

Spatial and Spectral Interpolation of Ground-Motion Intensity Measure Observations

by C. Bruce Worden, Eric M. Thompson, Jack W. Baker, Brendon A. Bradley,*
Nicolas Luco, and David J. Wald

Abstract Following a significant earthquake, ground-motion observations are available for a limited set of locations and intensity measures (IMs). Typically, however, it is desirable to know the ground motions for additional IMs and at locations where observations are unavailable. Various interpolation methods are available, but because IMs or their logarithms are normally distributed, spatially correlated, and correlated with each other at a given location, it is possible to apply the conditional multivariate normal (MVN) distribution to the problem of estimating unobserved IMs. In this article, we review the MVN and its application to general estimation problems, and then apply the MVN to the specific problem of ground-motion IM interpolation. In particular, we present (1) a formulation of the MVN for the simultaneous interpolation of IMs across space and IM type (most commonly, spectral response at different oscillator periods) and (2) the inclusion of uncertain observation data in the MVN formulation. These techniques, in combination with modern empirical ground-motion models and correlation functions, provide a flexible framework for estimating a variety of IMs at arbitrary locations.

Electronic Supplement: Demonstration Python script for the evaluation of the multivariate normal (MVN) with additional uncertainty.

Introduction

Following an earthquake of any significance, numerous sources of information can be used to constrain maps of shaking intensity. Examples include observations of ground motions from seismic instrumentation, online “Did You Feel It?” (DYFI) reports (Wald *et al.*, 2012), and field surveys of structural performance (e.g., Bommer and Stafford, 2012). The ground motions may be reported as a variety of intensity measures (IMs), such as peak ground acceleration (PGA), pseudoacceleration response spectra (SA) at selected oscillator periods (T), or macroseismic intensity (MI). These observations are made at specific locations (or, in the case of MI, for a specific localized area) and are not always available at the spectral period of interest. For many purposes, it is necessary to estimate IMs for additional locations and spectral periods. One example is ShakeMap (Worden and Wald, 2016), in which the IMs available in near-real time are reported for a limited set of spectral ordinates.

Empirical studies have demonstrated that ground-motion IMs are spatially correlated (e.g., Goda and Hong,

2008), and different IMs are also correlated with each other at the same location (e.g., Baker and Jayaram, 2008; Bradley, 2010, 2012). It is also reasonable to consider ground-motion IMs or their logarithms (e.g., MI or $\ln(\text{PGA})$) to be normally distributed, conditional upon rupture parameters such as magnitude and distance (Jayaram and Baker, 2008). These features make it possible to apply the conditional multivariate normal (MVN) to the problem of IM estimation.

In some cases, observed IMs may be available at a site of interest, but we wish to estimate IMs that are not included in the set of reported IMs. Baker (2011) summarizes the conditional spectrum approach, in which the distribution of SA at any period is conditioned upon an SA value at the same location. Bradley (2010, 2012) extended the method to account for the dependence of seismic response on non-spectral-acceleration IMs, such as those that are related to the duration or energy of the ground motion. Kishida (2017) makes use of the MVN to generalize the conditional spectrum approach to include multiple conditioning periods, which is a situation commonly encountered for near-real-time systems such as ShakeMap. The sites of interest, however, are most commonly not coincident with observations.

*Also at Department of Civil and Natural Resources Engineering, University of Canterbury, Private Bag 4800, Christchurch 8041, New Zealand.

In these cases, we wish to know the distribution of the IMs conditioned upon the available nearby observations. [Park et al. \(2007\)](#), [Stafford \(2012\)](#), and [Bradley \(2014\)](#) applied the MVN to the problem of interpolating IMs to sites of interest, based on the work of [Vanmarcke \(1983\)](#).

More generally, we wish to interpolate a vector of IMs over a spatial region, where the distribution of a particular IM is conditioned on observations of any IM at an arbitrary location. After reviewing the techniques discussed above, we give a formulation of the MVN that allows for simultaneous conditioning across multiple IMs and space, and provide examples for cases with SA at multiple periods and MI observations. Finally, the MVN (as typically formulated) assumes that observations are precise (i.e., that they have no uncertainty). It is often the case, however, that the conditioning data have nonzero uncertainty; this is especially important for MI data, but uncertainties in instrumental IM data can also arise due to differences in data processing, such as the manner in which orthogonal components are combined ([Boore and Kishida, 2017](#)). Here, we present a method that accounts for the uncertainty of the conditioning data for the MVN.

The Conditional Multivariate Normal Distribution

We begin with a brief review of the MVN. Typically, we have one or more estimates of the mean and variance of some quantity, a set of observations, and we wish to estimate the value of the parameter some distance from our observations. In this context, distance can refer to spatial separation, or separation in spectral period, time, or any other parameter(s) for which the correlation among observations may be determined.

For a variable of interest \mathbf{Y} , assume we wish to compute predictions (\mathbf{Y}_1) at M target ordinates, conditioned on N observations (\mathbf{Y}_2). The MVN is typically summarized (see e.g., [Johnson and Wichern, 2002](#)) as a function of a random vector partitioned into these two components:

$$\mathbf{Y} = \begin{Bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{Bmatrix}, \quad (1)$$

with mean

$$\mu_{\mathbf{Y}} = \begin{Bmatrix} \mu_{\mathbf{Y}_1} \\ \mu_{\mathbf{Y}_2} \end{Bmatrix}, \quad (2)$$

and covariance

$$\Sigma_{\mathbf{Y}} = \begin{bmatrix} \Sigma_{\mathbf{Y}_1\mathbf{Y}_1} & | & \Sigma_{\mathbf{Y}_1\mathbf{Y}_2} \\ \hline \Sigma_{\mathbf{Y}_2\mathbf{Y}_1} & | & \Sigma_{\mathbf{Y}_2\mathbf{Y}_2} \end{bmatrix}, \quad (3)$$

in which $M \times M$, $M \times N$, $N \times M$, and $N \times N$ give the dimensions of the partitioned arrays.

Given a set of observations $\mathbf{Y}_2 = \mathbf{y}_2$, we define a vector of residuals

$$\boldsymbol{\zeta} = \mathbf{y}_2 - \mu_{\mathbf{Y}_2}. \quad (4)$$

The distribution of \mathbf{Y}_1 , given that $\mathbf{Y}_2 = \mathbf{y}_2$, is MVN with mean and covariance

$$\mu_{\mathbf{Y}_1|\mathbf{y}_2} = \mu_{\mathbf{Y}_1} + \Sigma_{\mathbf{Y}_1\mathbf{Y}_2} \Sigma_{\mathbf{Y}_2\mathbf{Y}_2}^{-1} \boldsymbol{\zeta}, \quad \text{and} \quad (5)$$

$$\Sigma_{\mathbf{Y}_1\mathbf{Y}_1|\mathbf{y}_2} = \Sigma_{\mathbf{Y}_1\mathbf{Y}_1} - \Sigma_{\mathbf{Y}_1\mathbf{Y}_2} \Sigma_{\mathbf{Y}_2\mathbf{Y}_2}^{-1} \Sigma_{\mathbf{Y}_2\mathbf{Y}_1}. \quad (6)$$

The values of $\mu_{\mathbf{Y}}$ would typically be derived from a model of \mathbf{Y} (in this article, an empirical ground-motion model [GMM]). The elements of $\Sigma_{\mathbf{Y}}$ in equation (3) are

$$\Sigma_{Y_i Y_j} = \rho_{Y_i Y_j} \sigma_{Y_i} \sigma_{Y_j}, \quad (7)$$

in which σ_{Y_i} and σ_{Y_j} are taken from a GMM, and $\rho_{Y_i Y_j}$ is the correlation of the i th and j th elements of \mathbf{Y} conditional on the rupture and site parameters.

To aid with intuition in subsequent discussion, we note that in the bivariate case equations (5) and (6) reduce to

$$\mu_{Y_1|\mathbf{y}_2} = \mu_{Y_1} + \frac{\sigma_{Y_1}}{\sigma_{Y_2}} \rho_{Y_1 Y_2} (y_2 - \mu_{Y_2}), \quad \text{and} \quad (8)$$

$$\sigma_{Y_1|\mathbf{y}_2}^2 = (1 - \rho_{Y_1 Y_2}^2) \sigma_{Y_1}^2. \quad (9)$$

Spectral Interpolation

For P observed IMs at a site, we wish to estimate \mathbf{Y} at Q additional IMs. We follow the conditional spectrum framework ([Baker, 2011](#)) to compute the probability distribution of SA at an arbitrary T_1 conditioned on SA at T_2 . The conditional spectral equations in [Baker \(2011\)](#) are for a single conditioning period and are equivalent to equations (8) and (9). Therefore, equations (5) and (6) are the generalization of this approach to accommodate multiple conditioning periods, as employed by [Kishida \(2017\)](#).

Figure 1 is a synthetic example using the [Chiou and Youngs \(2014\)](#) GMM for an M 7.2 earthquake and V_{S30} of 760 m/s. We use synthetic instrumental data at periods of 0.3, 1.0, and 3.0 s. (The synthetic instrumental data were produced for illustrative purposes only and are not drawn from any specific model or distribution.) This situation often occurs with ShakeMap, in which SA data for these three periods are routinely supplied by seismic networks, but additional periods are desirable for loss modeling. Here, we use the interperiod correlation model of [Baker and Jayaram \(2008](#); see their fig. 4 for an illustration of the interperiod correlation structure). The gray line in Figure 1 represents the median values from the GMM, the dashed lines show

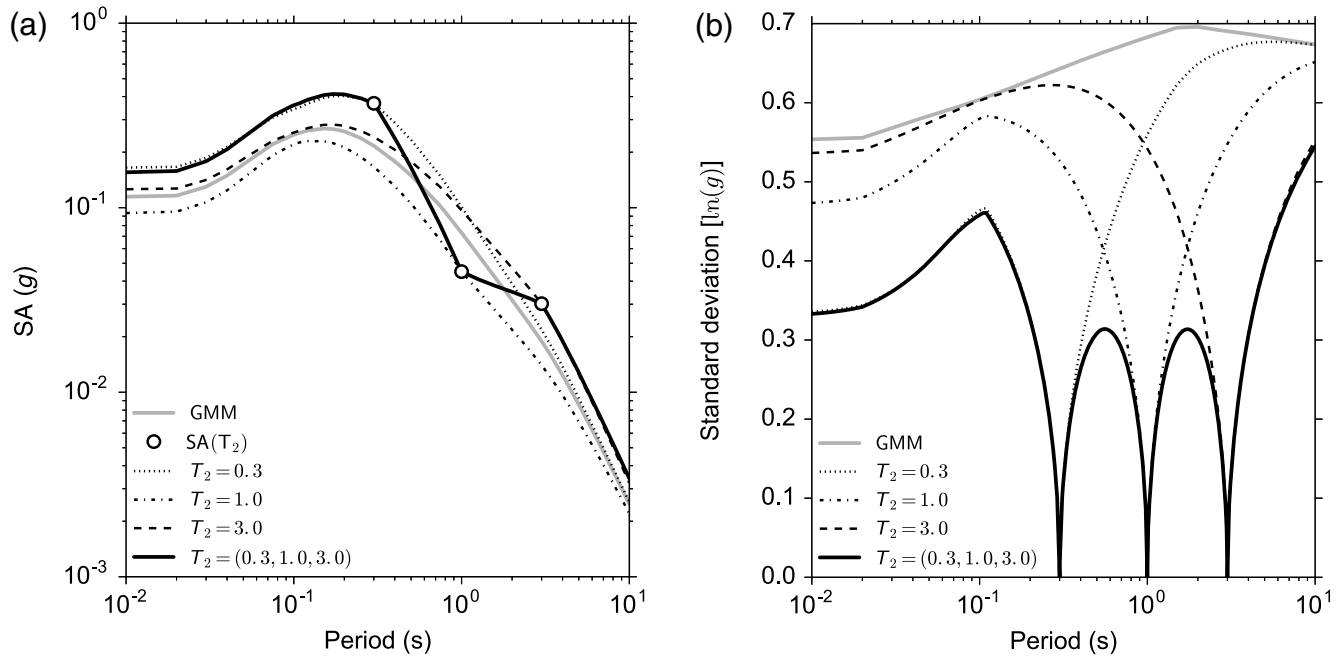


Figure 1. (a) Response spectra including the ground-motion model (GMM), the observed values $SA(T_2)$, the mean of the conditional spectrum distribution for three example T_2 , and the mean of the conditional spectrum distribution for T_2 . (b) The standard deviations for the response spectra curves on the left. SA, pseudoacceleration response spectra.

the single-period conditioning for each individual instrumental data point, and the solid black line shows the multiple-period conditioning result. This example demonstrates the value of conditioning on multiple periods simultaneously, in which the standard deviation of the estimate conditioned on multiple periods is always less than the standard deviation of the estimate conditioned on a single period (Fig. 1b). Figures such as this can help assess the impact of incorporating additional data at specific periods on the uncertainty of the estimates at other periods. The GMM uncertainty is often a function of input variables (i.e., it is heteroskedastic) and so multiple figures with varying input values are required for complete evaluation.

Spatial and Spectral Interpolation

Park *et al.* (2007), Stafford (2012), and Bradley (2014) each applied the MVN to the spatial estimation of ground-motion IMs. In general, GMMs define ground motion of the i th IM at the m th site as

$$IM_{i,m} = \mu_{IM_{i,m}} + \zeta_{i,m}, \quad (10)$$

in which $IM_{i,m}$ is the value of the i th IM at the m th site, and $\mu_{IM_{i,m}}$ is the mean value given by the GMM and is a function of the rupture and site parameters. For instrumental IMs, $IM_{i,m}$ is the logarithm of the intensity measure, whereas no transformation is required for MI. The total residual is usually treated as a linear random-intercept mixed-effect model $\zeta_{i,m} = \delta B_{i,m} + \delta W_{i,m}$, in which the between-event residual $\delta B_{i,m}$ is the event-specific random effect deviation

(i.e., the bias), and $\delta W_{i,m}$ is the remaining (within-event) residual. $\delta B_{i,m}$ is zero mean with a standard deviation of $\tau_{i,m}$ (i.e., the between-event standard deviation), and $\delta W_{i,m}$ is zero mean with a standard deviation of $\phi_{i,m}$ (i.e., the within-event standard deviation). To construct a ShakeMap for an IM for which no data are available, we need to be able to estimate the mean and variance of $\delta B_{i,m}$ for any IM from data at other IMs.

Heteroskedasticity commonly enters GMMs through dependencies of ϕ on magnitude, V_{S30} , and distance, and dependencies of τ on magnitude. Some GMMs additionally include heteroskedasticity in τ that varies from site to site with parameters such as V_{S30} and the amplitude of the ground motion for reference rock conditions (Al Atik and Abrahamson, 2010). Here, we modify the Jayaram and Baker (2010) equations for the mean and variance of the bias to include data from any period:

$$\mu_{\delta B_{i,m}} = \frac{\mathbf{Z}_i^T \Sigma_{\mathbf{Y}_2}^{-1} \boldsymbol{\zeta}_i}{\tau_{i,m}^{-2} + \mathbf{Z}_i^T \Sigma_{\mathbf{Y}_2}^{-1} \mathbf{Z}_i}, \quad \text{and} \quad (11)$$

$$\sigma_{\delta B_{i,m}}^2 = \frac{1}{\tau_{i,m}^{-2} + \mathbf{Z}_i^T \Sigma_{\mathbf{Y}_2}^{-1} \mathbf{Z}_i}, \quad (12)$$

in which $\tau_{i,m}$ is the τ corresponding to the i th IM and m th site, and $\Sigma_{\mathbf{Y}_2}$ is the covariance matrix of the within-event residuals (i.e., $(\Sigma_{\mathbf{Y}_2})_{i,j} = \rho_{i,j} \phi_i \phi_j$). For a linear event-specific bias, \mathbf{Z}_i is a column vector of ones (Jayaram and Baker, 2010). But evaluating the random effect for all data (at all IMs and locations) gives a single bias term across all

IMs. Typically, IMs are analyzed separately, and thus each IM has a different bias, and a single bias representative of all IMs is less useful. We can compute an IM-specific bias term by weighting the residuals (and \mathbf{Z}_i) by the correlation between the IM of the observation and the target IM.

The variance of the within-event residual, accounting for the uncertainty in the estimate of the between-event residual, is thus

$$\sigma_{\delta W_{i,m}}^2 = \phi_{i,m}^2 + \sigma_{\delta B_{i,m}}^2. \quad (13)$$

To ease later calculations, let us further transform the within-event residuals, normalizing by the total within-event standard deviations:

$$X_{i,m} \equiv \frac{\delta W_{i,m}}{\sigma_{\delta W_{i,m}}}, \quad (14)$$

thus ensuring $X_{i,m}$ is standard normal.

Assumptions

The correlation structure for IMs of differing types at differing locations can be reasonably assumed as Markovian in nature, which can be expressed as

$$\rho_{X_{i,m}X_{j,n}} = \rho_{X_{i,m}X_{j,m}} \times \rho_{X_{j,m}X_{j,n}}. \quad (15)$$

That is, the correlation between differing IMs at differing locations is the product of the cross correlation of IMs i and j at the same location ($\rho_{X_{i,m}X_{j,m}}$) and the spatial correlation due to the distance between sites m and n ($\rho_{X_{j,m}X_{j,n}}$). This is an approximation, but a reasonable one (Loth and Baker, 2010). This assumption implies a conditional independence of the inter-IM correlation and the spatial correlation of a given IM. This conditional independence will simplify some of the calculations below, but it is not needed in general. There is a slight nonuniqueness problem to this formulation, because the second term in the equation could be either $\rho_{X_{j,m}X_{j,n}}$ or $\rho_{X_{i,m}X_{i,n}}$; that is, the spatial correlation could be taken for either the i th or j th IM. Goda and Hong (2008) use the larger of the two values, and Loth and Baker (2010) find that choice to be more consistent with a full spatial cross-correlation model. Direct spatial cross-correlation models could also be used (e.g., Loth and Baker, 2010, for PGA and SA), and the assumption of conditional independence would then not be needed, though the convenience of conditional independence would also be lost.

The covariance matrices must be positive definite. As discussed by Loth and Baker (2010) and Bradley (2012), this is not guaranteed for all possible correlation models. Loth and Baker (2010) address this by selecting functional forms that ensure the resulting covariance matrices are positive definite. However, if one does encounter a non-positive-definite covariance matrix, Bradley (2012) suggests using the method by Higham (2002) to slightly modify the matrix to make it positive definite before proceeding with further calculations.

Conditioning on Observations

Consider \mathbf{X} (defined in equation 14) as the vector of all normalized within-event residuals of interest, which is composed of N observed (\mathbf{X}_2) residuals (from N locations) and M predicted residuals (\mathbf{X}_1) for M locations, for P observed and Q desired IMs, analogous to equation (1). This vector is standard MVN. The mean vector for the nonobserved locations is

$$\boldsymbol{\mu}_{\mathbf{X}_1} = \begin{pmatrix} \mu_{X_{1,1}} \\ \mu_{X_{1,2}} \\ \vdots \\ \mu_{X_{1,M}} \\ \mu_{X_{2,1}} \\ \mu_{X_{2,2}} \\ \vdots \\ \mu_{X_{2,M}} \\ \vdots \\ \mu_{X_{Q,M}} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \quad (16)$$

which has QM elements. The $\boldsymbol{\mu}_{\mathbf{X}_2}$ vector is also zeros and is constructed in a similar fashion as $\boldsymbol{\mu}_{\mathbf{X}_1}$. However, not all P -observed IMs are necessarily available at all N sites. For the sake of this discussion, we define the number of elements in $\boldsymbol{\mu}_{\mathbf{X}_2}$ as R .

The covariance matrix is a modification of $\Sigma_{\mathbf{Y}_1\mathbf{Y}_1}$ (given in equation 3):

$$\Sigma_{\mathbf{X}_1\mathbf{X}_1} = \begin{pmatrix} 1 & \rho_{X_{1,2}X_{2,1}} & \cdots & \rho_{X_{Q,M}X_{M,Q}} \\ \rho_{X_{2,1}X_{1,2}} & 1 & & \vdots \\ \vdots & & \ddots & \\ \rho_{X_{M,Q}X_{Q,M}} & \cdots & & 1 \end{pmatrix}. \quad (17)$$

Note that $\Sigma_{\mathbf{X}_2\mathbf{X}_2}$ is similarly constructed but has dimensions $R \times R$, whereas $\Sigma_{\mathbf{X}_1\mathbf{X}_2}$ has dimensions $(MQ) \times R$. The means of the MQ predictions of \mathbf{X} conditioned on the R observations are given by a simplification of equation (5):

$$\boldsymbol{\mu}_{\mathbf{X}_1|\mathbf{x}_2} = \begin{pmatrix} \Sigma_{\mathbf{X}_1\mathbf{X}_2} & \Sigma_{\mathbf{X}_2\mathbf{X}_2}^{-1} \end{pmatrix} \mathbf{x}_2. \quad (18)$$

The conditional covariance matrix is

$$\Sigma_{\mathbf{X}_1\mathbf{X}_1|\mathbf{x}_2} = \Sigma_{\mathbf{X}_1\mathbf{X}_1} - \Sigma_{\mathbf{X}_1\mathbf{X}_2} \Sigma_{\mathbf{X}_2\mathbf{X}_2}^{-1} \Sigma_{\mathbf{X}_2\mathbf{X}_1}. \quad (19)$$

To gain an intuitive understanding of the above equations, we can consider the bivariate case, in which we observed $X_2 = x_2$ and are interested in the conditional distribution of X_1 . Equation (18) simplifies to

$$\mu_{X_1|x_2} = \rho_{X_1X_2}x_2. \quad (20)$$

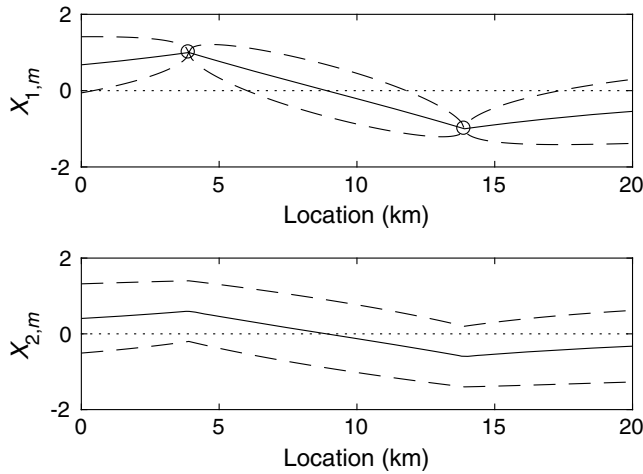


Figure 2. Example conditional ground-motion fields. Circles show observed values, solid lines show conditional mean values, and dashed lines show conditional mean $\pm 1\sigma$ values. Dotted lines at $X = 0$ indicate the original mean value before conditioning.

This is a simplification of equation (8), which is possible because of the normalization in equation (14). From equation (20), we see that the mean value of X_1 is influenced in proportion to $\rho_{X_1 X_2}$. For perfect correlation ($\rho_{X_1 X_2} = 1$), the conditional mean of X_1 will be equal to x_2 . For no correlation, the conditional mean of X_1 will be zero (i.e., the same as with no observation of X_2). Equation 19 simplifies to

$$\sigma_{X_1|x_2}^2 = 1 - \rho_{X_1 X_2}^2. \quad (21)$$

In this case, for perfect correlation, the conditional variance will be 0, and for no correlation, the conditional variance will be 1 (i.e., the same as with no observation of X_2).

Once $\mu_{X_1|x_2}$ and $\Sigma_{X_1 X_1|x_2}$ are computed, they must be scaled back into the appropriate natural IM variables, that is, $\mu_{\text{IM}_1|\text{im}_2}$ and $\Sigma_{\text{IM}_1|\text{im}_2}$. First, we define three MQ column vectors: μ_{IM} for the mean value of the IMs from the GMM, $\mu_{\delta B}$ for the between-event residual, and $\mu_{\delta W}$ for the within-event standard deviations. The i th element of each of these vectors corresponds to the IM and location of the i th element of \mathbf{X}_1 . The conditional mean is then

$$\mu_{\text{IM}_1|\text{im}_2} = \mu_{\text{IM}} + \mu_{\delta B} + \sigma_{\delta W} \odot \mu_{X_1|x_2}, \quad (22)$$

in which \odot represents the Hadamard (element-by-element) product. The conditional covariance matrix is given by

$$\Sigma_{\text{IM}_1|\text{im}_2} = \sigma_{\delta W} \sigma_{\delta W}^T \odot \Sigma_{X_1 X_1|x_2}. \quad (23)$$

Numerical Example

The following provides numerical examples based on the solution of equations (18) and (19), for slightly more realistic cases than the bivariate case above. MATLAB code pro-

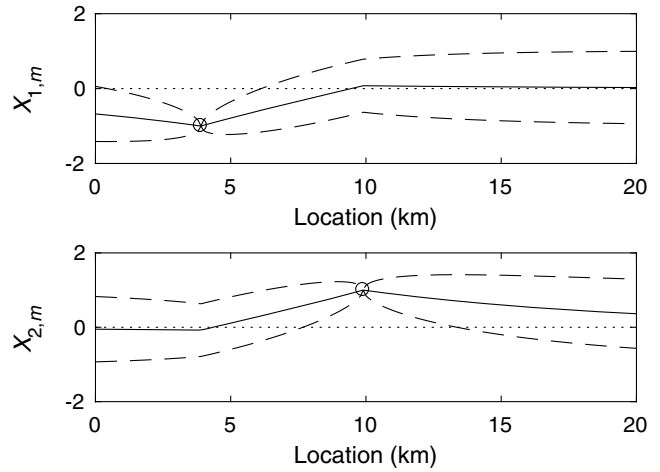


Figure 3. A second example of conditional ground-motion fields. Circles show observed values, solid lines show conditional mean values, and dashed lines show conditional mean $\pm \sigma$ values. Dotted lines at $X = 0$ indicate the original mean value before conditioning.

ducing these calculations is available for review and testing of other cases (see [Data and Resources](#)).

For these results, consider the following inputs: there are two IMs with $\rho_{X_{1,m} X_{2,m}} = 0.6$. For a given IM, the spatial correlation of residuals is given by

$$\rho_{X_{1,m} X_{1,n}} = \exp(-h/10), \quad (24)$$

in which h is the separation distance between the two sites, in units of kilometers. Using these two formulas, the covariance matrix for all locations of interest can be computed. (The correlation formulas themselves were chosen to be illustrative and do not represent any particular real-world model.)

Figure 2 shows example results to indicate the effect of these correlations. There are two observed values of $X_{1,m}$ in this case. This numerical example is representative of the practical situation in which $X_{1,m}$ is MI at sites m , and $X_{2,m}$ is PGA at sites m , in which MI is taken from the online DYFI reports, but no instrumental data are available. We see that $X_{1,m}$ passes through the observed values and decays toward $X = 0$, away from the observations. The $\pm 1\sigma$ values for $X_{1,m}$ pinch near the observations and broaden at other locations where the correlation is weaker. $X_{2,m}$ follows similar trends, though both the mean and σ are less affected by the observations, given the reduction in correlation across different IMs.

A second example is shown in Figure 3. In this case, a single observation of each IM type is shown, and the observations are opposite in sign from each other. We see that at location values between 4 and 10 km, both IMs interpolate between the observations in a relatively smooth manner. For $X_{1,m}$ values with location > 10 km, and $X_{2,m}$ values with location < 4 km, the conditional mean values are very close to zero, indicating the offsetting impacts of the two conditioning data points in those circumstances.

These two simple examples indicate that the conditioning scheme above produces intuitive results. The above mathematical formulations can easily be extended to 2D coordinate geometries and to more than two IM types. The simplifications of the examples were simply to ease the setup and plotting.

Accommodating Uncertain Measurements

The general MVN equations given earlier (i.e., equations 5 and 6) do not explicitly accommodate noisy or uncertain observations. It is not always the case, however, that the observations are known precisely. Frequently, the observed ground motions upon which we wish to condition our result have nonzero uncertainty, characterized by a mean and standard deviation. To accommodate a noisy observation (Z) we assume that its distribution with respect to the modeled value of the quantity Y_2 is given by

$$Z = Y_2 + \epsilon, \quad (25)$$

in which ϵ is an error term with zero mean and standard deviation σ_ϵ . This definition implicitly assumes an unbiased model. Such an assumption could be violated if, for instance, a model was used outside the geographic region for which it was developed. The problem then becomes one of how to unbiased the model, which is beyond the scope of this article. Given the noise model in equation (25), and assuming Y_2 and ϵ are independent, propagation of uncertainty results in

$$\mu_Z = \mu_{Y_2}, \quad \text{and} \quad (26)$$

$$\sigma_Z^2 = \sigma_{Y_2}^2 + \sigma_\epsilon^2. \quad (27)$$

Substituting Z for Y_2 , in equation (8), the bivariate mean can be written as

$$\mu_{Y_1|z} = \mu_{Y_1} + \frac{\sigma_{Y_1}}{\sigma_Z} \rho_{Y_1Z} (z - \mu_Z) \quad (28)$$

$$= \mu_{Y_1} + \frac{\sigma_{Y_1} \sigma_{Y_2} \sigma_Z}{\sigma_{Y_2} \sigma_Z^2} \rho_{Y_1Z} (z - \mu_Z). \quad (29)$$

Substituting equations (26) and (27) gives

$$\mu_{Y_1|z} = \mu_{Y_1} + \frac{\sigma_{Y_1} \sigma_{Y_2} \sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}}{\sigma_{Y_2} (\sigma_{Y_2}^2 + \sigma_\epsilon^2)} \rho_{Y_1Z} (z - \mu_{Y_2}) \quad (30)$$

$$= \mu_{Y_1} + \frac{\sigma_{Y_1}}{\sigma_{Y_2}} \frac{\sigma_{Y_2}}{\sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}} \rho_{Y_1Z} (z - \mu_{Y_2}). \quad (31)$$

This leads to the definition of an adjustment factor

$$\omega = \sqrt{\frac{\sigma_{Y_2}^2}{\sigma_{Y_2}^2 + \sigma_\epsilon^2}}, \quad (32)$$

which gives

$$\mu_{Y_1|z} = \mu_{Y_1} + \frac{\sigma_{Y_1}}{\sigma_{Y_2}} \omega \rho_{Y_1Z} (z - \mu_{Y_2}). \quad (33)$$

By definition

$$\rho_{Y_1Z} = \frac{\text{cov}(Y_1, Z)}{\sigma_{Y_1} \sigma_Z}, \quad (34)$$

substituting equations (25) and (27) and then simplifying gives

$$\rho_{Y_1Z} = \frac{\text{cov}(Y_1, Y_2 + \epsilon)}{\sigma_{Y_1} \sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}} \quad (35)$$

$$= \frac{\text{cov}(Y_1, Y_2) + \text{cov}(Y_1, \epsilon)}{\sigma_{Y_1} \sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}}. \quad (36)$$

But because Y_1 and ϵ are assumed to be uncorrelated, $\text{cov}(Y_1, \epsilon) = 0$. That gives

$$\rho_{Y_1Z} = \frac{\text{cov}(Y_1, Y_2)}{\sigma_{Y_1} \sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}} \quad (37)$$

$$= \frac{\sigma_{Y_2}}{\sqrt{\sigma_{Y_2}^2 + \sigma_\epsilon^2}} \frac{\text{cov}(Y_1, Y_2)}{\sigma_{Y_1} \sigma_{Y_2}} \quad (38)$$

$$= \omega \rho_{Y_1Y_2}. \quad (39)$$

Thus, the mean and variance of the bivariate normal distribution with noise can be written as

$$\mu_{Y_1|z} = \mu_{Y_1} + \frac{\sigma_{Y_1}}{\sigma_{Y_2}} \omega^2 \rho_{Y_1Y_2} (z - \mu_{Y_2}), \quad \text{and} \quad (40)$$

$$\sigma_{Y_1|y_2}^2 = (1 - \omega^2 \rho_{Y_1Y_2}^2) \sigma_{Y_1}^2, \quad (41)$$

which reduce to the familiar equations (8) and (9) when $\sigma_\epsilon = 0$.

To gain some intuitive understanding of these equations, we consider the bivariate case in which the observation is at the site we wish to estimate, so $\mu_{Y_1} = \mu_{Y_2} = \mu$, $\sigma_{Y_1} = \sigma_{Y_2} = \sigma$, and $\rho_{Y_1Y_2} = 1$. Equations (40) and (41) reduce to

$$\mu|z = \mu + \frac{\sigma^2}{\sigma^2 + \sigma_\epsilon^2} (z - \mu), \quad \text{and} \quad (42)$$

$$\sigma^2|z = \frac{\sigma^2 \sigma_\epsilon^2}{\sigma^2 + \sigma_\epsilon^2}. \quad (43)$$

This formulation of the mean and variance (i.e., equations 42 and 43) has a number of desirable features listed below and illustrated in Figure 4.

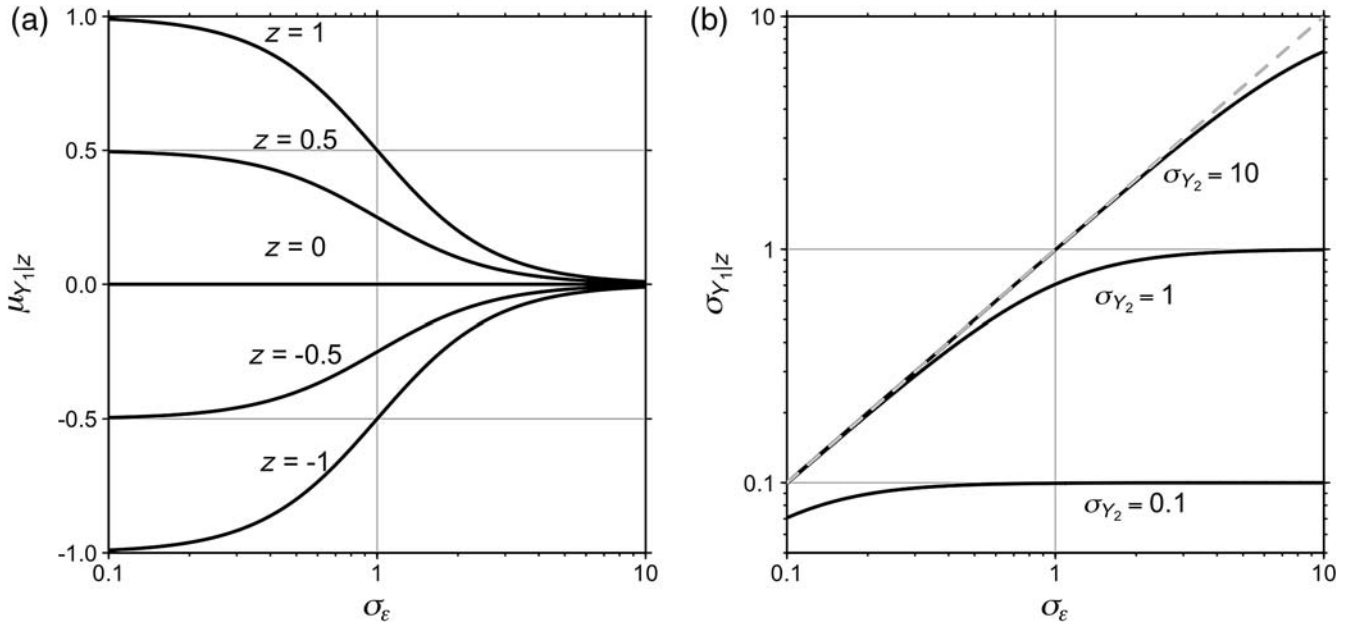


Figure 4. Illustration of the effects of σ_ϵ . (a) Plot of $\mu_{Y_1|z}$ as a function of σ_ϵ for a range of assumed values of z from -1 to 1 , assuming $\mu_{Y_2} = 0$ and $\sigma_{Y_2}^2 = 1$. (b) Plot of $\sigma_{Y_1|z}$ as a function of σ_ϵ for a range of assumed values of $\sigma_{Y_2}^2$ from 0.1 to 10 . The dashed gray line in (b) is for $\sigma_{Y_1|z} = \sigma_\epsilon$.

1. When σ_ϵ is zero (indicating a precise observation z), then equation (42) reduces to $\mu|z = z$. Similarly, equation (43) becomes $\sigma^2|z = 0$, confirming our precise knowledge of z .
2. When $\sigma_\epsilon \rightarrow \infty$ (i.e., an extremely unreliable observation), equation (42) reduces to $\mu|z = \mu$. Equation (43) becomes (in the limit) $\sigma^2|z = \sigma^2$. Thus, very unreliable observations have minimal impact on our initial estimates.
3. When $\sigma_\epsilon = \sigma$, the observation and estimate are given equal weight and equation (42) becomes $\mu|z = (\mu + z)/2$.
4. Equation (43) is symmetric in σ and σ_ϵ , as one would expect, given the interchangeability of the estimate and the observation when both are distributions.
5. From equation (43), $\sigma|z$ is always less than the lesser of σ and σ_ϵ , validating our intuition that an observation, even a noisy one, will improve our knowledge of the variable in question.

This result depends upon the independence of the additional uncertainty σ_ϵ and the correlation expressed through ρ . Thus, this formulation is less applicable when σ_ϵ stems from spatial variability than when it emerges from instrumental or processing uncertainty, for example. It may be possible to construct a model that better accommodates spatial uncertainty, but it would likely need to include that uncertainty more organically in the correlation model itself. This is analogous to eliminating the Markovian assumption of a compound correlation function (as discussed earlier) by directly modeling spatial cross correlations.

Generalizing equations (40) and (41) to the multivariate case requires generating a matrix of adjustment factors. An ω can be computed for each element in our random variable (equation 1):

$$\omega_{Y_i} = \sqrt{\frac{\sigma_{Y_i}^2}{\sigma_{Y_i}^2 + \sigma_{\epsilon, Y_i}^2}}, \quad (44)$$

in which σ_{ϵ, Y_i}^2 is the variance of the observation at Y_i , giving

$$\omega = \begin{Bmatrix} \omega_{Y_1} \\ \omega_{Y_2} \end{Bmatrix} = \begin{Bmatrix} J_{Y_1} \\ \omega_{Y_2} \end{Bmatrix}, \quad (45)$$

in which J_{Y_1} is a vector of ones the same length as Y_1 . The matrix of adjustment factors is then:

$$\Omega = \omega \omega^T = \begin{bmatrix} \Omega_{Y_1 Y_1} & | & \Omega_{Y_1 Y_2} \\ \hline \Omega_{Y_2 Y_1} & | & \Omega_{Y_2 Y_2} \end{bmatrix}, \quad (46)$$

in which the diagonal elements $\Omega_{i,i}$ should be set to 1 because the elements are perfectly correlated with themselves. We can then generate a modified covariance matrix Σ' by multiplication of the original covariance matrix Σ with the adjustment factors:

$$\Sigma' = \begin{bmatrix} \Sigma'_{Y_1 Y_1} & | & \Sigma'_{Y_1 Y_2} \\ \hline \Sigma'_{Y_2 Y_1} & | & \Sigma'_{Y_2 Y_2} \end{bmatrix} = \Omega \odot \Sigma. \quad (47)$$

The ground-motion residuals are also weighted by the corresponding adjustment factor

$$\zeta' = \omega_{Y_2} \odot \zeta. \quad (48)$$

Then, equations (5) and (6) become

$$\mu_{Y_1|Y_2} = \mu_{Y_1} + \Sigma'_{Y_1 Y_2} \Sigma'^{-1}_{Y_2 Y_2} \zeta', \quad \text{and} \quad (49)$$

$$\Sigma_{Y_1 Y_1|Y_2} = \Sigma_{Y_1 Y_1} - \Sigma'_{Y_1 Y_2} \Sigma'^{-1}_{Y_2 Y_2} \Sigma'_{Y_2 Y_1}. \quad (50)$$

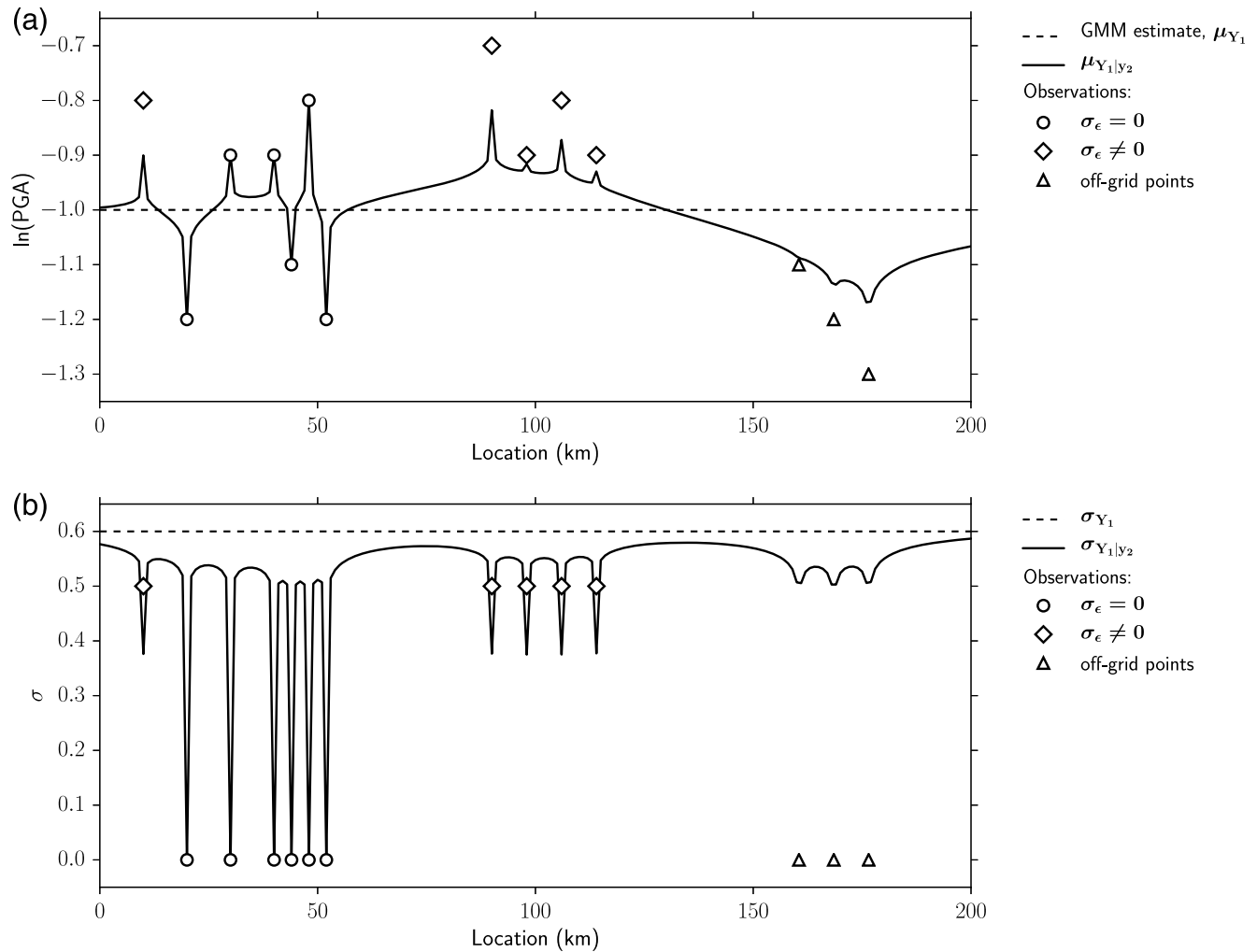



Figure 5. A 1D synthetic example of the multivariate normal (MVN) with additional uncertainty. (a) Ground motion $\ln(\text{PGA})$ as a function of position. (b) Uncertainty (standard deviation) as a function of position. See the [Accommodating Uncertain Measurements](#) section for a discussion. PGA, peak ground acceleration.

Note that $\Omega_{Y_1 Y_1}$ is all ones, and therefore does not modify $\Sigma_{Y_1 Y_1}$.

Figure 5 illustrates a 1D synthetic example of the conditional multivariate process in which some of the observations have additional uncertainty. Figure 5a shows predicted amplitudes, and Figure 5b shows their standard deviations. The GMM estimates are shown as a dashed black line ($\mu_{Y_1} = -1, \sigma_{Y_1} = 0.6$), whereas the conditional MVN distribution mean ($\mu_{Y_1|y_2}$) and standard deviation ($\sigma_{Y_1|y_2}$) are represented by the solid black lines. The markers indicate the location of station data: circles are stations with no additional uncertainty ($\sigma_\epsilon = 0$), and diamonds have additional uncertainty ($\sigma_\epsilon = 0.5$). Both the circles and diamonds represent stations that are collocated with points in the output grid. The triangles represent stations that do not have additional uncertainty but that are located halfway between output points (1 km spacing); in this case, the predicted values do not match the observations because the locations are not the same. In Figure 5a, the markers indicate the observed ampli-

tudes (y_2), and in Figure 5b they indicate the amount of additional uncertainty (σ_ϵ). The spatial correlation model used was that of [Goda and Atkinson \(2010\)](#).

The  electronic supplement to this article includes a Python program, similar to that which produced Figure 5, demonstrating the evaluation of the MVN with additional uncertainty.

Discussion

[Gehl et al. \(2017\)](#) propose a method for conditioning ground-motion estimates on observations using Bayesian networks (BNs). The BN approach addresses many of the important issues that we address with the MVN. [Gehl et al. \(2017\)](#) compare the BN with an implementation of the MVN, but that MVN implementation is different from ours, highlighting the importance of the details of the MVN calculation. The difference between the MVN implementations is particularly evident in the standard deviations given in figure 5 of [Gehl et al. \(2017\)](#). At the location of an observation,

where the within-event conditional covariance is zero, [Gehl et al. \(2017\)](#) add the between-event standard deviation (τ), resulting in the standard deviation of the conditional mean at an observation becoming equal to τ . In contrast, we use $\sigma_{\delta W_{i,m}}$ (equation 13) in the normalization of X (equation 14), which allows for the conditional variance ($\sigma_{X_1|x_2}^2$) to exhibit the following characteristics:

- $\sigma_{X_1|x_2}^2$ is $\tau^2 + \phi^2$ in the absence of data;
- $\sigma_{X_1|x_2}^2$ approaches zero as the prediction location approaches the observation location;
- $\sigma_{X_1|x_2}^2$ approaches ϕ^2 at locations distant to observations, for events with enough observation that the variance of the event term ($\sigma_{\delta B_{i,m}}^2$) approaches zero; and
- for situations intermediate to the above end members, $\sigma_{X_1|x_2}^2$ exhibits a smooth transition in terms of both the observation-prediction distance and the total number of observations.

These trends are consistent with the BN results demonstrated by [Gehl et al. \(2017\)](#) in their figure 5, except that the figure does not address the dependence of $\sigma_{X_1|x_2}^2$ on the number of observations. Other limitations of the MVN identified by [Gehl et al. \(2017\)](#), such as the inclusion of multiple IMs and macroseismic data, are exactly the limitations that we addressed in this article. [Gehl et al. \(2017\)](#) also raise the issue of including epistemic uncertainty; in our opinion, this can be handled in either approach, for example, by combining multiple GMMs into a single model that reflects the variability in the means and standard deviations of the individual GMMs.

We note that for computer memory or performance reasons, the computation of the conditional mean for the different IMs can proceed completely independently of one another with no loss of information. That is, one could divide the vector μ_{X_1} (equation 16) into any number of subsets, revise the mean and covariance matrices accordingly, and compute the components of $\mu_{X_1|x_2}$ sequentially or in parallel. The diagonal elements of the conditional covariance matrix (the variances) may also be calculated piecemeal. The full formulation is required only if the off-diagonal elements of the conditional covariance matrix are needed.

The MVN approach outlined here is appropriate for producing interpolated values at specific geographic points. If an areal average is required (e.g., for modeling the losses of aggregate inventories), then the method discussed by [Stafford \(2012\)](#), which revises the variances and covariances to account for the spatial averaging of the ground motions, may be required.

Summary and Conclusions

In situations where certain IMs are available at a set of sites, the MVN may be applied to estimate additional IMs and IMs at other sites. This approach requires the availability of a GMM, a cross-correlation model among the IMs in question, and the assumption that the IM residuals (relative to the GMM) have an MVN distribution. If some IMs are available

and others are desired at the same locations, then it is sufficient to estimate the additional IMs from GMM and an inter-period correlation model such as [Baker and Jayaram \(2008\)](#). Analogously, given a GMM and a spatial correlation model such as [Goda and Atkinson \(2010\)](#), a specific IM may be estimated at geographic locations remote from the observation sites. More generally, with a cross-correlation model such as [Loth and Baker \(2010\)](#), one may use the MVN to estimate ground-motion distributions for a variety of IMs and locations, using a mixture of observed IMs.

These techniques may also be applied to other IMs with normal or lognormal distributions, such as significant duration, Arias intensity, and cumulative average velocity. This is, however, not possible at this time because the necessary spatial cross-correlation relations have not yet been developed. Other parameters, such as the time-averaged shear-wave velocity to 30 m depth (i.e., V_{S30}), may also be interpolated by these techniques, though we have not investigated those applications here.

The MVN, as typically presented, treats the conditioning observations as exact (the variance goes to 0 as the correlation approaches unity). It is often the case, however, that our observations have significant levels of uncertainty. For example, MI observations represent an areal average compiled from a number of damage reports. Also, new types of mass-deployed seismic instruments (e.g., [Cochran et al., 2009](#); [Evans et al., 2014](#); [Clayton et al., 2015](#)) may provide accurate information in aggregate but will represent samples of a distribution when treated individually. One may also imagine other sources of noise, found either in the acquisition or processing of seismic data, that add uncertainty to the observations (uncertainty in the orientation or location of sensors, for instance). We envision that this method of interpolation could be useful for other variables, such as V_{S30} , in which different survey methods are associated with different levels of precision, and in extreme cases, V_{S30} could be converted from indirect information, such as standard penetration-test blow counts. An analogous example is the conversion between MI and SA ([Worden et al., 2012](#)). An alternative to converting one variable to another and accounting for the uncertainty of the conversion is to develop the relevant cross correlation models. These are, however, frequently unavailable, and so it is useful to maintain the flexibility to employ both approaches. The techniques presented here allow for the incorporation of these noisy data into the conditional MVN.

Data and Resources

No data were used in this article. The MATLAB code for constructing Figures 2 and 3 is available at <https://github.com/bakerjw/ground-motion-interpolation> (last accessed July 2017).

Acknowledgments

B. A. B. was supported by QuakeCore: The New Zealand Center for Earthquake Resilience. The authors thank Albert Kottke for alerting them to

Kishida (2017) prior to its publication and Tadahiro Kishida for providing them with a preprint of that paper. Discussions with Peter Stafford were helpful in understanding the role of heteroskedastic τ . The authors thank Sarah Minson, Adrian Rodriguez-Marek, and two anonymous reviewers for helpful and insightful reviews.

References

- Al Atik, L., and N. Abrahamson (2010). Nonlinear site response effects on the standard deviations of predicted ground motions, *Bull. Seismol. Soc. Am.* **100**, no. 3, 1288–1292.
- Baker, J. W. (2011). Conditional mean spectrum: Tool for ground-motion selection, *J. Struct. Eng.* **137**, no. 3, 322–331.
- Baker, J. W., and N. Jayaram (2008). Correlation of spectral acceleration values from NGA ground motion models, *Earthq. Spectra* **24**, no. 1, 299–317.
- Bommer, J., and P. Stafford (2012). Estimating ground motion levels in earthquake damage investigations: A framework for forensic engineering seismology, *Int. J. Forensic Eng.* **1**, 3–20.
- Boore, D. M., and T. Kishida (2017). Relations between some horizontal-component ground-motion intensity measures used in practice, *Bull. Seismol. Soc. Am.* **107**, no. 1, 334–343.
- Bradley, B. A. (2010). A generalized conditional intensity measure approach and holistic ground-motion selection, *Earthq. Eng. Struct. Dynam.* **39**, no. 12, 1321–1342.
- Bradley, B. A. (2012). A ground motion selection algorithm based on the generalized conditional intensity measure approach, *Soil Dynam. Earthq. Eng.* **40**, 48–61.
- Bradley, B. A. (2014). Site-specific and spatially-distributed ground-motion intensity estimation in the 2010–2011 Canterbury earthquakes, *Soil Dynam. Earthq. Eng.* **61/62**, 83–91.
- Chiou, B. S.-J., and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **30**, no. 3, 1117–1153.
- Clayton, R. W., T. Heaton, M. Kohler, M. Chandy, R. Guy, and J. Bunn (2015). Community seismic network: A dense array to sense earthquake strong motion, *Seismol. Res. Lett.* **86**, no. 5, 1354–1363, doi: [10.1785/0220150094](https://doi.org/10.1785/0220150094).
- Cochran, E. S., J. F. Lawrence, C. Christensen, and R. S. Jukka (2009). The quake-catcher network: Citizen science expanding seismic horizons, *Seismol. Res. Lett.* **80**, no. 1, 26–30.
- Evans, J., R. M. Allen, A. Chung, E. Cochran, R. Guy, M. Hellweg, and J. Lawrence (2014). Performance of several low-cost accelerometers, *Seismol. Res. Lett.* **85**, no. 1, 147–158.
- Gehl, P., J. Douglas, and D. D'Ayala (2017). Inferring earthquake ground-motion fields with Bayesian networks, *Bull. Seismol. Soc. Am.* doi: [10.1785/0120170073](https://doi.org/10.1785/0120170073).
- Goda, K., and G. M. Atkinson (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data, *Bull. Seismol. Soc. Am.* **100**, no. 6, 3055–3067.
- Goda, K., and H.-P. Hong (2008). Spatial correlation of peak ground motions and response spectra, *Bull. Seismol. Soc. Am.* **98**, no. 1, 354–365.
- Higham, N. J. (2002). Computing the nearest correlation matrix—A problem from finance, *IMA J. Numer. Anal.* **22**, no. 3, 329–343, doi: [10.1093/imanum/22.3.329](https://doi.org/10.1093/imanum/22.3.329).
- Jayaram, N., and J. W. Baker (2008). Statistical tests of the joint distribution of spectral acceleration values, *Bull. Seismol. Soc. Am.* **98**, no. 5, 2231–2243.
- Jayaram, N., and J. W. Baker (2010). Considering spatial correlation in mixed-effects regression and the impact on ground-motion models, *Bull. Seismol. Soc. Am.* **100**, no. 6, 3295–3303.
- Johnson, R. A., and D. W. Wichern (2002). *Applied Multivariate Statistical Analysis*, Vol. 5, Prentice Hall, Upper Saddle River, New Jersey.
- Kishida, T. (2017). Conditional mean spectra given a vector of spectral accelerations at multiple periods, *Earthq. Spectra* **33**, no. 2, 469–479.
- Loth, C., and J. W. Baker (2010). A spatial cross-correlation model of spectral accelerations at multiple periods, *Earthq. Eng. Struct. Dynam.* **42**, no. 3, 397–417.
- 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 *Applications of Statistics and Probability in Civil Engineering*, Taylor & Francis Group, London, England, 1–8.
- 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 Eng.* **1**, no. 1, 58–77.
- Vanmarcke, E. (1983). *Random Fields, Analysis and Synthesis*, The MIT Press, Cambridge, Massachusetts.
- Wald, D. J., V. Quitoriano, C. B. Worden, M. Hopper, and J. W. Dewey (2012). USGS “Did You Feel It?” Internet-based macroseismic intensity maps, *Ann. Geophys.* **54**, no. 6, doi: [10.4401/ag-5354](https://doi.org/10.4401/ag-5354).
- Worden, C. B., and D. J. Wald (2016). *ShakeMap Manual Online: Technical Manual, User's Guide, and Software Guide*, U.S. Geol. Surv. doi: [10.5066/F7D21VPQ](https://doi.org/10.5066/F7D21VPQ).
- Worden, C. B., M. C. Gerstenberger, D. A. Rhoades, and D. J. Wald (2012). Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California, *Bull. Seismol. Soc. Am.* **102**, no. 1, 204–221.

U.S. Geological Survey
Denver Federal Center
PO Box 25046, MS 966
Denver, Colorado 80225
cbworden@usgs.gov
(C.B.W., E.M.T., N.L., D.J.W.)

Department of Civil and Environmental Engineering
Stanford University
Stanford, California 94305-4020
(J.W.B.)

Stanford University
Stanford, California 94305-4020
(B.A.B.)

Manuscript received 14 July 2017;
Published Online 13 February 2018