Comparison of Bayesian Varying Coefficient Models for the Development of Nonergodic Ground-Motion Models

Nicolas Kuehn[*1]

[1]*University of California, Los Angeles*

June 13, 2021

Abstract

A nonrgodic ground-motion model explicitly takes systematic local source, path and site effects on the predicted ground-motion into account. With an increasing number of ground-motion records, it s possible to estimate these effects. Landwehr et al. (2016) proposed a varying coefficient model as a tool to estimate nonergodic ground-motion models, based on Gaussian processes. Gaussian processes are computationally expensive, so for large data sets, some approximations have to be made. Here, we compare different Bayesian implementations of varying coefficient models, using the probabilistic programming language Stan (Carpenter et al., 2017), and the integrated nested Laplace approximation (INLA) (Rue et al., 2009). The models are used to fit nonergodic models on the California subset of the NGA-West2 data set (Ancheta et al., 2014). We find that both implementations lead to very similar results, both in the estimated parameters and the predicted ground motions.

1 Introduction

The goal of probabilistic seismic hazard analysis (PSHA) is to estimate the expected annual rate of exceedance of a ground-motion parameter of interest for a particular site. For that purpose, in PSHA one combines a seismic source model (SSC), which describes the temporal and spatial occurrence of events, with a ground-motion model (GMM), which models the distribution of the ground-motion parameter, given a particular earthquake scenario. The value of the standard deviation of aleatory variability associated with the ground-motion distribution controls the slope of the hazard curve, in particular at low exceedance rates (Bommer and Abrahamson, 2006).

GMMs are typically estimated empirically using a regression approach. Since empirical data sets are sparse, one often relies on the ergodic assumption (Anderson and Brune, 1999) in some way, which states that the ground-motion distribution over time at one station is the same as the ground-motion distribution averaged over many stations in a particular region. Due to the increasing number of ground-motion recordings over the last decades, it is possible to relax the ergodic assumption, for example by estimating systematic station terms (e.g. Atkinson, 2006; Rodriguez-Marek et al., 2014). In addition, recent GMMs often include regional adjustment

---

terms (e.g. Stafford, 2014; Kotha et al., 2020; Sedaghati and Pezeshk, 2017; Parker et al., 2020; Kuehn et al., 2020) to account for regional variations in site and distance scaling. These adjustments, however, are still applicable on a large regional scale, and thus employ the ergodic assumption over the regions underlying the models.

Different methods have been proposed to further relax the ergodic assumption with respect to path effects (Sung and Lee, 2016; Anderson and Uchiyama, 2011; Kuehn and Abrahamson, 2020). Dawood and Rodriguez-Marek (2013) proposed to model source-to-site specific path effects as the sum of atteuation over small cells, which was extended by Kuehn et al. (2019) using Bayesian inference to account for the uncertainties associated with the cell-specific attenuation coefficients. Landwehr et al. (2016) proposed a fully nonergodic GMM for California, based on a varying coefficient model (VCM) (Gelfand et al., 2003; Bussas et al., 2015, 2017), where the coefficients of the GMM are functions of source and site location. In the VCM of Landwehr et al. (2016), however, the anelastic attenuation parameter was modeled as dependent only on the event location, and thus did not fully capture path effects, which should depend on both event and station locations. Abrahamson et al. (2019) combined the models of Landwehr et al. (2016) and Kuehn et al. (2019) to calculate nonergodic PSHA for three sites in California. They found that for sites with empirical data in the vicinity, differences between the (mean) ergodic and nonergodic hazard curves can be quite large, while the mean nonergodic hazard curve for a site with no empirical data close-by is similar to the mean ergodic hazard curve, but with much wider fractiles, due to the epistemic uncertainty associated with the nonergodic terms.

Any relaxation of the ergodic assumption means that some pars of the GMM are estimated with few data points; for example, when one calculates a regional adjustment to a coefficient, the region under study has less data than the overall data set, and the coefficient is thus less well constrained. It is thus very important to accurately estimate the epistemic uncertainties associated with any nonergodic term, since otherwise the hazard would be underestimated (Abrahamson et al., 2019).

In the VCM of Landwehr et al. (2016), the spatial dependence of the varying coefficients was modeled using Gaussian processes (GPs) (Rasmussen and Williams, 2006). GPs are a attractive choice for spatially varying parameters since they provide a flexible framework o model the spatial dependence, while also providing good quantification of the associated uncertainty. In Bayesian nonparametrics (Hjort et al., 2010), a GP can be thought of as a prior over functions. Hence, we do not need to know the exact functional form of the spatial dependence; it is estimated from data, with some constraints set by the form of the GP. The GP formulation also lends itself to a good quantification of uncertainty in ranges where data is sparse (Landwehr et al., 2016) – this uncertainty can be calculated as the posterior distribution of the function describing the spatial dependence.

While GPs provide an elegant mathematical framework for estimating nonergodic GMMs and assess the associated uncertainty, they do, however, have the drawback that they do not scale well to large data sets. In particular, estimating the parameters of a GP (as well as predicting new data points using a GP) requires the calculation of the inverse of large covariance matrices, which is computationally expensive. For large data sets, this requires some approximations to the covariance matrix. In the future, the size of strong-motion data sets will continue to increase, as more data becomes available. This offers many modeling opportunities, since more data allows one to better constrain nonergodic effects. On the other hand, this also poses computational challenges, as models with spatial components are harder to estimate. It is therefore necessary to explore how nonergodic models can be fit efficiently for large data sets.

Landwehr et al. (2016) estimated the parameters of their nonergodic GMM by maximizing the marginal likelihood of the data (cf. Chapter 5.4.1 in Rasmussen and Williams (2006)), which gives point estimates of the parameters. However, parameters are often associated with considerable uncertainty, which should not be ignored. Hence, instead of point estimates it is
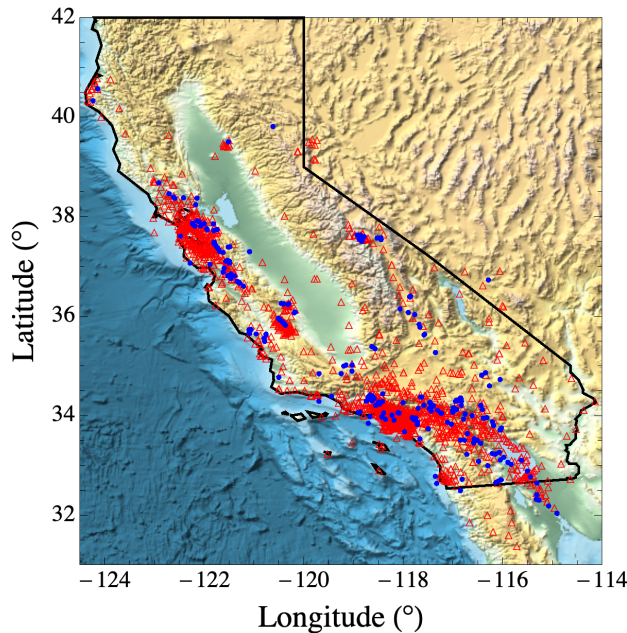
Figure 1: Caption

better to work with the posterior distribution of all parameters, and estimate the models via full Bayesian inference. Here, we compare different implementations of Bayesian nonergodic models (a combination of a VCM with the cell-specific attenuation model of (Kuehn et al., 2019)). We estimate nonergodic models using three different methods: (1) a full Bayesian approach without approximation to the GP, estimated via Hamiltonian Monte Carlo (HMC) sampling (Neal, 2011; Betancourt, 2017; Betancourt and Girolami, 2015); (2) a full Bayesian approach using the nearest neighbour GP, (Datta et al., 2016a), which is an approximation too a full GP, estimated via HMC; and (3) a full Bayesian approach using the integrated nested Laplace approximation (INLA) (Rue et al., 2009). We compare the resulting nonergodic models, in terms of their estimated parameters, and their predictions.

2   Data

The parameters of the nonergodic adjustment models are estimated using the Western US data from the model of Abrahamson et al. (2014). The dataset comprises 12018 records from 274 earthquakes and 1479 stations. For simplicity, we use the same data set for each of the three ergodic base GMMs (ASK14, BSSA14, CB14). The event and station locations are is shown in Figure 1. We estimate the nonergodic adjustment models for a pseudo-spectral acceleration (PSA) with a period of $T = 0.02$s. The median predictions for the ergodic base GMMs are calculated using the python package `pygmm` (Kottke, 2016).

For the cell-specific attenuation, we use the same cells as Kuehn et al. (2019), who used cells of about 25×25km. The cell-specific distances are calculated as a straight line from the closest point on the rupture to the station. Hence, the The coordinates of the events are the latitude and longitude of the hypocenter coordinates in degrees.

3 Nonergodic GMMs and PSHA

The general approach to performing nonergodic PSHA is outlined in e.g. Abrahamson et al. (2019); Villani and Abrahamson (2015); Kuehn and Abrahamson (2020). The key difference between an nonergodic and an ergodic ground-motion model is that the nonergodic GMM explicitly takes into account repeatable, systematic source, path and site effects. These systematic effects are location specific; hence, the median prediction of a GMM becomes location dependent

$$\mu_{nonergodic}(\vec{x}, \vec{t_e}, \vec{t_s}) = \mu_{ergodic}(\vec{x}) + f_{eq}(\vec{t_e}) + f_{path}(\vec{t_e}, \vec{t_s}) + f_{stat}(\vec{t_s}) \quad (1)$$

where $f_{eq}(\vec{t_e}), f_{path}(\vec{t_e}, \vec{t_s}), f_{stat}(\vec{t_s})$ are the location-specific source, path and site effects, respectively; and $\vec{t_e}$ and $\vec{t_s}$ are the event and site coordinates. $\mu_{ergodic}(\vec{x})$ is the prediction of an ergodic GMM, and $\vec{x}$ is a vector of predictor variables such as magnitude, distance, time averaged shear wave velocity in the upper 30m V_{S30}, and so on.

For the current study, location-specific nonergodic adjustment terms are calculated for the empirical GMMs of Abrahamson et al. (2014); Boore et al. (2014); Campbell and Bozorgnia (2014), hereafter called ASK14, BSSA14, and CB14. The nonergodic adjustment terms are based on a varying coefficient model Landwehr et al. (2016); Bussas et al. (2017), and the cell-specific attenuation model Dawood and Rodriguez-Marek (2013); Kuehn et al. (2019). We do not calculate adjustment terms for the model of Chiou and Youngs (2014), because it models the anelastic attenuation term as magnitude dependent, which is difficult to incorporate into the cell-specific attenuation model.

The nonergodic adjustment models are estimated via regression from California data. We use two main methods (different programs) to estimate the nonergodic models: (1) Bayesian inference via Markov Chain Monte Carlo (MCMC) sampling using the program Stan (Carpenter et al., 2017); (2) Bayesian inference vie the R-INLA package (Rue et al., 2009, 2017; Bivand et al., 2015). For Stan, we describe an exact GP model model, as well as the nearest neighbour Gaussian process approximation (Datta et al., 2016a). There are some differences in the methods, described below. From a practical point of view, the main differences are in the runtime of different approaches. The exact version of Stan takes about 120 hours for the California dataset, the nearest neighbour approximation runs for about 20 hours, and INLA runs about 2 hours.

3.1 Nonergodic Adjustment Models

This section describes the formulation of the nonergodic adjustment models.

We assume that we have an ergodic GMM, whose median prediction can be calculated by

$$\mu_{GMM} = f_{GMM}(\vec{c}; \vec{x}) \quad (2)$$

where $\vec{x}$ is a vector of predictor variables such as magnitude, distance and so on; and $\vec{c}$ is a vector of coefficients of the model. We want to make this GMM nonergodic, which means that we need to adjust the median prediction with systematic, location-specific nonergodic adjustment terms.

The adjustment terms are an earthquake-specific constant; a site-specific constant; and a cell-specific anelastic attenuation term. Following Landwehr et al. (2016), the constants are modeled as spatially correlated (i.e., they are distributed according to a Gaussian process). The geometrical spreading is modeled as ergodic (i.e. not spatially varying); this is in line with most partially nonergodic models, which only regionalize the constant, linear site scaling, and anelastic attenuation. The cell-specific anelastic attenuation coefficient is modeled as in Kuehn et al. (2019), following Dawood and Rodriguez-Marek (2013). We also include an intercept, to account for any model misspecification.

Hence, the full model can be written as

$$\ln Y = c_0 + f_{GMM}(\vec{c}; \vec{x}) + \delta c_{eq}(\vec{t}_e) + \delta c_{stat}(\vec{t}_s)$$

$$- f_{attn,GMM}(R_{RUP}) - \sum_{j=1}^{N_C} c_{ca,j} \Delta R_j$$

$$+ \delta B_0 + \delta W_0 + \delta S2S_0 \tag{3}$$

The $\delta c_{eq/stat}$-terms are the location-specific adjustment terms for evens and stations, and $\vec{t}_{e/s}$ are the event and station locations. The term $f_{attn,GMM}(R_{RUP})$ is the anelastic attenuation function in the GMM; it is taken out and replaced by the cell-specific attenuation, $\sum_{i=1}^{N_C} c_{ca,i} \Delta R_i$, where $c_{ca,i}$ is the coefficient for cell i, and R_i is the length of the path within the ith cell. We constrain the cell-specific attenuation coefficients to be positive, so the full attenuation term has a negative sign. $\delta B_0, \delta S2S_0, \delta W_0$ are the remaining, spatially uncorrelated between-event, site-to-site, and within-event/within-site residual. These residuals are assumed to follow normal distributions with between-event variability τ_0, site-to-site variability $\phi_{S2S,0}$, and within-event/within-site variability ϕ_0.

$$\delta B_0 \sim N(0, \tau_0) \tag{4}$$
$$\delta S2S_0 \sim N(0, \phi_{S2S,0}) \tag{5}$$
$$\delta W_0 \sim N(0, \phi_0) \tag{6}$$

The location-specific constants ($\delta c_{eq}(\vec{t}_e)$ and $\delta c_{site}(\vec{t}_s)$) are distributed according to a Gaussian process prior, as in Landwehr et al. (2016)

$$\delta c \sim GP(0, k(\vec{t}, \vec{t'}) \tag{7}$$

where $k(\vec{t}, \vec{t'})$ is the covariance function that determines how the adjustment terms can vary. Since the δc are adjustments to an ergodic base model, a zero mean function is appropriate. Depending on the method to estimate the ergodic adjustment terms, we either use an exponential covariance function (for estimation with Stan) or a Matérn covariance function (for estimation with INLA).

In CB14, the anelastic attenuation coefficient is applied to records wih $R_{RUP} > 80$km. We apply the same constraint to the cell-specific attenuation model, if the base model is CB14. Hence, in this case the cell-specific distance ΔR_j is zero if the total distance up that point is less than 80km.

A quick word on notation and terminology: In the following, the adjustment terms for events δc_{eq}, and stations δc_{stat}, are treated mathematically the same. Hence, we often refer to them in the text as δc, which implies that the corresponding explanation applies to both event and station adjustments. Similarly we use $\vec{t}$ for the locations when speaking generally about the model. Parameters that govern the distribution of other parameters (such as τ governs the distribution of event terms) are called hyperparameters. We refer to parameters whose distribution is controlled by hyperparameters as latent parameters or effects. Latent parameters are for example the event and station terms, as well as the adjustment terms δc.

Below, we describe the derivations of the nonergodic models, using both Stan and INLA. We provide code to run the models using the computer environment R (R Core Team, 2020) at https://github.com/nikuehn/NonergodicGMMS_STAN_INLA.

3.2 Derivation of Nonergodic Parameters using Stan

First, we describe how the parameters of the nonergodic adjustment model described in the previous section are estimated using full Bayesian inference, via MCMC sampling with the

program Stan (Carpenter et al., 2017). The covariance function we use is the exponential covariance function, as in Landwehr et al. (2016)

$$k(\vec{t}, \vec{t'}) = \psi^2 \exp\left(-\frac{|\vec{t} - \vec{t'}|}{\ell}\right) \tag{8}$$

where ψ determines the range how much the adjustment terms can vary, and ℓ is a spatial correlation length-scale. Depending on whether the GP is modeling the source or site-specific adjustment terms, we use either the event or site locations in Equation (8).

In general, one can integrate out all latent adjustment terms (non-ergodic adjustment terms as well as event and site terms) in Equation (3), which leads to a model where the target variable is distributed according to a multivariate normal distribution

$$\vec{Y} \sim N(\mu, \boldsymbol{\Sigma}) \tag{9}$$

where

$$\mu_i = c_0 + f_{GMM}(\vec{c}; \vec{x}_i) - f_{attn,GMM}(R_{RUP,i}) - \sum_{j=1}^{N_C} c_{ca,j} \Delta R_{i,j} \tag{10}$$

The covariance matrix is

$$\boldsymbol{Sigma}_{ij} = \delta_{ij}\phi_0^2 + \delta_{eq(i),eq(j)}\tau_0^2 + \delta_{stat(i),stat(j)}\phi_{S2S,0}^2 + k_{eq}(\vec{t}_e, \vec{t'}_e) + k_{stat}(\vec{t}_s, \vec{t'}_s) \tag{11}$$

where δ_{ij} equals 1 if $i == j$ and zero otherwise, and $eq(i)$ and $stat(i)$ are indices connecting an event or station with record i. However, $\boldsymbol{\Sigma}$ is of size $N \times N$, and is not sparse, so it is impractical to work with Equation (9), since one has to calculate the inverse of Σ multiple times during the estimation of the model parameters.

If we do not integrate out the latent parameters, but treat them as parameters to be estimated, then the likelihood of the data factorizes, conditional on the latent effects. Since the posterior distribution of the parameters ($\vec{\theta}$, which encompasses all parameters and latent parameters of the model) is proportional to the product of the likelihood and the prior distribution, we can write the full log-posterior distribution (up to a constant) as

$$\begin{aligned}
\ln p(\vec{\theta}|D) = &\ln p(c_0, \tau_0, \phi_{S2S,0}, \phi_{SS,0}, \psi_{eq}, \ell_{eq}, \psi_{stat}, \ell_{stat}, \mu_{ca}, \psi_{ca}) \\
&+ \sum_{i=1}^{N_E} \ln p(\delta B_{0,i}|0, \tau) + \sum_{j=1}^{N_S} p(\delta S2S_{0,j}|0, \phi_{S2S}) \\
&+ \ln p(\vec{\delta c}_{eq}|\vec{0}, \boldsymbol{K}_{eq}) + \ln p(\vec{\delta c}_{stat}|\vec{0}, \boldsymbol{K}_{stat}) \\
&+ \sum_{j=1}^{N_C} p(c_{ca,j}|\mu_{ca}, \sigma_{ca}) - \sum_{j=1}^{N_C} \ln(1 - cdf(0|\mu_{ca}, \psi_{ca})) \\
&+ \sum_{k=1}^{N} \ln p(y_k|\mu_k + \delta B_{eq(k)} + \delta S2S_{stat(k)} + \delta c_{eq(k)} + \delta c_{stat(k)}, \phi_{SS})
\end{aligned} \tag{12}$$

where D denotes the data, and $p(\cdot|m, s)$ is the probability density function of a normal distribution with mean m and standard deviation s (or a multivariate normal distribution if the argument is a vector). In Equation (12), the first line describes the prior distribution for the intercept c_0 and the hyper-parameters of the model (i.e., parameters that govern the distribution of the latent effects). The second line describes the prior distributions for the systematic, non-spatially varying event and station terms. These are independent of each other, and are normally distributed with mean zero and standard deviation τ_0 and $\phi_{S2S,0}$.

The third line in Equation (12) is the prior distribution of the systematic event and station terms, which is a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{K}_{eq/sat}$ defined by the covariance function (cf. Equation (8)). The covariance matrix $\boldsymbol{K}_{eq}$ has size $N_E \times N_E$.

The fourth line in Equation (12) is the prior distribution for the cell-specific attenuation models. We follow the formulation of Kuehn et al. (2019), and assume that the cell-specific attenuation coefficients are distributed according to a truncated normal distribution. Here, we constrain the parameters to be positive, so the prior distribution is lower truncated at zero, and the second term is the required adjustment to the untruncated density. A Stan program consists of statements that increments the log-density for each component of the model, so Equation (12) is straightforward ti implement.

For GPs, informative prior distributions are important, to constrain the range of the parameters, and help to avoid possible trade-offs. For the intercept c_0, we do not believe that it is very large, since the models should be approximately centered. Hence, we give it a normal prior with a small standard deviation

$$c_0 \sim N(0, 0.1) \tag{13}$$

For the standard deviation parameters $\phi_0, \tau_0, \phi_{S2S,0}$, we use a lognormal distribution that discourages values below 0.2 and above 1.2.

$$\phi_0 \sim \text{lognormal}(\ln(0.5), 0.5) \tag{14}$$
$$\tau_0 \sim \text{lognormal}(\ln(0.5), 0.5) \tag{15}$$
$$\phi_{S2S,0} \sim \text{lognormal}(\ln(0.5), 0.5) \tag{16}$$

The rationale for the prior distributions for the scale parameters of the GPs, ψ, is to penalize too large values of ψ, since these would be a strong deviation from the ergodic base model (Simpson et al., 2017). We do believe that the nonergodic values exist, but we need to see strong evidence in the data to model them. Hence, we use an exponential distribution as the prior for these parameters

$$\psi_{eq} \sim \text{exponential}(20) \tag{17}$$
$$\psi_{stat} \sim \text{exponential}(20) \tag{18}$$

Kuehn et al. (2020) compared prior distributions for regional partially nonergodic parameters based on an exponential distribution and a half-cauchy distribution, which is sometimes used as a prior for standard deviations in Bayesian hierarchical models (Gelman, 2006). They found that the exponential distribution gave better results.

For the length-scale ℓ, we use an inverse Gamma distribution as a prior distribution [1], since this discourages both very small and very large length-scales. We scale the parameters of the inverse Gamma distribution such that length scales below 0.06 and above 1.2 are discouraged. This leads to the following prior distributions

$$\ell_{eq} \sim \text{inv\_gamma}(3, 0.5) \tag{19}$$
$$\ell_{stat} \sim \text{inv\_gamma}(3, 0.5) \tag{20}$$

For the cell-specific attenuation coefficients, we constrain them to be positive, and then use a minus sign in Equation (3). The cell-specific coefficients c_{ca} are distributed according to a normal distribution with mean $\mu_{\psi\,ca}$ and standard deviation $\psi_{c\,ca}$. The prior distribution for the mean attenuation is a lognormal distribution (which is why we constrain the parameters to

---

[1] based on https://betanalpha.github.io/assets/case_studies/gp_part3/part3.html

be positive) with median $\ln 0.011$, while the prior distribution for the standard deviation is a truncated Student-t distribution with degrees of freedom $\nu = 3$.

$$c_{ca} \sim N(\mu_{c\,ca}, \psi_{c\,ca})\,T(0,) \tag{21}$$

$$\mu_{c\,ca} \sim LN(-4.5, 0.5) \tag{22}$$

$$\psi_{c\,ca} \sim \text{student-t}(3, 0, 0.01)\,T(0,) \tag{23}$$

3.2.1 Nearest Neighbour Gaussian Process

In the Stan code of the model, the computational bottleneck is the calculation of the covariance matrices $\boldsymbol{K}_{stat}$ and $\boldsymbol{K}_{eq}$ and their Cholesky decompositions. The size of these matrices (in particular $\boldsymbol{K}_{stat}$) leads to long runtimes of the Stan program. One way to overcome this issue is to use an approximation to the covariance matrices. Here, we use a nearest neighbour Gaussian process (NNGP) (Datta et al., 2016a,b) to approximate the original GP. The basic idea behind the NNGP is to represent the multivariate variables $\vec{\delta c}$ as a directed acyclic graph (DAG)

$$p(\vec{\delta c}) \sim p(\delta c_1) \prod_{i=2}^{N} p(\delta c_i | \text{Pa}[i]) \tag{24}$$

where $\delta c_i = \delta c(\vec{t_i})$, and $\text{Pa}[i]$ is the set of parents for δc_i. In the NNGP, the size of $\text{Pa}[i]$ is shrunk to a small number (corresponding to the nearest neighbors of $\vec{t_i}$), leading to a sparse covariance matrix, which can be exploited computationally. For details on the NNGP approximation, see (Datta et al., 2016a,b; Finley et al., 2018; Zhang et al., 2018; Zhang, 2018). The calculation of the likelihood of the NNGP is taken from Zhang (2018).

For our models, we use a number of 10 nearest neighbours for events and stations. We use the same prior distributions as in the Stan program without the NNGP approximation.

3.2.2 Running the Stan Programs

We run the Stan programs using the command line interface `cmdstan`. We run four chains for each case, with 1000 warm-up samples and 1000 samples post warm-up.

3.3 Derivation of Nonergodic Parameters using INLA

In the previous section, we described how the nonergodic parameters are calculated using Stan. Via Stan, the parameters are estimated using MCMC sampling, which means that the outcome of the program are samples of the full posterior distribution of all parameters. From these samples, one can assess summary statistics such as mean, median, standard deviation for each parameter, or correlations between different parameters.

By contrast, INLA (Rue et al., 2009) provides a deterministic method to perform Bayesian inference by approximating posterior marginal distributions. The INLA method is implemented in the R-package R-INLA (www.r-inla.org) (Rue et al., 2017). For an introduction to INLA and example models, see e.g. Krainski et al. (2018); Gomez-Rubio (2020). In particular, INLA can be used for efficiently estimating spatial models combining latent Gaussian models with stochastic partial differential equations (SPDEs) (Lindgren et al., 2011; Lindgren and Rue, 2015; Bakka et al., 2018).

Using INLA for the estimation of spatial models requires a slight change of the covariance function. In Equation (8), we have used an exponential covariance function. The covariance function used in INLA is a Matérn covariance function

$$K(\vec{t}, \vec{t'}) = \psi^2 \frac{2^{(1-\nu)}}{\Gamma(\nu)} (\kappa |\vec{t} - \vec{t'}|)^\nu K_\nu(\kappa |\vec{t} - \vec{t'}|) \tag{25}$$
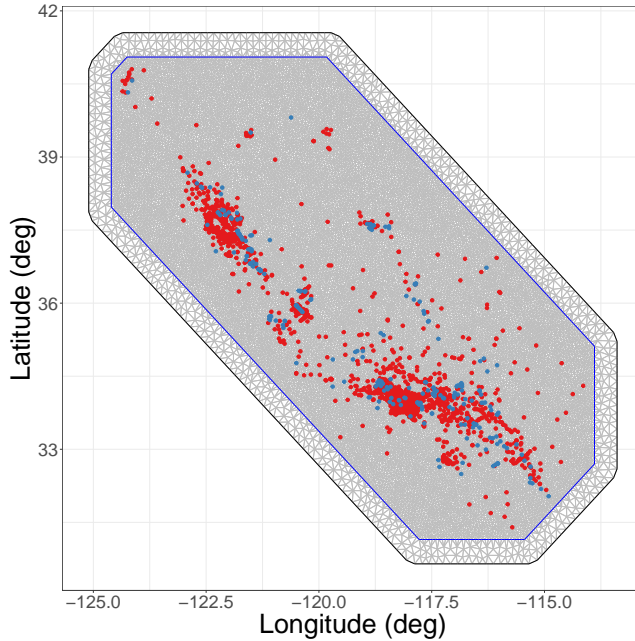
Figure 2: Mesh used for the INLA models. Red dots are the stations, blue dots are the events.

where 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). The default value in INLA for GPs with two-dimensional coordinates is $\nu = 1$ (Lindgren and Rue, 2015; Bakka et al., 2018). Other values are possible, but require an additional approximation.

Spatial modeling with INLA approximates a spatial process (in our case, the systematic event and station terms) with a basis function approximation (Lindgren et al., 2011; Krainski et al., 2018). The basis functions are evaluated on nodes of a triangular mesh. We generate a mesh that covers the state of California and the observed events and stations, We use a mesh that encompasses the observed data points, with an outer bound of 0.5 degrees. The maximum triangle edge length is 0.05 degrees, and 0.25 degrees in the outer bound. The final mesh consists of 17794 nodes and is shown in Figure 2.

Based on this mesh, we fit the model and estimate the parameters. The parameters are the same as before, but the systematic terms now are distributed according to a GP with Matérn covariance with $\nu = 1$. The prior specification is different compared to the Stan implementation. In INLA, the normal distribution is parameterized in terms of mean and precision, which is defined as $prec = \frac{1}{\sigma^2}$ (i.e. precision is one over variance). For the variance components of the non-spatially varying terms $(\tau_0, \phi_0, \phi_{S2S,0})$, we place Gamma distribution on the corresponding precision parameters. The prior distributions have shape parameter 2 and rate parameter 0.5

$$\ln prec_\tau \sim G(2, 0.5) \tag{26}$$

$$\ln prec_{\phi_0} \sim G(2, 0.5) \tag{27}$$

$$\ln prec_{\phi_{S2S,0}} \sim G(2, 0.5) \tag{28}$$

Thus, the prior distributions for the precisions have a mean of 4, which corresponds to a value of the standard deviation of 0.5, which is similar to the prior specification that was used in the Stan implementation.

We again assume that the base models should be approximately centered, and assign a normal distribution with mean zero and precision 100 to the intercept c_0.

The prior values for the parameters controlling the spatially varying parameters are so-called penalized complexity (PC) priors (Simpson et al., 2017; Fuglstad et al., 2019; Franco-Villoria

Table 1: Model selection criteria for ergodic and nonergodic NLA models.

GMM	WAIC (ergodic)	WAIC (VCM)	CPO (ergodic)	CPO (VCM)
ASK14	20044.2	18982.8	10074.4	9541.43
BSSA14	19923.3	18857.9	10016.	9480.
CB14	19922.4	19361.2	10014.2	9729.19

et al., 2018). These priors shrink the model towards a base model without a spatial component (i.e. they penalize the extra complexity of the spatial components). The simpler base model has zero variance and infinite range, where the range is defined as $\ell = \frac{\sqrt{8\nu}}{\kappa}$. The PC prior can be specified through the tail probabilities $P(\psi > \psi_0) = \alpha_1$ and $P(\ell < \ell) = \alpha_2$. We assume the same values for both the spatial model associated with the event and station terms. We use the following prior specification for the PC-prior of the spatially varying coefficients

$$P(\psi_{eq/stat} > 0.23) = 0.01 \tag{29}$$
$$P(\ell_{eq/stat} < 0.0.45) = 0.9 \tag{30}$$
$$\tag{31}$$

These values are derived from the prior distributions on the spatial hyperparameters used in the Stan program. For details on penalized complexity priors in the context of VCMs, see Franco-Villoria et al. (2019).

The anelastic attenuation term is implemented in the INLA model as a fixed effect (corresponding to μ_{ca} in Equation (22)) and cell-specific adjustment coeffients. The prior distribution for the fixed effect is a normal distribution with mean -0.01 and precision 10000. The cell-specific adjustment coefficients are implemented as generic random effects, since they can be written as

$$\sum_{i=1}^{N_C} c_{ca,i} \Delta R_i = \boldsymbol{Z} \vec{c}_{ca} \tag{32}$$

where $\boldsymbol{Z}$ is a matrix of size $N \times N_C$, whose i, jth entry corresponds to the fraction of path in cell j for record i. The R-INLA package contains a module to specify random effects of the form $\boldsymbol{Z}\vec{c}$, where $\boldsymbol{Z}$ is the design matrix of the random effects, and $\vec{c}$ is a vector of random effects.

4 Results

In this section, we briefly summarize results of the different models. We compare the posterior distributions of the estimated parameters, for the different base models and the different estimation procedures. For the Stan based models, we assess the posterior distribution via the post warm-up samples from the MCMC chains, while the outcome of INLA are approximations of the marginal posterior distributions for the different parameters. For comparison, we also estimated a model without spatially varying coefficients with INLA, i.e. a model that only accounts for an intercept, event terms δB and station terms δS. In traditional GMM terminology, such a model would be called partially ergodic, since it accounts for systematic station terms, but for brevity we call this model ergodic.

In Table 1, we compare the model fit of the ergodic INLA models and the INLA VCMs, based on the widely applicable information criterion (WAIC, Watanabe, 2013), and the Conditional Predictive Ordinate (CPO, Pettit, 1990). For both criteria, lower values indicate a better fit. In Table (1), the VCMs perform better than the ergodic models, which is in line with previous results by e.g. Landwehr et al. (2016); Kuehn et al. (2019), and also aligns with our expectation that moving from an ergodic to nonergodic models should lead to model improvement.
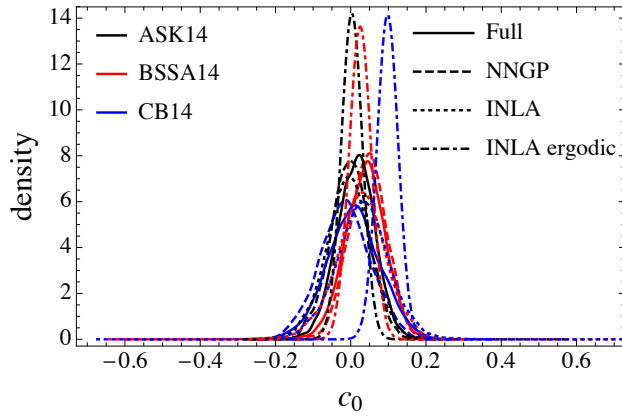
Figure 3: Posterior density of the intercept of nonergodic models for different base GMMs and estimation procedures. For comparison, the intercepts of ergodic models are also shown.

Overall, the results are quite similar between the full Stan model, the NNGP Stan model, and the INLA model. Depending on the parameter, there can be some differences between the adjustment models based on the different base models. This is not really surprising, since the NGA West 2 models can be quite different for small magnitudes, and thus the event and station terms for individual events/stations can be strongly affected by different scaling of the base models.

Figure 3 shows the posterior distributions of the intercept c_0. In general, all models show similar behavior. We can also see that the intercept is different between the ergodic and nonergodic models. This can be due to the fact that an ergodic models can be biased towards regions where data is more abundant. One can also see that the posterior distribution for the VCMs is a bit wider than for the ergodic model. This is not surprising, since the model is distributed over more terms, which leads to larger uncertainty.

Figure 4 shows the posterior distributions of the standard deviations of the non-spatially varying terms (δB_0, δS_0, δW_0), i.e. of τ_0, $\phi_{S2S,0}$, $\phi_{SS,0}$. Again, we also show the results of an ergodic INLA run for comparison. As expected, we get a reduction for the nonergodic models compared to the ergodic runs. The largest decrease occurs for ϕ_{S2S}, which is line with results of Chao et al. (2020). The results between the different approaches and base models are quite similar, in particular for within-event/within-station standard deviation ϕ_0, and site-to-site standard deviation $\phi_{S2S,0}$. For CB14, there is less of a reduction in ϕ_0, which is probably due to the fact that the anelastic attenuation term only applies for distances $R_{RUP} > 80$km in CB14. For the between-event standard deviation τ_0, the differences are a bit larger. On reason could be that the between-model variability of the three base models is larger a small magnitudes, where the bulk of the data is. Overall, though, the differences between the Stan results and INLA results is quite small. In particular, the full Bayesian model and the NNGP model produce almost identical results.

The posterior distributions of the hyperparameters describing the spatially varying event and station terms are shown in Figure 5 and 6. The standard deviations of the systematic event terms (ψ_{eq}) are quite consistent, but there are some differences in the posterior distributions of the length scales. In particular, the INLA models lead to larger length scales. This is probably due to the different prior specification, as the PC prior used in INLA tries to shrink the estimates towards a simpler base model with infinite spatial range. A similar picture emerges for the hyperparameters associated with the systematic station terms, with INLA models having larger length scales. In this case, the models based on CB14 also have lager systematic standard deviations ψ_{stat}. This might be due to the fact that the anelastic attenuation in CB14 only applies to records with distances $R > 80$km. For the other base models (ASK14 and BSSA14),
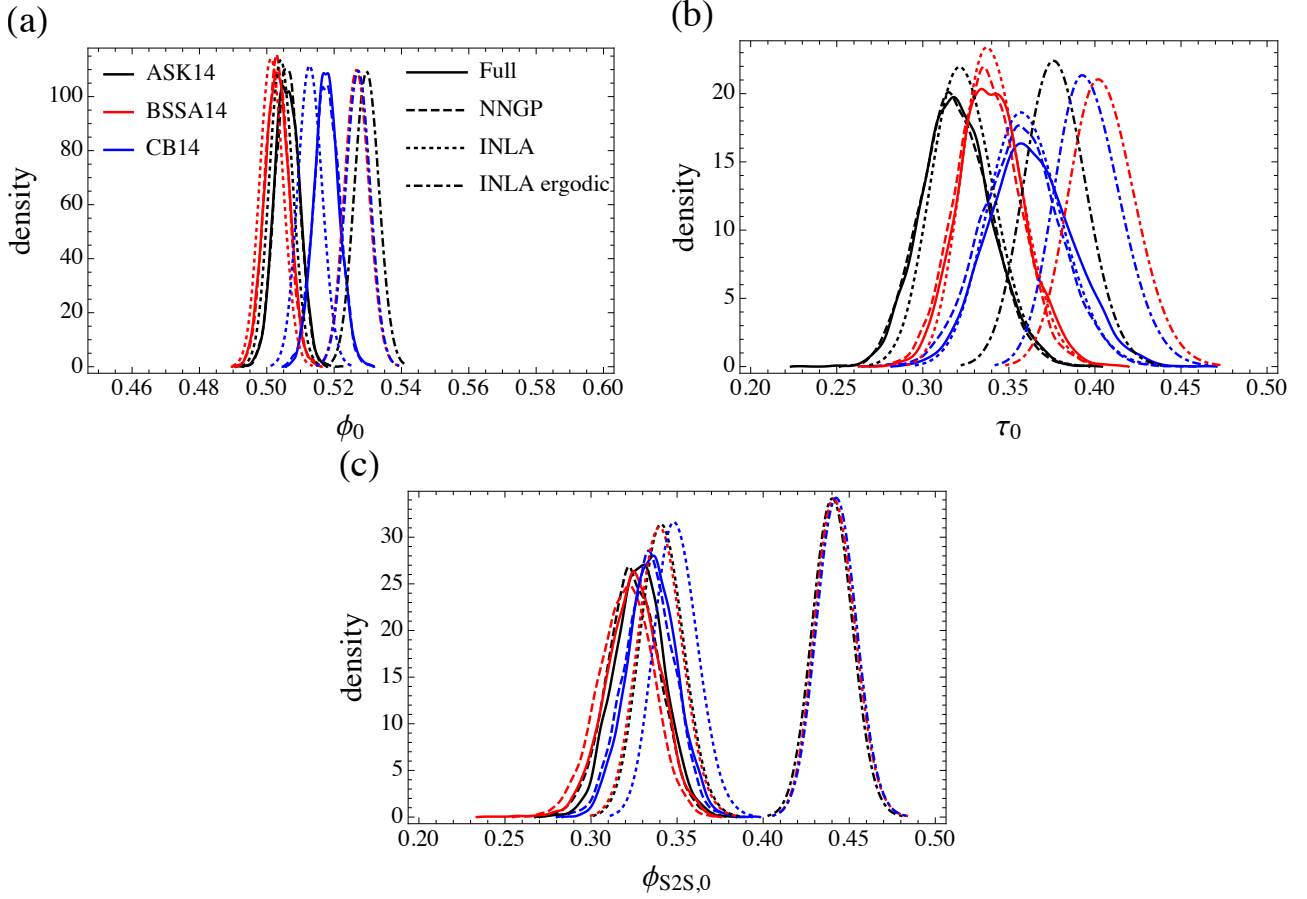
Figure 4: Posterior density of the standard deviations of the non-spatially varying terms, for different base GMMs and estimation procedures: (a) aleatory within-event/within-site variability ϕ_0; (b) aleatory between-event variability τ_0; (c) epistemic aleatory between-site variability $\phi_{S2S,0}$. For comparison, the distributions for ergodic models are shown as well.
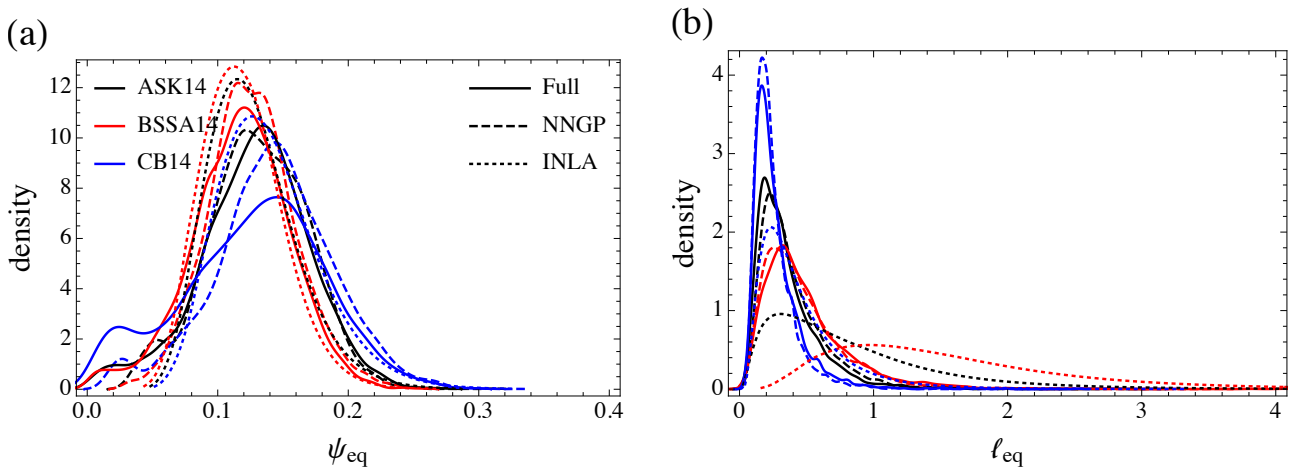

Figure 5: Posterior densities of (a) the standard deviation; and (b) the length scale associated with spatially varying event terms.
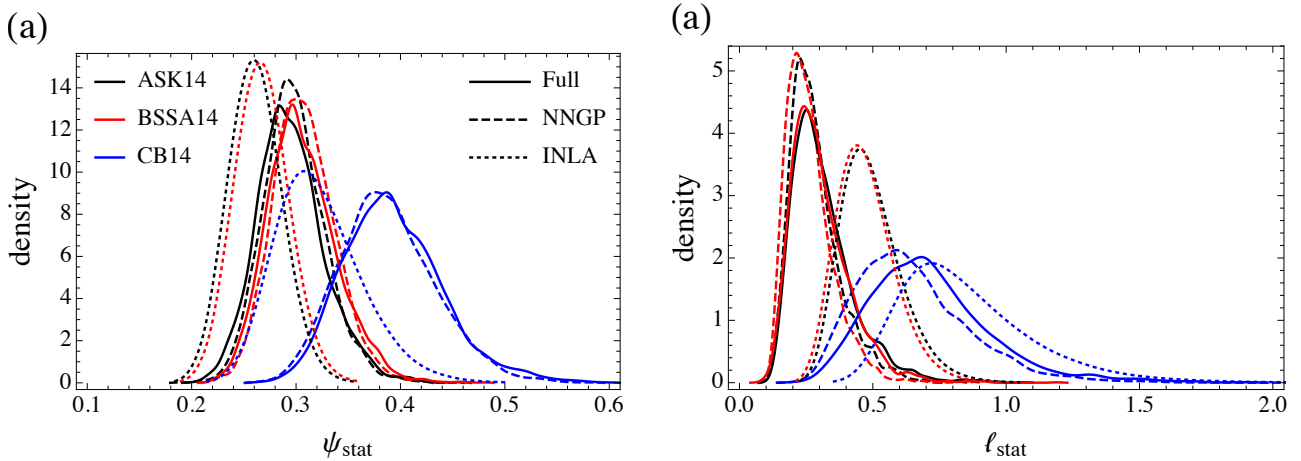
Figure 6: Posterior densities of (a) the standard deviation; and (b) the length scale associated with spatially varying station terms.
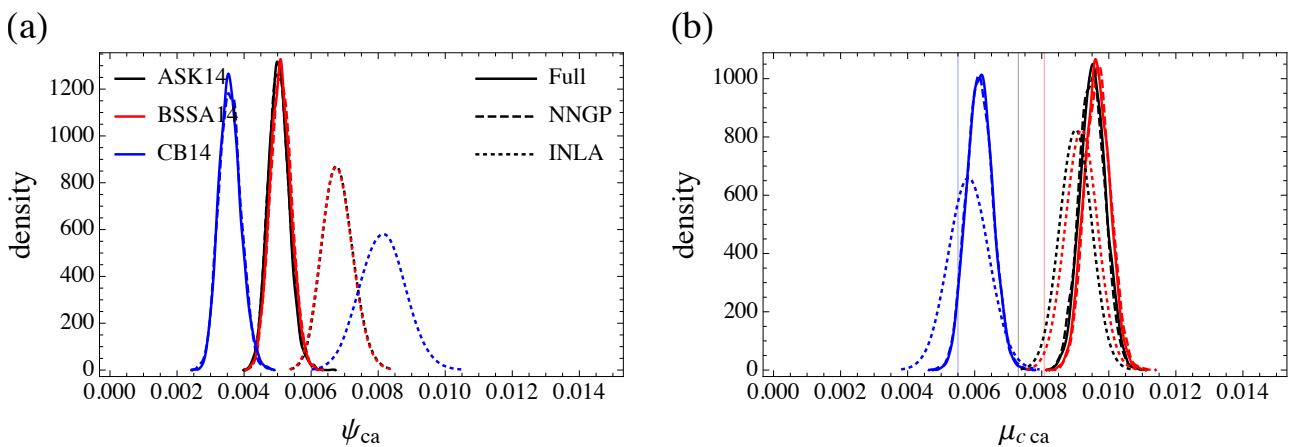

Figure 7: Posterior densities of (a) the standard deviation of the cell-specific atenuation coefficients; (b) the mean anelastic attenuation. Vertical bars in (b) are the anelastic attenuation coefficients of the ergodic base models.

some of those records are affected by the cell-specific attenuation, while in CB14 some of the systematic effects might be mapped into station terms.

The parameters controlling the anelastic attenuation are shown in Figure 7. We can see a similar picture as before, with almost identical results between the full Stan model and the NNGP approximation, and only small differences on the INLA models. As already observed by Kuehn et al. (2019), the differences for cell-specific parameters based on ASK14 and BSSA14 are almost negligible. Similar to other parameters the CB14 model shows a slightly different behavior, due to its different anelastic attenuation model. Figure 7(b) shows the mean attenuation $\mu_c a$, and for comparison the values of the (absolute) ergodic attenuation coefficients. These are different, due to the fact that the majority of the data is from Southern California, so the ergodic attenuation coefficients are biased towards Southern California. Kuehn et al. (2019) found that the ergodic attenuation coefficient is a weighted average of the cell-specific attenuation coefficients, with weights corresponding to the number of paths in each cell.

Figure 8 shows the difference in the cell-specific attenuation coefficients, based on the full Stan model, the NNGP model, and the INLA model. As previously seen, the full Stan model and the NNGP Stan model give very similar results. The INLA coefficients generally follow a similar trend; however, we can see that there are a few cell-specific coefficients that are negative, which leads to a positive anelastic attenuation, which is unphysical. For the Stan
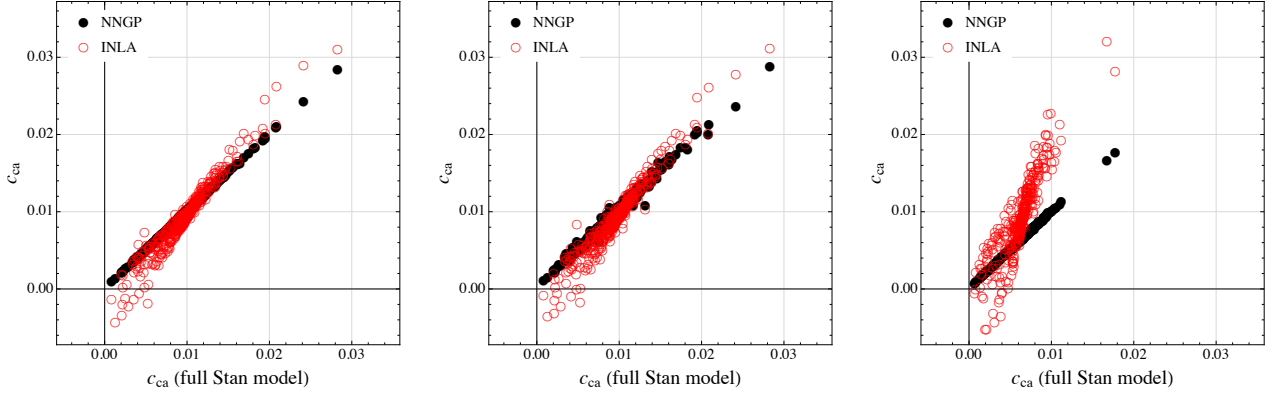
Figure 8: Comparison of mean cell-specific coefficients for different estimation preceduces. (a) ASK14 base model; (b) BSSA14 base model; (c) CB14 base model.
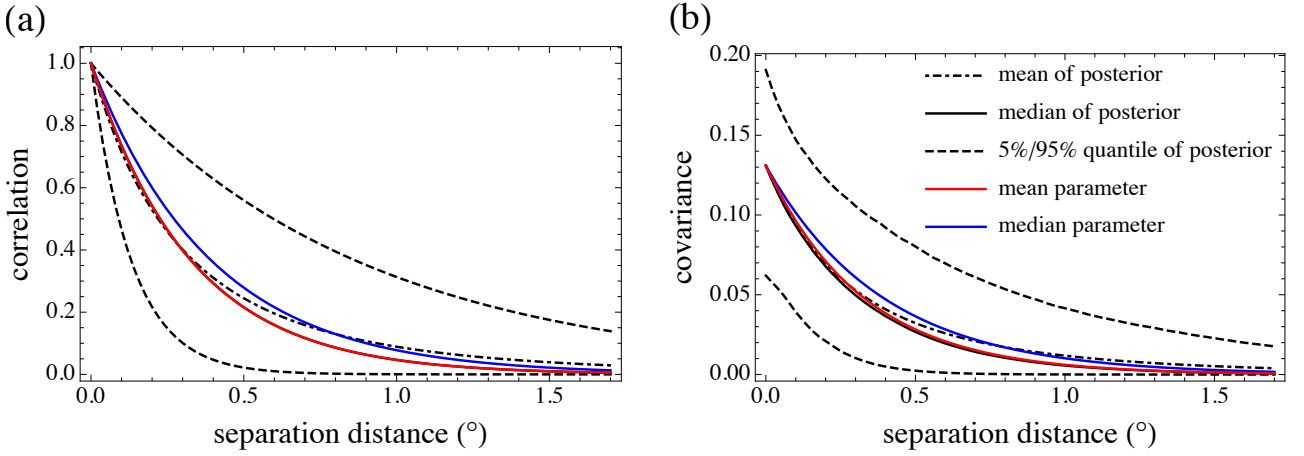

Figure 9: (a) Correlation of spatially varying event terms, depending on separation distance. Shown are the range of correlations basd on the posterior distribution of ℓ_{eq}, as well as correlations based on point estimates of the posterior (mean/median of posterior distribution of ℓ_{eq}). (b) Covariance of spatiall varying event terms, taking into account uncertainty of ψ_{eq}. The ergodic base model is ASK, the nonergodic adjustment model is NNGP.

models, we used a truncated normal distribution (cf. Equation (12)) as the prior distribution for the cell-specific effects, while we use an untruncated normal distribution in INLA. The number of negative cell-specific attenuation coefficients is small, and in a forward prediction they should be set to zero.

The posterior distributions of the hyperparameters ℓ_{eq} and ψ_{eq}, governing the distribution of the spatially varying even terms, are shown in Figure 5. The impact of the uncertainty in the hyperparameters on the correlation/covariance for different separation distances $|\vec{t_e} - \vec{t_e}'|$ is shown in Figure 9, for the ASK14 NNGP model. For each sample of the posterior distribution, we calculate the correlation and covariance for different separation distances according to Equation (8). The 5%,95% quantiles, as well as mean and median, of the resulting correlations/covariances are shown in Figure 9. We also show the correlations/covariances, calculated with the mean and median of the posterior distributions.

The range of possible correlation/covariance values is quite broad, reflecting the uncertainty in the values of ℓ_{eq} and ψ_{eq}. A full nonergodic PSHA would take the uncertainty in the correlation/covariance into account. However, we will show in Section Prediction of Spatial Terms that the impact of his uncertainty on prediction of spatial terms is very small.
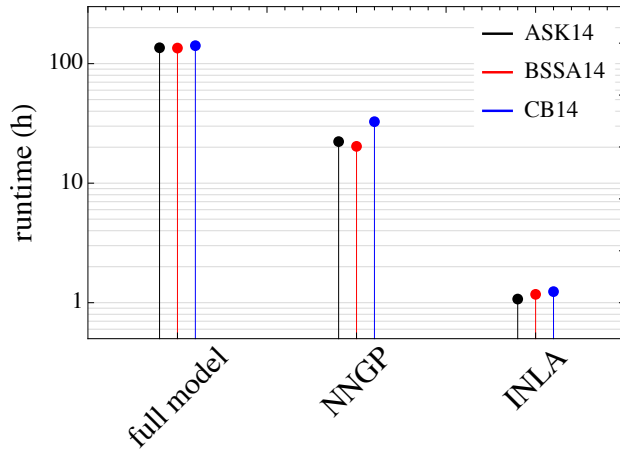
14

Figure 10: Runtime in hours for the different models. For the Stan models, the average time of four chains is plotted.

4.1 Runtime of Different Models

Figure 10 shows the absolute runtime of the different models in hours. We can see that using the NNGP approximation in Stan results in a five-fold speedup compared to the full Stan model. Since the estimated parameters between these two models are almost indistinguishable, this means that the use of the NNGP model can be used to estimate a Bayesian nonergodic VCM for larger data sets and/or more target variables. However, the runtime for the NNGP Stan model is sill quite long with about 20h. By contrast, the INLA model takes only about 1-2h to run, with very similar results to the other models.

One should not overinterpret absolute runtimes, as they depend on the load of the computer. We also ran 1000 warm-up and post warm-up samples for the Stan models; in general, the adaptation phase of HMC is quite fast, so one could potentially reduce the number of the warm-up iterations, which would lead to a smaller runtime.

5 Calculating Predictions

Following Landwehr et al. (2016), we have cast the nonergodic adjustment models as VCMs with GP priors on the adjustment terms. Hence, the predictive distribution of adjustment term δc_* for a new location $\vec{t}_*$ is a normal distribution with predictive mean μ_* and predictive standard deviation ψ_*. we can calculate the predictive distribution by conditioning on observed values (cf. Equations (2.18) and (2.19) in Rasmussen and Williams (2006))

$$p(\delta c_* | \vec{\delta c}) = N(\mu_*, \psi_*) \tag{33}$$

$$\mu_* = \vec{k}^T \boldsymbol{K}^{-1} \vec{\delta c} \tag{34}$$

$$\psi_*^2 = k_* - \vec{k}^T \boldsymbol{K}^{-1} \vec{k} \tag{35}$$

where $\boldsymbol{K}$ is the covariance matrix of the observed systematic effects, and $\vec{k}$ is a vector whose components are the covariance of the new location $\vec{t}_*$ with the observed locations $\vec{t}$. The entries of $\boldsymbol{K}$ and $\vec{k}$ can be calculated from the covariance function with the appropriate coordinates. In our case, δc is either the event or site adjustment, associated with the coordinates of the hypocenter or the site, respectively.

Note that the formulation in Equation (33) is different from the corresponding Equation in Landwehr et al. (2016). They integrated out all the nonergodic coefficients, and thus condi-

tioned on the observed PSA values to calculate the predictive distribution. This means that the matrix $\boldsymbol{K}$ in Equations (34) and (35) needs to include the aleatory variance, since the observations are not noise-free (cf. Equation (2.20) to (2.21) in (Rasmussen and Williams, 2006)). By contrast, in our case we explicitly estimate the adjustment terms $\vec{\delta c}$ as latent variables; they are not integrated out. Since the latent adjustment variables are theoretically considered noise-free, the matrix $\boldsymbol{K}$ does not include aleatory variability terms. The values of $\vec{\delta c}$, however, are estimated from data, and thus are associated with epistemic uncertainty. If this uncertainty is ignored, then the estimated predictive uncertainty will potentially be too small.

In general, one needs to integrate over the posterior distribution of the parameters governing the nonergodic adjustments to account for their uncertainty in the predictive distribution. Thus, the posterior predictive distribution of adjustment terms can be calculated by

$$p\left(\delta c_* | D\right) = \int p\left(\delta c_* | \vec{\delta c}, \psi, \ell\right) p\left(\vec{\delta c}, \psi, \ell | D\right) d\left(\vec{\delta c}, \psi, \ell\right) \tag{36}$$

Given samples from the posterior distribution, the integral can be approximated bu

$$p(\delta c_* | D) \approx \frac{1}{M} \sum_{m=1}^{M} p\left(\delta c_* | \vec{\delta c}^{(m)}, \psi^{(m)}, \ell^{(m)}\right) \tag{37}$$

where superscript (m) denotes the mth sample from the posterior distribution.

Calculating the predictive distribution $p(\delta c_* | D)$ from samples of the posterior distribution requires calculating and inverting the covariance matrix K for each sample of the posterior distribution, which can be computationally expensive. If one considers only uncertainty in the latent variables $\vec{\delta c}$, and assumes that this uncertainty is Gaussian, then one can calculate the predictive distribution as (Lavrentiadis et al., 2021)

$$\mu_* = \vec{k}^T \boldsymbol{K}^{-1} \vec{\mu}_{\delta c} \tag{38}$$

$$\psi_*^2 = k_* - \vec{k}^T \boldsymbol{K}^{-1} \vec{k} + \vec{k}^T \boldsymbol{K}^{-1} \boldsymbol{\Psi}_{\delta c} (\boldsymbol{K}^{-1} \vec{k})^T \tag{39}$$

where $\vec{\mu}_{\delta c}$ is a vector comprising the means of the posterior distributions of the estimated $\vec{\delta c}$ values, and $\boldsymbol{\Psi}_{\delta c}$ is a diagonal matrix whose entries are the variances of the posterior distributions of the estimated $\vec{\delta c}$ values.

For the Stan models, one can calculate the predictive distribution at a new location according to Equations (38) and (39) (using a point estimate of the hyperparameters to calculate the covariances k_*, $\vec{k}$, and $\boldsymbol{K}$), or via Equation (37) if one wants to take the full uncertainty into account. The R-INLA package provides tools to sample from the posterior distribution of the model at the mesh nodes, which can then be interpolated to a new location. The samples from the INLA model at new locations implicitly take the full uncertainty of all locations into account.

5.1   Prediction of Spatial Terms

Figure 11 shows maps of the mean of systematic event terms δc_{eq} over California, for the ASK14 base model. We calculate the values for the mesh nodes shown in Figure 2. For the INLA model, the output contains mean and quantiles of the marginal posterior distribution for the systematic terms at each mesh node. For the Stan models, we calculate the values of the systematic terms at each mesh node based on Equation (38) and (39); for the values of the hyperparameters ψ_{eq} and ℓ_{eq}, we use the median of the respective posterior distributions.

As before, Figure 11 shows relatively small differences between the Stan models and the INLA model. This is concordant with Figure 5, where the posterior distributions of ψ_{eq} and
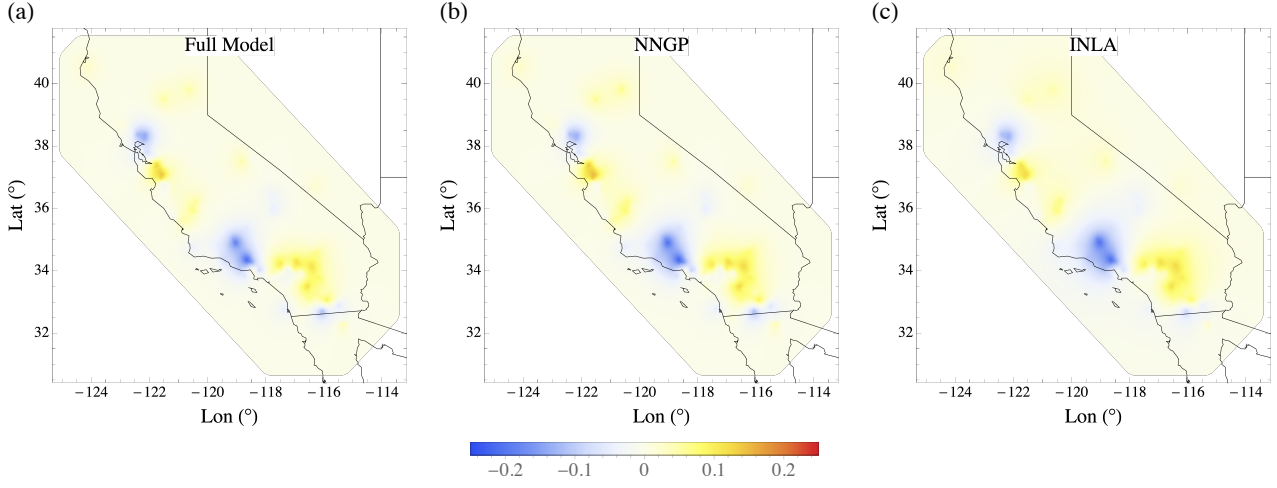
Figure 11: Spatially varying event terms δc_{eq}, for different estimation methods, and ASK14 as the base model.
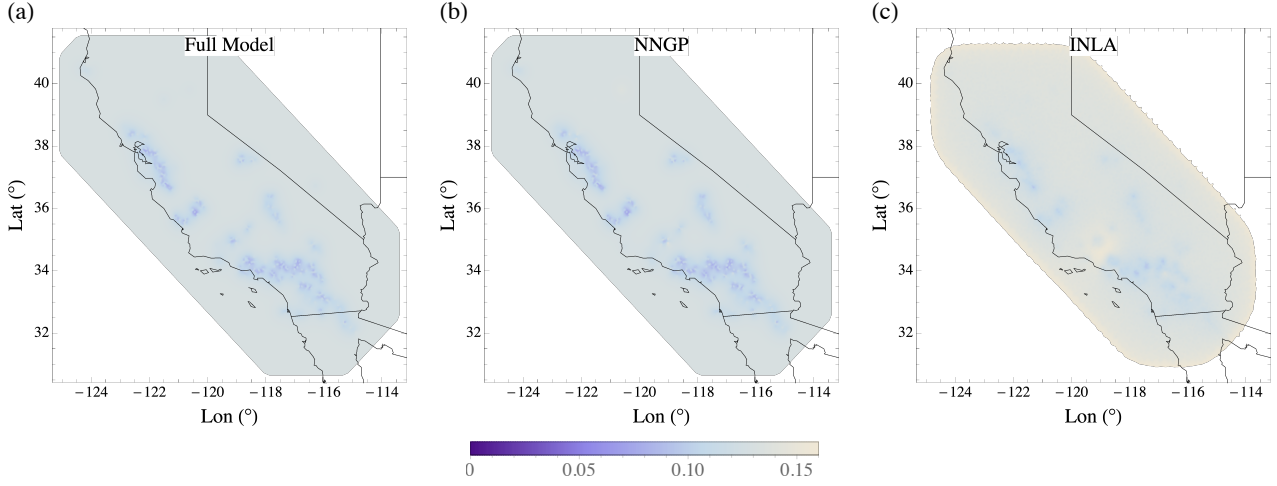

Figure 12: Epistemic uncertainty associated with spatially varying event terms δc_{eq}, for different estimation methods, and ASK14 as the base model.

ℓ_{eq} are similar. We show similar maps for the other ergodic base models, as well as maps for the systematic site terms δc_{stat}, in the electronic supplement.

The corresponding epistemic uncertainties are shown in Figure 12. The maps are quite similar, with INLA predicting slightly higher uncertainty. The uncertainty associated with the spatial event adjustment terms δc_{eq} ranges from 0.075 in the vicinity of observed events to 0.131 in regions where observed events are sparse.

In Figures 13 and 14 we compare the predictive mean μ_* and standard deviation ψ_* of the Stan ASK14 NNGP model, calculated for each mesh node, and based on different methods to take into account the epistemic uncertainty of the latent function values $\vec{\delta c}_{eq}$ and hyperparameters ψ_{eq} and ℓ_{eq}. We compare four different variations:

- Model A: Taking into account the epistemic uncertainty of the latent function values $\vec{\delta c}$, and the hyperparameters ψ_{eq} and ℓ_{eq}, based on posterior samples. The predictive distributions at each mesh node are calculated according to Equation (37).

- Model B: Similar to Model A, but now we only take into account epistemic uncertainty in the late function values $\vec{\delta c}_{eq}$. For the hyperparameters, we use the median of the posterior distribution. Predictions are again calculated according to Equation (37).
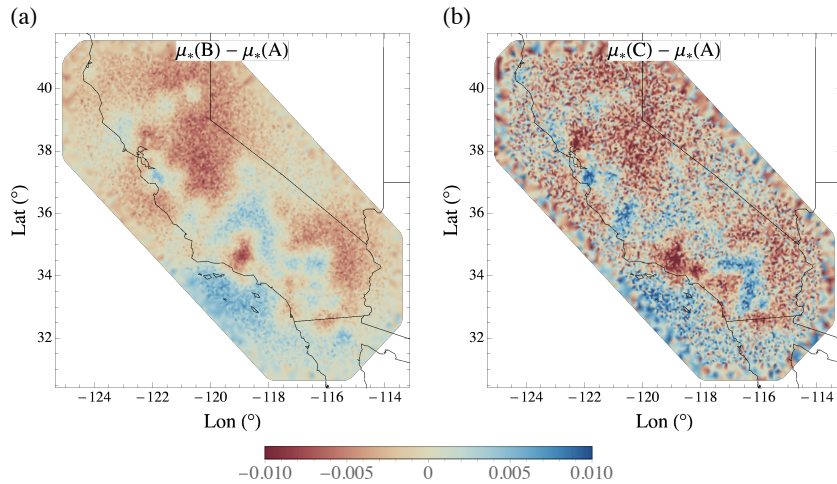
17

Figure 13: Difference in mean predictions of spatially varying event terms, for the ASK14 NNGP model, based on different methods to take into account uncertainty. (a) mean prediction μ_* of model B minus mean prediction μ_* of model A. (b) mean prediction μ_* of model C minus mean prediction μ_* of model A. See text for description of models A, B, C.

- Model C: The predictive distribution is calculated according to Equations (38) and (39). For the hyperparameters, we use the median of the posterior distributions. The matrix $\mathbf{\Psi}_{\delta c}$ is a diagonal matrix whose entries are the variances of the posterior distributions of the estimated $\vec{\delta c}$ values.

- Model D: Similar to Model B, but the $\mathbf{\Psi}_{\delta c}$ is the full covariance matrix of the estimated $\vec{\delta c}$.

We use model A as the reference model, since it includes uncertainties on all parameters, and compare the results from the other three models against the predictions from model A. The predictive means μ_* for model C and D are the same.

The comparisons are shown in Figure 13 for the predictive mean μ_* and Figure 14 for the predictive standard deviation ψ_*, for all mesh nodes. In general, the differences in the predictive means (Figure 13) are quite small, in particular when compared to the overall range of the nonergdic adjustments (Figure 11). The standard deviations for model A and model B are very similar, which implies that the uncertainty in the hyperparameters ψ_{eq} and ℓ_{eq} has only a small influence on the overall epistemic uncertainty of the predictions. There are larger differences in the standard deviations of model C, which shows lower values of ψ_* close to the observed data. Overall, these differences are still small compared to the overall predictive uncertainty. Taking into account the covariance between the estimated latent function values (model D) brings the values of ψ_* closer to the ones of model A, as seen in Figure 14 (c). It is important to note that the computation of the predictive distribution according to Equation (37) (model A and B) requires repeated calculations of $\mu_{*,k}$ and $\psi_{*,k}$, for different samples k of the posterior distribution. By contrast, model C and D require only one calculation step, and are thus more efficient computationally.

6 Discussion

With an increasing number of ground-motion recordings, it is possible to relax the ergodic assumption in GMM modeling, and consider systematic, repeatable source, site, and path effects. This provides challenges, both from a conceptual point of view (how should the systematic effects be modeled), as well as computational. Landwehr et al. (2016) introduced a VCM to
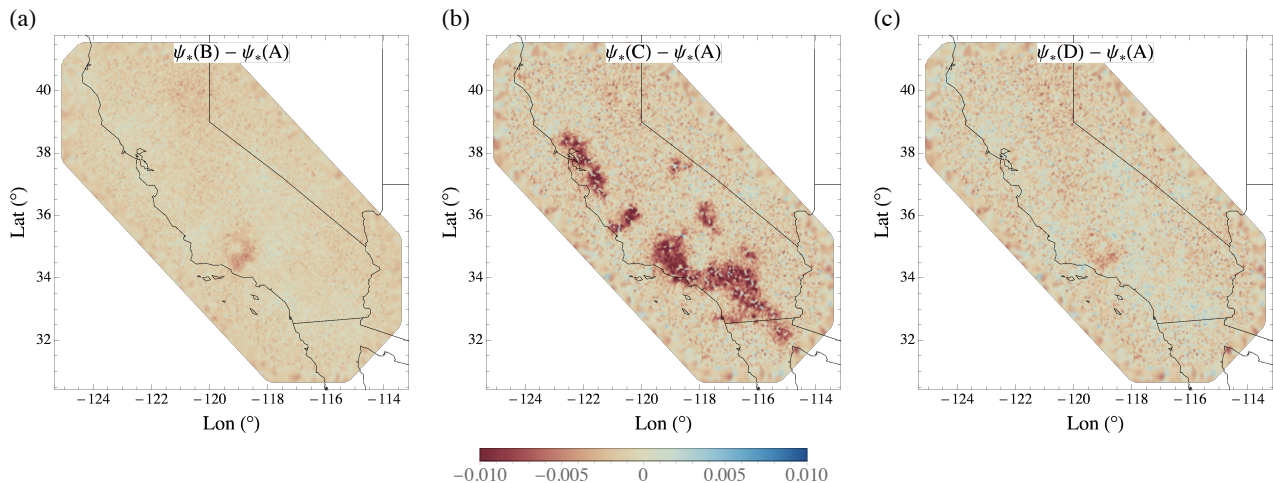
Figure 14: Difference in standard deviations of predictions of spatially varying event terms, for the ASK14 NNGP model, based on different methods to take into account uncertainty. (a) standard deviation of prediction ψ_* of model B minus ψ_* of model A. (b) standard deviation of prediction ψ_* of model C minus ψ_* of model A. (b) standard deviation of prediction ψ_* of model D minus ψ_* of model A. See text for description of models A, B, C, D.

model nonergodic GMMs, based on GPs. While conceptually attractive due to their simplicity, as well as the ease to estimate predictive uncertainties, GPs pose difficulties when applied to large data sets, since they do not scale well to large data sets. Here, we have investigated different methods to fit Bayesian nonergodic GMMs, based on INLA (Rue et al., 2009) and MCMC sampling via Stan Carpenter et al. (2017). For Stan, we fit both a full (without any approximation) as well as an approximated model based o NNGPs (Datta et al., 2016b). Since the development of full nonergodic GMMs is still in its infancy, it is important to investigate different methods to estimate them. Though the INLA and Stan models are based on slightly different assumptions, we find that the results are quite similar between the models.

In terms of runtime of the regression models, INLA provides a clear advantage in our experiments (cf. Figure 10). We also find that the NNGP approximation has a clear speed-up over over the full Bayesian MCMC model without any approximation. Further avenues of investigation can include other approximations, such as combination of varational inference and MCMC (Margossian et al., 2020) or reduced-rank GP regression (Solin and Särkkä, 2020; Riutort-Mayol et al., 2020). Recently, Dambon et al. (2021) proposed a method based regularized maximum likelihood estimation to estimate VCMs with many varying coefficients on large data sets, which can be explored as well. As an alternative to nonergodic GMMs based on GP VCMs, Caramenti et al. (2020) proposed a method based on geographically weighted regression (GWR).

Here, we are mainly concerned with the comparison of different estimation approaches for nonergodic GMMs. Both methods are Bayesian models. We believe that Bayesian models provide advantages, as they allow to incorporate prior information in a probabilistic way, and can be easily updated when new data becomes available (Stafford, 2019). We have cast the nonergodic models as adjustments to ergodic base models, but both INLA and Stan allow to easily combine the fitting of on-spatially varying ad spatially varying coefficients. As shown in Kuehn et al. (2019) (and also seen in Figure 7), the average anelastic attenuation coefficient changes when estimating cell-specific path effects, so taking into account nonergodic effects could have an effect on other parameters as well.

It is easy to extend the nonergodic models, both for INLA and for Stan, to incorporate other spatially varying terms. For example, Landwehr et al. (2016); Caramenti et al. (2020);

Sung et al. (2021) account for a spatially varying V_{S30}-scaling in their models. One could also incorporate different spatial length scales for the same parameter, to account for regional and local variation. In this case, one should use informative prior distributions to avoid problems with identifiability of the different length scales.

In our models, the spatially varying adjustment terms are modeled with a stationary and isotropic correlation function (cf. Equations (8) and (25)). There is evidence that spatial correlations of ground-motion within-event are better modeled with a non-stationary or anisotropic correlation function (Kuehn and Abrahamson, 2020; Abbasnejadfard et al., 2020), and one would expect this to be true for spatially varying coefficients as well. INLA models have been developed to account for spatial non-stationarity (Ingebrigtsen et al., 2014, 2015; Bakka et al., 2019); it should be investigated whether incorporating non-stationarity into the nonergodic models leads to improved performance.

One important aspect when relaxing the ergodic assumption is to take into account epistemic uncertainty. The nonergodic adjustments are by nature estimated from smaller amounts of data, and are thus less well constrained compared to ergodic GMMs. In a VCM, there is parameter uncertainty (cf. Figure 9), as well as uncertainty for predictions at new locations. Figures 13 and 14 indicate that uncertainty in the hyperparameters only has a small influence on the overall predictive uncertainty.

Acknowledgements

References

Abbasnejadfard, M., Bastami, M., and Fallah, A. (2020). "Investigation of anisotropic spatial correlations of intra-event residuals of multiple earthquake intensity measures using latent dimensions method." *Geophysical Journal International*, 222(2), 1449–1469.

Abrahamson, N., Kuehn, N., Walling, M., and 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.

Abrahamson, N. A., Silva, W. J., and Kamai, R. (2014). "Summary of the ASK14 Ground Motion Relation for Active Crustal Regions." *Earthquake Spectra*, 30(3), 1025–1055.

Ancheta, T. D., Darragh, R. B., Stewart, J. P., Seyhan, E., Silva, W. J., Chiou, B. S.-J., Wooddell, K. E., Graves, R. W., Kottke, A. R., Boore, D. M., Kishida, T., and Donahue, J. L. (2014). "NGA-West2 Database." *Earthquake Spectra*, 30(3), 989–1005.

Anderson, J. G. and Brune, J. N. (1999). "Probabilistic Seismic Hazard Analysis without the Ergodic Assumption." *Seismological Research Letters*, 70(1), 19–28.

Anderson, J. G. and Uchiyama, Y. (2011). "A Methodology to Improve Ground-Motion Prediction Equations by Including Path Corrections." *Bulletin of the Seismological Society of America*, 101(4), 1822–1846.

Atkinson, G. M. (2006). "Single-Station Sigma." *Bulletin of the Seismological Society of America*, 96(2), 446–455.

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). "Spatial modeling with R-INLA: A review." *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), e1443.

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. (2019). "Non-stationary Gaussian models with physical barriers." *Spatial Statistics*, 29, 268–288.

Betancourt, M. (2017). "A Conceptual Introduction to Hamiltonian Monte Carlo.

Betancourt, M. and Girolami, M. (2015). "Hamiltonian Monte Carlo for Hierarchical Models." *Current Trends in Bayesian Methodology with Applications*, Chapman and Hall/CRC, 79–101.

Bivand, R. S., Gómez-Rubio, V., and Rue, H. (2015). "Spatial Data Analysis with R - INLA with Some Extensions." *Journal of Statistical Software*, 63(20), 1–31.

Bommer, J. J. and Abrahamson, N. A. (2006). "Why Do Modern Probabilistic Seismic-Hazard Analyses Often Lead to Increased Hazard Estimates?." *Bulletin of the Seismological Society of America*, 96(6), 1967–1977.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M. (2014). "NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes." *Earthquake Spectra*, 30(3), 1057–1085.

Bürkner, P. C., Gabry, J., Kay, M., and Vehtari A. (2020). "posterior: Tools for Working with Posterior Distributions.." *Earthquake Spectra*, R package version 0.1.3, https://mc-stan.org/posterior

Bussas, M., Sawade, C., Kühn, N., Scheffer, T., and Landwehr, N. (2017). "Varying-coefficient models for geospatial transfer learning." *Machine Learning*, 106(9-10), 1419–1440.

Bussas, M., Sawade, C., Scheffer, T., and Landwehr, N. (2015). "Varying-coefficient models with isotropic Gaussian process priors." 0, 1–17.

Campbell, K. W. and Bozorgnia, Y. (2014). "NGA-West2 Ground Motion Model for the Average Horizontal Components of PGA, PGV, and 5% Damped Linear Acceleration Response Spectra." *Earthquake Spectra*, 30(3), 1087–1115.

Caramenti, L., Menafoglio, A., Sgobba, S., and Lanzano, G. (2020). "Multi-Source Geographically Weighted Regression for Regionalized Ground-Motion Models." *Report no.*, https://mox.polimi.it/publication-results/?id=917&tipo=add_qmox.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). "Stan : A Probabilistic Programming Language." *Journal of Statistical Software*, 76(1), 1–32.

Chao, S.-H., Lin, C.-M., Kuo, C.-H., Huang, J.-Y., Wen, K.-L., and Chen, Y.-H. (2020). "Implementing horizontal-to-vertical Fourier spectral ratios and spatial correlation in a ground-motion prediction equation to predict site effects." *Earthquake Spectra*, 875529302095244.

Chiou, B. S.-J. and Youngs, R. R. (2014). "Update of the Chiou and Youngs NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra." *Earthquake Spectra*, 30(3), 1117–1153.

Dambon, J. A., Sigrist, F., and 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.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016a). "Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets." *Journal of the American Statistical Association*, 111(514), 800–812.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016b). "On nearest-neighbor Gaussian process models for massive spatial data." *Wiley Interdisciplinary Reviews: Computational Statistics*, 8(5), 162–171.

Dawood, H. M. and 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.

Finley, A. O., Datta, A., Cook, B. D., Morton, D. C., Andersen, H. E., and Banerjee, S. (2018). "Efficient Algorithms for Bayesian Nearest Neighbor Gaussian Processes." *Journal of Computational and Graphical Statistics*, 111(514), 1–14.

Franco-Villoria, M., Ventrucci, M., and Rue, H. (2018). "Bayesian varying coefficient models using PC priors." 1–21.

Franco-Villoria, M., Ventrucci, M., and Rue, H. (2019). "A unified view on Bayesian varying coefficient models." *Electronic Journal of Statistics*, 13(2), 5334–5359.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2019). "Constructing Priors that Penalize the Complexity of Gaussian Random Fields." *Journal of the American Statistical Association*, 114(525), 445–452.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003). "Spatial Modeling With Spatially Varying Coefficient Processes." *Journal of the American Statistical Association*, 98(462), 387–396.

Gelman, A. (2006). "Prior distributions for variance parameters in hierarchical models." *Bayesian Analysis*, 1(3), 515–534.

Gómez-Rubio, V. (2020). *Bayesian inference with INLA* gitbook, https://becarioprecario.bitbucket.io/inla-gitbook/index.html.

Hjort, N. L., Holmes, C., Müller, P., Walker, S. G., Ghosal, S., Lijoi, A., Prünster, I., Teh, Y. W., Jordan, M. I., Griffin, J., Dunson, D. B., and Quintana, F. (2010). *Bayesian Nonparametrics*. Cambridge University Press.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). "Spatial models with explanatory variables in the dependence structure." *Spatial Statistics*, 8(C), 20–38.

Ingebrigtsen, R., Lindgren, F., Steinsland, I., and Martino, S. (2015). "Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field." *Spatial Statistics*, 14, 338–364.

Kotha, S. R., Weatherill, G., Bindi, D., and Cotton, F. (2020). "A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe." *Bulletin of Earthquake Engineering*, 18(9), 4091–4125.

Kottke, A. (2016). "pygmm: Ground motion models implemented in Python (Version 0.4.1)", *Zenodo* http://doi.org/10.5281/zenodo.53814

Krainski, E. T., Gómez-Rubio, V., Bakka, H., Lenzi, A., Castro-Camilo, D., Simpson, D., Lindgre, F., and Rue, H. (2018). *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA* gitbook, https://becarioprecario.bitbucket.io/spde-gitbook/index.html.

Kuehn, N., Bozorgnia, Y., Campbell, K. W., and Gregor, N. (2020). "Partially Non-Ergodic Ground-Motion Model for Subduction Regions using the NGA-Subduction Database." *Report No. August*, Pacific Earthquake Engineerig Research Center.

Kuehn, N. M. and Abrahamson, N. A. (2020). "Spatial correlations of ground motion for nonergodic seismic hazard analysis." *Earthquake Engineering & Structural Dynamics*, 49(1), 4–23.

Kuehn, N. M., Abrahamson, N. A., and Walling, M. A. (2019). "Incorporating Nonergodic Path Effects into the NGA-West2 Ground-Motion Prediction Equations." *Bulletin of the Seismological Society of America*, 109(2), 575–585.

Kumar, R., Carroll, C., Hartikainen, A., and Martin, O. (2019). "ArviZ a unified library for exploratory analysis of Bayesian models in Python." *Journal of Open Source Software*, 4(33), 1143.

Landwehr, N., Kuehn, N. M., Scheffer, T., and 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.

Lavrentiadis, G., Abrahamson, N.A., and Kuehn, N. M. (2021). "A Non-ergodic Effective Amplitude Ground-Motion Model for California." *submitted to Bulletin of Earthquake Engineering*.

Lindgren, F. and Rue, H. (2015). "Bayesian Spatial Modelling with R - INLA." *Journal of Statistical Software*, 63(19), 1–25.

Lindgren, F., Rue, H., and 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.

Margossian, C. C., Vehtari, A., Simpson, D., and Agrawal, R. (2020). "Hamiltonian Monte Carlo using an adjoint-differentiated Laplace approximation: Bayesian inference for latent Gaussian models and beyond." (Mcmc).

Neal, R. (2011). "MCMC Using Hamiltonian Dynamics." *Handbook of Markov Chain Monte Carlo*, 113–162.

Parker, G. A., Stewart, J. P., Boore, D. M., Atkinson, G. M., and Hassani, B. (2020). "NGA-Subduction Global Ground-Motion Models with Regional Adjustment Factors." *Report No. August*.

Pettit, L. I. (1990). "The Conditional Predictive Ordinate for the Normal Distribution." *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1), 175–184.

R Core Team (2020) "R: A language and environment for statistical computing. R Foundation for Statistical Computing." Vienna, Austria. https://www.R-project.org.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, http://www.gaussianprocess.org/gpml.

Riutort-Mayol, G., Bürkner, P.-C., Andersen, M. R., Solin, A., and Vehtari, A. (2020). "Practical Hilbert space approximate Bayesian Gaussian processes for probabilistic programming." 1–33.

Rodriguez-Marek, A., Rathje, E. M., Bommer, J. J., Scherbaum, F., and Stafford, P. J. (2014). "Application of single-station sigma and site-response characterization in a probabilistic seismic-hazard analysis for a new nuclear site." *Bulletin of the Seismological Society of America*, 104(4), 1601–1619.

Rue, H., Martino, S., and 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.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). "Bayesian Computing with INLA: A Review." *Annual Review of Statistics and Its Application*, 4(1), 395–421.

Sedaghati, F. and Pezeshk, S. (2017). "Partially Nonergodic Empirical Ground-Motion Models for Predicting Horizontal and Vertical PGV, PGA, and 5% Damped Linear Acceleration Response Spectra Using Data from the Iranian Plateau." *Bulletin of the Seismological Society of America*, 107(2), 934–948.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). "Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors." *Statistical Science*, 32(1), 1–28.

Solin, A. and Särkkä, S. (2020). "Hilbert space methods for reduced-rank Gaussian process regression." *Statistics and Computing*, 30(2), 419–446.

Stafford, P. J. (2014). "Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models." *Bulletin of the Seismological Society of America*, 104(2), 702–719.

Stafford, P. J. (2019). "Continuous integration of data into ground-motion models using Bayesian updating." *Journal of Seismology*, 23(1), 39–57.

Sung, C. and Lee, C. (2016). "A New Methodology for Quantification of the Systematic Path Effects on Ground-Motion Variability." *Bulletin of the Seismological Society of America*, 106(6), 2796–2810.

Sung, C., Abrahamson, N.A., Kuehn, N. M., Traversa3, P., and Zentner4, I. (2021). "A non-ergodic ground-motion model of Fourier amplitude spectra for France." *submitted to Bulletin of Earthquake Engineering.*

Villani, M. and Abrahamson, N. A. (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.

Watanabe, S. (2013). "A widely applicable bayesian information criterion." *Journal of Machine Learning Research*, 14(1), 867–897.

Zhang, L., Datta, A., and Banerjee, S. (2018). "Practical Bayesian Modeling and Inference for Massive Spatial Datasets On Modest Computing Environments." 1–20.

Zhang, L. (2018). "Nearest Neighbor Gaussian Processes (NNGP) based models in Stan." Stan case study https://mc-stan.org/users/documentation/case-studies/nngp.html