A COMPARISON OF NONERGODIC GROUND-MOTION MODELS BASED ON GEOGRAPHICALLY WEIGHTED REGRESSION AND THE INTEGRATED NESTED LAPLACE APPROXIMATION

NICOLAS KUEHN^{*1}

¹*University of California, Los Angeles*

September 15, 2021

Abstract

Different nonergodic Ground-Motion Models based on spatially varying coefficient models are compared for ground-motion data in Italy. The models are based different methodologies: Multi-source geographically weighted regression ([Caramenti et al., 2020](#)), and Bayesian hierarchical models estimated with the integrated nested Laplace approximation ([Rue et al., 2009](#)). The different models are compared in terms of their predictive performance, their spatial coefficients, and their predictions. Models that include spatial terms perform slightly better than a simple base model that includes only event and station terms, in terms of out-of sample error based on cross-validation. The Bayesian spatial models have slightly lower generalization error, which can be attributed to the fact that they can include random effects for events and stations. The different methodologies give rise to different dependencies of the spatially varying terms on event and station locations, leading to between-model uncertainty in their predictions, which should be accommodated in a nonergodic seismic hazard assessment.

^{*}kuehn@ucla.edu

1 Introduction

In recent years, there has been a push to relax the ergodic assumption ([Anderson and Brune, 1999](#)) in the development of empirical ground-motion models (GMMs). Loosely speaking, the ergodic assumption states that one can trade time for space, and that the ground-motion distribution at a particular site is the same as the ground-motion distribution averaged over many sites. This means that an ergodic GMM ignores repeatable systematic source, path, and site effects. Due to increasing numbers of observed data, it is possible to estimate these effects. Much has been written about nonergodic GMMs and their use in probabilistic seismic hazard analysis (PSHA). For a schematic overview of nonergodic PSHA, see e.g. [Villani and Abrahamson \(2015\)](#) or [Walling and Abrahamson \(2012\)](#). [Abrahamson et al. \(2019\)](#) conducted a nonergodic PSHA for three sites in California, and showed that there can be large differences in PSHA results due to the inclusion of nonergodic effects on ground motion.

[Landwehr et al. \(2016\)](#) introduced the concept of spatially varying coefficient models (SVCs, [Gelfand et al., 2003](#)) for the estimation of nonergodic GMMs. In an SVC, several coefficients are functions of location, thus encoding spatial variation of ground-motion scaling. The model of [Landwehr et al. \(2016\)](#) is an example of a Bayesian hierarchical model, with the covariance of the random effects modeled by a Gaussian process (GP, [Rasmussen and Williams, 2006](#)). The GP can be understood as a prior distribution over the functional form of the spatial dependence of the coefficients, with the covariance function providing some constraints on the shape of the function. The GMM of ([Landwehr et al., 2016](#)) was based on the SVC formulation of [Bussas et al. \(2015, 2017\)](#), which relied on marginalization of the random effects. [Kuehn \(2021b\)](#) reformulated the model, exploiting that the number of stations and events is much smaller than the number of records. In this formulation, the size of the covariance matrix for the spatial event and site effects depends only on the number of events and stations, while it depends on the number of records when the effects are integrated out. This formulation is also used in recent nonergodic GMMs based on GPs ([Lavrentiadis et al., 2021](#); [Sung et al., 2021](#)).

Recently, [Caramenti et al. \(2020\)](#) introduced an alternative to GPs for the estimation of spatially varying GMMs, based on geographically weighted regression (GWR) ([Fotheringham et al., 2002](#)), called multi-source geographically weighted regression (MS-GWR). The method of

Caramenti et al. (2020) extends GWR to account for spatially varying coefficients depending on both event and station coordinates. They showed an example based on peak ground acceleration (PGA) data recorded in Italy, using the ITA2018 GMM (Lanzano et al., 2019) as a reference model. Lanzano et al. (2021) investigated MS-GWR further, extending the model to different response spectral periods.

GWR and GP based SVCMs are based on different underlying assumptions, which naturally raises the question how they compare. GWR constructs local models for different locations, based on weighting data according to their distance to the location. The weights are determined by a spatial kernel. By contrast, GP based models assume an underlying functional dependence of the spatial coefficients, and place a GP prior on the functions. There have been several studies comparing Bayesian SVCMs and GWR based models (e.g. Finley, 2011; Wolf et al., 2018), which generally find that both methods can produce similar results. None of the comparisons, though, deal with the peculiarities of empirical ground-motion data sets, with spatial processes based on different coordinates (event and station), and non-spatially varying random effects.

Here, I compare different GP based nonergodic models with each other, as well as with the MS-GWR model of (Caramenti et al., 2020). The same data set as in Caramenti et al. (2020) is used, which comprises the Italian data of the model of Lanzano et al. (2019). The GP based nonergodic models are estimated via Bayesian inference using the *integrated nested Laplace approximation* (INLA, Rue et al., 2009). INLA provides an efficient way to estimate spatially varying models, and can be used to estimate nonergodic models (Kuehn, 2021b).

The paper is organized as follows: I give a short overview of SVCMs and place them in the context of nonergodic GMMs. Then follows a brief introduction to the Italian data and the specific nonergodic GMM formulation (functional form, which nonergodic terms are included). I then provide some details on the implementation of the models with INLA and MS-GWR. Finally, different models are compared in terms of their predictive accuracy, and their estimated spatially varying coefficients. Code to run the analysis in the computer program R (R Core Team, 2021) is provided at https://github.com/nikuehn/NonErgGMM_Italy.

2 Spatially Varying Coefficient Models

In this section, I provide a brief introduction to SVCs, and the differences between Bayesian GP based models and GWR models. For more details, see e.g. [Finley \(2011\)](#); [Gelfand et al. \(2003\)](#) or [Fotheringham et al. \(2002, 2017\)](#). These models are then connected to nonergodic GMMs.

A general linear model has the following form

$$y = \vec{\beta} \vec{\xi} + \epsilon \quad (1)$$

where y is the target variable, β is a vector of coefficients, ξ is the vector of predictor variables, and ϵ is the residual. For a spatial model, where the coefficients depend on location $\vec{x}$, the model becomes

$$y(\vec{x}) = \vec{\beta}'(\vec{x}) \vec{\xi}(\vec{x}) + \epsilon(\vec{x}) \quad (2)$$

In a Bayesian GP based SVC, the spatial coefficients are typically written as

$$\vec{\beta}'(\vec{x}) = \vec{\beta} + \delta\vec{\beta}(\vec{x}) \quad (3)$$

where $\delta\vec{\beta}(\vec{x})$ can be understood as spatial adjustment terms to the base coefficients $\vec{\beta}$. The adjustment terms are assumed to be distributed according to a GP. For computational reasons, the GPs for the different adjustments are often assumed to be independent ([Finley, 2011](#)), but a multivariate specification is possible as well ([Gelfand et al., 2004](#)). The GPs have mean zero, so the specification for the i th coefficient becomes

$$\delta\beta_i \sim GP(0, k_i(\vec{x}, \vec{x}')) \quad (4)$$

where $k_i(\vec{x}, \vec{x}')$ is the covariance function (or kernel function). Chapter 4 of [Rasmussen and Williams \(2006\)](#) provides an overview of covariance functions. Typically, a stationary and isotropic covariance function is employed (such as an exponential or squared exponential kernel), but nonisotropic and nonstationary models are possible ([Finley, 2011](#); [Paciorek and Schervish,](#)

2006). One can understand the Bayesian SVCMM as a model where the coefficients are spatially correlated, and the spatial correlation structure is determined by the covariance function. The parameters of the Bayesian SVCMM can be estimated via Markov Chain Monte Carlo (MCMC) sampling (Finley, 2011; Wheeler and Calder, 2007), or by optimizing the marginal likelihood (Dambon et al., 2021; Rasmussen and Williams, 2006).

In GWR, the coefficients for location $\vec{x}$ are determined through a weighted regression, where the weights depend on the distance to the location, and are determined by a spatial kernel. Thus, the regression coefficients are estimated as

$$\vec{\beta}'(\vec{x}) = [\mathbf{X}^T \mathbf{W}(\vec{x}) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(\vec{x}) \vec{y} \quad (5)$$

where $\mathbf{X}$ is the design matrix, $\vec{y}$ is the vector of target variables, and $\mathbf{W}(\vec{x})$ is a diagonal matrix of weights. The weight matrix is location specific, and its elements are determined by a spatial kernel, $W_{ij} = h(\vec{x}_i, \vec{x}_j)$. The elements of $\mathbf{W}(\vec{x})$ place greater weight on observations that are close to $\vec{x}$. Different kernels may be used (Fotheringham et al., 2002; Wheeler, 2014), with popular choices being the exponential or the Gaussian kernel. Apart from choosing a kernel, one must also choose a kernel bandwidth, which determines the range of influence of the observations. This parameter is typically determined by cross-validation. In Equation (5), the bandwidth is the same for all spatial coefficients. This assumption can be relaxed (Fotheringham et al., 2017).

In Equation (2), all coefficients $\vec{\beta}$ are spatially varying. A more flexible model is one where only some of the coefficients vary with $\vec{x}$

$$y(\vec{x}) = \vec{\beta}_c \vec{\xi}_c(\vec{x}) + \vec{\beta}'_v \vec{\xi}_v(\vec{x}) + \epsilon(\vec{x}) \quad (6)$$

where subscript c denotes the spatially constant coefficients, and subscript v denotes the spatially varying coefficients. Fotheringham et al. (2002) extended simple GWR to this case. In a Bayesian SVCMM, this can be considered a special case of the more general formulation of Equation (2), where some of the adjustment terms are zero.

2.1 Nonergodic GMMs and Spatially Varying Coefficient Models

A general GMM has the form

$$y_{es} = f(\vec{c}; \vec{\xi}_{es}) + \delta B_e + \delta S2S_s + \delta W S_{es} \quad (7)$$

where y is the ground-motion parameter of interest, $f(\vec{c}; \vec{\xi})$ is a function describing the scaling of the y with the predictor variables $\vec{\xi}$, $\vec{c}$ is a vector of coefficients, δB_e and $\delta S2S_s$ are the event term and site term, and subscripts e and s denote event e and station s , respectively. The event and site term are random effects which are introduced to account for correlation of records from the same event or station. $\delta W S_{es}$ is the remaining within-event/within-station residual, corresponding to ϵ in the linear model of Equation (1).

It has long been recognized that the formulation of Equation (7) is overly simplistic, as it does not account for systematic, repeatable, location specific source and path effects. Equation (7) does account for systematic site effects via $\delta S2S$, but these are specific to each particular station, and technically do not allow to extrapolate to locations close to the station. If one accounts for systematic terms, then the GMM becomes nonergodic, which is typically written as

$$y(\vec{x}_e, \vec{x}_s) = f(\vec{c}; \vec{\xi}(\vec{x}_e, \vec{x}_s)) + \delta B_{0,e} + \delta S2S_s + \delta L2L(\vec{x}_e) + \delta P2P(\vec{x}_e, \vec{x}_s) + \delta W S_{0,es} \quad (8)$$

where $\delta L2L(\vec{x}_e)$ and $\delta P2P(\vec{x}_e, \vec{x}_s)$ are systematic source and path effects, and the dependence of the different variables (target, predictors, systematic terms) on event locations $\vec{x}_e$ and station locations $\vec{x}_s$ is made explicit. In this formulation, a nonergodic model consists of local adjustment terms to an ergodic base model. Often it is assumed that the event term can be decomposed into the systematic source term and the remaining event term, $\delta B = \delta L2L + \delta B_0$ (Al-Atik et al., 2010; Lin et al., 2011). Similarly, the within-event/within-site residual is decomposed as $\delta W S = \delta W S_0 + \delta P2P$. Accounting for spatial correlation in site terms $\delta S2S$ can lead to improved predictions for sites that do not have observed records (Chao et al., 2020).

In an SVCN framework, the terms associated with spatially varying coefficients can be

thought of as the nonergodic, systematic terms. The interpretation outlined above in Equation (8) fits neatly with the Bayesian SVCM, as the spatially varying coefficients are modeled as adjustments to a global coefficient. The Bayesian SVCM is a natural extension to the simple random effects model, which accounts for spatial effects with a more complicated covariance structure. A GWR model is different, and does not lend itself to the “nonergodic adjustment to an ergodic base model” formulation. Instead, a GWR model constructs a local model for each location under consideration. One should note that neither framework is a priori better or more suited to the development of nonergodic GMMs.

3 Nonergodic Ground-Motion Models for Italy

The functional form of the ITA18 model, which serves as the reference model, is as follows (Lanzano et al., 2019)

$$\begin{aligned}
\log_{10} PGA = & a + b_1(M_W - M_h) 1_{(M_w \leq M_h)} + b_2(M_W - M_h) 1_{(M_w > M_h)} \\
& + [c_2 + c_1(M_W - M_h)] \log_{10} \sqrt{R_{JB}^2 + h^2} + c_3 \sqrt{R_{JB}^2 + h^2} \\
& + k \left[\log_{10} \frac{V_{S30}}{800} 1_{(V_{S30} \leq 1500)} + \frac{1500}{800} 1_{(V_{S30} > 1500)} \right] + f_1 F_{SS} + f_2 F_R \\
& + \delta B + \delta S2S + \delta WS
\end{aligned} \tag{9}$$

where $1_{(M_w \leq M_h)}$ is an indicator function that evaluates to one if the condition is true, and zero otherwise. M_W is the moment magnitude, R_{JB} is the Joyner-Boore distance in km , and V_{S30} is the time-average shear wave velocity in the upper 30m with units of m/s . F_{SS} and F_R are indicators for strike-slip and reverse-faulting events, respectively. δB , $\delta S2S$, and δWS are the event term, site term, and within-event/within-site residual, which are distributed according to normal distributions with mean zero and standard deviations τ , ϕ_{S2S} , and ϕ_{SS} , respectively.

In their nonergodic GMM for California, Landwehr et al. (2016) made the constant coefficient (intercept), the geometrical spreading coefficient, and the anelastic attenuation coefficient

dependent on the event location $\vec{x}_e$. In addition, they had an intercept and the linear site scaling coefficient that was dependent on site location $\vec{x}_s$. The model of Landwehr et al. (2016) did not include an event term δB or a site term $\delta S2S$.

The MS-GWR model of Caramenti et al. (2020) follows the model of Landwehr et al. (2016), and includes a spatially varying geometrical spreading, anelastic attenuation, and linear site scaling coefficient. It also does not include event or site terms, and does not account for spatially varying intercepts.

One can rewrite the reference model of Equation (9) with spatially varying coefficients as

$$\begin{aligned}
\log_{10} PGA = & \tilde{a}_e(\vec{x}_e) + \tilde{a}_s(\vec{x}_s) + b_1(M_W - M_h) 1_{(M_w \leq M_h)} + b_2(M_W - M_h) 1_{(M_w > M_h)} \\
& + [\tilde{c}_2(\vec{x}_e) + c_1(M_W - M_h)] \log_{10} \sqrt{R_{JB}^2 + h^2} + \tilde{c}_3(\vec{x}_e) \sqrt{R_{JB}^2 + h^2} \\
& + \tilde{k}(\vec{x}_s) \left[\log_{10} \frac{V_{S30}}{800} 1_{(V_{S30} \leq 1500)} + \frac{1500}{800} 1_{(V_{S30} > 1500)} \right] + f_1 F_{SS} + f_2 F_R \\
& + \delta B_0 + \delta S2S_0 + \delta W S_0
\end{aligned} \tag{10}$$

where two spatial intercepts (dependent on event and station location), spatially varying geometrical spreading, anelastic attenuation, and linear site scaling coefficients are included.

If one defines a spatially varying coefficient as an adjustment to the ergodic base coefficient, the full spatial model can be written as

$$\begin{aligned}
\log_{10} PGA = & a + \delta a_e(\vec{x}_e) + \delta a_s(\vec{x}_s) + b_1(M_W - M_h) 1_{(M_w \leq M_h)} + b_2(M_W - M_h) 1_{(M_w > M_h)} \\
& + [(c_2 + \delta c_2(\vec{x}_e)) + c_1(M_W - M_h)] \log_{10} \sqrt{R_{JB}^2 + h^2} + (c_3 + \delta c_3(\vec{x}_{e/es})) \sqrt{R_{JB}^2 + h^2} \\
& + (k + \delta k(\vec{x}_s)) \left[\log_{10} \frac{V_{S30}}{800} 1_{(V_{S30} \leq 1500)} + \frac{1500}{800} 1_{(V_{S30} > 1500)} \right] + f_1 F_{SS} + f_2 F_R \\
& + \delta B_0 + \delta S2S_0 + \delta W S_0
\end{aligned} \tag{11}$$

where $\delta a_e(\vec{x}_e)$ is the spatially varying adjustment for the constant coefficient a , dependent on

event location. The other adjustment coefficients are denoted similarly. The subscript 0 is added to the event and site term and the within-event/within-site residual to indicate that they are the terms after the systematic, repeatable nonergodic terms are taken out.

The adjustment to anelastic attenuation coefficient, δc_3 , is modeled as event location dependent in [Landwehr et al. \(2016\)](#) and [Caramenti et al. \(2020\)](#). This implies that the attenuation for a specific event is isotropic, i.e. is the same in all directions. This is an improvement over an ergodic model, but is unrealistic. Hence, the anelastic attenuation coefficient is made dependent on the midpoint between event and station location (denoted by $\vec{x}_{es}$), to account for possible directional differences in attenuation. I compare the midpoint model to an event-location model, which is why the coefficient is denoted as $\delta c_3(\vec{x}_{e/es})$, to specify that the adjustment depends either on event or midpoint location.

Even if the anelastic attenuation is modeled dependent on the midpoint, it neglects that the anelastic attenuation is an accumulated effect over the whole travel path. This could be modeled by a dependence on both event and station locations, but this runs into problems that there not many close points in four dimensions. [Dawood and Rodriguez-Marek \(2013\)](#) introduced a cell-specific attenuation model in which the anelastic attenuation is modeled as the sum over distance fractions in small cells. [Kuehn et al. \(2019\)](#) cast the cell-specific model as a Bayesian model to account for uncertainty in cells that are not well sampled. In the cell-specific model, the adjustment to the anelastic attenuation term becomes

$$\begin{aligned}\delta c_{attn} &= \sum_{i=1}^{N_C} \delta c_{ca,i} \Delta R_i \\ &= \overrightarrow{\Delta R} \cdot \vec{\delta c}_{ca}\end{aligned}\tag{12}$$

where ΔR_i is the fraction of the path in the i th cell.

For the Bayesian SVCs, the spatially varying adjustment coefficients are modeled as a Gaussian process (or a spatial Gaussian field, GF) with mean zero and Matérn covariance

function

$$\delta a \sim GP(0, \omega_a^2 k(\vec{x}, \vec{x}')) \quad (13)$$

$$k(\vec{x}, \vec{x}') = \frac{2^{(1-\nu)}}{\Gamma(\nu)} (\kappa |\vec{x} - \vec{x}'|)^\nu K_\nu(\kappa |\vec{x} - \vec{x}'|) \quad (14)$$

where ω denotes the standard deviation of the GP, which is a measure of how much the coefficient can vary across the region under study. In Equation (14), Γ is the gamma function, K_ν is the modified Bessel function of the second kind, κ is a scale parameter and ν is a smoothness parameter. For $\nu = 0.5$, the Matérn covariance function becomes the exponential covariance function (Rasmussen and Williams, 2006), which is extensively used in spatial correlation models for ground motion (e.g. Chao et al., 2020; Esposito and Iervolino, 2011, 2012; Huang and Galasso, 2019; Jayaram and Baker, 2009; Kuehn and Abrahamson, 2020). The practical spatial range is defined as $\ell = \sqrt{8\nu}/\kappa$, which corresponds to the distance where the correlation becomes 0.14 (Krainski et al., 2019).

Here, the Matérn covariance function with $\nu = 1$ is used. This makes it possible to estimate the model parameters using Bayesian inference, via the INLA method (Rue et al., 2009). Together with the *stochastic partial differential equation* (SPDE) approach (Bakka et al., 2018; Lindgren and Rue, 2015; Lindgren et al., 2011) INLA allows for fast and efficient inference in spatial models. Some further details on spatial models and INLA are provided in Section [Implementation](#). An SVCMM can be easily combined with the cell-specific attenuation model in INLA (Kuehn, 2021b).

In Equation (10), the ergodic coefficients $(a, b_1, b_2, c_1, c_2, c_3, k, f_1, f_2)$ are fixed effects. The adjustment terms (both the spatial adjustment terms, as well as the event terms δB and site term $\delta S2S$), are random effects. The parameters that control the distribution of the random effects are called hyperparameters. These are the standard deviations τ and ϕ_{S2S} for event terms and site terms, and the standard deviation ω and spatial range ℓ for the spatial terms, or the standard deviation ω_{ca} for the cell-specific adjustment coefficients. All parameters (fixed effects, random effects, and hyperparameters) are estimated at the same time during the fitting process.

(a) Magnitude-distance distribution.

(b) Map of data. Stations are triangles, events are circles.

Figure 1: Data used in the analysis.

In the following section, I consider different nonergodic SVCMs, implemented in INLA. The models differ in the nonergodic adjustment terms that are included. The INLA models are compared with each other, as well as the MS-GWR model of (Caramenti et al., 2020).

4 Data

The analysis is performed on the same data set as in Caramenti et al. (2020). This data set comprises the Italian subset of the data used to derive the ITA18 GMM of Lanzano et al. (2019). For details on the data set see Lanzano et al. (2019) and Caramenti et al. (2020). In total, there are 4784 records from 137 events, recorded at 923 stations. I consider stations with the same location as the same station. In the data set of (Caramenti et al., 2020), there were three stations (as identified by their location) with slightly different values of V_{S30} . I used the geometric mean of the V_{S30} -values of all records from each station as the final V_{S30} value for each station. The magnitude ranges from $4 \leq 6.87$, the distance range is $0 \leq R_{JB} \leq 199.7\text{km}$. Figure 1 shows a magnitude-distance scatterplot, as well as a map with event and station locations. For the analysis, the event and station locations are converted to UTM coordinates (UTM zone 33). The target value is horizontal PGA in units of cm/s^2 . The two horizontal components are combined as RotD50 (Boore, 2010).

5 Implementation

Here, I briefly describe details on the implementation of the INLA models. All models are implemented in the R computing environment (R Core Team, 2021). The implementation of the MS-GWR model follows the one provided by Caramenti et al. (2020) at <https://github.com/lucaramenti/ms-gwr/>. Code to fit the INLA models can be found at https://github.com/nikuehn/NonErgGMM_Italy.

5.1 INLA

The INLA models are implemented using the R-INLA¹ package (Bivand et al., 2015; Lindgren and Rue, 2015; Rue et al., 2017). R-INLA is a package for approximate Bayesian inference in latent Gaussian models. For details and mathematical background on INLA, see Rue et al. (2009). For introductions to computation with R-INLA, see e.g. Gómez-Rubio (2020) or Krainski et al. (2019). For a brief introduction to INLA in the context of estimation of GMMs, see Kuehn (2021a). Kuehn (2021b) showed how one can use R-INLA to efficiently estimate a nonergodic GMM which is a combination of an SVCMM and the cell-specific attenuation approach.

In the SPDE approach to modeling spatial data, the spatial Gaussian field is approximated by basis functions which are evaluated on a triangular mesh (Lindgren and Rue, 2015; Lindgren et al., 2011). For a Matérn covariance function with $\nu = 1$, the field becomes a Gaussian Markov random field, which means that the precision matrix (the inverse of the covariance matrix) becomes sparse, which can be exploited numerically. Thus, instead of modeling the spatially varying coefficients at the observation locations, they are modeled at the mesh nodes, and projected onto the observation locations.

As an example, the model for the spatial adjustment term $\delta a_e(\vec{x}_e)$ is actually

$$\begin{aligned}\vec{\delta a}_e(\vec{x}_{mesh}) &= N(0, \mathbf{P}^{-1}) \\ \vec{\delta a}_e(\vec{x}_e) &= \mathbf{A}_e \vec{\delta a}_e(\vec{x}_{mesh})\end{aligned}\tag{15}$$

¹<https://www.r-inla.org/>

Figure 2: Triangular used to estimate the spatial effects in the INLA models, generated for the Italian data, together with stations (triangles) and epicenters (circles).

where $\mathbf{P}^{-1}$ is the covariance matrix describing the covariance of the adjustment term between the mesh nodes $\vec{x}_{mesh}$, $\mathbf{P}$ is the associated precision matrix, and $\mathbf{A}_e$ is a projector matrix connecting the mesh nodes to the observation (event) locations. The precision matrix $\mathbf{P}$ is sparse, which can be exploited numerically (Simpson et al., 2012).

Figure 2 shows the mesh used for the Italian data. For simplicity, the same mesh is used for station and event locations. The mesh has 7505 nodes, with a maximum edge length of 10km, and an outer boundary of 50km. In general, the model complexity increases with a larger number of nodes (and smaller edge lengths). Meshes with too large edge lengths will lead to biased inferences².

²See <https://haakonbakkagit.github.io/btopic114.html> for some informal advice on mesh generation.

5.1.1 Prior Distributions

INLA is a Bayesian method, which means one needs to set prior distributions. For the prior distributions of the fixed effects, the following priors are used:

$$a, b_1, b_2, c_1, c_2, k \sim N(0, 5)$$

$$c_3 \sim N(-0.005, 0.01)$$

$$f_1 \sim N(0.02, 0.025)$$

$$f_2 \sim N(0.1, 0.04)$$

The priors for the coefficients associated with different types of faulting (f_1, f_2) are based on [Bommer et al. \(2003\)](#). The anelastic attenuation coefficient is typically small and negative, which informs the prior distribution for c_3 . For the other coefficients, I choose a weakly informative prior distribution, which means I use a prior that is not too restrictive on the parameter space.

The prior distributions for the standard deviations τ , ϕ_{S2S} , and ϕ_{SS} have to be set on the log-precisions, which are internally used in INLA. A log-Gamma distribution with shape parameter 0.9 and rate parameter 0.007 is used:

$$\ln prec_{\tau} \sim \log - Gamma(0.9, 0.007)$$

$$\ln prec_{\phi_{S2S}} \sim \log - Gamma(0.9, 0.007)$$

$$\ln prec_{\phi_{SS}} \sim \log - Gamma(0.9, 0.007)$$

The parameters are chosen in such a way that the 5% and 95% quantiles for the prior distribution of the standard deviations (if the log-Gamma distribution is suitably transformed) are 0.05 and 0.45, respectively.

The prior distributions for the spatial field parameters are based on a penalized complexity

(PC) prior derived by [Fuglstad et al. \(2019\)](#) for SPDE models. The PC prior encodes a notion of parsimony, treating the spatial model as a more complex extension of a simpler base model (in this case the ergodic base model without spatial terms). The PC prior shrinks the parameters towards the simpler base model, where the spatial terms have zero marginal variance and infinite range. For details, see [Franco-Villoria et al. \(2019\)](#); [Fuglstad et al. \(2019\)](#); [Simpson et al. \(2017\)](#). The parameters of the PC prior for spatial models are two value u and α such that $P(\omega > u) = \alpha$, and two values v, β such that $P(\ell < b) = \beta$. Thus, the PC prior is specified by setting our prior belief that the standard deviation of the spatial field is larger than a certain value, and its spatial range smaller than a certain value. For the spatial range ℓ of all spatially varying adjustment coefficients, I assume the following prior

$$P(\ell < 30) = 0.5$$

meaning I assume a prior probability of 50% that the range is smaller than 30km. For the standard deviations, I assume

$$P(\omega > 0.3) = 0.01$$

for the constants δa_e , δa_s , geometrical spreading adjustment δc_2 , and linear site scaling adjustment δk . Since the scale of the anelastic attenuation coefficient is very different, I use a prior distribution for its standard deviation of

$$P(\omega > 0.01) = 0.01$$

For models that use a cell-specific attenuation model, I also use a PC prior for the standard deviation ω_{ca} , which is specified in the same way and uses the same values. The prior distributions for the spatial terms are informed by results of [Landwehr et al. \(2016\)](#), [Kuehn et al. \(2019\)](#), and [Kuehn \(2021b\)](#), based on Californian data.

5.2 MS-GWR

Caramenti et al. (2020) extended the simple spatial model of Equation (2) to the case where different coefficients can depend on different sets of coordinates (like event and station coordinates), i.e. a model like

$$y = \vec{\beta}_c \cdot \vec{\xi}_c + \vec{\beta}_e(\vec{x})_e \cdot \vec{\xi}_e(\vec{x})_e + \vec{\beta}_s(\vec{x})_s \cdot \vec{\xi}_s(\vec{x})_s \quad (16)$$

where subscript c , e , s denote the sets of coefficients and predictor variables that are spatially constant or depend on event and station coordinates. The different sets of coefficients are estimated sequentially: the spatial coefficients dependent on the station location $\vec{\beta}_s(\vec{x}_s)$ are estimated first; then, the spatial coefficients dependent on the event location, $\vec{\beta}_e(\vec{x}_e)$ are estimated based on partial residuals calculated from the first step, and finally the constant coefficients $\vec{\beta}_c$ from partial residuals of the second step. Note that the order of the sequence is arbitrary, but this is the one found to be preferred by Caramenti et al. (2020), and used in Lanzano et al. (2021). For the exact derivation of the coefficients, see (Caramenti et al., 2020). The event-related spatial coefficients are the geometrical spreading coefficient c_2 and the anelastic attenuation coefficient c_3 . The spatial coefficient dependent on the site location is k , the coefficient controlling the V_{S30} scaling. The spatial weights are calculated based on a Gaussian kernel. The bandwidths used for the MS-GWR model are 25km for the event-related coefficients, and 75km for the station related coefficient.

The MS-GWR model in the specification of Caramenti et al. (2020) does not include systematic event and site terms. Following Lanzano et al. (2021), the residuals are partitioned into between-event, station-to-station, and within-event/within-station residuals, which allows one to calculate ϕ_{SS} , τ , and ϕ_{S2S} . This model is called “MS-GWR B”. The partitioning of the residuals is done using INLA.

Table 1: Comparison of different INLA models. Column *Random effects* denotes which non-spatially effect is included, column *Spatial effects* which spatially varying effect is included. A spatially varying intercept is denoted with c , spatially varying geometrical spreading with gs , spatially varying linear site term with vs , and spatially varying anelastic attenuation with r . Subscript e , s , es denote dependence on event, station, and midpoint coordinate. Column *Cells* indicates whether cell-specific attenuation is included.

Model	Random Effects	Spatial Effects	Cells	WAIC	runtime (s) [†]
Model 1	NA	NA	no	3407.34	12.51
Model 2	δB	NA	no	2398.99	34.10
Model 3	$\delta B, \delta S2S$	NA	no	-777.41	22.05
Model 4	$\delta B, \delta S2S$	c_e, c_s	no	-1521.88	477.44
Model 5	$\delta B, \delta S2S$	c_e, c_s, r_{es}	no	-1627.29	409.91
Model 6	$\delta B, \delta S2S$	c_e, c_s, r_e	no	-1382.83	269.69
Model 7	$\delta B, \delta S2S$	$c_e, c_s, gs_e, r_{es}, vs_s$	no	1758.34	1926.42
Model 8	$\delta B, \delta S2S$	c_e, c_s, gs_e, vs_s	yes	-1953.83	617.32
Model 9	$\delta B, \delta S2S$	c_e, c_s	yes	-1841.30	907.45
Model 10	NA	gs_e, r_{es}, vs_s	no	-187.91	1199.44
MS-GWR	NA	gs_e, r_e, vs_s	no	NA	11434.796

[†] Runtimes depend on the computational load of the computer, and can be variable even for different runs on the same computer. All runs were carried out on a Macbook Pro with 2.8Ghz and 16GB RAM.

6 Results

6.1 Model Comparison based on Predictive Measures

First, I compare different (spatial and non-spatial) models fit with INLA in terms of the widely applicable information criterion³ (WAIC, [Watanabe, 2010, 2013](#)), which is an information criterion for Bayesian models. Similar to the Akaike information criterion ([Akaike, 1973](#)), lower values indicate a better fit. The different models, their specifications, and the WAIC and runtimes are listed in Table 1. The simplest (Model 1) model is a linear model of Equation (9), without any random effects or spatially varying coefficients. I then add event and site terms, as well as the spatially varying terms for the intercept (both dependent on event and station coordinates), geometrical spreading, linear site scaling, and anelastic attenuation. I also replace the spatially varying anelastic attenuation coefficient with a cell-specific attenuation term.

As expected, the simplest model performs worst in terms of WAIC. Adding event terms, and in particular site terms, leads to a major improvement. Incorporating spatially varying

³also called Watanabe-Akaike information criterion

coefficients leads to lower WAIC values, with the inclusion of the spatially varying intercepts leading to the most improvement. One can see that the model with a spatially varying linear distance term dependent on the midpoint between the event and station coordinate has lower WAIC than when the coefficient depends on event coordinate. This is somewhat expected, as the anelastic attenuation should be anisotropic for a given event. The models with cell-specific anelastic attenuation coefficients perform best. This is due to the fact that the cell model treats the anelastic attenuation as a sum over the path.

I also fit a model that only includes spatially varying geometrical spreading, linear site scaling, and anelastic attenuation, but no intercepts or event/site terms. Such a model is closest to the MS-GWR model of [Caramenti et al. \(2020\)](#), which includes only these terms. In terms of WAIC, the INLA model with these terms performs better than the simplest models, but actually worse than a model without spatial terms, but site terms $\delta S^2 S$. One should not draw conclusions about the MS-GWR model based on this comparison, though, since the WAIC values are tied to the INLA models and their parameterization.

The SVCMM models lead to a significant increase in runtime of the models. However, the runtime is still manageable, with the longest running model finishing in under 40 minutes. In fact, even the longest running INLA model compares favorably with the MS-GWR model in terms of runtime. One thing to note here is that INLA is run with a simple integration scheme and a simpler approximation to the posterior marginals⁴. Running Model 9 with default options (which uses a composite integration scheme over the hyperparameters, and the simplified Laplace approximation to the posterior marginals ([Martino and Riebler, 2020](#))) increases the runtime to 1596.59s for Model 9.

The WAIC values listed in Table 1 are based on calculating the likelihood of the observations (while integrating over the posterior distribution of the parameters), adjusted by the effective number of parameters ([Vehtari et al., 2017](#)). The likelihood of each observation is calculated conditional on all random effects, which means the median prediction for an observation includes the event and site term, and the standard deviation is ϕ_{SS} or ϕ_0 . For ground-motion prediction, one typically wants to predict observations from new events, which means the event term is

⁴using `control.inla = list(int.strategy = "eb", strategy = "gaussian")` in the INLA call

(a) RMSE values for different models.

(b) Loglikelihood values for different models.

Figure 3: Cross-validation results. For model abbreviations, see Table 1.

unknown, and the aleatory standard deviation is $\sigma_{SS} = \sqrt{\phi_{SS}^2 + \tau^2}$ (or σ_0 , calculated the same way). To compare the models how they can predict new events, a 5-fold cross-validation (CV) is performed. The events are split into 5 sets. In each iteration, the observations from one set of events are left out and used as the test set, while the other sets are used as the training set. In the prediction of the test set, I include all random effects, i.e. if there are stations with records in both the training and test, the estimated site term is used for the prediction. The root mean squared error (RMSE) is calculated over all combined test data points.

Since the accurate quantification of the predictive uncertainty is important in PSHA, I also calculate the log-likelihood of each test observation. This indicates how well the predictive uncertainty is captured in a model, which is important for PSHA. Note that the predictive uncertainty contains both aleatory variability and epistemic uncertainty: there is aleatory variability because the predictions are made for previously unseen records and events, and there is epistemic uncertainty because the exact values of the nonergodic model terms are not known, in particular if the new data is far from observations. There is also uncertainty in the values of the fixed effects. The loglikelihood values are calculated for each observation in the test set for all 5 folds, and the total loglikelihood is computed as the sum of all observations.

For the CV calculations, the coefficients associated with style of faulting (f_1 and f_2 in Equations (9) and (10)) are dropped. I only perform the CV calculations for a subset of the models of Table 1, mainly to reduce computation time. The two versions of the MS-GWR model are included. For the second model, the site terms estimated from the residuals to the

corresponding site terms in the test set are added to the predictions. The loglikelihood values are not calculated for the second model, since it becomes difficult to combine the uncertainty from the first model with the one from the second model.

The RMSE and loglikelihood values based on CV are shown in Figure 3. The results in terms of ranking of the models are similar to the ones based on WAIC. However, one can see that the spatial models only provide a small improvement (slightly smaller RMSE, slightly larger loglikelihood) over the base model that includes event and site terms (Model 3). Based on the CV results, Model 9 performs best, which includes spatially varying intercepts and cell-specific attenuation, but not spatially varying geometrical spreading or V_{S30} -scaling; whereas based on WAIC, Model 8, which includes these terms, performs better. An explanation could be that an event-specific geometrical spreading term leads to a better prediction of records within the event, but that it is not well represented by a simple spatial correlation model. Another explanation could be that there are not enough events in the data set to resolve the spatially varying geometrical spreading well enough that it proves beneficial for predicting new events.

Another conclusion one can draw from Figure 3 is that while RMSE and loglikelihood lead to a similar ranking of models, there is more separation between models in the loglikelihood values. Models 4 and 8 have essentially the same RMSE, but different loglikelihood values, which indicates that both lead to similar median predictions, but Model 4 has a better quantification of the predictive uncertainty.

In terms of RMSE and loglikelihood, the MS-GWR model provides an improvement over the linear model without random effects (Model 1), but performs worse than the simple random-effects model without spatial terms. Adding site terms to the predictions leads to improvements, with model MS-GWR B having similar RMSE values as the other models.

6.2 Standard Deviations

Figure 4 compares the standard deviations of different models, for the event, station and record levels (τ , ϕ_{S2S} , and ϕ_{SS}). The median of the posterior distributions are plotted, as well as the

Figure 4: Estimated value of the between-event standard deviation τ , site-to-site standard deviation ϕ_{S2S} , and within-event/within-site standard deviation ϕ_{SS} . For models that do not partition the residuals, the estimated total standard deviation σ is shown. Error bars indicate the 2.5% and 97.5% quantiles of the posterior distribution.

2.5% and 97.5% quantiles. For the models that do not include event and site terms, the total standard deviation σ is plotted. In general, there is a reduction in the values of ϕ_{SS} , ϕ_{S2S} , and τ due to the inclusion of spatial terms, which is expected, and has been observed before (Kuehn et al., 2019; Lavrentiadis et al., 2021; Lin et al., 2011). For the MS-GWR model, the total variability is about $\sigma = 0.3$, which is larger than for the other spatial models. This might in part be explained that the model does not include event and site terms, which are accounted for by the spatial terms. Since the bandwidths of the spatial kernels are quite long (75km and 25km), they cannot accommodate the non-spatially varying random effects, which would have a bandwidth (or spatial correlation length) of zero. The model that partitions the residuals of the MS-GWR model into between-event, station-to-station, and within-event residuals leads to values of τ that are comparable with the spatial INLA models, while ϕ_{SS} and ϕ_{S2S} are larger, and are comparable with the values from Model 3, which is the non-spatial model. These results are in line with the findings of Lanzano et al. (2021).

6.3 Spatially Varying Coefficients

Figures 5, 6, and 7 show maps of the spatially varying geometrical spreading, V_{S30} -scaling, and anelastic attenuation coefficients c_2 , k , and c_3 . These are shown for Model 7 (all spatially varying coefficients), Model 8 (cell-specific attenuation), Model 10 (no random effects

or spatially varying intercepts), and the MS-GWR model. Maps for other models, as well as maps of associated uncertainties and the spatially varying intercepts, can be found at https://github.com/nikuehn/NonErgGMM_Italy. Main differences can be seen for Model 10; since this model does not include non-spatially varying event and site terms, the model tries to accommodate these terms by very short spatial ranges for the geometrical spreading and V_{S30} -scaling. Differences between Model 7 and Model 8 are small for the geometrical spreading and V_{S30} -coefficient, as are differences between the anelastic attenuation coefficient for Model 7 and Model 10.

Differences between the INLA models and the MS-GWR model are more pronounced, in particular for the geometrical spreading and anelastic attenuation coefficient. There are some similarities, such as low values for k in the Northwest at the border to France, and higher values in and around Tuscany. All models have positive values of the anelastic attenuation coefficient in the Northern part of Italy. This is an area where the Moho bounce is prominent (Lanzano et al., 2016), which biases the estimates of the anelastic attenuation if not properly modeled. The cell-specific attenuation coefficients (Model 8 in Figure 7) follow a similar pattern as the c_3 values of Models 8 and 10, with lower values along the Tyrrhenian coast. They also show some unphysical behavior: 15 out of 603 cells have a positive cell-specific attenuation coefficient.

6.4 Prediction

Here, models are compared in terms of their predictions. Predictions are calculated for an event in Middle Italy, for seismic waves traveling to the Northwest and Southeast. The geometry is shown in Figure 8. Predictions are calculated for 1km to 200km, using Model 3 (only event and site terms), Model 9 (spatially varying intercepts and cell-specific attenuation), and the MS-GWR model. The predictions and associated epistemic uncertainties, denoted ψ_μ , are also shown in Figure 8. At short distances, the predictions of both SVCMs are similar, slightly higher than the ergodic base predictions from Model 3. This is due to the combined effect of a systematic event and site term, which both models capture with their spatially varying coefficients. The SVCM predictions show a different attenuation with distance than the base model. For the MS-GWR model, where the anelastic attenuation coefficient is dependent on

(a) Model 7

(b) Model 9

(c) Model 10 (no event/site terms)

(d) MS-GWR model.

Figure 5: Values of the spatially varying geometrical spreading coefficient, for different models.

(a) Model 7

(b) Model 8

(c) Model 10 (no event/site terms)

(d) MS-GWR model

Figure 6: Values of the spatially varying V_{S30} -scaling coefficient, for different models.

(a) Model 7

(b) Model 8, cell-specific attenuation coefficients.

(c) Model 10 (no event/site terms)

(d) MS-GWR model

Figure 7: Values of the spatially varying anelastic attenuation coefficient, for different models.

event coordinate, the predictions to the North and South are very similar; small differences arise due to the spatially varying V_{S30} -scaling term at the different stations. By contrast, the INLA model predictions diverge, due to the cell-specific attenuation, with predictions to the South being more similar to the MS-GWR predictions. One can also see that the INLA predictions towards the South sometimes exhibit an increase in amplitudes with distance. This happens because of a positive cell-specific attenuation coefficient, which is unphysical, and should be corrected in a real application.

Figure 8 also shows the standard deviation of epistemic uncertainty associated with the predictions, denoted as ψ_μ . Two plots are shown: in the first one, the standard deviation associated with the systematic site terms, ϕ_{S2S} , is included in ψ_μ , while in the second plot it is not. If ϕ_{S2S} is included in the uncertainty, then the base model (Model 3) exhibits larger values of ψ_μ than the spatial models. The reason is that the value of ϕ_{S2S} is reduced due to the inclusion of the spatial terms (cf. Figure 4), while the additional uncertainty due to the spatial terms is reduced, since the predictions are made for locations close to observed data. On the other hand, if ϕ_{S2S} is not included, then ψ_μ is determined only by the uncertainty of the fixed effects, and the uncertainty due to the spatial terms. In this case, the spatial models show larger uncertainty than the base model.

The values of ψ_μ for the INLA and MS-GWR models are similar. The value of ϕ_{S2S} for the MS-GWR B model is larger than for the INLA model, which leads to a value closer to the base model when ϕ_{S2S} is included. In the INLA model, the uncertainty increases strongly with increasing distance. This is due to the cell-specific attenuation model.

7 Discussion and Conclusions

The main point of the present paper is to compare SVCMs for the development of nonergodic GMMs. Two methods to model the SVCMs are compared, Bayesian hierarchical models based on GPs, and local models based on GWR, all on the Italian data of (Lanzano et al., 2019). Based on the analyses carried out in this work, both methods lead to broadly similar results, but with some differences in the details. Differences between methods should be embraced,

(a) Mesh, data, and event/stations for which predictions are calculated. The event location is denoted by the red circle, and predictions are calculated for stations along the red and blue lines.

(b) Median prediction for the different models along the different directions.

(c) Standard deviation ψ_μ of epistemic uncertainty associated with the predictions. ψ_μ includes epistemic uncertainty due to spatial effects, uncertainty in fixed effects, and uncertainty in the site terms (ϕ_{S2S}).

(d) Standard deviation ψ_μ of epistemic uncertainty associated with the predictions. ψ_μ does not include ϕ_{S2S} .

Figure 8: Median predictions and associated epistemic uncertainty for different models. The solid lines are for INLA Model 9, the dashed lines are for the MS-GWR model. The black line corresponds to Model 3 (the non-spatial base model).

as none of the models are a “true” representation of reality. Epistemic uncertainty, both on model parameters as well as on different models, is typically included in a standard PSHA. For nonergodic GMMs, the importance of accounting for epistemic uncertainty in the nonergodic terms has long been recognized (e.g. [Walling and Abrahamson, 2012](#)). Nonergodic GMMs typically provide a method to calculate within-model uncertainty due to the spatial effects (e.g. [Caramenti et al., 2020](#); [Landwehr et al., 2016](#); [Lavrentiadis et al., 2021](#)), which can be incorporated into a nonergodic PSHA ([Abrahamson et al., 2019](#)). However, between-model uncertainty is generally not

Both GWR and GP based SVCMM methods have their strengths and weaknesses. [Caramenti et al. \(2020\)](#) and [Lanzano et al. \(2021\)](#) argue that GWR as a method is easier to understand. This is true, as in its simplest form it is a simple weighted linear regression, and the extension to MS-GWR, which takes into account multiple sets of coordinates, is relatively straightforward. A Bayesian SVCMM, on the other hand, places a GP prior on the functional dependence of the spatial adjustment on the locations. However, one can understand the GP as a framework to extend the simple random effects structure of a typical GMM to a more general structure that takes more complicated correlations between records into account. Random effects models are common and familiar in GMM development, so this perspective should make the model easier to understand. Results based on CV have shown that inclusion of site terms $\delta S2S$ is important, with Model 3 (a non-spatial model with site terms) having similar performance as the spatial models. The GP based models can include random effects such as event and site terms directly, while they need to be estimated from partitioned residuals for MS-GWR. Figure 4 shows that the values of the resulting standard deviations ϕ_{SS} , τ , and ϕ_{S2S} are different. This could be attributed to just a case of between-model uncertainty. However, the Bayesian models show a reduction in all components of the standard deviation (ϕ_{SS} , ϕ_{S2S} , and τ), while for the MS-GWR model the main reduction is in the value of τ . Given that the spatial models are modeling nonergodic source, path, and site effects, one would expect a decrease in all components ([Lin et al., 2011](#)).

Even if one accepts that nonergodic adjustment terms (and SVCMMs) are just a simple extension to a traditional random effects model, GP based models are not easy to fit. Often,

these models are fit with MCMC sampling (Finley, 2011). Since GPs scale poorly to large data sets, approximations need to be made with increasing number of observations, such as a nearest neighbourhood GP (Datta et al., 2016; Finley et al., 2018, NNGP,). Here, I have used INLA, which allows to fit the spatial models efficiently. Neither INLA nor MCMC are as easy to apply as a simple (weighted) linear regression, and require some experience. However, this may be offset by the rich inferential framework supplied by the Bayesian methods, in particular the possibility to model complicated random effects structures. This includes the cell-specific attenuation term, which can be formulated as a random effect, is a more physical representation of the path effect than a simple spatially varying coefficient. In addition, one can set prior distributions on all parameters, which can help constrain the model. Informative prior distributions on model coefficients have been proven to be successful to improve model performance (Kowsari et al., 2020).

In terms of predictive performance, the spatial models estimated with INLA outperform the MS-GWR models. Once site terms are accounted for in model MS-GWR B, the difference in terms of RMSE based on CV is small. However, the different models exhibit different spatial patterns of the spatially varying coefficients (cf. Figures 5, 6, 7), which can lead to quite different predictions (Figure 8), in particular at large distances. Given the similarity in RMSE, one should probably consider the notion of using different models based on different methodologies in a PSHA.

Both GWR and Bayesian SVCs employed in this work are relatively simple. The MS-GWR model uses a single kernel, and the same bandwidth for the spatially varying coefficients that depend on the same location (geometrical spreading and anelastic attenuation). Murakami et al. (2019) note “the importance of scale in spatially Varying Coefficient Modeling” and note that the inclusion of different bandwidths for different coefficients is generally beneficial. This can be incorporated into a GWR model based on the methodology of Fotheringham et al. (2017). In the INLA models, different values of ℓ (the practical spatial range) are estimated for the different spatial effects, which adds flexibility to the model. Other extensions to increase flexibility are the use of adaptive kernels (where bandwidth is not measured in term of length, but in the number of nearest neighbours (Wolf et al., 2018)) for GWR, or the use of non-

stationary and non-isotropic covariance function in Bayesian SVCMs.

Apart from a comparison of different SVCMs, this paper also highlights the usefulness of INLA to estimate nonergodic GMMs. A range of different models were fit with INLA, with different numbers of spatial random effects. INLA is quite fast (cf. Table 1), which is important for scaling the method to larger data sets. The Italian data set used here is relatively small (4784 records) compared with other regions, which means that model fitting can be done relatively easy. For comparison, the Californian data of the NGA-W2 data set [Ancheta et al. \(2014\)](#); [Bozorgnia et al. \(2014\)](#) comprises more than 12000 records. A nonergodic model for California similar to the ones estimated here takes about 1-2 hours to fit with INLA, compared to 20 hours using MCMC sampling ([Kuehn, 2021b](#)). Data sets are going to increase in the next years, both due newly recorded data, as well as lowering magnitude thresholds. This is an opportunity to develop better nonergodic models (more data means that the nonergodic terms can be better constrained), but this requires estimation methods that can handle large spatial data sets. INLA can fit a nonergodic GMM for an induced data set in Oklahoma, which contains more than 90000 records from about 3600 earthquakes ([Rennolet et al., 2018](#)), in about 3 hours (M. Walling, personal communication). Thus, I believe INLA is a promising tool for the development of nonergodic GMMs for large data sets. Code to fit the INLA models presented herein is available at https://github.com/nikuehn/NonErgGMM_Italy.

Acknowledgements

CV/MSLL and loglikelihood calculation follows ideas from the INLA Google group (<https://groups.google.com/g/r-inla-discussion-group/c/K49zsVx0Yiw/m/GwEQGk1dAwAJ>). Plots are made using the R package `ggplot2` ([Wickham, 2016](#)). The mesh plot was made via the package `inlabru` ([Bachl et al., 2019](#)). Other packages used are `sp` ([Bivand et al., 2013](#)), `sf` ([Pebesma, 2018](#)), `rgdal` ([Bivand et al., 2021](#)). The MS-GWR code uses internally the package `GWmodel` ([Gollini et al., 2015](#)). Spatial plots (Figures 5, 7, and 6) follow the style of [Caramenti et al. \(2020\)](#), from <https://github.com/lucaramenti/ms-gwr/>.

I would like to thank Melanie Walling, Greg Lavrentiadis, Yousef Bozorgnia, Xiaofeng Ming,

and Elnaz Seylabi, and the nonergodic working group at UC Berkeley/UCLA for fruitful discussions about INLA and nonergodic GMMs.

Partial support of Pacific Gas & Electric Company and California Department of Transportation are gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect those of the sponsors.

References

- Abrahamson N, Kuehn N, Walling M, Landwehr N (2019) Probabilistic Seismic Hazard Analysis in California Using Nonergodic Ground Motion Models. *Bulletin of the Seismological Society of America* 109(4):1235–1249, DOI 10.1785/0120190030, URL <https://pubs.geoscienceworld.org/ssa/bssa/article/109/4/1235/571959/>
- Akaike H (1973) Information Theory And An Extension Of The Maximum Likelihood Principle. *Second International Symposium on Information Theory* pp 267–281, DOI 10.1007/978-1-4612-1694-0{_}15
- Al-Atik L, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F, Kuehn N (2010) The Variability of Ground-Motion Prediction Models and Its Components. *Seismological Research Letters* 81(5):794–801, DOI 10.1785/gssrl.81.5.794, URL <http://srl.geoscienceworld.org/cgi/doi/10.1785/gssrl.81.5.794>
- Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T, Donahue JL (2014) NGA-West2 Database. *Earthquake Spectra* 30(3):989–1005, DOI 10.1193/070913EQS197M, URL <http://earthquakespectra.org/doi/abs/10.1193/070913EQS197M>
- Anderson JG, Brune JN (1999) Probabilistic Seismic Hazard Analysis without the Ergodic Assumption. *Seismological Research Letters* 70(1):19–28, DOI 10.1785/gssrl.70.1.19, URL <http://srl.geoscienceworld.org/cgi/doi/10.1785/gssrl.70.1.19>
- Bachl FE, Lindgren F, Borchers DL, Illian JB (2019) inlabru: an R package for Bayesian

spatial modelling from ecological survey data. *Methods in Ecology and Evolution* 10(6):760–766, DOI 10.1111/2041-210X.13168

Bakka H, Rue H, Fuglstad GA, Riebler A, Bolin D, Illian J, Krainski E, Simpson D, Lindgren F (2018) Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics* 10(6):e1443, DOI 10.1002/wics.1443, URL <http://doi.wiley.com/10.1002/wics.1443>

Bivand R, Keitt T, Rowlingson B (2021) rgdal: Bindings for the 'Geospatial' Data Abstraction Library. URL <https://CRAN.R-project.org/package=rgdal>, r package version 1.5-23

Bivand RS, Pebesma E, Gomez-Rubio V (2013) *Applied spatial data analysis with R*, Second edition. Springer, NY, URL <https://asdar-book.org/>

Bivand RS, Gómez-Rubio V, Rue H (2015) Spatial Data Analysis with R - INLA with Some Extensions. *Journal of Statistical Software* 63(20):1–31, DOI 10.18637/jss.v063.i20, URL <http://www.jstatsoft.org/v63/i20/>

Bommer JJ, Douglas J, Strasser FO (2003) Style-of-faulting in ground-motion prediction equations. *Bulletin of Earthquake Engineering* 1(2):171–203, DOI 10.1023/A:1026323123154, URL <http://www.springerlink.com/index/VW58123785485R5R.pdf>

Boore DM (2010) Orientation-Independent, Nongeometric-Mean Measures of Seismic Intensity from Two Horizontal Components of Motion. *Bulletin of the Seismological Society of America* 100(4):1830–1835, DOI 10.1785/0120090400, URL <https://pubs.geoscienceworld.org/bssa/article/100/4/1830-1835/349553>

Bozorgnia Y, Abrahamson NA, Atik LA, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BSJ, Darragh R, Day S, Donahue J, Graves RW, Gregor N, Hanks T, Idriss IM, Kamai R, Kishida T, Kottke A, Mahin SA, Rezaeian S, Rowshandel B, Seyhan E, Shahi S, Shantz T, Silva W, Spudich P, Stewart JP, Watson-Lamprey J, Wooddell K, Youngs R (2014) NGA-West2 Research Project. *Earthquake Spectra* 30(3):973–987, DOI 10.1193/072113EQS209M, URL <http://earthquakespectra.org/doi/abs/10.1193/072113EQS209M>

- Bussas M, Sawade C, Scheffer T, Landwehr N (2015) Varying-coefficient models with isotropic Gaussian process priors. arXiv 0:1–17, URL <http://arxiv.org/abs/1508.07192>
- Bussas M, Sawade C, Kühn N, Scheffer T, Landwehr N (2017) Varying-coefficient models for geospatial transfer learning. *Machine Learning* 106(9-10):1419–1440, DOI 10.1007/s10994-017-5639-3, URL <http://link.springer.com/10.1007/s10994-017-5639-3>
- Caramenti L, Menafoglio A, Sgobba S, Lanzano G (2020) Multi-Source Geographically Weighted Regression for Regionalized Ground-Motion Models. Tech. rep., URL https://mox.polimi.it/publication-results/?id=917%5C&tipo=add_qmox
- Chao SH, Lin CM, Kuo CH, Huang JY, Wen KL, Chen YH (2020) Implementing horizontal-to-vertical Fourier spectral ratios and spatial correlation in a ground-motion prediction equation to predict site effects. *Earthquake Spectra* p 875529302095244, DOI 10.1177/8755293020952449, URL <http://journals.sagepub.com/doi/10.1177/8755293020952449>
- Dambon JA, Sigrist F, Furrer R (2021) Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. *Spatial Statistics* 41:100470, DOI 10.1016/j.spasta.2020.100470, URL <https://linkinghub.elsevier.com/retrieve/pii/S2211675320300646>
- Datta A, Banerjee S, Finley AO, Gelfand AE (2016) On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdisciplinary Reviews: Computational Statistics* 8(5):162–171, DOI 10.1002/wics.1383
- Dawood HM, Rodriguez-Marek A (2013) A Method for Including Path Effects in Ground-Motion Prediction Equations: An Example Using the Mw 9.0 Tohoku Earthquake Aftershocks. *Bulletin of the Seismological Society of America* 103(2B):1360–1372, DOI 10.1785/0120120125, URL <http://www.bssaonline.org/cgi/doi/10.1785/0120120125>
- Esposito S, Iervolino I (2011) PGA and PGV spatial correlation models based on European multievent datasets. *Bulletin of the Seismological Society of America* 101(5):2532–2541, DOI 10.1785/0120110117

- Esposito S, Iervolino I (2012) Spatial Correlation of Spectral Acceleration in European Data. *Bulletin of the Seismological Society of America* 102(6):2781–2788, DOI 10.1785/0120120068, URL <http://www.bssaonline.org/cgi/doi/10.1785/0120120068>
- Finley AO (2011) Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution* 2(2):143–154, DOI 10.1111/j.2041-210X.2010.00060.x
- Finley AO, Datta A, Cook BD, Morton DC, Andersen HE, Banerjee S (2018) Efficient Algorithms for Bayesian Nearest Neighbor Gaussian Processes. *Journal of Computational and Graphical Statistics* 111(514):1–14, DOI 10.1080/10618600.2018.1537924, URL <http://arxiv.org/abs/1702.00434><https://www.tandfonline.com/doi/full/10.1080/10618600.2018.1537924>
- Fotheringham AS, Brunsdon C, Charlton M (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley
- Fotheringham AS, Yang W, Kang W (2017) Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers* 107(6):1247–1265, DOI 10.1080/24694452.2017.1352480
- Franco-Villoria M, Ventrucci M, Rue H (2019) A unified view on Bayesian varying coefficient models. *Electronic Journal of Statistics* 13(2):5334–5359, DOI 10.1214/19-EJS1653
- Fuglstad GA, Simpson D, Lindgren F, Rue H (2019) Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *Journal of the American Statistical Association* 114(525):445–452, DOI 10.1080/01621459.2017.1415907, URL <https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1415907>
- Gelfand AE, Kim HJ, Sirmans CF, Banerjee S (2003) Spatial Modeling With Spatially Varying Coefficient Processes. *Journal of the American Statistical Association* 98(462):387–396, DOI 10.1198/016214503000170
- Gelfand AE, Schmidt AM, Banerjee S, Sirmans CF (2004) Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* 13(2):263–312, DOI 10.1007/BF02595775, URL <http://link.springer.com/10.1007/BF02595775>

- Gollini I, Lu B, Charlton M, Brunsdon C, Harris P (2015) GWmodel : an R package for exploring spatial heterogeneity. *Journal of Statistical Software* 63(17):1–50
- Gómez-Rubio V (2020) Bayesian inference with INLA. Chapman and Hall/CRC, Boca Raton, FL.
- Huang C, Galasso C (2019) Ground-motion intensity measure correlations observed in Italian strong-motion records. *Earthquake Engineering and Structural Dynamics* 48(15):1634–1660, DOI 10.1002/eqe.3216
- Jayaram N, Baker JW (2009) Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics* 38(15):1687–1708, DOI 10.1002/eqe.922, URL <http://onlinelibrary.wiley.com/doi/10.1002/eqe.2230/full><http://doi.wiley.com/10.1002/eqe.922>
- Kowsari M, Sonnemann T, Halldorsson B, Hrafnkelsson B, Snaebjörnsson JP, Jonsson S (2020) Bayesian inference of empirical ground motion models to pseudo-spectral accelerations of south Iceland seismic zone earthquakes based on informative priors. *Soil Dynamics and Earthquake Engineering* 132:106075, DOI 10.1016/j.soildyn.2020.106075, URL <https://linkinghub.elsevier.com/retrieve/pii/S0267726119307432>
- Krainski E, Gómez-Rubio V, Bakka H, Lenzi A, Castro-Camilo D, Simpson D, Lindgren FK, Rue H (2019) Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA. Chapman and Hall/CRC, Boca-Raton, FL
- Kuehn N (2021a) A Primer for using INLA to Estimate Ground-Motion Models. *Engrxiv* pp 1–30, DOI 10.31224/osf.io/6ut3p
- Kuehn N (2021b) Comparison of Bayesian Varying Coefficient Models for the Development of Nonergodic Ground-Motion Models. *Engrxiv* pp 1–25, DOI 10.31224/osf.io/tjxa3
- Kuehn NM, Abrahamson NA (2020) Spatial correlations of ground motion for non-ergodic seismic hazard analysis. *Earthquake Engineering & Structural Dynamics* 49(1):4–23, DOI 10.1002/eqe.3221, URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/eqe.3221><https://onlinelibrary.wiley.com/doi/10.1002/eqe.3221>

- Kuehn NM, Abrahamson NA, Walling MA (2019) Incorporating Nonergodic Path Effects into the NGA-West2 Ground-Motion Prediction Equations. *Bulletin of the Seismological Society of America* 109(2):575–585, DOI 10.1785/0120180260, URL <https://pubs.geoscienceworld.org/ssa/bssa/article/569193/Incorporating-Nonergodic-Path-Effects-into-thehttps://pubs.geoscienceworld.org/ssa/bssa/article/109/2/575/569193/Incorporating-Nonergodic-Path-Effects-into-the>
- Landwehr N, Kuehn NM, Scheffer T, Abrahamson N (2016) A Nonergodic Ground-Motion Model for California with Spatially Varying Coefficients. *Bulletin of the Seismological Society of America* 106(6):2574–2583, DOI 10.1785/0120160118, URL <https://pubs.geoscienceworld.org/bssa/article/106/6/2574-2583/324828>
- Lanzano G, D’Amico M, Felicetta C, Puglia R, Luzi L, Pacor F, Bindi D (2016) Ground-Motion Prediction Equations for Region-Specific Probabilistic Seismic-Hazard Analysis. *Bulletin of the Seismological Society of America* 106(1):73–92, DOI 10.1785/0120150096, URL <http://www.bssaonline.org/lookup/doi/10.1785/0120150096>
- Lanzano G, Luzi L, Pacor F, Felicetta C, Puglia R, Sgobba S, D’Amico M (2019) A Revised Ground-Motion Prediction Model for Shallow Crustal Earthquakes in Italy. *Bulletin of the Seismological Society of America* 109(2):525–540, DOI 10.1785/0120180210, URL <https://pubs.geoscienceworld.org/ssa/bssa/article/568639/A-Revised-GroundMotion-Prediction-Model-forhttps://pubs.geoscienceworld.org/ssa/bssa/article/109/2/525/568639/A-Revised-GroundMotion-Prediction-Model-for>
- Lanzano G, Sgobba S, Caramenti L, Menafoglio A (2021) Ground-Motion Model for Crustal Events in Italy by Applying the Multisource Geographically Weighted Regression (MS-GWR) Method. *Bulletin of the Seismological Society of America* pp 1–17, DOI 10.1785/0120210044, URL <https://pubs.geoscienceworld.org/bssa/article/doi/10.1785/0120210044/606656/Ground-Motion-Model-for-Crustal-Events-in-Italy-by>
- Lavrentiadis G, Abrahamson NA, Kuehn NM (2021) A Non-ergodic Effective Amplitude

- Ground-Motion Model for California. 0123456789, Springer Netherlands, DOI 10.1007/s10518-021-01206-w, URL <http://arxiv.org/abs/2106.07834>
- Lin PS, Chiou B, Abrahamson N, Walling M, Lee CT, Cheng CT (2011) Repeatable Source, Site, and Path Effects on the Standard Deviation for Empirical Ground-Motion Prediction Models. *Bulletin of the Seismological Society of America* 101(5):2281–2295, DOI 10.1785/0120090312, URL <http://www.bssaonline.org/cgi/doi/10.1785/0120090312>
- Lindgren F, Rue H (2015) Bayesian Spatial Modelling with R - INLA. *Journal of Statistical Software* 63(19):1–25, DOI 10.18637/jss.v063.i19, URL <http://www.jstatsoft.org/v63/i19/>
- Lindgren F, Rue H, Lindström J (2011) An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 73(4):423–498, DOI 10.1111/j.1467-9868.2011.00777.x
- Martino S, Riebler A (2020) Integrated Nested Laplace Approximations (INLA) . Wiley StatsRef: Statistics Reference Online pp 1–19, DOI 10.1002/9781118445112.stat08212
- Murakami D, Lu B, Harris P, Brunsdon C, Charlton M, Nakaya T, Griffith DA (2019) The Importance of Scale in Spatially Varying Coefficient Modeling. *Annals of the American Association of Geographers* 109(1):50–70, DOI 10.1080/24694452.2018.1462691, URL <https://doi.org/10.1080/24694452.2018.1462691>
- Paciorek CJ, Schervish MJ (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17(5):483–506, DOI 10.1002/env.785, URL <http://doi.wiley.com/10.1002/env.785>
- Pebesma E (2018) Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal* 10(1):439–446, DOI 10.32614/RJ-2018-009, URL <https://doi.org/10.32614/RJ-2018-009>
- R Core Team (2021) R: A Language and Environment for Statistical Computing. URL <https://www.r-project.org/>

- Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. MIT Press, Cambridge, URL <http://www.gaussianprocess.org/gpml/>
- Rennolet SB, Moschetti MP, Thompson EM, Yeck WL (2018) A Flatfile of Ground Motion Intensity Measurements from Induced Earthquakes in Oklahoma and Kansas. *Earthquake Spectra* 34(1):1–20, DOI 10.1193/101916EQS175DP, URL <http://earthquakespectra.org/doi/10.1193/101916EQS175DP>
- Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71(2):319–392, DOI 10.1111/j.1467-9868.2008.00700.x, URL <http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x>
- Rue H, Riebler A, Sørbye SH, Illian JB, Simpson DP, Lindgren FK (2017) Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application* 4(1):395–421, DOI 10.1146/annurev-statistics-060116-054045, URL <http://www.annualreviews.org/doi/10.1146/annurev-statistics-060116-054045>
- Simpson D, Lindgren F, Rue H (2012) In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* 23(1):65–74, DOI 10.1002/env.1137, URL <http://doi.wiley.com/10.1002/env.1137>
- Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017) Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science* 32(1):1–28, DOI 10.1214/16-STS576, URL <http://arxiv.org/abs/1403.4630><http://projecteuclid.org/euclid.ss/1491465621>
- Sung CH, Abrahamson NA, Kuehn NM, Traversa P, Zentner I (2021) A non-ergodic ground-motion model of fourier amplitude spectra for france. submitted to *Bulletin of Earthquake Engineering*
- Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* 27(5):1413–1432, DOI 10.1007/s11222-016-9696-4, URL <http://link.springer.com/10.1007/s11222-016-9696-4>

- Villani M, Abrahamson NA (2015) Repeatable Site and Path Effects on the Ground-Motion Sigma Based on Empirical Data from Southern California and Simulated Waveforms from the CyberShake Platform. *Bulletin of the Seismological Society of America* 105(5):2681–2695, DOI 10.1785/0120140359, URL <http://www.bssaonline.org/lookup/doi/10.1785/0120140359>
- Walling M, Abrahamson NA (2012) Non-Ergodic Probabilistic Seismic Hazard Analyses. In: 15th World Conference on Earthquake Engineering (15WCEE)
- Watanabe S (2010) Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory. *Journal of Machine Learning Research* 11:3571–3594, URL <http://arxiv.org/abs/1004.2316>
- Watanabe S (2013) A widely applicable bayesian information criterion. *Journal of Machine Learning Research* 14(1):867–897
- Wheeler DC (2014) *Geographically Weighted Regression*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 1435–1459. DOI 10.1007/978-3-642-23430-9_77, URL https://doi.org/10.1007/978-3-642-23430-9_77
- Wheeler DC, Calder CA (2007) An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems* 9(2):145–166, DOI 10.1007/s10109-006-0040-y
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, URL <https://ggplot2.tidyverse.org>
- Wolf LJ, Oshan TM, Fotheringham AS (2018) Single and Multiscale Models of Process Spatial Heterogeneity. *Geographical Analysis* 50(3):223–246, DOI 10.1111/gean.12147