

Bayesian spatial modelling for high-dimensional seismic inverse problems

Ran Zhang

Center for Mathematical Science, Technische Universität München, Munich, Germany

Claudia Czado

Center for Mathematical Science, Technische Universität München, Munich, Germany

Karin Sigloch

Department of Earth Sciences, University of Oxford, Oxford, United Kingdom

Summary. We study the application of Bayesian spatial modelling to seismic tomography, a geophysical, high-dimensional, linearised inverse problem that infers the three-dimensional structure of the earth's interior. We develop a spatial dependency model of seismic wave velocity variations in the earth's mantle based on a Gaussian Matérn field approximation. Using the theory of stochastic partial differential equations, this model quantifies the uncertainties of the parameter space by means of the integrated nested Laplace approximation. In resolution tests using simulated data and in inversions using real data, our model matches the performance of conventional deterministic optimisation approaches in retrieving three-dimensional structure of the earth's mantle. In addition it delivers estimates of the full parameter covariance matrix. Our model substantially improves on previous work relying on Markov Chain Monte Carlo methods in terms of statistical misfits and computing time.

Keywords: high-dimensional inverse problems, integrated nested Laplace approximation, seismic tomography, spatial modelling on 3-D manifolds, stochastic partial differential equations

1. Introduction

Seismic tomography is a geophysical inverse problem that uses observations recorded at the earth's surface for imaging the three-dimensional (3-D) heterogeneity structure of the subsurface. Seismometers on the earth's surface record ground motion caused by seismic waves that originate from moderate or large earthquakes. These waveforms are sensitive to variations of seismic wave propagation velocities, which in turn are functions of the elastic parameters of rock in the earth's crust and mantle.

To first order, global mantle structure is layered and spherically symmetric because it is caused by the force of gravity acting on density anomalies. The earth's mantle, comprising the upper mantle (at depths of 30-670 km) and the lower mantle (670-2889 km), behaves like solid rock on the time scales of seismic wave propagation (seconds to hours), but acts like a viscous fluid over geologic time scales. The motion of the rock masses, on the order of centimetres per year, leads to lateral variations of material properties of typically a few

Address for correspondence: Ran Zhang, Technische Universität München, Boltzmannstrasse 3, Garching bei München, Germany
E-mail: ran.zhang@ma.tum.de

per cent relative to the spherically symmetric background structure. Such anomalies are caused by downwellings of dense, cool rock from the surface and upwellings of hot, buoyant rock from near the core, as the earth is slowly losing its internal heat. Seismic tomography provides 3-D "snapshots" of these flows that characterise the planet's heat budget and internal convection modes, but are too inaccessible to be observed directly.

Robust estimation of the spherically symmetric component of mantle structure had been accomplished by the 1980's (e.g., PREM, the Preliminary Reference Earth Model of Dziewonski and Anderson (1981), or model IASP91 of Kennett and Engdahl (1991), which differ only in detail). The computationally much more demanding goal of seismic tomography is to estimate lateral deviations from these reference models, especially 3-D variations of seismic waves velocities, which are the most robustly estimated subsurface properties by far. A brief history of the field can be found in Nolet (2008).

Given that seismic wave velocities deviate from their layered reference structure by only a few per cent, it is common practice to formulate seismic tomography as a linear(ised) inverse problem. The earth's interior is parameterised by thousands to millions of wave velocity parameters on a discrete mesh. The data to assimilate are thousands to millions of seismic traveltimes measurements, which indicate whether a seismic wave arrived slightly earlier or later than expected for a spherically symmetric reference earth. The large, sparse and underconstrained matrix that relates traveltimes measurements to the velocity structure of the interior, the so-called sensitivity matrix, is unusually "inverted" in an optimisation problem, using the method of least squares and Tikhonov regularisation (Nolet, 1987; Tian et al., 2009; Sigloch, 2011).

Bayesian approaches to seismic tomography have been formulated mainly as linear or nonlinear problems of low dimensionality with Gaussian errors (Mosegaard and Tarantola, 1995; Bodin and Sambridge, 2009; Stähler and Sigloch, 2014). For high-dimensional linear inverse problems, Zhang et al. (2013) developed an efficient sampling scheme using Gibbs-Metropolis Markov Chain Monte Carlo methods (MCMC) to exploit the sparse structure of the sensitivity matrix. They demonstrated it on a realistic seismic tomography problem featuring about 11,000 unknowns that had previously been solved by optimisation (Sigloch (2011)).

In this paper we present a Bayesian approach to estimate seismic wave velocities considering their spatial structure in 3-D. Parameterisation of the earth's interior is achieved through a highly irregular tetrahedral mesh of thousands of vertices spaced by 60-200 km, identical to the mesh used by Sigloch (2011) and Zhang et al. (2013). Tetrahedra sizes are adapted to the locations of wave sources and receivers so that mantle volumes with denser seismic ray coverage are covered by more densely spaced grids than regions with sparser ray coverage. The resulting large number of velocity parameters represents an approximation to the continuous seismic velocity field. For modelling the 3-D structure of this velocity field, Zhang et al. (2013) defined a neighbourhood within a fixed distance of the velocity parameters, but the number of neighbouring vertices within a fixed distance obviously depends on the geometry of the triangulation. Here we work in a more general setting, i.e., independent of mesh geometry. We develop a model for the velocity field on a continuous 3-D domain by means of the Gaussian field approximation, based on the theory of stochastic partial differential equations (SPDE) introduced by Lindgren et al. (2011). In the SPDE approach, the dense covariance matrix of a Gaussian field (GF) from the Matérn class is approximated by the sparse structure of the Gaussian Markov random field (GMRF) using a finite-dimensional basis function representation based on the finite element method. The sparse precision matrix of the GMRF arising from the SPDE approximation provides

a huge computational advantage when dealing with Bayesian inference, since efficient numerical methods, such as fast matrix factorisation, can be applied (Rue and Held, 2005). We show an application that quantifies 3-D structure of the earth’s upper mantle down to 700 km depth under the western United States by estimating 8977 velocity parameters.

Furthermore, our model allows for spatial correlations in the data errors. Generally, real seismic data observed on the surface are spatially dependent. Unknown spatial errors are for example caused by imperfect knowledge of the location and timing of the wave sources. Here we model spatially correlated data errors of both sources (earthquake) and receivers (seismic station) locations by means of the SPDE approach of Lindgren et al. (2011). The random field over sources is defined on a curved space over the entire surface of the earth, and the receiver field is defined on a curved space that covers the western United States. With prediction maps we can identify locations at which the data may not be well explained by the physical model and may show a systematic error at the sources or receivers. As in the Kriging methodology, we produce maps of optimal predictions and associated prediction standard errors from incomplete and noisy spatial data (Cressie, 1993). General Kriging methods in large-scale data sets can be found in Furrer et al. (2007) and Banerjee et al. (2008).

Along with the SPDE approach, the Bayesian inference in our application is carried out by the integrated nested Laplace approximation (INLA) algorithm developed by Rue et al. (2009). Tailored to the class of latent Gaussian models, INLA exploits deterministic nested Laplace approximations and provides a faster and more accurate alternative to stochastic simulations. It is computationally more efficient than MCMC while yielding accurate approximations to the posterior distributions. Incorporating the powerful properties of the SPDE approach, INLA has become very popular in Bayesian modelling of large-scale spatial data over the past years due to its computational advantage. For example, Schrödle and Held (2011) applied INLA in a spatio-temporal disease mapping problem. Simpson et al. (2011) proposed a new formulation of the log-Gaussian Cox processes using SPDE/INLA and conducted inference for a data set on a globe. Cameletti et al. (2012) considered a hierarchical spatio-temporal model for particulate matter concentration in northern Italy. A detailed description on the theory of the SPDE and INLA approaches, as well as many code examples can be found in Simpson et al. (2012); Lindgren (2012); Illian et al. (2012), or on the webpage of the R-INLA package (www.r-inla.org). So far, the SPDE approach within the INLA framework has been mainly applied to spatial modelling of large data sets on \mathbb{R}^2 or \mathbb{S}^2 manifolds. Here we demonstrate a novel application in a full 3-D space and deploy the R-INLA program for our Bayesian inference of about 13,000 seismic velocity parameters, assimilating over 53,000 global observations. We show that the INLA algorithm achieves a significant computational advantage over the MCMC algorithm.

Hence we improve on several aspects of earlier work on uncertainty quantification in high-dimensional seismic tomography:

- (a) We calculate the precision matrix for the 3-D Gaussian Matérn field explicitly based on the theory of the SPDEs introduced by Lindgren et al. (2011) so that the continuous seismic velocity field in 3-D is modeled using the Gaussian field approximation.
- (b) In estimating the wave velocity parameters of the earth’s interior, we allow for spatial correlated data errors depending on the wave source and receiver locations at its surface.
- (c) We adopt the INLA algorithm to improve the computational efficiency of the Bayesian inference.

Section 2 describes the physical forward model and the properties of the seismological traveltime data. Section 3 discusses some essentials of the SPDE approach for approximating Gaussian random fields and provides the INLA specification to the approximation. The performance of this approach is examined in simulation studies in Section 4 under various scenarios. Section 5 applies the SPDE approach with INLA to traveltime data previously studied by least-squares optimisation (Sigloch, 2011) and in a MCMC approach (Zhang et al., 2013). This allows for performance comparison of the different estimation methods. Section 6 discusses the advantages, limitations, and possible extensions of our new approach.

2. The physical forward model, the data and the statistical spatial model

In this section we introduce the physics and the seismological data, as well as established modelling techniques in seismic tomography. Then we present our statistical model tailored to this type of tomography problem. For a more comprehensive description of seismic tomography for statisticians we refer to Zhang et al. (2013) and Tian et al. (2009).

2.1. Spatial characteristics of the data and the geophysical model

Seismograms are time series recording ground displacement on discrete surface locations, and are thus discrete spatial samples of continuous seismic wave fields that exist everywhere inside the earth and on its surface. Every modest or large earthquake can be recorded by tens to thousands seismic stations around the globe, and every station can record earthquakes occurring anywhere if they are strong enough. In practice, every tomographic study exploits only a subset of the rich information contained in modern waveform data sets, in terms of wave types, sources, and receivers used. Furthermore, a tomographic study also depends on the spatial scope of the study (local, regional, or global), the abundance of exploitable wave types in different locations, the earth properties of interest, and computational modelling capabilities. What seismic tomography problems have in common is that they usually involve high-dimensional data and parameter spaces. Here for example we use teleseismic primary waves or P-waves in order to image several hundred kilometres deep beneath the North American continent. The source and receiver distributions are shown in Figure 1. P-waves are the fastest waves that travel through a solid, and teleseismic means that the earthquakes occurred at more than 3500 km distance from the receivers and thus sampled relatively deep mantle regions beneath the receivers.

Seismic data most often consist of measurements of traveltime anomalies (time delays), rather than amplitudes or entire waveforms, for example. Each traveltime delay datum y_i is given as the difference between the observed and predicted arrival time, i.e. $y_i := y_i^{\text{obs}} - y_i^{\text{pred}}$, where i denotes the wave path (source-receiver combination). The predicted arrival times are obtained by simulating wave propagation through the spherically symmetric reference earth model IASP91. The observed and predicted P-wave traveltimes are typically in the range of 600-1000 seconds depending on the distance between earthquake and seismic station. The traveltime delays y_i are commonly on the order of one second. We use 53,270 observations that were recorded mainly by the USArray seismometer network, which is a subset of the data measured by Sigloch (2008).

Time delays (positive y_i) indicate that the recorded wave traveled through earth structure that on average featured slower wave propagation velocities than the reference model,

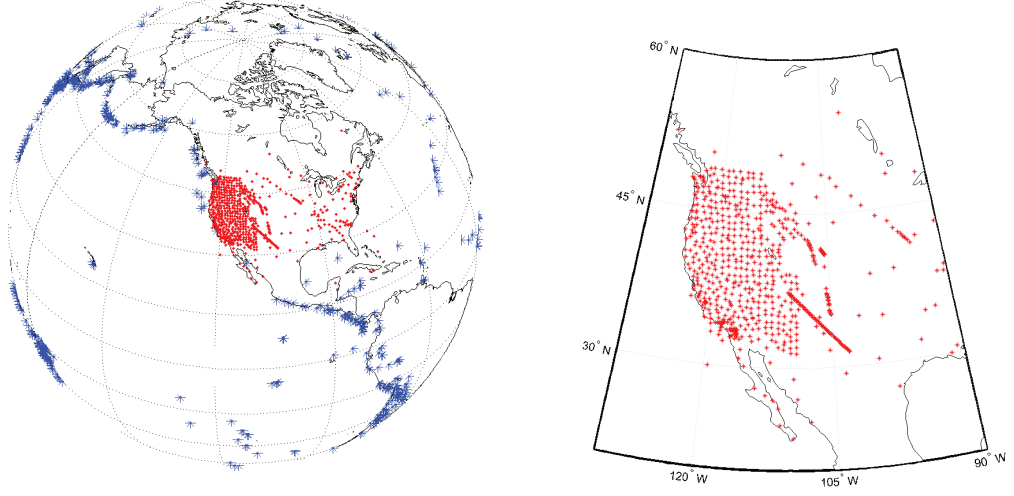


Fig. 1. Left: spatial distribution of the wave sources (large earthquakes, blue) that created our data, and the receivers (seismic broadband stations, red) that recorded them. Right: the target region of tomographic inversion, including a more detailed view of the receiver distribution.

whereas traveltime advances (negative y_i) indicate anomalously fast seismic structure somewhere along the wave path. Seismic velocity variations as a function of three spatial coordinates are the unknowns to be estimated. The measure of misfit is the sum of the squared traveltime anomalies. Following Sigloch (2008) and Zhang et al. (2013), the earth's mantle is parameterised by a mesh of irregular tetrahedra with $p = 8977$ mesh vertices. Each vertex represents the P-wave velocity deviation of the mantle with respect to the IASP91 reference model. We represent the continuous velocity deviation field (or velocity field) by a finite set of velocity deviations $\boldsymbol{\beta} := (\beta_1(\mathbf{r}_1), \dots, \beta_p(\mathbf{r}_p), \mathbf{r}_1, \dots, \mathbf{r}_p \in M_{\text{Earth}}) \in \mathbb{R}^{p=8977}$, where M_{Earth} represents the set of mesh vertices inside the earth. The finite element discretisation on a tetrahedral mesh with basis function $b_j(\mathbf{r})$ at location \mathbf{r} is given by

$$\beta(\mathbf{r}) = \sum_{j=1}^p b_j(\mathbf{r})\beta_j, \quad b_j(\mathbf{r}_i) := \begin{cases} 1 & : i = j \\ 0 & : i \neq j. \end{cases} \quad (1)$$

at the vertices \mathbf{r}_i and linearly interpolated for other locations using the tetrahedron containing \mathbf{r} (Sambridge and Gudmundsson, 1998).

Since the magnitude of $\beta(\mathbf{r})$ is only on the order of few percent, the wave equation can be linearised around the spherical symmetric reference earth model using finite-frequency theory (Dahlen et al., 2000):

$$y_i = \iiint_{\text{Earth}} x_i(\mathbf{r})\beta(\mathbf{r})d^3\mathbf{r}, \quad (2)$$

where $x_i(\mathbf{r}) \in \mathbb{R}^3$ represents the Fréchet sensitivity kernel of the i th wave path. After numerical integration of kernel $x_i(\mathbf{r})$ onto the mesh, (2) can be written as

$$y_i = \sum_{j=1}^p x_{ij}\beta_j.$$

Here, x_{ij} is the Fréchet kernel sensitivity at mesh node j for the i th wave path, i.e., the partial derivative of the delay datum y_i with respect to velocity field β . For N traveltime delay data points we obtain a linear system $\mathbf{y} = X\beta$ with the sensitivity matrix $X \in \mathbb{R}^{N \times p}$ containing N sensitivity kernels. Since the matrix X is sparse and never has full rank, regularisation must be applied in order to solve the linear system. One of the most widely used solvers for large linear sparse system is the Sparse Equations and Least Squares (LSQR) algorithm by Paige and Saunders (1982), which minimises $\|\mathbf{y} - X\beta\|^2 + \lambda\|\beta\|^2$, where λ is a positive regularisation parameter. The effect is to remove non-uniqueness from the system, essentially by adding a multiple of the identity matrix onto the normal equations (Nolet, 1987; Tian et al., 2009).

2.2. Setup of the statistical spatial model

Our goal is to estimate volumetric anomalies in seismic wave velocities, i.e., the shapes and magnitudes of regions in the earth’s interior where seismic waves travel faster or slower than predicted by a simple, layered earth reference model. The resulting, 3-D earth model represents the primary output of seismic tomography. In practice things are complicated by the fact that traveltime (or other) measurements are also influenced by near-source and/or near-receiver effects not accounted for by 3-D mantle structure. Earthquake characteristics are unknown a priori. We extract information about their timing and locations from routinely assembled catalogues, but such estimates are usually not accurate enough for our purposes. Earthquake mislocations cause traveltime anomalies, which if unchecked would propagate into the solution for 3-D mantle structure. Hence tomographers commonly introduce “source corrections”, typically four free parameters per earthquake (latitude, longitude, depth, and time), which are free to adjust optimally during inversion. We take a different approach and introduce a 2-D random field of traveltime anomalies, the “source field”, to estimate traveltime anomalies due to source mislocations. We introduce a second 2-D random field, the “receiver field” to correct for traveltime anomalies near each receiver, which are mostly caused by poorly known crustal structure, although malfunctioning of the station (e.g., clock errors) might also contribute. For simplicity we ignore the depth of the sources and the altitudes of the receivers and assume they are located at the surface. In contrast to conventional corrections, which are associated with discrete source and receiver locations, our source and receiver fields are continuous over (parts of) the earth’s surface. We proceed to formulate the model for the joint Bayesian estimation of velocity, source and receiver fields.

An n -sphere of radius r centred at the origin, denoted by \mathbb{S}^n , is embedded in the Euclidean space \mathbb{R}^{n+1} and defined by $\mathbb{S}^n := \{\mathbf{x} \in \mathbb{R}^{n+1} : \|\mathbf{x}\| = r\}$. Our target region of estimation is the mantle beneath North America, a sub-mesh of the global 3-D domain (\mathbb{R}^3), which contains 8977 free velocity parameters, denoted as β , located between latitudes 20° - 60° N, longitudes 90° - 130° W, and 0-800 km depth. Tetrahedral nodes of the sub-mesh are spaced by 60-150 km. From here on, β is also called the “velocity field”. To model the spatial dependency of errors in seismic traveltime data, we consider the following fields. For the data from the sources (the source field), we have a mesh covering the entire earth on a curved space in \mathbb{S}^2 . For the data from the receivers (the receiver field), the mesh is generated over the North America region again in \mathbb{S}^2 space with vertices close to the locations of the sources and receivers as shown in Figure 2.

Let y_i denote the i th traveltime delay datum for $i = 1, \dots, N$. Further let $r(i)$ be the

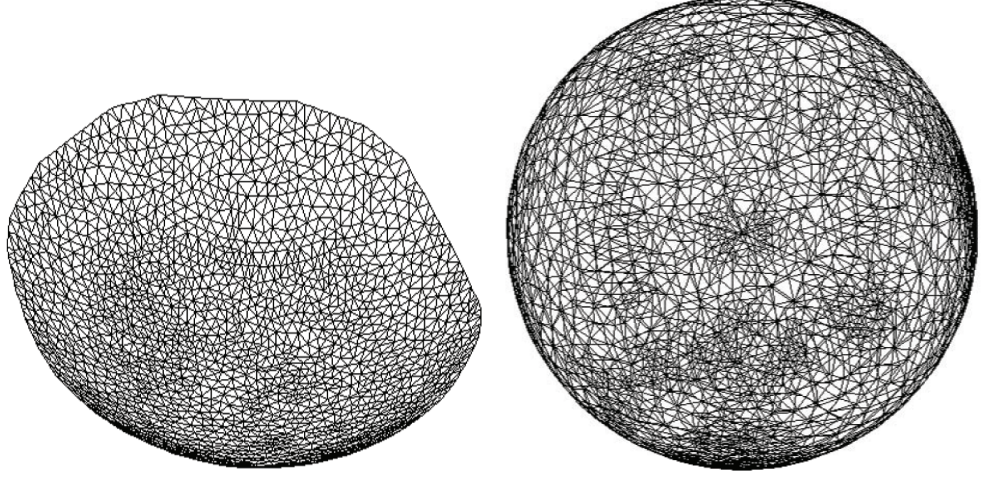


Fig. 2. Triangular meshes on a \mathbb{S}^2 space for estimating the receiver field (left) and the source field (right). Mesh vertices are located close to source and receiver locations. Left: receivers were deployed mainly in western North America, as reflected by the shape of the densified (darker) region of nodes. Right: Earthquakes (sources) occur globally but cluster in certain (plate boundary) locations.

receiver that recorded the i th observation y_i , $r = 1, \dots, R$. Similarly, let $s(i)$ be the source that caused the i th observation, $s = 1, \dots, S$. Let p_s and p_r denote the numbers of mesh vertices of the source and receiver fields on \mathbb{S}^2 , respectively. Define the mapping matrices between data and mesh vertices as $B_s := Z^s A^s \in \mathbb{R}^{N \times p_s}$ and $B_r := Z^r A^r \in \mathbb{R}^{N \times p_r}$. Here, the matrices $A^s \in \mathbb{R}^{S \times p_s}$ and $A^r \in \mathbb{R}^{R \times p_r}$ map from the vertex locations to the positions of the sources and receivers, respectively. The matrices $Z^s \in \mathbb{R}^{N \times S}$ and $Z^r \in \mathbb{R}^{N \times R}$ are incident matrices defined as

$$Z_{is}^s := \begin{cases} 1 & : s = s(i) \\ 0 & : \text{otherwise.} \end{cases} \quad \text{and} \quad Z_{ir}^r := \begin{cases} 1 & : r = r(i) \\ 0 & : \text{otherwise.} \end{cases} \quad (3)$$

Note that our data are ordered by stations. For illustration consider Figure 3 as a toy example, where seven data \mathbf{y} arise from (source, receiver)-combinations (1,1), (1,3), (2,1), (2,2), (3,2), (4,1) and (4,3). The corresponding Z^s and Z^r for the seven observations are therefore

$$Z^s = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Z^r = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Let $\boldsymbol{\gamma} \in \mathbb{R}^{p_s}$ and $\boldsymbol{\xi} \in \mathbb{R}^{p_r}$ be the Gaussian processes on those mesh vertices including the sources and receivers, where p_s and p_r are the number of vertices on the meshes for the sources and receivers. Let p be number of parameters of the velocity field $\boldsymbol{\beta}$. The full statistical model (VSR) defined on *mesh vertices* of the velocity field, receivers and sources

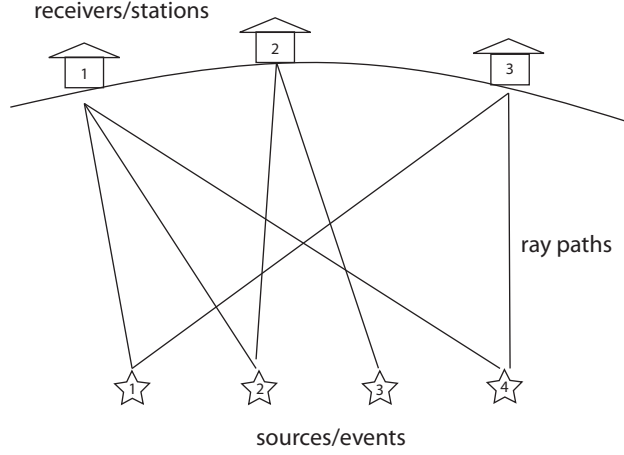


Fig. 3. An earthquake source produces as many traveltimes observations as there are receivers successfully recording the event. Over time, every receiver records earthquakes from a variety of earthquake locations. As demonstrated three receivers record earthquakes from four different earthquake sources.

can be written as

$$Y_i = \sum_{j=1}^p x_{ij} \beta_j + \sum_{k=1}^{p_s} B_{s,ik} \gamma_k + \sum_{l=1}^{p_r} B_{r,il} \xi_l + \epsilon_i$$

or in matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}_s\boldsymbol{\gamma} + \mathbf{B}_r\boldsymbol{\xi} + \boldsymbol{\epsilon}, \quad (4)$$

where $\boldsymbol{\epsilon}$ denotes the vector of data errors. “VSR” stands for the combination of velocity, source and receiver fields. Since the parameters are linear we can gather all parameters and covariates together as

$$\mathbf{Y} = \mathbf{H}\mathbf{m} + \boldsymbol{\epsilon}, \quad (5)$$

with $\mathbf{H} := [\mathbf{X}, \mathbf{B}_s, \mathbf{B}_r] \in \mathbb{R}^{N \times D}$ and the random vector $\mathbf{m} := [\boldsymbol{\beta}^t, \boldsymbol{\gamma}^t, \boldsymbol{\xi}^t]^t \in \mathbb{R}^D$ with $D = p + p_s + p_r$.

Source field $\boldsymbol{\gamma}$ and receiver field $\boldsymbol{\xi}$ may either be used to estimate the traveltime effects of mislocated earthquakes and uncertain near-receiver structure from scratch, or to assess the quality of source and receiver corrections from earlier, deterministic solutions (including their spatial correlations). We take the first approach in our synthetic study of Section 4, which tests the proper functioning of our method, and the latter approach in our study of real data (Section 5). There we evaluate and improve on the source and receiver corrections obtained by Sigloch (2008), allowing for direct comparison to the Bayesian study by Zhang et al. (2013), who also used the conventional corrections.

To summarise, we have constructed a statistical model that takes into account and estimates the seismic velocity field $\boldsymbol{\beta}$ inside the earth, jointly with traveltime corrections at both the earthquake sources (source field $\boldsymbol{\gamma}$) and the seismic stations (receiver field $\boldsymbol{\xi}$). In the following Section we present the estimation methods developed for this statistical model.

3. Estimation methods

3.1. An approximation to Gaussian fields with continuous Markovian structures on manifolds

We adopt the stochastic partial differential equation formulation of Lindgren et al. (2011) for the approximation of the Gaussian random field in order to specify our structural parameters β as well as the Gaussian process parameters γ and ξ on tetrahedral meshes in the \mathbb{R}^3 and \mathbb{S}^2 spaces, respectively. Suppose U is a subset of \mathbb{R}^d . For $d = 2$, $\mathbf{u} := (u_x, u_y) \in U$ is characterised by fixed geographic coordinates, and if $d = 3$, we have $\mathbf{u} := (u_x, u_y, u_z) \in U$ with depth u_z . Let $\{Z(\mathbf{u}) : \mathbf{u} \in U \text{ and } U \subseteq \mathbb{R}^d\}$ be a stochastic process or a spatial process for $d > 1$ (Banerjee et al., 2003). In practice, the observations will be a partial realisation of that spatial process measured at a finite set of locations $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$. The process is said to be Gaussian if, for any $p \geq 1$ and any set of sites $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$, $\mathbf{Z} := (Z(\mathbf{u}_1), \dots, Z(\mathbf{u}_p))^t$ has a multivariate normal distribution (Banerjee et al., 2003). A spatial process \mathbf{Z} is *weakly stationary* if the process has a constant mean, and $\text{Cov}(Z(\mathbf{0}), Z(\mathbf{u})) = c(\mathbf{u})$ for all $\mathbf{u} \in U$ (Cressie, 1993). If the covariance function $c(\cdot)$ only depends on the length $\|\mathbf{u}\|$ of vector \mathbf{u} , then we say that the process is *isotropic*. Here our spatial processes are set to be weakly stationary and isotropic with mean zero. We refer to the Gaussian distributed spatial process as the Gaussian field (GF) and consider a particular class of GFs, called Matérn random fields (MF). (We will later consider only a special case of the MF.) Traditionally, Matérn random fields are zero-mean, Gaussian, stationary, isotropic random fields. The covariance function at location $\mathbf{u} \in \mathbb{R}^d$ is given by

$$\text{Cov}(Z(\mathbf{0}), Z(\mathbf{u})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\mathbf{u}\|)^\nu K_\nu(\kappa \|\mathbf{u}\|) \quad (6)$$

(Cressie, 1993; Stein, 1999). The function $K_\nu(\cdot)$ is the modified Bessel function of the second kind with smoothing parameter $\nu > 0$. $\Gamma(\cdot)$ denotes the Gamma function and ν controls the smoothness of the random field. For all \mathbf{u} the marginal variance $\sigma^2 := \text{Var}[Z(\mathbf{u})] > 0$ is defined as

$$\sigma^2 := \frac{\Gamma(\nu)}{\tau^2 \Gamma(\nu + d/2) (4\pi)^{\frac{d}{2}} \kappa^{2\nu}}. \quad (7)$$

Here, τ is with a scaling parameter proportional to $1/\sigma^2$ if other parameters are fixed. d is the dimensionality of the random field which is either $d = 2$ for source/receiver fields or $d = 3$ for the velocity field. Parameter κ controls the degree of spatial dependency of the random field. For more intuitive interpretability, κ may also be specified through a function that depends on the range parameter ρ , i.e., $\kappa := \frac{\sqrt{8\nu}}{\rho}$, which was derived empirically by Lindgren et al. (2011). The significance of parameter ρ is that for all ν the spatial correlation is about 0.1 at a distance ρ (Lindgren et al., 2011). Given a fixed smoothing parameter ν , the larger the value of ρ , the stronger the spatial correlation. Here, we will fix ν at $\nu + d/2 = 2$ depending on the dimensionality d of the random field. For $d = 2$, the smoothness is fixed at $\nu = 1$ and the variance is $\sigma^2 := 1/(4\pi\kappa^2\tau^2)$, whereas for $d = 3$, we have $\nu = 1/2$ and $\sigma^2 = 1/(8\pi\kappa\tau^2)$. The latter case yields the exponential covariance function, which is given by $\text{Cov}(Z(\mathbf{0}), Z(\mathbf{u})) = \sigma^2 \exp(-\kappa\|\mathbf{u}\|)$.

Gaussian fields (GFs) are the most important model class in spatial statistics with well studied, interpretable statistical properties. However, if the dimension of the random vector $\mathbf{Z} := (Z(\mathbf{u}_1), \dots, Z(\mathbf{u}_p))^t$ is high, its covariance function yields a very large, dense

covariance matrix, which is computationally not feasible for inference. Instead Gaussian Markov random fields (GMRFs) are used as computationally efficient models (Rue and Held, 2005; Simpson et al., 2012). A GMRF is a GF which satisfies the conditional independence assumptions, also called the Markovian property. The Markovian property is related to the definition of a neighbourhood structure. This assumption says that the conditional distribution of a random variable, say Z_i , given all other variables Z_j , $j \neq i$, only depends on the neighbours. That is, $P(Z_i | \mathbf{Z}_{-i}) = P(Z_i | \mathbf{Z}_{\delta_i})$ with $\mathbf{Z}_{-i} := (Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_p)^t$ or in general $\mathbf{Z}_{-S} := Z_i, i \notin S$. The notation \mathbf{Z}_{δ_i} denotes the set of random variables in the neighbourhood of Z_i . Following the notations of Rue and Held (2005) and Cameletti et al. (2012), the conditional independence condition can be written as $Z_i \perp\!\!\!\perp \mathbf{Z}_{-i, \delta_i} \mid \mathbf{Z}_{\delta_i}$ for $i = 1, \dots, p$. This condition is closely related to the precision matrix Q , in particular, $Z_i \perp\!\!\!\perp Z_j \mid \mathbf{Z}_{-i, j} \Leftrightarrow Q_{ij} = 0$. The nonzero elements of Q are given by the neighbourhood structure of the random field. Therefore, the precision matrix Q of a GMRF is sparse, enabling fast computations.

In classical spatial statistics, GMRF models are mainly used for modelling areal data. In order to benefit from the fast computational property of the GMRFs while still modelling the continuously indexed GFs, Lindgren et al. (2011) developed the link between GFs and GMRFs through a type of stochastic partial differential equations (SPDE). We briefly review their results. Let $(Z_1, \dots, Z_p) := (Z(\mathbf{u}_1), \dots, Z(\mathbf{u}_p))^t$ be multivariate Gaussian at random locations $\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$. Lindgren et al. (2011) proposed to use a mesh for parameterising the GF (the same mesh as adopted for the physical model). A finite-dimensional basis function expansion for a continuous specification of the GF \mathbf{Z} is given by

$$Z(\mathbf{u}) \approx \sum_{j=1}^p \varphi_j(\mathbf{u}) Z_j, \quad \text{with } \varphi_j(\mathbf{u}_i) := \begin{cases} 1 & : i = j \\ 0 & : i \neq j \end{cases}. \quad (8)$$

The $\varphi_j(\mathbf{u}_i)$ are deterministic, piecewise linear basis functions. Linear interpolation is applied to the basis function $\varphi_j(\mathbf{u})$ at points $\mathbf{u} \neq \mathbf{u}_i$. Lindgren et al. (2011) showed that a subset of Matérn fields gives a piecewise linear representation of the GF $Z(\mathbf{u})$. Define $\alpha := \nu + d/2$. When α in (9) is an integer, the Matérn field has the piecewise linear representation as in (8). In this case, the piecewise linear representation of $Z(\mathbf{u})$ in (8) is the stationary solution to the stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} (\tau Z(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \alpha := \nu + d/2 \quad (9)$$

where Δ is the Laplacian defined as $\Delta := \sum_{i=1}^d \frac{\partial^2}{\partial u_i^2}$. Here, $\mathcal{W}(\mathbf{u})$ denotes a Gaussian white noise process. Whittle (1954, 1963) showed that Matérn fields $Z(\mathbf{u})$ on \mathbb{R}^d , with the covariance in (6), are the only stationary solutions to the SPDE in (9). Lindgren et al. (2011) further showed that for $\alpha = 2$ the random vector $\mathbf{Z} := (Z_1, \dots, Z_p)^t$ in (8) is a Gaussian Markov random field (GMRF) with mean zero, and its precision matrix Q is given by

$$Q := \tau^2(\kappa^4 \tilde{C} + 2\kappa^2 G + G \tilde{C}^{-1} G), \quad (10)$$

with elements of \tilde{C} and G given by

$$\tilde{C}_{ii} := \int_U \varphi_i(\mathbf{u}) d\mathbf{u} \quad \text{and} \quad G_{ij} := \int_U \nabla \varphi_i(\mathbf{u})^t \nabla \varphi_j(\mathbf{u}) d\mathbf{u},$$

respectively. \tilde{C} is a diagonal matrix. We integrate over the observation domain U . For the velocity field, U is the earth's interior. For the source field, U is the earth's surface, and

for the receiver field, U is the curved space over the western U.S. (Figure 2). In subsequent sections we only consider the model with $\alpha = 2$ due to computational simplicity and software limitations. As mentioned in the previous section, for source/receiver fields ($d = 2$) we have $\nu = 1$, and for the velocity field ($d = 3$), we have $\nu = 1/2$ which yields an exponential covariance function. For theoretical results for $\alpha > 2$ and $d = 2$ we refer to Lindgren et al. (2011).

The technique for calculating Q is based on the finite element method (FEM). For the theory of the FEM we refer to Hughes (1987) and Gockenbach (2006). In the context of our model the Z_j 's in (8) correspond to the velocity field given by the β_j 's in (1). For modelling β we derive the explicit form of the precision matrix Q on the \mathbb{R}^3 manifold. The calculation technique of the 3-D field is similar to the method used in the 2-D case as described in Lindgren et al. (2011). The difference is that each element of a 3-D mesh is a tetrahedron instead of a triangle in a 2-D mesh and the vertices of the tetrahedra are in \mathbb{R}^3 . In this case, the volume and surface of each tetrahedron are needed for the calculation of the precision matrix. Explicit expressions for the \tilde{C} and G matrices for one building block of Q using tetrahedral meshes are provided in the Appendix.

3.2. Model specification using the Integrated Nested Laplace Approximations (INLA)

Bayesian inference of the model (4) is facilitated by the MCMC-free method INLA. The INLA method performs deterministic inference for latent Gaussian models, where the latent field is Gaussian, parameterised by a few hyper-parameters, and the response can be Gaussian or non-Gaussian observations. For Gaussian observations, the accuracy of such deterministic inference is limited only by the error in the numerical integration, whereas for non-Gaussian observations, the approximation can be accurately carried out by a set of Laplace approximations (Rue and Martino, 2007; Rue et al., 2009; Simpson et al., 2012). All calculations in our study are performed in R with the INLA software packages (www.r-inla.org).

Recall the linear model in (5), $\mathbf{Y} = H\mathbf{m} + \epsilon$ with $\mathbf{m} := (\beta^t, \gamma^t, \xi^t)^t =: (m_1, \dots, m_D)^t$. Assuming stationary Gaussian fields for the parameters \mathbf{m} , we denote $\boldsymbol{\theta} := (\tau_\beta, \kappa_\beta, \tau_\gamma, \kappa_\gamma, \tau_\xi, \kappa_\xi, \phi)$ to be a set of hyper-parameters where ϕ denotes the precision parameter of the data errors. Then, the posterior distribution is specified by

$$\pi(\mathbf{m}, \boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{m} | \boldsymbol{\theta}) \pi(\mathbf{y} | \mathbf{m}, \boldsymbol{\theta}).$$

Assuming the hierarchical model,

$$\begin{aligned} \mathbf{Y} | \mathbf{m}, \boldsymbol{\theta} &\sim \mathcal{N}_N(H\mathbf{m}, \frac{1}{\phi} I_N), \\ \mathbf{m} | \boldsymbol{\theta} &\sim \mathcal{N}_D(\mathbf{m}_0, Q(\boldsymbol{\theta})^{-1}), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}). \end{aligned} \tag{11}$$

The prior precision matrix $Q(\boldsymbol{\theta})$ is given by

$$Q(\boldsymbol{\theta}) := \begin{pmatrix} Q_\beta(\kappa_\beta, \tau_\beta) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & Q_\gamma(\kappa_\gamma, \tau_\gamma) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & Q_\xi(\kappa_\xi, \tau_\xi) \end{pmatrix}. \tag{12}$$

$Q_\beta(\kappa_\beta, \tau_\beta)$ is the precision matrix for β as defined in (10) with the corresponding \tilde{C} and G for dimensionality $d = 3$, whereas $Q_\gamma(\kappa_\gamma, \tau_\gamma)$ and $Q_\xi(\kappa_\xi, \tau_\xi)$ are the precision matrices for

γ and ξ for $d = 2$, respectively. The explicit expression of $Q_\beta(\kappa_\beta, \tau_\beta)$ can be found in the Appendix, and for the expression of $Q_\gamma(\kappa_\gamma, \tau_\gamma)$ and $Q_\xi(\kappa_\xi, \tau_\xi)$ on 2-D we refer to Lindgren et al. (2011). The prior mean is denoted as $\mathbf{m}_0 := (\beta_0^t, \mathbf{0}^t, \mathbf{0}^t)^t$. For β_0 we use the LSQR solution β^{LSQR} obtained by minimising the objective function $\|\mathbf{y} - X\beta\|^2 + \lambda\|\beta\|^2$ using the LSQR algorithm (Sigloch, 2008). The full conditional of \mathbf{m} can be expressed as

$$\mathbf{m} | \mathbf{y}, \boldsymbol{\theta} \sim \mathcal{N}_D(\Omega^{-1}\boldsymbol{\zeta}, \Omega^{-1}), \quad \text{with } \Omega := Q(\boldsymbol{\theta}) + \phi H^t H, \quad \boldsymbol{\zeta} := Q(\boldsymbol{\theta})\mathbf{m}_0 + \phi H^t \mathbf{y}. \quad (13)$$

Traditionally, the marginal posterior distributions $\pi(\mathbf{m}|\mathbf{y})$ and $\pi(\boldsymbol{\theta}|\mathbf{y})$ are obtained by sampling, using MCMC methods. For example, (13) can be sampled using the efficient Cholesky decomposition method if the posterior precision matrix Ω is sparse. The sparsity of Ω is guaranteed by the sparsity of matrix H (Rue and Held, 2005; Zhang et al., 2013). This is important in the INLA context. Since the sparsity of H ensures the Markovian property of the latent field \mathbf{m} , fast approximation using the INLA scheme is valid and possible for this model (Simpson et al., 2012).

Following the description in Rue et al. (2009) and Simpson et al. (2012), the main goal of INLA is to compute the marginal posterior distribution $\pi(m_i|\mathbf{y})$ of the latent Gaussian field \mathbf{m} . In particular,

$$\pi(m_i|\mathbf{y}) = \int \pi(m_i|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta}, \quad i = 1, \dots, D. \quad (14)$$

In the INLA software three approximations, denoted as $\tilde{\pi}(m_i|\boldsymbol{\theta}, \mathbf{y})$, are available for approximating the true posterior distributions $\pi(m_i|\boldsymbol{\theta}, \mathbf{y})$: the Gaussian, the Laplace, and a simplified Laplace approximation (Rue et al., 2009). Further, the posterior distribution of the hyper-priors $\pi(\boldsymbol{\theta}|\mathbf{y})$ is approximated again by the Laplace approximation given by (Tierney and Kadane, 1986),

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{m}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{m}, \boldsymbol{\theta})}{\pi(\mathbf{m}|\boldsymbol{\theta}, \mathbf{y})} \approx \frac{\pi(\boldsymbol{\theta})\pi(\mathbf{m}|\boldsymbol{\theta})\pi(\mathbf{y}|\mathbf{m}, \boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{m}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{m}=\mathbf{m}^*(\boldsymbol{\theta})},$$

where $\mathbf{m}^*(\boldsymbol{\theta})$ is typically the conditional mode $E[\mathbf{m}|\boldsymbol{\theta}, \mathbf{y}]$, and $\tilde{\pi}_G(\mathbf{m}|\boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to $\pi(\mathbf{m}|\boldsymbol{\theta}, \mathbf{y})$. Using the integration points $\boldsymbol{\theta}_k$, $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is calculated numerically. The approximation of the marginal posterior in (14), denoted as $\tilde{\pi}(m_i|\mathbf{y})$, is calculated using numerical integration, i.e.,

$$\tilde{\pi}(m_i|\mathbf{y}) = \int \tilde{\pi}(m_i|\boldsymbol{\theta}, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \approx \sum_{k=1}^K \tilde{\pi}(m_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k,$$

where Δ_k are integration weights and $\boldsymbol{\theta}_k$ are the integration points.

To wrap up, we develop the Gaussian Matérn field approximation for a \mathbb{R}^3 manifold by means of the SPDE approach using finite-dimensional linear basis functions. Thus, spatial dependency of the velocity field as well as the source and receiver fields are specified using the SPDE representations. The INLA approach is then applied to the GMRF to facilitate Bayesian inference in our high-dimensional parameter space.

4. Simulation study

4.1. Simulation setup

In this section we investigate the performance of our approach. We assume that the measurement errors ϵ are independently identically normal distributed with a variance $1/\phi = 1$.

We choose a synthetic variance of 1 since it is similar to the variance we expect from the real data. We consider the full model in (4) and three reduced models as follows:

$$\begin{aligned} \text{VSR} : \mathbf{y} &= X\boldsymbol{\beta} + B_s\boldsymbol{\gamma} + B_r\boldsymbol{\xi} + \boldsymbol{\epsilon} \\ \text{VS} : \mathbf{y} &= X\boldsymbol{\beta} + B_s\boldsymbol{\gamma} + \boldsymbol{\epsilon} \\ \text{VR} : \mathbf{y} &= X\boldsymbol{\beta} + B_r\boldsymbol{\xi} + \boldsymbol{\epsilon} \\ \text{V} : \mathbf{y} &= X\boldsymbol{\beta} + \boldsymbol{\epsilon} \end{aligned}$$

The synthetic traveltime data \mathbf{y} are generated using the full model (VSR: velocity, source and receiver fields) and in turn estimated using the VSR, VS (velocity and source fields) and VR (velocity and receiver fields) models, since we want to examine the performance of these models. We simulate the synthetic random fields $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ using (11) for two spatial correlations. The true hyper parameters are specified as follows.

- *Simulated velocity field $\boldsymbol{\beta}$ in \mathbb{R}^3 :*
 For dimensionality $d = 3$ and variance of the random field $\sigma_\beta^2 = 1$, we have $\kappa_\beta := 2/\rho_\beta$ and $\tau_\beta := 1/\sqrt{8\pi\kappa_\beta}$. We investigate the scenario with the correlation range $\rho_\beta = 0.8$. Recall that in the Matérn covariance function ρ_β indicates the distance between vertices at which the spatial correlation is about 0.1. It follows that the true values for κ_β and τ_β are fixed at $\kappa_\beta = 2.5$ and $\tau_\beta = 0.126$, respectively.
- *Simulated source and receiver fields $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ in \mathbb{S}^2 :*
 For $d = 2$ and $\sigma_\gamma^2 = 1$, the hyper-parameters of source field parameters are given as $\kappa_\gamma = 2\sqrt{2}/\rho_\gamma$ and $\tau_\gamma = 1/\sqrt{4\pi\kappa_\gamma^2}$ (the same equations are applied for the receiver field parameters).
 - *Source field $\boldsymbol{\gamma}$ with $\rho_\gamma = 0.8$:* The true values of τ_γ and κ_γ are given by $\tau_\gamma = 0.080$ and $\kappa_\gamma = 3.536$, respectively.
 - *Receiver field $\boldsymbol{\xi}$ with $\rho_\xi = 0.3$:* The true values of τ_ξ and κ_ξ are $\tau_\xi = 0.030$ and $\kappa_\xi = 9.428$, respectively.

Although the true values of the hyper-parameters can be chosen arbitrarily, the range parameter ρ should not exceed the range of the entire domain (the maximum distance between two arbitrary mesh vertices or in general between locations), as discussed by Simpson et al. (2012). The maximum of pairwise distances are 1.23 km for the receiver field, and 2 km for the source field. These are scaled values because the sphere with filling interior used for the velocity field $\boldsymbol{\beta}$ is also re-scaled from a radius of 6420 km to a radius of 20 km in order to avoid numerical instability causing by the INLA-functions. (6420 km is slightly larger than the earth's radius to ensure a convex mesh.) The estimated values of the hyper-parameters κ_β and τ_β are adapted to the radius of the sphere. The parameters could be re-scaled to obtain a model for any radius.

4.2. Performance evaluation measures

To assess the performance of the models, we evaluate the misfit using the Mahalanobis distance, defined by $\|\mathbf{x}\|_\Sigma := \sqrt{(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}$ with the covariance matrix Σ . The deviance information criterion (DIC) and the effective number of parameters (denoted as p_D) are computed using the INLA program. The model with the smallest DIC value is

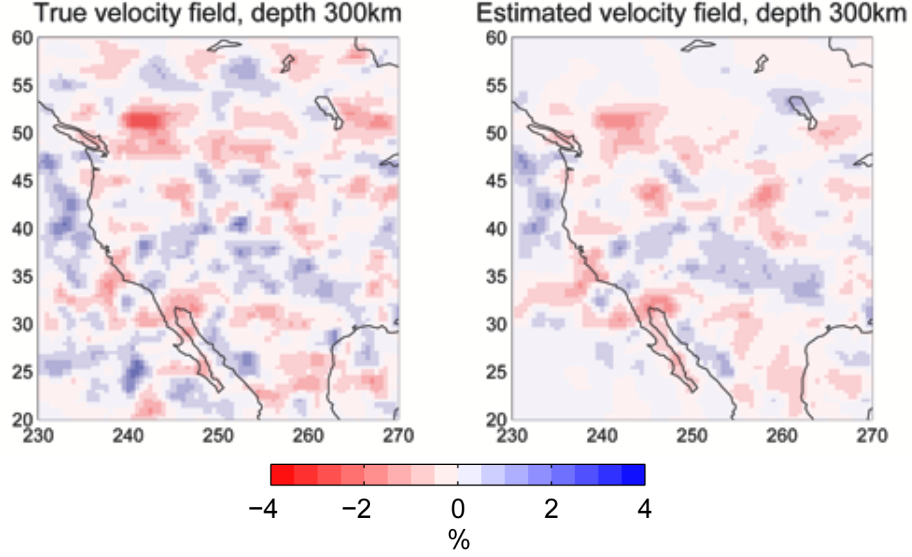


Fig. 4. Bayesian tomography using simulated data. Synthetic true velocity field β (left) and its posterior mean as estimated by the VSR model, using $\beta_0 = 0$ as the prior (right). Colour indicates P-wave velocity anomalies β in % relative to the spherically symmetric reference model IASP91. Blue areas map out faster-than-expected seismic wave velocities, red indicates slower than expected.

preferred with regard to both model fit and model complexity (Spiegelhalter et al., 2002). For p_D , we obtain $p_D \approx D - \text{tr}(Q(\theta)Q^*(\theta)^{-1})$, where $Q(\theta)$ is the prior precision matrix and $Q^*(\theta)$ is the posterior covariance matrix and $D := -2\log(\pi(\mathbf{y}|\theta))$. The term p_D can be seen as a Bayesian measure of fit. As pointed out by Rue et al. (2009), the smaller the value p_D is compared to the number of the data, the better the model fits the data.

4.3. Simulation results and interpretation

Recall that the synthetic data are generated using the VSR model and estimated using VSR, VS and VR models. We show an exemplary 2-D section through the 3-D velocity field β in Figure 4, comparing the true synthetic β field – seismic P-wave velocity anomalies at a depth of 300 km – to its estimated posterior mean from the VSR model. The images show that our model successfully retrieves the major structural features, especially the portion of the seismic velocity field located beneath the densely instrumented western United States (c.f. Figure 1).

Figure 5 illustrates the true source and receiver fields (top row), as well as their posterior mean estimates from the VSR model (bottom row). It shows that the VSR model can recover the basic structure of the true source and receiver fields. However, for areas with fewer data or less dense mesh points, structure cannot be reconstructed as well, such as in parts of the Indian and Southern Oceans. This is explained by the lack of sources in these regions, because the maximum distance of earthquakes from North America was 85° in this teleseismic P-wave tomography.

The summary statistics resulting from the VSR, VS and VR models in Table 1 show

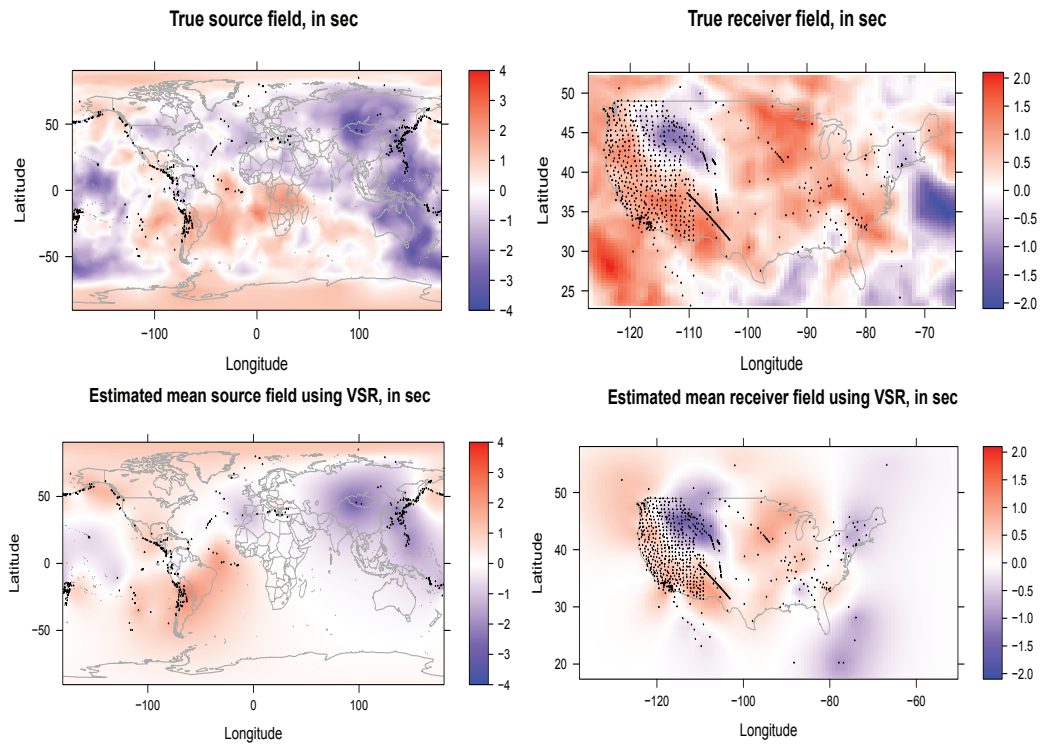


Fig. 5. Bayesian tomography using simulated data: estimation of source and receiver fields using the VSR model. Top left: true synthetic source field. Bottom left: posterior mean of the source field as estimated from the data. Top right: true synthetic receiver field. Bottom right: posterior mean of the receiver field as estimated from the data. Black dots are source and receiver locations. Colour shows traveltime anomalies in seconds, originating from near-source and near-receiver effects rather than from seismic velocity anomalies deep in the mantle. Blue means that a source or receiver causes an advanced arrival time y_i^{obs} relative to its prediction y_i^{pred} ; red indicates a traveltime delay.

Table 1. Summary statistics of the posterior estimates of the hyper parameters ϕ , τ and κ of models VSR, VR und VS. Simulated data are generated using model VSR.

Summary statistics								
Model	Parameters	ϕ	τ_β	κ_β	τ_γ	κ_γ	τ_ξ	κ_ξ
VSR	True	1.000	0.126	2.5	0.080	3.536	0.030	9.428
	mean	1.001	0.122	2.343	0.074	3.581	0.031	15.720
	2.5%	0.989	0.113	1.984	0.065	1.774	0.024	10.714
	97.5%	1.015	0.130	2.876	0.088	5.347	0.039	23.226
VS	mean	0.981	0.122	2.207	0.080	2.995		
	2.5%	0.969	0.115	1.896	0.070	2.002	—	—
	97.5%	0.998	0.132	2.513	0.095	4.340		
	mean	0.669	0.064	0.564			0.031	13.066
VR	2.5%	0.661	0.061	0.280	—	—	0.024	8.626
	97.5%	0.680	0.067	0.977			0.039	19.320

Table 2. Performance measures DIC, p_D and MaxLik of synthetic random field models VSR, VR and VS.

Performance measures						
Model	$\ \mathbf{y} - H\mathbf{m}\ _{\Sigma_y}$	$\ \mathbf{y} - H\mathbf{m}_L\ _{\Sigma_y}$	$\ \mathbf{y} - H\mathbf{m}_U\ _{\Sigma_y}$	DIC	p_D	MaxLik
VSR	226.99	633.09	634.22	152,917	1813	-428,397
VS	232.39	750.28	748.27	153,929	1709	-428,852
VR	184.93	891.80	891.98	174,707	2162	-439,986

that the true hyper-parameters can be recovered within the 95% credible intervals from the estimates of the VSR and VS models. Estimates of VR can only correctly identify the receiver field while the velocity estimate is polluted by the attempt to fit traveltimes originating from the unmodelled source field. We also examine the performance of the VR model if the synthetic receiver field is simulated with a high spatial correlation parameter (not shown in Table 1), i.e., $\rho_\xi = 0.8$, instead of the listed value of $\rho_\xi = 0.3$. In this case the true hyper-parameters of the velocity field can be recovered within 95% credible intervals, but the estimated mean cannot identify the true value.

The DIC and p_D values measuring Bayesian complexity and data fit, as well as the maximum likelihood value MaxLik in Table 2 provide the same conclusion: VSR and VS perform about equally well according to their DIC, p_D and MaxLik values, whereas performance of VR is worse on all accounts. Furthermore, the criteria ($\|\mathbf{y} - H\mathbf{m}_L\|_{\Sigma_y}$ and $\|\mathbf{y} - H\mathbf{m}_U\|_{\Sigma_y}$) show the misfit given the posterior 90% lower and upper 90% quantiles estimates \mathbf{m}_L and \mathbf{m}_U , respectively. They indicate that the range of data misfits for both the VSR and VS models is smaller than for VR.

5. Application to the seismic traveltime data

In this section we apply the INLA algorithm to infer the seismic P-wave velocity field as well as the source and receiver fields using real traveltime delay data measured by Sigloch (2008). We use only broadband traveltime observations from $N = 53,270$ unique source-receiver paths whereas the original data set contained additional bandpassed measurements for the same source-receiver paths. Similar data sets for the western United States have been studied by, e.g., Burdick et al. (2008), Sigloch et al. (2008), Schmandt and Humphreys (2010), and Zhang et al. (2013). Traditionally, the velocity field is estimated by minimising the L2-norm using the LSQR algorithm (Nolet, 1987; Sigloch et al., 2008; Tian et al., 2009), whereas Zhang et al. (2013) applied the MCMC algorithm to provide estimates with uncertainty assessment. We estimate over $D = p + p_s + p_r \approx 13,000$ parameters consisting

Table 3. Performance measures of posterior estimation results using real seismic data show that VSR and VS models achieve better estimates than VR or V models.

Model	Performance measure						
	$\ \mathbf{y} - H\mathbf{m}\ _{\Sigma_y}$	$\ \mathbf{y} - H\mathbf{m}_L\ _{\Sigma_y}$	$\ \mathbf{y} - H\mathbf{m}_U\ _{\Sigma_y}$	DIC	p_D	MaxLik	Time in hours
VSR	147.60	288.44	288.20	198,666	1,282	-450,623	33.8
VS	146.61	319.78	319.53	198,817	1,253	-450,677	10.5
VR	144.96	277.70	277.73	200,607	967	-451,374	8.7
V	143.21	333.47	333.56	200,633	1,035	-451,362	4.18

of $p = 8977$ tetrahedral nodes, $p_s = 2200$ source field parameters and $p_r = 1800$ receiver field parameters. With Gaussian error assumptions we fit the data using the models VSR, VS, VR or V (see Section 4) within the INLA framework.

As prior information, we used the solution obtained by Sigloch (2008) using iterative least-squares optimization (LSQR). This solution included a velocity field (our prior β_0), but also estimates for source corrections. We converted these corrections (for earthquake location and time) into traveltimes anomalies, and applied them to our data \mathbf{Y} . Hence our source field γ effectively evaluates the quality of their corrections, rather than correcting for earthquake mislocations from scratch. We made this choice in the interest of direct comparison with the results of Zhang et al. (2013), who also accepted the corrections obtained by Sigloch (2008). By contrast, receiver corrections were not treated as free parameters by Sigloch (2008). Instead they used prior models of the earth’s crustal structure, topography, and ellipticity to compute receiver corrections. We apply these same traveltimes corrections to our data \mathbf{Y} , so that effectively our receiver field ξ assesses prior information about surface structure around the receiver, especially crustal structure, which is relatively uncertain.

5.1. Posterior results and interpretation of the velocity, source and receiver fields

Table 3 provides an overview of the performance measures of the models. Results from the VSR and VS models yield similar estimates, which are considerably better than the results from the VR and V models with respect to DIC and the maximum likelihood estimate (MaxLik). Due to the high number of parameters, the computing time of VSR is 3 to 6 times higher than for the models with only one or two random fields. In summary, VSR or VS models would estimate the parameters and fit the data better from the perspective of DIC and p_D . In terms of computing time, the reduced models VS and V are significantly less expensive than the VSR model and could therefore be preferable.

In Figure 6, the posterior mean of the 3-D velocity fields (at 400 km depth) from the VSR and V models display strong similarities. But their respective credible intervals (third and fourth columns) are different from the results of the MCMC approach in Zhang et al. (2013), particularly in the north-eastern part of the domain. The maps of significant regions show posterior modes rendered on regions that differ from the spherically symmetric reference model ($\beta = \mathbf{0}$) with 95% posterior probability. These maps provide independent confirmation of geological inference: red areas map out particularly hot upper mantle under the volcanic provinces and plate boundary regions of western North America (Juan de Fuca spreading ridge, Basin and Range province, Yellowstone), where seismic waves travel more slowly than predicted by the globally averaged IASP91 earth model; blue anomalies map out the western edge of the old and cool North American craton (Zhang et al., 2013).

Figure 7 illustrates the significant regions of the estimated posterior mean for the source and receiver fields from the VSR model, including nonsignificant receiver positions. Sig-

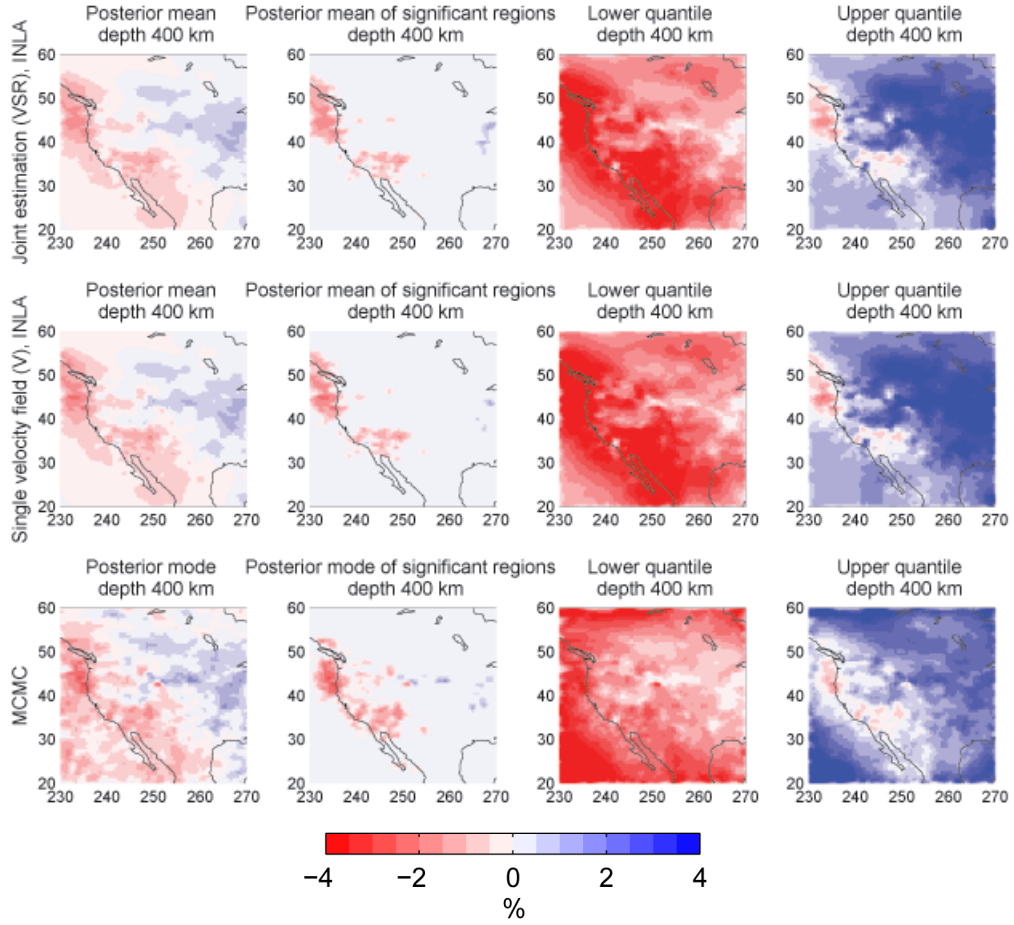


Fig. 6. Tomographic images from the Bayesian inversion of real data: Comparison of results from the VSR model (top row), V model (middle row), and the MCMC inversion of Zhang et al. (2013) (bottom row). Colour indicates seismic velocity anomalies in per cent with respect to the spherically symmetric reference model IASP91, at a depth of 400 km under North America. Red regions indicate slower-than-expected mantle structure, blue is faster than expected. Physically, blue (fast) usually implies a colder-than average region, whereas red (slow) means hotter. Four columns show (1) the posterior mean (or posterior mode in the case of MCMC); (2) the posterior mean or mode only in regions where is rises above the 95% significance level; (3) the lower quantile of the velocity anomaly distribution; (4) the upper quantile.

nificant regions map out the areas that differ from a zero mean field with 95% posterior probability. Within such significant regions, coloured dots near sources or receivers show that delay times originate there with high probability, indicating that velocity anomalies β in the mantle are not sufficient to explain all traveltime observations. Since the β -field is already corrected for the deterministic source and receiver effects estimated by Sigloch (2008), coloured dots in significant regions diagnose sources and receivers for which the deterministic corrections seems to have been not completely adequate. (In effect we chose to estimate required deviations from the corrections of Sigloch (2008), rather than the cor-

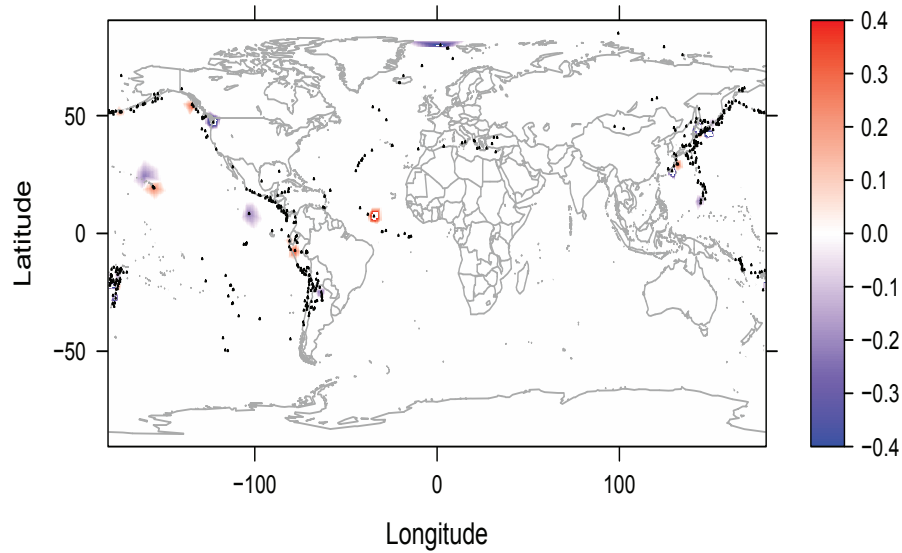
rections themselves. We return to this point below.) The obtained source field looks rather “white”, indicating only a few problematic earthquakes (mainly in the northern Pacific), and they are associated with relatively small time discrepancies of a few tenths of a second. Since earthquake origin times are typically uncertain by several seconds, the deterministic source corrections appear to have been largely effective. Similarly, relatively few receivers are affected by a significant delay time discrepancy, and even for them, the time discrepancies are small (Figure 7, bottom right). This evaluation of deterministic corrections is a valuable output of our probabilistic approach. We experimented with alternative parameterisations for the source and receivers fields, where mesh points did not coincide with source and receiver locations. The estimation results were very similar to those shown in Figure 7.

In the following we compare estimated posterior results from INLA with results of the MCMC method by Zhang et al. (2013). The statistical model in Zhang et al. (2013) is formulated as $\mathbf{y} = X\boldsymbol{\beta} + X_{\text{hyp}}\boldsymbol{\beta}_{\text{hyp}} + X_{\text{time}}\boldsymbol{\beta}_{\text{time}} + \boldsymbol{\epsilon}$, where X_{hyp} and X_{time} are sensitivity kernels for earthquake hypocenter and origin time corrections. The quantities $\boldsymbol{\beta}_{\text{hyp}}$ and $\boldsymbol{\beta}_{\text{time}}$ are the source parameters to be estimated. The model error $\boldsymbol{\epsilon}$ is assumed to be i.i.d. Gaussian distributed with zero mean and a constant variance. In Zhang et al. (2013) an ellipsoidal spatial prior on $\boldsymbol{\beta}$ is constructed, which follows a conditional autoregressive model as defined in Pettitt et al. (2002). This model is non-isotropic and non-stationary. The same prior mean – the LSQR solution of $\boldsymbol{\beta}_0$ – is applied in both Zhang et al. (2013) and in this study. A Gibbs-sampling with a MH-step is adopted for estimation. The parameter dimension of the earlier study was about 11,000, which is less than the $\approx 13,000$ model parameters of this study.

$\boldsymbol{\beta}_{\text{hyp}}$ (three free parameters per source) and $\boldsymbol{\beta}_{\text{time}}$ (one parameter per source) have the physical meaning of source location and origin time, respectively, whereas $X_{\text{hyp}}\boldsymbol{\beta}_{\text{hyp}}$ and $X_{\text{time}}\boldsymbol{\beta}_{\text{time}}$ are traveltime anomalies. Both Sigloch (2008) and Zhang et al. (2013) estimated $\boldsymbol{\beta}_{\text{hyp}}$ and $\boldsymbol{\beta}_{\text{time}}$. In principle, the INLA method is equally capable of handling such spatially discrete corrections, but we did not attempt this due to the large number of additional parameters and considerations of computing time. Instead, we chose the approach of estimating the continuous source field $\boldsymbol{\gamma}$ of traveltime anomalies, and in addition we corrected our traveltime observations \mathbf{Y} *a priori*, using the deterministic source and receiver corrections obtained by Sigloch (2008). The fields $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ exist at all vertices of the source and receiver field meshes, and their spatial correlations are modelled, whereas $\boldsymbol{\beta}_{\text{hyp}}$ and $\boldsymbol{\beta}_{\text{time}}$ assumes no spatial correlations and exist only at source and receiver locations. As a practical note on this rather technical subject, different tomographers choose to treat source and receiver corrections differently. Our main purpose here was not (yet) to quantitatively compare to various existing practices, but to introduce an alternative concept for how these corrections could be undertaken, namely by continuous and spatially correlated 2-D random fields, and to deliver a proof of concept.

The comparison of the INLA results to earlier MCMC and LSQR inversions, e.g. the models by Zhang et al. (2013) and Sigloch (2008), confirms that Bayesian inference using INLA successfully retrieves the first-order structures of earth’s upper mantle under the densely instrumented and thus well illuminated western United States. Values at the boundaries of the images from INLA are smoother than the results from MCMC. Maps of significant regions resulting from VSR, V and MCMC are similar. The MCMC algorithm is programmed in MATLAB using parallelized built-in Cholesky decomposition for sampling from a multivariate normal. On a Intel Xeon E7-4860 (2.27GHz) with 40 cores, it

Significant region of estimated mean source field using VSR, in sec



Estimated mean receiver field using VSR, in sec

Significant region of estimated mean receiver field, in sec

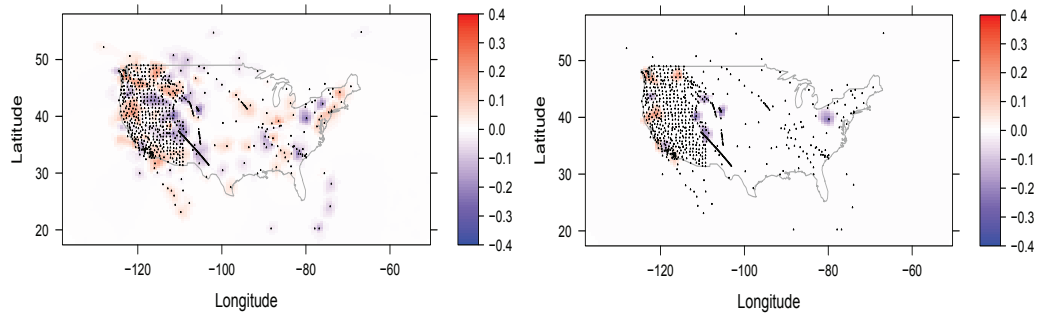


Fig. 7. Inversion of real data using the VSR model: estimated posterior mean of the source field (top panel) and the receiver field (bottom panels). Black dots indicate source and receiver locations. Blue colour indicates earlier-than-predicted wave arrival times, red indicates later-than-predicted arrivals. For the source field (top), time anomalies are shown only if located in 95% significant areas, but most areas are 95% significant. For the receiver field, the bottom left panel shows all time anomalies, whereas the bottom right panel shows only those anomalies located in 95% significant areas.

runs for over 32 CPU hours to estimate about 11,000 parameters (velocity parameters + source correction parameters), and about 7.5 CPU hours to estimate about 8977 velocity parameters (using 10,000 MCMC runs). On the same infrastructure, the R-INLA program (INLA build date: Dec 8. 2012, hgid 6100f01c2e05) takes about 33 CPU hours to run the full model VSR with about 13,000 parameters (velocity parameters + source and receiver fields on S^2 -grids). Estimating model V, which contains only the 8977 velocity parameters, takes about 4 CPU hours with INLA. This shows that for this type of seismic tomography problems, the INLA algorithm has a computational advantage over the MCMC method

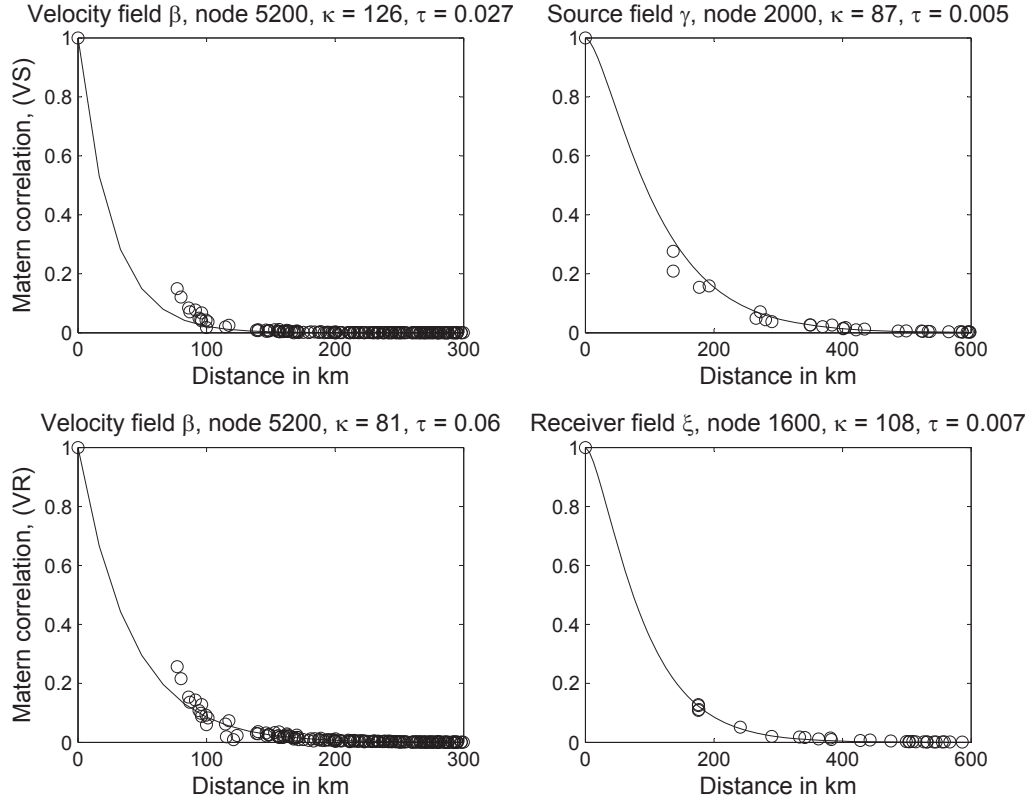


Fig. 8. Matérn correlations of the velocity, source, and receiver fields, as estimated from real data by the VS model (top row) and the VR model (bottom row). Results for three exemplary mesh nodes are shown. For velocity field β at node #5200 (located at 160 km depth), posterior mean values are $\kappa_\beta = 126$, $\tau_\beta = 0.027$ for the VS model (top left) and $\kappa_\beta = 81$, $\tau_\beta = 0.06$ for the VR model (bottom left). For the source field γ at node #2000, $\kappa_\gamma = 87$ and $\tau_\gamma = 0.005$ (top right). For the receiver field ξ at node #1600, $\kappa_\xi = 108$ and $\tau_\xi = 0.007$ (bottom right). Circles are estimated correlations obtained from the approximated Matérn correlation matrix using the SPDE approach. They are close to their theoretical values (solid lines), which confirms the approximate correctness of our 3-D Matérn model.

while delivering similar posterior mean images of the velocity fields.

5.2. Estimated Matérn correlation of the velocity, source and receiver fields

Figure 8 shows the Matérn correlation of one vertex in each of the estimated velocity, source and receiver fields from the VS and VR models. The Matérn correlation of an exemplary node #5200 of the velocity field with posterior mean estimates reveals spatial correlation within 150 km, and the correlation falls to close to zero at a distance of around 200 km (left panels).

The source and receiver fields (with estimates $\kappa_\gamma = 87$ and $\tau_\gamma = 0.005$ for the source field and $\kappa_\xi = 108$ and $\tau_\xi = 0.007$ for the receiver field) feature similar correlation length

scales as the 3-D velocity field, which can be also confirmed by comparing the relative magnitude of κ 's: the larger the value of κ is, the weaker the spatial correlation. Weak spatial correlations in the source and receiver fields imply that the corrections that had been applied a priori, based on a 3-D model of the continental, crust and source location corrections, explained the data well at most locations.

The two mesh vertices for which empirical Matérn correlations are illustrated in Figure 8 are selected arbitrarily from the well-resolved central areas, away from mesh boundaries. We note that the empirical correlations can vary from the theoretical correlations if the vertices are located at the mesh boundaries. Generally, variances at vertices close to boundaries are much larger due to the Neumann boundary conditions (Lindgren et al., 2011; Simpson et al., 2011). This phenomenon means that the approximation of the 3-D GF is a solution to the SPDE, but not a stationary solution (constant variance everywhere inside the domain). One possibility to overcome this non-stationarity is to construct a larger meshed domain than the target domain to run the inference. The original radius of our tetrahedral mesh for the globe is 6420 km, 50 km larger than the true radius of the earth (6371 km). However, this may not be sufficient to generate a purely stationary field for our target region. Theoretical solutions to avoid variance inhomogeneity are topics of current research.

Concerning identifiability, proper posterior estimates of the velocity and source fields can be achieved by the full model VSR but also by the reduced model VS. Within the ranges investigated here, posterior estimates were not very sensitive to different mesh grids or prior setups. The finding that for all practical purposes, tomographic images are similar regardless of the prior used, indicates that posterior estimates are well identified and mainly explained by the data. If computing resources are a limiting factor, the VS model is a pragmatic and viable alternative to the full VSR model.

6. Discussion and outlook

This paper studies high-dimensional linear inverse problems by means of Bayesian spatial modelling techniques. We construct a new, computationally efficient approximation to the velocity field on a \mathbb{R}^3 -manifold using the stochastic partial differential equations (SPDE) approach by Lindgren et al. (2011). In our motivating application, we estimate the 3-D seismic wave velocity field in the upper mantle beneath the western United States, while also allowing for spatial correlation structure of the data errors, which are expressed through the source and receiver fields over the earth's surface. In our tomographic problem, Bayesian uncertainty quantification of the random fields is carried out jointly by the integrated nested Laplace approximation (INLA) of Rue et al. (2009) with the R-INLA software. Our application to upper mantle structure involves about 13,000 parameters and over 53,000 observations. Simulation studies and inversion of real data show that INLA has a huge computational advantage over the MCMC method in Zhang et al. (2013), while delivering comparable posterior tomographic images with better statistical fits.

Our current model using INLA assumes that the posterior distribution of the latent model is close to Gaussian, in which case the INLA approximation is exact (Rue et al., 2009). For non-Gaussian data errors, such as t-distributed error, INLA is computationally still too time consuming for our application. Martin et al. (2012) extends the INLA theory beyond the scope of latent Gaussian models, where the latent field can have a near-Gaussian distribution. Furthermore, our spatial model in 3-D using the SPDE approach can be flexibly extended to non-stationary, non-isotropic Gaussian fields (Simpson et al., 2012;

Fuglstad et al., 2013; Lindgren and Rue, 2014; Ingebrigtsen et al., 2014). Following the notations in Ingebrigtsen et al. (2014) for these models the hyper-parameters of a non-stationary model $\tau := \tau(\mathbf{u})$ and $\kappa := \kappa(\mathbf{u})$ are functions of the spatial locations $\mathbf{u} \in U$. The logarithm of these parameters are formulated as sums of basis functions with a set of weight parameters, i.e., $\log(\tau(\mathbf{u})) = \theta_1^\tau + \sum_{k=2}^N b_k^\tau(\mathbf{u})\theta_k^\tau$ and $\log(\kappa(\mathbf{u})) = \theta_1^\kappa + \sum_{k=2}^N b_k^\kappa(\mathbf{u})\theta_k^\kappa$, where b_k^τ and b_k^κ are additional space-dependent basis functions (for example using low order spherical harmonics and B-splines as basis functions). The parameters $\theta_1^\tau, \dots, \theta_N^\tau$ and $\theta_1^\kappa, \dots, \theta_N^\kappa$ are to be estimated. If the random field is stationary, N equals to 1, otherwise N is the squares of the degrees of the basis functions. For instance, for a degree of 2, we have 3 additional parameters for each hyper-parameter to estimate. Incorporating these space-dependent hyper-parameters the precision matrix Q in (10) is modified to

$$Q(\boldsymbol{\theta}) = T(K^2 \tilde{C} K^2 + K^2 G + G K^2 + G \tilde{C}^{-1} G),$$

where $T_{ii} := \tau(\mathbf{u}_i)$ and $K_{ii} := \kappa(\mathbf{u}_i)$. In the non-stationary setting computational cost is expected to be higher due to the high number of hyper-parameters. Additional research is needed here. Replacing a single SPDE with a system of SPDEs, multivariate SPDEs were investigated by Hu et al. (2013a,b). Other research focuses on the extension of Matérn class approximations to other classes of the continuous Gaussian fields, for example, constructing other Markov field approximations when α in the SPDE equation (9) is not an integer. For a comprehensive discussion on the SPDE approach we refer to Lindgren (2012); Simpson et al. (2012).

In conclusion, our application is the first example of modelling continuous Gaussian fields in a 3-D space using the SPDE approach. Our application demonstrates an example of very high-dimensional parameter estimation carried out within the INLA framework that captures spatial dependency of data errors caused by different regions simultaneously.

Acknowledgement

We thank Dr Finn Lindgren, two anonymous reviewers and the editor for discussions and constructive comments. The authors acknowledge the support of the computational facilities of the Geophysics department at the Ludwig-Maximilians Universität München and the Leibniz-Supercomputing in Garching. The first author is funded by the project B3 of the Munich Center of Advanced Computing (MAC/IGSSE), Technische Universität München.

Appendix: Calculation of the C and G matrices on 3-D tetrahedral mesh

The calculation of the \tilde{C} and G matrices is accomplished by using the theory of the finite element method (FEM). Advanced implementation techniques can be seen in Albery et al. (2002). Here, we show the explicit expression of the \tilde{C} and G matrices for one tetrahedron as shown in Figure 9. First, we denote the four vertices in a tetrahedron as

$$\mathbf{a}_1 := (x_1, y_1, z_1)^t, \mathbf{a}_2 := (x_2, y_2, z_2)^t, \mathbf{a}_3 := (x_3, y_3, z_3)^t \text{ and } \mathbf{a}_4 := (x_4, y_4, z_4)^t.$$

We also define the *directional* vectors \mathbf{v}_{ij} as $\mathbf{v}_{ij} := \mathbf{a}_i - \mathbf{a}_j$ for $i, j = 1, \dots, 4$ as illustrated in Figure 9. Every tetrahedral element in the original mesh will be calculated based on a reference tetrahedron with the (ξ, γ, η) -coordinates. Furthermore, the cross product of two vectors $\mathbf{u} := (u_x, u_y, u_z)^t$ and $\mathbf{v} := (v_x, v_y, v_z)^t$ is defined by

$$\mathbf{u} \times \mathbf{v} := \mathbf{e}_1(u_y v_z - u_z v_y) + \mathbf{e}_2(u_z v_x - u_x v_z) + \mathbf{e}_3(u_x v_y - u_y v_x),$$

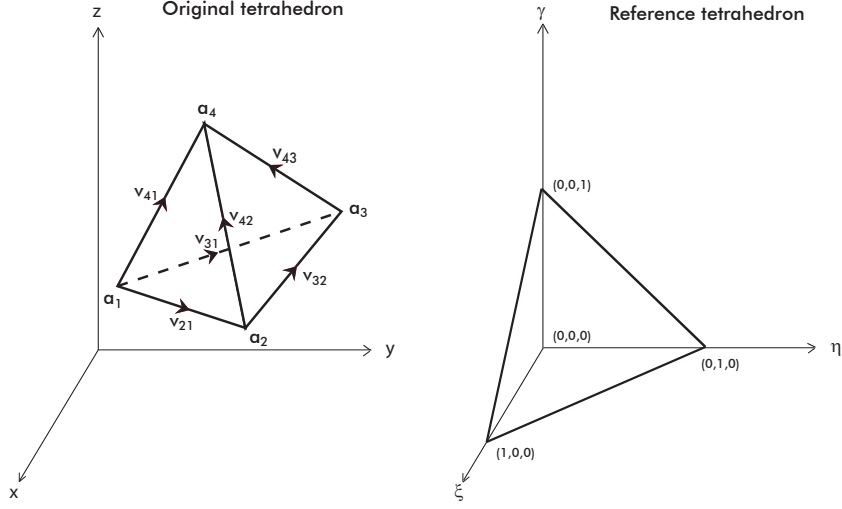


Fig. 9. Left: original tetrahedron; right: reference tetrahedron

where $\mathbf{e}_1 := (1, 0, 0)^t$, $\mathbf{e}_2 := (0, 1, 0)^t$ and $\mathbf{e}_3 := (0, 0, 1)^t$ are standard basis vectors. The cross product of two vectors is always perpendicular to both vectors with the orientation determined by the right-hand rule (Arfken, 1985; Bronshtein et al., 2007).

Definitions of the basis functions and their gradients.

Let $\varphi_i(x, y, z) : \mathbb{R}^3 \rightarrow \mathbb{R}$ be the piecewise linear basis function in the original coordinates as defined in (8). Further, we denote the piecewise linear basis functions in the reference coordinate as $\tilde{\varphi}_i(\xi, \eta, \gamma) : \mathbb{R}^3 \rightarrow \mathbb{R}$, $i = 1, \dots, 4$. They are defined as

$$\tilde{\varphi}_1(\xi, \eta, \gamma) := 1 - \xi - \eta - \gamma, \quad \tilde{\varphi}_2(\xi, \eta, \gamma) := \xi, \quad \tilde{\varphi}_3(\xi, \eta, \gamma) := \eta, \quad \tilde{\varphi}_4(\xi, \eta, \gamma) := \gamma. \quad (15)$$

For expressing the four vertices of the tetrahedron in the (x, y, z) -coordinates by the standard (ξ, η, γ) -coordinates, we consider the mapping $F : \mathbb{R}^3 \rightarrow \mathbb{R}^3$. In particular for $\mathbf{p} := (x, y, z)^t$ define $\mathbf{p} = F(\tilde{\mathbf{p}})$ with $\tilde{\mathbf{p}} := (\xi, \eta, \gamma)^t$, then F is given by,

$$F : \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_2 - x_1 & x_3 - y_1 & x_4 - z_1 \\ y_2 - y_1 & y_3 - y_1 & y_4 - z_1 \\ z_2 - z_1 & z_3 - y_1 & z_4 - z_1 \end{pmatrix} \begin{pmatrix} \xi \\ \eta \\ \gamma \end{pmatrix} + \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} =: J\tilde{\mathbf{p}} + \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix}$$

where J is the Jacobian matrix. The inverse of F is $\tilde{\mathbf{p}} = F^{-1}(\mathbf{p})$. Then the generic basis function in the original coordinates can be also written as $\varphi_i(\mathbf{p}) = \tilde{\varphi}_i(F^{-1}(\mathbf{p})) = \tilde{\varphi}_i(\tilde{\mathbf{p}})$. The gradient of the basis functions φ_i can be obtained using the chain rules, that is,

$$\nabla \varphi_i(\mathbf{p}) = \left(\frac{\partial F}{\partial \tilde{\mathbf{p}}} \right)^{-1} \nabla \tilde{\varphi}_i(F^{-1}(\mathbf{p})) := J^{-1} \nabla \tilde{\varphi}_i(\tilde{\mathbf{p}}). \quad (16)$$

It can also be rewritten as

$$J = \begin{bmatrix} \mathbf{v}_{21} & \mathbf{v}_{31} & \mathbf{v}_{41} \end{bmatrix}^t$$

with directional vectors $\mathbf{v}_{21} := \mathbf{a}_2 - \mathbf{a}_1$, $\mathbf{v}_{31} := \mathbf{a}_3 - \mathbf{a}_1$ and $\mathbf{v}_{41} := \mathbf{a}_4 - \mathbf{a}_1$ in \mathbb{R}^3 , as shown in Figure 9. The inverse of the Jacobian matrix can be expressed using the cross products of vectors, i.e.,

$$J^{-1} = \frac{1}{6|K|} [\mathbf{b}_2 \ \mathbf{b}_3 \ \mathbf{b}_4],$$

due to the identity $\det(J) = 6|K|$, where $|K|$ denote the volume of the tetrahedron K (Bronshtein et al., 2007). Here, $\mathbf{b}_i \in \mathbb{R}^3$ is the cross product of the two vectors that span the triangle opposite to the vertex i in a tetrahedron, $i = 1, \dots, 4$, (Bronshtein et al., 2007). Further, we denote by K and \tilde{K} the tetrahedra in the original and reference coordinates, respectively. By inserting J^{-1} into (16) we then have

$$\nabla \varphi_i(\mathbf{p}) = \frac{1}{6|K|} [\mathbf{b}_2 \ \mathbf{b}_3 \ \mathbf{b}_4] \nabla \tilde{\varphi}_i(\mathbf{p}).$$

Calculation of the \tilde{C}_{ii} element.

Recall that $\varphi_i := \varphi_i(x, y, z) : \mathbb{R}^3 \rightarrow \mathbb{R}$ and $\tilde{\varphi}_i := \tilde{\varphi}_i(\xi, \eta, \gamma) : \mathbb{R}^3 \rightarrow \mathbb{R}$ are the basis functions in the original coordinates (x, y, z) and in the reference coordinates (ξ, η, γ) , respectively, as defined in (15). The corresponding gradients of φ_i and $\tilde{\varphi}_i$ are $\nabla \varphi_i : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ and $\nabla \tilde{\varphi}_i : \mathbb{R}^3 \rightarrow \mathbb{R}^3$, respectively. For the precision matrix Q_{ij} with one tetrahedron as shown in Figure 9 we need the matrix \tilde{C} :

$$\tilde{C}_{ii} := \iiint_K \varphi_i dz dy dx.$$

In particular, we have

$$\tilde{C}_{ii} = \iiint_K \varphi_i dz dy dx = \iiint_{\tilde{K}} \tilde{\varphi}_i \det(J) d\gamma d\eta d\xi = \int_0^1 \left[\int_0^{1-\xi} \left[\int_0^{1-\xi-\eta} \tilde{\varphi}_i \det(J) d\gamma \right] d\eta \right] d\xi$$

For $i = 1$, this simplifies to $\tilde{C}_{11} = \det(J) \int_0^1 \left[\int_0^{1-\xi} \left[\int_0^{1-\xi-\eta} (1 - \xi - \eta - \gamma) d\gamma \right] d\eta \right] d\xi = \frac{|K|}{4}$. After similar algebraical calculation for $i = 2, 3, 4$ it follows that

$$\tilde{C} := \frac{|K|}{4} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Calculation of the G_{ij} element.

Given the gradients of $\tilde{\varphi}$, i.e. $\nabla \tilde{\varphi}_1 = \nabla(1 - \xi - \eta - \gamma)^t = (-1, -1, -1)^t$, $\nabla \tilde{\varphi}_2 = (1, 0, 0)^t$, $\nabla \tilde{\varphi}_3 = (0, 1, 0)^t$ and $\nabla \tilde{\varphi}_4 = (0, 0, 1)^t$, the gradient of φ_i is $\nabla \varphi_i = J^{-1} \nabla \tilde{\varphi}_i$, $i = 1, \dots, 4$,

which simplifies to:

$$\begin{aligned}
i = 1 : \quad \nabla \varphi_1 &= J^{-1} \nabla \tilde{\varphi}_1 = -\frac{1}{\det(J)} (\mathbf{b}_2 + \mathbf{b}_3 + \mathbf{b}_4) = -\frac{1}{\det(J)} (\mathbf{v}_{43} \times \mathbf{v}_{23}) = \frac{1}{\det(J)} \mathbf{b}_1 \\
&\quad \text{with } \mathbf{b}_1 := -(\mathbf{v}_{43} \times \mathbf{v}_{23}) = \mathbf{v}_{43} \times \mathbf{v}_{32}, \\
i = 2 : \quad \nabla \varphi_2 &= J^{-1} \nabla \tilde{\varphi}_2 = \frac{1}{\det(J)} \mathbf{b}_2, \quad \mathbf{b}_2 := \mathbf{v}_{31} \times \mathbf{v}_{41}, \\
i = 3 : \quad \nabla \varphi_3 &= J^{-1} \nabla \tilde{\varphi}_3 = \frac{1}{\det(J)} \mathbf{b}_3, \quad \mathbf{b}_3 := \mathbf{v}_{41} \times \mathbf{v}_{21}, \\
i = 4 : \quad \nabla \varphi_4 &= J^{-1} \nabla \tilde{\varphi}_4 = \frac{1}{\det(J)} \mathbf{b}_4, \quad \mathbf{b}_4 := \mathbf{v}_{21} \times \mathbf{v}_{31}.
\end{aligned}$$

Then, the (i, j) th element of the G matrix is given by,

$$\begin{aligned}
G_{ij} &:= \iiint_K \nabla \varphi_i^t \nabla \varphi_j \, dz dy dx \\
&= \iiint_{\tilde{K}} (J^{-1} \nabla \tilde{\varphi}_i)^t (J^{-1} \nabla \tilde{\varphi}_j) \det(J) \, d\gamma d\eta d\xi = \frac{1}{\det(J)} \mathbf{b}_i^t \mathbf{b}_j \iiint_{\tilde{K}} d\gamma d\eta d\xi.
\end{aligned}$$

Since the term $\iiint_{\tilde{K}} d\gamma d\eta d\xi = \frac{1}{6}$ represents the volume of the standard reference tetrahedron, we have $G_{ij} = \frac{1}{36|K|} \mathbf{b}_i^t \mathbf{b}_j$ or

$$G := \frac{1}{36|K|} \begin{pmatrix} \|\mathbf{b}_1\|^2 & \mathbf{b}_1^t \mathbf{b}_2 & \mathbf{b}_1^t \mathbf{b}_3 & \mathbf{b}_1^t \mathbf{b}_4 \\ \mathbf{b}_2^t \mathbf{b}_1 & \|\mathbf{b}_2\|^2 & \mathbf{b}_2^t \mathbf{b}_3 & \mathbf{b}_2^t \mathbf{b}_4 \\ \mathbf{b}_3^t \mathbf{b}_1 & \mathbf{b}_3^t \mathbf{b}_2 & \|\mathbf{b}_3\|^2 & \mathbf{b}_3^t \mathbf{b}_4 \\ \mathbf{b}_4^t \mathbf{b}_1 & \mathbf{b}_4^t \mathbf{b}_2 & \mathbf{b}_4^t \mathbf{b}_3 & \|\mathbf{b}_4\|^2 \end{pmatrix}.$$

References

- Alberty, J., C. Carstensen, S. A. Funken, and R. Klose (2002). Matlab implementation of the finite element method in elasticity. *Computing* 69, 239 – 263.
- Arfken, G. (1985). *Mathematical Methods for Physicists*, Chapter 1.4, pp. 18 – 26. Orlando, FL: Academic Press.
- Banerjee, S., A. E. Gelfand, and B. P. Carlin (2003). *Hierarchical Modeling and Analysis for Spatial Data. Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.
- Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang (2008). Gaussian predictive process models for large spatial data sets. *Journal of Royal Statistical Society: Series B* 70(4), 825 – 848.
- Bodin, T. and M. Sambridge (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International* 178(3), 1411 – 1436.
- Bronshstein, I. N., K. A. Semendyayev, G. Musiol, and H. Mühlig (2007). *Handbook of Mathematics* (5 ed.), Chapter 3, pp. 213. Springer.

- Burdick, S., C. Li, V. Martynov, T. Cox, J. Eakins, T. Mulder, L. Astiz, F. L. Vernon, G. L. Pavlis, and R. D. van der Hilst (2008). Upper mantle heterogeneity beneath North America from travel time tomography with global and USArray transportable array data. *Seismological Research Letters* 79(3), 389 – 392.
- Cameletti, M., F. Lindgren, D. Simpson, and H. Rue (2012). Spatio-temporal modeling of particular matter concentration through the SPDE approach. *AStA Adv. Stat. Anal.*. DOI 10.1007/s10182-012-0196-3.
- Cressie, N. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics.
- Dahlen, F. A., S.-H. Hung, and G. Nolet (2000). Fréchet kernels for finite-frequency traveltimes-I. Theory. *Geophysical Journal International* 141(1), 157 – 174.
- Dziewonski, A. M. and D. L. Anderson (1981). Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors* 25, 297 – 356.
- Fuglstad, G.-A., D. Simpson, F. Lindgren, and H. Rue (2013). Non-stationary spatial modelling with applications to spatial prediction of precipitation. Technical report, Department of Mathematical Sciences, NTNU, Norway.
- Furrer, R., R. Knutti, S. R. Sain, D. W. Nychka, and G. A. Meehl (2007). Spatial patterns of probabilistic temperature change projections from a multivariate Bayesian analysis. *Geophys. Res. Lett.* 34(6), L06711.
- Gockenbach, M. S. (2006). *Understanding and Implementing the Finite Element Method*. SIAM.
- Hu, X. P., F. Lindgren, D. Simpson, S. Martino, and H. Rue (2013a). Multivariate Gaussian Random Fields Using Systems of Stochastic Partial Differential Equations. *arXiv:1307.1379*.
- Hu, X. P., F. Lindgren, D. Simpson, S. Martino, and H. Rue (2013b). Multivariate Gaussian Random Fields with Oscillating Covariance Functions using Systems of Stochastic Partial Differential Equations. *arXiv:1307.1384*.
- Hughes, T. J. R. (1987). *The Finite Element Method. Linear Static and Dynamic Finite Element Analysis*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Illian, J. B., S. H. Sørbye, and H. Rue (2012). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *Ann. Appl. Statist.* 6(4), 1499 – 1530.
- Ingebrigtsen, R., F. Lindgren, and I. Steinsland (2014). Spatial models with explanatory variables in the dependence structure. *Spatial Statistics* 8, 20 – 38.
- Kennett, B. L. N. and E. R. Engdahl (1991). Traveletimes for global earthquake location and phase identification. *Geophysical Journal International* 105, 429 – 465.
- Lindgren, F. (2012). Continuous domain spatial models in R-INLA. *The ISBA Bulletin* 19(4).

- Lindgren, F. and H. Rue (2014). Bayesian spatial modelling with R-INLA. Technical report, University of Bath, United Kingdom.
- Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the SPDE approach (with discussion). *Journal of Royal Statistical Society: Series B* 73(4), 423 – 498.
- Martin, J., L. C. Wilcox, C. Burstedde, and O. Ghattas (2012). A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing* 34(3), A1460 – A1487.
- Mosegaard, K. and A. Tarantola (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research* 100(B7), 12431 – 12447.
- Nolet, G. (1987). Seismic wave propagation and seismic tomography. In G. Nolet (Ed.), *Seismic Tomography*, pp. 1 – 24. Dordrecht: D. Reidel Publishing Company.
- Nolet, G. (2008). *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*. Cambridge University Press.
- Paige, C. C. and M. A. Saunders (1982). LSQR: An algorithm for sparse linear equations and sparse least squares. *TOMS* 8(1), 43–71.
- Pettitt, N. A., I. S. Weir, and A. G. Hart (2002). A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modelling large sets of binary data. *Statistics and Computing* 12, 353 – 367.
- Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC.
- Rue, H. and S. Martino (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference* 137(10), 3177 – 3192. Special Issue: Bayesian Inference for Stochastic Processes.
- Rue, H., S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B* 71(2), 319 – 392.
- Sambridge, M. and O. Gudmundsson (1998). Tomographic systems of equations with irregular cells. *Journal of Geophysical Research* 103(B1), 773 – 781.
- Schmandt, B. and E. Humphreys (2010). Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth and Planetary Science Letters* 297(3-4), 435 – 445.
- Schrödle, B. and L. Held (2011). Spatio-temporal disease mapping using INLA. *Environmetrics* 22(6), 725 – 734.
- Sigloch, K. (2008). *Multiple-frequency body-wave tomography*. Ph. D. thesis, Princeton University.
- Sigloch, K. (2011). Mantle provinces under North America from multi-frequency P-wave tomography. *Geochemistry, Geophysics, Geosystems* 12(Q02W08).

- Sigloch, K., N. McQuarrie, and G. Nolet (2008). Two-stage subduction history under North America inferred from finite-frequency tomography. *Nature GEO* 1, 458 – 462.
- Simpson, D. P., J. B. Illian, F. Lindgren, S. H. Sørbye, and H. Rue (2011). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Preprint statistics no.10/2011, Department of Mathematical Sciences, Norwegian University of Science and Technology.
- Simpson, D. P., F. Lindgren, and H. Rue (2011). In order to make spatial statistics computationally feasible, we need to forget about the covariance function. *Environmetrics* 23(1), 65 – 74. Special Issue: Spatio-Temporal Stochastic Modelling.
- Simpson, D. P., F. Lindgren, and H. Rue (2012). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics* 1, 16 – 29.
- Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B* 64(4), 583 – 639.
- Stähler, S. and K. Sigloch (2014). Full probabilistic seismic source inversion Part 1: Efficient parameterisation. *Solid Earth* 5(2), 1055 – 1069.
- Stein, M. L. (1999). *Interpolation of Spatial Data*. New York: Springer-Verlag.
- Tian, Y., K. Sigloch, and G. Nolet (2009). Multiple-frequency SH-wave tomography of the western US upper mantle. *Geophysical Journal International* 178(3), 1384 – 1402.
- Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association* 81(393), 82 – 86.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika* 41, 434 – 449.
- Whittle, P. (1963). Stochastic processes in several dimensions. *Bull. Inst. Internat. Statist.* 40, 974 – 994.
- Zhang, R., C. Czado, and K. Sigloch (2013). A Bayesian linear model for the high-dimensional inverse problem of seismic tomography. *Annals of Applied Statistics* 7(2), 1111 – 1138.