the vicinity of each of the two hypothetical stations. For instance, the closest grid point to observation #1 is heavily influenced by the parent variable U_1 according to the corresponding element in the transformation matrix **T** (i.e., $t_{1,1}$ = 1). Therefore, the posterior distribution of U_1 is shifted towards the left to represent over-estimation of PGA by the initial GMPE prediction when compared to the observation. The same effect is observed for the grid points close to observation #2, where the recorded PGA is higher than the initial GMPE

prediction: the distribution of U_8, which has a strong weight in the transformation matrix (i.e., $t_{8,8}$ = 0.488) with respect to grid point *Y(8)*, is therefore shifted towards the right.

[Figure 4 about here]

Comparison with current ground-motion map methods

The BN-updated ground-motion field is first compared with the ShakeMap algorithm (Worden and Wald, 2016), developed by the U.S. Geological Survey, which has proven its operational abilities to deliver ground-motion maps in near real-time. The main principles of this algorithm are summarized as follows, in the case of a basic ground-motion map using strong-motion data only (i.e., no conversion between macroseismic intensity and ground-motion parameters):

- Removal of the potential site amplification factors from the observed ground motions (i.e., correction to "rock" site).
- Computation of the global bias introduced by the recorded ground motions with respect to the initial GMPE estimates, and use of a bias-adjusted GMPE for the prediction at the grid points. This adjustment is achieved by finding the M_w magnitude that reduces the errors between the observed and the predicted ground-motions, when the GMPE is evaluated for the adjusted magnitude.
- Interpolation of the observations to the grid points.
- At each grid point, updating of the ground motion through a weighted average between the bias-adjusted GMPE estimate and the observations (Worden et al., 2010). The GMPE estimate is weighted by the inverse of the variance provided by the GMPE, while each observation is weighted by the term $1/\sigma^2_{obs}$ (i.e., σ_{obs} is the standard deviation assigned to the observation — it increases with the distance between the observation and the grid point based on a correlation model).
- Application of potential site amplification factors at the grid points.

In the ShakeMap method, the total standard deviation associated to each grid point is obtained as a byproduct of the interpolation process (Worden et al., 2010):

$$\sigma_{\ln PGA} = \sqrt{\frac{1}{\frac{1}{\sigma_{GMPE}^2} + \sum_{j=1}^{k} \frac{1}{\sigma_{obs,j}^2}}} \tag{6}$$

where $\sigma_{obs,j}$ is the standard deviation assigned to the j^{th} observation and σ_{GMPE} is the standard deviation of the GMPE estimate. As stated by Worden et al. (2010), if enough observations are used to update the event magnitude, σ_{GMPE} may be taken as equal to σ_{ζ} (i.e., intra-event variability only), which has been assumed here.

In order to apply the ShakeMap algorithm, one must ensure that comparable correlation models are applied to both ShakeMap and BN-based approaches (Equation 2). In the ShakeMap method, the influence of the observations on the grid predictions is modelled by a weighting function $F = \sigma_{obs}/\sigma_{\text{GMPE}}$, which tends towards zero when the inter-site distance is zero, and towards infinity for large distances. However, it appears that using the influence distances that are advocated in the ShakeMap manual (i.e., r_{ROI} = 10 km and r_{MAX} = 15 km) corresponds to correlation coefficients that are much larger than the ones generated by Equation 2, with b = 13.5 km. Some numerical tests have shown that, in order to yield comparable results, the weighting function should follow the same shape as the proposed correlation model, with an exponentially decreasing rate. The mathematical link between the weighting function F and the correlation coefficient ρ is investigated in Appendix A, where equivalent models for both the ShakeMap and the BN approaches are presented.

On the other hand, the use of the spatial correlation matrix in the BN approach in order to update the intra- and inter-event error distributions is conceptually similar to the analytical resolution of a conditional multivariate normal distribution, as proposed by Vanmarcke (1983) and Stafford (2012). This method directly computes the means and standard deviations of the intra- and inter-event error terms through vector and matrix multiplications. Therefore, it is straightforward to implement and fast to execute. All three methods are applied to the synthetic example, for the updating of PGA distributions across a 3 x 3 grid: the updated ground-motion values and their associated uncertainties are detailed in Table 2.

[Table 2 about here]

Regarding the prediction of the ground-motion means, it appears first that the BN results are identical to the analytical solution, thus validating the accuracy of the proposed approach. The ShakeMap method, however, does not provide exactly the same means, which might be due to the way the updating is modelled, i.e., through an interpolation instead of accounting for the full spatial correlation matrix. Moreover, both BN and analytical methods provide the same value for the inter-event error term η and its standard-deviation, while the ShakeMap approach results in a lower value with zero uncertainty: this is explained by the bias removal through an optimization process, without accounting for the prior distribution and the associated likelihood function, as opposed to conditional updating methods.

On the other hand, the uncertainties in Table 2 are not exactly identical between the different methods, since the analytical method yields higher standard-deviations than both the BN and ShakeMap methods, especially for grid points that are close to observations. In the BN method, the global uncertainty appears to be sometimes lower than the inter-event standard-deviation σ_η, which is in contradiction to the widespread assumption of the quadratic combination of inter- and intra-event dispersions. In order to investigate this aspect, another synthetic example is considered, where points are fixed along a line at increasing distances from a given observation, thus measuring the evolution of the ground-motion uncertainty with distance (see Figure 5).

[Figure 5 about here]

Many noteworthy observations can be made from this result:

- Far from the observation, both BN and analytical solutions converge to the same asymptote, which correspond to the quadratic combination of the updated inter-event dispersion and the far-field intra-event dispersion. Conversely, the global dispersion is underestimated by the

ShakeMap method, which converges toward the intra-event dispersion only: this result is due to the assumption that sufficient observation points lead to an inter-event term with zero dispersion (Worden et al., 2010), while the alternative methods have shown that this is not necessarily the case.

- Close to the observation, both BN and ShakeMap methods yield a global dispersion that tends towards zero: this behavior is consistent with a correlated Gaussian random field, where predictions in the immediate vicinity of an observation are almost certain, with negligible dispersion. On the other hand, if the analytical solution is used, the dispersion tends towards the inter-event standard-deviation σ_η, which results in predictions that keep a significant dispersion even when very close to an observation. This discrepancy is explained by the following rationale:
  - The analytical solution uses a two-step set of separate equations to compute the updated distributions of inter- and intra-event error terms.
  - These terms may then be used to compute the global dispersion thanks to a quadratic combination, under the assumption that the variables are independent.
  - However, it appears that the intra-event error terms are dependent on the estimation of the inter-event error, thus breaking the independency assumption and preventing the use of the quadratic combination. Therefore, while the analytical solution is perfectly valid for the separate estimation of inter- and intra-event error terms, it does not provide any means of accurately computing the global dispersion.
  - The BN method, on the other hand, implicitly accounts for the correlation between the intra- and inter-event residuals through the multiplication of conditional probabilities: as a result, a stronger correlation close to an observation leads to a smaller global dispersion, which ultimately tends towards zero.

Therefore, the comparison between the different approaches has demonstrated that the BN method is as accurate as the analytical solution for the updating of mean values, while the ShakeMap method cannot provide the same values due to its interpolation scheme that is conceptually different to the use of spatial correlation models. Moreover, in terms of uncertainties, the BN method also provides the best solution to account for both inter- and intra-event dispersions, whether the predictions are made close or far from an observation.

COMPUTATIONAL PERFORMANCE

The following sub-sections study the feasibility of the proposed approach for large spatial grids, while different strategies are investigated in order to facilitate its use in real-world applications. If such a method is to be used to develop an operational decision support system, it has to be accurate enough (i.e., high resolution grid) over a spatial extent that covers most of the earthquake's effects, thus possibly leading to a huge correlation matrix. Moreover, such a system is expected to deliver updated ground-motion fields almost immediately after the occurrence of an earthquake, in order to provide situational awareness to emergency responders.

Scalability

As stated above, the BN has been implemented in the Bayes Net toolbox, which enables the inference of GBNs through a junction-tree algorithm. Thanks to the Gaussian formulation that enables the use of continuous variables, the computation time is expected to remain much lower than the same BN structure with discrete variables, which would lead to the creation of conditional probability tables and clique potentials with an intractable number of elements. The execution time of a single inference operation (i.e., updating of one **Y** node) is detailed in Figure 6 for different grid sizes, for a single IM prediction (i.e., no secondary IM). As expected, the computational load increases exponentially with the number of grid points, even if the execution time remains tractable for a large grid containing 400 points. In Figure 6 (right), the computation time is represented with respect to the number of $\mathbf{U} \rightarrow \mathbf{Y}$ links that are required in the BN. The almost-linear relation between these two indicators shows that they are closely related; therefore, the explosion in computational times is mostly due to the proposed BN formulation, which is associated with an exponential increase of links with respect to the number of nodes.

[Figure 6 about here]

As a result, even with the use of GBNs, the proposed BN approach is eventually bound to reach its limits for very large grids, usually due to elongated computation times that no longer meet the demands of a near real-time information system. This issue becomes especially pressing when high-resolution maps are required, e.g. grid steps around 1 km for areas spanning several hundreds of kilometers, which would lead to tens of thousands of grid points.

Optimization strategies

To make the problem tractable, several optimized BN formulations for correlated Gaussian random fields have been proposed by Bensi et al. (2011b), who have found that a numerical optimization of an approximate transformation matrix $\hat{\mathbf{T}}$ results in a better computational performance than a Cholesky decomposition. This optimization starts by specifying a number m of **U** nodes to keep in the BN, so that the approximation of the correlated Gaussian random field can be expressed as follows:

$$\hat{\mathbf{Z}} = \hat{\mathbf{T}} \cdot \mathbf{U}_m + \mathbf{S} \cdot \mathbf{V} \tag{7}$$

where $\hat{\mathbf{T}}$ is the approximated $n \times m$ transformation matrix, **V** is a $n \times 1$ vector of independent standard normal variables and it is multiplied by a diagonal $n \times n$ transformation matrix **S**, whose elements s_i are used to correct the global variance of the variables in $\hat{\mathbf{Z}}$:

$$s_i = \sqrt{1 - \sum_{k=1}^{m} \hat{t}_{ik}^2} \tag{8}$$

This approximation may be seen as a generalization of a Dunnett–Sobel (DS) class of Gaussian random variables: the $\hat{t}_{ik}$ elements are found through a numerical optimization, with the objective of minimizing the difference between the actual correlation matrix **R** and its approximation $\hat{\mathbf{T}} \cdot \hat{\mathbf{T}}^t$. Finally, once the transformation matrix has been optimized, further simplifications may be carried out, such as the removal of nodes (i.e., columns in $\hat{\mathbf{T}}$) or links (i.e., elements in $\hat{\mathbf{T}}$) from the BN: all these elimination strategies, based on importance measures checking the respective influence of each variable, are detailed in Bensi et al. (2011b). This strategy is tested on the synthetic example detailed above, with the aim of comparing its accuracy and computational efficiency. The scalability is also investigated by increasing the grid extent with different scenarios (see Table 3). The discrepancy between the exact solution (i.e., BN with Cholesky decomposition over the full grid) and the various approximations is measured with two metrics, namely the average of the absolute errors in predicting the PGA mean over the grid points, and the maximum error.

[Table 3 about here]

It appears that the optimization strategy does not lead to any computational time gain, even at the cost of less accurate results (i.e., around 20% error rate on the prediction of the mean PGA for larger grids). For smaller grids, the accuracy loss is negligible; however, as the number of grid points increases relatively to the number m, the quality of the approximation diminishes greatly, unless a very costly optimization is carried out with a large enough m value. This observation is in strong contrast with the original findings by Bensi et al. (2011b), who have shown that this optimization significantly reduces the computational time while maintaining a reasonable accuracy. In the present concept, two main conceptual differences with the work by Bensi et al. (2011b) explain this difference:

- Continuous GBNs are implemented here instead of discrete BNs, so that the initial computational bottlenecks (i.e., size of CPTs and cliques) are now removed to some extent and that the size of the correlation matrix is one of the main issues. The initialization time is displayed in Table 3, thus revealing how difficult it is to perform the numerical optimization of a large matrix with a large number m of variables. The BN inference time becomes less essential, although it should be noted that adding the **V** nodes has the effect of slightly increasing computational costs.
- The objective here is to perform a backward analysis (i.e., inference from an observation to other **Y** nodes), while only the accuracy of a forward analysis has been investigated in Bensi et al. (2011b). Backward analyses are more complex (i.e., message passing through many nodes) and they require a highly accurate correlation matrix.

Alternatively, a more radical and straightforward strategy is proposed in the present paper, where the initial grid containing $n \times n$ points is divided into k sub-grids of $m \times m$ points, where $k = (n / m)^2$. As a result, k BNs need to be created and solved before all the predictions at the grid points are aggregated and projected on the same map (Figure 7).

[Figure 7 about here]

As shown in Figure 7, all observations must be used as evidence within each BN in order to ensure that the updating of the variables is at the same level for each sub-grid. This approach is justified by the fact that the BN inference appears to be robust with respect to the number and location of the Y_i variables, as long as the evidence nodes remain unchanged (i.e., whatever the extent of the grid, the updated ground-motion field should remain stable). This grid subdivision may be seen as an extreme case of the numerical optimization detailed above, in the sense that grid points from two different sub-grids are similar as nodes between which links have removed. However, the main difference lies in the fact that all observations are kept for all sub-grids in order to maintain the same inference across all sub-grids. This strategy is tested on the largest grid of the synthetic example (i.e., 24 x 24 grid with 16 observations), which is divided in different sets of sub-grids: the average and maximum error measures are estimated for both the PGA mean and standard-deviation (see Table 4).

[Table 4 about here]

The grid sub-division provides a dramatic decrease in computational times, while the accuracy of the estimations remains very high and stable, i.e. always below 0.1% of maximum error. This preliminary observation is investigated further by checking the evolution of the error rate with the correlation length, defined as $d_{corr} = b/3$ if the correlation model from Equation 2 is used. The spatial extent of the sub-grid with respect to the correlation distance appears to govern slightly the evolution of the error rate. For small correlation distances, there are almost no differences between the various grids; while greater error rates, albeit still very small, may be observed when the grid extent (i.e., its total dimension) becomes much smaller than the correlation distance. Globally, these small deviations from the initial grid appear to be negligible, especially when considering that such a strategy enables almost any map size and resolution to be handled, with few computational constraints. Additional overlapping sub-grids might also be considered in order to correct any boundary effects; however the present configuration, with standard correlation lengths (e.g., up to a couple of dozen km) and 1 km grid steps, results in excellent accuracy, especially when compared to the much larger error rates obtained with the optimized transformation matrix. One significant caveat, however, is that the use of sub-grids requires the construction of independent BNs, thus preventing the computation of joint statistics for locations that do not belong to the same sub-grid: such a feature is essential in the context of infrastructure risk analysis (i.e., presence of interdependent assets at various locations). It may be overlooked, however, if the main objective is to generate a ground-motion map following an earthquake.

APPLICATION TO THE M_w 6.2 KUMAMATO EARTHQUAKE (APRIL 14 2016)

The inference abilities of the proposed BN approach are demonstrated in the following sub-sections, where strong-motion data from the M_w 6.2 earthquake that occurred near Kumamoto (Japan) on April 14th 2016 (this was the foreshock of the destructive M_w 7.0 event that occurred two days later in the same region) are exploited. This earthquake was recorded by a dense network of strong-motion stations in the near field. Its smaller magnitude than the mainshock enables a point-source event to be assumed.

Single-IM Bayesian inference

The M_w 6.2 Kumamoto earthquake (see Data and Resources) was recorded by a total of 192 local strong-motion instruments. For demonstration purposes, a distributed ground-motion field is predicted across a 100 by 100 km square area, which contains 26 strong-motion observations (Table 5 and Figure 8). In this section, the BN approach is first demonstrated for a single-IM prediction (peak ground acceleration, PGA, only), without the cross-correlation with other IMs (e.g., response spectral ordinates). Therefore only 25 observations are exploited, since the PGA from station #8 was not available.

[Table 5 about here]

The prior ground-motion field is computed with the GMPE of Chiou and Youngs (2008), assuming a strike-slip faulting mechanism and a depth to top of rupture Z_{TOR} = 5 km. Before the PGA observations are entered in the BN, they are converted to rock conditions by removing the amplification factors that are modelled in the GMPE using the time-averaged velocity of the top 30m, $V_{s,30}$. The $V_{s,30}$ value for each seismic station is obtained from the K-NET database, while an extrapolation for profiles that are shallower than 30 m has been performed using the relationships provided by Boore et al. (2011).

[Figure 8 about here]

A 48 × 48 global grid is used for the prediction of the ground motions, while a subdivision into 16 12 × 12 sub-grids is adopted to reduce the computation time. As a result, the total number of points within each sub-grid equals 144 (+ 25 observation points), which leads to a 169 × 169 correlation matrix. The resulting BN contains a total of 339 nodes and around 14,000 directed links between the variables.

By substituting $Y_{1,i}$ = ln PGA_i in Equation 4 and by setting $\sigma_\zeta = 0.518$ and $\sigma_\eta = 0.296$ (i.e., intra- and inter-event standard deviation provided by Chiou and Youngs, 2008), the updated PGA field is computed using Bayesian inference (Figure 9, left). The spatial correlation model from Equation 2 with b = 13.5 km for PGA is adopted here, although the choice of the spatial correlation model and its corresponding correlation distance remains a crucial issue and may have a large impact on the resulting ground-motion map, as noted by Crowley et al. (2008b). Other studies (e.g., Sokolov et al., 2010) have shown the significant variations in correlation lengths that may be deduced from different seismic arrays, even for the same geographical area. The total time taken for the generation of the ground-motion field is less than 3 minutes on a personal computer.

[Figure 9 about here]

The updated ground-motion field from the ShakeMap method is also displayed on Figure 9 right. It can be seen that the outcomes from both approaches are very similar: over all grid points, the averaged error rate between the ShakeMap and BN results is 6.5%, with a maximum of 32.2%. There is a slight over-estimation by the BN method, since the initial inter-event variability of the GMPE (i.e., prior distribution) tends to constrain the updating of the distribution from the relatively small number of observations. Conversely, the ShakeMap algorithm has adjusted the event magnitude down to 5.952 to even out the global bias introduced by the observations. In the GMPE used, the relation between M_w and the Y_i estimates is not linear, so lowering the magnitude is not exactly the same as lowering the inter-event error. Other differences are due to the fact that the spatial correlation between grid points in not taken into account by the ShakeMap method, which relies on interpolation only, as discussed in Section “Demonstration of the Bayesian network approach”.

The total standard-deviation of the PGA estimates by the two methods is also displayed in Figure 10. The results confirm the discussion in Section “Demonstration of the Bayesian network approach” (see Figure 5): the dispersion of the predictions far from the observations is lower for the ShakeMap, due to the assumption that the inter-event standard-deviation can be set to zero if enough observations are present. On the contrary, the BN method provides an updated inter-event standard-deviation of 0.101 (instead of the initial value of 0.296), which has to be included in the field of intra-event dispersions.

[Figure 10 about here]

To summarize, the discrepancy in the estimation of the uncertainty fields derives from the way posterior distributions are computed in the BN: the ground-motion inference relies entirely on the updating of the intra- and inter-event error terms, which are globally affected by the number and the spatial distribution of observations. On the other hand, the interpolation that is performed in the ShakeMap algorithm is strongly influenced by the observations in the immediate vicinity.

Joint inference on two cross-correlated IMs

The M_w 6.2 Kumamoto earthquake is used again to demonstrate the inference of cross-correlated ground-motion fields, namely PGA as the primary IM and SA(1.0s) as the secondary IM. Therefore the vector-valued ground-motion field may be updated from 25 PGAs and 26 values of SA(1.0s), according to Table 5. Assuming a correlation distance of 20 km for SA(1.0s) and a period-to-period cross-correlation coefficient of ρ_{12} = 0.587 (Baker & Cornell, 2006), the inferred ground motions are displayed in Figure 11 for both cross-correlated IMs. It should be noted that another BN configuration has been tested, where SA(1.0s) becomes the primary IM and PGA the secondary one: the results are identical whatever the selected order of IMs, thanks to the message passing ability of BNs (i.e., the propagation of evidence is not necessarily influenced by the direction of the link between two variables).

[Figure 11 about here]

Slight differences may be observed between the PGA field that has been estimated as a single-IM prediction (Figure 9) and the one that is cross-correlated with a secondary IM (Figure 11). In particular, the PGA field appears to be altered at the location of station #8, which has no record of PGA, thanks to the contribution of the SA(1.0s) observations, which provide additional constraints. This effect is demonstrated through a cross-validation study (see Table 6) on the 25 stations for which PGA observations are available: for each station, the PGA observation is removed from the analysis and the prediction at this station's location. This process is repeated for three approaches, namely the ShakeMap method, the BN inference with PGA only and the BN inference with both PGA and SA(1.0s), in order to compare their predictive abilities.

[Table 6 about here]

The difference between the ShakeMap method and the single-IM BN inference is not very significant, as already suggested by the comparison of the respective ground-motion maps in the previous sub-section. The multi-IM BN approach, however, introduces non-negligible changes in the PGA field and improves the prediction with respect to the observation in most cases. Aside from better constraining the ground-motion map, the ability of the BN approach to generate multiple-IM fields is very useful for the rapid post-earthquake damage assessment of different types of exposed assets.

Integration of macroseismic intensities and site conditions

To demonstrate the operational capabilities of the proposed BN approach, a ground-motion map is generated for a wider area, i.e. a 200 by 200 km square surrounding the epicenter of the M_w 6.2

Kumamoto foreshock, with a step grid of around 2 km. Within this area, 90 strong-motion observations are found, along with 14 aggregated reports of macroseismic intensity. As with the ShakeMap algorithm, macroseismic data may be exploited in complement to strong-motion data, through the use of ground-motion intensity conversion equations (GMICEs) (Wald et al., 1999). Starting from the BN in Figure 1, another set of BN nodes representing the macroseismic intensity is created, with a link pointing from each primary IM (i.e., PGA) node to each macroseismic intensity node. In the present example, the global GMICE developed by Caprio et al. (2015) has been used, thus the expression of the modified Mercalli intensity (MMI) takes the following form:

$$MMI = \alpha + \beta \cdot \ln PGA + \varepsilon_{MMI} \tag{9}$$

where α and β are GMICE coefficients and ε_{MMI} represents the error term of the regression, which follows a normal distribution with zero mean and standard deviation σ_{MMI}.

Therefore, in the BN, the conditional probability distribution of each MM_i node, which is the child of a $Y_{1,i}$ node representing PGA, can be expressed as:

$$p\left(MMI_i \middle| Y_{1,i}\right) = N\left(\alpha + \beta \cdot Y_{1,i}, \sigma_{MMI}^2\right) \tag{10}$$

As a result, the BN is able to collect evidence from various sources and pass the inference message in a two-way manner, i.e. (*i*) from a PGA observation up to the neighboring grid points and finally to the converted intensities on the grid, or (*ii*) from the reported intensity up to the converted PGA at the same location and finally to the neighboring grid points. The generated ground-motion maps for both PGA and MMI are displayed in Figure 12, after a site correction has been applied to the inferred variables at “rock” conditions.

[Figure 12 about here]

CONCLUSIONS

The BN formulation presented in this paper, which makes use of the spatial distribution of the intra- and inter-event errors in the GMPE, has been successfully tested on a real-world example, thus validating the way the ground-motion inference is performed in the proposed Bayesian framework. Therefore, such a result lays a solid foundation for the development of more elaborate BNs that integrate damage and loss assessments, which may be used as part of an operational decision support system for emergency responders.

The comparison with the ShakeMap algorithm has provided valuable lessons on the respective merits of each approach. Although computationally costlier, the BN method offers a different philosophy when treating uncertainties because a more refined estimation of the posterior distribution of the inter-event error is possible. It may be imagined to use such an approach in complement to the current ShakeMap algorithm, to adjust the value of inter-event error with respect to the number of observations, for instance. On the other hand, there is no obvious link between the weighted interpolation used in the ShakeMap algorithm and the spatial correlation coefficient used in the BN method, which complicates the direct comparison of the two approaches. However, the analysis and the comparison of maps generated with these two complementary approaches could be useful to help constrain the current correlation models.

Moreover, the ability of the BN approach to compute vector-valued IM fields and to access the joint probabilities of IMs across several locations should prove highly beneficial when dealing with the loss prediction of infrastructure systems, whose components are often susceptible to different IMs. Such inferences come at a high computational cost, which are currently not suitable for the near real-time applications that are covered by the ShakeMap framework. Conversely, in the case of the risk management of spatially-distributed infrastructure systems, where the ground-motion prediction has to be carried out for a limited number of sites, the BN approach might provide a rigorous probabilistic framework for the rapid loss assessment of interdependent components.

Finally, the proposed BN has mainly been focused on the treatment of aleatory variabilities (i.e., GMPE error terms); however, other variables representing epistemic uncertainties may be added to the BN, such as different GMPE candidates or different source or site assumptions. Provided that sufficient field observations are gathered, the BN inference would then be able to better constrain these parameters.

DATA AND RESOURCES

The metadata on the M_w 6.2 Kumamoto earthquake for the generation of the ground-motion map have been taken from the USGS ShakeMap webpage (http://earthquake.usgs.gov/earthquakes/eventpage/us20005hzn#shakemap). Website last accessed on August 1st 2017.

The information on the soil profiles of the seismic stations has been taken from the K-NET network webpage (http://www.kyoshin.bosai.go.jp/). Website last accessed on August 1st 2017.

The Bayes Net toolbox has been written by Kevin Murphy and it is available from the webpage https://github.com/bayesnet/bnt. Website last accessed on August 1st 2017.

ACKNOWLEDGMENTS

This research has been partially supported by the internal research program PSO VULNERABILITE at BRGM, France, and by the European Commission's FP7 project INFRARISK (Grant Agreement No. 603960) at University College London, UK. We thank two anonymous reviewers for their detailed and careful comments that led to significant improvements to this study.

REFERENCES

Baker, J. W., and C. A. Cornell (2006). Correlation of Response Spectral Values for Multicomponents Ground Motions, *Bull. Seism. Soc. Am.* **96**(1) 215-227.

Bensi, M., A. Der Kiureghian, and D. Straub (2011a). A Bayesian network methodology for infrastructure seismic risk assessment and decision-support, *PEER Report 2011/02*, Pacific Earthquake Engineering Research Center, University of California, Berkeley, CA.

Bensi, M., A. Der Kiureghian A, and D. Straub (2011b). Bayesian network modeling of correlated random variables drawn from a Gaussian random field, *Struct. Saf.* **33**(6) 317-332.

Bensi, M., A. Der Kiureghian, and D. Straub (2013). Efficient Bayesian network modeling of systems, *Reliab. Eng. Syst. Safe.* **112** 200-213.

Boore, D. M., E. M. Thompson, and H. Cadet (2011). Regional correlations of V_{S30} and velocities averaged over depths less than and greater than 30 meters, *Bul. Seism. Soc. Am.* **106**(6) 3046-3059.

Cavalieri, F., P. Franchin, P. Gehl, and D. D'Ayala (2017). Bayesian Networks and Infrastructure Systems: Computational and Methodological Challenges, in *Risk and Reliability Analysis: Theory and Applications*, P. Gardoni (Editor), Springer, 385-415.

Caprio, M., B. Tarigan, B. C. Worden, S. Wiemer, and D. J. Wald (2015). Ground motion to intensity conversion equations (GMICEs): A global relationship and evaluation of regional dependency, *Bull. Seism. Soc. Am.* **105**(3) 1476-1490.

Chiou, B. S., and R. R. Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**(1) 173-215.

Crowley, H., and J. J. Bommer (2006). Modelling seismic hazard in earthquake loss models with spatially distributed exposure, *Bull. Earthquake Eng.* **4**(3) 249-273.

Crowley, H., J. J. Bommer, and P. J. Stafford (2008a). Recent developments in the treatment of ground-motion variability in earthquake loss models, *J. Earthquake Eng.* **12**(2) 71-80.

Crowley, H., J. J. Bommer, and P. J. Stafford (2008b). Can earthquake loss models be validated using field observations? *J. Earthquake Eng.* **12**(7) 1078-1104.

Douglas, J. (2007). Inferred ground motions on Guadeloupe during the 2004 Les Saintes earthquake, *Bull. Earthquake Eng.* **5**(3) 363-376.

Douglas, J. and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation, *Earth Sci. Rev.* **160** 203-219.

Ebel, J.E., and D. J. Wald (2003). Bayesian estimations of peak ground acceleration and 5% damped spectral acceleration from Modified Mercalli Intensity data, *Earthquake Spectra* **19**(3) 511-529.

Erdik, M., K. Şeşetyan, M. B. Demircioğlu, U. Hancılar, and C. Zülfikar (2011). Rapid earthquake loss assessment after damaging earthquakes, *Soil Dyn. Earthquake Eng.* **31**(2) 247-266.

Esposito, S., and I. Iervolino (2011). PGA and PGV spatial correlation models based on European multievent datasets, *Bull. Seism. Soc. Am.* **101**(5) 2532-2541.

Gehl, P., F. Cavalieri, P. Franchin, and C. Negulescu (2017). Robustness of a hybrid simulation-based/Bayesian approach for the risk assessment of a real-world road network, in *Proceedings of the 12th International Conference on Structural Safety and Reliability*, Vienna, Austria.

Jaiswal, K., D. J. Wald, and D. D'Ayala (2011). Developing empirical collapse fragility functions for global building types, *Earthquake Spectra* **27**(3) 775-795.

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, *Earthquake Eng. Struct. Dyn.* **38**(15) 1687-1708.

Murphy, K. P. (2002). *Dynamic Bayesian networks: representation, inference and learning*, PhD dissertation, University of California, Berkeley.

Oliver, D. S. (2003). Gaussian cosimulation modelling of the cross-covariance, *Math. Geol.* **356** 681-698.

Park, J., P. Bazzurro, and J. W. Baker (2007). Modeling spatial correlation of ground motion Intensity Measures for regional seismic hazard and portfolio loss estimation, in *Proceedings of the 10th International Conference on Applied Statistics and Probability*, Tokyo, Japan.

Pozzi, M., and A. Der Kiureghian (2013). Gaussian Bayesian network for reliability analysis of a system of bridges, in *Proceedings of the 11th International Conference on Structural Safety and Reliability*, New York, United States.

Sokolov, V., F. Wenzel, W. Y. Jean, and K. L. Wen (2010). Uncertainty and spatial correlation of earthquake ground motion in Taiwan, *Terr. Atmos. Ocean. Sci.* **21**(6) 905-921.

Stafford, P. J. (2012). Evaluation of structural performance in the immediate aftermath of an earthquake: a case study of the 2011 Christchurch earthquake, *Int. J. Forensic Engineering* **1**(1) 58-77.

Vanmarcke, E. (1983). *Random Fields, Analysis and Synthesis*, The MIT Press, Cambridge, Massachussets.

Wald, D. J., V. Quitoriano, T. H. Heaton, and H. Kanamori (1999). Relationships between peak ground acceleration, peak ground velocity and modified Mercalli intensity in California, *Earthquake Spectra* **15**(3) 557-564.

Wald, D. J., B. C. Worden, V. Quitoriano, and K. L. Pankow (2005). ShakeMap manual: technical manual, user's guide, and software guide, *U.S. Geol. Surv. Tech. Methods 12-A1, 132 pp.*

Wald, D. J., K. W. Lin, K. Porter, and L. Turner (2008). ShakeCast: Automating and improving the use of ShakeMap for post-earthquake decision-making and response, *Earthquake Spectra* **24**(2) 533-553.

Weatherill, G., S. Esposito, I. Iervolino, P. Franchin, and F. Cavalieri (2014). Framework for seismic hazard analysis of spatially distributed systems, in *SYNER-G: Systemic Seismic Vulnerability and Risk Assessment of Complex Urban, Utility, Lifeline Systems and Critical Facilities - Methodology and Applications*, K. Pitilakis, P. Franchin, B. Khazai, H.Wenzel (Editors), Springer, 57-88.

Worden, B. C., and D. J. Wald (2016). ShakeMap Manual Online: technical manual, user's guide, and software guide, DOI: 10.1234/012345678. Available: http://usgs.github.io/shakemap. Accessed July 2017.

Worden, B. C., D. J. Wald, T.I. Allen, K. W. Lin, D. Garcia, and G. Cua (2010). A revised ground-motion and intensity interpolation scheme for ShakeMap, *Bull. Seism. Soc. Am.* **100**(6) 3083-3096.

FULL MAILING ADDRESS FOR EACH AUTHOR

BRGM/DRP, 3 avenue C. Guillemin, BP 36009, 45060 Orléans Cedex 2, France [PG]

University of Strathclyde, Dept of Civil and Environmental Engineering, James Weir Building, 75 Montrose Street, G11XJ Glasgow, UK [JD]

University College London, Dept of Civil, Environmental and Geomatic Engineering, Gower Street, WC1E 6BT London, UK [DDA]

TABLES

Table 1: Prior and posterior Gaussian distribution parameters of the BN variables corresponding to the synthetic example.

Variables	Prior		Posterior	
	μ	σ	μ	σ
U1	0	1	-0,2499	0,5877
U2	0	1	0,0628	0,9667
U3	0	1	0,1174	0,9667
U4	0	1	0,0780	0,9674
U5	0	1	0,1512	0,9379
U6	0	1	0,1468	0,9498
U7	0	1	0,1116	0,9720
U8	0	1	0,1983	0,9050
U9	0	1	0,1361	0,9564
U10	0	1	-0,1992	0,9194
U11	0	1	0,1897	0,9126
W	0	1	-0,0300	0,8434
Y(1)	-1,6377	0,6508	-1,7884	0,2163
Y(2)	-1,5960	0,6508	-1,6974	0,2976
Y(3)	-1,5648	0,6508	-1,6083	0,3906
Y(4)	-1,6701	0,6508	-1,7595	0,2884
Y(5)	-1,6283	0,6508	-1,6645	0,2762
Y(6)	-1,5958	0,6508	-1,5843	0,3325
Y(7)	-1,7113	0,6508	-1,7275	0,3591
Y(8)	-1,6700	0,6508	-1,6213	0,2418
Y(9)	-1,6375	0,6508	-1,5610	0,2528
Yobs1	-1,6335	0,6508	**-1,7961**	**0**
Yobs2	-1,6526	0,6508	**-1,5573**	**0**

The numbers in bold correspond to the evidence from the observations.

Table 2: Posterior ground-motion parameters of the synthetic example, obtained with the three methods (BN, ShakeMap and analytical solution).

	Analytical solution		BN method		ShakeMap method	
	PGA [m/s^2]	σ_{tot_lnPGA}	PGA [m/s^2]	σ_{tot_lnPGA}	PGA [m/s^2]	σ_{tot_lnPGA}
Y(1)	0.1672	0.3477	0.1672	0.2163	0.1710	0.2080
Y(2)	0.1832	0.4030	0.1832	0.2976	0.1858	0.2808
Y(3)	0.2002	0.4736	0.2002	0.3906	0.2004	0.3568
Y(4)	0.1721	0.3967	0.1721	0.2884	0.1732	0.2758
Y(5)	0.1893	0.3883	0.1893	0.2762	0.1891	0.2750
Y(6)	0.2051	0.4291	0.2051	0.3325	0.2020	0.3106
Y(7)	0.1777	0.4495	0.1777	0.3591	0.1763	0.3319
Y(8)	0.1977	0.3645	0.1977	0.2418	0.1959	0.2350
Y(9)	0.2099	0.3712	0.2099	0.2528	0.2040	0.2404
	η	σ_{η}	η	σ_{η}	η	σ_{η}
	-0.0097	0.2730	-0.0097	0.2730	-0.0336	0.0000

Table 3: Computational cost and accuracy of the optimization strategy, with different values of m (number of **U** *nodes).*

Grid size	Measure	Cholesky decomposition	Optimization			
			m = 20	*m* = 10	*m* = 5	*m* = 2
3x3 (+ 2 obs.)	mean error	*exact*	N/A	< 0.1%	0.1%	0.5%
	max. error	*exact*		< 0.1%	0.3%	1.1%
	init. time	< 0.1 s		7.7 s	0.5 s	0.2 s
	BN time	< 0.1 s		< 0.1 s	< 0.1 s	< 0.1 s
6x6 (+ 4 obs.)	mean error	*exact*	< 0.1%	0.4%	2.6%	7.7%
	max. error	*exact*	0.2%	1.0%	6.1%	13.2%
	init. time	< 0.1 s	173.6 s	20.5 s	2.7 s	0.6 s
	BN time	0.3 s	0.9 s	0.8 s	0.7 s	0.7 s
12x12 (+ 8 obs.)	mean error	*exact*	1.4%	2.2%	3.3%	3.6%
	max. error	*exact*	6.3%	9.9%	13.1%	12.8%
	init. time	< 0.1 s	4237.6 s	606.5 s	117.6 s	16.9 s
	BN time	5.9 s	14.9 s	14.1 s	13.4 s	12.9 s
24x24 (+ 16 obs.)	mean error	*exact*	out of memory	out of memory	out of memory	5.9%
	max. error	*exact*				15.8%
	init. time	< 0.1 s				1141.1 s
	BN time	1213.1 s				2588.4 s

"init. time" refers to the initialization time, corresponding to the construction of the transformation matrix **T** or its approximation $\hat{\mathbf{T}}$, "BN time" refers to the total duration of the Bayesian execution (i.e., construction of the junction tree and Bayesian inference for all grid points) on a standard PC. The mean and maximum error measures refer to the PGA mean value.

Table 4: Computational cost and accuracy of the grid subdivision strategy, with different sub-grid sizes, for the large 24 x 24 grid tested in Table 3.

Measure	64 x (3 x 3) sub-grids		16 x (6 x 6) sub-grids		4 x (12 x 12) sub-grids	
init. time	< 0.1 s		< 0.1 s		< 0.1 s	
BN time	10.3 s		8.8 s		27.9 s	
d_{corr} = 4.5 km	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}
mean error	2.14E-4%	9.37E-4%	1.81E-4%	8.82E-4%	1.57E-4%	8.50E-4%
max. error	1.83E-3%	4.33E-3%	1.63E-3%	4.41E-3%	1.54E-3%	3.72E-3%
d_{corr} = 9 km	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}
mean error	1.93E-4%	7.51E-4%	1.80E-4%	7.80E-4%	1.48E-4%	7.29E-4%
max. error	2.13E-3%	4.75E-3%	3.53E-3%	7.92E-3%	1.84E-3%	4.75E-3%
d_{corr} = 18 km	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}	*mean PGA*	σ_{tot_lnPGA}
mean error	1.52E-4%	8.12E-4%	1.41E-4%	8.24E-4%	1.42E-4%	8.12E-4%
max. error	2.90E-3%	1.32E-2%	2.25E-3%	8.45E-3%	1.96E-3%	9.46E-3%

The error measures, applied to both the mean PGA and global standard-deviation, are computed with respect to the full grid solution. The accuracy is quantified for three correlation lengths d_{corr}, the first one corresponding to the standard case (i.e., b = 3*d_{corr} = 13.5 km).

Table 5: Recording K-Net stations used and corresponding PGA and SA(1.0s) values corrected at a rock site, for the M_w 6.2 Kumamoto earthquake.

Station #	Station ID	Estimated $V_{s,30}$ [m/s]	Recorded PGA_{rock} [m/s^2]	GMPE-Predicted PGA_{rock} [m/s^2]	Recorded $SA(1.0s)_{rock}$ [m/s^2]	GMPE-Predicted $SA(1.0s)_{rock}$ [m/s^2]
1	KMM006	195	4.03	3.26	1.93	1.42
2	KMM008	160	2.34	2.39	1.10	1.01
3	KMM005	287	1.34	1.46	0.64	0.60
4	KMM003	239	0.50	1.22	0.18	0.49
5	KMM011	185	2.66	1.05	0.23	0.43
6	KMM002	190	0.70	0.98	0.18	0.40
7	KMM010	149	0.38	0.86	0.25	0.35
8	KMM009	348	-	-	0.19	0.34
9	KMM012	205	0.75	0.70	0.35	0.29
10	NGS012	466	0.25	0.67	0.27	0.28
11	FKO016	363	0.44	0.63	0.11	0.26
12	KMM007	239	1.01	0.53	0.14	0.22
13	FKO014	858	0.34	0.49	0.17	0.21
14	KMM004	211	0.20	0.48	0.08	0.21
15	KMM014	641	0.57	0.44	0.07	0.19
16	NGS011	518	0.16	0.42	0.24	0.18
17	FKO015	134	0.42	0.38	0.16	0.17
18	KMM001	223	0.22	0.38	0.09	0.17
19	FKO013	259	0.37	0.38	0.08	0.16
20	KMM013	220	0.30	0.38	0.15	0.16
21	NGS008	547	0.18	0.37	0.11	0.16
22	NGS014	143	0.26	0.36	0.08	0.16
23	KMM018	287	0.29	0.35	0.03	0.16
24	MYZ020	256	0.44	0.33	0.06	0.15
25	KMM019	490	0.26	0.28	0.03	0.13
26	KMM020	386	0.15	0.25	0.05	0.12

Table 6: Updated prediction of PGA for the 25 stations, when sequentially removing the PGA observation at the given station.

Station #	ShakeMap		BN (PGA only)		BN (PGA and SA)	
	PGA_{rock} [m/s^2]	Prediction error	PGA_{rock} [m/s^2]	Prediction error	PGA_{rock} [m/s^2]	Prediction error
1	2.95	-27.0%	2.68	-33.4%	3.46	-14.2%
2	2.11	-9.7%	1.97	-15.8%	2.21	-5.8%
3	1.25	-7.1%	1.22	-9.4%	1.34	-0.4%
4	1.03	107.2%	1.00	102.5%	0.64	29.2%
5	0.88	-66.9%	0.83	-68.7%	0.67	-74.7%
6	0.81	16.0%	0.79	12.1%	0.58	-17.3%
7	0.72	87.8%	0.73	90.8%	0.65	70.1%
9	0.57	-23.9%	0.56	-25.5%	0.69	-8.1%
10	0.55	123.2%	0.54	119.7%	0.57	132.5%
11	0.51	17.9%	0.50	15.5%	0.37	-15.3%
12	0.43	-57.5%	0.40	-60.0%	0.36	-64.1%
13	0.40	15.4%	0.41	17.8%	0.40	15.5%
14	0.39	98.9%	0.43	119.0%	0.29	44.7%
15	0.35	-37.9%	0.36	-37.2%	0.24	-57.6%
16	0.33	100.6%	0.33	99.5%	0.41	147.5%
17	0.30	-28.3%	0.31	-26.7%	0.34	-18.2%

18	0.30	35.2%	0.31	40.2%	0.25	13.2%
19	0.30	-19.3%	0.31	-16.7%	0.22	-40.1%
20	0.30	-1.9%	0.31	1.6%	0.33	8.6%
21	0.29	63.9%	0.30	67.8%	0.25	40.4%
22	0.28	7.1%	0.29	10.1%	0.21	-18.6%
23	0.28	-2.8%	0.29	0.2%	0.14	-52.1%
24	0.26	-40.8%	0.26	-39.6%	0.19	-57.3%
25	0.22	-15.9%	0.23	-14.9%	0.13	-50.7%
26	0.19	24.0%	0.21	33.2%	0.17	10.8%

The prediction error measures the relative error rate with the actual observation.

LIST OF FIGURE CAPTIONS

FIGURES

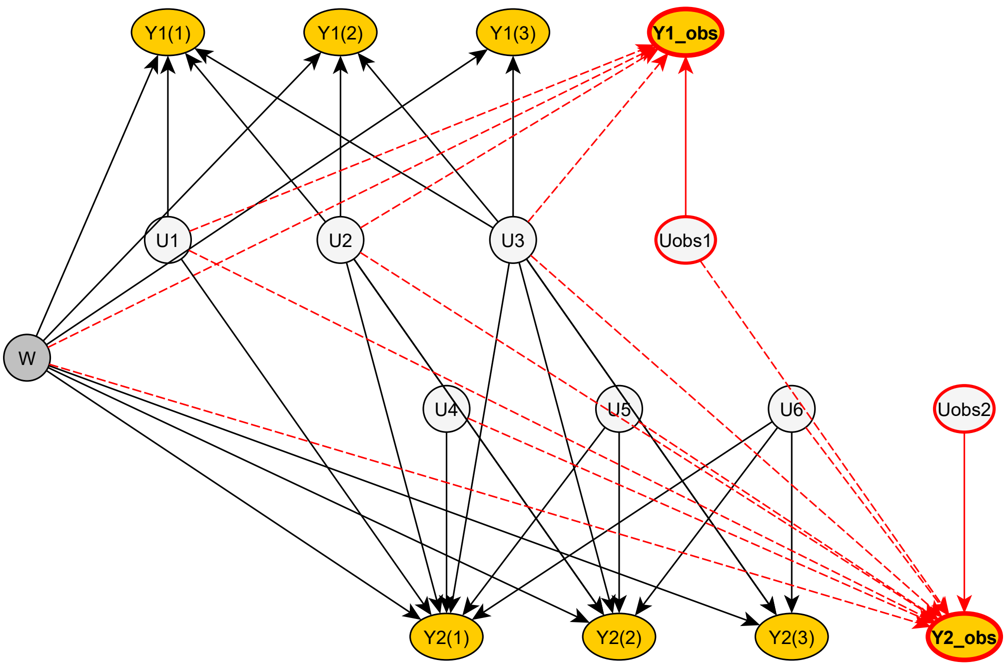

Figure 1: Example of a BN structure for the prediction of the spatial ground-motion distribution for three grid points (Y1 represents the principal IM and Y2 the secondary IM). The nodes Y1_obs and Y2_obs in bold represent an observation (i.e., evidence) of the two IMs at a given location.

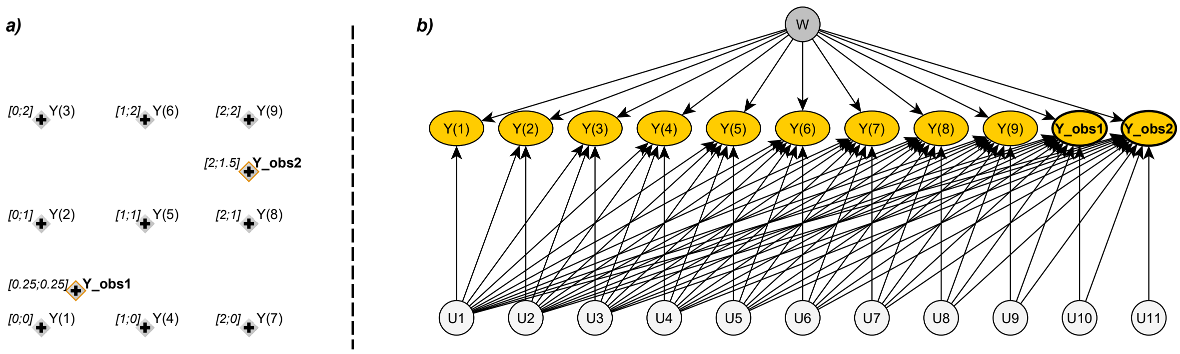

Figure 2: (a) Spatial configuration of the synthetic example used in the demonstration and (b) corresponding BN formulation.

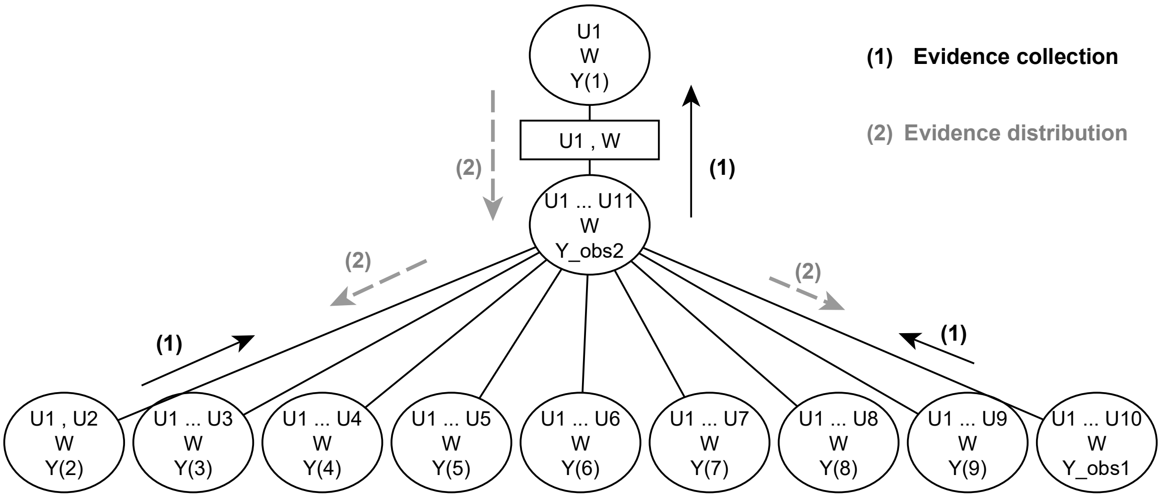

Figure 3: Junction tree corresponding to the example BN. The circles represent the cliques and the rectangular box is an example of a clique separator (i.e., set of nodes that are common to two connected cliques). The top circle is the root clique and the bottom ones represent the leaves of the junction tree.

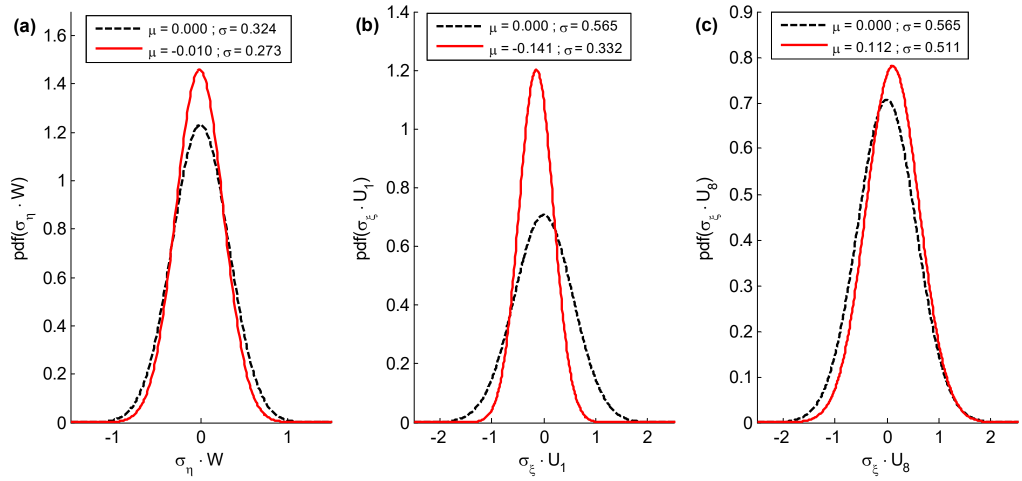

Figure 4: Prior (dashed line) and posterior (solid line) distributions for variables W, U_1 and U_8, representing respectively the inter-event error (left), an overestimated intra-event error (middle) and an underestimated intra-event error (right). The normal variables U_1 and U_8 have a strong link in the BN with the sites close to virtual stations #1 and #2, respectively.

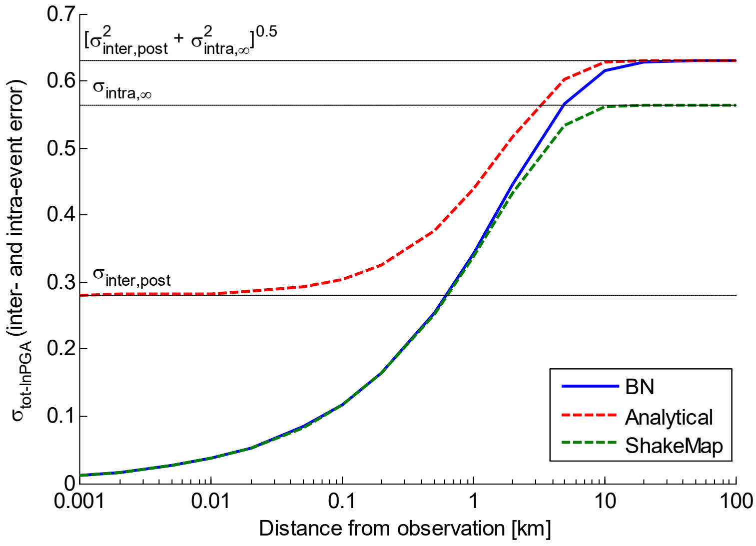

Figure 5: Evolution of the global error term (intra- and inter-event) as a function of the distance from an observation, for the three methods. $\sigma_{inter,post}$ represents the updated inter-event standard-deviation and $\sigma_{intra,\infty}$ the updated intra-event standard-deviation very far from the observation (i.e., equivalent to the prior intra-event standard-deviation).

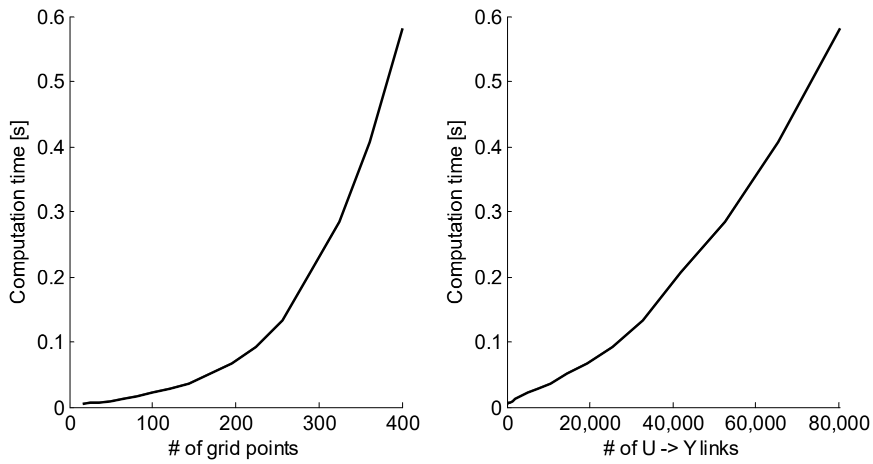

*Figure 6: Computation time (on an Intel(R) Core(TM) i5 processor with 4 GB RAM) for the Bayesian updating of one **Y** node, with respect to the number of points in the grid (left) and the number of **U** → **Y** links in the corresponding BN.*

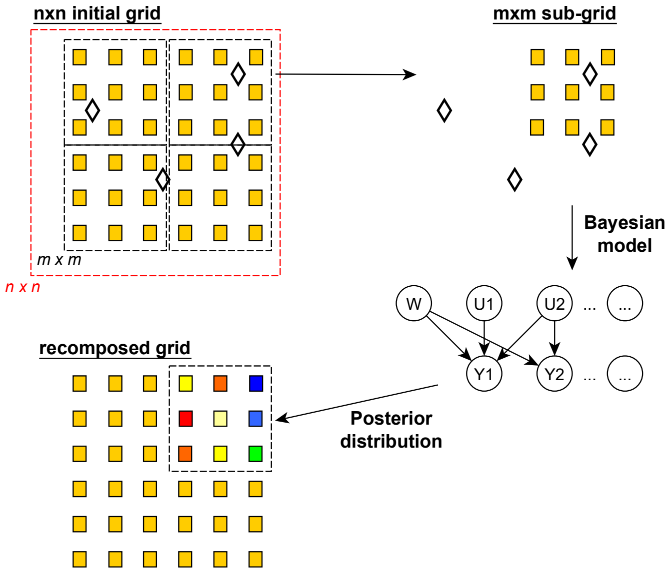

Figure 7: Illustration of the grid sub-division strategy, where the diamonds represent observations.

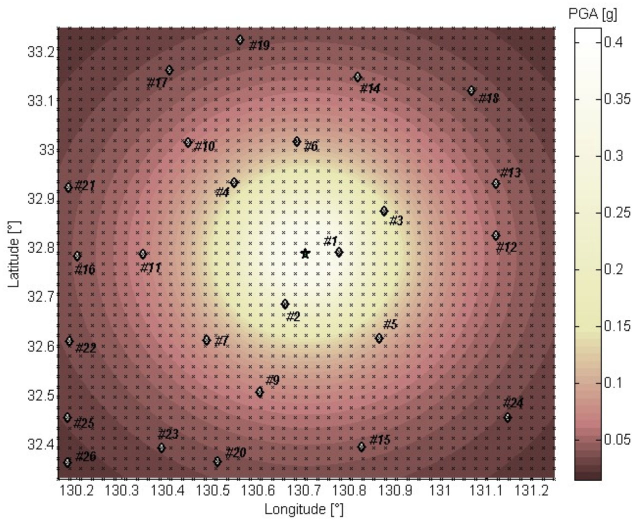

Figure 8: Prior estimation of PGA_{rock} using the source parameter and the GMPE. The recording stations are represented by diamonds and the earthquake epicenter by a star. The small black crosses represent the 2 025 grid points.

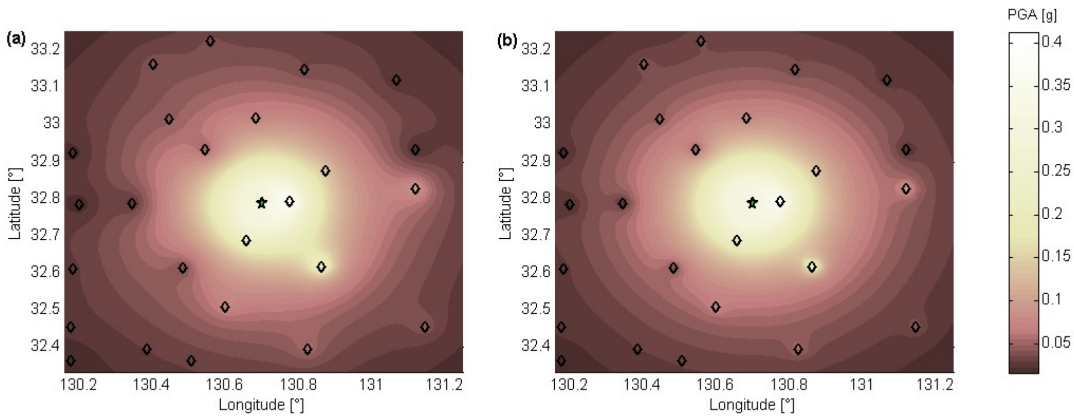

Figure 9: Updated shake-map for PGA_{rock} using the BN approach (left) and the ShakeMap algorithm (right). The recording stations are represented by diamonds and the earthquake epicenter by a star.

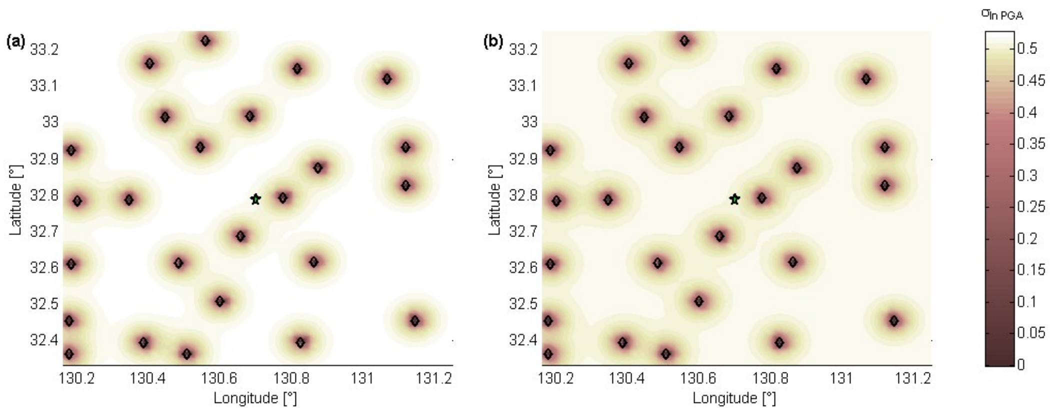

Figure 10: Updated field of $\sigma_{\ln PGA}$ using the BN approach (left) and the ShakeMap algorithm (right).

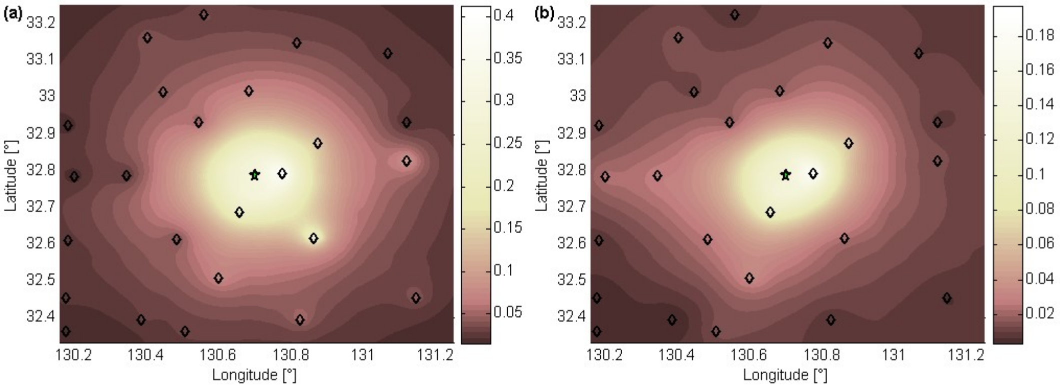

Figure 11: Updated BN based shake-map for PGA_{rock} (left) and SA_{rock} at 1.0s (right) in g, using all observations from Table 5. Both sets of observations are used for the generation of each of the maps.

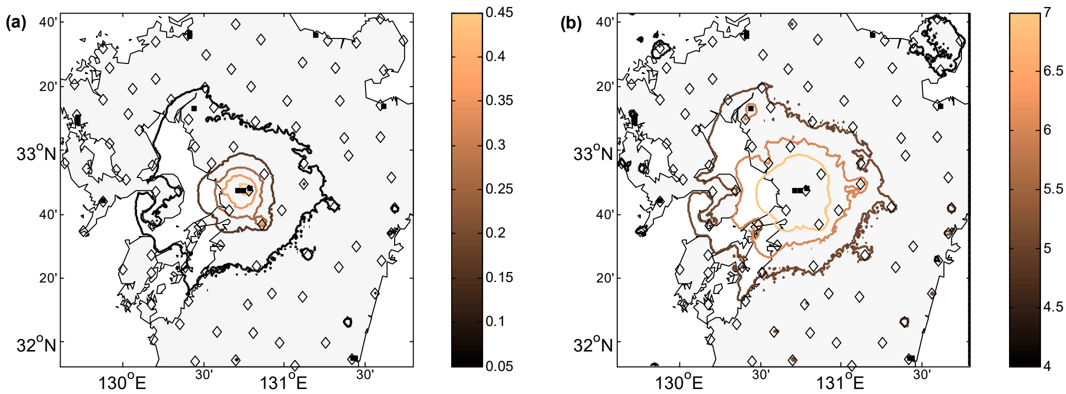

Figure 12: Updated BN based shake-map for PGA (left) in g and MMI (right), accounting for sites conditions and all available data (strong-motion data and macroseismic intensities). Strong-motion stations are represented by diamonds and intensity reports by black full squares.

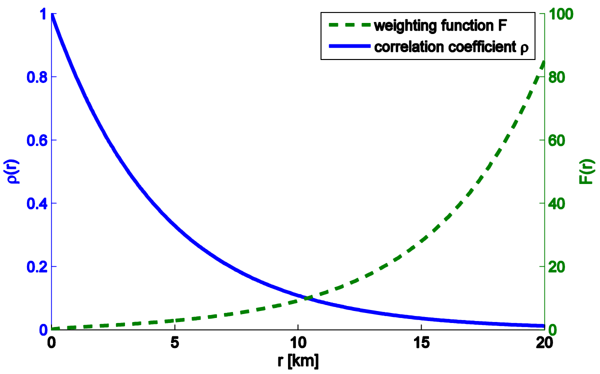

Figure A.1: Adopted spatial correlation model and weighting function, for b = 13.5 km.

APPENDIX A

The objective of this Appendix is to establish a correspondence between the weighting function used in the ShakeMap interpolation and the spatial correlation coefficient used in the BN approach.

Consider two independent and identically distributed normal variables *X* and *Y*, where *X* represents the initial GMPE estimate at a given grid point and *Y* represents an observation at a distance *r*.

According to Worden et al. (2010), the interpolated value X_{int} at the grid point may be computed from the following expression:

$$X_{\text{int}} = \frac{\frac{X}{\sigma_X^2} + \frac{Y}{\sigma_Y^2}}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_Y^2}} = \frac{\frac{X}{\sigma_X^2} + \frac{Y}{\sigma_X^2 \cdot F(r)^2}}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_X^2 \cdot F(r)^2}} = \frac{X + \frac{Y}{F(r)^2}}{1 + \frac{1}{F(r)^2}} = \frac{F(r)^2 \cdot X + Y}{F(r)^2 + 1} \tag{A.1}$$

where $F(r) = \sigma_Y/\sigma_X$ is the weighting function defined by Worden et al. (2010), and X_{int} is assumed to have the following standard deviation:

$$\sigma_{X\text{int}} = \sqrt{\frac{1}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_X^2 \cdot F(r)^2}}} = \frac{\sigma_X \cdot F(r)}{\sqrt{F(r)^2 + 1}} \tag{A.2}$$

According to the above definitions, the correlation coefficient between the variables *Y* and X_{int} must correspond to the spatial correlation coefficient ρ between two sites separated by a distance *r*. Therefore we can write:

$$\begin{aligned} \rho(r) &= \frac{\operatorname{cov}(X_{\text{int}}, Y)}{\sigma_{X\text{int}} \cdot \sigma_Y} \\ &= \frac{1}{\sigma_{X\text{int}} \cdot \sigma_Y \cdot \left(F(r)^2 + 1\right)} \cdot \operatorname{cov}\left(F(r)^2 \cdot X + Y, Y\right) \\ &= \frac{F(r)^2}{\sigma_{X\text{int}} \cdot \sigma_Y \cdot \left(F(r)^2 + 1\right)} . \operatorname{cov}(X, Y) + \frac{1}{\sigma_{X\text{int}} \cdot \sigma_Y \cdot \left(F(r)^2 + 1\right)} . \operatorname{cov}(Y, Y) \end{aligned} \tag{A.3}$$

By definition, we have cov(Y,Y) = σ_Y^2 and cov(X,Y) = 0, due to the independence assumption. Therefore the expression of the spatial correlation coefficient becomes:

$$\begin{aligned} \rho(r) &= \frac{\sigma_Y}{\sigma_{X\,\mathrm{int}} \cdot \left(F(r)^2 + 1\right)} \\ &= \frac{\sigma_X \cdot F(r)}{\sigma_{X\,\mathrm{int}} \cdot \left(F(r)^2 + 1\right)} \\ &= \frac{1}{\sqrt{F(r)^2 + 1}} \end{aligned} \quad \text{(A.4)}$$

If a spatial correlation model with an exponential decrease rate is used here (i.e., see Equation 8), then the weighting function *F(r)* that is proposed for the ShakeMap algorithm becomes:

$$F(r) = \sqrt{\frac{1}{\rho^2} - 1} = \sqrt{\left[\exp\left(\frac{3 \cdot r}{b}\right)\right]^2 - 1} \quad \text{(A.5)}$$

where b is the correlation length.

The evolution of the weighting function *F(r)* and of the correlation coefficient *ρ(r)* with respect to inter-site distance r is represented in Figure A.1, for b = 13.5 km.

[Figure A.1 about here]