von Specht, S., Cotton, F. (2020): A Link between Machine Learning and Optimization in Ground-Motion Model Development: Weighted Mixed-Effects Regression with Data-Driven Probabilistic Earthquake Classification. - Bulletin of the Seismological Society of America, 110, 6, 2777-2800.

https://doi.org/10.1785/0120190133

# A Link between Machine Learning and Optimization in Ground-Motion Model Development: Weighted Mixed-Effects Regression with Data-Driven Probabilistic Earthquake Classification

**Sebastian von Specht*[1,2] and Fabrice Cotton[2,3]**

**ABSTRACT**

The steady increase of ground-motion data not only allows new possibilities but also comes with new challenges in the development of ground-motion models (GMMs). Data classification techniques (e.g., cluster analysis) do not only produce deterministic classifications but also probabilistic classifications (e.g., probabilities for each datum to belong to a given class or cluster). One challenge is the integration of such continuous classification in regressions for GMM development such as the widely used mixed-effects model. We address this issue by introducing an extension of the mixed-effects model to incorporate data weighting. The parameter estimation of the mixed-effects model, that is, fixed-effects coefficients of the GMMs and the random-effects variances, are based on the weighted likelihood function, which also provides analytic uncertainty estimates. The data weighting permits for earthquake classification beyond the classical, expert-driven, binary classification based, for example, on event depth, distance to trench, style of faulting, and fault dip angle. We apply Angular Classification with Expectation–maximization, an algorithm to identify clusters of nodal planes from focal mechanisms to differentiate between, for example, interface- and intraslab-type events. Classification is continuous, that is, no event belongs completely to one class, which is considered in the ground-motion modeling. The theoretical framework described in this article allows for a fully automatic calibration of ground-motion models using large databases with automated classification and processing of earthquake and ground-motion data. As an example, we developed a GMM on the basis of the GMM by Montalva et al. (2017) with data from the strong-motion flat file of Bastías and Montalva (2016) with ~2400 records from 319 events in the Chilean subduction zone. Our GMMwith the data-driven classification is comparable to the expert-classification-based model. Furthermore, the model shows temporal variations of the between-event residuals before and after large earthquakes in the region.

**KEY POINTS**

- We incorporate data-driven classification into mixed-effects regressions for ground-motion models.
- Continuous data weighting requires a regression based on the weighted likelihood.
- The incorporation of data weighting into mixed-effects models allows for more transparent data selection.

Supplemental Material

## INTRODUCTION

The rapid progress and expansion of strong-motion networks result in a steady increase of available ground-motion data that are beyond the limits of manual processing and necessitate automated processing. The increasing growth of data and the need for reproducible, objective (i.e., independent of expert judgment) seismic hazard and ground-motion models (GMMs) make classical classification methods difficult to use. Indeed, current classifications used in seismic hazard or ground-motion modeling are deterministic in the sense that earthquakes or sites are unambiguously placed into a single category (e.g., intraplate or interface earthquake). With the growth of data, properties and classifications of thousands of sites and earthquakes cannot anymore solely be based on expert opinions and deterministic classification, which are neglecting the uncertainties associated with the classification. Recently, the community is moving to data-driven and probabilistic classifications: for example, for GMM regionalization (Chen et al., 2018) or site classification (Kotha et al., 2018). These classifications have the advantage to be fully data driven, transparent, and probabilistic (i.e., uncertainties related to the classification are quantified). However, probabilistic classifications come new challenges for ground-motion modelers, because they require not only more generalized ground-motion regression techniques that consider both measurement errors (uncertainties of the ground-motion values) but also new types of uncertainties resulting from machine-learning and data-driven classifications. These new uncertainty

1. Earthquake Disaster & Risk Evaluation and Management (E-DREaM) Center, National Central University, Taoyuan City, Taiwan; 2. Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Potsdam, Germany; 3. Institute of Geosciences, University of Potsdam, Potsdam, Germany

*Corresponding author: specht@gfz-potsdam.de

evaluations have to be incorporated in the regression methods used not only to develop GMMs but also to quantify correlations for the GMM coefficients and the various components of the aleatory variability (e.g., between-event, within-event, and between-path variabilities).

The limitations of classical earthquake classifications and the need for new GMMs are particularly important for subduction areas, because GMMs for subduction-zone earthquakes show large aleatory variability and epistemic uncertainties (Douglas, 2010). One of the key challenges of the development of new ground-shaking models for subduction zones relates to the event classification. Subduction-zone GMMs differentiate between interface-type events occurring at the coupled interface of the subducting and the overriding plate, and intraslab-type events that take place within the downgoing slab. Classification into different source types into interface- and intraslab-type events is performed manually and is usually based on the depth of an event, the distance to the trench, style of faulting (SoF), and dip angle of the fault plane (e.g., Haendel et al., 2015; Abrahamson et al., 2016; Bastías and Montalva, 2016). These classifications are discrete, that is, an event is unambiguously assigned to a single category ("intraplate" or "interface") neglecting any uncertainties or degree of belief.

Although most reverse events in a subduction zone will be related to the interface and most normal events to the slab, there are possible exceptions. The behavior of earthquake slip in relation to the underlying stress field can be well approximated by the Wallace–Bott hypothesis, which states that slip on a pre-existing fault will be in the direction of maximum resolved shear stress (Wallace, 1951; Bott, 1959; Lisle, 2013). However, the resolved shear stress decreases slowly with angular distance from its maximum, as it is related to the cosine of the angle between slip orientation and shear stress orientation. It follows that the slip orientation can be in a range of $\pm 15°$ to the maximum shear stress orientation and still be at 90% or more of the amplitude of the maximum resolved shear stress (Lisle, 2013), resulting in a wide variety of potential slip orientations even on plain interfaces in homogeneous stress field. Many additional factors contribute to even larger deviations from an ideal slip orientation on the interface, in particular, uneven patches on the interface and local variations in the stresses (possibly related to other prior major earthquakes, asperities, and barriers).

In case of faulting in the slab, it is even more complicated, as faults have several origins (formation close to the mid-ocean ridges, during the bending before subduction initiation, and so on) resulting in a wide distribution of fault orientations. A slab dipping at a high angle can show stress tensor orientations very different from the time of fault generation when the slab was still oceanic crust near the surface, which may result in reverse faulting at large depths that are actually normal faulting at an oblique angle. Such apparent faulting reversal has been observed in the slab of the Philippine Sea plate under the Ryukyu microplate northeast of Taiwan where reverse faulting occurs at depths far below the plate interface (Kao and Rau, 1999).

In this study, we couple a machine-learning-based earthquake classification and a generalized ground-motion regression model to address data classification and selection for complex tectonic environments and the proper implementation of the classified data into the regression model. This new framework is tested and applied to the development of GMMs for the Chilean subduction zone. We first apply Angular Classification with Expectation–maximization (ACE), a method to efficiently identify clusters of nodal planes from focal mechanisms (FMs) as well as the SoF in one purely data-driven algorithm based on expectation–maximization with extensions in optimizing the number of clusters (Specht et al., 2017). ACE only requires the strike, rake, and dip of both nodal planes from FM catalogs to identify clusters belonging to different SoF. The method has been applied by von Specht et al. (2018) to stress tensor inversion in northern Chile to reduce uncertainties of the stress tensor estimates by selecting events associated with one SoF for the inversion.

ACE classification is continuous, that is, no event belongs completely to one class, and classification uncertainty is then evaluated. As stated earlier, it is currently not possible to identify with 100% accuracy the SoF and associate it with interface and intraslab activity, particularly for exotic events (i.e., with unusual FMs or in unusual stress environments). Therefore, with ACE we can say for instance that some events with differing FMs are to some extent interface as they fit to the bulk group of interfaces (intraslabs) but are also outliers in that group. This alleviates problems, in particular, with expert-defined hard cuts for rake intervals (e.g., in the 45°–135° range) when subduction is more complicated, by being oblique and/or several subduction systems that are spatially close.

In this article, the mixed-effects model is reviewed from its theoretical foundation not only to incorporate continuous data weights (like the ones from ACE) into the governing equations but also to provide a consistent work flow in the development of mixed-effects-based GMMs, as they are used in strong-motion seismology, in general. The parameter estimators of the mixed-effects model are derived from its basic assumptions and for different random effects (in GMMs namely the between-event, between-site, and between-path [region] variabilities) following the works of Henderson et al. (1959), Harville (1977), Lindstrom and Bates (1988), and Bates et al. (2015). Their work forms the basis of the state-of-the-art implementations of mixed-effects regression, in general.

The data weights are introduced into the model through the weighted likelihood (Field and Smith, 1994), which avoids weight-related scaling in the model uncertainties (the "sigma" of the GMM) as would be the case when the weights are incorporated in the data covariance matrix (e.g., Bates et al., 2015). Only data weights directly related to the data (i.e., with the same unit [e.g., acceleration] like in measurement errors or residual scaling for robust regression) should be included in the covariance matrix, as they directly influence and scale the model uncertainties. Data weights such as the ACE-based classification are not directly linked to the data (different units; probability vs. acceleration) and thus would inadvertently alter the outcome of the uncertainty estimates if included in the covariance matrix. Incorporation of such data weights in the likelihood function instead of the covariance matrix alleviates this alteration of the uncertainty estimates.

The algorithmic formulation given here is a generalization of the linear mixed-effects model of which the mixed-effects model by Abrahamson and Youngs (1992) is a special case. Furthermore, the formulations of the mixed-effects model used in this article also form the basis for the derivation of analytic solutions of parameter uncertainties and correlations for the GMM coefficients and the random-effects variances. The analytic uncertainty estimates are based on the Fisher information matrix and the Cramér–Rao inequality. This allows to investigate the role and interactions of GMM parameters for a given model and its random-effects variances.

We present results of the event classification for Chile showing the identification of clusters of thrust (associated with plate interface activity) and normal events (associated with intraslab seismicity). In the following step, we derive GMMs with events selected by expert judgment and ACE as a data-driven method and analyze the correlations not only among GMM parameters, variances but also between parameters and variances. We then compare the models with each other and also to the model of Montalva et al. (2017). Finally, we investigate the residuals and following Socquet et al. (2017), Bindi et al. (2018), and Piña-Valdés et al. (2018), we analyze the temporal variations of between-event residuals.

## METHOD

### *The weighted mixed-effects model*

The general form of the linear mixed-effects model is

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{p}) + \boldsymbol{B}\boldsymbol{q} + \epsilon \tag{1}$$

in which $\boldsymbol{f}(\boldsymbol{p})$ is the fixed-effects model with parameters $\boldsymbol{p}$. The model $\boldsymbol{f}(\boldsymbol{p})$ can be a nonlinear function of $\boldsymbol{p}$. The random effects $\boldsymbol{q}$ are linearly linked to the design matrix $\boldsymbol{B}$. This model is stated under the assumption that residuals of both the random and the fixed model are normally distributed, expressed by $\epsilon$. The form in equation (1) is the most commonly form encountered in strong-motion modeling.

The random effects are normally distributed with $\boldsymbol{q} \sim \mathcal{N}(0, \sigma^2\boldsymbol{D})$, in which $\sigma^2\boldsymbol{D}$ is the random-effects covariance matrix. The data $\boldsymbol{y}$ conditioned on the random effects $\boldsymbol{q}$ are assumed to be $\boldsymbol{y}|\boldsymbol{q} \sim \mathcal{N}(f(\boldsymbol{p}) + \boldsymbol{B}\boldsymbol{q}, \sigma^2\boldsymbol{C})$, in which $\sigma^2\boldsymbol{C}$ is the data covariance matrix. From the theorem related to the partition of sums of squares follows that the sum of variance(s) associated with $\boldsymbol{q}$ (explained sum(s) of squares; random-effects variances [between-event residuals, between-site residuals, and so forth]) and the variance of $\epsilon$ (residual sum of squares; within-event variance) is the total variance (total sum of squares) (e.g., Sahai and Ageel, 2000).

The goal of the mixed-effects regression is to maximize the likelihood of both the fixed-effects and random-effects models (Henderson et al., 1959; Harville, 1976). The joint probability is given by

$$\begin{aligned} &P(\boldsymbol{y}, \boldsymbol{q}|\boldsymbol{f}(\boldsymbol{p}), \sigma^2, \boldsymbol{C}, \boldsymbol{D}) \\ &\quad = \frac{1}{\sqrt{(2\pi)^{N+M}|\sigma^2\boldsymbol{C}||\sigma^2\boldsymbol{D}|}} \\ &\quad \times e^{\frac{1}{2\sigma^2}((\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{p})-\boldsymbol{B}\boldsymbol{q})^T\boldsymbol{C}^{-1}(\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{p})-\boldsymbol{B}\boldsymbol{q})^T\boldsymbol{q}^T\boldsymbol{D}^{-1}\boldsymbol{q})} \end{aligned} \tag{2}$$

The maximum-likelihood (ML) estimation of its parameters is given by Henderson et al. (1959), Harville (1977), Lindstrom and Bates (1988), and Bates et al. (2015).

At this point, we generalize the likelihood by introducing weights into the likelihood function. A weighted likelihood (e.g., Field and Smith, 1994; Markatou et al., 1998) is given by

$$\mathcal{L} = \prod_{i=1}^{N} P_i^{w_i}, \tag{3}$$

and reduces to the ordinary likelihood, if $w_i = 1$. A generalized likelihood for mixed effects has been proposed by (e.g., Wolfinger and O'Connell, 1993). This approach however has a very broad scope, for example, modifying the assumption of normally distributed residuals. Another widely used approach to weight incorporation in mixed-effects models is the penalized or weighted least-squares for mixed-effects model by Bates and DebRoy (2004) and Bates et al. (2015). They account for weights by stating them explicitly in the covariance matrix C. This approach requires that the magnitude of the weights is linked to the data, that is, the weights have the same unit of measurement as the data, for example, the error measurements of the modeled quantity at hand. Nonetheless, the weighted

**TABLE 1**
**Expression for Weighted Mean and Variance Based on Covariance Weighting (Left Column) and Weighted Likelihood (Right Column)**

| Parameter | Covariance Weights | Likelihood Weights |
|---|---|---|
| Mean $\hat{\mu}$ | $\frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$ | $\frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}$ |
| Variance $\hat{\sigma}^2$ | $\frac{\sum_{i=1}^{N} w_i (x_i - \hat{\mu})^2}{N}$ | $\frac{\sum_{i=1}^{N} w_i (x_i - \hat{\mu})^2}{\sum_{i=1}^{N} w_i}$ |

While the weighted mean estimators are identical, there is a difference in the normalization of the weighted variances. The covariance-based estimator of the variance scales with the weights $w_i$, the weighted likelihood based not. Derivation of the estimators is given in Appendix B.

likelihood does not require such direct link between data and their weights, because it is less constrained on the relation between weights and data, as the weights in equation (3) are not applied to the data but their probabilities. This less constrained relation between weights and data in the weighted likelihood is paramount when combining weights and data with different measurement units, for instance, probabilities as weights and ground-shaking data.

An illustrative example for the difference of weights in the covariance matrix and in the likelihood is given in Table 1 for the estimation of the weighted mean and variance, respectively (derivation is in Appendix A). The example represents the two fundamental components of the mixed model, namely the mean as representative for the fixed-effects part and the variance for the random-effects part. Although the mean estimators are identical under both weighting methods, the variance estimators show a small but important difference: for both variance estimators scale, the squared sums linearly with the weights, but the covariance-based weighted estimator normalizes by the number data, whereas the likelihood-based weighted estimator normalizes by the sum of weights. This means that the covariance weighted variance scales with the weights, whereas the weighted likelihood does not show such scaling behavior; a property of high importance in the development of GMMs and their subsequent application in probabilistic seismic hazard analysis.

The difference in both weighting methods implies that they are not mutually exclusive. On the contrary, both can be used simultaneously, each representing a different aspect of uncertainty, for example, measurement error and degree of belief.

Here, we specifically introduce weights for earthquakes, illustrated with a set of earthquakes, but the idea of weighting can be applied to any other setting. The occurrence of an earthquake can be viewed as a random effect (event) $e_i$, and it is recorded at $N_i$ random stations, that is, we have $N_i$ records $\boldsymbol{y} = \left(y_1, y_2, \dots, y_{N_i}\right)^T$. If we observe $E$ earthquakes, we have for the $i$th earthquake $N_i$ records $\boldsymbol{y}_i$. The number of records of an earthquake depends on several factors, for example, earthquake location, magnitude, and seismic station network, and the record quality can vary with each event and to some extent also with each station. When developing a GMM from these $E$ earthquakes, the records $\boldsymbol{y}$ are usually selected per earthquake. For GMMs, the selection of earthquakes often relates to the event type; particularly interface and intraslab events. Therefore, the seismologist (or some algorithm like ACE) applies a weighting per earthquake, that is, the $i$th event gets weight $w_i$. This weighting represents the seismologist's (or ACE's) degree of belief whether an event is interface or intraslab (or something else). The main difference between the seismologist and ACE is that the seismologist selects events on binary basis - an event is in or out - whereas ACE assigns continuous weights based on a probabilistic model. To summarize, the seismologist wishes to develop a GMM from $E$ earthquakes, in which the $i$th earthquake is recorded at $N_i$ stations, that is, with the total number of records $N = \sum_{i=1}^{E} N_i$, and all records of the $i$th event are weighted with weight $w_i$. This exemplifying scenario is probably the most widely encountered in ground-motion modeling in terms of mixed-effects regression.

The variations between earthquakes and between stations are random. Let $\boldsymbol{e} = (e_1, e_2, \dots, e_E)^T$ be the random effects of the earthquakes, and $\boldsymbol{s} = (s_1, s_2, \dots, s_S)^T$ the random effects of all seismic stations, where $S$ is the total number of stations. All random effects are assumed to have arisen from a normal distribution: $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma_E^2 \boldsymbol{D}_E)$ and $\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{0}, \sigma_S^2 \boldsymbol{D}_S)$. The random-effects covariance matrices $\sigma_E^2 \boldsymbol{D}_E$, $\sigma_S^2 \boldsymbol{D}_S$ are diagonal matrix with $\boldsymbol{D_E} = \boldsymbol{I}_E$ and $\boldsymbol{D_S} = \boldsymbol{I}_S$ ($\boldsymbol{I}_X$ is a unit matrix of size $X \times X$), that is, $\sigma_E^2$ is the variance between earthquakes (the variance associated with the between-event residuals), and $\sigma_S^2$ is the variance between stations (the station-to-station residuals). In line with these definitions, the covariance matrix $\sigma^2 \boldsymbol{C}$ - introduced in the beginning and equation (2) - is set to $\sigma^2 \boldsymbol{I}$. The variance $\sigma^2$ is therefore the variance of the data $\boldsymbol{y}$ conditioned on the random effects $\boldsymbol{e}$ and $\boldsymbol{s}$; the remaining variance in the data after accounting for earthquake and station variations.

The $K$ random effects $\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_K$ can be summarized into a vector:

$$\boldsymbol{q} = \begin{pmatrix} \boldsymbol{q}_1 \\ \boldsymbol{q}_2 \\ \vdots \\ \boldsymbol{q_K} \end{pmatrix}, \tag{4}$$

and $M = N_{q_1} + N_{q_2} + \cdots + N_{q_K}$ is the total number of random effects. The summarized covariance matrix is a block diagonal matrix:

$$\sigma^2\boldsymbol{D} = \sigma^2\begin{pmatrix} \tau_1\boldsymbol{I} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \tau_2\boldsymbol{I} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \tau_K\boldsymbol{I} \end{pmatrix}, \tag{5}$$

in which the $i$th block matrix is of size $N_{q_i} \times N_{q_i}$. The elements $\tau_1$, $\tau_2$, ..., $\tau_k$ are variance factors for the respective random effects; introduced for computational convenience. The $i$th random-effects variance is given by $\sigma_i^2 = \sigma^2\tau_i$. The random effects design matrix is

$$\boldsymbol{B} = (\boldsymbol{B}_1 \quad \boldsymbol{B}_2 \quad \cdots \quad \boldsymbol{B}_K) \tag{6}$$

in which the $i$th design matrix is of size $N \times N_{q_i}$. With covariance matrices defined and the incorporation of weights according to equation (3), the event-weighted likelihood of equation (2) is

$$\begin{aligned} &\mathcal{L}_w(\boldsymbol{f}(\boldsymbol{p}), \sigma^2, \boldsymbol{C}, \boldsymbol{D}|\boldsymbol{y}, \boldsymbol{q}) \\ &= \prod_{i=1}^{E}\left[\frac{1}{(2\pi|\sigma^2\boldsymbol{D}_i|)^{\frac{w_i}{2}}} e^{-\frac{1}{2\sigma^2}(\boldsymbol{V}_{q,i}\boldsymbol{q}_i)^T\boldsymbol{D}_i^{-1}\boldsymbol{V}_{q,i}\boldsymbol{q}_i}\right] \\ &\times \prod_{i=1}^{E}\left[\frac{1}{((2\pi)^{N_i}|\sigma^2\boldsymbol{C}_i|)^{\frac{w_i}{2}}}\right. \\ &\left.\times\, e^{-\frac{1}{2\sigma^2}(\boldsymbol{V}_i(\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{p}))-\boldsymbol{B}_i\boldsymbol{q}_i)^T\boldsymbol{C}_i^{-1}(\boldsymbol{V}_i(\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{p}))-\boldsymbol{B}_i\boldsymbol{q}_i)}\right], \end{aligned} \tag{7}$$

in which the event-based weights are expressed in matrix form $\boldsymbol{V}_i = \sqrt{w_i}\boldsymbol{I}_{N_i}$ for the residuals and $\boldsymbol{V}_i = \sqrt{w_i}$ for the random effects. See Appendix C for a detailed description of the weight matrices.

The ML estimator for $\boldsymbol{p}$ and ML predictor for $\boldsymbol{q}$ are derived from the derivatives of the log likelihood in equation (7) (detailed derivation in Appendix B). The estimators for the GMM parameters are given as the parameters of the fixed-effects model $\boldsymbol{f}(\boldsymbol{p})$. Because $\boldsymbol{f}(\boldsymbol{p})$ can be a nonlinear function of $\boldsymbol{p}$, the parameter estimators $\hat{\boldsymbol{p}}$ are found by nonlinear regression:

$$\Delta\boldsymbol{p} = \boldsymbol{X}\boldsymbol{r}, \tag{8}$$

With

$$\boldsymbol{X} = (\boldsymbol{J}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{J})^{-1}\boldsymbol{J}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}, \tag{9}$$

in which $\boldsymbol{J}$ is the Jacobian $\boldsymbol{J} = \frac{d\boldsymbol{f}(\boldsymbol{p})}{d\boldsymbol{p}}$ and $\boldsymbol{X}$ is the generalized least-squares estimator with covariance matrix $\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}$ (Aitken, 1936), and the best linear unbiased estimator (Henderson, 1975).

The predictor for the random effects, $\boldsymbol{q}$, is based on the fixed-effects model parameters:

$$\hat{\boldsymbol{q}} = \boldsymbol{V}_q^{-1}\boldsymbol{D}\boldsymbol{B}^T\boldsymbol{S}^{-1}\boldsymbol{V}(\boldsymbol{r} - \boldsymbol{J}\Delta\boldsymbol{p}) \tag{10}$$

which is the best linear unbiased predictor of $\boldsymbol{q}$ (Henderson, 1975).

The variance estimators of the mixed-effects model dependent on the estimator $\hat{\boldsymbol{p}}$ and predictor $\hat{\boldsymbol{q}}$ of the fixed-effects model. It is well established that ML variance estimates are biased downward by $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$, that is, variances are underestimated Patterson and Thompson (1971), Harville (1974, 1976), Lindstrom and Bates (1988), Bates and DebRoy (2004), and Bates et al. (2015). To obtain unbiased variance estimates, the likelihood is defined in terms of error contrasts, resulting in the reduced ML (RML). Details of the concept are described in full detail by Patterson and Thompson (1971) and Harville (1976).

The RML estimator of the residual variance (within-event variance) $\sigma^2$ is found from the derivative of equation (B25):

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{r}}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}}{\left(1 - \frac{P}{N}\right)tr(\boldsymbol{V}\boldsymbol{V})}. \tag{11}$$

By normalizing the weights such that it holds, that

$$tr(\boldsymbol{V}\boldsymbol{V}) = \sum_{i=1}^{N} V_{ii}^2 = N, \tag{12}$$

then the variance estimator can be stated as

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{r}}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}}{N - P}. \tag{13}$$

This is the unbiased estimator of the variance and appears frequently with the denominator $N - 1$, that is, for $P - 1$ (resulting from the sample mean estimation). The estimator corrects the bias for the loss of degree of freedoms, which is equal to the number of parameters $P$ in the GMM parameter estimator $\hat{\boldsymbol{p}}$. For generality, the variance formulation in equation (B29) is used in the following equations.

The RML estimators of the variance factors $\tau_k$ are linked to the variance $\hat{\sigma}^2$. There are no analytic solutions for the random-effects variances, and the likelihood must be maximized numerically, for example, by gradient ascent based on the log-likelihood derivatives of the variance factors:

$$\frac{\partial \ln \mathcal{L}_w}{\partial \tau_k} = \frac{1}{2}\left[-tr\left(\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{B}_k\boldsymbol{B}_k^T\boldsymbol{V}\right) + tr\left(\left(\boldsymbol{J}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{J}\right)^{-1}\boldsymbol{J}^T\boldsymbol{S}^{-1}\boldsymbol{B}_k\boldsymbol{B}_k^T\boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{J}\right)\frac{tr(\boldsymbol{V}\boldsymbol{V})}{N} + \frac{1}{\hat{\sigma}^2}\hat{\boldsymbol{r}}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{B}_k\boldsymbol{B}_k^T\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}\right]. \quad (14)$$

The first and last term correspond to the basic ML estimator. The second term is the effect of the RML estimation, that is, this term corrects for the bias of the ML estimator, if GMM parameters and random-effects variances are derived from the data.

The variance factor estimates are found by iteration of

$$\boldsymbol{\tau}^{(i+1)} = \boldsymbol{\tau}^{(i)} + \gamma \nabla \mathcal{L}_w, \quad (15)$$

in which $\boldsymbol{\tau}$ is the vector of variance factors, that is, with equation (5) $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_K)^T$. The factor $\gamma$ is chosen such to guarantee convergence and is updated at each iteration (Barzilai and Borwein, 1988). The algorithm starts from some initial values and is repeated until sufficient convergence of the parameters is reached. At each iteration, $\hat{\sigma}^2$ and $\hat{\boldsymbol{p}}$ are updated before the factors $\boldsymbol{\tau}$ are updated. Finally, the ML estimates of the variances are as follows:

$$\hat{\boldsymbol{v}} = \hat{\sigma}^2\hat{\boldsymbol{\tau}}. \quad (16)$$

***Estimator uncertainties of GMM parameters and random-effects variances***

The mixed-effects regression assumes that any kind of residuals are normally distributed. This property allows for the derivation of analytic expressions for parameter uncertainty estimates, that is, computationally intensive methods such as bootstrapping and jackknifing are not necessary. The (co)variance of a parameter is related to the Fisher information $\boldsymbol{I}$ of the parameters, expressed by the Cramér–Rao bound:

$$\mathrm{cov}\left(\hat{\boldsymbol{\theta}}\right) > \boldsymbol{I}\left(\hat{\boldsymbol{\theta}}\right)^{-1}. \quad (17)$$

If the previous relation becomes an equality, the estimator $\hat{\boldsymbol{\theta}}$ is said to be an efficient estimator. Hartley and Rao (1967) showed for the mixed-effects model that the parameter estimates $\hat{\boldsymbol{p}}$ are efficient and the variance estimators $\hat{\boldsymbol{\tau}}$ are asymptotically efficient, that is, converge toward efficiency with increasing number of data. This (asymptotic) equality of the Cramér–Rao bound allows the derivation of parameter uncertainties from the Fisher information.

The Fisher information $\boldsymbol{I}$ of a parameter estimator $\hat{\boldsymbol{\theta}}$ is a matrix related to the likelihood $\mathcal{L}$. Its components are given by

$$I_{ij} = \mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}}{\partial\hat{\theta}_i \partial\hat{\theta}_j}\right], \quad (18)$$

in which $\mathbb{E}[\cdot]$ is the expected value operator. For the mixed-effects model, the Fisher information is given as $3 \times 3$ block matrices:

$$\boldsymbol{I}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}}) = -\mathbb{E}\begin{bmatrix} \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{p}}\partial\hat{\boldsymbol{p}}} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{p}}\partial\hat{\sigma}^2} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{p}}\partial\hat{\boldsymbol{v}}} \\ \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\sigma}^2\partial\hat{\boldsymbol{p}}} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\sigma}^2\partial\hat{\sigma}^2} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\sigma}^2\partial\hat{\boldsymbol{v}}} \\ \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{v}}\partial\hat{\boldsymbol{p}}} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{v}}\partial\hat{\sigma}^2} & \frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{v}}\partial\hat{\boldsymbol{v}}} \end{bmatrix} \quad (19)$$

The inverse $\boldsymbol{I}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}})^{-1}$ - which is asymptotically equivalent to the covariance matrix as $N \to \infty$ - is stated as

$$\boldsymbol{I}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}})^{-1} \sim \mathrm{cov}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}}) = \begin{pmatrix} \hat{\sigma}^2(\boldsymbol{J}^T\boldsymbol{P}\boldsymbol{J})^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{\widehat{2\sigma}^4}{tr(\boldsymbol{V}\boldsymbol{V})} + \boldsymbol{T}\boldsymbol{M}\boldsymbol{T}^T & \boldsymbol{T}\boldsymbol{M} \\ \boldsymbol{0} & \boldsymbol{M}\boldsymbol{T}^T & \boldsymbol{M} \end{pmatrix}, \quad (20)$$

in which $\boldsymbol{P}$, $\boldsymbol{T}$, and $\boldsymbol{M}$ are matrices related to the random-effects variances and the first and second derivatives of the mixed-effects likelihood (see Appendix D for the definitions and a detailed derivation of the covariance matrix). The terms related to the covariance for a single random effect without weighting reduces to the terms used by Abrahamson and Youngs (1992; equations 11–13). The covariance matrix shows that the covariance of the model parameters (upper left block) is independent of the covariance of the random-effects variances (four lower right blocks), that is, the model parameters are independent of the variances. Furthermore, equation (20) shows that the variances of the random-effects variances are not necessarily independent from each other, as there are nonzero off-diagonal terms relating to the variances.

The correlations among the parameters (the off-diagonal elements) can be expressed by Pearson's correlation coefficient. The correlation between parameters $\theta_i$ and $\theta_j$ is given by

$$\rho_{\theta_i,\theta_j} = \frac{I_{ij}^{-1}}{\sqrt{I_{ii}^{-1}I_{jj}^{-1}}}, \quad (21)$$

in which $I_{ij}^{-1}$ is the element in the $i$th row and $j$th column of the inverse of the Fisher information matrix. The Fisher information matrix provides therefore a powerful tool to

assess the linear dependence of parameters and thus helps with model design and evaluation and allows to check for the linear independence of the variance components.

## DATA

We use the strong-motion flat file for Chile published by Bastías and Montalva (2016) for the exemplary application of weighted ground-motion modeling. Seismicity in South America is dominated by two linked processes: (1) thrust events at the interface between the South American plate and Nazca plate and (2) normal faulting in the subducting slab of the Nazca plate. The events associated with either process are clearly separated spatially and by their FM and provide, therefore, a suitable data set for the investigation of different data classification methodologies.

The flat file of Bastías and Montalva (2016) contains in total 477 events range from $M_w$ 4.6 to $M_w$ 8.8 with 3572 three-component records, with distances measured with several metrics (rupture plane distance, $R_{rup}$, and hypocentral distance, $R_{hyp}$). Peak ground acceleration (PGA) and pseudospectral accelerations (PSAs) with 5% damping are available for each accelerometer component for oscillator periods between 0.01 and 10 s. Bastías and Montalva (2016) classify events in the flat file as interface, intraslab, and crustal. Interface events are defined by a depth of less than 50 km and a maximum distance of 2.5° from the trench (equivalent to 260 km at 20°S and 210 km at 40°S) and with a reverse SoF. Intraslab events are all events deeper than 50 km and events shallower than 50 km with normal faulting within 2.5° distance to the trench. Crustal events are shallower than 50 km and have a distance more than 2.5° from the trench.

Because ACE requires FM for classification, we use a subset of the flat file events for which also Global Centroid Moment Tensor (CMT) catalog solutions exist (Ekström et al., 2012). This subset contains 319 events with 2443 records with a magnitude range of $M_w$ 4.9–8.8 (Fig. 1).

## DATA-DRIVEN EVENT CLASSIFICATION

The weighted mixed-effects regression introduced in this article allows for a more general treatment of data selection. Besides the classical approach of deterministic selection, more sophisticated data selection is feasible with the weighted mixed-effects regression. We use ACE to classify earthquakes on the basis of their FMs (Specht et al., 2017; von Specht et al., 2018). ACE is based on expectation–maximization and fits a mixture distribution to the FM data in the strike–rake–dip domain. Because ACE considers both nodal planes, a separation into rupture and auxiliary plane is not required. The number of mixture components is optimized as part of the parameter estimation.

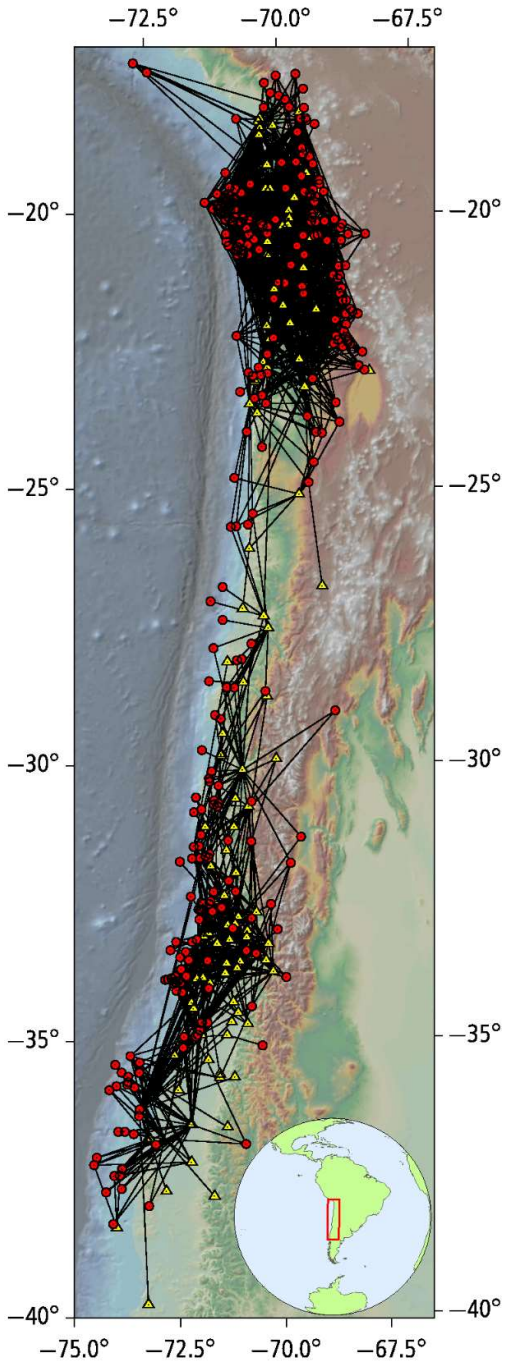


***Figure 1.*** *Map from Chile and adjacent regions showing events used in this study from the flat file of Bastías and Montalva (2016). Our catalog is a subset of the flat file with reported focal mechanisms (FMs) and contains 319 events (circles) recorded at 176 stations (triangles) with a total of 2443 records (lines). The map projection is Universal Transverse Mercator (UTM) zone 19S. The inset shows the study area in South America as green box.*

The underlying principle of ACE is the assumption that in a homogeneous background stress field only random variations in the stress field exist locally and FMs of a given SoF will tend to be similar to each other. Kagan (1991) showed that the minimal rotation angle between

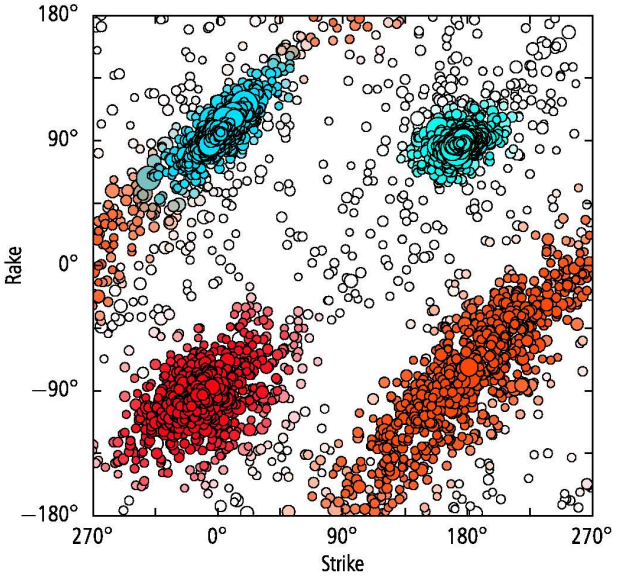


***Figure 2.*** *Distribution of nodal planes of FM for Chile from the Global Centroid Moment Tensor catalog between 1976 and 2019. FMs are represented in the strike–rake plane (lower right). The shape of the clusters depends on the dip, the shallower the dip, the more elongated the clusters. The model based on Angular Classification with Expectation–maximization (ACE) identifies four clusters: two for interface (in the upper half, with positive rakes) and intraslab (in the lower half, with negative rakes), respectively. The remaining scattered data are unclassified. Those events have mainly hypocenters in the South American crust. The color saturation corresponds to the probability of a nodal plane to be in a certain cluster. For each event, the probabilities of both nodal planes are averaged and used as weights for the event classification used in ground-motion modeling.*

two FM in a homogeneous stress field follows a double couple rotational Cauchy distribution. FMs from regional catalogs tend to have small rotation angles, that is, the FMs of a particular stress field form clusters in terms of their parameters (strike, dip, and rake).

For the Chile region, ACE identifies three event types in five clusters: two clusters for either nodal planes of interface and intraslab events, respectively, and one cluster for unclassified events (Fig. 2). The probability densities of the mixture distribution are used as weights in the development of the GMM and their square roots populate the matrix $\boldsymbol{V}$.

Event classification of the Chile database based on expert judgment and ACE are shown in Figure 3a,c. Both event classifications are near identical with few differences, and only six out of 319 events have opposite SoF assigned (Fig. 3b).

## GROUND-MOTION MODEL

We apply the GMM of Montalva et al. (2017) to model the strong-motion data of Bastías and Montalva (2016). The model is related to the BC Hydro model by Abrahamson et al. (2016) with few modifications. The model consists of several terms related to the source, path, event type, event depth, and site, and is defined for interface and intraslab earthquakes:

$$y = \theta_1 + f_{source} + f_{path} + f_{event/depth} + f_{site}, \tag{22}$$

$$f_{source} = \theta_4 \Delta C_1 + f_{mag}, \tag{23}$$

$$f_{mag} = \begin{cases} \theta_4\left(M_w - (C_1 + \Delta C_1)\right) & \text{if } M_w \le C_1 + \Delta C_1 \\ \theta_5\left(M_w - (C_1 + \Delta C_1)\right) & \text{if } M_w > C_1 + \Delta C_1 \end{cases}, \tag{24}$$

$$f_{path} = \left(\theta_2 + \theta_{14} f_{event} + \theta_3 (M_w - 7.2)\right) \ln\left(R + C_4 e^{\theta_9 (M_w - 6)}\right) + \theta_6 R, \tag{25}$$

$$f_{event/depth} = \left(\theta_{10} + \theta_{11}(\min(Z_h, 120) - 60)\right) f_{event}, \tag{26}$$

$$f_{site} = \theta_{12} \ln \frac{V_S^*}{V_{lin}} + \begin{cases} b \ln \dfrac{PGA_{1000} + c\left(\frac{V_S^*}{V_{lin}}\right)}{PGA_{1000} + c} & \text{if } V_{S30} < V \\ b \ln \dfrac{V_S^*}{V_{lin}} & \text{if } V_{S30} \ge V_{lin} \end{cases}, \tag{27}$$

$$V_S^* = \min(V_{S30}, 1000), \tag{28}$$

in which $y$ is either $\frac{1}{2}\ln(PGA_{EW} PGA_{NS})$ or $\frac{1}{2}\ln\left(PSA_{EW}(T) PSA_{NS}(T)\right)$ with $T$ as the oscillator period. PGA and PSA are given in units of $g = 9.80665 \text{m/s}^2$. The parameters $\theta$ are fitted to PGA and PSA depending on:

1. moment magnitude ($M_w$) as reported in the Global CMT catalog (Ekström et al., 2012),
2. hypocentral depth in kilometers ($Z_h$),
3. source-to-site distance in km ($R$), defined as:
   a) closest distance to the rupture plane ($R_{rup}$) for interface earthquakes
   b) hypocentral distance ($R_{hyp}$) for intraslab earthquakes
4. PGA1000 is the median PGA value for a site with $V_{S30} = 1000$ m/s,
5. event type ($f_{event}$); for interface $f_{event} = 0$, for intraslab $f_{event} = 1$,
6. modified $V_{S30}$ with a maximum value of 1000 m/s ($V_S^*$).

The parameters $\Delta C_1$, $C_1 = 7.2$, and $C_4 = 10$ are taken from the BC Hydro model. The site-related parameters $V_{lin}$, $b$, $c$, and $n$ are computed directly from the nonlinear site-response function of the Peninsular model of Walling et al. (2008) (as has been done in the BC Hydro model).

The BC Hydro model also includes a fore-arc (back-arc) term, but Montalva et al. (2017) noted that no back-arc events are present in the flat file. We, therefore, drop that term from the BC Hydro model and refer to Abrahamson et al. (2016) and Montalva et al. (2017) if a fore-arc

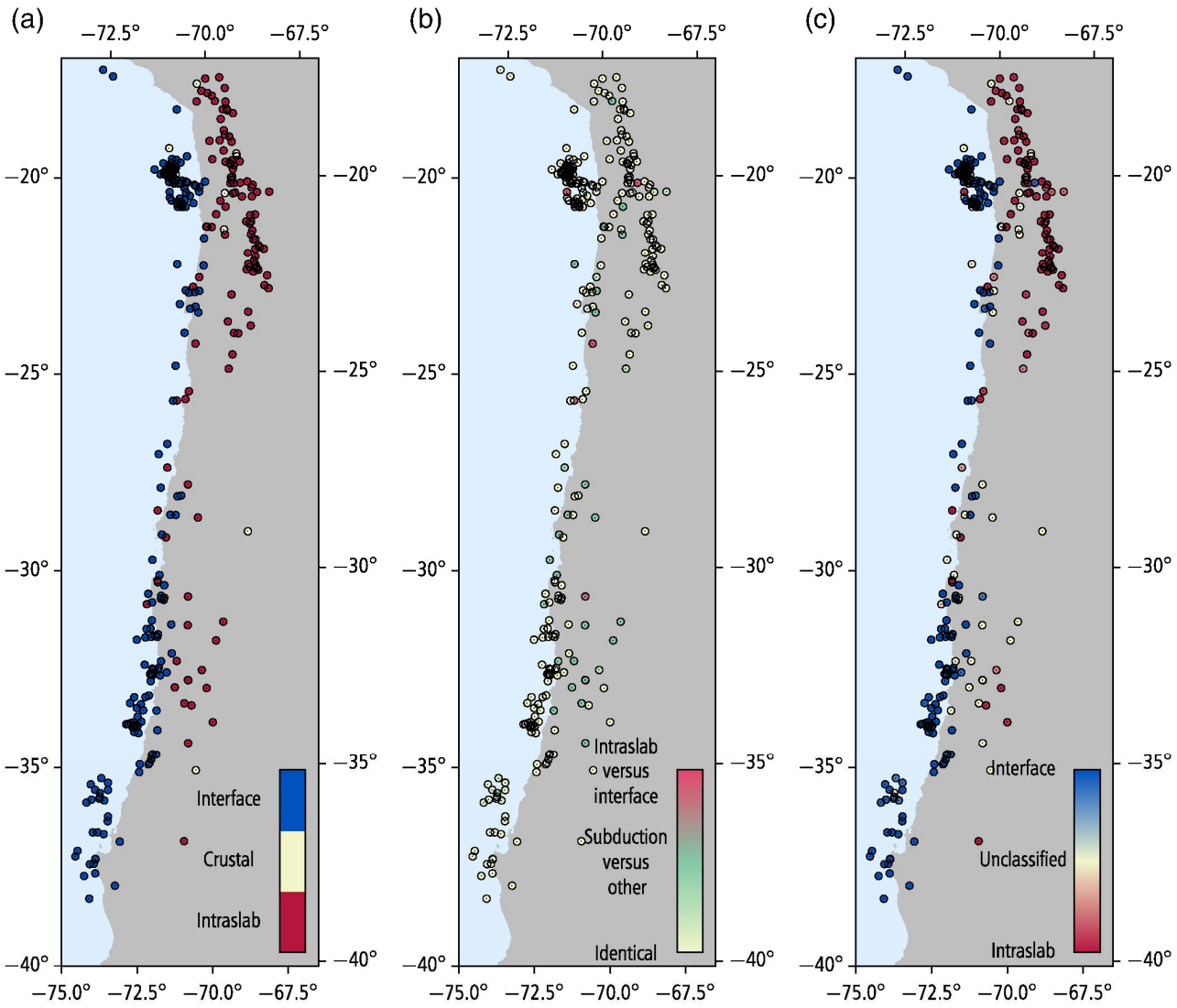


***Figure 3.*** *(a) Style-of-fault (SoF) classification of events according to the flat file (deterministic expert judgment). (b) Difference between expert-based and ACE classifications ranging from identical classification, through subduction interface or intraslab) against other (crustal or unclassified), to interface against intraslab. (c) Event classification based on ACE. SoF classification is continuous, that is, no event belongs completely to one class. This is implemented by assigning weights to each event and shown as color saturation. Automated SoF assignments are very similar to the expert-based classification. The map projection is UTM zone 19S.*

(back-arc) term is required. Consistent with Montalva et al. (2017), we do not include the quadratic magnitude of the original BC Hydro model.

The choice of event type is straightforward in the expert classification by setting $F_{event} = 0$ for interface (IF) events, $F_{event} = 1$ for intraslab (IS) events, as provided in the flat file. For the ACE-based classification, the $F_{eve}$ term is defined as the probability of an event being interface ($P(IF)$) or intraslab ($P(IS)$) conditioned on the probability of an event being related to subduction ($P(S) = P(IF) + P(IS)$, i.e., either interface or intraslab):

$$F_{event} = P(IS|S) = \frac{P(IS)}{P(IF) + P(IS)}, \tag{29}$$

$$1 - F_{event} = P(IF|S) = \frac{P(IF)}{P(IF) + P(IS)}. \tag{30}$$

In case of $\Delta C_1$, which takes different values for interface and intraslab events, we define the weighted average:

$$\Delta C_1^{ACE} = F_{event}\Delta C_1^{intraslab} + (1 - F_{event})\Delta C_1^{interface}. \tag{31}$$

As in Montalva et al. (2017), we augment the model with two random effects for between-event residuals, with standard deviation $\tau$, and between-station residuals,

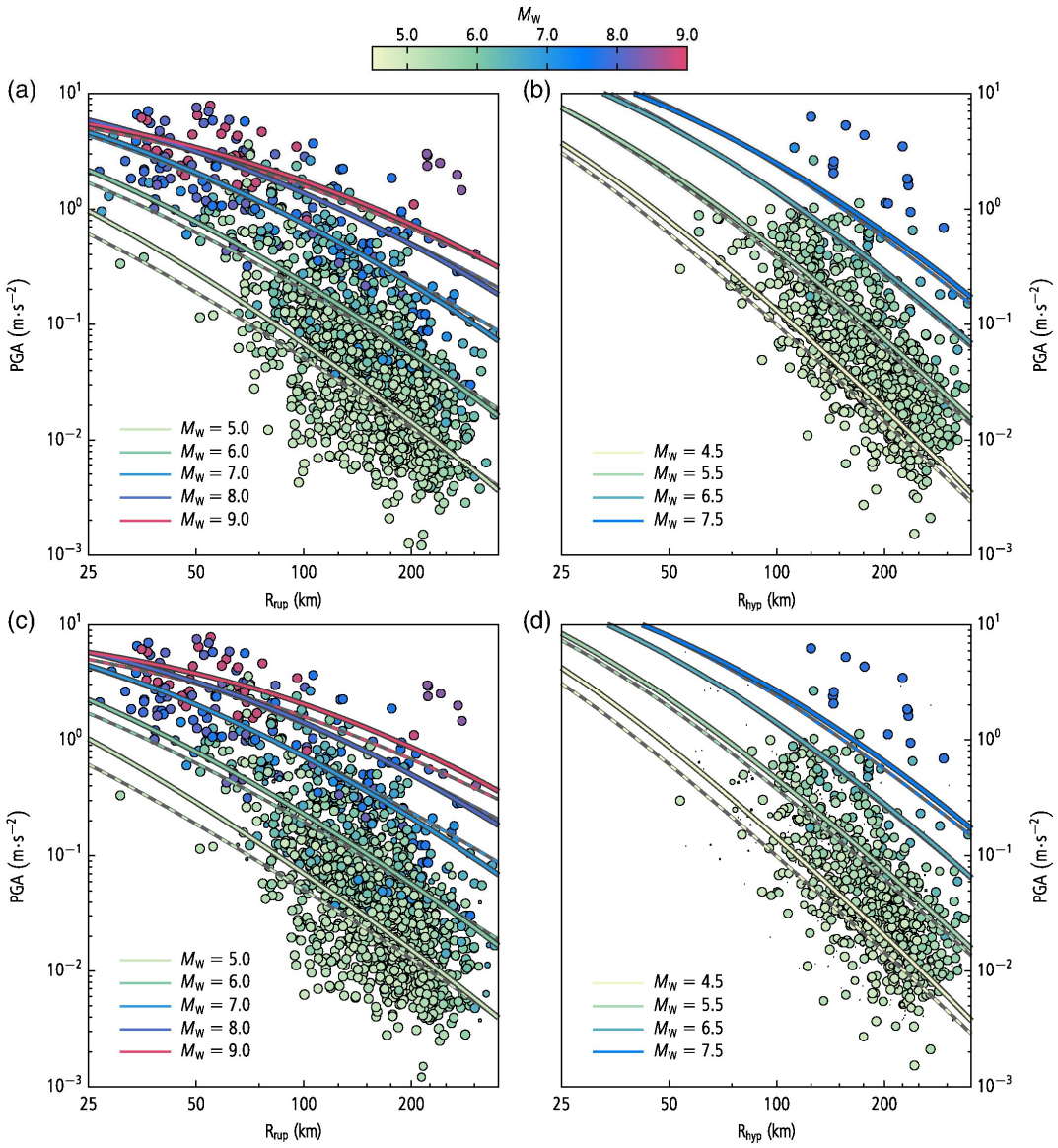


***Figure 4.*** *(Top row) Ground-motion model (GMM) for Chile obtained for expert-based data classification for (a) interface and (b) intraslab. Only interface and intraslab events are used. (Bottom row) GMM for Chile obtained for ACE-based data classification for (c) interface and (d) intraslab. All events are included with weights proportional to the color saturation in the map of Figure 3 (middle). The size of the events corresponds to their weights (only visible in bottom row). The thin lines show the model of Montalva et al. (2017) for comparison, which is based on the entire data set in the flat file of Bastías and Montalva (2016), while we used only a subset with available FMs. All models are shown for $V_{S30}$ = 500 m/s and a hypocentral depth of 25 km for interface events and 50 km for intraslab events. PGA, peak ground acceleration.*

with standard deviation $\phi_{S2S}$. The residual standard deviation is $\phi$.

Montalva et al. (2017) treated $\theta_9$, one of the parameters related to the magnitude-scaled distance term, as a fixed value taken from Abrahamson et al. (2016). We treat $\theta_9$ as free parameter in the parameter estimation for PGA and similarly to Abrahamson et al. (2016) constrain it in the parameter estimations for the PSAs. However, instead of fixing $\theta_9$ for all spectral periods to the $\theta_9$ of PGA, we use $\theta_9$ of PGA as a priori information with variance estimated from the Fisher information matrix. Thus, $\theta_9$ will diverge into unreasonable ranges as it would to if left unconstrained, as we have seen from trials. Other than the treatment of $\theta_9$, we follow the procedure of Montalva et al. (2017) and invert all $\theta$ parameters without any smoothing or conditioning as has been done by Abrahamson et al. (2016).

For outlier detection, we use the extreme studentized deviate test for multiple outliers Rosner (1975, 1983). At the 5% significance level, we identified a single event with outlying acceleration values and remove the event completely from our catalog and rerun all parameter inversions.

**TABLE 2**

**Ground-Motion Model (GMM) Coefficients with Error Standard Deviation in Parentheses**

| $T$ (s) | $\theta_1$ ($\times 10^0$) | $\theta_4$ ($\times 10^0$) | $\theta_5$ ($\times 10^{-1}$) | $\theta_2$ ($\times 10^0$) | $\theta_3$ ($\times 10^{-1}$) | $\theta_{14}$ ($\times 10^{-1}$) |
|---|---|---|---|---|---|---|
| PGA | 5.8674 (1.2740) | 0.4950 (0.5866) | −2.2046 (8.0350) | −1.7698 (0.2649) | 1.6360 (1.0567) | −5.2715 (1.0541) |
| 0.010 | 5.8161 (1.2150) | 0.4810 (0.5557) | −1.8416 (7.8436) | −1.7632 (0.2547) | 1.6441 (1.0069) | −5.2814 (1.0498) |
| 0.015 | 5.8208 (1.2210) | 0.4737 (0.5588) | −1.8212 (7.8682) | −1.7682 (0.2559) | 1.6514 (1.0122) | −5.3293 (1.0538) |
| 0.020 | 5.9893 (1.2235) | 0.4923 (0.5596) | −1.5992 (7.8718) | −1.8018 (0.2565) | 1.6094 (1.0136) | −5.3288 (1.0584) |
| 0.025 | 6.4368 (1.2218) | 0.5940 (0.5574) | −0.4667 (7.8105) | −1.9225 (0.2564) | 1.4171 (1.0099) | −5.4280 (1.0648) |
| 0.030 | 6.8574 (1.2114) | 0.5688 (0.5506) | −0.8310 (7.8052) | −1.9879 (0.2546) | 1.4685 (0.9994) | −5.1905 (1.0634) |
| 0.035 | 7.3422 (1.2103) | 0.5580 (0.5491) | −0.8042 (7.8025) | −2.0946 (0.2544) | 1.4878 (0.9971) | −5.0808 (1.0650) |
| 0.040 | 7.8314 (1.2122) | 0.5415 (0.5498) | −0.7798 (7.8009) | −2.2023 (0.2548) | 1.5087 (0.9978) | −4.9494 (1.0664) |
| 0.045 | 8.2305 (1.2114) | 0.5553 (0.5497) | −0.7216 (7.7680) | −2.2879 (0.2546) | 1.4775 (0.9965) | −4.8228 (1.0628) |
| 0.050 | 8.4476 (1.1988) | 0.6339 (0.5414) | 0.0320 (7.7317) | −2.3203 (0.2523) | 1.3314 (0.9830) | −4.3758 (1.0604) |
| 0.055 | 8.5743 (1.2000) | 0.6415 (0.5421) | 0.0419 (7.7253) | −2.3337 (0.2525) | 1.3160 (0.9835) | −4.0219 (1.0590) |
| 0.060 | 8.8950 (1.1919) | 0.6735 (0.5365) | 0.2996 (7.6921) | −2.3886 (0.2510) | 1.2494 (0.9742) | −3.3732 (1.0569) |
| 0.065 | 8.9273 (1.1806) | 0.7279 (0.5298) | 0.9664 (7.6721) | −2.3797 (0.2488) | 1.1427 (0.9633) | −3.4130 (1.0525) |
| 0.070 | 8.7764 (1.1759) | 0.7512 (0.5265) | 1.4416 (7.6841) | −2.3264 (0.2480) | 1.0912 (0.9584) | −3.3608 (1.0536) |
| 0.075 | 8.5963 (1.1699) | 0.7284 (0.5243) | 1.3214 (7.6411) | −2.2752 (0.2466) | 1.1269 (0.9537) | −3.1250 (1.0435) |
| 0.080 | 8.6032 (1.1771) | 0.6716 (0.5281) | 0.9820 (7.6637) | −2.2663 (0.2480) | 1.2265 (0.9601) | −2.7275 (1.0466) |
| 0.085 | 8.2809 (1.1921) | 0.5969 (0.5378) | 0.1143 (7.6956) | −2.1835 (0.2507) | 1.3628 (0.9752) | −2.5891 (1.0472) |
| 0.090 | 8.1957 (1.1902) | 0.6219 (0.5366) | 0.2989 (7.6768) | −2.1520 (0.2504) | 1.3079 (0.9730) | −2.3130 (1.0465) |
| 0.100 | 8.3754 (1.1441) | 0.8456 (0.5078) | 2.9281 (7.5367) | −2.1583 (0.2419) | 0.8785 (0.9278) | −1.9401 (1.0405) |
| 0.150 | 7.2582 (1.1650) | 0.8739 (0.5166) | 3.2938 (7.6380) | −1.8534 (0.2465) | 0.7890 (0.9453) | −1.3569 (1.0661) |
| 0.200 | 5.7616 (1.2188) | 0.8596 (0.5478) | 2.9418 (7.8989) | −1.5154 (0.2571) | 0.8593 (0.9998) | −3.4052 (1.0972) |
| 0.250 | 5.6332 (1.2254) | 0.8377 (0.5506) | 2.8095 (8.0245) | −1.4842 (0.2586) | 0.9351 (1.0065) | −3.7360 (1.1092) |
| 0.300 | 4.9557 (1.2771) | 0.7201 (0.5817) | 0.6172 (8.2869) | −1.3344 (0.2682) | 1.2317 (1.0584) | −4.4337 (1.1232) |
| 0.350 | 4.7701 (1.2898) | 0.8106 (0.5888) | 1.0139 (8.3252) | −1.2999 (0.2710) | 1.1308 (1.0727) | −5.6683 (1.1377) |
| 0.400 | 4.4297 (1.3066) | 0.7826 (0.6007) | 0.3432 (8.3403) | −1.2346 (0.2739) | 1.2519 (1.0922) | −6.0518 (1.1380) |
| 0.450 | 5.0066 (1.2652) | 0.9050 (0.5739) | 2.0495 (8.1713) | −1.3903 (0.2669) | 1.0699 (1.0523) | −6.4750 (1.1487) |
| 0.500 | 5.0001 (1.3061) | 0.8548 (0.5984) | 0.7301 (8.2260) | −1.4041 (0.2745) | 1.2468 (1.0912) | −6.9011 (1.1596) |
| 0.550 | 5.0753 (1.2612) | 1.0224 (0.5739) | 2.3350 (8.0092) | −1.4361 (0.2660) | 1.0068 (1.0512) | −7.5842 (1.1459) |
| 0.600 | 5.0252 (1.2637) | 1.0755 (0.5774) | 1.9901 (7.9155) | −1.4443 (0.2662) | 0.9743 (1.0550) | −8.4424 (1.1413) |
| 0.650 | 4.5525 (1.2735) | 1.1247 (0.5843) | 1.9026 (7.9155) | −1.3516 (0.2678) | 0.9388 (1.0652) | −8.7213 (1.1431) |
| 0.700 | 4.3106 (1.2770) | 1.1714 (0.5867) | 2.4084 (7.9005) | −1.3163 (0.2685) | 0.8949 (1.0689) | −9.0489 (1.1468) |
| 0.750 | 3.8901 (1.2747) | 1.2216 (0.5867) | 2.9205 (7.8362) | −1.2353 (0.2679) | 0.8195 (1.0674) | −8.8897 (1.1417) |
| 0.800 | 3.6938 (1.2725) | 1.2229 (0.5871) | 2.9233 (7.7854) | −1.2143 (0.2671) | 0.8456 (1.0667) | −8.9814 (1.1346) |
| 0.850 | 3.5633 (1.2700) | 1.2362 (0.5878) | 2.7339 (7.7263) | −1.2048 (0.2663) | 0.8681 (1.0660) | −9.2110 (1.1275) |
| 0.900 | 3.2986 (1.2659) | 1.2481 (0.5876) | 2.7478 (7.6786) | −1.1619 (0.2652) | 0.8841 (1.0641) | −9.5303 (1.1208) |
| 1.000 | 3.3155 (1.2449) | 1.3389 (0.5757) | 3.2240 (7.5239) | −1.1932 (0.2612) | 0.7967 (1.0440) | −9.4543 (1.1089) |
| 1.500 | 2.6015 (1.2033) | 1.3838 (0.5605) | 1.4172 (7.1554) | −1.1624 (0.2517) | 1.0609 (1.0129) | −9.2607 (1.0541) |
| 2.000 | 2.4436 (1.1320) | 1.4227 (0.5219) | 1.9606 (6.8129) | −1.2189 (0.2375) | 1.1946 (0.9501) | −7.9194 (1.0004) |
| 2.500 | 2.2367 (1.1250) | 1.2530 (0.5243) | 0.3137 (6.7084) | −1.2370 (0.2349) | 1.6664 (0.9502) | −6.8786 (0.9542) |
| 3.000 | 2.5146 (1.0952) | 1.2553 (0.5122) | −0.9866 (6.4283) | −1.3512 (0.2283) | 1.8296 (0.9258) | −7.0437 (0.9159) |
| 3.500 | 2.6802 (1.1046) | 1.1771 (0.5202) | −2.1898 (6.4795) | −1.4582 (0.2300) | 2.0883 (0.9378) | −7.4402 (0.9112) |
| 4.000 | 2.4227 (1.1596) | 1.0033 (0.5465) | −4.4772 (6.7164) | −1.4491 (0.2421) | 2.4946 (0.9812) | −7.2274 (0.9282) |
| 4.500 | 2.3244 (1.1932) | 0.8978 (0.5654) | −5.2766 (6.8931) | −1.4829 (0.2485) | 2.7164 (1.0129) | −7.2125 (0.9449) |
| 5.000 | 2.2971 (1.2098) | 0.8539 (0.5733) | −5.2659 (6.9956) | −1.5287 (0.2520) | 2.8025 (1.0274) | −7.4060 (0.9637) |
| 5.500 | 2.2951 (1.2154) | 0.9456 (0.5746) | −4.0559 (7.0018) | −1.5647 (0.2536) | 2.6083 (1.0311) | −7.5188 (0.9809) |
| 6.000 | 2.5763 (1.2296) | 0.8563 (0.5817) | −4.8554 (7.0255) | −1.6801 (0.2562) | 2.8225 (1.0419) | −7.7617 (0.9857) |
| 6.500 | 2.7858 (1.2401) | 0.9014 (0.5860) | −4.5313 (7.0705) | −1.7585 (0.2588) | 2.7583 (1.0509) | −7.8432 (0.9998) |
| 7.000 | 2.7613 (1.2371) | 1.0118 (0.5857) | −3.3955 (7.0366) | −1.7762 (0.2582) | 2.5457 (1.0498) | −7.9832 (1.0030) |
| 7.500 | 2.5966 (1.2374) | 1.0584 (0.5859) | −2.5575 (7.0372) | −1.7637 (0.2584) | 2.4468 (1.0501) | −8.3864 (1.0059) |
| 8.000 | 2.6657 (1.2552) | 1.1398 (0.5840) | −1.2996 (7.0720) | −1.8105 (0.2634) | 2.3318 (1.0489) | −8.8907 (1.0690) |
| 8.500 | 2.4218 (1.2730) | 1.0996 (0.5963) | −0.9746 (7.2275) | −1.7836 (0.2669) | 2.3663 (1.0711) | −9.1883 (1.0806) |
| 9.000 | 2.1284 (1.2772) | 1.0666 (0.5993) | −0.6946 (7.2672) | −1.7389 (0.2677) | 2.4033 (1.0761) | −9.1426 (1.0809) |
| 10.000 | 1.8182 (1.2780) | 1.0765 (0.5995) | 0.2892 (7.2782) | −1.7097 (0.2681) | 2.3489 (1.0774) | −9.4484 (1.0982) |

(continued)

**TABLE 2 (Continued)**

| $T$ (s) | $\theta_6$ ($\times 10^{-3}$) | $\theta_9$ ($\times 10^{-1}$) | $\theta_{10}$ ($\times 10^{0}$) | $\theta_{11}$ ($\times 10^{-3}$) | $\theta_{12}$ ($\times 10^{0}$) |
|---|---|---|---|---|---|
| PGA | −1.2238 (0.9858) | 4.9767 (3.6185) | 3.5868 (0.5572) | 4.3855 (3.3146) | 0.7695 (0.0829) |
| 0.010 | −1.2464 (0.9651) | 4.9188 (3.3413) | 3.5856 (0.5540) | 4.5251 (3.2998) | 0.7651 (0.0865) |
| 0.015 | −1.1860 (0.9693) | 4.8861 (3.3648) | 3.5929 (0.5559) | 4.4666 (3.2944) | 0.7369 (0.0868) |
| 0.020 | −1.0913 (0.9731) | 4.8268 (3.3208) | 3.6035 (0.5584) | 4.6072 (3.3099) | 0.7418 (0.0871) |
| 0.025 | −0.6731 (0.9782) | 4.5928 (3.1567) | 3.6771 (0.5617) | 4.8661 (3.3257) | 0.9283 (0.0877) |
| 0.030 | −0.3573 (0.9737) | 4.7907 (2.9395) | 3.6099 (0.5617) | 4.7691 (3.3611) | 0.8402 (0.0873) |
| 0.035 | 0.0864 (0.9747) | 4.7752 (2.7825) | 3.5984 (0.5632) | 5.1243 (3.4068) | 0.8999 (0.0874) |
| 0.040 | 0.5188 (0.9769) | 4.6889 (2.6747) | 3.5955 (0.5647) | 5.3125 (3.4569) | 0.9467 (0.0875) |
| 0.045 | 0.8522 (0.9762) | 4.5188 (2.6260) | 3.5665 (0.5640) | 5.8549 (3.5112) | 0.9739 (0.0875) |
| 0.050 | 0.7933 (0.9724) | 4.5993 (2.5116) | 3.3830 (0.5639) | 6.0284 (3.5677) | 1.0034 (0.0879) |
| 0.055 | 0.6624 (0.9729) | 4.5020 (2.5321) | 3.2453 (0.5641) | 5.8933 (3.6170) | 1.0848 (0.0879) |
| 0.060 | 0.6999 (0.9705) | 4.5417 (2.4311) | 2.9286 (0.5639) | 5.9331 (3.6485) | 1.1272 (0.0884) |
| 0.065 | 0.4999 (0.9652) | 4.6465 (2.3693) | 2.9548 (0.5626) | 5.9506 (3.6963) | 1.1691 (0.0883) |
| 0.070 | 0.0861 (0.9643) | 4.7961 (2.3593) | 2.8994 (0.5633) | 6.5910 (3.7025) | 1.2144 (0.0885) |
| 0.075 | −0.1803 (0.9574) | 4.7581 (2.4216) | 2.7869 (0.5586) | 6.2094 (3.7068) | 1.2324 (0.0882) |
| 0.080 | −0.2574 (0.9613) | 4.7443 (2.4634) | 2.5900 (0.5598) | 5.9205 (3.6829) | 1.2716 (0.0886) |
| 0.085 | −0.6946 (0.9662) | 4.5748 (2.6807) | 2.5474 (0.5596) | 5.2657 (3.6558) | 1.2990 (0.0888) |
| 0.090 | −0.9089 (0.9658) | 4.5535 (2.7193) | 2.4230 (0.5593) | 4.4761 (3.6521) | 1.3255 (0.0891) |
| 0.100 | −1.2061 (0.9490) | 5.1153 (2.3475) | 2.1577 (0.5558) | 5.0288 (3.6078) | 1.4195 (0.0889) |
| 0.150 | −2.5819 (0.9688) | 5.2204 (2.7466) | 1.6818 (0.5662) | 3.4786 (3.4937) | 1.6790 (0.0909) |
| 0.200 | −3.3174 (0.9981) | 5.0754 (3.6576) | 2.5305 (0.5773) | 3.2097 (3.2837) | 1.8329 (0.0924) |
| 0.250 | −3.0028 (1.0033) | 5.3831 (3.6179) | 2.5617 (0.5831) | 2.8218 (3.2998) | 2.0038 (0.0917) |
| 0.300 | −2.9509 (1.0228) | 5.3015 (4.3939) | 2.8167 (0.5879) | 1.7432 (3.1887) | 2.0825 (0.0922) |
| 0.350 | −2.4565 (1.0344) | 5.3074 (4.5422) | 3.2966 (0.5936) | 2.0886 (3.1169) | 2.0823 (0.0936) |
| 0.400 | −2.3093 (1.0368) | 5.2575 (4.9625) | 3.4168 (0.5923) | 1.9612 (3.0302) | 2.0806 (0.0923) |
| 0.450 | −1.1773 (1.0296) | 5.7167 (3.8948) | 3.5859 (0.5971) | 1.8007 (2.9802) | 2.1086 (0.0916) |
| 0.500 | −0.7313 (1.0464) | 5.2249 (4.3093) | 3.7784 (0.6019) | 1.9144 (2.9651) | 2.1462 (0.0904) |
| 0.550 | −0.3505 (1.0248) | 5.4198 (3.8778) | 4.1098 (0.5955) | 1.8721 (2.9931) | 2.1187 (0.0885) |
| 0.600 | −0.0647 (1.0234) | 5.0993 (4.0247) | 4.5210 (0.5926) | 2.0976 (2.9638) | 2.0859 (0.0874) |
| 0.650 | −0.3912 (1.0276) | 4.9361 (4.4455) | 4.6488 (0.5933) | 2.0772 (2.9719) | 2.0109 (0.0874) |
| 0.700 | −0.3798 (1.0308) | 4.8600 (4.6174) | 4.8012 (0.5951) | 1.9504 (2.9801) | 1.9422 (0.0873) |
| 0.750 | −0.6758 (1.0288) | 4.7312 (4.9954) | 4.6981 (0.5925) | 1.8820 (2.9720) | 1.8932 (0.0876) |
| 0.800 | −0.5877 (1.0242) | 4.6562 (5.1449) | 4.7533 (0.5890) | 1.3784 (2.9703) | 1.8054 (0.0870) |
| 0.850 | −0.4422 (1.0190) | 4.5382 (5.2731) | 4.8885 (0.5855) | 0.7438 (2.9833) | 1.7187 (0.0858) |
| 0.900 | −0.5595 (1.0140) | 4.5003 (5.5016) | 5.0596 (0.5821) | 0.4064 (2.9769) | 1.6474 (0.0853) |
| 1.000 | −0.3677 (1.0031) | 4.4912 (5.2274) | 5.0329 (0.5760) | −0.0075 (2.9378) | 1.4923 (0.0856) |
| 1.500 | −0.1422 (0.9582) | 4.5107 (5.3076) | 4.8824 (0.5474) | 0.4718 (2.7923) | 0.6833 (0.0810) |
| 2.000 | −0.0189 (0.9018) | 5.1552 (4.3872) | 4.2165 (0.5200) | 0.2934 (2.6677) | −0.1004 (0.0756) |
| 2.500 | −0.2975 (0.8723) | 5.2047 (4.4814) | 3.7073 (0.4969) | 0.0802 (2.6111) | −0.4225 (0.0750) |
| 3.000 | 0.2826 (0.8450) | 4.9119 (4.1666) | 3.7782 (0.4778) | 0.1242 (2.5626) | −0.4079 (0.0733) |
| 3.500 | 1.2547 (0.8508) | 4.8181 (4.0037) | 3.9819 (0.4752) | −0.0077 (2.5380) | −0.4145 (0.0729) |
| 4.000 | 1.3908 (0.9380) | 4.8722 (4.2993) | 3.8955 (0.4829) | −0.2255 (2.4982) | −0.4439 (0.0750) |
| 4.500 | 1.7688 (0.9571) | 4.8395 (4.4007) | 3.8911 (0.4905) | −0.1485 (2.5014) | −0.4584 (0.0770) |
| 5.000 | 2.1116 (0.9720) | 4.8548 (4.3236) | 4.0259 (0.4993) | −0.7652 (2.5205) | −0.4643 (0.0782) |
| 5.500 | 2.2480 (0.9813) | 4.6786 (4.2584) | 4.0979 (0.5080) | −0.8335 (2.5393) | −0.5014 (0.0786) |
| 6.000 | 3.1817 (0.9949) | 4.5478 (4.0776) | 4.1635 (0.5152) | 0.5619 (2.6354) | −0.4961 (0.0790) |
| 6.500 | 3.6233 (1.0092) | 4.4139 (3.9542) | 4.2051 (0.5219) | 0.4891 (2.6306) | −0.4862 (0.0805) |
| 7.000 | 3.5816 (1.0115) | 4.1887 (3.9868) | 4.2891 (0.5237) | 0.0857 (2.6543) | −0.4653 (0.0815) |
| 7.500 | 3.5051 (1.0153) | 4.1406 (4.0215) | 4.4842 (0.5257) | 0.1012 (2.6944) | −0.4559 (0.0817) |
| 8.000 | 3.8776 (1.0466) | 4.2468 (3.8273) | 4.7190 (0.5588) | −0.1272 (2.7277) | −0.4092 (0.0823) |
| 8.500 | 3.8256 (1.0579) | 4.2901 (3.9610) | 4.8394 (0.5638) | 0.1772 (2.6961) | −0.4048 (0.0842) |
| 9.000 | 3.5976 (1.0610) | 4.3540 (4.0723) | 4.8069 (0.5633) | 0.1897 (2.6534) | −0.4082 (0.0855) |
| 10.000 | 3.5188 (1.0670) | 4.3974 (4.1084) | 4.9927 (0.5699) | −0.5129 (2.7015) | −0.4146 (0.0864) |

PGA, peak ground acceleration.

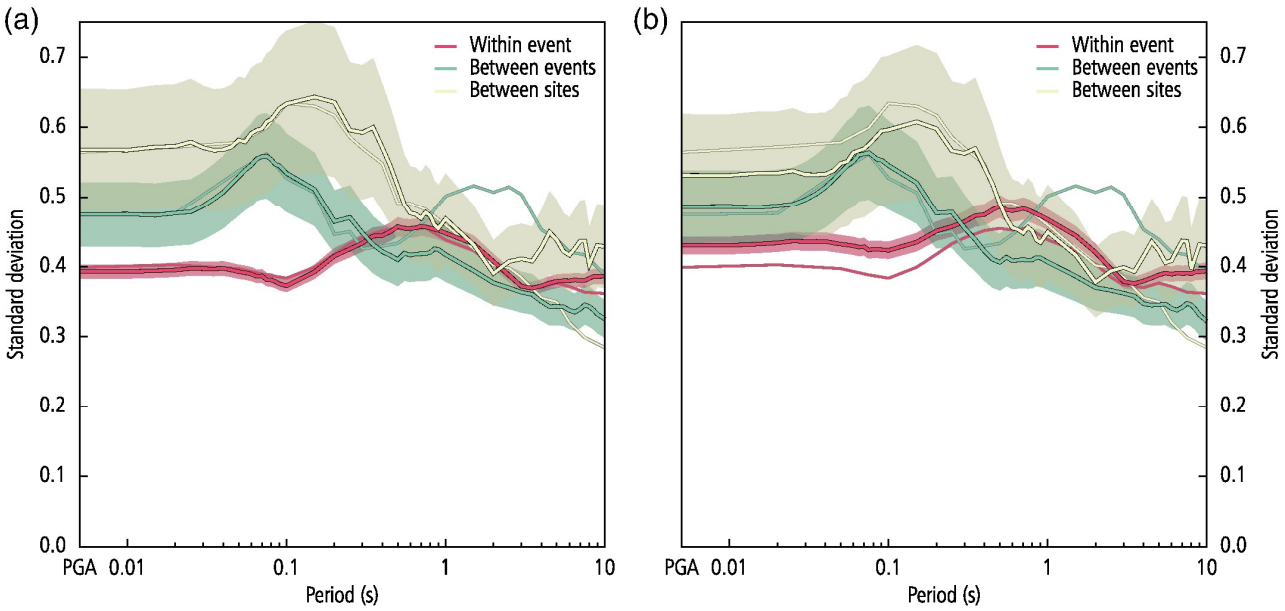


***Figure 5.*** *Model standard deviations (random-effects variances + residual variance) with oscillatory period of the mixed-effects regression with (a) deterministic weights and (b) probabilistic weights (ACE). Thick lines are standard deviations from this study with 95% confidence regions shown as bands. Thin lines are the standard deviations of Montalva et al. (2017).*

## RESULTS AND DISCUSSION

The GMMs for Chile based on deterministic and ACE classifications (Fig. 3) show minor variations (Fig. 4). These two models based on events with reported FMs only are also in agreement (Fig. 4) with the model of Montalva et al. (2017) for the entire flat file data.

The random-effects variances of the new models are also comparable (Table 2) with variances of Montalva et al. (2017). Although the uncertainty estimates of the parameters are stable with frequency, the variance uncertainties are frequency dependent (shown in terms of standard deviations in Fig. 5 and Table 3) and peak together with the random-effects variances around 0.1 s. The total variances are near identical for deterministic and ACE classifications (Fig. 6), whereas the random-effects variances show differences in the partition, which can be attributed to the different weighting methodologies. Although the between-event residuals have comparable variances, the between-site and within-event variances are decreased and increased for all periods, respectively. The difference ratios are nearly constant for all periods (Fig. 5).

To test the predictive power of the two weighting methodologies, we performed $K$-fold cross validation with $K = 10$ to estimate the predictive error of the models when applied to data not used in the parameter estimation (Hastie et al., 2009). The overall predictive errors reflect the total uncertainties of the GMMs (Fig. 6). Differences in the predictive error of the expert- and ACE-based models are small and decrease with increasing period and are within each other error bands. The slightly increased predictive error may be attributed to the fact that ACE discards more data than the expert, because the ACE-based inversions uses effectively only 85% of the data that are used in the expert-classification-based inversions.

The new methods developed in this article give also the opportunity to analyze the correlations among GMM parameters and variances. The parameter covariance shows intricate correlations among GMM parameters, while virtually no correlations exist between the random-effects variances (Fig. 7). The two random-effects variances ($\tau^2$, $\phi_{S2S}^2$) correlate with the residual variance $\phi^2$, which is a consequence of equation (20). The correlation of random effects is usually assumed to be negligible but is rarely tested (e.g. Al Atik et al., 2010). For the model presented here, the random-effects variances have negligible correlations. As can be seen from equation (20), the covariance matrix (and thus the correlation) is dependent on the data set and the GMM, therefore systematic evaluation of the model covariance matrix is important. The GMM parameters and the variances have vanishing correlations, as implied by the covariance matrix in equation (20) and shown in Figure 7. The correlation matrix indicates the trade-off between the model parameters and help identify those that are poorly resolved. Ground-motion parameters related to the magnitude scaling and near-source ground-motion saturation ($\theta_1$ to $\theta_5$) are strongly correlated as expected. The correlation matrix also confirms the fact that the site term (parameter $\theta_{12}$ in equation 28) is weakly correlated with other parameters, as suggested by Fourier spectra parameter inversion (e.g., Drouet et al., 2008).

**TABLE 3**
**GMM Random-Effects Standard Deviations with Error Standard Deviation in Parentheses**

| $T$ ($s$) | $\phi$ | $\tau$ | $\phi_{S2S}$ |
|---|---|---|---|
| PGA | 0.4304 (0.0064) | 0.4862 (0.0264) | 0.5315 (0.0444) |
| 0.010 | 0.4304 (0.0064) | 0.4865 (0.0265) | 0.5314 (0.0443) |
| 0.015 | 0.4322 (0.0064) | 0.4849 (0.0263) | 0.5343 (0.0448) |
| 0.020 | 0.4341 (0.0065) | 0.4873 (0.0266) | 0.5344 (0.0449) |
| 0.025 | 0.4366 (0.0065) | 0.4895 (0.0268) | 0.5380 (0.0455) |
| 0.030 | 0.4359 (0.0065) | 0.4964 (0.0275) | 0.5325 (0.0447) |
| 0.035 | 0.4361 (0.0065) | 0.5049 (0.0283) | 0.5323 (0.0447) |
| 0.040 | 0.4361 (0.0065) | 0.5142 (0.0292) | 0.5330 (0.0448) |
| 0.045 | 0.4338 (0.0065) | 0.5250 (0.0303) | 0.5381 (0.0454) |
| 0.050 | 0.4323 (0.0064) | 0.5358 (0.0314) | 0.5484 (0.0469) |
| 0.055 | 0.4311 (0.0064) | 0.5454 (0.0323) | 0.5512 (0.0472) |
| 0.060 | 0.4296 (0.0063) | 0.5515 (0.0329) | 0.5638 (0.0490) |
| 0.065 | 0.4277 (0.0063) | 0.5609 (0.0339) | 0.5667 (0.0494) |
| 0.070 | 0.4284 (0.0063) | 0.5618 (0.0340) | 0.5693 (0.0498) |
| 0.075 | 0.4238 (0.0062) | 0.5639 (0.0341) | 0.5753 (0.0505) |
| 0.080 | 0.4249 (0.0062) | 0.5591 (0.0336) | 0.5819 (0.0516) |
| 0.085 | 0.4247 (0.0062) | 0.5541 (0.0331) | 0.5884 (0.0525) |
| 0.090 | 0.4242 (0.0062) | 0.5535 (0.0330) | 0.5948 (0.0535) |
| 0.100 | 0.4233 (0.0062) | 0.5455 (0.0322) | 0.5965 (0.0537) |
| 0.150 | 0.4351 (0.0065) | 0.5208 (0.0298) | 0.6078 (0.0558) |
| 0.200 | 0.4511 (0.0070) | 0.4769 (0.0259) | 0.5979 (0.0548) |
| 0.250 | 0.4573 (0.0072) | 0.4783 (0.0261) | 0.5666 (0.0503) |
| 0.300 | 0.4646 (0.0074) | 0.4552 (0.0242) | 0.5628 (0.0499) |
| 0.350 | 0.4728 (0.0077) | 0.4389 (0.0229) | 0.5698 (0.0512) |
| 0.400 | 0.4746 (0.0077) | 0.4221 (0.0215) | 0.5429 (0.0473) |
| 0.450 | 0.4815 (0.0079) | 0.4108 (0.0207) | 0.5188 (0.0441) |
| 0.500 | 0.4868 (0.0081) | 0.4064 (0.0204) | 0.4903 (0.0404) |
| 0.550 | 0.4814 (0.0079) | 0.4136 (0.0209) | 0.4691 (0.0374) |
| 0.600 | 0.4803 (0.0079) | 0.4084 (0.0204) | 0.4555 (0.0357) |
| 0.650 | 0.4815 (0.0079) | 0.4094 (0.0206) | 0.4529 (0.0354) |
| 0.700 | 0.4837 (0.0080) | 0.4102 (0.0207) | 0.4479 (0.0348) |
| 0.750 | 0.4815 (0.0079) | 0.4091 (0.0205) | 0.4563 (0.0358) |
| 0.800 | 0.4784 (0.0078) | 0.4097 (0.0205) | 0.4513 (0.0351) |
| 0.850 | 0.4754 (0.0077) | 0.4131 (0.0208) | 0.4380 (0.0334) |
| 0.900 | 0.4730 (0.0076) | 0.4125 (0.0207) | 0.4346 (0.0329) |
| 1.000 | 0.4674 (0.0075) | 0.4067 (0.0201) | 0.4495 (0.0346) |
| 1.500 | 0.4444 (0.0068) | 0.3864 (0.0182) | 0.4166 (0.0301) |
| 2.000 | 0.4203 (0.0061) | 0.3703 (0.0167) | 0.3779 (0.0253) |
| 2.500 | 0.3966 (0.0054) | 0.3662 (0.0161) | 0.3941 (0.0266) |
| 3.000 | 0.3790 (0.0050) | 0.3622 (0.0156) | 0.3988 (0.0267) |
| 3.500 | 0.3757 (0.0049) | 0.3583 (0.0153) | 0.3929 (0.0262) |
| 4.000 | 0.3797 (0.0050) | 0.3473 (0.0146) | 0.4146 (0.0289) |
| 4.500 | 0.3840 (0.0052) | 0.3450 (0.0145) | 0.4354 (0.0315) |
| 5.000 | 0.3883 (0.0053) | 0.3456 (0.0147) | 0.4228 (0.0304) |
| 5.500 | 0.3909 (0.0055) | 0.3465 (0.0150) | 0.4059 (0.0288) |
| 6.000 | 0.3893 (0.0055) | 0.3406 (0.0146) | 0.4085 (0.0292) |
| 6.500 | 0.3909 (0.0056) | 0.3383 (0.0145) | 0.4235 (0.0312) |
| 7.000 | 0.3902 (0.0056) | 0.3421 (0.0148) | 0.4369 (0.0329) |
| 7.500 | 0.3905 (0.0056) | 0.3478 (0.0153) | 0.4360 (0.0329) |
| 8.000 | 0.3902 (0.0058) | 0.3454 (0.0154) | 0.3968 (0.0290) |
| 8.500 | 0.3929 (0.0059) | 0.3389 (0.0151) | 0.4152 (0.0314) |
| 9.000 | 0.3927 (0.0059) | 0.3312 (0.0145) | 0.4313 (0.0335) |
| 10.000 | 0.3943 (0.0060) | 0.3223 (0.0140) | 0.4301 (0.0334) |

PGA, peak ground acceleration.

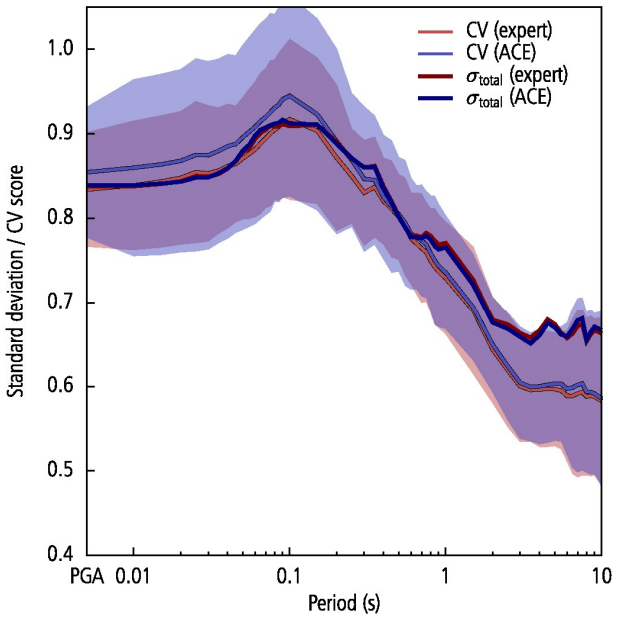


***Figure 6.*** *Cross-validation (CV) based estimates of the predictive errors of the GMM for Chile with deterministic (expert-based) weights in red and probabilistic (ACE) weights in blue. Bands indicate uncertainties (two standard deviation) of the predictive error estimates. The predictive errors are similar to the total standard deviations of the GMMs for all investigated periods (Fig. 5) and both weighting methodologies perform virtually identically and are within each other uncertainty ranges. The slightly broadened uncertainty bands for the ACE-based cross validation are most likely due to the slightly lower data size used, as ACE excludes approximately 12% less data than in the expert-based catalog.*

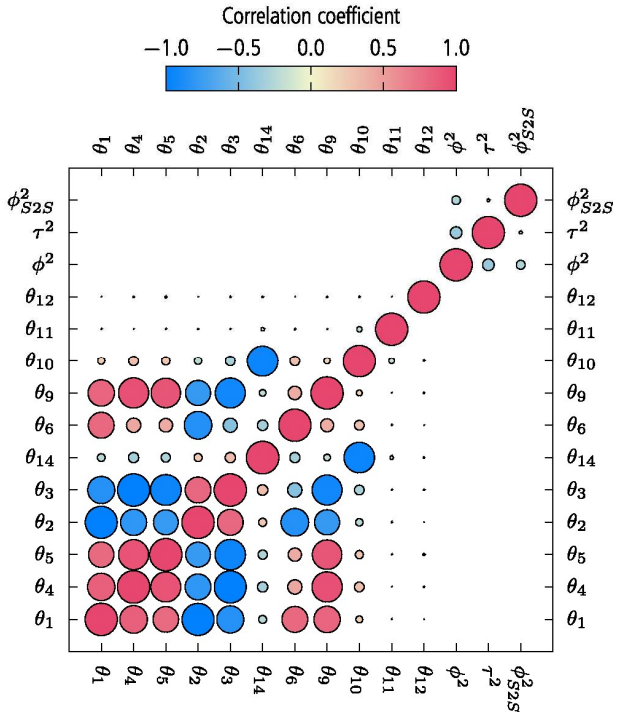


***Figure 7.*** *Correlation coefficients of the mixed-effects regression for PGA from the covariance matrix (equation 20). No correlations between the GMM parameters and the model variances exist by definition. However, many parameters of the GMM are correlated with each other, while some are not: the event-type-dependent path parameter $\theta_{14}$ and the event-type parameter $\theta_{10}$ only correlate with each other, while the depth parameter $\theta_{11}$ and the site parameter $\theta_{12}$ correlate with no other parameter.*

Several past crustal earthquakes studies (e.g., Abrahamson and Silva, 2008) have suggested that aftershocks generate weaker ground motions than the associated mainshock. Some recent analysis of ground-motion residuals have not only confirmed such lower shaking associated with aftershocks but have also suggested that ground-motion residuals computed using a backbone model may change several months before large earthquakes (Piña-Valdés et al., 2018). The detection and quantification of these time dependencies are important, because they may reveal changes in the signature of earthquake spectra, the long-term preparation phase of large earthquakes or postseismic healing processes (Socquet et al., 2017; Bindi et al., 2018; Piña-Valdés et al., 2018). Because of this recent and stimulating literature, we have computed (Fig. 8) the time dependencies of between-event residuals 15 yr before the Maule earthquake and 5 yr after. The results suggest a slight decrease of between events before the mainshock, which is consistent with the progressive decrease of the released energy at high frequencies observed before the Iquique earthquake (Piña-Valdés et al., 2018). These observations may indicate a change on the subduction interface that may be related to a long-term nucleation process of the megathrust earthquake (Socquet et al., 2017). Figure 8 shows, however, that these preseismic changes are still poorly constrained because of the lack of data and the variability of the observed residuals.

One particular aspect usually neglected in mixed-effects regression within seismology is the scaling of the variances. When performing mixed-effects regression for both parameters and random-effects variances, then the RML should be used to avoid an underestimation of the random-effects variances by accounting for the reduction of degrees of freedom (Patterson and Thompson, 1971; Harville, 1974). The reduction of degrees of freedom depends on the number of free parameters in the model (e.g., equation B29); and probably the most famous example is the $N-1$ divisor in the unbiased estimate of the sample variance with unknown mean (the mean is the free parameter, hence the degrees of freedom are reduced by one). Thus, when neglecting the reduced freedom of degrees, the underestimation decreases with increasing data size for a given number of model parameters. The introduction of the mixed-effects model into the strong-motion seismological community is closely linked to the work by Abrahamson and Youngs (1992). Their widely cited mixed-effects algorithm is based on Searle (1971, ch. 8b, p. 462). However, RML for mixed effects in a general

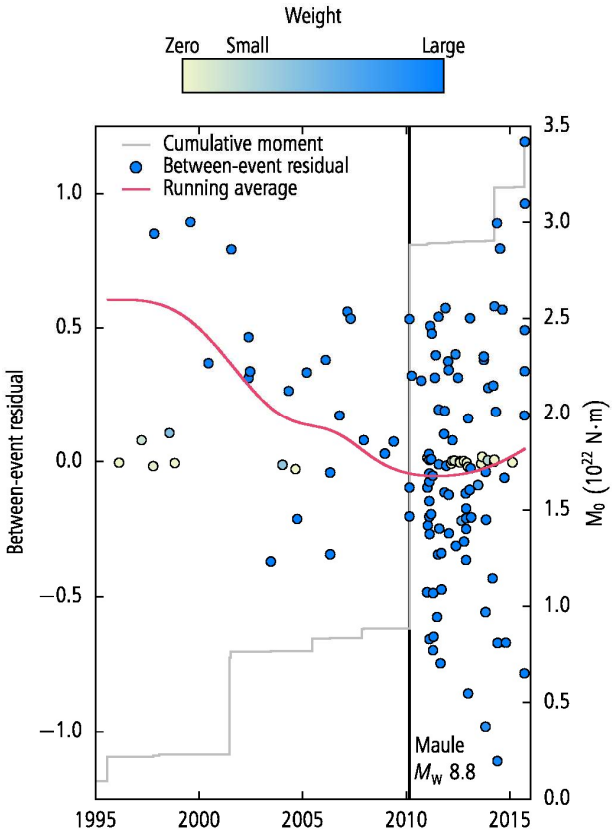


***Figure 8.*** *Temporal variations of the between-event residuals of the GMMs based on ACE-weighted mixed-effects regression. The residuals follow the same trend and drop after major earthquakes, in particular, after the 2010 Maule earthquake. The cumulative seismic moment of the earthquake catalog as step-line plot with the 2010 Maule earthquake having the single highest increase in released moment.*

definition has been introduced by Patterson and Thompson (1971), that is, simultaneously to Searle (1971). Several works related to the RML for mixed effects followed Patterson and Thompson (1971) (e.g., Harville, 1974; Lindstrom and Bates, 1988), and it is widely applied in many fields (Sahai and Ageel, 2000). Implementations of both types of likelihoods have been realized in several software packages: lme4 in R, statsmodels in python, fitlme in MATLAB (see Data and Resources) to name a few. In case of lme4, RML is automatically included when crossed random effects are used (as is the usual case with site and event random effects). Although the bias decreases with increasing data size, using the RML ensures better comparability between different GMMs.

## CONCLUSION

This article introduces the basis for a generalized mixed-effects regression by incorporating data weights through a weighted likelihood. The type of weights is more general and can be used in addition to data weights in the data covariance matrix. Although the latter type of weights usually represents uncertainties related to the dependent variable, for example, measurement errors (uncertainties of the ground-motion values) or a different weighting of residuals (e.g., robust regression), the former type of weights can represent not directly related uncertainties, for example, degree of belief, event classification (either data-driven or expert judgment). With ACE-based data selection as a data-driven and automatic determination of SoF classes, we provide a fully machine-learning-based GMM. The performance of the automatically derived GMM with regard to its parameters is comparable to the expert-based event classification of Montalva et al. (2017).

The total variances are near identical for the models based on deterministic and ACE classifications at all investigated periods that demonstrates that the data weighting in the likelihood function does not alter the total model uncertainty. The difference between covariance-based weighting and likelihood-based weighting for the mixed-effects model is analogous to the basic weighted mean and weighted variance estimators in Table 1. Although there is no difference in the estimators of the weighted mean (compare with the fixed-effects estimator), the variance estimators differ by their normalization constant: sum of weights versus number of data (compare with the estimators for the within-event variance and the random-effects variances). In the particular case of ACE-based weights with a value range between 0 and 1, we would underestimate all model uncertainties by approximately 12% for the shown Chile example in the covariance-weighted approach, because the sum of all ACE weights is 88% of the total number of data used in the regression. This difference in uncertainty estimates - which can be of arbitrary size - demonstrates the necessity of the formal incorporation of data weights through the weighted likelihood if the weight units (e.g., no unit as in case of ACE) differ from the data units (e.g., meters per second squared for ground shaking).

The cluster model of ACE can be directly implemented in modeling ground motion as weights for the events in the catalog. Once the mixture model has been learned from FM data set, new data can be classified instantly without a new cluster analysis. This opens the possibility to update ground-motion models in an automated way and can augment other updating routines as the Bayesian GMM approach of Stafford (2019).

We provide here a weight-augmented mixed-effects model that can be applied to both deterministic classification (expert judgment) and probabilistic classification (ACE). However, ACE treats data objectively and requires FM data alone, whereas expert judgment is subjective and may require additionally the event location. For large-magnitude events, hypocenters may not be suitable for location-dependent event classification due to the large extent of the rupture plane and the lack of any preferred

nucleation point of the rupture (Mai et al., 2005), which may result in mislabeling events.

The formulation used for the mixed-effects model is kept as general as possible and allows for an arbitrary number of linearly combined random effects and is therefore suitable to design GMMs with a single framework. This facilitates hybrid models that partially incorporate mixed-effects regression (e.g., Anderson and Brune, 1999; Sahakian et al., 2018). The incorporation of weights allows for smooth continuous spatial variations in regional GMMs - similar to the varying coefficient model by Landwehr et al. (2016) or the mixed-effects model with spatial correlation by Ming et al. (2019) - thus opening the possibility for the development of nonergodic GMMs with mixed-effects regression.

## DATA AND RESOURCES

The Chile strong-motion data file (Bastías and Montalva, 2016) is available at https://datacenterhub.org/resources/14342 (last accessed January 2020). The focal mechanism data of the Global Centroid Moment Tensor (CMT) catalog was downloaded at https://www.globalcmt.org/CMTsearch.html (last accessed January 2020). The supplemental material contains the coefficients and standard deviations of the ground-motion model (GMM) in equation (28) for peak ground acceleration (PGA) and periods between 0.01 and 10 s. The MATLAB is available at www.mathworks.com/products/matlab (last accessed May 2020). lme4 is available at https://cran.r-project.org/web/packages/lme4/index.html (last accessed May 2020) for the R language (https://www.2r-project.org/, last accessed May 2020). Statsmodels is available at https://www.statsmodels.org/stable/index.html (last accessed May 2020) for the python language (https://www.python.org/, last accessed May 2020). All ground-motion models derived in this study were computed with the linear algebra library armadillo (http://arma.sourceforge.net/docs.html, last accessed May 2020) for C++. All graphics were designed in gnuplot (http://www.gnuplot.info/, last accessed May 2020) and LaTeX (https://www.latex-project.org/, last accessed May 2020).

## ACKNOWLEDGMENTS

The authors are very thankful to John G. Anderson and John Douglas as well as two anonymous reviewers for their helpful comments to improve this article.

## REFERENCES

Abrahamson, N., and W. Silva (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, Earthq. Spectra 24, no. 1, 67–97.

Abrahamson, N., N. Gregor, and K. Addo (2016). BC Hydro ground motion prediction equations for subduction earthquakes, Earthq. Spectra 32, no. 1, 23–44.

Abrahamson, N. A., and R. R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, Bull. Seismol. Soc. Am. 82, no. 1, 505.

Aitken, A. C. (1936). IV––On least squares and linear combination of observations, Proc. Roy. Soc. Edinb. 55, 42–48.

Al Atik, L., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components, Seismol. Res. Lett. 81, no. 5, 794–801.

Anderson, J. G., and J. N. Brune (1999). Probabilistic seismic hazard analysis without the ergodic assumption, Seismol. Res. Lett. 70, no. 1, 19–28.

Barzilai, J., and J. M. Borwein (1988). Two-point step size gradient methods, IMA J. Numer. Anal. 8, no. 1, 141–148.

Bastías, N., and G. A. Montalva (2016). Chile strong ground motion flatfile, Earthq. Spectra 32, no. 4, 2549–2566.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4, J. Stat. Software 67, no. 1, 251–264.

Bates, D. M., and S. DebRoy (2004). Linear mixed models and penalized least squares, J. Multivariate Anal. 91, no. 1, 1–17.

Bindi, D., F. Cotton, D. Spallarossa, M. Picozzi, and E. Rivalta (2018). Temporal variability of ground shaking and stress drop in Central Italy: A hint for fault healing? Bull. Seismol. Soc. Am. 108, no. 4, 1853–1863.

Bott, M. H. P. (1959). The mechanics of oblique slip faulting, Geol. Mag. 96, no. 2, 109.

Chen, Y.-S., G. Weatherill, M. Pagani, and F. Cotton (2018). A transparent and data-driven global tectonic regionalization model for seismic hazard assessment, Geophys. J. Int. 213, no. 2, 1263–1280.

Demidenko, E. (2013). Mixed Models, Second Ed., Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., Hoboken, New Jersey.

Douglas, J. (2010). Consistency of ground-motion predictions from the past four decades, Bull. Earthq. Eng. 8, no. 6, 1515–1526.

Drouet, S., S. Chevrot, F. Cotton, and A. Souriau (2008). Simultaneous inversion of source spectra, attenuation parameters, and site responses: Application to the data of the French Accelerometric Network, Bull. Seismol. Soc. Am. 98, no. 1, 198–219.

Ekström, G., M. Nettles, and A. M. Dziewoński (2012). The global CMT project 2004-2010: Centroid-moment tensors for 13,017 earthquakes, Phys. Earth Planet. In. 200/201, 1–9.

Field, C., and B. Smith (1994). Robust estimation: A weighted maximum likelihood approach, Int. Stat. Rev. 62, no. 3, 405.

Haendel, A., S. Specht, N. M. Kuehn, and F. Scherbaum (2015). Mixtures of ground-motion prediction equations as backbone models for a logic tree: An application to the subduction zone in Northern Chile, Bull. Earthq. Eng. 13, no. 2, 483–501.

Hartley, H. O., and J. N. K. Rao (1967). Maximum-likelihood estimation for the mixed analysis of variance model, Biometrika 54, nos. 1/2, 93.

Harville, D. (1976). Extension of the Gauss-Markov theorem to include the estimation of random effects, Ann. Stat. 4, no. 2, 384–395.

Harville, D. A. (1974). Bayesian inference for variance components using only error contrasts, Biometrika 61, no. 2, 383–385.

Harville, D. A. (1977). Maximum likelihood approaches to variance component estimation and to related problems, J. Am. Stat. Assoc. 72, no. 358, 320–338.

Harville, D. A. (1997). Matrix Algebra from a Statistician's Perspective, Springer, New York, New York.

Hastie, T., R. Tibshirani, and J. Friedman (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Ed., Springer Science & Business Media, New York, New York.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model, Biometrics 31, no. 2, 423–447.

Henderson, C. R., O. Kempthorne, S. R. Searle, and C. M. von Krosigk (1959). The estimation of environmental and genetic trends from records subject to culling, Biometrics 15, no. 2, 192–218.

Henderson, H. V., and S. R. Searle (1981). On deriving the inverse of a sum of matrices, SIAM Rev. 23, no. 1, 53–60.

Kagan, Y. Y. (1991). 3-D rotation of double-couple earthquake sources, Geophys. J. Int. 106, no. 3, 709–716.

Kao, H., and R.-J. Rau (1999). Detailed structures of the subducted Philippine Sea Plate beneath northeast Taiwan: A new type of double seismic zone, J. Geophys. Res. 104, no. B1, 1015–1033.

Kotha, S. R., F. Cotton, and D. Bindi (2018). A new approach to site classification: Mixed-effects ground motion prediction equation with spectral clustering of site amplification functions, Soil Dynam. Earthq. Eng. 110, January, 318–329.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). A

nonergodic ground-motion model for California with spatially varying coefficients, Bull. Seismol. Soc. Am. 106, no. 6, 2574–2583.
Lindstrom, M. J., and D. M. Bates (1988). Newton-Raphson and EM algorithms for linear models for repeated-measures data, J. Am. Stat. Assoc. 83, no. 404, 1014–1022.
Lisle, R. J. (2013). A critical look at the Wallace-Bott hypothesis in fault-slip analysis, Bull. Soc. Geol. France 184, nos. 4/5, 299–306.
Mai, P. M., P. Spudich, and J. Boatwright (2005). Hypocenter locations in finite-source rupture models, Bull. Seismol. Soc. Am. 95, no. 3, 965–980.
Markatou, M., A. Basu, and B. G. Lindsay (1998). Weighted likelihood equations with bootstrap root search, J. Am. Stat. Assoc. 93, no. 442, 740–750.
Ming, D., C. Huang, G. W. Peters, and C. Galasso (2019). An advanced estimation algorithm for ground-motion models with spatial correlation, Bull. Seismol. Soc. Am. 109, no. 2, 541–566.
Montalva, G. A., N. Bastías, and A. Rodriguez-Marek (2017). Ground-motion prediction equation for the Chilean subduction zone, Bull. Seismol. Soc. Am. 107, no. 2, 901–911.
Patterson, H. D., and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal, Biometrika 58, no. 3, 545–554.
Piña-Valdés, J., A. Socquet, F. Cotton, and S. Specht (2018). Spatiotemporal variations of ground motion in Northern Chile before and after the 2014 Mw 8.1 Iquique megathrust event, Bull. Seismol. Soc. Am. 108, no. 2, 801–814.
Rosner, B. (1975). On the detection of many outliers, Technometrics 17, no. 2, 221.
Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure, Technometrics 25, no. 2, 165.
Sahai, H., and M. I. Ageel (2000). The Analysis of Variance, First Ed., Birkhäuser, Boston, Massachusetts.
Sahakian, V., A. Baltay, T. Hanks, J. Buehler, F. Vernon, D. Kilb, and N. Abrahamson (2018). Decomposing leftovers: Event, path, and site residuals for a small-magnitude Anza region GMPE, Bull. Seismol. Soc. Am. 108, doi: 10.1785/0120170376.
Searle, S. (1971). Linear Models, John Wiley & Sons, New York, New York.
Socquet, A., J. P. Valdes, J. Jara, F. Cotton, A. Walpersdorf, N. Cotte, S. Specht, F. Ortega-Culaciati, D. Carrizo, and E. Norabuena (2017). An 8 month slow slip event triggers progressive nucleation of the 2014 Chile megathrust, Geophys. Res. Lett. 44, no. 9, 4046–4053.
Specht, S., O. Heidbach, F. Cotton, and A. Zang (2017). Data-driven earthquake focal mechanism cluster analysis, Tech. Report, GFZ German Research Centre for Geosciences, Potsdam, Germany.
Stafford, P. J. (2019). Continuous integration of data into ground-motion models using Bayesian updating, J. Seismol. 23, no. 1, 39–57.
von Specht, S., O. Heidbach, F. Cotton, and A. Zang (2018). Uncertainty reduction of stress tensor inversion with data-driven catalogue selection, Geophys. J. Int. 214, no. 3, 2250–2263.
Wallace, R. E. (1951). Geometry of shearing stress and relation to faulting, J. Geol. 59, no. 2, 118–130.
Walling, M., W. Silva, and N. Abrahamson (2008). Nonlinear site amplification factors for constraining the NGA models, Earthq. Spectra 24, no. 1, 243–255.
Wolfinger, R., and M. O'Connell (1993). Generalized linear mixed models a pseudo-likelihood approach, J. Stat. Comput. Simul. 48, nos. 3/4, 233–243.

## APPENDIX A

### *Effect of different weighting methods*

To show the difference between generalized (weighted) least-squares and weighted likelihood estimators, both types of estimators for the mean and variance are derived, respectively. The weighted least-squares estimator of a parameter minimizes the log likelihood for $N$ independent and identically distributed samples each with probability $p_i$:

$$\ln \mathcal{L} = \sum_{i=1}^{N} \ln p_i \tag{A1}$$

Least squares are based on the assumption of normally distributed variates and the log likelihood is

$$\ln \mathcal{L} = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{x}-\mu)^T\boldsymbol{W}(\boldsymbol{x}-\mu). \tag{A2}$$

The minimum of the log likelihood is attained when the derivatives
with respect to mean and variance vanish:

$$\frac{\partial \ln \mathcal{L}}{\partial \mu} = 0, \tag{A3}$$

$$= \boldsymbol{W}(\boldsymbol{x}-\mu), \tag{A4}$$

and

$$\frac{\partial \ln \mathcal{L}}{\partial \sigma^2} = 0, \tag{A5}$$

$$= -\frac{N}{\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{x}-\mu)^T\boldsymbol{W}(\boldsymbol{x}-\mu). \tag{A6}$$

Using sum representation, the weighted mean follows

$$\hat{\mu} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \tag{A7}$$

and the weighted variance

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} w_i (x_i - \hat{\mu})^2. \tag{A7}$$

Although the weighted mean is normalized by the weights, the weighted variance is not.

Now, we derive the estimators from weighted likelihood:

$$\ln \mathcal{L}_w = \sum_{i=1}^{N} w_i \ln p_i. \tag{A9}$$

For the normal distribution holds

$$\ln \mathcal{L} = -\frac{1}{2}\sum_{i=1}^{N} w_i \left(\ln(2\pi\sigma^2) + \frac{(x_i-\mu)^2}{\sigma^2}\right). \tag{A10}$$

The minimum of the log likelihood is attained when the

derivative vanishes:

$$\frac{\partial \ln \mathcal{L}_w}{\partial \mu} = 0, \tag{A11}$$

$$= \sum_{i=1}^{N} w_i (x_i - \mu), \tag{A12}$$

and

$$\frac{\partial \ln \mathcal{L}_w}{\partial \sigma^2} = 0, \tag{A13}$$

$$= \sum_{i=1}^{N} w_i \left( \frac{1}{\sigma^2} - \frac{(x_i - \mu)^2}{\sigma^4} \right). \tag{A14}$$

The weighted mean follows as

$$\hat{\mu} = \frac{\sum_{i=1}^{N} w_i x_i}{\sum_{i=1}^{N} w_i}, \tag{A15}$$

and the weighted variance

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{N} w_i (x_i - \hat{\mu})^2}{\sum_{i=1}^{N} w_i}. \tag{A16}$$

The weighted mean is identical for both approaches, while only the weighted likelihood variance is (properly) scaled by the sum of weights and not by the number of samples.

## APPENDIX B

### *Derivation of the coefficients of the mixed-effects model based on weighted likelihood*

*The estimators for the ground-motion model parameters* $\boldsymbol{p}$ *and predictors for the random effects* $\boldsymbol{q}$. The maximum-likelihood (ML) estimator for $\boldsymbol{p}$ and ML predictor $\boldsymbol{q}$ are derived from the derivatives of the log likelihood in equation (7). Replacing the constant factors independent of $\boldsymbol{p}$ and $\boldsymbol{q}$ by $c$, the log likelihood is

$$\begin{aligned} \ln \mathcal{L} = c \\ &- \frac{1}{2\sigma^2} \Big( \boldsymbol{q}^T \boldsymbol{V}_q \boldsymbol{D}^{-1} \boldsymbol{V}_q \boldsymbol{q} \\ &+ \big( \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p})) - \boldsymbol{B}\boldsymbol{V}_q \boldsymbol{q} \big)^T \boldsymbol{C}^{-1} \big( \boldsymbol{V}(\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p})) \\ &- \boldsymbol{B}\boldsymbol{V}_q \boldsymbol{q} \big) \Big). \end{aligned} \tag{B1}$$

The weight matrices for the random effects, $\boldsymbol{V}_q$, and the residuals, $\boldsymbol{V}$, are related to each other (see Appendix C). Because $\boldsymbol{f}(\boldsymbol{p})$ can be nonlinear, the problem cannot be solved directly. The least-squares solution of a nonlinear problem is obtained by iteratively updating the parameter estimates from a start condition. The parameter update, $\Delta\boldsymbol{p}$, is calculated from the Taylor expansion of the model function truncated after the first derivative term. Thus, the residual term $\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p})$ is approximately linearized:

$$\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}) \approx \Delta\boldsymbol{y} - \boldsymbol{J}\Delta\boldsymbol{p} \tag{B2}$$

in which $\boldsymbol{\Delta y} = \boldsymbol{y} - \boldsymbol{f}(\boldsymbol{p}^{(i)})$ with $\boldsymbol{p}^{(i)}$ as the parameter estimate in the $i$th iteration. The matrix $\boldsymbol{J}$ is the Jacobian with the derivatives of $\boldsymbol{f}(\boldsymbol{p}^{(i)})$ with respect to $\boldsymbol{p}$. The parameter update $\Delta\boldsymbol{p}$ describes the difference between the current parameter estimate $\boldsymbol{f}(\boldsymbol{p}^{(i)})$ and the updated parameter estimate $\boldsymbol{f}(\boldsymbol{p}^{(i+1)})$. Using the relation in equation (B2), the derivatives of equation (B1) with respect to $\Delta\boldsymbol{p}$ and $\boldsymbol{q}$ are given by

$$\frac{\partial \ln \mathcal{L}}{\partial \Delta} = \frac{1}{\sigma^2} \boldsymbol{J}^T \boldsymbol{V} \boldsymbol{C}^{-1} \big( \boldsymbol{V}(\Delta\boldsymbol{y} - \boldsymbol{J}\Delta\boldsymbol{p}) - \boldsymbol{B}\boldsymbol{V}_q \boldsymbol{q} \big), \tag{B3}$$

$$\begin{aligned} \frac{\partial \ln \mathcal{L}}{\partial \boldsymbol{q}} = \frac{1}{\sigma^2} \big( \boldsymbol{B}^T \boldsymbol{C}^{-1} \big( \boldsymbol{V}(\Delta\boldsymbol{y} - \boldsymbol{J}\Delta\boldsymbol{p}) - \boldsymbol{B}\boldsymbol{V}_q \boldsymbol{q} \big) \\ - \boldsymbol{V}_q \boldsymbol{D}^{-1} \boldsymbol{V}_q \boldsymbol{q} \big). \end{aligned} \tag{B4}$$

Setting both derivatives to zero leads to the equations for the weighted mixed-effects model:

$$\begin{aligned} &\begin{pmatrix} \boldsymbol{J}^T \boldsymbol{V} \boldsymbol{C}^{-1} \boldsymbol{V} \Delta\boldsymbol{y} \\ \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{V} \Delta\boldsymbol{y} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{J}^T \boldsymbol{V} \boldsymbol{C}^{-1} \boldsymbol{V} \boldsymbol{J} & \boldsymbol{J}^T \boldsymbol{V} \boldsymbol{C}^{-1} \boldsymbol{B} \boldsymbol{V}_q \\ \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{V} \boldsymbol{J} & \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{B} \boldsymbol{V}_q + \boldsymbol{V}_q \boldsymbol{D}^{-1} \boldsymbol{V}_q \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{p} \\ \hat{\boldsymbol{q}} \end{pmatrix} \end{aligned} \tag{B5}$$

Solving the second row for the ML predictor of the random
effects

$$\hat{\boldsymbol{q}} = \big( \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{B} \boldsymbol{V}_q + \boldsymbol{V}_q \boldsymbol{D}^{-1} \boldsymbol{V}_q \big)^{-1} \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{V} \hat{\boldsymbol{r}}, \tag{B6}$$

with $\hat{\boldsymbol{r}} = \Delta\boldsymbol{y} - \boldsymbol{K}\Delta\boldsymbol{p}$. The following matrix substitution is given by Henderson and Searle (1981)

$$\begin{aligned} &\big( \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \boldsymbol{B} \boldsymbol{V}_q + \boldsymbol{V}_q \boldsymbol{D}^{-1} \boldsymbol{V}_q \big)^{-1} \boldsymbol{V}_q \boldsymbol{B}^T \boldsymbol{C}^{-1} \\ &\quad = \boldsymbol{V}_q^{-1} \boldsymbol{D} \boldsymbol{V}_q^{-1} \boldsymbol{V}_q \boldsymbol{B}^T \big( \boldsymbol{C} \\ &\quad + \boldsymbol{B} \boldsymbol{V}_q \boldsymbol{V}_q^{-1} \boldsymbol{D} \boldsymbol{V}_q^{-1} \boldsymbol{V}_q \boldsymbol{B}^T \big)^{-1}, \end{aligned} \tag{B7}$$

$$= \boldsymbol{V}_q^{-1} \boldsymbol{D} \boldsymbol{B}^T (\boldsymbol{C} + \boldsymbol{B} \boldsymbol{D} \boldsymbol{B}^T)^{-1}, \tag{B8}$$

and with

$$\boldsymbol{S} := \boldsymbol{C} + \boldsymbol{B} \boldsymbol{D} \boldsymbol{B}^T, \tag{B9}$$

the predictor for $\boldsymbol{q}$ is

$$\hat{\boldsymbol{q}} = \boldsymbol{V}_{\boldsymbol{q}}^{-1}\boldsymbol{D}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}} \tag{B10}$$

which is the best linear unbiased predictor of $\boldsymbol{q}$ (Henderson, 1975). Substituting $\hat{\boldsymbol{q}}$ in the first row of equation (B5):

$$\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{C}^{-1}(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{S}^{-1})\boldsymbol{V}\hat{\boldsymbol{r}} = 0, \tag{B11}$$

with $\boldsymbol{I}$ being the identity matrix,

$$\begin{gathered}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{C}^{-1}(\boldsymbol{S} - \boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{\boldsymbol{T}})\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}} = 0 \\ \boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{C}^{-1}\boldsymbol{C}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}} = 0 \\ \boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}} = 0\end{gathered} \tag{B12}$$

and solving for the parameters of the fixed-effects model:

$$\Delta\boldsymbol{p} = (\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\boldsymbol{J})^{-1}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}, \tag{B13}$$

which is the generalized least-squares estimator with covariance matrix $\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}$ (Aitken, 1936), and the best linear unbiased estimator (Henderson, 1975).

*The variance estimators.* The variance estimators of the mixed-effects model dependent on the estimator $\hat{\boldsymbol{p}}$ and predictor $\hat{\boldsymbol{q}}$ of the fixed-effects model. It is well established that ML variance estimates are biased downward by $\hat{\boldsymbol{p}}$ and $\hat{\boldsymbol{q}}$, that is, variances are underestimated Patterson and Thompson (1971), Harville (1974, 1976), Lindstrom and Bates (1988), Bates and DebRoy (2004), and Bates et al. (2015). To obtain unbiased variance estimates, the likelihood is de_ned in terms of error contrasts. Details of the concept are described in full detail by Patterson and Thompson (1971) and Harville (1976).

The basis for the variance estimators is to marginalize out the random effects $\boldsymbol{q}$ of the probability function from equation (2). First, completing the square for $\boldsymbol{q}$ in the exponential part:

$$\begin{gathered}(\boldsymbol{r} - \boldsymbol{B}\boldsymbol{q})^{\boldsymbol{T}}\boldsymbol{C}^{-1}(\boldsymbol{r} - \boldsymbol{B}\boldsymbol{q}) + \boldsymbol{q}^{\boldsymbol{T}}\boldsymbol{D}^{-1}\boldsymbol{q} \\ = (\boldsymbol{q} - \boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{P}(\boldsymbol{q} - \boldsymbol{u}) + \boldsymbol{v},\end{gathered} \tag{B14}$$

with

$$\begin{gathered}\boldsymbol{P} = \boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B} + \boldsymbol{D}^{-1} \\ \boldsymbol{u} = \boldsymbol{P}^{-1}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{r} \\ \boldsymbol{v} = \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{r} - \boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{r}.\end{gathered}$$

The integral over $\boldsymbol{q}$ is given by

$$\int_{\mathbb{R}^M} e^{-\frac{1}{2}(\boldsymbol{q}-\boldsymbol{u})^{\boldsymbol{T}}\boldsymbol{P}(\boldsymbol{q}-\boldsymbol{u})}\, d\boldsymbol{q} = \left(\frac{2\pi}{|\boldsymbol{P}|}\right)^{\frac{M}{2}}, \tag{B15}$$

and the marginal probability of equation (2) is

$$\begin{gathered}P(\boldsymbol{y}|\boldsymbol{f}(\boldsymbol{p}), \sigma^2, \boldsymbol{C}, \boldsymbol{D}) \\ = \frac{1}{\sqrt{(2\pi)^N|\sigma^2\boldsymbol{C}||\sigma^2\boldsymbol{D}||\sigma^{-2}(\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B} + \boldsymbol{D}^{-1})|}} \\ \times e^{-\frac{1}{2\sigma^2}\boldsymbol{r}^{\boldsymbol{T}}\left(\boldsymbol{C}^{-1} - \boldsymbol{C}^{-1}\boldsymbol{B}(\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B} + \boldsymbol{D}^{-1})^{-1}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\right)\boldsymbol{r}}.\end{gathered} \tag{B16}$$

Using the identities in equation (B8), equation (B9), as well as equations (B11) and (B12), the matrix expression in the exponential term reduces to $\boldsymbol{S}^{-1}$.With the generalized matrix determinant lemma (e.g., Harville, 1997, theorem 13.3.8) and the identity $|k\boldsymbol{M}| = k^n|\boldsymbol{M}|$ (in which $\boldsymbol{M}$ is of size $n \times n$), the determinant is restated as

$$\begin{gathered}|\sigma^2\boldsymbol{C}||\sigma^2\boldsymbol{D}||\sigma^{-2}(\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B} + \boldsymbol{D}^{-1})| \\ = \sigma^{2N}|\boldsymbol{C}||\boldsymbol{D}\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{C}^{-1}\boldsymbol{B} + \boldsymbol{I}| \\ = \sigma^{2N}|\boldsymbol{C} + \boldsymbol{B}\boldsymbol{D}\boldsymbol{B}^{\boldsymbol{T}}| \\ = \sigma^{2N}|\boldsymbol{S}|.\end{gathered} \tag{B17}$$

Thus, the marginal probability is

$$P(\boldsymbol{y}|\boldsymbol{f}(\boldsymbol{p}), \sigma^2, \boldsymbol{S}) = \frac{1}{\sqrt{(2\pi\sigma^2)^N|\boldsymbol{S}|}} e^{-\frac{1}{2\sigma^2}\boldsymbol{r}^{\boldsymbol{T}}\boldsymbol{S}^{-1}\boldsymbol{r}}. \tag{B18}$$

This probability distribution is given in terms of $\boldsymbol{p}$ – the unknown population parameter - and any ML estimates of the variances will be biased (Patterson and Thompson, 1971). Because $\boldsymbol{p}$ is unknown, its estimator $\hat{\boldsymbol{p}}$ (equation B13) is used instead, which is based on the same data as used for the variance estimators which also introduces a bias (Harville, 1974). This dependence reduces the degree of freedoms by the number of parameters $P$ in $\boldsymbol{p}(\hat{\boldsymbol{p}})$. Patterson and Thompson (1971) and Harville (1974) introduced the reduced ML (RML), which provides unbiased estimates of the variances by transforming the data to linearly independent error contrasts. Furthermore, the RML definition of Harville (1974) also expresses the likelihood in terms of the known estimator $\hat{\boldsymbol{p}}$ instead of the unknown population parameter $\boldsymbol{p}$. The RML is commonly used in mixed-effects regression and should be used when both fixed model parameters and random-effects variances are estimated from the data (e.g., Lindstrom and Bates, 1988; Demidenko, 2013; Bates et al., 2015).

The (unweighted) likelihood of equation (B18) according to Harville (1974, 1976) is

$$\begin{gathered}\mathcal{L}(\boldsymbol{y}|\boldsymbol{f}(\hat{\boldsymbol{p}}), \sigma^2, \boldsymbol{S}) \\ = \frac{\sqrt{|\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J}|}}{\sqrt{(2\pi)^{N-P}|\sigma^2\boldsymbol{S}||\sigma^{-2}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{S}^{-1}\boldsymbol{J}|}} e^{-\frac{1}{2\sigma^2}\hat{\boldsymbol{r}}^{\boldsymbol{T}}\boldsymbol{S}^{-1}\hat{\boldsymbol{r}}},\end{gathered} \tag{B19}$$

in which $\hat{\boldsymbol{r}} = \boldsymbol{y} - \boldsymbol{f}(\hat{\boldsymbol{p}})$ are the residuals with respect to

the estimator $\hat{\boldsymbol{p}}$. The determinant $|\sigma^{-2}\boldsymbol{J}^T\boldsymbol{S}^{-1}\boldsymbol{J}|$ results from the transformation of the likelihood as a function of p to the likelihood as a function of $\hat{\boldsymbol{p}}$. The likelihood in equation (B19) cannot be split into event-wise terms as shown in equation (7), because the covariance matrix $\boldsymbol{S}$ has off-diagonal elements (equation B9). An eigendecomposition of $\boldsymbol{S}$ provides the necessary transformation, such that the likelihood is compliant with equation (3). To ensure that the weight matrix $\boldsymbol{V}$ remains diagonal as well, both matrices are eigendecomposed with

$$\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}\boldsymbol{Q}^T, \tag{B20}$$

in which $\boldsymbol{Q}$ is the eigenvector matrix of $\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}$ with the $i$th column corresponding to the $i$th eigenvalue in $\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}$. The eigenvector matrix $\boldsymbol{Q}$ of a symmetric matrix is orthonormal, that is, it holds $\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$ and with

$$\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^T, \tag{B21}$$

the eigendecomposition is factorized. Furthermore, the eigenvector matrix simultaneously diagonalizes $\boldsymbol{S}^{-1}$ and $\boldsymbol{V}$, that is

$$\boldsymbol{V} = \boldsymbol{Q}\boldsymbol{K}\boldsymbol{Q}^T, \tag{B22}$$

$$\boldsymbol{S}^{-1} = \boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T. \tag{B23}$$

Replacing $\boldsymbol{S}$ with its eigendecomposition gives

$$\begin{aligned}\mathcal{L}(\boldsymbol{Q}^T\boldsymbol{y}|\boldsymbol{Q}^T\boldsymbol{J}\hat{\boldsymbol{p}}, \sigma^2, \boldsymbol{\Lambda}) \\ = \frac{\sqrt{|\boldsymbol{J}^T\boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{J}|}}{\sqrt{(2\pi)^{N-P}|\sigma^2\boldsymbol{\Lambda}||\sigma^{-2}\boldsymbol{J}^T\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\boldsymbol{J}|}} e \\ -\frac{1}{2\sigma^2}\hat{\boldsymbol{r}}^T\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\hat{\boldsymbol{r}}.\end{aligned} \tag{B24}$$

This can be expressed by the product:

$$\begin{aligned}\mathcal{L}(\boldsymbol{Q}^T\boldsymbol{y}|\boldsymbol{Q}^T\boldsymbol{J}\hat{\boldsymbol{p}}, \sigma^2, \boldsymbol{\Lambda}) \\ = \frac{\sqrt{|\boldsymbol{J}^T\boldsymbol{Q}\boldsymbol{Q}^T\boldsymbol{J}|}}{\sqrt{(2\pi)^{N-P}|\sigma^{-2}\boldsymbol{J}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\boldsymbol{J}|}} \\ \times \left[\prod_{i=1}^{N}\frac{1}{\sqrt{\sigma^2\boldsymbol{\Lambda}_i}}\right] e^{-\frac{1}{2\sigma^2}\hat{\boldsymbol{r}}^T\boldsymbol{Q}\boldsymbol{\Lambda}^{-1}\boldsymbol{Q}^T\hat{\boldsymbol{r}}}.\end{aligned} \tag{B25}$$

From the previous expression, the record-wise weight for the weighted likelihood can be inferred by comparing equation (B25) with equation (2) and equation (7), and is given by the diagonal elements of $\boldsymbol{K}$, that is, $\boldsymbol{K}_i$. Thus, the weighted likelihood in terms of its logarithm is expressed as (summarizing constant terms):

$$\ln \mathcal{L}_w\left(\boldsymbol{Q}^T\boldsymbol{y}|\boldsymbol{Q}^T\boldsymbol{f}(\hat{\boldsymbol{p}}), \sigma^2, \boldsymbol{\Lambda}\right) = \text{const.} \tag{B26}$$

$$\begin{aligned}-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{K}_i^2\left(\frac{1}{N}\ln|\sigma^{-2}\boldsymbol{J}^T\boldsymbol{Q}\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}\boldsymbol{Q}^T\boldsymbol{J}| + \ln|\sigma^2\boldsymbol{\Lambda}_i|\right) \\ -\frac{1}{2\sigma^2}\hat{\boldsymbol{r}}^T\boldsymbol{Q}\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}\boldsymbol{Q}^T\hat{\boldsymbol{r}}.\end{aligned}$$

The RML estimator of $\sigma^2$ is found from the derivative of equation (B26):

$$\frac{\partial \ln \mathcal{L}_w}{\partial \sigma^2} = \frac{1}{2\sigma^4}\boldsymbol{s}^T\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}\boldsymbol{s} - \frac{1-\frac{P}{N}}{2\sigma^2}\sum_{i=1}^{N}\boldsymbol{K}_i^2. \tag{B27}$$

Setting the derivative to zero, solving for $\sigma^2$ and substitute the expression using equation (B20):

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{r}}^T\boldsymbol{Q}\boldsymbol{K}\boldsymbol{\Lambda}^{-1}\boldsymbol{K}\boldsymbol{Q}^T\hat{\boldsymbol{r}}}{\left(1-\frac{P}{N}\right)tr(\boldsymbol{K}\boldsymbol{K})}, \tag{B28}$$

$$= \frac{\hat{\boldsymbol{r}}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}}{\left(1-\frac{P}{N}\right)tr(\boldsymbol{V}\boldsymbol{V})}, \tag{B29}$$

the trace is used in equation (B29) and results from the identity

$$tr(\boldsymbol{K}\boldsymbol{K}) = \sum_{i=1}^{N}\boldsymbol{K}_i^2 = tr(\boldsymbol{V}\boldsymbol{V}) = \sum_{i=1}^{N}\boldsymbol{V}_i^2. \tag{B30}$$

If the sum of weights is equal to the number of data and the variance can also be expressed as

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{r}}^T\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}\hat{\boldsymbol{r}}}{N-P}. \tag{B31}$$

This is the unbiased estimator of the variance and appears frequently with the denominator $N-1$, that is, for $P=1$. The estimator corrects the bias for the loss of degree of freedoms, which is equal to the number of parameters $P$ in the ground-motion model parameter estimator $\hat{\boldsymbol{p}}$. For generality, the variance formulation in equation (B29) is used in the following equations.

The RML estimators of the variance factors $\tau_k$ are linked to the variance $\hat{\sigma}^2$ and to the eigenvalue matrix, because it is a function of the variance contrasts, that is, $\boldsymbol{\Lambda}(\tau_1, \tau_2, \ldots \tau_k)$. The derivative of equation (B26) with respect to $\tau_k$ is

$$\frac{\partial \ln \mathcal{L}_w}{\partial \tau_k} = \frac{1}{2}\sum_{i=1}^{N} \boldsymbol{K_i}^2 \left(\frac{1}{N}\frac{\partial}{\partial \tau_k}\ln|\boldsymbol{J^T Q K \Lambda^{-1} K Q^T J}| + \frac{\partial}{\partial \tau_k}\ln|\boldsymbol{\Lambda_i}|\right) - \frac{1}{2\sigma^2}\hat{\boldsymbol{r}}^T \boldsymbol{Q K}\frac{\partial \boldsymbol{\Lambda^{-1}}}{\partial \tau_k}\boldsymbol{K Q^T}\hat{\boldsymbol{r}}. \quad \text{(B32)}$$

The derivative of a determinant is given by Jacobi's formula, which expresses the derivative of the determinant. With the chain rule, the derivatives of the log determinants are

$$\frac{\partial \ln|\boldsymbol{\Lambda_i}|}{\partial \tau_k} = tr\left(\boldsymbol{\Lambda_i^{-1}}\frac{\partial \boldsymbol{\Lambda_i}}{\partial \tau_k}\right), \quad \text{(B33)}$$

and

$$\frac{\partial \ln|\boldsymbol{J^T Q K \Lambda^{-1} K Q^T J}|}{\partial \tau_k} = tr\left((\boldsymbol{J^T Q K \Lambda^{-1} K Q^T J})^{-1}\boldsymbol{J^T Q K}\frac{\partial \boldsymbol{\Lambda^{-1}}}{\partial \tau_k}\boldsymbol{K Q^T J}\right) \quad \text{(B34)}$$

From equations (5), (6), (B9), and (B23), the derivative of $\boldsymbol{\Lambda_i}$ is

$$\frac{\partial \boldsymbol{\Lambda_i}}{\partial \tau_k} = \boldsymbol{Q_i^T B_k B_k^T Q_i}, \quad \text{(B35)}$$

and the derivative of the inverse is

$$\frac{\partial \boldsymbol{\Lambda^{-1}}}{\partial \tau_k} = -\boldsymbol{\Lambda^{-1}}\frac{\partial \boldsymbol{\Lambda}}{\partial \tau_k}\boldsymbol{\Lambda^{-1}} = -\boldsymbol{\Lambda^{-1} Q^T B_k B_k^T Q \Lambda^{-1}}. \quad \text{(B36)}$$

Because $\boldsymbol{\Lambda_i}$ and $\mathbf{K_i}$ are of size $1 \times 1$ the sum of weights in equation (B32) and the trace of the derivative of the determinant in equation (B35) can be combined

$$\sum_{i=1}^{N} \boldsymbol{K_i^2}\frac{\partial}{\partial \tau_k}\ln|\boldsymbol{\Lambda_i}| = \sum_{i=1}^{N} \boldsymbol{K_i}\,\boldsymbol{\Lambda_i^{-1} Q_i^T B_k B_k^T Q_i K_i} = tr\left(\mathbf{K}\boldsymbol{\Lambda^{-1} Q^T B_k B_k^T Q K}\right). \quad \text{(B37)}$$

With the diagonalized matrices replaced according to equation (B20) and the fact that the trace is invariant under circular permutation, the eigenvector matrix $\boldsymbol{Q}$ disappears

$$\begin{aligned} &tr\left(\mathbf{K}\boldsymbol{\Lambda^{-1} Q^T B_k B_k^T Q K}\right) \\ &= tr\left(\boldsymbol{Q^T V S^{-1} B_k B_k^T V Q}\right) \\ &= tr\left(\boldsymbol{Q Q^T V S^{-1} B_k B_k^T V}\right) \\ &= tr\left(\boldsymbol{V S^{-1} B_k B_k^T V}\right). \end{aligned} \quad \text{(B38)}$$

Thus, the derivative can be expressed without the eigendecompositions by replacing the remaining diagonalized matrices in equation (B32) according to equation (B20):

$$\begin{aligned} \frac{\partial \ln \mathcal{L}_w}{\partial \tau_k} &= \frac{1}{2}\Big[-tr\left(\boldsymbol{V S^{-1} B_k B_k^T V}\right) \\ &+ tr\left((\boldsymbol{J^T S^{-1} V J})^{-1}\boldsymbol{J^T V S^{-1} B_k B_k^T S^{-1} V J}\right)\frac{tr(\boldsymbol{VV})}{N} \\ &+ \frac{1}{\hat{\sigma}^2}\hat{\boldsymbol{r}}^T \boldsymbol{V S^{-1} B_k B_k^T S^{-1} V}\hat{\boldsymbol{r}}\Big]. \end{aligned} \quad \text{(B39)}$$

The second term is closely related to the fixed model estimator $\hat{\boldsymbol{p}}$ (equation B13).

The derivative with respect to the variance factors has no analytic solution for maximization and must be maximized numerically, for example, by gradient ascent:

$$\tau^{(i+1)} = \tau^{(i)} + \gamma \nabla \mathcal{L}_w, \quad \text{(B40)}$$

in which $\tau$ is the vector of variance factors, that is, with equation (5) $\tau = (\tau_1, \tau_2, \ldots \tau_k)^T$. The factor $\gamma$ is chosen such to guarantee convergence and is updated at each iteration (Barzilai and Borwein, 1988):

$$\gamma = \frac{\left(\tau^{(i)} - \tau^{(i-1)}\right)^T \left(\nabla \mathcal{L}_w(\tau^{(i)}) - \nabla \mathcal{L}_w(\tau^{(i-1)})\right)}{\|\nabla \mathcal{L}_w(\tau^{(i)}) - \nabla \mathcal{L}_w(\tau^{(i-1)})\|^2}. \quad \text{(B41)}$$

The algorithm starts from some initial values and is repeated until sufficient convergence of the parameters is reached. At each iteration, $\hat{\sigma}^2$ and $\hat{\boldsymbol{p}}$ are updated before the factors $\tau$ are updated. Finally, the ML estimates of the variances are as follows:

$$\hat{\boldsymbol{v}} = \hat{\sigma}^2 \hat{\boldsymbol{\tau}} \quad \text{(B42)}$$

## APPENDIX C

### *Data weighting matrices*

In equation (B1), we introduced two weight matrices: one for the random effects, $\boldsymbol{V_q}$, and one for the residuals, $\boldsymbol{V}$. The two matrices are related via their entries. First, we consider the random-effects weights: we can therefore split the random-effects weight matrix into $K$ submatrices of different sizes, each for one type of random effect:

$$\boldsymbol{V_q} = \begin{pmatrix} \boldsymbol{V_1} & 0 & 0 & \cdots & 0 \\ 0 & \boldsymbol{V_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boldsymbol{V_K} \end{pmatrix} \quad \text{(C1)}$$

The random effects used in ground-motion modeling usually include - but are not restricted - to random station

effects and random event effects. In this case, the weight matrix looks like

$$\boldsymbol{V_q} = \begin{pmatrix} \boldsymbol{V_E} & 0 \\ 0 & \boldsymbol{V_S} \end{pmatrix} \tag{C2}$$

in which $\boldsymbol{V_S}$ is the weight matrix for the station random effects and $\boldsymbol{V_E}$ for is the weight matrix for the event random effects. Because the data weighting introduced here is weighting the events only, we set $\boldsymbol{V_S} = \boldsymbol{I}$, that is, all stations are unweighted. The entries of $\boldsymbol{V_E}$ are the data weights for each event: If Angular Classification with Expectation–maximization is used for event weighting, then each entry on the diagonal is related to the probability given in equation (30).

The relation of $\boldsymbol{V_E}$ to the residual weight matrix $\boldsymbol{V}$ can be illustrated by the following expressions:

$$\boldsymbol{V_E} = \begin{pmatrix} v_1 & 0 & 0 & \cdots & 0 \\ 0 & v_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & v_M \end{pmatrix}, \tag{C3}$$

$$\boldsymbol{V} = \begin{pmatrix} v_1\boldsymbol{I_1} & 0 & 0 & \cdots & 0 \\ 0 & v_2\boldsymbol{I_2} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & v_M\boldsymbol{I_M} \end{pmatrix}, \tag{C4}$$

in which $v_i$ is the square root of the $i$th event weight and $\boldsymbol{I_i}$ is an identity matrix with the size of the number of records of the $i$th event. This formulation assumes that records are sorted by events, however, any permutation of the data is possible and the ordered representation is used for illustrative purposes only.

## APPENDIX D

### *Fisher information matrix*

Here the entries of the Fisher information matrix in equation (19) are given. The matrix has six independent derivatives (the matrix is symmetric and the off-diagonal elements are related by their transpose). The residual vector is given by

$$\hat{\boldsymbol{r}} = \boldsymbol{y} - \boldsymbol{f}(\hat{\boldsymbol{p}}) \approx \Delta\boldsymbol{y} - \boldsymbol{J}\Delta\boldsymbol{p}. \tag{D1}$$

All derivatives of the log likelihood are related to the weighted precision matrix $\boldsymbol{P}$ and/or its derivatives:

$$\boldsymbol{P} = \boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{V}, \tag{D2}$$

$$\frac{\partial \boldsymbol{P}}{\partial \hat{\tau}_k} = -\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{B_k}\boldsymbol{B_k^T}\boldsymbol{S}^{-1}\boldsymbol{V}, \tag{D3}$$

$$\frac{\partial^2 \boldsymbol{P}}{\partial \hat{\tau}_k \hat{\tau}_l} = 2\boldsymbol{V}\boldsymbol{S}^{-1}\boldsymbol{B_l}\boldsymbol{B_l^T}\boldsymbol{S}^{-1}\boldsymbol{B_k}\boldsymbol{B_k^T}\boldsymbol{S}^{-1}\boldsymbol{V}. \tag{D4}$$

For the variance factors, $\hat{\boldsymbol{\tau}}$, follows from equation (B42):

$$\hat{\boldsymbol{\tau}} = \frac{1}{\hat{\sigma}^2}\hat{\boldsymbol{v}}. \tag{D5}$$

With the chain rule, the derivatives can be represented in terms of the random-effects variance vector $\hat{\boldsymbol{v}}$:

$$\frac{\partial \boldsymbol{P}}{\partial \hat{v}_k} = \frac{\partial \boldsymbol{P}}{\partial \hat{\tau}_k}\frac{\partial \hat{\tau}_k}{\partial \hat{v}_k}, \tag{D6}$$

$$= \frac{1}{\hat{\sigma}^2}\frac{\partial \boldsymbol{P}}{\partial \hat{\tau}_k}, \tag{D7}$$

$$\frac{\partial^2 \boldsymbol{P}}{\partial \hat{v}_k \partial \hat{v}_l} = \frac{\partial}{\partial \hat{\tau}_l}\left(\frac{\partial \boldsymbol{P}}{\partial \hat{\tau}_k}\frac{\partial \hat{\tau}_k}{\partial \hat{v}_k}\right)\frac{\partial \hat{\tau}_l}{\partial \hat{v}_l}, \tag{D8}$$

$$= \left(\frac{\partial^2 \boldsymbol{P}}{\partial \hat{\tau}_k \partial \hat{\tau}_l}\frac{\partial \hat{\tau}_k}{\partial \hat{v}_k} + \frac{\partial \boldsymbol{P}}{\partial \hat{\tau}_k}\underbrace{\frac{\partial^2 \hat{\tau}_k}{\partial \hat{v}_k \hat{v}_l}}_{=0}\right)\frac{\partial \hat{\tau}_l}{\partial \hat{v}_l}, \tag{D9}$$

$$= \frac{1}{\hat{\sigma}^4}\frac{\partial^2 \boldsymbol{P}}{\partial \hat{\tau}_k \partial \hat{\tau}_l}, \tag{D10}$$

and the second derivatives of the log likelihood are

$$\frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{\boldsymbol{p}} \partial \hat{\boldsymbol{p}}} = \frac{1}{\hat{\sigma}^2}\boldsymbol{J}^T\boldsymbol{P}\boldsymbol{J}, \tag{D11}$$

$$\frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{\boldsymbol{p}} \partial \hat{\sigma}^2} = \frac{1}{\hat{\sigma}^4}\boldsymbol{J}^T\boldsymbol{P}\hat{\boldsymbol{r}}, \tag{D12}$$

$$\frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{\boldsymbol{p}} \partial \hat{v}_k} = \frac{1}{\hat{\sigma}^2}\boldsymbol{J}^T\frac{\partial \boldsymbol{P}}{\partial \hat{v}_k}\hat{\boldsymbol{r}}, \tag{D13}$$

$$\frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{\sigma}^2 \partial \hat{\sigma}^2} = \frac{tr(\boldsymbol{V}\boldsymbol{V})\left(1 - \frac{P}{N}\right)}{2\hat{\sigma}^4} - \frac{1}{\hat{\sigma}^6}\hat{\boldsymbol{r}}^T\boldsymbol{P}\hat{\boldsymbol{r}}, \tag{D14}$$

$$\frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{\sigma}^2 \partial \hat{v}_k} = \frac{1}{2\hat{\sigma}^4}\hat{\boldsymbol{r}}^T\frac{\partial \boldsymbol{P}}{\partial \hat{v}_k}\hat{\boldsymbol{r}}, \tag{D15}$$

$$\begin{aligned} \frac{\partial^2 \ln \mathcal{L}_w}{\partial \hat{v}_k \partial \hat{v}_l} &= \frac{1}{4}tr\left(\frac{\partial \boldsymbol{P}}{\partial \hat{v}_k \partial \hat{v}_l}\boldsymbol{S}\right) - \frac{1}{2\hat{\sigma}^2}\hat{\boldsymbol{r}}^T\frac{\partial^2 \boldsymbol{P}}{\partial \hat{v}_k \hat{v}_l}\hat{\boldsymbol{r}} \\ &\quad - \frac{tr(\boldsymbol{V}\boldsymbol{V})}{2N}tr\left(\boldsymbol{H}\frac{\partial \boldsymbol{P}}{\partial \hat{v}_l}\boldsymbol{H}\frac{\partial \boldsymbol{P}}{\partial \hat{v}_k} + \boldsymbol{H}\frac{\partial^2 \boldsymbol{P}}{\partial \hat{v}_k \hat{v}_l}\right), \end{aligned} \tag{D16}$$

$$\text{with } \boldsymbol{H} = \boldsymbol{J}(\boldsymbol{J}^T\boldsymbol{P}\boldsymbol{J})^{-1}\boldsymbol{J}^T. \tag{D17}$$

For the expected value, it holds:

$$\mathbb{E}[\boldsymbol{J}\Delta\boldsymbol{p}] = \hat{\boldsymbol{r}} \tag{D18}$$

$$\mathbb{E}[\hat{\boldsymbol{r}}^{\boldsymbol{T}}\boldsymbol{\Omega}\boldsymbol{r}] = \hat{\sigma}\ tr(\boldsymbol{\Omega}\boldsymbol{S}) \tag{D19}$$

for some nonstochastic matrix $\boldsymbol{\Omega}$. From and equations (D1) and (D18) follows immediately for equations (D12) and (D13) that their expected values vanish:

$$\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{p}}\partial\hat{\sigma}^2}\right] = 0 \tag{D20}$$

$$\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\boldsymbol{p}}\partial\hat{v}_k}\right] = 0 \tag{D21}$$

that is, the model parameters are independent from the variances. With equation (D19), the expected value of the "normal term" reduces to

$$\mathbb{E}[\hat{\boldsymbol{r}}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{r}] = \hat{\sigma}^2\ tr(\boldsymbol{P}\boldsymbol{S}) = \hat{\sigma}^2\ tr(\boldsymbol{V}\boldsymbol{V}). \tag{D22}$$

Let $\boldsymbol{T}$ be the $1 \times K$ vector associated with the expected value of the derivatives of equation (D15):

$$-\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{\sigma}^2\partial\hat{v}_k}\right] = -\frac{1}{2\hat{\sigma}^2} tr\left(\frac{\partial^2\boldsymbol{P}}{\partial\hat{v}_k}\boldsymbol{S}\right), \tag{D23}$$

$$= -\frac{tr(\boldsymbol{V}\boldsymbol{V})}{2\hat{\sigma}^4} T_k, \tag{D24}$$

$$T_k = \frac{\hat{\sigma}^2}{tr(\boldsymbol{V}\boldsymbol{V})} tr\left(\frac{\partial\boldsymbol{P}}{\partial\hat{v}_k}\right), \tag{D25}$$

and $\boldsymbol{U}$ is the $K \times K$ matrix associated with the expected value of the derivatives of equation (D16):

$$U_{kl} = -\mathbb{E}\left[\frac{\partial^2 \ln \mathcal{L}_w}{\partial\hat{v}_k\partial\hat{v}_l}\right], \tag{D26}$$

$$= \frac{1}{4} tr\left(\frac{\partial^2\boldsymbol{P}}{\partial\hat{v}_k\partial\hat{v}_l}\boldsymbol{S}\right). \tag{D27}$$

The Fisher information matrix is

$$\boldsymbol{I}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}}) = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{J} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{tr(\boldsymbol{V}\boldsymbol{V})}{2\hat{\sigma}^4} & -\frac{tr(\boldsymbol{V}\boldsymbol{V})}{2\hat{\sigma}^4}\boldsymbol{T} \\ \boldsymbol{0} & -\frac{tr(\boldsymbol{V}\boldsymbol{V})}{2\hat{\sigma}^4}\boldsymbol{T}^{\boldsymbol{T}} & \boldsymbol{U} \end{pmatrix}. \tag{D28}$$

This matrix can be seen as a block diagonal matrix of two blocks (with sizes $1 \times 1$ and $2 \times 2$). The inverse of a block diagonal matrix is again a block diagonal matrix in which each block is inverted separately, allowing a simplified computation of the entire matrix inverse. An arbitrary $2 \times 2$ block matrix has the inverse (e.g., Harville, 1997)

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{G}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{G}^{-1} \\ -\boldsymbol{G}^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \boldsymbol{G}^{-1} \end{pmatrix}^{-1} \tag{D29}$$

in which $\boldsymbol{G} = \boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$ is a nonsingular matrix, and $\boldsymbol{A}$ and $\boldsymbol{D}$ are square matrices. The inverse of the block submatrix of the variances (bold italics block in equation D28) is

$$\begin{pmatrix} \frac{tr(\boldsymbol{V}\boldsymbol{V})}{2\hat{\sigma}^4} & -\frac{1}{\hat{\sigma}^2}\boldsymbol{T} \\ -\frac{1}{\hat{\sigma}^2}\boldsymbol{T}^{\boldsymbol{T}} & \boldsymbol{U} \end{pmatrix}^{-1} = \begin{pmatrix} \frac{2\hat{\sigma}^4}{tr(\boldsymbol{V}\boldsymbol{V})} + \boldsymbol{T}\boldsymbol{M}\boldsymbol{T}^{\boldsymbol{T}} & \boldsymbol{T}\boldsymbol{M} \\ \boldsymbol{M}\boldsymbol{T}^{\boldsymbol{T}} & \boldsymbol{M} \end{pmatrix}, \tag{D30}$$

$$\text{with } \boldsymbol{M} = \left(\boldsymbol{U} - \tfrac{tr(\boldsymbol{V}\boldsymbol{V})}{\hat{\sigma}^4}\boldsymbol{T}^{\boldsymbol{T}}\boldsymbol{T}\right)^{-1}. \tag{D31}$$

The inverse of the Fisher information - which is asymptotically equivalent to the covariance matrix as $N \to \infty$ - is stated as

$$\boldsymbol{I}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}})^{-1} \sim \text{cov}(\hat{\boldsymbol{p}}, \hat{\sigma}^2, \hat{\boldsymbol{v}}) = \begin{pmatrix} \hat{\sigma}^2(\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{J})^{-1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{2\hat{\sigma}^4}{tr(\boldsymbol{V}\boldsymbol{V})} + TMT^T & \boldsymbol{T}\boldsymbol{M} \\ \boldsymbol{0} & MT^T & \boldsymbol{M} \end{pmatrix}. \tag{D32}$$

The upper left block matrix is equivalent to the model parameter covariance matrix; only the covariance matrix of the random-effects variances is asymptotically equivalent to previous expression.