An Advanced Estimation Algorithm for Ground-Motion Models with Spatial Correlation

Deyu Ming[1,*], Chen Huang[2], Gareth W. Peters[3], and Carmine Galasso[2]

[1]Department of Statistical Science, University College London, London, UK.

[2]Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK.

[3]Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Edinburgh, UK.

*Corresponding author: Deyu Ming

Department of Statistical Science

University College London

London, England, UK WC1E 6BT

Email: deyu.ming.16@ucl.ac.uk

Abstract

Ground-motion prediction equations (GMPEs), also called ground-motion models and attenuation relations, are empirical models widely used in Probabilistic Seismic Hazard Analysis (PSHA). They estimate the conditional distribution of ground shaking at a site given an earthquake of a certain magnitude occurring at a nearby location. In the past decade, the increasing interest in assessing earthquake risk and resilience of spatially distributed portfolios of buildings and infrastructures has motivated the modeling of ground-motion spatial correlation. This introduces further challenges for researchers to develop statistically rigorous and computationally efficient algorithms to perform ground-motion model estimation with spatial correlation. To this goal, we introduce a one-stage ground-motion estimation algorithm, called the Scoring estimation approach, to fit ground-motion models with spatial correlation. The Scoring estimation approach is introduced theoretically and numerically, and it is proven to have desirable properties on convergence and computation. It is a statistically robust method, producing consistent and statistically efficient estimators of inter- and intra-event variances and parameters in spatial correlation functions. The performance of the Scoring estimation approach is assessed through a comparison with the multi-stage algorithm proposed by Jayaram and Baker (2010) in a simulation-based application. The results of the simulation study show that the proposed Scoring estimation approach presents comparable or higher accuracy in estimating ground-motion model parameters, especially when the spatial correlation becomes smoother. The simulation study also shows that ground-motion models with spatial correlation built via the Scoring estimation approach can be used for reliable ground shaking intensity predictions and, ultimately, as an accurate input for the earthquake risk assessment of spatially distributed systems. The performance of the Scoring estimation approach is further discussed under the ignorance of spatial correlation and we find that neglecting spatial correlation in ground-motion models may result in overestimation of inter-event variance, underestimation of intra-event variance, and thus inaccurate predictions.

Introduction

Ground-motion models, also known as ground-motion prediction equations (GMPEs) and attenuation relationships, are empirical models widely used in probabilistic seismic hazard analysis (PSHA), to predict ground-motion intensity measures (IMs) occurring at sites due to a nearby earthquake of a certain magnitude. Ground-motion models require robust estimation techniques. The accuracy of the estimated ground-motion models is important for assessing earthquake risk and resilience of engineered systems.

Initial ground-motion models were formulated as fixed-effects models without considering variations across different events. To further characterize the aleatory variability in ground shaking intensities, the uncertainties are separated into the inter-event and the intra-event components, where the inter-event components were introduced as random effects to the ground-motion model (Brillinger and Preisler, 1984a). The modern ground-motion model is thus constructed as a mixed-effects model in the following form,

$$Y_{ij} = f(\mathbf{X}_{ij},\, \mathbf{b}) + \eta_i + \varepsilon_{ij}\,, \quad i = 1, \ldots, N,\ j = 1, \ldots, n_i\,, \tag{1}$$

where $Y_{ij} = \log \mathrm{IM}_{ij}$ is the logarithm of the IM of interest (e.g., peak ground acceleration (PGA), peak ground velocity (PGV), spectral ordinates) at site j during earthquake i; $f(\mathbf{X}_{ij},\, \mathbf{b})$ is the ground-motion prediction function of $\mathbf{b}$, a vector of unknown parameters, and $\mathbf{X}_{ij}$, a vector of predictors (e.g., magnitude, source-to-site distance, soil type at site) for site j during event i; η_i and ε_{ij} are the inter-event error and the intra-event error, respectively; N is the total number of earthquakes and n_i is the number of recording sites during the i-th earthquake.

Traditionally, the ground-motion model in equation (1) is treated without spatial correlation by assuming the intra-event errors are spatially independent of each other, and is primarily estimated by algorithms proposed by Abrahamson and Youngs (1992) and Joyner and Boore (1993). However, it is well known that the intra-event errors are spatially correlated because of the common source and wave traveling paths, and similar site conditions (Goda

and Hong, 2008; Jayaram and Baker, 2009). Hong *et al.* (2009) investigated the effects of spatial correlation on ground-motion model estimation and observed that the estimates of variances for inter-event and intra-event errors change significantly when spatial correlation is considered. Jayaram and Baker (2010) confirmed the results and demonstrated that the changes in variances of inter-event and intra-event errors have important implications for the seismic risk assessment of spatially distributed systems. Hence, we argue that it is crucial to develop an efficient and accurate estimation method for ground-motion models with spatial correlation.

Indeed, the consideration of spatial correlation complicates the estimation of ground-motion models. In particular, Hong *et al.* (2009) illustrated how to incorporate the spatial correlation into a ground-motion model and performed estimation using the method under the framework proposed by Joyner and Boore (1993). However, the estimation method proposed by Hong *et al.* (2009) uses the linearization of the ground-motion prediction function, an inefficient technique that can add bias because of model misspecifications and was subsequently criticized by Draper and Smith (2014) for its slow convergence, wide oscillation and possibility of divergence. Based on the framework of Abrahamson and Youngs (1992), Jayaram and Baker (2010) introduced a multi-stage algorithm to account for the spatial correlation by adopting the idea of the classical geostatistical analysis (Zimmerman and Stein, 2010). However, this algorithm (reviewed in Section Jayaram and Baker's Multi-stage Algorithm) may not be statistically optimal and can result in inefficient parameter estimation and poor conclusions on model structure and variable selection, which in turn affects predictions of spatially distributed ground-motion intensities. This will ultimately affect the reliability of seismic risk assessment and loss estimation for portfolios of spatially distributed buildings and lifelines.

In addition to the bespoke algorithms mentioned earlier, there is also a more generic existing computer package, namely nlme in R, available to fit ground-motion models with or without spatial correlation. However, this package is based on the method proposed by Lindstrom and Bates (1990) for mixed-effects models with nonlinear random effects and

thus introduces excessive computational expenses during its implementation. Besides, the package may experience numerical instabilities when spatial correlation is considered even though the estimation is performed on a small number of events. Jayaram and Baker (2010) also reported the numerical instability of the package. We argue that the failure of the package is due to the numerical issues that can arise when working with the Hessian matrices during its implementation of the Newton-Raphson algorithm. Furthermore, the package only considers limited types of spatial correlation structures (Pinheiro and Bates, 2000).

In this article, a one-stage estimation method based on the method of Scoring (Fisher, 1925) under the maximum likelihood estimation framework is developed as a specialized alternative procedure for fitting ground-motion models with spatial correlation. Although the method of Scoring applied to the maximum likelihood estimation is a well-developed statistical technique, to the best of authors' knowledge, this is the first attempt to use it in the ground-motion model estimations, particularly when the spatial correlation is considered. This article first illustrates in detail the specifications and assumptions of the considered ground-motion model with spatial correlation. The multi-stage algorithm introduced by Jayaram and Baker (2010) is then reviewed, and its limitations are highlighted. The new method, referred to as the Scoring estimation approach, is then formally introduced. Numerical considerations for the Scoring estimation approach are also discussed. A simulation study is followed to measure the performances of the Scoring estimation approach by comparing against those of the multi-stage algorithm. Finally, we discuss the performance of the Scoring estimation approach when spatial correlation structure is neglected in the ground-motion model.

The Ground-Motion Model

The ground-motion model in this article is expressed as the vector form of equation (1):

$$\mathbf{Y}_i = \mathbf{f}(\mathbf{X}_i,\, \mathbf{b}) + \boldsymbol{\eta}_i + \boldsymbol{\varepsilon}_i\,, \quad i = 1, \ldots, N\,, \tag{2}$$

in which

- $\mathbf{Y}_i = \log \mathbf{IM}_i = (\log \mathrm{IM}_{i1}, \ldots, \log \mathrm{IM}_{ij}, \ldots, \log \mathrm{IM}_{in_i})^\top$ is an $n_i \times 1$ vector of logarithmic IMs of interest at all sites $j \in \{1, \ldots, n_i\}$ during earthquake i;
- $\mathbf{f}(\mathbf{X}_i, \mathbf{b}) = (f(\mathbf{X}_{i1}, \mathbf{b}), \ldots, f(\mathbf{X}_{in_i}, \mathbf{b}))^\top$ is an $n_i \times 1$ vector of ground-motion prediction functions $f(\mathbf{X}_{ij}, \mathbf{b})$ at all sites $j \in \{1, \ldots, n_i\}$ during earthquake i;
- $\mathbf{X}_{ij}$ represents a vector of predictors (e.g., magnitude, source-to-site distance, soil type at site) for site j during earthquake i;
- $\mathbf{b} \in \mathbb{R}^{p_1}$ is a vector of unknown model parameters;
- $\boldsymbol{\eta}_i = \eta_i \mathbf{1}_{n_i}$ for all $i \in \{1, \ldots, N\}$ and $(\eta_i)_{i=1,\ldots,N}$ are independent and identically distributed inter-event errors with $\mathbb{E}(\eta_i) = 0$ and $\mathrm{var}(\eta_i) = \tau^2$ for all $i \in \{1, \ldots, N\}$, in which $\mathbf{1}_{n_i}$ is an $n_i \times 1$ vector of ones;
- $(\boldsymbol{\varepsilon}_i)_{i=1,\ldots,N}$ are independent intra-event error vectors of size $n_i \times 1$ with $\mathbb{E}(\boldsymbol{\varepsilon}_i) = \mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon}_i) = \sigma^2 \boldsymbol{\Omega}_i(\boldsymbol{\omega})$, in which $\boldsymbol{\Omega}_i(\boldsymbol{\omega})$ is the correlation matrix corresponding to earthquake i with $\boldsymbol{\omega}$, a vector of unknown parameters;
- $(\eta_i)_{i=1,\ldots,N}$ and $(\boldsymbol{\varepsilon}_i)_{i=1,\ldots,N}$ are mutually independent.

To take the spatial correlation into account, the jj'-th entry $\boldsymbol{\Omega}_{i,jj'}(\boldsymbol{\omega})$ of $\boldsymbol{\Omega}_i(\boldsymbol{\omega})$ is specified as

$$\boldsymbol{\Omega}_{i,jj'}(\boldsymbol{\omega}) = k(\mathbf{s}_{ij}, \mathbf{s}_{ij'}) \tag{3}$$

for all $i \in \{1, \ldots, N\}$ and $j, j' \in \{1, \ldots, n_i\}$, in which $k(\mathbf{s}_{ij}, \mathbf{s}_{ij'})$ gives the correlation $\rho(\varepsilon_{ij}, \varepsilon_{ij'})$ between ε_{ij} and $\varepsilon_{ij'}$ at locations $\mathbf{s}_{ij}$ and $\mathbf{s}_{ij'}$ of sites j and j' during earthquake i:

$$k(\mathbf{s}_{ij}, \mathbf{s}_{ij'}) = \rho(\varepsilon_{ij}, \varepsilon_{ij'}). \tag{4}$$

There are many choices for $k(\mathbf{s}_{ij}, \mathbf{s}_{ij'})$ (Rasmussen and Williams, 2006). For independent intra-event errors (i.e., no spatial correlation is incorporated),

$$k(\mathbf{s}_{ij}, \mathbf{s}_{ij'}) = 0 \tag{5}$$

for all sites $j \neq j'$ during earthquake i. For stationary (i.e., invariant to translations) and isotropic (i.e., invariant to rigid motions) process of intra-event errors, the correlation $\rho(\varepsilon_{ij}, \varepsilon_{ij'})$ only depends on $d_{i,jj'} = \|\mathbf{s}_{ij} - \mathbf{s}_{ij'}\|_2$, the Euclidean distance between sites j and j' during earthquake i, such that

$$k(\mathbf{s}_{ij}, \mathbf{s}_{ij'}) = k(d_{i,jj'}) \,. \tag{6}$$

Examples of this class of correlation functions $k(\cdot)$ include (see Rasmussen and Williams (2006) or Zimmerman and Stein (2010)):

- **Matérn:**

$$k(d) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} d}{h} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu} d}{h} \right) \tag{7}$$

  with positive parameters ν and h, in which $\Gamma(\cdot)$ is the gamma function and $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind. The Matérn correlation function can be simplified to exponential and squared exponential correlation functions by setting $\nu = 1/2$ and $\nu \to \infty$, respectively;

- **Exponential:**

$$k(d) = \exp\left(-\frac{d}{h} \right) \tag{8}$$

  with a positive range parameter h, which indicates the distance at which the correlation is around 0.37. It is worth noting that the exponential correlation function (8) has a slightly different form from the one used by Jayaram and Baker (2010), Esposito and Iervolino (2011, 2012), and others. These studies defined the exponential correlation function with the following form:

$$k(d) = \exp\left(-\frac{3d}{h} \right) , \tag{9}$$

  where h now indicates the distance at which the correlation is approximately 0.05. In fact, the exponential correlation function can be defined with a more general form

$$k(d) = \exp\left(-\frac{cd}{h} \right) , \tag{10}$$

in which c is a given positive constant and h indicates the distance at which the correlation is $\exp(-c)$. One should note that the choice of c only affects the interpretation of h in terms of the correlation at $d = h$ but has no influences on the spatial information implied by the ultimately estimated correlation structure;

- **Squared Exponential:**

$$k(d) = \exp\left(-\frac{d^2}{2h^2}\right) \tag{11}$$

with range parameter h defining the characteristic length-scale. This type of correlation function is sometimes called Gaussian in references such as Jayaram and Baker (2009).

Examples of other types of correlation functions (including non-stationary or anisotropic ones) are illustrated in Rasmussen and Williams (2006).

In the rest of this article, we denote $\boldsymbol{\alpha} = (\mathbf{b}^\top, \boldsymbol{\theta}^\top)^\top \in \mathbb{R}^p$ as the complete vector of model parameters, in which $\boldsymbol{\theta} = (\tau^2, \sigma^2, \boldsymbol{\omega}^\top)^\top \in \mathbb{R}^{p_2}$ with $\boldsymbol{\omega}$ being a vector of the parameters (e.g., h in exponential and squared exponential correlation functions) contained in the correlation function $k(\mathbf{s}_{ij}, \mathbf{s}_{ij'})$.

Jayaram and Baker's Multi-stage Algorithm

In this section, we review the multi-stage algorithm proposed by Jayaram and Baker (2010) to estimate ground-motion models with spatial correlation. This algorithm will serve as the current best benchmark procedure for our new proposed method, so it is important to discuss its properties and compare its approach to our proposed Scoring estimation approach. The algorithm consists of three stages (see Figure 1) and follows the framework of the classical geostatistical method (Zimmerman and Stein, 2010). In the preliminary stage, the algorithm provisionally estimates the model parameters ignoring the spatial correlation. In the second stage, the residuals from the estimated provisional ground-motion prediction function are used to estimate the parameters in the correlation function by fitting a parametric semivariogram model to the empirical semivariogram. In the final stage, the preliminary estimates of model

parameters from the first stage are updated given the spatial correlation structure fitted in the second stage. We proceed to outline each stage in detail below.

[Figure 1 about here.]

The preliminary stage

The preliminary stage of the algorithm aims at estimating ground-motion models requiring no knowledge about the spatial correlation. Because the spatial correlation is being ignored at this stage, authors such as Goda and Hong (2008), Goda and Atkinson (2009, 2010), and Sokolov *et al.* (2010) adopted estimation methods introduced by Abrahamson and Youngs (1992) or Joyner and Boore (1993) to obtain the estimates of unknown model parameters $\mathbf{b}$, τ^2, and σ^2. Other authors such as Wang and Takada (2005), Jayaram and Baker (2009), and Esposito and Iervolino (2011, 2012) obtained the estimates of $\mathbf{b}$, τ^2, and σ^2 by simply adopting existing ground-motion models developed without consideration of spatial correlation.

The spatial correlation stage

The spatial correlation stage is designed to estimate $\boldsymbol{\omega}$, a vector of unknown parameters in the correlation function, from the total residuals

$$e_{ij}^{(t)} = Y_{ij} - f(\mathbf{X}_{ij}, \widehat{\mathbf{b}}), \tag{12}$$

in which $\widehat{\mathbf{b}}$ is the estimate of $\mathbf{b}$ given by the preliminary stage. Because the total error term

$$\varepsilon_{ij}^{(t)} = \varepsilon_{ij} + \eta_i \tag{13}$$

consists of intra-event errors ε_{ij} and inter-event errors η_i, the total residuals can be represented by intra-event residuals $\widehat{\varepsilon}_{ij}$ and inter-event residuals $\widehat{\eta}_i$:

$$e_{ij}^{(t)} = \widehat{\varepsilon}_{ij} + \widehat{\eta}_i . \tag{14}$$

Then one defines a random process of the standardized intra-event errors

$$\widetilde{\boldsymbol{\varepsilon}} = \frac{\boldsymbol{\varepsilon}}{\sigma} \tag{15}$$

with $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, \ldots, \boldsymbol{\varepsilon}_N^\top)^\top$ and $\widetilde{\boldsymbol{\varepsilon}} = (\widetilde{\boldsymbol{\varepsilon}}_1^\top, \ldots, \widetilde{\boldsymbol{\varepsilon}}_N^\top)^\top$. Assuming that the process of intra-event errors is second-order stationary and isotropic, Jayaram and Baker (2009) constructed for each earthquake i the empirical semivariogram $\widehat{\gamma}_i(d)$, a moment-based estimator defined by Cressie (1993), of $\widetilde{\boldsymbol{\varepsilon}}_i$ from the scaled intra-event residuals:

$$\widehat{\widetilde{\varepsilon}}_{ij} = \frac{\widehat{\varepsilon}_{ij}}{\widehat{\sigma}}\,. \tag{16}$$

The empirical semivariogram $\widehat{\gamma}_i(d)$ is calculated by

$$\begin{aligned}
\widehat{\gamma}_i(d) &= \frac{1}{2|N_{i,\delta}(d)|} \sum_{N_{i,\delta}(d)} \left(\widehat{\widetilde{\varepsilon}}_{ij} - \widehat{\widetilde{\varepsilon}}_{ij'}\right)^2 \\
&= \frac{1}{2|N_{i,\delta}(d)|} \sum_{N_{i,\delta}(d)} \left(\frac{e_{ij}^{(t)} - \widehat{\eta}_i}{\widehat{\sigma}} - \frac{e_{ij'}^{(t)} - \widehat{\eta}_i}{\widehat{\sigma}}\right)^2 \\
&= \frac{1}{2|N_{i,\delta}(d)|} \sum_{N_{i,\delta}(d)} \left(\frac{e_{ij}^{(t)} - e_{ij'}^{(t)}}{\widehat{\sigma}}\right)^2\,,
\end{aligned} \tag{17}$$

in which $N_{i,\delta}(d)$ is a δ-neighborhood set consisting of all site pairs $(j,\, j')$ such that

$$d - \delta < \|\mathbf{s}_{ij} - \mathbf{s}_{ij'}\|_2 < d + \delta \tag{18}$$

during earthquake i and $|N_{i,\delta}(d)|$ is the number of distinct pairs in $N_{i,\delta}(d)$.

Each empirical seimivariogram $\widehat{\gamma}_i(d)$ is then fitted by a common parametric semivariogram model $\gamma(d)$ constructed from a stationary and isotropic correlation function $k(d)$ according to the relationship given by

$$\gamma(d) = 1 - k(d)\,. \tag{19}$$

Equation (19) holds because of the assumed second-order stationarity of the process of the intra-event errors $\boldsymbol{\varepsilon}$ and the corresponding proof is available in the Proof of Equation (19) of the Main Article section of the electronic supplement to this article. One can then obtain the estimate $\widehat{\boldsymbol{\omega}}_i$ of $\boldsymbol{\omega}$ for each earthquake i by fitting $\gamma(d)$ to the sample estimator given by $\widehat{\gamma}_i(d)$ via estimation methods such as least-squares and trial-and-error methods. Jayaram and Baker (2009) then computed the estimates $\widehat{\boldsymbol{\omega}}_{i=1,\ldots,N}$ for spectral accelerations at different structural periods and built linear regression models to obtain the estimate of $\boldsymbol{\omega}$ for a given structural period.

Unlike Jayaram and Baker (2009) who estimated $\boldsymbol{\omega}$ by constructing empirical semivariogram for each earthquake i, Esposito and Iervolino (2011, 2012) built a pooled empirical semivariogram $\widehat{\gamma}(d)$ given by

$$\widehat{\gamma}(d) = \frac{1}{2|N_\delta(d)|} \sum_{N_\delta(d)} \left(\frac{e_{ij}^{(t)} - e_{ij'}^{(t)}}{\widehat{\sigma}} \right)^2 , \tag{20}$$

in which $N_\delta(d)$ is a δ-neighborhood set consisting of all site pairs $(j,\, j')$ such that

$$d - \delta < \|\mathbf{s}_{ij} - \mathbf{s}_{ij'}\|_2 < d + \delta \tag{21}$$

across all earthquakes $i \in \{1, \ldots, N\}$. The estimate of $\boldsymbol{\omega}$ is then obtained by fitting a parametric semivariogram model $\gamma(d)$ to $\widehat{\gamma}(d)$ via least-squares and trial-and-error methods.

Jayaram and Baker (2009) discussed the method of least squares and the trial-and-error method (i.e., a manual fitting method focusing on fitting the empirical semivariogram at short separation distances d) and suggested that the trial-and-error method is a better choice because of its simplicity and better fit at separation "distances that are of practical interest" (Jayaram and Baker, 2009).

The re-estimation stage

The objective of the re-estimation stage is to update the estimates $\widehat{\mathbf{b}}$, $\widehat{\sigma^2}$ and $\widehat{\tau^2}$ obtained in the preliminary stage by considering the spatial correlation structure established in the spatial correlation stage. Algorithm 1 illustrates the re-estimation procedure proposed by Jayaram and Baker (2010). However, Jayaram and Baker (2010) did not report any convergence properties of the procedure. In the Alternative Construction of the Re-estimation Procedure section of the electronic supplement to this article, we demonstrate that the re-estimation procedure can be alternatively constructed based on the idea of the Expectation-Maximization (EM) algorithm (Lahiri and Ware, 1982; Brillinger and Preisler, 1984b; Lahiri *et al.*, 1987). Therefore, the re-estimation procedure is a non-decreasing algorithm as long as the fixed-effects regression algorithm (step 4 of the Algorithm 1) solves the following generalized least

squares problem with respect to $\mathbf{b}$:

$$\widehat{\mathbf{b}}^{(k+1)} = \arg\min \sum_{i=1}^{N} [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i,\, \mathbf{b}) - \widehat{\eta}_i \mathbf{1}_{n_i}]^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i,\, \mathbf{b}) - \widehat{\eta}_i \mathbf{1}_{n_i}] \,. \tag{22}$$

Algorithm 1 The re-estimation procedure (Jayaram and Baker, 2010)

Input: 1) $\mathbf{Y}_i$, $\mathbf{X}_{ij}$ and $\mathbf{s}_{ij}$ for $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, n_i\}$;

2) Estimate $\widehat{\boldsymbol{\omega}}$ of $\boldsymbol{\omega}$ obtained in the spatial correlation stage.

Output: Updated estimates of $\mathbf{b}$, σ^2 and τ^2.

1: **Initialization:**

1) obtain the initial estimate $\widehat{\mathbf{b}}^{(1)}$ of $\mathbf{b}$ by a fixed-effects regression algorithm setting $\eta_{i=1,\ldots,N} = 0$;

2) obtain the initial estimates $\widehat{\sigma^2}^{(1)}$ and $\widehat{\tau^2}^{(1)}$ by maximizing the log-likelihood function:

$$l\left(\sigma^2,\, \tau^2 \middle| \mathbf{b} = \widehat{\mathbf{b}}^{(1)},\, \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}\right) = -\frac{\sum_{i=1}^{N} n_i}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \ln \left| \tau^2 \mathbf{1}_{n_i \times n_i} + \sigma^2 \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}}) \right|$$
$$- \frac{1}{2} \sum_{i=1}^{N} [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i,\, \widehat{\mathbf{b}}^{(1)})]^\top \left(\tau^2 \mathbf{1}_{n_i \times n_i} + \sigma^2 \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}}) \right)^{-1} [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i,\, \widehat{\mathbf{b}}^{(1)})] \,; \tag{23}$$

2: **repeat**

3: Given $\widehat{\mathbf{b}}^{(k)}$, $\widehat{\sigma^2}^{(k)}$, $\widehat{\tau^2}^{(k)}$ and $\widehat{\boldsymbol{\omega}}$, obtain $\widehat{\eta}_{i=1,\ldots,N}$ from

$$\widehat{\eta}_i = \frac{\frac{1}{\widehat{\sigma^2}^{(k)}} \mathbf{1}_{n_i}^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) \, [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i,\, \widehat{\mathbf{b}}^{(k)})]}{\frac{1}{\widehat{\tau^2}^{(k)}} + \frac{1}{\widehat{\sigma^2}^{(k)}} \mathbf{1}_{n_i}^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) \, \mathbf{1}_{n_i}} \,; \tag{24}$$

4: Given $\widehat{\eta}_{i=1,\ldots,N}$, obtain $\widehat{\mathbf{b}}^{(k+1)}$, the estimate of $\mathbf{b}$ at iteration $k+1$, using a fixed-effects regression algorithm by setting $\eta_i = \widehat{\eta}_i$ for all $i \in \{1, \ldots, N\}$;

5: Given $\widehat{\mathbf{b}}^{(k+1)}$ and $\widehat{\boldsymbol{\omega}}$, obtain $\widehat{\sigma^2}^{(k+1)}$ and $\widehat{\tau^2}^{(k+1)}$ by maximizing the log-likelihood function $l(\sigma^2,\, \tau^2 | \mathbf{b} = \widehat{\mathbf{b}}^{(k+1)},\, \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})$;

6: **until** $l(\sigma^2, \tau^2, \mathbf{b} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})$ is maximized and parameter estimates converge.

Problems of the multi-stage algorithm

Although the multi-stage algorithm is feasible in practice and may be numerically stable by estimating the spatial correlation function in separate steps (i.e., the preliminary and spatial correlation stages), it is not optimal in various aspects from a statistical estimation perspective.

First, the least squares estimator of $\boldsymbol{\omega}$ produced by the first two stages of the algorithm is inconsistent (i.e., $\widehat{\boldsymbol{\omega}}$ does not converge in probability to the true value of $\boldsymbol{\omega}$). Lahiri *et al.* (2002) and Kerby (2016) discussed the conditions for the consistency of the least squares estimator of $\boldsymbol{\omega}$. To have a consistent least squares estimator of $\boldsymbol{\omega}$, we need the empirical semivariogram $\widehat{\gamma}(d)$ to be a consistent estimator of $\gamma(d)$. However, this consistency typically requires very restrictive asymptotic conditions in which "not only the number of locations increases but the distance between them decreases" (Kerby, 2016). Furthermore, Kerby (2016) showed that observation locations must not be heavily clustered (which is the case in reality where the recording sites are indeed clustered, especially at near-fault locations) and the bandwidth δ need to be carefully chosen so that the consistency of the empirical semivariogram $\widehat{\gamma}(d)$ is ensured. In addition, the consistency of the empirical semivariogram $\widehat{\gamma}(d)$ requires the estimators of $\mathbf{b}$ and σ^2 obtained from the preliminary stage to be consistent. However, it can be shown mathematically that although the estimator of $\mathbf{b}$ obtained in the preliminary stage is consistent, the estimator of σ^2 is not. Consequently, the least squares estimator of $\boldsymbol{\omega}$ obtained at the spatial correlation stage is not consistent. Finally, the least squares estimator of $\boldsymbol{\omega}$ can be statistically inefficient (Lahiri *et al.*, 2002), and naively using the formula of asymptotic standard error estimate produced by software packages based on ordinary least squares can cause incorrect confidence interval on $\boldsymbol{\omega}$.

With regard to the trial-and-error method, although it fits the parametric semivariogram model to the empirical semivariograms better than the least squares at short separation distances, Stein (1999) illustrated in a simulation study that this eyeball procedure leads to substantial prediction errors, especially when the spatial correlation structure is misspecified.

Besides, this manual fitting procedure makes it impossible to evaluate the asymptotic properties of the estimator of $\boldsymbol{\omega}$. Therefore, such a heuristic procedure should not become the standard.

Moreover, the first two stages are only capable of estimations of isotropic and stationary correlation structures and inflexible in considering more advanced (e.g., non-stationary) spatial correlation functions.

In addition, the re-estimation procedure maximizes the conditional log-likelihood function $l(\sigma^2, \tau^2, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})$ given the pre-computed estimate $\widehat{\boldsymbol{\omega}}$. Because the least squares estimator of $\boldsymbol{\omega}$ is inconsistent, the resulting estimators of $\mathbf{b}$ (although consistent) are statistically inefficient, and estimators of τ^2 and σ^2 are both inconsistent and statistically inefficient.

Additionally, because the re-estimation procedure can be interpreted via the idea of the EM algorithm, it suffers from the "hopelessly slow linear convergence" (Couvreur, 1997) and is very sensitive to the initial parameter values (Gao and Wang, 2013).

Furthermore, unlike the Scoring estimation approach introduce in Section A One-Stage Algorithm: Scoring Estimation Approach, the multi-stage algorithm does not produce asymptotic standard error estimates of model parameters as by-products. As a consequence, the multi-stage algorithm requires extra computations and complexities in its implementation when asymptotic standard error estimates are desired. Finally, it is worth noting that the equations provided by Jayaram and Baker (2010) for asymptotic standard error estimates of τ^2 and σ^2 are only valid when estimators of τ^2 and σ^2 are asymptotically independent. However, $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ are not asymptotically independent, thus, their asymptotic variance estimates should be obtained by taking the first and the second diagonal entry of

$$2\left[\begin{array}{cc} \mathrm{tr}\left\{\left(\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\tau^2)}\right)^2\right\} & \mathrm{tr}\left\{\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\tau^2)}\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\sigma^2)}\right\} \\ \mathrm{tr}\left\{\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\sigma^2)}\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\tau^2)}\right\} & \mathrm{tr}\left\{\left(\mathbf{C}(\boldsymbol{\theta})^{-1}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial(\sigma^2)}\right)^2\right\} \end{array}\right]^{-1}_{\boldsymbol{\theta}=(\widehat{\tau^2},\widehat{\sigma^2},\widehat{\boldsymbol{\omega}}^\top)^\top}, \quad (25)$$

in which

$$\mathbf{C}(\boldsymbol{\theta}) = \begin{bmatrix} \tau^2\mathbf{1}_{n_1\times n_1} + \sigma^2\boldsymbol{\Omega}_1(\boldsymbol{\omega}) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \tau^2\mathbf{1}_{n_2\times n_2} + \sigma^2\boldsymbol{\Omega}_2(\boldsymbol{\omega}) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \tau^2\mathbf{1}_{n_N\times n_N} + \sigma^2\boldsymbol{\Omega}_N(\boldsymbol{\omega}) \end{bmatrix} . \quad (26)$$

However, even matrix (25) may not give the correct asymptotic standard error estimates of $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ because the least squares estimator of $\boldsymbol{\omega}$ is inconsistent and asymptotic variances of $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ depend on that of $\widehat{\boldsymbol{\omega}}$.

To avoid the above complications and statistical deficiencies inherent in the Jayaram and Baker (2010) multi-stage estimation procedure, we introduce the Scoring estimation approach, a method based on maximum likelihood estimation framework. The proposed Scoring estimation approach produces model parameter estimators consistently in a single stage algorithm, which admits any parametric class of correlation functions and associated spatial correlation properties, including anisotropic or non-stationary choices.

A One-Stage Algorithm: The Scoring Estimation Approach

The one-stage estimation approach we propose here aims at obtaining the maximum likelihood estimate of $\boldsymbol{\alpha}$ by maximizing the following log-likelihood function:

$$\begin{aligned} l(\boldsymbol{\alpha}) &= \ln L(\boldsymbol{\alpha}) \\ &= -\frac{\sum_{i=1}^{N} n_i}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\mathbf{C}(\boldsymbol{\theta})\right| - \frac{1}{2}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^\top \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]\,, \end{aligned} \quad (27)$$

in which $L(\boldsymbol{\alpha}|\mathbf{Y})$ is the likelihood function, $\mathbf{f}(\mathbf{X},\, \mathbf{b}) = \left(\mathbf{f}(\mathbf{X}_1,\, \mathbf{b})^\top, \ldots, \mathbf{f}(\mathbf{X}_N,\, \mathbf{b})^\top\right)^\top$ and $\mathbf{Y} = (\mathbf{Y}_1^\top, \ldots, \mathbf{Y}_N^\top)^\top$.

The classic statistical method to maximize the log-likelihood function (27) is via the Newton-Raphson algorithm. The Newton-Raphson algorithm finds the estimate of $\boldsymbol{\alpha}$ that maximizes the log-likelihood function (27) via the updating equation:

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \widehat{\boldsymbol{\alpha}}^{(k)} - \mathbf{H}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})\mathbf{S}(\widehat{\boldsymbol{\alpha}}^{(k)})\,, \quad (28)$$

in which $\widehat{\boldsymbol{\alpha}}^{(k)}$ denotes the estimate of $\boldsymbol{\alpha}$ at iteration step k, and

$$\mathbf{S}(\boldsymbol{\alpha}) = \frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \quad \text{and} \quad \mathbf{H}(\boldsymbol{\alpha}) = \frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top} \tag{29}$$

represent the gradient and Hessian matrix of $l(\boldsymbol{\alpha})$, respectively. In general, however, the Newton-Raphson algorithm may not be a robust maximization algorithm when applied directly to applications such as the one in this study. There are numerous reasons for this. First, even though the Hessian matrix is negative definite at the local maximum, the Hessian matrix may not be negative definite at every iteration. Thus, the algorithm does not guarantee an ascent direction of the log-likelihood function and may converge to a local minimum if positive definite Hessian matrices are encountered during the updates. Second, the Hessian matrix can sometimes have poor sparsity and thus can be computationally expensive to evaluate at each iteration. Finally, the Hessian matrix can be indefinite or even singular (Seber and Wild, 2003), causing numerical instabilities in the Newton-Raphson algorithm.

To overcome these issues, the Scoring estimation approach is proposed in this article to obtain the maximum likelihood estimate of $\boldsymbol{\alpha}$. The Scoring estimation approach is based on the method of Scoring introduced by Fisher (1925), which is a modified version of the Newton-Raphson algorithm. The updating equation for the Scoring estimation approach is obtained by replacing the negative Hessian matrix, $-\mathbf{H}(\boldsymbol{\alpha})$, by the expected (or Fisher) information matrix, $\mathbf{I}(\boldsymbol{\alpha})$:

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \widehat{\boldsymbol{\alpha}}^{(k)} + \mathbf{I}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)}) \mathbf{S}(\widehat{\boldsymbol{\alpha}}^{(k)}) \tag{30}$$

with

$$\mathbf{I}(\boldsymbol{\alpha}) = -\mathbb{E}\left[\mathbf{H}(\boldsymbol{\alpha})\right] = -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^\top}\right]. \tag{31}$$

Let $\boldsymbol{\alpha}_0$ be the true parameter value of $\boldsymbol{\alpha}$ and assume that $L(\boldsymbol{\alpha})$ and its first derivatives with respect to $\boldsymbol{\alpha}$ are continuous in the domains of $\boldsymbol{\alpha}$ and $\mathbf{Y}$. Then it can be shown (Wooldridge, 2010) that

$$\mathbf{I}(\boldsymbol{\alpha}_0) = \mathbf{A}(\boldsymbol{\alpha}_0) \tag{32}$$

with

$$\mathbf{A}(\boldsymbol{\alpha}) = \mathbb{E}\left[\frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^{\top}}\right], \tag{33}$$

which is positive-definite. This result states that the expected information matrix $\mathbf{I}(\boldsymbol{\alpha}_0)$ is always positive-definite, meaning that if we replace $\boldsymbol{\alpha}_0$ in $\mathbf{I}(\boldsymbol{\alpha}_0)$ by $\widehat{\boldsymbol{\alpha}}^{(k)}$, then each iteration of the approach will lead the log-likelihood function in an uphill direction. Therefore, the Scoring estimation approach is more numerically stable than the Newton-Raphson algorithm. Furthermore, equation (32) states that only the gradient of $l(\boldsymbol{\alpha})$ is required for the calculation of the expected information matrix $\mathbf{I}(\boldsymbol{\alpha})$, implying that computation in each iteration of the approach is usually quicker than that of Newton-Raphson.

Denote the gradient $\mathbf{S}(\boldsymbol{\alpha})$ and expected information matrix $\mathbf{I}(\boldsymbol{\alpha})$ of $l(\boldsymbol{\alpha})$ by the partitions

$$\mathbf{S}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{S}_{\mathbf{b}}(\boldsymbol{\alpha}) \\ \\ \mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \end{bmatrix} \tag{34}$$

and

$$\mathbf{I}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{I}_{\mathbf{bb}}(\boldsymbol{\alpha}) & \mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}(\boldsymbol{\alpha}) \\ \\ \mathbf{I}_{\boldsymbol{\theta}\mathbf{b}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\alpha}) \end{bmatrix}. \tag{35}$$

Then, the Scoring estimation approach obtains the maximum likelihood estimate of $\boldsymbol{\alpha}$ by the updating equations

$$\widehat{\mathbf{b}}^{(k+1)} = \widehat{\mathbf{b}}^{(k)} + \mathbf{I}_{\mathbf{bb}}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})\, \mathbf{S}_{\mathbf{b}}(\widehat{\boldsymbol{\alpha}}^{(k)}), \tag{36}$$

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})\, \mathbf{S}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}^{(k)}), \tag{37}$$

in which

- the i-th element of $\mathbf{S}_{\mathbf{b}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{S}_{\mathbf{b}}(\boldsymbol{\alpha})]_i = \left[\frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]; \tag{38}$$

- the i-th element of $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$ is given by

$$\begin{aligned}[\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})]_i = &-\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} \\ &+ \frac{1}{2}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]; \end{aligned} \tag{39}$$

- the ij-th element of $\mathbf{I_{bb}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I_{bb}}(\boldsymbol{\alpha})]_{ij} = \left[\frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})\, \frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_j}\,; \tag{40}$$

- the ij-th element of $\mathbf{I_{\boldsymbol{\theta\theta}}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I_{\boldsymbol{\theta\theta}}}(\boldsymbol{\alpha})]_{ij} = \frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\right\}. \tag{41}$$

The proof for equation (36) to (41) can be found in the Proof of Equations (36)- (41) of the Main Article section of the electronic supplement to this article.

It can be seen from the updating equations (36) and (37) that the Scoring estimation approach is able to update the estimates of $\mathbf{b}$ and $\boldsymbol{\theta}$ by separate equations. This separation has two advantages. For the Newton-Raphson update equation (28), it requires at each iteration the complexity (i.e., a concept in computer sciences describing the amount of time required for running an algorithm) of $\mathcal{O}(p^3)$ dominated by the inversion of the Hessian matrix $\mathbf{H}(\widehat{\boldsymbol{\alpha}}^{(k)})$. However, thanks to the separation, the Scoring estimation approach only requires at each iteration the complexity of $\mathcal{O}(p_1^3 + p_2^3)$ dominated by inversions of

$$\mathbf{I_{bb}}(\widehat{\boldsymbol{\alpha}}^{(k)}) \in \mathbb{R}^{p_1 \times p_1} \quad \text{and} \quad \mathbf{I_{\boldsymbol{\theta\theta}}}(\widehat{\boldsymbol{\alpha}}^{(k)}) \in \mathbb{R}^{p_2 \times p_2}\,, \tag{42}$$

in which $p_1 + p_2 = p$ and p_1 and p_2 are dimensions of $\mathbf{b}$ and $\boldsymbol{\theta}$, respectively. Therefore, the separate updating equations in the Scoring estimation approach reduce computational expenses. In addition, equations (36) and (37) indicate that the Scoring estimation approach only requires inversions of $\mathbf{I_{bb}}(\widehat{\boldsymbol{\alpha}}^{(k)})$ and $\mathbf{I_{\boldsymbol{\theta\theta}}}(\widehat{\boldsymbol{\alpha}}^{(k)})$, each of which has a smaller size than the Hessian matrix $\mathbf{H}(\widehat{\boldsymbol{\alpha}}^{(k)})$ in the Newton-Raphson algorithm. Pyzara *et al.* (2011) showed that the size of a matrix is positively connected to its condition number, and the condition number of an ill-conditioned matrix (e.g., a Hilbert matrix) can grow at a remarkably higher rate than that of a well-conditioned matrix as its size increases. Thus, inversions of matrices of smaller sizes in the Scoring estimation approach mitigate the risk of developing large condition numbers, which reduces the effects of round-off error and thus improves the computational stability.

Asymptotic properties of the maximum likelihood estimator $\widehat{\boldsymbol{\alpha}}$

Applying the asymptotic results of M-estimator (Wooldridge, 2010; Demidenko, 2013), we have that the maximum likelihood estimator $\widehat{\boldsymbol{\alpha}}$ is consistent, asymptotically normal, and statistically efficient when $N \to \infty$. The asymptotic standard error estimate $\widehat{\text{se}}(\widehat{\boldsymbol{\alpha}})$ of $\widehat{\boldsymbol{\alpha}} = (\widehat{\mathbf{b}}^\top, \widehat{\boldsymbol{\theta}}^\top)^\top$ can be obtained by

$$\widehat{\text{se}}(\widehat{\mathbf{b}}) = \sqrt{\text{diag}\left[\mathbf{I}_{\mathbf{bb}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\right]} \tag{43}$$

and

$$\widehat{\text{se}}(\widehat{\boldsymbol{\theta}}) = \sqrt{\text{diag}\left[\mathbf{I}_{\boldsymbol{\theta\theta}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\right]}, \tag{44}$$

in which $\widehat{\boldsymbol{\alpha}}^{(K)}$ is the final estimate of $\boldsymbol{\alpha}$ (i.e., the estimate of $\boldsymbol{\alpha}$ given by the Scoring estimation approach at iteration K where the convergence is reached).

Because $\mathbf{I}_{\mathbf{bb}}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})$ and $\mathbf{I}_{\boldsymbol{\theta\theta}}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})$ are involved in the updating equations of the Scoring estimation approach, the asymptotic standard error estimates are by-products of the approach and can be obtained easily after the final iteration K.

Implementing the Scoring estimation approach

Algorithm 2 illustrates the implementation procedure of the Scoring estimation approach. The convergence criterion can be defined either as absolute distance or relative distance between estimate $\widehat{\boldsymbol{\alpha}}^{(k+1)}$ and $\widehat{\boldsymbol{\alpha}}^{(k)}$. According to Golub and Van Loan (2012), the absolute convergence criterion in q-norm can be defined as

$$\kappa_{\text{abs}} = \|\widehat{\boldsymbol{\alpha}}^{(k+1)} - \widehat{\boldsymbol{\alpha}}^{(k)}\|_q . \tag{45}$$

However, when magnitudes of model parameters in $\boldsymbol{\alpha}$ differ widely, a sufficient low tolerance level is required to achieve a satisfactory accuracy at the cost of speed. In such a case and if $\widehat{\boldsymbol{\alpha}}^{(k)} \neq \mathbf{0}$, the relative convergence criterion in q-norm defined by

$$\kappa_{\text{rel}} = \frac{\|\widehat{\boldsymbol{\alpha}}^{(k+1)} - \widehat{\boldsymbol{\alpha}}^{(k)}\|_q}{\|\widehat{\boldsymbol{\alpha}}^{(k)}\|_q} \tag{46}$$

is preferred. The choice of tolerance levels for κ_{abs} and κ_{rel} depends on problems under consideration and trade-offs between accuracy and speed.

Algorithm 2 Scoring estimation approach

Input: $\mathbf{Y}_i$, $\mathbf{X}_{ij}$ and $\mathbf{s}_{ij}$ for $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, n_i\}$.

Output: Estimates of $\mathbf{b}$ and $\boldsymbol{\theta}$ with corresponding asymptotic standard error estimates.

1: Initialization: choose values for $\widehat{\mathbf{b}}^{(1)}$ and $\widehat{\boldsymbol{\theta}}^{(1)}$;
2: **repeat**
3: Update the estimate of $\boldsymbol{\alpha} = (\mathbf{b}^\top, \boldsymbol{\theta}^\top)^\top$ by equations (36) and (37);
4: **until** the convergence criterion is met;
5: Obtain estimates of asymptotic standard errors of $\widehat{\mathbf{b}}$ and $\widehat{\boldsymbol{\theta}}$ by equations (43) and (44).

Numerical considerations

Many ground-motion prediction functions contain both linear and nonlinear parameters in $\mathbf{b}$. When the dimension of $\mathbf{b}$ is large, it can be more computationally effective to separate linear and nonlinear parameters and update their estimates separately to make the Scoring estimation approach better-conditioned and faster to maximize the log-likelihood function. This can be achieved in many families of ground-motion prediction functions, which contain combinations of linear and nonlinear components in the parameters.

To carry out updates for the linear and nonlinear parameter estimates separately (i.e., dimension reduction) in the Scoring estimation approach, the ground-motion prediction function $\mathbf{f}(\mathbf{X}_i, \mathbf{b})$ is decomposed as

$$\mathbf{f}(\mathbf{X}_i, \mathbf{b}) = \mathbf{g}(\mathbf{X}_i, \boldsymbol{\gamma})\boldsymbol{\beta}\,, \tag{47}$$

in which $\boldsymbol{\beta} \in \mathbb{R}^{p_{11}}$ represents a vector of linear parameters in $\mathbf{b}$ with its design matrix $\mathbf{g}(\mathbf{X}_i, \boldsymbol{\gamma})$ and $\boldsymbol{\gamma} \in \mathbb{R}^{p_{12}}$ is a vector of the nonlinear parameters in $\mathbf{b}$. It then can be demonstrated (see Appendix for details) that the Scoring estimation approach with dimension reduction provides a faster and better conditioned estimation procedure than the ordinary Scoring estimation approach represented by the updating equations (36) and (37).

Although the Scoring estimation approach with dimension reduction is generally fast to converge and numerically stable, it can be improved to further speed up the computation and reduce the chances of numerical errors. For example, we can perform inexact line search to promote the convergence by adding a step length $\varphi^{(k)}$ to the updating equation (30) of the Scoring estimation approach:

$$\widehat{\boldsymbol{\alpha}}^{(k+1)} = \widehat{\boldsymbol{\alpha}}^{(k)} + \varphi^{(k)}\mathbf{I}^{-1}(\widehat{\boldsymbol{\alpha}}^{(k)})\mathbf{S}(\widehat{\boldsymbol{\alpha}}^{(k)}) \tag{48}$$

and identify an appropriate value of $\varphi^{(k)}$ at each iteration k such that the log-likelihood function value is increased adequately at minimum cost. Desirable values for step lengths can be searched by algorithms that terminate upon certain conditions, such as the Wolfe conditions (Wolfe, 1969, 1971). For details of the inexact line search, its implementation algorithms as well as other optimization techniques that may be applied to improve the numerical performances of the Scoring estimation approach, readers can refer to Gill *et al.* (1981) and Nocedal and Wright (2006).

Simulation Study

The purpose of this section is to quantify and compare the performances of the multi-stage algorithm and the Scoring estimation approach. The performance of an estimation method can be measured by the accuracy of the obtained model parameter estimates and the resulting predictions. However, this requires knowledge about the true underlying model that is unknown in reality, causing the evaluation of an estimation method difficult in terms of its true performance. To resolve this issue, simulation studies can be implemented. Simulation studies are synthetic experiments conducted on computers under planned conditions, meaning that the generator of the ground-motion data (i.e., the true underlying ground-motion model and its parameter values) is chosen by experimenters and thus fully informative. As a result, the performance of an estimation method can be tested. Simulation studies have been used previously in earthquake modelling in work such as Chen and Tsai (2002), Arroyo and Ordaz

(2010), and Worden *et al.* (2018).

Generator settings

The first step of the simulation study is to specify the underlying generator (i.e., the true ground-motion model) of the considered IM. Specifically, in this simulation study, PGA is used as the considered ground-motion IM. To eliminate the effects of model misspecification, the true ground-motion model is chosen to have the same model representation as the hypothetical ground-motion model specified in Section The Ground-Motion Model with the ground-motion prediction function (proposed by Akkar and Bommer (2010)):

$$f(\mathbf{X}_{ij},\, \mathbf{b}) = b_1 + b_2\, M_i + b_3\, M_i^2 + (b_4 + b_5\, M_i) \log_{10} \sqrt{R_{ij}^2 + b_6^2} + b_7\, S_{S,ij} + b_8\, S_{A,ij} + b_9\, F_{N,i} + b_{10}\, F_{R,i}\,, \quad (49)$$

in which

- M_i is the moment magnitude (M_W) of earthquake i;
- R_{ij} is the Joyner-Boore distance (R_{JB}) (i.e., the closest distance to the surface projection of the rupture plane) in kilometers of site j in earthquake i;
- $S_{S,ij}$ and $S_{A,ij}$ are dummy variables determining the soil type at site j during earthquake i according to

$$(S_{S,ij},\, S_{A,ij}) = \begin{cases} (1,\, 0)\,, & \text{soft soil,} \\ (0,\, 1)\,, & \text{stiff soil,} \\ (0,\, 0)\,, & \text{rock;} \end{cases} \quad (50)$$

- $F_{N,i}$ and $F_{R,i}$ are dummy variables indicating the faulting type of earthquake i according to

$$(F_{N,i},\, F_{R,i}) = \begin{cases} (1,\, 0)\,, & \text{normal fault,} \\ (0,\, 1)\,, & \text{reverse fault,} \\ (0,\, 0)\,, & \text{strike-slip fault.} \end{cases} \quad (51)$$

Two correlation functions are selected for illustrative purposes:

$$k_1(d) = \exp\left(-\frac{d}{h}\right) \tag{52}$$

and

$$k_2(d) = \left(1 + \frac{\sqrt{3}d}{h}\right) \exp\left(-\frac{\sqrt{3}d}{h}\right), \tag{53}$$

which are special cases of Matérn correlation function with $\nu = 0.5$ and $\nu = 1.5$, respectively. The first correlation function (52) (i.e., exponential correlation function) represents a type of spatial correlation structure that is commonly used in works such as Jayaram and Baker (2009, 2010); Esposito and Iervolino (2011, 2012) and allows for an instructive comparison between the two estimation methods. The second correlation function (53) is smoother than the correlation function (52) and admits the comparison between the two estimation approaches when the logarithmic PGA field is smooth.

The parameter values in the true ground-motion model are outlined in Table 1. The values for $b_1, \ldots, b_{10}$, τ^2 and σ^2 are chosen based on the regression results given by Akkar and Bommer (2010) for the ground-motion model of PGA. The value of the range parameter h in the correlation function (52) is set arbitrarily to 11.5 km. This value of h corresponds to $d = 34.45$ km when $\rho = 0.05$ with the correlation function (52). To get the same ρ value at the same distance $d = 34.45$ km, it is found that $h = 12.58$ km for the correlation function (53).

[Table 1 about here.]

Choice for covariates

Before synthetic PGA datasets can be generated, the information of covariates needs to be known. The information of covariates includes the number of earthquakes N, the number of recording sites n_i during each event (i.e., earthquake) as well as their locations $\mathbf{s}_{ij}$, and the values of predictors

$$\mathbf{X}_{ij} = (M_i,\, R_{ij},\, S_{S,ij},\, S_{A,ij},\, F_{N,i},\, F_{R,i})\,. \tag{54}$$

In this simulation study, the information of covariates is extracted from a historical

ground-motion database, the European Strong-Motion (ESM) database (see Section Data and Resources), which ensures the generation of realistic scenarios for comparison of the two estimation methods. In using this database, we apply to the database the selection criteria detailed below so that the proposed simulation study can be independently verified and reproduced:

- retain events occurred within Italy;
- retain events with moment magnitude $M_W \geq 5$, removing events without M_W information;
- remove events without information of fault types;
- retain recording sites with epicentral distance $R_{epi} \leq 250$ km;
- remove recording sites without information of V_{S30} , the average shear-wave velocity (in m/s) in the upper 30 meters of the soil;
- remove recording sites that are not free-field;
- remove recording sites with redundant site information (e.g., co-located recording sites) in a single event; and
- retain events with at least two recording sites.

After the implementation of the above selection criteria, the resulting catalog used in this simulation study consists of 2150 entries of recording sites (in which the same recording site may appear in different earthquakes) from 62 earthquakes of $5 \leq M_W \leq 6.9$ in Italy from 1976 to 2016. The geographical distribution of the 62 earthquakes with their moment magnitudes, and the distribution of inter-site distance in each earthquake are shown in Figure 2.

[Figure 2 about here.]

The R_{JB} of each recording site in each earthquake is calculated based on the corresponding fault geometry (e.g., strike angle, dip angle, rake angle, length, and width), if information of the finite-fault model is available. Otherwise, R_{JB} is estimated by the empirical relationship

between R_{epi} and R_{JB} (Stucchi *et al.*, 2011) if the corresponding earthquake is with $M_W > 5.5$ and is set to be R_{epi} if the corresponding earthquake is with $M_W \leq 5.5$. The obtained R_{JB} for each recording site of each earthquake in the resulting catalog for this simulation study is less than 250 km. The site classification of each recording site in each earthquake is obtained based on the information of V_{S30} from the ESM database. In ESM database, V_{S30} is either obtained from in-situ experiments or inferred from the topographic slope according to Wald and Allen (2007). It is preferable to use V_{S30} from the experimental measurements, and if that is not available, the inferred V_{S30} is used instead. The soil type of each recording site of each earthquake in the catalog for this simulation study is then classified (according to Akkar and Bommer (2010)) as soft soil if $V_{S30} < 360$ m/s, stiff soil if 360 m/s $\leq V_{S30} \leq 750$ m/s, and rock if $V_{S30} > 750$ m/s.

PGA data generation

Given the true ground-motion model and information of covariates, we can simulate synthetic datasets of logarithmic PGAs through Algorithm 3.

Algorithm 3 Synthetic logarithmic PGA dataset generation

Input: Specified true ground-motion model and information of covariates.

Output: A synthetic dataset of logarithmic PGAs (denoted by $\mathbf{y}$).

1: Compute the covariance matrix $\mathbf{C}(\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\tau^2, \sigma^2, h)^\top$;

2: Compute the Cholesky factor $\mathbf{L}$ such that $\mathbf{L}\mathbf{L}^\top = \mathbf{C}(\boldsymbol{\theta})$;

3: Compute the value of $\mathbf{f}(\mathbf{X}, \mathbf{b})$;

4: Generate independently $G = \sum_{i=1}^{N} n_i$ standard normal random numbers $\mathbf{v} = (v_1, \ldots, v_G)^\top$;

5: Return a synthetic dataset of logarithmic PGAs by $\mathbf{y} = \mathbf{f}(\mathbf{X}, \mathbf{b}) + \mathbf{L}\mathbf{v}$.

Evaluation of the estimation performance

In this section, estimation performances of the multi-stage algorithm and the Scoring estimation approach are evaluated and compared. We first generate $\mathcal{T} = 1000$ synthetic datasets of logarithmic PGAs via Algorithm 3. Then for each of the synthetic dataset, the multi-stage algorithm and the Scoring estimation approach are implemented. Let $\widehat{\alpha}_t$ and $\widehat{\text{se}}(\widehat{\alpha}_t)$ represent, respectively, the estimate and the asymptotic standard error estimate of a model parameter $\alpha \in \{\mathbf{b}, \tau^2, \sigma^2, h\}$ produced by one of the two estimation methods on some synthetic dataset $t \in \{1, \ldots, \mathcal{T}\}$. The estimation performance of either method then can be evaluated by computing the following criteria:

- **root mean squared error (RMSE)**, computed by

$$\text{RMSE} = \sqrt{\frac{1}{\mathcal{T}} \sum_{t=1}^{\mathcal{T}} (\widehat{\alpha}_t - \alpha_0)^2}, \tag{55}$$

  in which α_0 is the true parameter value (given in Table 1) of α;

- **coverage rate (CR)**, defined by the percentage of $\mathcal{T}$ synthetic datasets in which the true parameter value α_0 falls into the 95% confidence interval constructed from $\widehat{\alpha}_t$ and $\widehat{\text{se}}(\widehat{\alpha}_t)$.

[Table 2 about here.]

Table 2 illustrates the estimation criteria of the parameter estimators produced by the multi-stage algorithm and the Scoring estimation approach under the correlation functions (52) and (53). It can be observed that the RMSEs of all parameter estimators from the Scoring estimation approach are less than those from the multi-stage algorithm under both types of correlation functions. Although the RMSEs of estimators of $b_1, \ldots, b_{10}$ produced by the multi-stage algorithm are not significantly higher than those produced by the Scoring estimation approach, the RMSEs of $\widehat{\tau^2}$, $\widehat{\sigma^2}$ and $\widehat{h}$ are noticeably different between the two methods. For $\widehat{\tau^2}$, the multi-stage algorithm produces 50% higher RMSE than the Scoring estimation approach under the correlation function (52) and two times larger RMSE than the Scoring

estimation approach under the correlation function (53). With regard to $\widehat{\sigma^2}$, the RMSE from the multi-stage algorithm is around eight times larger than that from the Scoring estimation approach under the correlation function (52) and more than 30 times larger than that from the Scoring estimation approach under the correlation function (53). Similar observations can be seen regarding the estimator of h, whose RMSE from the multi-stage algorithm is 12 times higher than that from the Scoring estimation approach under the correlation function (52) and about 26 times larger than that from the Scoring estimation approach under the correlation function (53). These findings imply that the estimators, particularly the estimators of τ^2, σ^2, and h, given by the Scoring estimation approach are more robust.

Finally, it can be found that the CRs under the Scoring estimation approach are relatively stable across different model parameters, the CRs for τ^2, σ^2, and h under the multi-stage algorithm are remarkably lower than the expected 95% confidence level, indicating that the constructed confidence interval from the multi-stage algorithm is biased in a non-conservative manner, that is, too narrow on average, and there exist risks of wrong decisions on hypothesis tests relating to model structure for the resulting GMPE under such an estimation procedure. The low CRs of τ^2 and σ^2 are partly due to the non-optimal formulas of asymptotic standard error estimates given by Jayaram and Baker (2010) and partly due to the separate estimation of h and the inconsistency of $\widehat{h}$. The low CR of h is because of the naive use of the asymptotic standard error formula for ordinary least squares and the inconsistency of $\widehat{\sigma^2}$ produced from the preliminary stage.

To examine how the estimation performances of the multi-stage algorithm and the Scoring estimation approach change, when the sample (i.e., event) size N varies, we extract two sub-catalogs from the full catalog described in Section Choice for covariates. One sub-catalog has the size of $N = 46$, which includes the events occurred by the end of the year 2010. Another sub-catalog has the size of $N = 29$, which includes the events occurred by the end of the year 2000. We then generate 1000 synthetic datasets of logarithmic PGAs for both sub-catalogs and implement the multi-stage algorithm and the Scoring estimation approach,

which provides 1000 sets of estimates for each sub-catalog under each estimation method. Figure 3 and 4 present the sampling distributions of $\widehat{b}_1, \ldots, \widehat{b}_{10}$ under correlation function (52) and (53), respectively. As we expected in Section Problems of the multi-stage algorithm, both the multi-stage algorithm and the Scoring estimation approach produce consistent estimators of $b_1, \ldots, b_{10}$ (i.e., the sampling distributions of $\widehat{b}_1, \ldots, \widehat{b}_{10}$ converge to the true parameter values as N increases).

[Figure 3 about here.]

[Figure 4 about here.]

We emphasize in Section Problems of the multi-stage algorithm that $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$ produced by the multi-stage algorithm are inconsistent, meaning that the sampling distribution of $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$ from the multi-stage algorithm will not converge to the true parameter values as N grows. This statement is illustrated in Figure 5. Under both the correlation function (52) and (53), the sampling distributions of $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$ produced by the Scoring estimation approach converge to the true parameter values as N increases. In contrast, the sampling distributions of $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$ produced by the multi-stage algorithm are biased. Moreover, the sampling distributions of $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ produced by the multi-stage algorithm under the correlation function (53) behave worse than those under the correlation function (52) because increasing sampling variances and a larger number of outliers are observed.

[Figure 5 about here.]

Evaluation of the predictive performance

The estimated ground-motion models allow one to perform ground-motion predictions at locations where recording sites are unavailable (e.g., generate a ground-motion shaking intensity map). Therefore, it is vital to assess the predictive performances of the ground-motion models estimated by the multi-stage algorithm and the Scoring estimation approach. To this goal, we examine the prediction accuracy for a selected event with ID 'IT-1997-0137',

which corresponds to the earthquake with $M_W = 5.6$ occurred in the regions of Umbria and Marche in 1997 and has $n_e = 15$ recording sites. This particular event is selected because it is included in both the full catalog (events by the end of the year 2016) and the two sub-catalogs (events by the end of the year 2000 and 2010) described in Section Evaluation of the estimation performance. This allows us to examine how the predictive performance of an estimation method changes as the number of events used for estimation varies. The prediction region of the event is set to be within a distance of 250 km from the epicenter (see Figure 6). The ground-motion models used for predictions are those estimated from the full catalog and the two sub-catalogs in Section Evaluation of the estimation performance.

[Figure 6 about here.]

We first discretize the prediction region of the event by fine square grids with mesh size $\Delta = 5$ km and treat the resulting $K = 5228$ grid points as prediction locations. Then, for each estimation method and each catalog (i.e., the full catalog and the two sub-catalogs) we proceed with the following steps:

1. For each synthetic dataset t, compute the predictions $\widehat{\mathbf{z}}_t = (\widehat{z}_{1,t}, \ldots, \widehat{z}_{K,t})$ on all grid points $k \in \{1, \ldots, K\}$ by the plug-in predictor (Stein, 1999)

$$\widehat{\mathbf{z}}_t = \mathbf{f}(\mathbf{W}, \widehat{\mathbf{b}}_t) + \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_t)\mathbf{c}^{-1}(\widehat{\boldsymbol{\theta}}_t)\left(\mathbf{y}_t - \mathbf{f}(\mathbf{X}_e, \widehat{\mathbf{b}}_t)\right), \tag{56}$$

   in which

   - $\widehat{\mathbf{b}}_t$ and $\widehat{\boldsymbol{\theta}}_t = (\widehat{\tau^2}_t, \widehat{\sigma^2}_t, \widehat{h}_t)$ are parameter estimates obtained from synthetic dataset t;
   - $\mathbf{f}(\mathbf{W}, \widehat{\mathbf{b}}_t) = (\mathbf{f}(\mathbf{W}_1, \widehat{\mathbf{b}}_t), \ldots, \mathbf{f}(\mathbf{W}_K, \widehat{\mathbf{b}}_t))^\top$ is a $K \times 1$ vector of mean logarithmic PGAs with $\mathbf{W}_k$ being a vector of predictors at grid point k. The soil types at grid points are obtained from the U.S. Geological Survey global V_{S30} database (see Section Data and Resources);
   - $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \text{cov}(\mathbf{Z}, \mathbf{Y})$ and $\mathbf{c}(\boldsymbol{\theta}) = \text{var}(\mathbf{Y})$ with $\mathbf{Z}$ and $\mathbf{Y}$ representing vectors of logarithmic PGAs at grid points and recording sites, respectively;

- $\mathbf{y}_t$ is an $n_e \times 1$ vector of logarithmic PGAs at recording sites and is obtained from the the t-th synthetic dataset of logarithmic PGAs simulated in Section Evaluation of the estimation performance;
- $\mathbf{f}(\mathbf{X}_e, \widehat{\mathbf{b}}_t) = (\mathbf{f}(\mathbf{X}_{e,1}, \widehat{\mathbf{b}}_t), \ldots, \mathbf{f}(\mathbf{X}_{e,n_e}, \widehat{\mathbf{b}}_t))^\top$ is an $n_e \times 1$ vector of mean PGAs with $\mathbf{X}_{e,j}$ being a vector of predictors at the recording site $j \in \{1, \ldots, n_e\}$ of the event.

In this step, a ground-motion shaking intensity map can be generated from the obtained $\widehat{\mathbf{z}}_t$, which represent the logarithmic PGAs on grid points predicted by the estimated ground-motion model given the synthetic observations $\mathbf{y}_t$;

2. For each $\mathbf{y}_t$, generate a synthetic logarithmic PGA dataset $\mathbf{z}_t = (z_{1,t}, \ldots, z_{K,t})$ on all grid points $k \in \{1, \ldots, K\}$ from the multivariate normal distribution

$$\mathcal{N}\left(\mathbf{f}(\mathbf{W}, \mathbf{b}_0) + \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\mathbf{c}^{-1}(\boldsymbol{\theta}_0)\left(\mathbf{y}_t - \mathbf{f}(\mathbf{X}_e, \mathbf{b}_0)\right), \boldsymbol{\Psi}(\mathbf{b}_0) - \boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\mathbf{c}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^\top(\boldsymbol{\theta}_0)\right), \quad (57)$$

in which $\boldsymbol{\Psi}(\boldsymbol{\theta}) = \mathrm{var}(\mathbf{Z})$, and $\mathbf{b}_0$ and $\boldsymbol{\theta}_0$ are true parameter values chosen for $\mathbf{b}$ and $\boldsymbol{\theta}$ in Section Generator settings. To assess the quality of the ground-motion shaking intensity map (i.e., the accuracy of the predictions $\widehat{\mathbf{z}}_t$) produced by the estimated ground-motion model in the last step, this step generates the benchmark logarithmic PGAs (i.e., $\mathbf{z}_t$) on grid points using the underlying true ground-motion model given the synthetic observations $\mathbf{y}_t$;

3. At each grind point k, compute the root mean squared error of predictions (RMSEP) by

$$\mathrm{RMSEP}_k = \sqrt{\frac{1}{\mathcal{T}}\sum_{t=1}^{\mathcal{T}}(\widehat{z}_{k,t} - z_{k,t})^2}, \quad (58)$$

which measures the predictive accuracy of the estimated ground-motion model at each grid point k.

In Figure 7, we plot at each grid point the percentage increase in RMSEP from the multi-stage algorithm relative to that from the Scoring estimation approach under three sample sizes of $N = 29$, 46 and 62 (corresponding to events by the end of the year 2000, 2010

and 2016) with correlation function (52) and (53). It can be seen that for both correlation function (52) and (53), as N increases, the region where the RMSEP from the multi-stage algorithm is greater than that from the Scoring estimation approach expands. When the correlation function (52) is considered, we find that the RMSEP from the Scoring estimation approach are smaller than those from the multi-stage algorithm, especially around the recording sites (triangles in Figure 7). This is because the spatial correlation structure in the ground-motion model is estimated with higher accuracy by the Scoring estimation approach. Because recording sites are often concentrated in the near-fault regions, the difference between the RMSEP from the Scoring estimation approach and that from the multi-stage algorithm becomes more distinct within the near-field (the region bounded by the dashed circle in Figure 7). This observation becomes remarkable when the correlation function (53) is considered, in which the RMSEP from the multi-stage algorithm can exceed that from the Scoring estimation approach by more than 10% near the recording sites. Furthermore, Figure 7 also indicates that the Scoring estimation approach is less sensitive to the overfitting problem than the multi-stage algorithm. As we can observe from (a) and (b) in Figure 7, even the number of events is scare (i.e., $N = 29$), the predictive performance of the Scoring estimation approach is still comparable or better than that of the multi-stage algorithm over the region, especially when the underlying spatial correlation follows the correlation function (53).

[Figure 7 about here.]

Performance of the Scoring Estimation Approach under the Ignorance of Spatial Correlation

We have demonstrated that the Scoring estimation approach outperforms the multi-stage algorithm in terms of estimation and prediction. However, if the spatial correlation structure is neglected from the ground-motion model while the spatial correlation is significant in

the ground-motion data, the performance of Scoring estimation approach may be degraded. Because most of the existing ground-motion models (e.g., Akkar and Bommer (2010); Abrahamson *et al.* (2014); Bindi *et al.* (2014); Boore *et al.* (2014); Campbell and Bozorgnia (2014); Chiou and Youngs (2014); Idriss (2014)) are proposed without any form of spatial correlation structure, we investigate in this section the performance of the Scoring estimation approach when the ground-motion model ignores spatial correlation.

Estimation performance

To assess the estimation performance of the Scoring approach, when the spatial correlation structure is ignored in the ground-motion model, 1000 synthetic datasets of logarithmic PGAs, which form a training set, are generated using the correlation function (52) with $h = 11.50$ km. The Scoring estimation approach is then applied to estimate, respectively, the ground-motion model with well-specified spatial correlation structure (i.e., with the correlation function (52)) and the ground-motion model without spatial correlation structure (i.e., with the correlation function (5)). The sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$ obtained under the two ground-motion models are shown in Figure 8. It can be seen that although the estimators of $b_1, \ldots, b_{10}$ produced by the Scoring estimation approach are generally unbiased for both models, estimators such as $\widehat{b}_5, \ldots, \widehat{b}_8$ exhibit larger variances when the spatial correlation structure is ignored in the ground-motion model. Comparisons between the sampling distributions for $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ under the two models are presented in Figure 9. We observe that when the training set is generated by the correlation function (52) with $h = 11.50$ km, the estimates of the inter-event variance τ^2 from the ground-motion model without spatial correlation structure are overestimated by the Scoring estimation approach, but the estimates of the intra-event variance σ^2 are underestimated. For the ground-motion model with well-specified spatial correlation structure, however, the estimates of τ^2 and σ^2 produced by the Scoring estimation approach essentially match their true values. To further investigate the Scoring estimation approach's overestimation on τ^2 and underestimation

on σ^2 when spatial correlation is ignored from the ground-motion model, we refit the two ground-motion models to two additional training sets, each of which consists of 1000 synthetic datasets of logarithmic PGAs, generated using the correlation function (52) with $h = 30.00$ and 60.00 km, respectively. From Figure 9, it can be seen that as the value of h increases (i.e., the spatial correlation implied by the training data becomes stronger), the overestimation on τ^2 and underestimation on σ^2 due to the ignorance of spatial correlation are amplified. On the contrary, the estimates of τ^2 and σ^2 from the ground-motion model with well-specified spatial correlation structure are still concentrated around the true parameter values.

[Figure 8 about here.]

[Figure 9 about here.]

We repeated the above procedure using the training sets generated by the correlation function (53). The sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$, $\widehat{\tau^2}$, and $\widehat{\sigma^2}$ under the two completing ground-motion models are visualized in Figure 10 and 11. Figure 10 indicates that the loss of statistical efficiency on the estimator of $\mathbf{b}$ becomes more apparent when the ground-motion model without spatial correlation structure is fitted to the training data with smoother spatial correlation. From Figure 11, we find that fitting the ground-motion model without spatial correlation structure to the training data with smoother spatial correlations will cause severer overestimation on τ^2 and underestimation on σ^2 . In contrast, the changed smoothness of the spatial correlation in the training data does not influence the accuracy of estimating τ^2 and σ^2 in the ground-motion model with well-specified spatial correlation structure.

[Figure 10 about here.]

[Figure 11 about here.]

Predictive performance

In this section, we consider the predictive performance of the estimated (via the Scoring estimation approach) ground-motion model without spatial correlation structure for the event

selected in Section Evaluation of the predictive performance. To investigate the predictive performance when observations are available in the far-field, 15 artificial recording sites are added to the event (see Figure 12). The addition of the 15 artificial recording sites increases the entries of recording sites in the catalog, which is described in Section Choice for covariates, from 2150 to 2165. On the basis of the updated catalog, we then generate six training sets, each of which includes 1000 synthetic datasets of logarithmic PGAs, using the generator specified in Section Generator settings with $h = 11.50$, 30.00 and 60.00 km for the correlation function (52) and with $h = 12.58$, 32.81 and 65.63 km for the correlation function (53). For each training set, we estimate the ground-motion model with well-specified spatial correlation structure (i.e., with the same correlation function as the underlying generator) and the ground-motion model with no spatial correlation structure by the Scoring estimation approach. The predictive performances of the estimated ground-motion models are subsequently assessed by the RMSEP obtained via the procedure detailed in Section Evaluation of the predictive performance. The RMSEPs produced by the estimated ground-motion models with and without spatial correlation are plotted in Figure 13 and 14. These figures show that when the spatial correlation structure is ignored from the ground-motion model, the resulting predictions are poor across the study region regardless of the strength (i.e., the magnitude of h) and the smoothness (i.e., the choice between the correlation function (52) and (53)) of the spatial correlation implied by the training data. In addition, we find that whereas the RMSEP around the recording sites are only weakly improved when the spatial correlation is ignored from the ground-motion model, the RMSEP near the recording sites are significantly reduced when the spatial correlation is well-specified in the ground-motion model. For example, when the spatial correlation implied by the training data are characterized by the correlation function (52) with $h = 60$ km, little reductions in RMSEP can be observed around the recording sites if the data are fitted by the ground-motion model without spatial correlation structure (see (f) in Figure 13). However, the improvement of predictions near the recording sites is obvious when the spatial correlation structure is well-specified in the ground-motion

model (see (e) in Figure 13). Furthermore, it is found that the reductions of RMSEP caused by the availability of recording sites are consistent in near-field and far-field, when the ground-motion model with well-specified spatial correlation structure is considered. However, under the ground-motion model without spatial correlation structure, the improvement of predictions caused by the proximity to the recording sites is clearer in the near-field than in the far-field, which suffers high RMSEP in all considered scenarios.

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

Conclusions

In this article, a one-stage algorithm, namely the Scoring estimation approach, is introduced under the maximum likelihood estimation framework. It is capable of estimating all parameters in ground-motion models with spatial correlation simultaneously and can be readily extended to accommodate a wide range of correlation functions (e.g., site-related correlation functions). The estimators produced by the approach have good statistical properties such as consistency, statistical efficiency and asymptotic normality. In addition, to yield consistent, statistically efficient and asymptotically normal estimators, the approach requires only a large number of events (that can be assumed to be independent) even with a small number of records per event, something that is historically relevant to earthquake records. The simulation study demonstrates that the Scoring estimation approach generally outperforms the multi-stage algorithm proposed by Jayaram and Baker (2010) in terms of estimation and prediction. With regard to estimation, the Scoring estimation approach produces parameter estimators in an accurate and stable manner under both smooth (e.g., correlation function (53)) and less smooth (e.g., correlation function (52)) correlation functions. Regarding the predictive performance, the simulation study indicates that the ground-motion model with

spatial correlation estimated via the Scoring estimation approach produces smaller prediction errors than the multi-stage algorithm does, especially at locations around the recording sites and when the spatial correlation is smooth. Because the estimation of ground-motion models with spatial correlation is a key ingredient in developing GMPEs for use in PHSA, the Scoring estimation approach provides a statistically robust way that increases the estimation accuracy in ground-motion model construction and has the potential to reduce prediction errors in ground-motion shaking intensity maps, which in turn can improve the earthquake-induced loss assessment process.

The performance of the Scoring estimation approach is also assessed under the condition that spatial correlation structure is ignored in ground-motion models. It is demonstrated that neglecting spatial correlation structure in ground-motion models can cause the Scoring estimation approach to produce inconsistent and statistically inefficient estimators, and inaccurate predictions. This investigation provides two important implications for seismic risk assessment. First, as any estimation technique, the Scoring estimation approach is only as good as the proposed ground-motion model. Therefore, a rigorous assessment of spatial correlation in the ground-motion data should be addressed during the GMPE construction such that the resulting ground-motion model is a good representation of the underlying data. In return, the Scoring estimation approach can serve as a competitive method for accurate ground-motion model estimation and shaking intensity map generation. Second, we show that ignoring spatial correlation in ground-motion models can result in overestimation of the inter-event variance and underestimation of the intra-event variance, and such biases increase when the spatial correlation implied by the underlying data becomes stronger and smoother. These results generalize the findings of Jayaram and Baker (2010) and further emphasize the importance to accurately estimate the inter-event and intra-event variances as their changes "have implications for risk assessments of spatially-distributed systems" (Jayaram and Baker, 2010).

Finally, because the Scoring estimation approach provides a relatively accurate estimation

of the spatial correlation parameters (e.g., h in the exponential correlation function), as a by-product of the ground-motion model estimation, this approach could be applied to areas that do not have well-recorded events, giving the opportunity to provide a first estimate of a spatial correlation model.

Data and Resources

The Engineering Strong-Motion database was searched using http://esm.mi.ingv.it (last accessed May 2018). The U.S. Geological Survey global V_{S30} database was obtained from https://earthquake.usgs.gov/data/vs30/ (last accessed February 2018). The multi-stage algorithm and the Scoring estimation approach are implemented in MATLAB® version R2018a and the code is available at https://github.com/mingdeyu/GMPE-estimation/ (last accessed December 2018).

Acknowledgments

The authors thank John Douglas, Editor-in-Chief Thomas Pratt, and two anonymous reviewers for their helpful reviews of this article. Thanks to Irmela Zentner and Zhiyi Wang for helpful discussions about the method proposed by Jayaram and Baker (2010). Deyu Ming and Chen Huang gratefully acknowledge the financial support of the China Scholarship Council (Grant No. 201608170005 and 201608440273). This work is partly funded by the UK Natural Environment Research Council (Grant Number NE/P01660X/1).

References

Abrahamson, N. A., W. J. Silva and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra* **30**, no. 3, 1025–1055.

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, 505–510.

Akkar, S. and J. J. Bommer (2010). Empirical equations for the prediction of PGA, PGV, and spectral accelerations in Europe, the Mediterranean region, and the Middle East, *Seismol. Res. Lett.* **81**, no. 2, 195–206.

Arroyo, D. and M. Ordaz (2010). Multivariate bayesian regression analysis applied to ground-motion prediction equations, part 1: theory and synthetic example, *Bull. Seismol. Soc. Am.* **100**, no. 4, 1551–1567.

Bindi, D., M. Massa, L. Luzi, G. Ameri, F. Pacor, R. Puglia, and P. Augliera (2014). Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset, *Bull. Seismol. Soc. Am.* **12**, no. 1, 391–430.

Boore, D. M., J. P. Stewart, E. Seyhan and G. M. Atkinson (2014). NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes, *Earthq. Spectra* **30**, no. 3, 1057–1085.

Brillinger, D. R. and H. K. Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data, *Bull. Seismol. Soc. Am.* **74**, no. 4, 1441–1450.

Brillinger, D. R. and H. K. Preisler (1984). Further analysis of the Joyner-Boore attenuation data, *Bull. Seismol. Soc. Am.* **75**, no. 2, 611–614.

Campbell, K. W. and Y. Bozorgnia (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra, *Earthq. Spectra* **30**, no. 3, 1087–1115.

Chen, Y. H. and C. C. P. Tsai (2002). A new method for estimation of the attenuation relationship with variance components, *Bull. Seismol. Soc. Am.* **92**, no. 5, 1984–1991.

Chiou, B. S. J. and R. R. Youngs (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **30**, no. 3, 1117–1153.

Couvreur, C. (1997). The EM algorithm: a guided tour, in *Computer Intensive Methods in Control and Signal Processing*, K. Warwick and M. Kárný (Editors), Birkhäuser, Boston, Massachusetts, 209–222.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*, Revised Ed., John Wiley & Sons, New York, New York, 900 pp.

Demidenko, E. (2013). *Mixed Models: Theory and Applications with R*, John Wiley & Sons, Hoboken, New Jersey, 717 pp.

Draper, N. R. and H. Smith (2014). *Applied Regression Analysis*, John Wiley & Sons, New York, New York, 736 pp.

Esposito, S. and I. Iervolino (2011). PGA and PGV spatial correlation models based on European multievent datasets, *Bull. Seismol. Soc. Am.* **101**, no. 5, 2532–2541.

Esposito, S. and I. Iervolino (2012). Spatial correlation of spectral acceleration in European data, *Bull. Seismol. Soc. Am.* **102**, no. 6, 2781–2788.

Fisher, R. A. (1925). Theory of statistical estimation, *Math. Proc. Camb. Phil. Soc.*, **22**, no. 5, 700–725.

Gao, X. and Y. Wang (2013). Application of EM algorithm in statistics natural language processing, *Res. J. Appl. Sci. Eng. Tech.* **5**, no. 10, 2969–2943.

Gill, P. E., W. Murray, and M. H. Wright (1981). *Practical Optimization*, Emerald, Bingley, United Kingdom, 401 pp.

Goda, K. and G. M. Atkinson (2009). Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan, *Bull. Seismol. Soc. Am.* **99**, no. 5, 3003–3020.

Goda, K. and G. M. Atkinson (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data, *Bull. Seismol. Soc. Am.* **100**, no. 6, 3055–3067.

Goda, K. and H. P. Hong (2008). Spatial correlation of peak ground motions and response spectra, *Bull. Seismol. Soc. Am.* **98**, no. 1, 354–365.

Golub, G. H. and C. F. Van Loan (2012). *Matrix Computations*, 4 Ed., JHU Press, 784 pp.

Hong, H. P., Y. Zhang, and K. Goda (2009). Effect of spatial correlation on estimated ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **99**, no. 2A, 928–934.

Idriss, I. M. (2014). An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, *Earthq. Spectra* **30**, no. 3, 1155–1177.

Jayaram, N. and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, *Earthq. Eng. Struct. Dynam.* **38**, no. 15, 1687–1708.

Jayaram, N. and J. W. Baker (2010). Considering spatial correlation in mixed-effects regression and the impact on ground-motion models, *Bull. Seismol. Soc. Am.* **100**, no. 6, 3295–3303.

Joyner, W. B. and D. M. Boore (1993). Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* **83**, no. 2, 469–487.

Kerby, B. (2016). Semivariogram Estimation: Asymptotic Theory and Applications, *Ph.D. thesis*, The University of Utah, Salt Lake City, Utah, 125 pp.

Lahiri, S. N., N. Lange, and D. Stram (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm, *J. Am. Stat. Assoc.* **82**, no. 397, 97–105.

Lahiri, S. N., Y. Lee, and N. Cressie (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters, *J. Stat. Plann. Infer.* **103**, no. 1–2, 65–85.

Lahiri, S. N. and J. N. Ware (1982). Random-effects methods for longitudinal data, *Biometrics* **38**, no. 4, 963–974.

Lindstrom, M. J. and D. M. Bates (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics* **46**, 673–687.

Nocedal, J. and S. J. Wright (2006). *Numerical Optimization*, Springer, New York, New York, 664 pp.

Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and S-PLUS*, Springer, Madison, Wisconsin, 528 pp.

Pyzara, A., B. Bylina, and J. Bylina (2011). The influence of a matrix condition number on iterative methods' convergence, in *IEEE Federated Conference on Computer Science and Information Systems*, Szczecin, Poland, 18–21 September 2011, 459–464.

Rasmussen, C. E. and C. K. Williams (2006). *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts, 248 pp.

Seber, G. and C. Wild (2003). *Nonlinear Regression*, John Wiley & Sons, Hoboken, New Jersey, 768 pp.

Sokolov, V., F. Wenzel, W. Y. Jean, and K. L. Wen (2010). Uncertainty and spatial correlation of earthquake ground motion in Taiwan, *Terr. Atmos. Ocean. Sci.* **21**, no. 6, 905–921.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, New York, 247 pp.

Stucchi, M., C. Meletti, V. Montaldo, H. Crowley, G. M. Calvi, and E. Boschi (2011). Seismic hazard assessment (2003-2009) for the Italian building code, *Bull. Seismol. Soc. Am.* **101**, no. 4, 1885–1911.

Wald, D. J. and T. I. Allen (2007). Topographic slope as a proxy for seismic site conditions and amplification, *Bull. Seismol. Soc. Am.* **97**, no. 5, 1379–1395.

Wang, M. and T. Takada (2005). Macrospatial correlation model of seismic ground motions, *Earthq. Spectra* **21**, no. 4, 1137–1156.

Wolfe, P. (1969). Convergence conditions for ascent methods, *SIAM Rev.* **11**, no. 2, 226–235.

Wolfe, P. (1971). Convergence conditions for ascent methods. II: some corrections, *SIAM Rev.* **13**, no. 2, 185–188.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, Massachusetts, 1064 pp.

Worden C. B., E. M. Thompson, J. W. Baker, B. A. Bradley, N. Luco, and D. J. Wald (2018). Spatial and spectral interpolation of ground–motion intensity measure observations, *Bull. Seismol. Soc. Am.* **108**, no. 2, 866–875.

Zimmerman, D. L. and M. Stein (2010). Classical geostatistical methods, in *Handbook of Spatial Statistics*, A. E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes (Editors), CRC Press, Boca Raton, Florida, 29–44.

Author Mailing Addresses

Department of Statistical Science

University College London

London, England, UK WC1E 6BT

deyu.ming.16@ucl.ac.uk

(D.M.)

Department of Civil, Environmental and Geomatic Engineering

University College London

London, England, UK WC1E 6BT

(C.H., C.G.)

Department of Actuarial Mathematics and Statistics

Heriot-Watt University

Edinburgh, Scotland, UK EH14 4AS

(G.W.P.)

Table 1

Parameter values chosen for the assumed true ground-motion model

Parameter	**Value**	**Parameter**	**Value**
b_1	1.0416	b_8	0.0153
b_2	0.9133	b_9	-0.0419
b_3	-0.0814	b_{10}	0.0802
b_4	-2.9273	τ^2	0.0099
b_5	0.2812	σ^2	0.0681
b_6	7.8664	$h\,(\nu = 0.5)^*$	11.50 km
b_7	0.0875	$h\,(\nu = 1.5)^\dagger$	12.58 km

* The range parameter h in the correlation function (52) (i.e., Matérn type with $\nu = 0.5$).

† The range parameter h in the correlation function (53) (i.e., Matérn type with $\nu = 1.5$).

Table 2

Comparison of the estimation performance between the multi-stage algorithm and the Scoring estimation approach

	Multi-Stage Algorithm*				Scoring Estimation Approach			
	$\nu = 0.5$†		$\nu = 1.5$‡		$\nu = 0.5$		$\nu = 1.5$	
	RMSE§	CR‖	RMSE	CR	RMSE	CR	RMSE	CR
b_1	2.5540	94.8	2.8160	95.7	2.5156	94.4	2.6551	92.8
b_2	0.8875	94.0	0.9795	94.8	0.8749	94.0	0.9234	92.8
b_3	0.0780	93.6	0.0862	94.0	0.0769	93.6	0.0811	92.3
b_4	0.3184	98.3	0.3529	99.9	0.3013	94.4	0.3071	95.0
b_5	0.0573	98.4	0.0634	99.8	0.0541	93.9	0.0551	94.7
b_6	0.8631	96.6	0.9250	89.7	0.8438	95.9	0.8092	94.3
b_7	0.0158	93.3	0.0055	80.3	0.0154	95.3	0.0054	94.5
b_8	0.0087	91.6	0.0017	83.3	0.0085	94.3	0.0016	96.2
b_9	0.0661	92.4	0.0723	92.9	0.0649	92.4	0.0651	92.8
b_{10}	0.0712	91.2	0.0740	92.9	0.0701	91.0	0.0683	92.7
τ^2	0.0052	51.3	0.0076	26.5	0.0034	88.9	0.0035	89.2
σ^2	0.0197	1.6	0.0790	0.0	0.0025	94.2	0.0026	94.9
h	8.6122	0.2	9.8763	0.0	0.7582	93.7	0.3773	94.3

* Jayaram and Baker (2010).

† Corresponding to the correlation function (52) (i.e., Matérn type with $\nu = 0.5$) with $h = 11.50$ km.

‡ Corresponding to the correlation function (53) (i.e., Matérn type with $\nu = 1.5$) with $h = 12.58$ km.

§ Root mean squared error of the corresponding parameter estimator.

‖ Coverage rate (in percentage and rounded to one decimal place) of the corresponding parameter.

List of Figure Captions

Figure 1. Flowchart of the multi-stage algorithm proposed by Jayaram and Baker (2010).

Figure 2. (a) The geographical distribution of 62 earthquakes of $5 \leq M_W \leq 6.9$ in Italy from 1976 to 2016. The epicenter of each event is labeled by a filled circle ($\circ$), whose size is scaled by the moment magnitude (M_W) of the event. (b) The distribution of inter-site distance in each earthquake (represented by its corresponding moment magnitude) on a log scale. The color version of this figure is available only in the electronic edition.

Figure 3. Sampling distributions for estimators $\widehat{b}_1, \ldots, \widehat{b}_{10}$ under the correlation function (52) with $h = 11.50$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

Figure 4. Sampling distributions for estimators $\widehat{b}_1, \ldots, \widehat{b}_{10}$ under the correlation function (53) with $h = 12.58$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

Figure 5. Sampling distributions for estimators $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$: (a), (c) and (e) correspond to

the correlation function (52) with $h = 11.50$ km; (b), (d) and (f) correspond to the correlation function (53) with $h = 12.58$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

Figure 6. The region (within a distance of 250 km from the epicenter) of the selected event with ID 'IT-1997-0137'. The epicenter of the event is labeled by a filled star ($\bigstar$); triangles ($\triangle$) represent the recording sites whose logarithmic peak ground acceleration (PGA) records (generated in Section Evaluation of the estimation performance) are observed and used for predictions. The color version of this figure is available only in the electronic edition.

Figure 7. Maps of percentage increases in root mean squared error of predictions (RMSEP) from the multi-stage algorithm relative to those from the Scoring estimation approach at grid points: (a), (c) and (e) correspond to the correlation function (52) with $h = 11.50$ km when $N = 29$, 46, and 62, reading from top to bottom; (b), (d) and (f) correspond to the correlation function (53) with $h = 12.58$ km when $N = 29$, 46, and 62, reading from top to bottom. Triangles ($\triangle$) are recording sites and the dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

Figure 8. Sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$ of ground-motion models with (S) and without (NS) spatial correlation structure. The estimates are obtained from 1000 synthetic datasets generated under the correlation function (52) with $h = 11.50$ km. The left boxplot in each panel corresponds to the ground-motion model with spatial correlation structure;

the right boxplot in each panel corresponds to the ground-motion model without spatial correlation structure. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

Figure 9. Sampling distributions for $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ of ground-motion models with (S) and without (NS) spatial correlation structure (specified by the correlation function (52)). The estimates are obtained from 1000 synthetic datasets generated under the correlation function (52) with $h = 11.50$, 30.00, and 60.00 km, respectively. (a), (c) and (e) correspond to the estimates of τ^2; (b), (d) and (f) correspond to the estimates of σ^2. The color version of this figure is available only in the electronic edition.

Figure 10. Sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$ of ground-motion models with (S) and without (NS) spatial correlation structure. The estimates are obtained from 1000 synthetic datasets generated under the correlation function (53) with $h = 12.58$ km. The left boxplot in each panel corresponds to the ground-motion model with spatial correlation structure; the right boxplot in each panel corresponds to the ground-motion model without spatial correlation structure. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

Figure 11. Sampling distributions for $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ of ground-motion models with (S) and without (NS) spatial correlation structure (specified by the correlation function (53)). The estimates are obtained from 1000 synthetic datasets generated under the correlation function (53) with $h = 12.58$, 32.81, and 65.63 km, respectively. (a), (c) and (e) correspond to the estimates of τ^2; (b), (d) and (f) correspond to the estimates of σ^2. The color version of this figure is available only in the electronic edition.

Figure 12. The region (within a distance of 250 km from the epicenter) of the selected event with ID 'IT-1997-0137', to which artificial recording sites are added. The epicenter of the event is labeled by a filled star (☆); triangles (△) represent the historical recording sites of

the selected event described in Section Evaluation of the predictive performance. Inverted triangles ($\bigtriangledown$) represent the artificial recording sites that are added to the selected event. The observations at both historical and artificial recording sites are used for prediction. The color version of this figure is available only in the electronic edition.

Figure 13. Maps of RMSEP from ground-motion models with and without spatial correlation structure (specified by the correlation function (52)). Ground-motion models are fitted to synthetic datasets generated under the correlation function (52) with $h = 11.50$, 30.00, and 60.00 km. (a), (c) and (e) correspond to the ground-motion model with spatial correlation structure; (b), (d) and (f) correspond to the ground-motion model without spatial correlation structure. Triangles ($\bigtriangleup$) and inverted triangles ($\bigtriangledown$) are historical and artificial recording sites, respectively. The dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

Figure 14. Maps of RMSEP from ground-motion models with and without spatial correlation structure (specified by the correlation function (53)). Ground-motion models are fitted to synthetic datasets generated under the correlation function (53) with $h = 12.58$, 32.81, and 65.63 km. (a), (c) and (e) correspond to the ground-motion model with spatial correlation structure; (b), (d) and (f) correspond to the ground-motion model without spatial correlation structure. Triangles ($\bigtriangleup$) and inverted triangles ($\bigtriangledown$) are historical and artificial recording sites, respectively. The dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

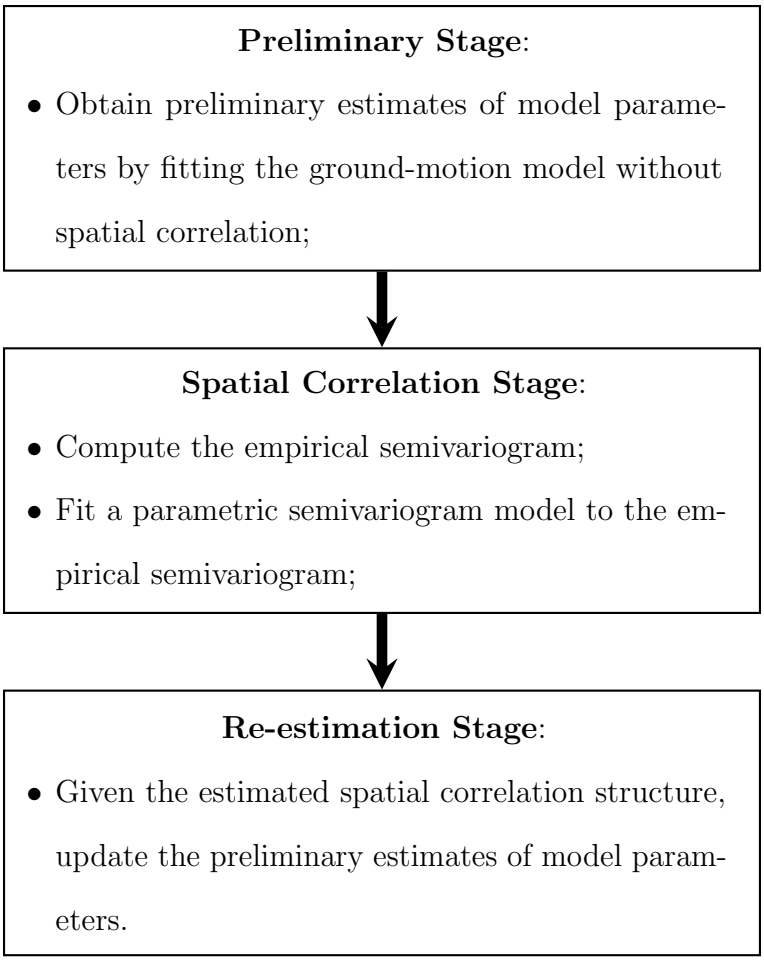

Figure 1. Flowchart of the multi-stage algorithm proposed by Jayaram and Baker (2010).

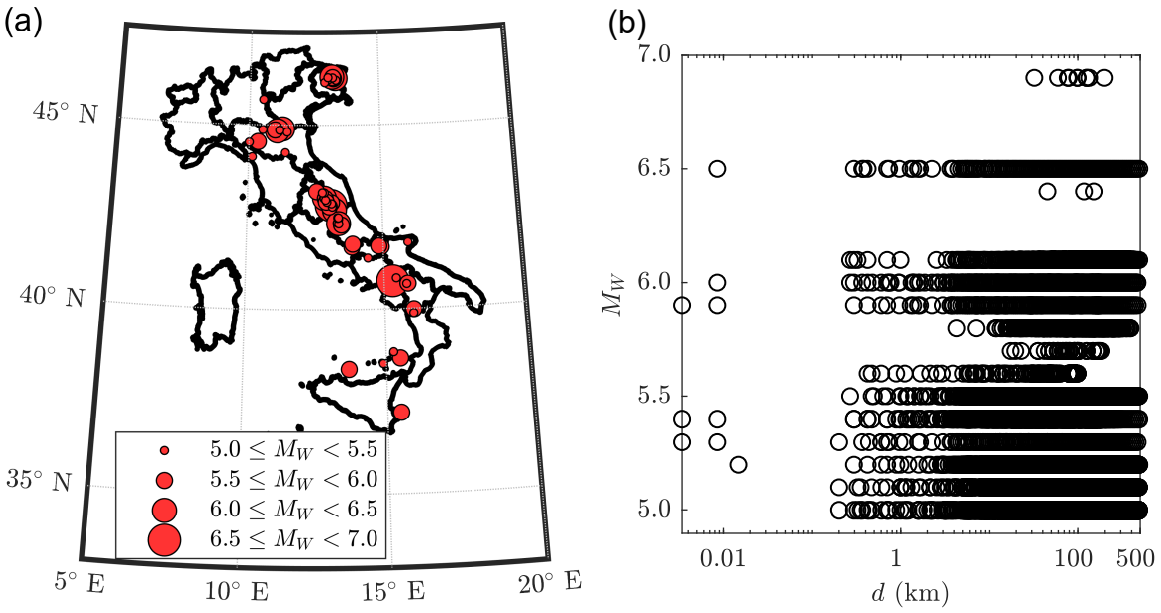

Figure 2. (a) The geographical distribution of 62 earthquakes of $5 \leq M_W \leq 6.9$ in Italy from 1976 to 2016. The epicenter of each event is labeled by a filled circle (∘), whose size is scaled by the moment magnitude (M_W) of the event. (b) The distribution of inter-site distance in each earthquake (represented by its corresponding moment magnitude) on a log scale. The color version of this figure is available only in the electronic edition.

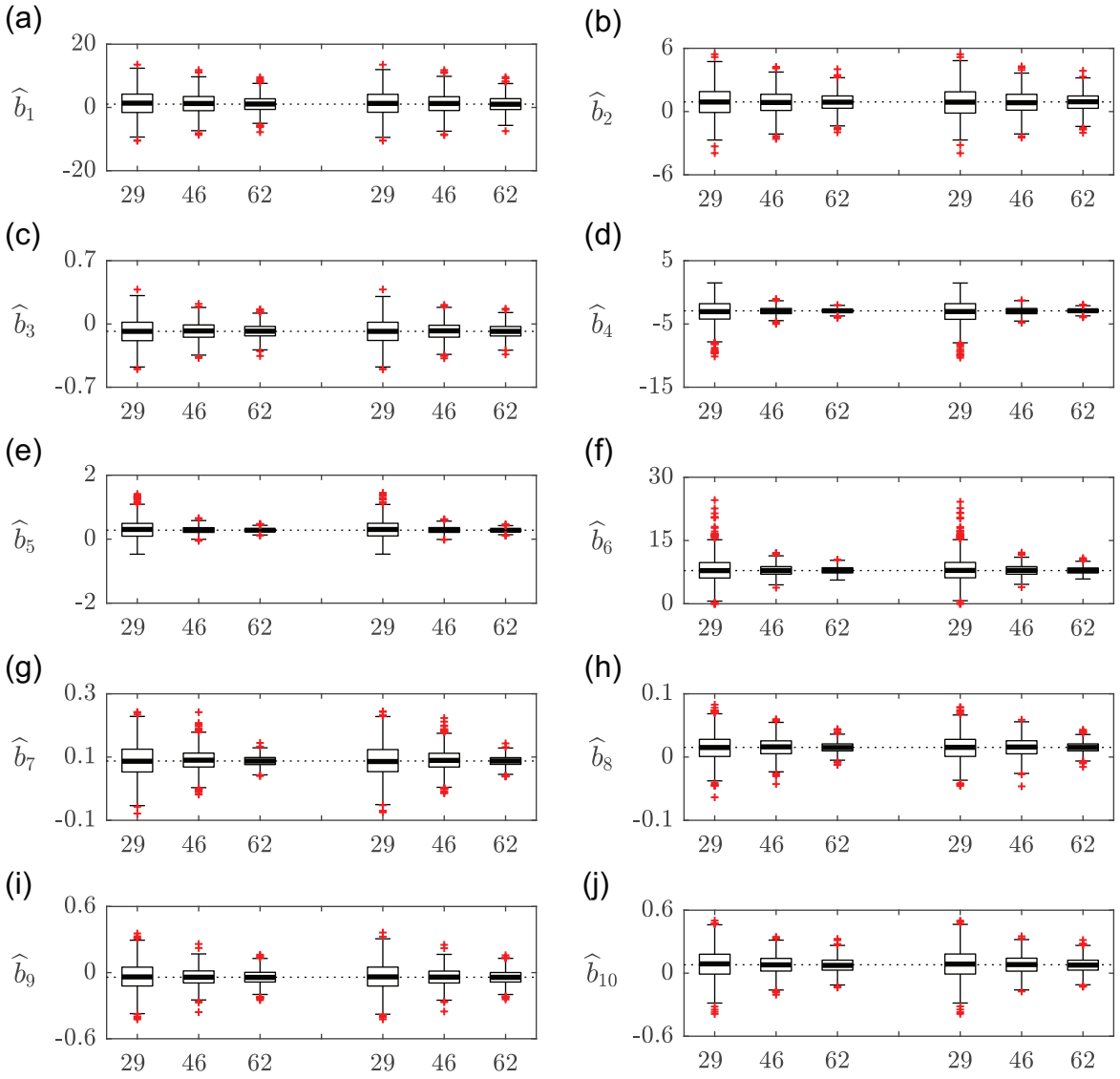

Figure 3. Sampling distributions for estimators $\widehat{b}_1, \ldots, \widehat{b}_{10}$ under the correlation function (52) with $h = 11.50$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

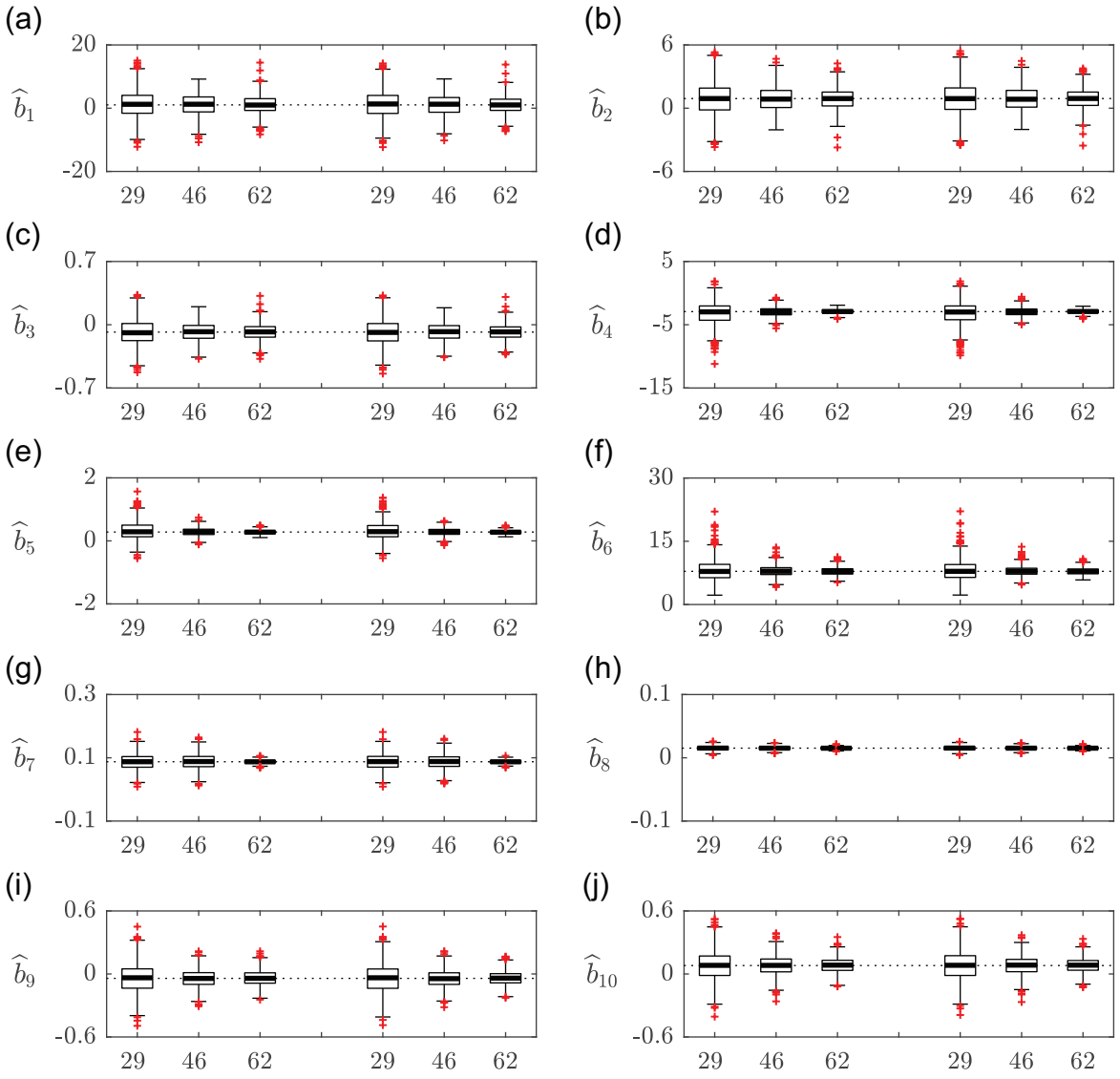

Figure 4. Sampling distributions for estimators $\widehat{b}_1, \ldots, \widehat{b}_{10}$ under the correlation function (53) with $h = 12.58$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

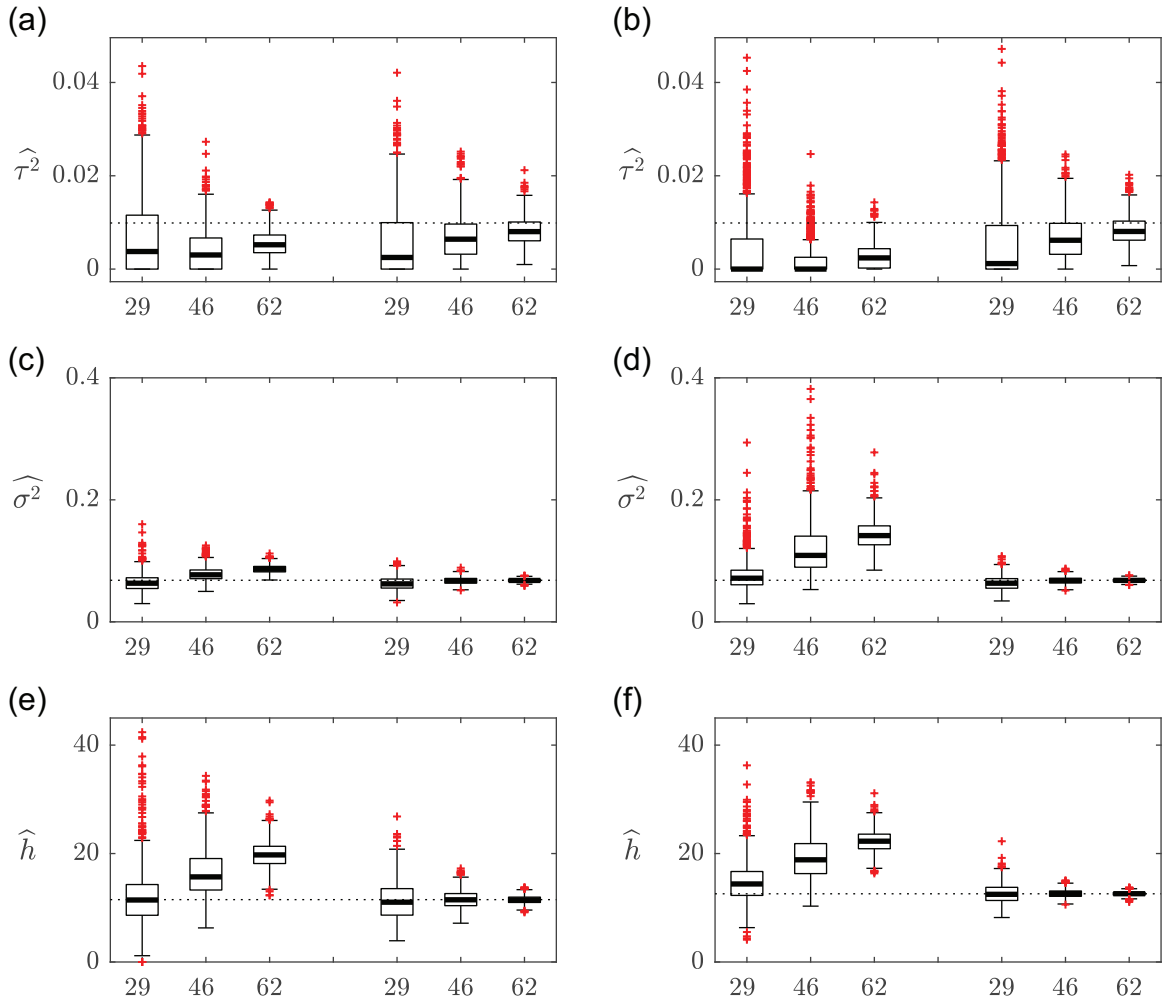

Figure 5. Sampling distributions for estimators $\widehat{\tau^2}$, $\widehat{\sigma^2}$, and $\widehat{h}$: (a), (c) and (e) correspond to the correlation function (52) with $h = 11.50$ km; (b), (d) and (f) correspond to the correlation function (53) with $h = 12.58$ km. The left three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the multi-stage algorithm, respectively; the right three boxplots (reading from left to right) in each panel correspond to event sizes of $N = 29$, 46, and 62 under the Scoring estimation approach, respectively; the three event sizes correspond to events by the end of the year 2000, 2010, and 2016, respectively. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

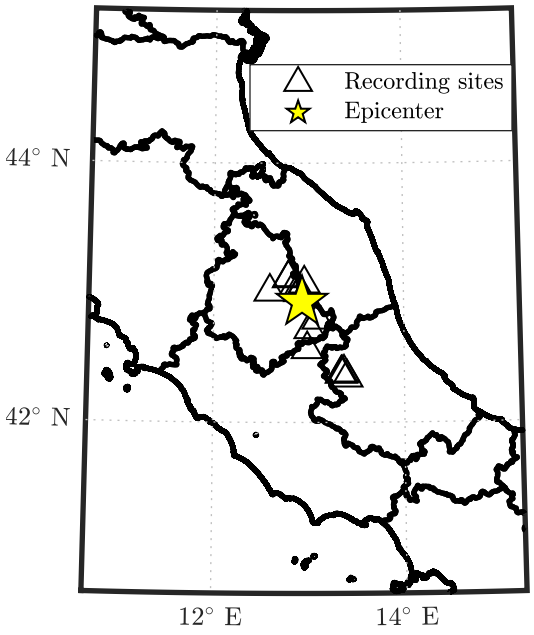

Figure 6. The region (within a distance of 250 km from the epicenter) of the selected event with ID 'IT-1997-0137'. The epicenter of the event is labeled by a filled star (☆); triangles (△) represent the recording sites whose logarithmic peak ground acceleration (PGA) records (generated in Section Evaluation of the estimation performance) are observed and used for predictions. The color version of this figure is available only in the electronic edition.

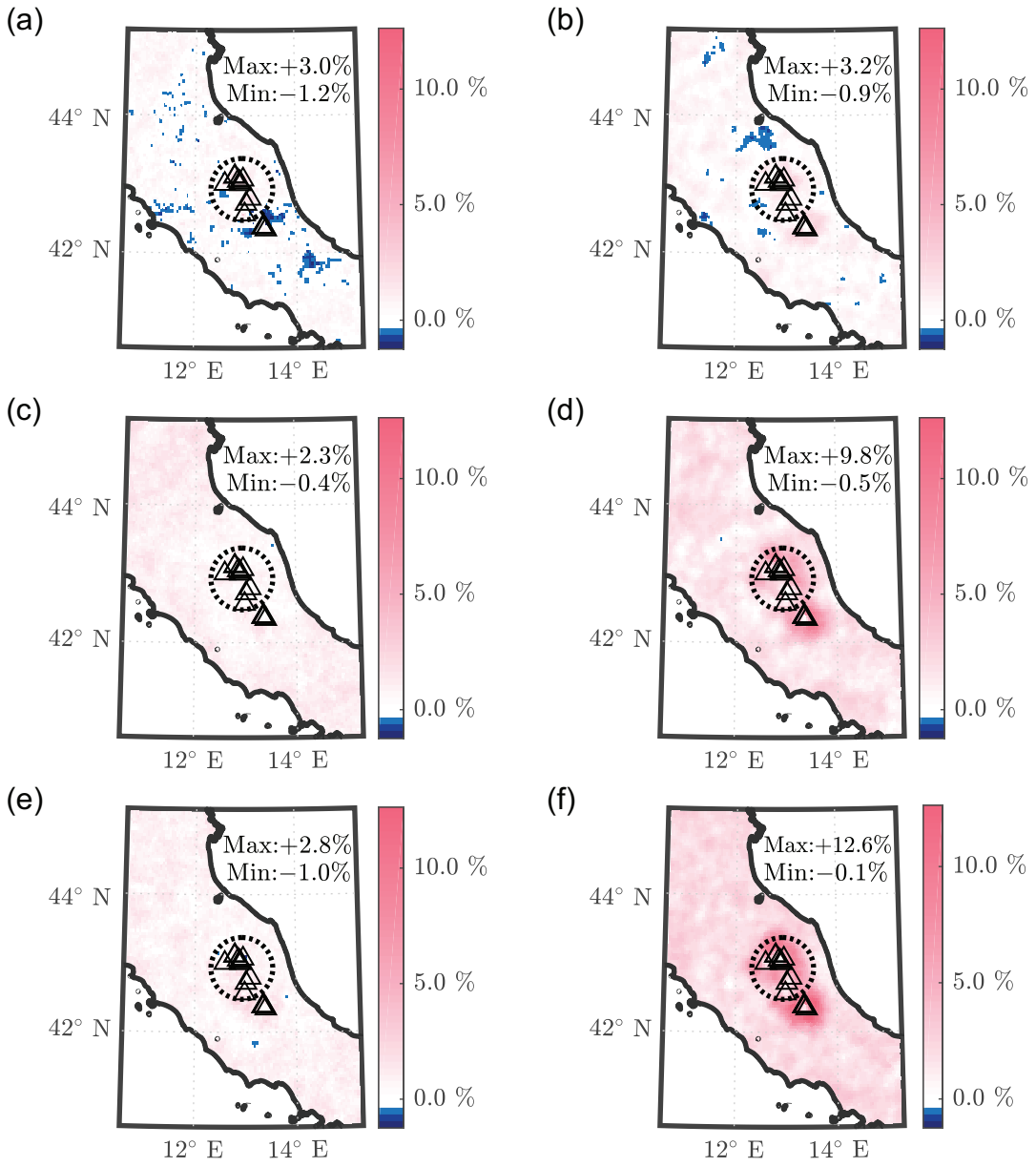

Figure 7. Maps of percentage increases in root mean squared error of predictions (RMSEP) from the multi-stage algorithm relative to those from the Scoring estimation approach at grid points: (a), (c) and (e) correspond to the correlation function (52) with $h = 11.50$ km when $N = 29$, 46, and 62, reading from top to bottom; (b), (d) and (f) correspond to the correlation function (53) with $h = 12.58$ km when $N = 29$, 46, and 62, reading from top to bottom. Triangles ($\triangle$) are recording sites and the dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

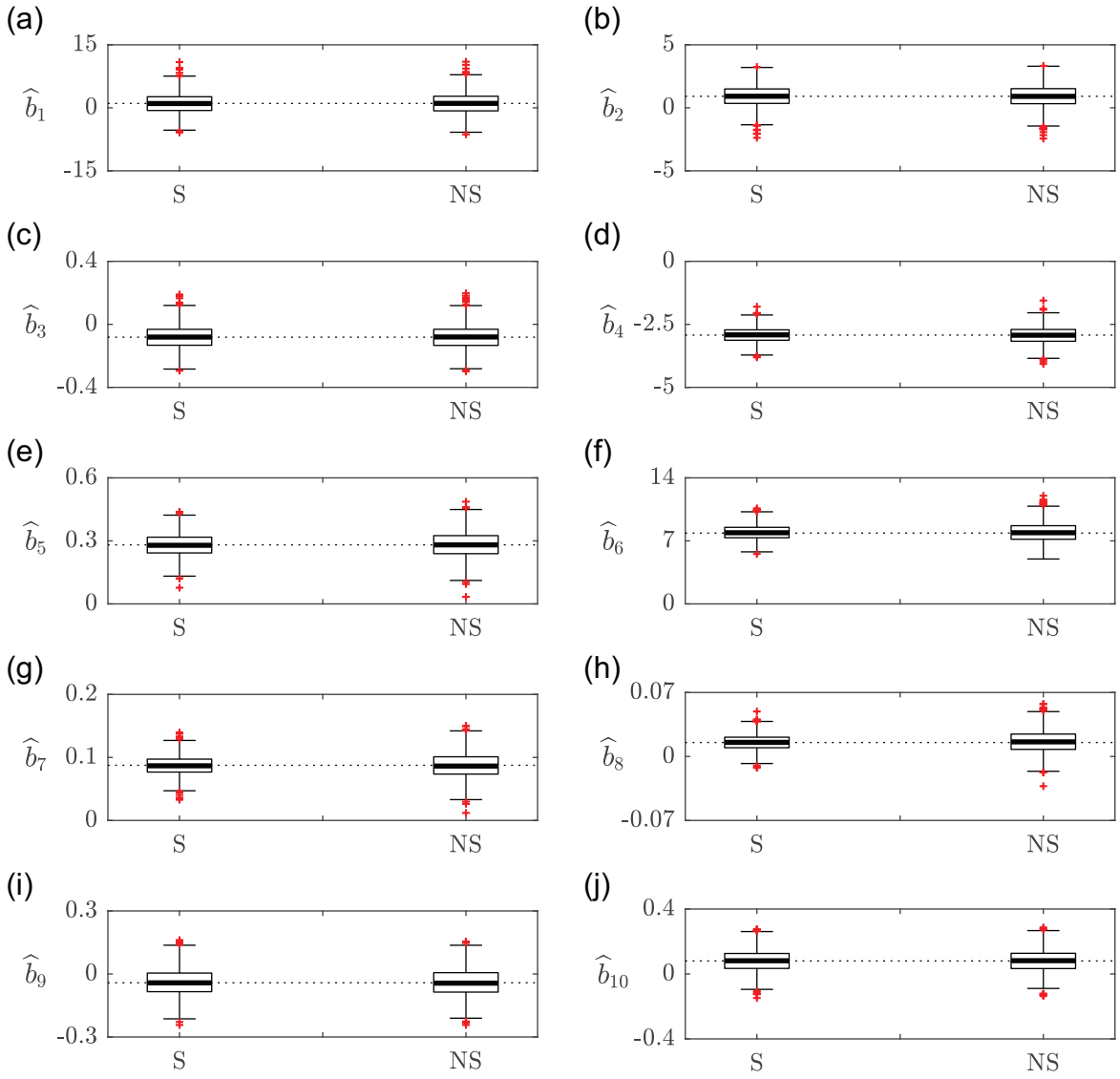

Figure 8. Sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$ of ground-motion models with (S) and without (NS) spatial correlation structure. The estimates are obtained from 1000 synthetic datasets generated under the correlation function (52) with $h = 11.50$ km. The left boxplot in each panel corresponds to the ground-motion model with spatial correlation structure; the right boxplot in each panel corresponds to the ground-motion model without spatial correlation structure. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

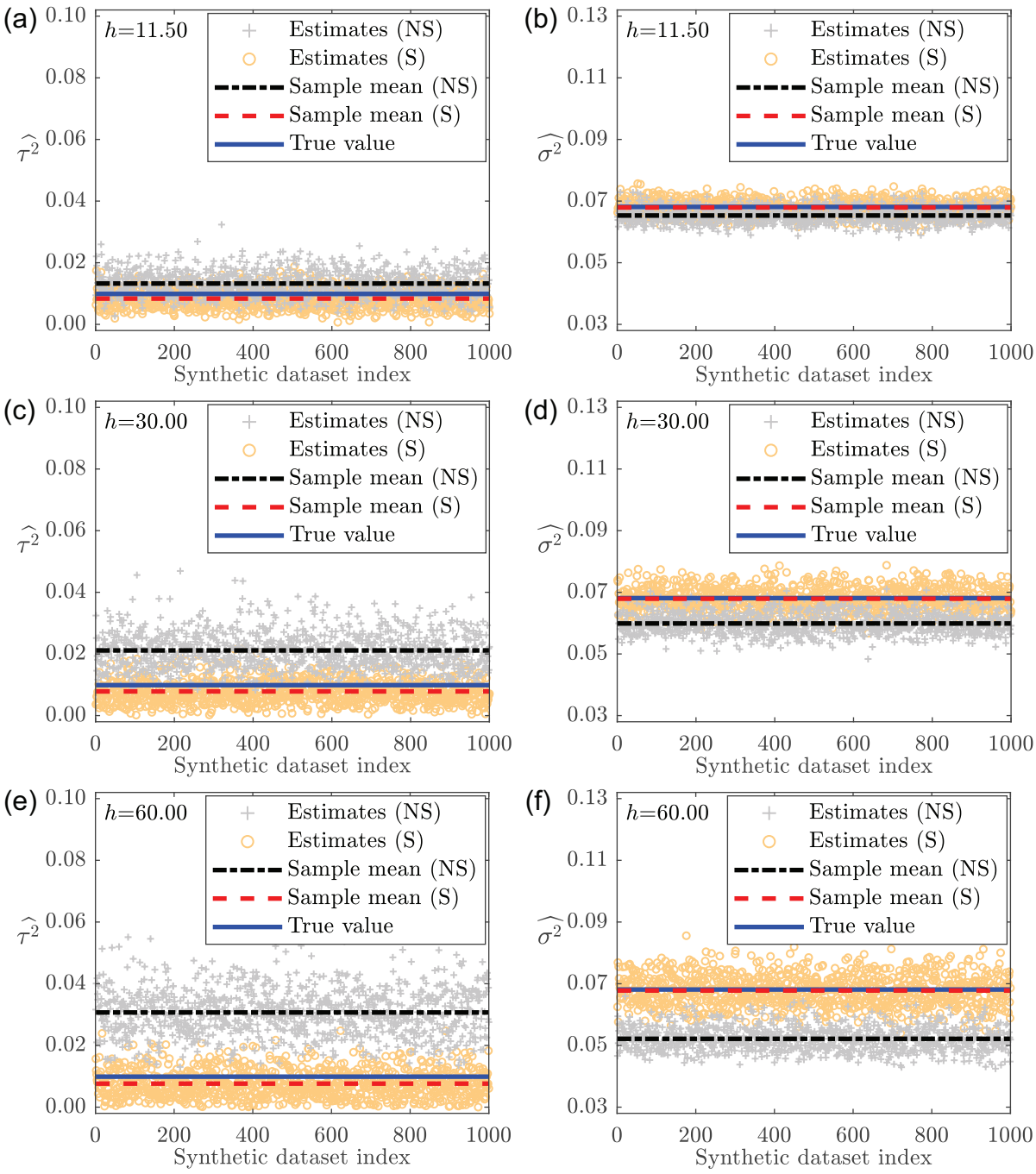

Figure 9. Sampling distributions for $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ of ground-motion models with (S) and without (NS) spatial correlation structure (specified by the correlation function (52)). The estimates are obtained from 1000 synthetic datasets generated under the correlation function (52) with $h = 11.50$, 30.00, and 60.00 km, respectively. (a), (c) and (e) correspond to the estimates of τ^2; (b), (d) and (f) correspond to the estimates of σ^2. The color version of this figure is available only in the electronic edition.

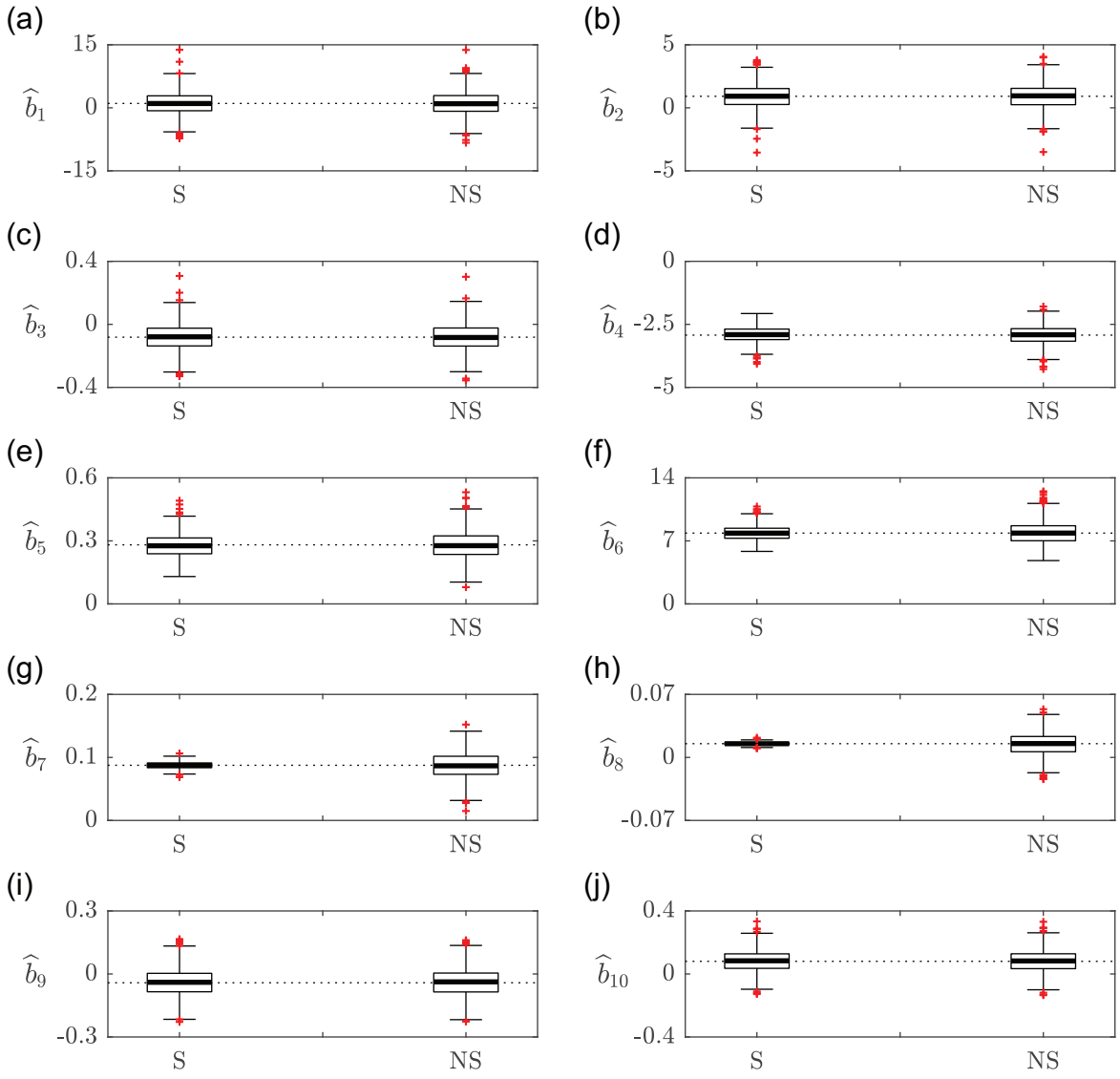

Figure 10. Sampling distributions for $\widehat{b}_1, \ldots, \widehat{b}_{10}$ of ground-motion models with (S) and without (NS) spatial correlation structure. The estimates are obtained from 1000 synthetic datasets generated under the correlation function (53) with $h = 12.58$ km. The left boxplot in each panel corresponds to the ground-motion model with spatial correlation structure; the right boxplot in each panel corresponds to the ground-motion model without spatial correlation structure. The dashed line in each panel represents the true parameter value. The color version of this figure is available only in the electronic edition.

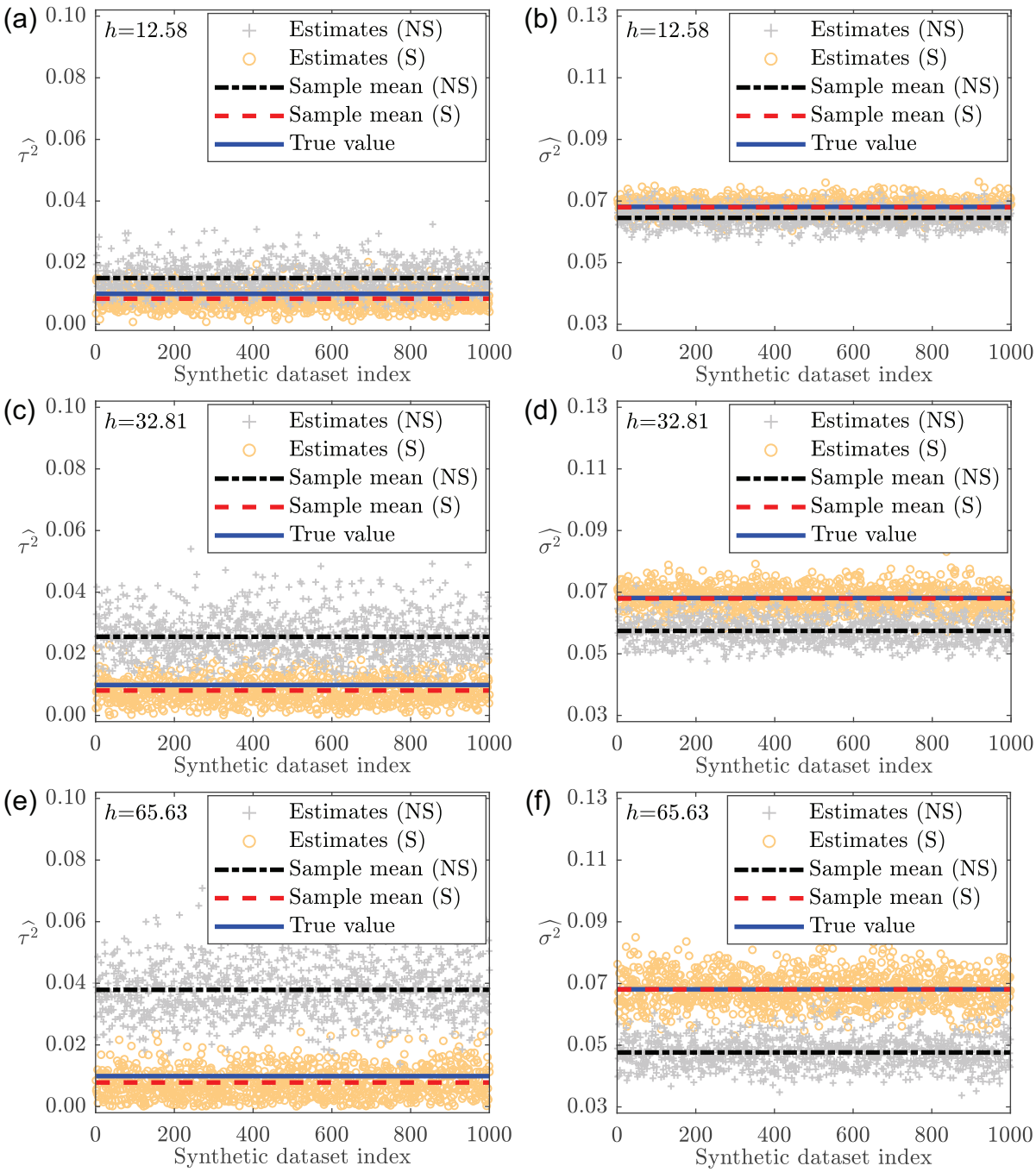

Figure 11. Sampling distributions for $\widehat{\tau^2}$ and $\widehat{\sigma^2}$ of ground-motion models with (S) and without (NS) spatial correlation structure (specified by the correlation function (53)). The estimates are obtained from 1000 synthetic datasets generated under the correlation function (53) with $h = 12.58$, 32.81, and 65.63 km, respectively. (a), (c) and (e) correspond to the estimates of τ^2; (b), (d) and (f) correspond to the estimates of σ^2. The color version of this figure is available only in the electronic edition.

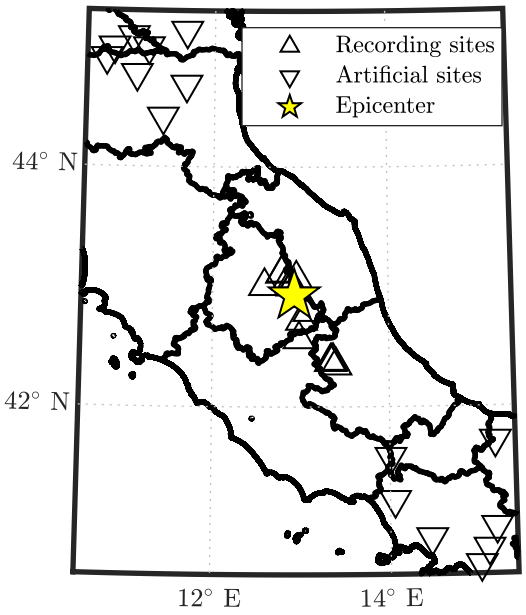

Figure 12. The region (within a distance of 250 km from the epicenter) of the selected event with ID 'IT-1997-0137', to which artificial recording sites are added. The epicenter of the event is labeled by a filled star (☆); triangles (△) represent the historical recording sites of the selected event described in Section Evaluation of the predictive performance. Inverted triangles (▽) represent the artificial recording sites that are added to the selected event. The observations at both historical and artificial recording sites are used for prediction. The color version of this figure is available only in the electronic edition.

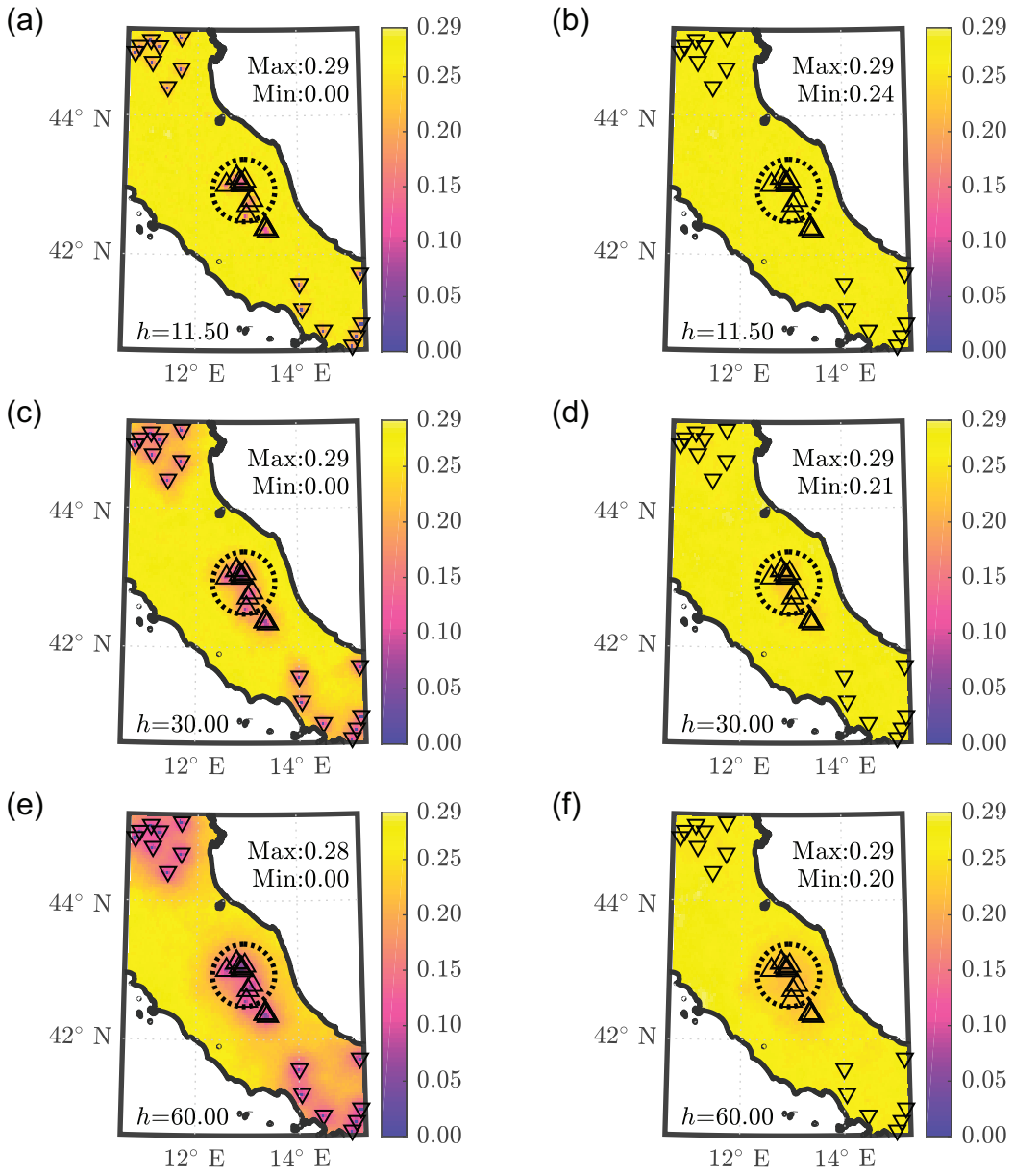

Figure 13. Maps of RMSEP from ground-motion models with and without spatial correlation structure (specified by the correlation function (52)). Ground-motion models are fitted to synthetic datasets generated under the correlation function (52) with $h = 11.50$, 30.00, and 60.00 km. (a), (c) and (e) correspond to the ground-motion model with spatial correlation structure; (b), (d) and (f) correspond to the ground-motion model without spatial correlation structure. Triangles ($\triangle$) and inverted triangles ($\triangledown$) are historical and artificial recording sites, respectively. The dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

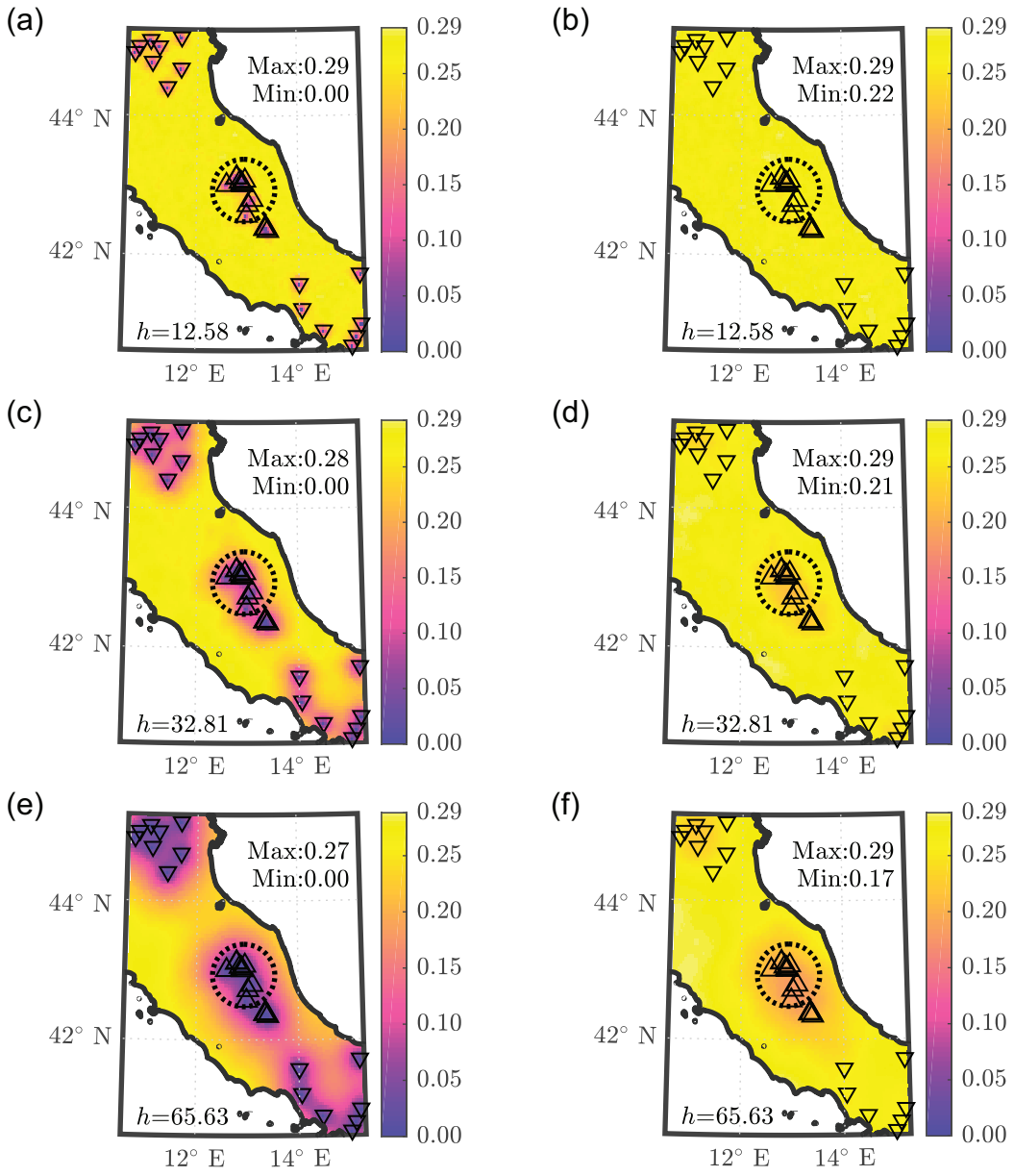

Figure 14. Maps of RMSEP from ground-motion models with and without spatial correlation structure (specified by the correlation function (53)). Ground-motion models are fitted to synthetic datasets generated under the correlation function (53) with $h = 12.58$, 32.81, and 65.63 km. (a), (c) and (e) correspond to the ground-motion model with spatial correlation structure; (b), (d) and (f) correspond to the ground-motion model without spatial correlation structure. Triangles ($\triangle$) and inverted triangles ($\triangledown$) are historical and artificial recording sites, respectively. The dashed circle defines the border of the near-field (within 50 km from the epicenter). The color version of this figure is available only in the electronic edition.

Appendix

This appendix details the dimension reduction in the Scoring estimation approach that is emphasized in Section Numerical considerations.

Let $\boldsymbol{\alpha} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ and denote the gradient $\mathbf{S}(\boldsymbol{\alpha})$ and expected information matrix $\mathbf{I}(\boldsymbol{\alpha})$ of $l(\boldsymbol{\alpha})$ by the partitions

$$\mathbf{S}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha}) \\ \mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) \\ \mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \end{bmatrix} \tag{A.1}$$

and

$$\mathbf{I}(\boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{I}_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\gamma\beta}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\gamma\theta}}(\boldsymbol{\alpha}) \\ \mathbf{I}_{\boldsymbol{\beta\gamma}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\beta\beta}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\beta\theta}}(\boldsymbol{\alpha}) \\ \mathbf{I}_{\boldsymbol{\theta\gamma}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\theta\beta}}(\boldsymbol{\alpha}) & \mathbf{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\alpha}) \end{bmatrix} . \tag{A.2}$$

Then, the maximum likelihood estimate of $\boldsymbol{\alpha}$ can be obtained by the Scoring estimation approach with dimension reduction using the following updating equations (the proof is available in the Proof of Equations (A.3)–(A.11) of the Main Article section of the electronic supplement to this article):

$$\widehat{\boldsymbol{\gamma}}^{(k+1)} = \widehat{\boldsymbol{\gamma}}^{(k)} + \left(\mathbf{I}_{\boldsymbol{\gamma\gamma}}(\widehat{\boldsymbol{\alpha}}^{(k)}) - \mathbf{I}_{\boldsymbol{\gamma\beta}}(\widehat{\boldsymbol{\alpha}}^{(k)})\, \mathbf{I}^{-1}_{\boldsymbol{\beta\beta}}(\widehat{\boldsymbol{\alpha}}^{(k)})\, \mathbf{I}_{\boldsymbol{\beta\gamma}}(\widehat{\boldsymbol{\alpha}}^{(k)})\right)^{-1} \mathbf{S}_{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\alpha}}^{(k)}) , \tag{A.3}$$

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \mathbf{I}^{-1}_{\boldsymbol{\theta\theta}}(\widehat{\boldsymbol{\alpha}}^{(k)})\, \mathbf{S}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}^{(k)}) , \tag{A.4}$$

$$\widehat{\boldsymbol{\beta}}^{(k+1)} = \mathbf{I}^{-1}_{\boldsymbol{\beta\beta}}(\widehat{\boldsymbol{\alpha}}^{(k)}) \left[\mathbf{g}^\top(\mathbf{X}, \widehat{\boldsymbol{\gamma}}^{(k+1)}) \mathbf{C}^{-1} \left(\widehat{\boldsymbol{\theta}}^{(k+1)}\right) \mathbf{Y}\right] , \tag{A.5}$$

in which $\mathbf{g}(\mathbf{X}, \boldsymbol{\gamma}) = \left(\mathbf{g}(\mathbf{X}_1, \boldsymbol{\gamma})^\top, \ldots, \mathbf{g}(\mathbf{X}_N, \boldsymbol{\gamma})^\top\right)^\top$ and

- the i-th element of $\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha})]_i = \left[\frac{\partial \mathbf{g}(\mathbf{X}, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i} \boldsymbol{\beta}\right]^\top \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{g}(\mathbf{X}, \boldsymbol{\gamma})\boldsymbol{\beta}] ; \tag{A.6}$$

- The i-th element of $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})]_i = -\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} + \frac{1}{2}[\mathbf{Y}-\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]^\top \mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y}-\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]\,; \tag{A.7}$$

- $\mathbf{I}_{\boldsymbol{\beta\beta}}(\boldsymbol{\alpha})$ is given by

$$\mathbf{I}_{\boldsymbol{\beta\beta}}(\boldsymbol{\alpha}) = \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})^\top \mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\,; \tag{A.8}$$

- the ij-th element of $\mathbf{I}_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I}_{\boldsymbol{\gamma\gamma}}(\boldsymbol{\alpha})]_{ij} = \left[\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i}\boldsymbol{\beta}\right]^\top \mathbf{C}^{-1}(\boldsymbol{\theta})\,\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_j}\boldsymbol{\beta}\,; \tag{A.9}$$

- the ij-th element of $\mathbf{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\alpha})]_{ij} = \frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\right\}\,; \tag{A.10}$$

- the i-th row of $\mathbf{I}_{\boldsymbol{\gamma\beta}}(\boldsymbol{\alpha})$ $\big($or the i-th column of $\mathbf{I}_{\boldsymbol{\beta\gamma}}(\boldsymbol{\alpha})\big)$ is given by

$$[\mathbf{I}_{\boldsymbol{\gamma\beta}}(\boldsymbol{\alpha})]_{i*} = [\mathbf{I}_{\boldsymbol{\beta\gamma}}(\boldsymbol{\alpha})]_{*i}^\top = \left[\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i}\boldsymbol{\beta}\right]^\top \mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\,. \tag{A.11}$$

It can be seen from equation (A.3)-(A.5), that after separating the linear and nonlinear parameters in ground-motion prediction functions via decomposition in equation (47), the Scoring estimation approach amounts to three updating equations in each iteration. The updating equation (A.5) for $\boldsymbol{\beta}$ has an analytical form given the estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ obtained from updating equations (A.3) and (A.4). The further separation of the update scheme caused by the isolation between linear and nonlinear parameters reduces the complexity of each iteration from $\mathcal{O}(p_1^3+p_2^3)$ (in the ordinary Scoring estimation approach) to $\mathcal{O}(p_{11}^3+p_{12}^3+p_2^3+p_{11}^2p_{12}+p_{12}^2p_{11})$, in which $p_{11}+p_{12}=p_1$ and p_{11} and p_{12} are dimensions of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. Another advantage of the dimension reduction in the Scoring estimation approach is that the conditioning of the algorithm is improved because of the further separation. Finally, the Scoring estimation approach with dimension reduction only requires initial values of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ to be set because the initial value $\widehat{\boldsymbol{\beta}}^{(1)}$ of $\boldsymbol{\beta}$ can be obtained by (A.5) using $\widehat{\boldsymbol{\gamma}}^{(1)}$ and

$\widehat{\boldsymbol{\theta}}^{(1)}$. Consequently, the convergence criterion is only required for $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, implying that the convergence may be achieved with fewer iterations. Define

$$\mathbf{I}_{-\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\alpha}) - \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\beta}}(\boldsymbol{\alpha})\,\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\alpha})\,\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\gamma}}(\boldsymbol{\alpha}) \tag{A.12}$$

and apply block matrix inversion on equation (43), the asymptotic standard error estimates of $\widehat{\boldsymbol{\gamma}}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$ are then given by

$$\widehat{\mathrm{se}}(\widehat{\boldsymbol{\gamma}}) = \sqrt{\mathrm{diag}\left[\mathbf{I}^{-1}_{-\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\right]}, \tag{A.13}$$

$$\widehat{\mathrm{se}}(\widehat{\boldsymbol{\beta}}) = \sqrt{\mathrm{diag}\left[\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right) + \mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\mathbf{I}^{-1}_{-\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\right]}, \tag{A.14}$$

$$\widehat{\mathrm{se}}(\widehat{\boldsymbol{\theta}}) = \sqrt{\mathrm{diag}\left[\mathbf{I}^{-1}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\alpha}}^{(K)}\right)\right]}. \tag{A.15}$$

The Algorithm A1 outlines the implementation procedure for the Scoring estimation approach with dimension reduction.

Algorithm A1 Scoring estimation approach with dimension reduction

Input: $\mathbf{Y}_i$, $\mathbf{X}_{ij}$ and $\mathbf{s}_{ij}$ for $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, n_i\}$.

Output: Estimates of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ with corresponding asymptotic standard error estimates.

1: Initialization:

   1) choose values for $\widehat{\boldsymbol{\gamma}}^{(1)}$ and $\widehat{\boldsymbol{\theta}}^{(1)}$;

   2) compute the value of $\widehat{\boldsymbol{\beta}}^{(1)}$ by equation (A.5);

2: **repeat**

3: Update the estimates of $\boldsymbol{\alpha} = (\boldsymbol{\gamma}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\theta}^\top)^\top$ by equation (A.3) to (A.5);

4: **until** the convergence criterion is met;

5: Obtain the asymptotic standard error estimate of $\widehat{\boldsymbol{\alpha}}$ by equation (A.13) to (A.15).

Electronic Supplement to An Advanced Estimation Algorithm for Ground-Motion Models with Spatial Correlation

This electronic supplement presents the proofs associated with the statements in the main article. These proofs are for the relation between the semivariogram and the correlation function, the connection between the re-estimation procedure in Jayaram and Baker (2010) and the expectation–maximization (EM) algorithm, the updating equations for the Scoring estimation approach, and the updating equations for the Scoring estimation approach with dimension reduction.

The proofs in the electronic supplement are presented in the following four sections: (1) the proof of equation (19) of the main article, (2) an alternative interpretation of the re-estimation procedure of Jayaram and Baker (2010) based on the EM algorithm of Jayaram and Baker (2010) based on the EM algorithm, (3) the proof for equations (36)–(41) of the main article, and (4) the proof of equations (A3)–(A11) of the main article.

Proof of Equation (19)

In this section, we prove the equation (19):

$$\gamma(d) = 1 - k(d) \tag{S.1}$$

in the manuscript.

The semivariogram of $\widetilde{\boldsymbol{\varepsilon}}$ is defined by

$$\gamma(\widetilde{\varepsilon}_{ij},\, \widetilde{\varepsilon}_{ij'}) = \frac{1}{2}\mathrm{var}(\widetilde{\varepsilon}_{ij} - \widetilde{\varepsilon}_{ij'})\,. \tag{S.2}$$

Then, we have

$$\begin{aligned}
\gamma(\widetilde{\varepsilon}_{ij},\, \widetilde{\varepsilon}_{ij'}) =& \frac{1}{2}\mathbb{E}\left[\left(\frac{\varepsilon_{ij}}{\sigma} - \frac{\varepsilon_{ij'}}{\sigma}\right)^2\right] \\
=& \frac{1}{2\sigma^2}\mathbb{E}\left[(\varepsilon_{ij} - \varepsilon_{ij'})^2\right] \\
=& \frac{1}{2\sigma^2}\left(\mathbb{E}[\varepsilon_{ij}^2] + \mathbb{E}[\varepsilon_{ij'}^2] - 2\mathbb{E}[\varepsilon_{ij}\varepsilon_{ij'}]\right) \\
=& \frac{1}{2\sigma^2}\mathrm{var}(\varepsilon_{ij}) + \frac{1}{2\sigma^2}\mathrm{var}(\varepsilon_{ij'}) - \frac{1}{\sigma^2}\mathrm{cov}(\varepsilon_{ij},\, \varepsilon_{ij'}) \\
=& 1 - k(\mathbf{s}_{ij},\, \mathbf{s}_{ij'})\,.
\end{aligned} \tag{S.3}$$

Since the correlation function is stationary and isotropic, we have

$$k(\mathbf{s}_{ij},\, \mathbf{s}_{ij'}) = k(d_{i,jj'}) \tag{S.4}$$

with $d_{i,jj'} = \|\mathbf{s}_{ij} - \mathbf{s}_{ij'}\|_2$. Thus, the semivariogram of $\widetilde{\boldsymbol{\varepsilon}}$ is a function of $d_{i,jj'}$:

$$\begin{aligned}
\gamma(\widetilde{\varepsilon}_{ij},\, \widetilde{\varepsilon}_{ij'}) &= \gamma(d_{i,jj'}) \\
&= 1 - k(d_{i,jj'})\,.
\end{aligned} \tag{S.5}$$

Then, for all site pairs $(j,\, j')$ such that $d_{i,jj'} = d$ we have

$$\gamma(d) = 1 - k(d)\,. \tag{S.6}$$

Alternative Construction of the Re-Estimation Procedure

In this section, we show how to reconstruct the re-estimation procedure based on the idea of the EM algorithm. This alternative construction of the re-estimation procedure is useful because it allows us to demonstrate that the re-estimation procedure increases $l(\sigma^2, \tau^2, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})$ at each iteration and understand the properties of the re-estimation procedure through the well-studied EM algorithm.

Treating the random effects $\eta_{i=1,\ldots,N}$ as unobservable, at iteration $k+1$ we first increase $l(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})$ with respect to $\mathbf{b}$ via one Expectation-Maximization (EM) step, which consists of an E-step and a M-step:

- **E-step:** find the expected log-likelihood function

$$Q(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) = \sum_{i=1}^{N} \mathbb{E}\left[l_i^F(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \right], \tag{S.7}$$

  where the expectation is taken with respect to $\eta_{i=1,\ldots,N}$ conditional on $\mathbf{Y}_{i=1,\ldots,N}$ and estimates $\widehat{\sigma^2}^{(k)}$, $\widehat{\tau^2}^{(k)}$ and $\widehat{\mathbf{b}}^{(k)}$; and

$$\begin{aligned}
& l_i^F(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \mathbf{b}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \\
&= \ln f(\mathbf{Y}_i|\eta_i) f(\eta_i)\big|_{\sigma^2=\widehat{\sigma^2}^{(k)}, \tau^2=\widehat{\tau^2}^{(k)}, \boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}} \\
&\propto -\frac{1}{2}\ln \widehat{\tau^2}^{(k)} - \frac{1}{2}\ln|\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})| - \frac{1}{2\widehat{\tau^2}^{(k)}}\eta_i^2 \\
&\quad - \frac{1}{2\widehat{\sigma^2}^{(k)}}[\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{b}) - \eta_i \mathbf{1}_{n_i}]^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}})[\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{b}) - \eta_i \mathbf{1}_{n_i}]\,;
\end{aligned} \tag{S.8}$$

- **M-step:** obtain the estimate $\widehat{\mathbf{b}}^{(k+1)}$ such that

$$Q(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k+1)}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \geq Q(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k)}|\boldsymbol{\omega} = \widehat{\boldsymbol{\omega}})\,. \tag{S.9}$$

Up to a constant, the expected log-likelihood function can be written as

$$\begin{aligned}
&Q(\widehat{\sigma^2}^{(k)},\, \widehat{\tau^2}^{(k)},\, \mathbf{b}|\boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}) \\
\propto &-\frac{N}{2}\ln\widehat{\tau^2}^{(k)} - \frac{1}{2}\sum_{i=1}^{N}\ln|\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})| \\
&-\frac{1}{2\widehat{\tau^2}^{(k)}}\sum_{i=1}^{N}\widehat{\eta_i^2} - \frac{1}{2\widehat{\sigma^2}^{(k)}}\sum_{i=1}^{N}\mathrm{tr}\left\{\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}})\mathbf{V}_i\right\} \\
&-\frac{1}{2\widehat{\sigma^2}^{(k)}}\sum_{i=1}^{N}[\mathbf{Y}_i-\mathbf{f}(\mathbf{X}_i,\,\mathbf{b})-\widehat{\eta}_i\mathbf{1}_{n_i}]^\top\boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}})[\mathbf{Y}_i-\mathbf{f}(\mathbf{X}_i,\,\mathbf{b})-\widehat{\eta}_i\mathbf{1}_{n_i}]\,,
\end{aligned} \tag{S.10}$$

where

$$\mathbf{V}_i = \mathrm{var}(\eta_i|\mathbf{Y}_i,\, \widehat{\sigma^2}^{(k)},\, \widehat{\tau^2}^{(k)},\, \widehat{\mathbf{b}}^{(k)},\, \boldsymbol{\omega}=\widehat{\boldsymbol{\omega}})\mathbf{1}_{n_i\times n_i}\,, \tag{S.11}$$

$$\widehat{\eta_i^2} = \mathbb{E}[\eta_i^2|\mathbf{Y}_i,\, \widehat{\sigma^2}^{(k)},\, \widehat{\tau^2}^{(k)},\, \widehat{\mathbf{b}}^{(k)},\, \boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}] \tag{S.12}$$

and

$$\widehat{\eta}_i = \mathbb{E}[\eta_i|\mathbf{Y}_i,\, \widehat{\sigma^2}^{(k)},\, \widehat{\tau^2}^{(k)},\, \widehat{\mathbf{b}}^{(k)},\, \boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}]\,. \tag{S.13}$$

Note that

$$\begin{aligned}
\widehat{\eta}_i =&\mathbb{E}[\eta_i|\mathbf{Y}_i,\, \widehat{\sigma^2}^{(k)},\, \widehat{\tau^2}^{(k)},\, \widehat{\mathbf{b}}^{(k)},\, \boldsymbol{\omega}=\widehat{\boldsymbol{\omega}}] \\
=&\widehat{\tau^2}^{(k)}\mathbf{1}_{n_i}^\top\left(\widehat{\tau^2}^{(k)}\mathbf{1}_{n_i\times n_i}+\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}[\mathbf{Y}_i-\mathbf{f}(\mathbf{X}_i,\,\widehat{\mathbf{b}}^{(k)})]\,,
\end{aligned} \tag{S.14}$$

where the second equality is given by the formula for the expectation of the conditional multivariate normal distribution (Flury, 2013). Also note that

$$\begin{aligned}
&\left(\widehat{\tau^2}^{(k)}\mathbf{1}_{n_i\times n_i}+\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1} \\
=&\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1} \\
&-\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}\mathbf{1}_{n_i}\left(\frac{1}{\widehat{\tau^2}^{(k)}}+\mathbf{1}_{n_i}^\top\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}\mathbf{1}_{n_i}\right)^{-1}\mathbf{1}_{n_i}^\top\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1} \\
=&\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}-\frac{\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}\mathbf{1}_{n_i}\mathbf{1}_{n_i}^\top\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}}{\frac{1}{\widehat{\tau^2}^{(k)}}+\mathbf{1}_{n_i}^\top\left(\widehat{\sigma^2}^{(k)}\boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}\mathbf{1}_{n_i}}\,,
\end{aligned} \tag{S.15}$$

where the first step uses the Woodbury identity (Petersen and Pedersen, 2012). Plugging equation (S.15) into (S.14), we have

$$\begin{aligned}\widehat{\eta_i} &= \frac{\mathbf{1}_{n_i}^\top \left(\widehat{\sigma^2}^{(k)} \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1}}{\frac{1}{\widehat{\tau^2}^{(k)}} + \mathbf{1}_{n_i}^\top \left(\widehat{\sigma^2}^{(k)} \boldsymbol{\Omega}_i(\widehat{\boldsymbol{\omega}})\right)^{-1} \mathbf{1}_{n_i}} [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \widehat{\mathbf{b}}^{(k)})] \\ &= \frac{\frac{1}{\widehat{\sigma^2}^{(k)}} \mathbf{1}_{n_i}^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) \left[\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \widehat{\mathbf{b}}^{(k)})\right]}{\frac{1}{\widehat{\tau^2}^{(k)}} + \frac{1}{\widehat{\sigma^2}^{(k)}} \mathbf{1}_{n_i}^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) \mathbf{1}_{n_i}}, \end{aligned} \tag{S.16}$$

which equals to equation (24) in step 3 of Algorithm 1 (re-estimation procedure) in the manuscript.

In M-step (corresponding to step 4 in the Algorithm 1) we obtain the estimate $\mathbf{b}^{(k+1)}$ by solving the generalized least squares problem:

$$\widehat{\mathbf{b}}^{(k+1)} = \arg\min \sum_{i=1}^{N} [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{b}) - \widehat{\eta}_i \mathbf{1}_{n_i}]^\top \boldsymbol{\Omega}_i^{-1}(\widehat{\boldsymbol{\omega}}) [\mathbf{Y}_i - \mathbf{f}(\mathbf{X}_i, \mathbf{b}) - \widehat{\eta}_i \mathbf{1}_{n_i}]. \tag{S.17}$$

Then we have

$$Q(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k+1)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \geq Q(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \tag{S.18}$$

and subsequently, by monotonicity one obtains that

$$l(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k+1)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \geq l(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}). \tag{S.19}$$

Finally, we obtain estimates $\widehat{\sigma^2}^{(k+1)}$ and $\widehat{\tau^2}^{(k+1)}$ by solving

$$(\widehat{\sigma^2}^{(k+1)}, \widehat{\tau^2}^{(k+1)}) = \arg\max l\left(\sigma^2, \tau^2 \middle| \mathbf{b} = \widehat{\mathbf{b}}^{(k+1)}, \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}\right), \tag{S.20}$$

which is the step 5 in the Algorithm 1 and implies that

$$l(\widehat{\sigma^2}^{(k+1)}, \widehat{\tau^2}^{(k+1)}, \widehat{\mathbf{b}}^{(k+1)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \geq l(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k+1)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}), \tag{S.21}$$

and by inequality (S.19)

$$l(\widehat{\sigma^2}^{(k+1)}, \widehat{\tau^2}^{(k+1)}, \widehat{\mathbf{b}}^{(k+1)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}) \geq l(\widehat{\sigma^2}^{(k)}, \widehat{\tau^2}^{(k)}, \widehat{\mathbf{b}}^{(k)} | \boldsymbol{\omega} = \widehat{\boldsymbol{\omega}}). \tag{S.22}$$

Proof of Equation (36) - (41)

The elements of gradient $\mathbf{S}(\boldsymbol{\alpha})$ and expected information matrix $\mathbf{I}(\boldsymbol{\alpha})$ can be calculated as follow:

- the i-th element of $\mathbf{S}_{\mathbf{b}}(\boldsymbol{\alpha})$:

$$\begin{aligned}
\left[\mathbf{S}_{\mathbf{b}}(\boldsymbol{\alpha})\right]_i =& \frac{\partial l(\boldsymbol{\alpha})}{\partial \mathbf{b}_i} \\
=& \frac{1}{2}\left\{ \left[\frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})] + [\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta}) \left[\frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_i}\right] \right\} \\
=& \left[\frac{\partial \mathbf{f}(\mathbf{X},\, \mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]\,, \qquad \text{(S.23)}
\end{aligned}$$

  where the last equality uses the fact that the transpose of a scalar is the same scalar;

- the i-th element of $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$:

$$\begin{aligned}
\left[\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})\right]_i =& \frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}_i} \\
=& -\frac{1}{2}\frac{1}{\left|\mathbf{C}(\boldsymbol{\theta})\right|}\frac{\partial \left|\mathbf{C}(\boldsymbol{\theta})\right|}{\partial \boldsymbol{\theta}_i} - \frac{1}{2}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^{\top}\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})] \\
=& -\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} - \frac{1}{2}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^{\top}\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})] \\
=& -\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} \\
&+ \frac{1}{2}[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{f}(\mathbf{X},\, \mathbf{b})]\,, \qquad \text{(S.24)}
\end{aligned}$$

  where the third and last steps use the following two matrix derivative identities:

$$\frac{\partial \left|\mathbf{C}(\boldsymbol{\theta})\right|}{\partial \boldsymbol{\theta}_i} = \left|\mathbf{C}(\boldsymbol{\theta})\right|\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} \qquad \text{(S.25)}$$

  and

$$\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} = -\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta}) \qquad \text{(S.26)}$$

  respectively from Petersen and Pedersen (2012);

- the ij-th element of $\mathbf{I_{bb}}(\boldsymbol{\alpha})$:

$$
\begin{aligned}
[\mathbf{I_{bb}}(\boldsymbol{\alpha})]_{ij} =& \mathbb{E}\left[\frac{\partial l(\boldsymbol{\alpha})}{\partial \mathbf{b}_i}\frac{\partial l(\boldsymbol{\alpha})}{\partial \mathbf{b}_j}\right] \\
=& \mathbb{E}\left\{\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_j}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]\right\} \\
=& \mathbb{E}\left\{\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})][\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_j}\right]\right\} \\
=& \left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\,\mathbb{E}\left\{[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})][\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\right\}\mathbf{C}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_j}\right] \\
=& \left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\,\mathbf{C}(\boldsymbol{\theta})\,\mathbf{C}^{-1}(\boldsymbol{\theta})\left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_j}\right] \\
=& \left[\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_i}\right]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\,\frac{\partial \mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial \mathbf{b}_j}\,; \qquad \text{(S.27)}
\end{aligned}
$$

- the ij-th element of $\mathbf{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\alpha})$:

$$
\begin{aligned}
[\mathbf{I}_{\boldsymbol{\theta\theta}}(\boldsymbol{\alpha})]_{ij} =& \mathbb{E}\left[\frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}_i}\frac{\partial l(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}_j}\right] \\
=& -\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right] \\
=& \frac{1}{2}\mathbb{E}\left[\mathrm{tr}\left\{\frac{\partial \mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2 \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right\}\right] \\
& -\frac{1}{2}\mathbb{E}\left[[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\frac{\partial}{\partial \boldsymbol{\theta}_j}\left(\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\right)[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]\right] \\
=& \frac{1}{2}\mathrm{tr}\left\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2 \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right\} \\
& -\frac{1}{2}\mathbb{E}\left[\mathrm{tr}\left\{[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\frac{\partial}{\partial \boldsymbol{\theta}_j}\left(\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\right)[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]\right\}\right] \\
=& \frac{1}{2}\mathrm{tr}\left\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2 \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right\} \\
& -\frac{1}{2}\mathbb{E}\left[\mathrm{tr}\left\{\frac{\partial}{\partial \boldsymbol{\theta}_j}\left(\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\right)[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})][\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\right\}\right] \\
=& \frac{1}{2}\mathrm{tr}\left\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2 \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right\} \\
& -\frac{1}{2}\mathrm{tr}\left\{\frac{\partial}{\partial \boldsymbol{\theta}_j}\left(\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\right)\mathbb{E}\left\{[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})][\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]^{\top}\right\}\right\} \\
=& \frac{1}{2}\mathrm{tr}\left\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2 \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i\partial \boldsymbol{\theta}_j}\right\}
\end{aligned}
$$

$$
\begin{aligned}
&-\frac{1}{2}\mathrm{tr}\left\{\frac{\partial}{\partial\boldsymbol{\theta}_j}\left(\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\right)\mathbf{C}(\boldsymbol{\theta})\right\}\\
=&\frac{1}{2}\mathrm{tr}\left\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\right\}\\
&-\frac{1}{2}\mathrm{tr}\Bigg\{-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}+\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial^2\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i\partial\boldsymbol{\theta}_j}\\
&-\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\Bigg\}\\
=&\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial\mathbf{C}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\right\};
\end{aligned}
\tag{S.28}
$$

- the ij-th element of $\mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}(\boldsymbol{\alpha})$:

$$
\begin{aligned}
[\mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}(\boldsymbol{\alpha})]_{ij}=&\mathbb{E}\left[\frac{\partial l(\boldsymbol{\alpha})}{\partial\mathbf{b}_i}\frac{\partial l(\boldsymbol{\alpha})}{\partial\boldsymbol{\theta}_j}\right]\\
=&-\mathbb{E}\left[\frac{\partial^2 l(\boldsymbol{\alpha})}{\partial\mathbf{b}_i\partial\boldsymbol{\theta}_j}\right]\\
=&-\mathbb{E}\left\{\left[\frac{\partial\mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial\mathbf{b}_i}\right]^{\top}\frac{\partial\mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})]\right\}\\
=&-\left[\frac{\partial\mathbf{f}(\mathbf{X},\,\mathbf{b})}{\partial\mathbf{b}_i}\right]^{\top}\frac{\partial\mathbf{C}^{-1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j}\mathbb{E}\left[\mathbf{Y}-\mathbf{f}(\mathbf{X},\,\mathbf{b})\right]\\
=&0\,;
\end{aligned}
\tag{S.29}
$$

- Since $\mathbf{I}_{\boldsymbol{\theta}\mathbf{b}}(\boldsymbol{\alpha})$ and $\mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}(\boldsymbol{\alpha})$ are symmetric, we have

$$
\mathbf{I}_{\boldsymbol{\theta}\mathbf{b}}(\boldsymbol{\alpha})=\mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}^{\top}(\boldsymbol{\alpha})=\mathbf{0}\,.
\tag{S.30}
$$

Replacing $\mathbf{I}_{\boldsymbol{\theta}\mathbf{b}}(\widehat{\boldsymbol{\alpha}}^{(k)})$ and $\mathbf{I}_{\mathbf{b}\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}^{(k)})$ by $\mathbf{0}$ in equation (30) of the manuscript gives the updating equations (36) and (37) in the manuscript.

Proof of Equation (A.3) - (A.11)

Denote by the following vector the model parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$, according to

$$\mathbf{a} = \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\theta} \end{bmatrix} . \tag{S.31}$$

Then, we have $l(\boldsymbol{\alpha}) = l(\mathbf{a}, \boldsymbol{\beta})$, and given fixed $\mathbf{a}$ the log-likelihood function $l(\mathbf{a}, \boldsymbol{\beta})$ is maximized when

$$\boldsymbol{\beta} = \left[\mathbf{g}^\top(\mathbf{X}, \boldsymbol{\gamma})\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X}, \boldsymbol{\gamma})\right]^{-1} \left[\mathbf{g}^\top(\mathbf{X}, \boldsymbol{\gamma})\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{Y}\right] . \tag{S.32}$$

This allows the profile log-likelihood function

$$M(\mathbf{a}) = l(\mathbf{a}, h(\mathbf{a})) , \tag{S.33}$$

where

$$h(\mathbf{a}) \stackrel{\text{def}}{=} \left[\mathbf{g}^\top(\mathbf{X}, \boldsymbol{\gamma})\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X}, \boldsymbol{\gamma})\right]^{-1} \left[\mathbf{g}^\top(\mathbf{X}, \boldsymbol{\gamma})\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{Y}\right] . \tag{S.34}$$

Taking first order derivative with respect to $\mathbf{a}$ on both sides of (S.33), we have

$$\begin{aligned} \frac{\partial M(\mathbf{a})}{\partial \mathbf{a}} =& \left.\frac{\partial l(\mathbf{a}, \boldsymbol{\beta})}{\partial \mathbf{a}}\right|_{\boldsymbol{\beta}=h(\mathbf{a})} + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \left[\frac{\partial l(\mathbf{a}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]_{\boldsymbol{\beta}=h(\mathbf{a})} \\ =& \left.\frac{\partial l(\mathbf{a}, \boldsymbol{\beta})}{\partial \mathbf{a}}\right|_{\boldsymbol{\beta}=h(\mathbf{a})} , \end{aligned} \tag{S.35}$$

where the last equality uses the fact that $\boldsymbol{\beta} = h(\mathbf{a})$ is the solution of

$$\frac{\partial l(\mathbf{a}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0} . \tag{S.36}$$

Now if we evaluate $\mathbf{a}$ at its estimate $\widehat{\mathbf{a}}^{(k)}$ at iteration k, we have

$$\widehat{\boldsymbol{\beta}}^{(k)} = h\left(\widehat{\mathbf{a}}^{(k)}\right) . \tag{S.37}$$

Then from equality (S.35), we obtain that

$$\left.\frac{\partial M(\mathbf{a})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} = \left.\frac{\partial l(\mathbf{a}, \boldsymbol{\beta})}{\partial \mathbf{a}}\right|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}, \boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}} . \tag{S.38}$$

Denote the Score function of $M(\mathbf{a})$ by

$$\mathbf{S}_M(\mathbf{a}) = \frac{\partial M(\mathbf{a})}{\partial \mathbf{a}}\,. \tag{S.39}$$

Then, we have from equality (S.38) that

$$\mathbf{S}_M(\widehat{\mathbf{a}}^{(k)}) = \mathbf{S}_{\mathbf{a}}(\widehat{\boldsymbol{\alpha}}^{(k)})\,. \tag{S.40}$$

Since

$$\left.\frac{\partial l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right|_{\boldsymbol{\beta}=h(\mathbf{a})} = \mathbf{0}\,, \tag{S.41}$$

taking derivative with respect to $\mathbf{a}$ gives

$$\left.\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \mathbf{a}\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=h(\mathbf{a})} + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}^\top}\right]_{\boldsymbol{\beta}=h(\mathbf{a})} = \mathbf{0}\,. \tag{S.42}$$

Taking expectation on both sides of (S.42), we have

$$\mathbb{E}\left[\left.\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \mathbf{a}\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=h(\mathbf{a})}\right] + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \mathbb{E}\left[\left.\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}^\top}\right|_{\boldsymbol{\beta}=h(\mathbf{a})}\right] = \mathbf{0}\,. \tag{S.43}$$

Evaluating $\mathbf{a}$ at $\widehat{\mathbf{a}}^{(k)}$ with

$$\widehat{\boldsymbol{\beta}}^{(k)} = h\left(\widehat{\mathbf{a}}^{(k)}\right)\,, \tag{S.44}$$

we obtain

$$\mathbb{E}\left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \mathbf{a}\partial \boldsymbol{\beta}^\top}\right]_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)},\,\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}} + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \Bigg|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} \mathbb{E}\left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\,\partial \boldsymbol{\beta}^\top}\right]_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)},\,\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}} = \mathbf{0}\,. \tag{S.45}$$

Thus,

$$\mathbf{I}_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \Bigg|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) = \mathbf{0}\,, \tag{S.46}$$

which gives

$$\left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \Bigg|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} = -\mathbf{I}_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\left[\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\right]^{-1} \tag{S.47}$$

Taking derivative with respect to $\mathbf{a}$ on both sides of (S.35), we have

$$\frac{\partial^2 M(\mathbf{a})}{\partial \mathbf{a}\partial \mathbf{a}^\top} = \left.\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \mathbf{a}\partial \mathbf{a}^\top}\right|_{\boldsymbol{\beta}=h(\mathbf{a})} + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\,\partial \mathbf{a}^\top}\right]_{\boldsymbol{\beta}=h(\mathbf{a})} \tag{S.48}$$

Taking expectation on both sides of (S.48) and evaluating $\mathbf{a}$ at $\widehat{\mathbf{a}}^{(k)}$ with

$$\widehat{\boldsymbol{\beta}}^{(k)} = h\left(\widehat{\mathbf{a}}^{(k)}\right)\,, \tag{S.49}$$

we obtain

$$\mathbb{E}\left[\frac{\partial^2 M(\mathbf{a})}{\partial \mathbf{a}\partial \mathbf{a}^\top}\right]_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} = \mathbb{E}\left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \mathbf{a}\partial \mathbf{a}^\top}\right]_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)},\,\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}} + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \Bigg|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} \mathbb{E}\left[\frac{\partial^2 l(\mathbf{a},\,\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\,\partial \mathbf{a}^\top}\right]_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)},\,\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}^{(k)}} . \tag{S.50}$$

Denote the expected information matrix of $M(\mathbf{a})$ by

$$\mathbf{I}_M(\mathbf{a}) = -\mathbb{E}\left[\frac{\partial^2 M(\mathbf{a})}{\partial \mathbf{a}\partial \mathbf{a}^\top}\right] . \tag{S.51}$$

Then, equation (S.50) gives

$$\mathbf{I}_M\left(\widehat{\mathbf{a}}^{(k)}\right) = \mathbf{I}_{\mathbf{aa}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) + \left(\frac{\partial h(\mathbf{a})}{\partial \mathbf{a}^\top}\right)^\top \Bigg|_{\mathbf{a}=\widehat{\mathbf{a}}^{(k)}} \mathbf{I}_{\boldsymbol{\beta}\mathbf{a}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) . \tag{S.52}$$

Plugging (S.47) into (S.52), we have

$$\mathbf{I}_M\left(\widehat{\mathbf{a}}^{(k)}\right) = \mathbf{I}_{\mathbf{aa}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) - \mathbf{I}_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta}\mathbf{a}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) . \tag{S.53}$$

The Scoring update scheme to find the estimate of $\mathbf{a}$ that maximizes $M(\mathbf{a})$ is then given by

$$\begin{aligned}\widehat{\mathbf{a}}^{(k+1)} &= \widehat{\mathbf{a}}^{(k)} + \mathbf{I}_M^{-1}\left(\widehat{\mathbf{a}}^{(k)}\right)\mathbf{S}_M\left(\widehat{\mathbf{a}}^{(k)}\right) \\ &= \widehat{\mathbf{a}}^{(k)} + \left[\mathbf{I}_{\mathbf{aa}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) - \mathbf{I}_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}^{-1}_{\boldsymbol{\beta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta}\mathbf{a}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\right]^{-1}\mathbf{S}_{\mathbf{a}}(\widehat{\boldsymbol{\alpha}}^{(k)}) ,\end{aligned} \tag{S.54}$$

where

$$\mathbf{I}_{\mathbf{aa}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) = \begin{bmatrix}\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) & \mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \\ \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\end{bmatrix} , \tag{S.55}$$

$$\mathbf{I}_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) = \begin{bmatrix}\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \\ \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\end{bmatrix} , \tag{S.56}$$

$$\mathbf{I}_{\boldsymbol{\beta}\mathbf{a}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) = \mathbf{I}^\top_{\mathbf{a}\boldsymbol{\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \tag{S.57}$$

and

$$\mathbf{S}_{\mathbf{a}}(\widehat{\boldsymbol{\alpha}}^{(k)}) = \begin{bmatrix}\mathbf{S}_{\boldsymbol{\gamma}}(\widehat{\boldsymbol{\alpha}}^{(k)}) \\ \mathbf{S}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\alpha}}^{(k)})\end{bmatrix} . \tag{S.58}$$

By elementary calculations analogous to those used in the third section of this supplement, we have that

- the i-th element of $\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{S}_{\boldsymbol{\gamma}}(\boldsymbol{\alpha})]_i = \left[\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i}\boldsymbol{\beta}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]\,; \tag{S.59}$$

- The i-th element of $\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$ is given by

$$\begin{aligned}[\mathbf{S}_{\boldsymbol{\theta}}(\boldsymbol{\alpha})]_i = &-\frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\right\} \\ &+\frac{1}{2}[\mathbf{Y} - \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]\,;\end{aligned} \tag{S.60}$$

- $\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is given by

$$\mathbf{S}_{\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})[\mathbf{Y} - \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\boldsymbol{\beta}]\,; \tag{S.61}$$

- the ij-th element of $\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\boldsymbol{\alpha})]_{ij} = \left[\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i}\boldsymbol{\beta}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})\,\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_j}\boldsymbol{\beta}\,; \tag{S.62}$$

- the ij-th element of $\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\alpha})$ is given by

$$[\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\boldsymbol{\alpha})]_{ij} = \frac{1}{2}\mathrm{tr}\left\{\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i}\mathbf{C}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{C}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j}\right\}\,; \tag{S.63}$$

- $\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\alpha})$ is given by

$$\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})^{\top}\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\,; \tag{S.64}$$

- the i-th row of $\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\beta}}(\boldsymbol{\alpha})$ $\big($or the i-th column of $\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\gamma}}(\boldsymbol{\alpha})\big)$ is given by

$$[\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\beta}}(\boldsymbol{\alpha})]_{i*} = [\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\gamma}}(\boldsymbol{\alpha})]_{*i}^{\top} = \left[\frac{\partial \mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_i}\boldsymbol{\beta}\right]^{\top} \mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{g}(\mathbf{X},\,\boldsymbol{\gamma})\,; \tag{S.65}$$

- $\mathbf{I}_{\boldsymbol{\gamma}\boldsymbol{\theta}}(\boldsymbol{\alpha}) = \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\gamma}}^{\top}(\boldsymbol{\alpha}) = \mathbf{0}\,;$

- $\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\beta}}(\boldsymbol{\alpha}) = \mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\theta}}^{\top}(\boldsymbol{\alpha}) = \mathbf{0}\,.$

Replacing $\mathbf{I}_{\boldsymbol{\gamma\theta}}(\boldsymbol{\alpha})$, $\mathbf{I}_{\boldsymbol{\theta\gamma}}(\boldsymbol{\alpha})$, $\mathbf{I}_{\boldsymbol{\theta\beta}}(\boldsymbol{\alpha})$ and $\mathbf{I}_{\boldsymbol{\beta\theta}}(\boldsymbol{\alpha})$ by $\mathbf{0}$ in (S.54), we obtain

$$\begin{bmatrix} \widehat{\boldsymbol{\gamma}}^{(k+1)} \\ \widehat{\boldsymbol{\theta}}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \widehat{\boldsymbol{\gamma}}^{(k)} \\ \widehat{\boldsymbol{\theta}}^{(k)} \end{bmatrix} + \begin{bmatrix} \left[\mathbf{I}_{\boldsymbol{\gamma\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) - \mathbf{I}_{\boldsymbol{\gamma\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta\beta}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\right]^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{\boldsymbol{\theta\theta}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \end{bmatrix} \begin{bmatrix} \mathbf{S}_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \\ \mathbf{S}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \end{bmatrix}, \tag{S.66}$$

which yields

$$\widehat{\boldsymbol{\gamma}}^{(k+1)} = \widehat{\boldsymbol{\gamma}}^{(k)} + \left[\mathbf{I}_{\boldsymbol{\gamma\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) - \mathbf{I}_{\boldsymbol{\gamma\beta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta\beta}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{I}_{\boldsymbol{\beta\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\right]^{-1}\mathbf{S}_{\boldsymbol{\gamma}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right) \tag{S.67}$$

$$\widehat{\boldsymbol{\theta}}^{(k+1)} = \widehat{\boldsymbol{\theta}}^{(k)} + \mathbf{I}_{\boldsymbol{\theta\theta}}^{-1}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right)\mathbf{S}_{\boldsymbol{\theta}}\left(\widehat{\boldsymbol{\alpha}}^{(k)}\right). \tag{S.68}$$

Plugging estimates $\widehat{\boldsymbol{\gamma}}^{(k+1)}$ and $\widehat{\boldsymbol{\theta}}^{(k+1)}$ in (S.67) and (S.68) into (S.32), we obtain the updating equation for the estimate of $\boldsymbol{\beta}$, which concludes the proof.

References

Flury, B. (2013). *A First Course in Multivariate Statistics*, Springer, New York, New York, 715 pp.

Jayaram, N. and J. W. Baker (2010). Considering spatial correlation in mixed-effects regression and the impact on ground-motion models, *Bull. Seismol. Soc. Am.* **100**, no. 6, 3295–3303.

Petersen, K. B. and M. S. Pedersen (2012). *The Matrix Cookbook*, Technical University of Denmark, Lyngby, Denmark, 72 pp.