Math Geosci
DOI 10.1007/s11004-014-9567-5

A General Probabilistic Approach for Inference of Gaussian Model Parameters from Noisy Data of Point and Volume Support

Thomas Mejer Hansen · Knud Skou Cordua · Klaus Mosegaard

Received: 15 January 2014 / Accepted: 9 October 2014

Abstract Methods that rely on Gaussian statistics require a choice of a mean and covariance to describe a Gaussian probability distribution. This is the case using for example kriging, sequential Gaussian simulation, least-squares collocation, and least-squares-based inversion, to name a few examples. Here, an approach is presented that provides a general description of a likelihood function that describes the probability that a set of, possibly noisy, data of both point and/or volume support is a realization from a Gaussian probability distribution with a specific set of Gaussian model parameters. Using this likelihood function, the problem of inferring the parameters of a Gaussian model is posed as a non-linear inverse problem using a general probabilistic formulation. The solution to the inverse problem is then the a posteriori probability distribution over the parameters describing a Gaussian model, from which a sample can be obtained using, e.g., the extended Metropolis algorithm. This approach allows detailed uncertainty and resolution analysis of the Gaussian model parameters. The method is tested on noisy data of both point and volume support, mimicking data from remote sensing and cross-hole tomography.

Keywords Covariance parameters · Linear inversion · Kriging · Collocation

1 Introduction

Consider N data $\mathbf{d} = [d_1, d_2, \ldots, d_N]^t$ that are linearly related with M model parameters $\mathbf{m} = [m_1, m_2, \ldots, m_M]^t$ through the linear operator $\mathbf{G}$

$$\mathbf{d} = \mathbf{G}\mathbf{m}. \tag{1}$$

T. M. Hansen (✉) · K. S. Cordua · K. Mosegaard
Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
e-mail: tmeha@nbi.dk

Published online: 12 November 2014

Further, assume that $\mathbf{m}$ is a realization of a multivariate Gaussian model described by a mean $\mathbf{m}_0$ and a covariance $\mathbf{C}_\mathrm{m}$, $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$. The data $\mathbf{d}$ can be seen as linear weighed averages of the model parameters $\mathbf{m}$, where each row g_i in $\mathbf{G}$ defines the weight of each model parameter associated to the corresponding data d_i. The model parameters $\mathbf{m}$ typically, for geophysical related problems, represent some physical property that one is interested in, and the data $\mathbf{d}$ represent some geophysical observations.

The area or volume over which a value is measured/integrated is sometimes referred to as the support of the value. Model parameters $\mathbf{m}$ most often represent values related to a point, i.e., values of point support. Observed data $\mathbf{d}$ often represent values integrated over some volume, which we refer to as data of volume support or linear average data. In the simplest case, data $\mathbf{d}$ are direct observations of model parameters $\mathbf{m}$ of point support. Hence, there is no averaging in the forward operator $\mathbf{G}$. For this case, when the number of data and model parameters is the same ($N = M$), $\mathbf{G}$ is then simply an identity matrix.

Equation (1) describes the forward relation between model parameters and data for a number of research fields such as linear inverse theory (Tarantola and Valette 1982a), 2-point geostatistics (Journel and Huijbregts 1978), and least-squares collocation (Krarup 1969; Herzfeld 1992), to name a few. The main focus is typically to infer information about $\mathbf{m}$, while $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ is assumed to be known a priori (independent of the collected data). Here, we consider how to infer information about $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$, given $\mathbf{d}$, $\mathbf{G}$ and a Gaussian model describing the noise on data, $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$. Different properties of this problem have been dealt with in a number of studies for data of both point and volume support.

1.1 Inference of Properties of $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ from Data of Point Support

Inference of properties of $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ from data of point support has been analyzed in detail in the geostatistical community since the 60s, see, e.g., Journel and Huijbregts (1978), and many methods exists for inferring information about $\mathbf{C}_\mathrm{m}$. Most methods assume that the Gaussian model $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ can be described by a set of parameters, which we will refer to as Gaussian model parameters. The Gaussian model parameters describing the covariance model $\mathbf{C}_\mathrm{m}$ are referred to as covariance model parameters. Some methods rely on computing an experimental covariance model from observed data, followed by inference of covariance model parameters that fits the experimental model based one some criteria, such as for example weighed least-squares (Cressie 1985; Knudsen 1987; Pardo-Igúzquiza 1999; Lark and Papritz 2003; Arabelos and Tscherning 2003; Emery 2010; Desassis and Renard 2013). Other methods rely on maximum-likelihood estimation of these covariance model parameters (Kitanidis and Lane 1985; Warnes and Ripley 1987; Pardo-Igúzquiza 1997, 1998; Pardo-Igúzquiza and Dowd 2005; Jarmołowski and Bakuła 2014).

1.2 Inference of the Covaraince Model Parameters from Data of Volume Support

In reality, most measurements provide data of volume support, even if the volume may be small. At the smallest scale, core plugs provide average measurements of properties on centimeter scale, while seismic data provide average measurements on

meter scale (Frykman and Deutsch 2002). Such information of volume support can be integrated into a Gaussian model of point support, if the averaging of the parameter is linear, using estimation and simulation techniques (Hansen et al. 2006; Hansen and Mosegaard 2008; Liu and Journel 2009). Likewise, block kriging can be used to integrate such information into a model of volume support (Isaaks and Srivastava 1989). The general formulation of kriging, see, e.g., Journel and Huijbregts (1978), allows seamless integration of Gaussian distributed data at any support, which is utilized by Goovaerts (2010) and Atkinson (2013). In any of these cases, a covariance model describing the spatial variability at point support, $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$, is needed.

A number of scaling laws have been suggested that allow scaling of covariance model parameters inferred at one volume support to covariance model parameters inferred at another volume support (Kupfersberger et al. 1998; Frykman and Deutsch 1999, 2002). In principle, this allows to estimate covariance model parameters related to data of point support, given covariance model parameters inferred from data of volume support, using the methods described above. The scaling laws require that each observed data has the same volume of support, and that the volumes of support are non-overlapping. This makes it applicable to some data related to remote sensing, but not to, for example, linear inverse problems in general, where the size of volume of support may differ for each data and where volume of support is often overlapping.

Kyriakidis (2004) discusses how the covariance between two data of volume support can be obtained from convolution of the covariance between points. He then suggests to infer the covariance between points, from the covariance between data of volume support, using deconvolution. This idea was further developed by Goovaerts (2008) who considered arbitrarily complex voxel/volume shapes.

Other approaches have been developed in the field of inverse problem theory. Asli et al. (2000) suggest to compute the covariance model of the data at volume support, $\mathbf{C}_v^{\text{theory}}$, using the linear forward operator, $\mathbf{G}$, and an assumed a priori covariance model, $\mathbf{C}_\mathrm{m}$. This is compared to the experimental covariance obtained from the observed data of volume support, $\mathbf{C}_v^{exp}$ using the so-called V–V plot. The optimal choice of $\mathbf{C}_\mathrm{m}$ is selected as the one for which $\mathbf{C}_v^{exp}$ best matches $\mathbf{C}_v^{theory}$. This method has since been applied by, e.g., Gloaguen et al. (2005, 2007), Giroux et al. (2007), and Shamsipour et al. (2010). Shamsipour et al. (2010) describe in more detail how the V–V plot works, and demonstrate the performance of the method through an application to a synthetic data set of point support.

Hansen et al. (2008) suggest an approach to infer properties of the prior model related to an inverse problem. They compute a probabilistic measure of the overlap between the a priori probability density and the a posteriori probability density, and argue that the prior model with most overlap provides the prior model most consistent with observed data. Looms et al. (2010) further developed this idea and demonstrate the method by inferring the covariance model parameters of a Gaussian model, for both synthetic and real data related to tomographic cross-borehole ground-penetrating radar (GPR) data.

Kelsall and Wakefield (2002) propose a probabilistic approach to infer the mean, variance and range of an isotropic cubic covariance model of point support, given data of volume support.

Hansen et al. (2013a, b) propose a parameterization of Gaussian non-linear inverse problems where Gaussian covariance model parameters are treated as model parameters as opposed to being treated as a priori chosen parameters. This allows inferring information about the properties of the Gaussian model, as well as the model parameters themselves, directly as part of a stochastic inversion.

Here, we present a general approach to compute the likelihood that a set of possibly noisy data, of both point and volume support (i.e., linear average data), as in Eq. (1), is a realization from a Gaussian probability distribution with specific Gaussian model parameters. The problem of inferring information about $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ is formulated as a probabilistic inverse problem, which allows a priori information on the covariance model parameters, and a Gaussian model describing uncertainty of the observed data. The solution to the inverse problem is the a posteriori probability density over the covariance model parameters describing the properties of the Gaussian model. A sample of this posterior probability density, obtained using, e.g., the rejection or Metropolis sampler, allows full-resolution analysis of the inferred covariance model parameters. The method differs from the one proposed in Hansen et al. (2013a, b) in that only information about the Gaussian model parameters are inferred, but not information about the model parameters themselves. Initially, we lay out the theory of the method using notation from probabilistic inverse theory.

2 Theory

Consider a physical system that can be represented by some model parameters $\mathbf{m}$, that are related to a number of data $\mathbf{d}$, through the linear mapping operator $\mathbf{G}$, as in Eq. (1). Further, assume that $\mathbf{m}$ is a realization from a Gaussian probability distribution $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ with the corresponding Gaussian probability density

$$\rho_\mathrm{M}(\mathbf{m}) = ((2\pi)^M \ |\mathbf{C}_\mathrm{m}|)^{-0.5} \ \exp\left(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_0)^t \ \mathbf{C}_\mathrm{m}^{-1} \ (\mathbf{m}-\mathbf{m}_0)\right). \tag{2}$$

As $\mathbf{d}$ is linearly related to $\mathbf{m}$ through the linear operator $\mathbf{G}$, the multivariate Gaussian distribution describing the spatial distribution of model parameters $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ can be transformed into a Gaussian distribution in the data space $\mathcal{N}(\mathbf{v}_0, \mathbf{C}_\mathrm{v})$ by (see, e.g., Georgii 2008; Kay 2006)

$$\mathbf{v}_0 = \mathbf{G}\mathbf{m}_0 \tag{3}$$

$$\mathbf{C}_\mathrm{v} = \mathbf{G}\mathbf{C}_\mathrm{m}\mathbf{G}', \tag{4}$$

where $\mathcal{N}(\mathbf{v}_0, \mathbf{C}_\mathrm{v})$ describes the spatial distribution of noise-free data given $\mathbf{G}$ and $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$. $\mathbf{C}_\mathrm{v}$ is equivalent to the so-called *data-to-data* covariance used as part of the kriging system for solving kriging problems, see e.g. Journel and Huijbregts (1978), Hansen and Mosegaard (2008).

Data are typically contaminated with noise. Here noise is considered that can be described by a multivariate Gaussian distribution, $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$. Often, the noise is considered unbiased and $\mathbf{n}_0$ is set to zero. In addition, off-diagonal elements of $\mathbf{C}_\mathrm{n}$ are

often set to zero, implying that only uncorrelated data errors are considered. In reality, correlated data errors may often exist, with a significant amplitude, due to for example model imperfections, and we therefore keep the general description of the Gaussian noise model given above, see, e.g., Cordua et al. (2008), Hansen et al. (2014).

The corresponding complete Gaussian description of the data $\mathbf{d}$ can now be obtained by combining $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$ and $\mathcal{N}(\mathbf{v}_0, \mathbf{C}_\mathrm{v})$ through addition of the mean and the covariances as $\mathcal{N}(\mathbf{f}_0, \mathbf{C}_\mathrm{f}) = \mathcal{N}(\mathbf{v}_0 + \mathbf{n}_0, \mathbf{C}_\mathrm{v} + \mathbf{C}_\mathrm{n})$, see, e.g., Hansen et al. (2014). In other words, for a linear Gaussian forward problem where data $\mathbf{d}$ are linearly related to $\mathbf{m}$ through the linear operator $\mathbf{G}$, and the model parameters can be described by $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$, and the noise model on the data by $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$, a complete statistical description of data $\mathbf{d}$ can be given by the Gaussian distribution $\mathcal{N}(\mathbf{f}_0, \mathbf{C}_\mathrm{f})$ where

$$\mathbf{f}_0 = \mathbf{v}_0 + \mathbf{n}_0 = \mathbf{G}\mathbf{m}_0 + \mathbf{n}_0 \tag{5}$$

$$\mathbf{C}_\mathrm{f} = \mathbf{C}_\mathrm{v} + \mathbf{C}_\mathrm{n} = \mathbf{G}\mathbf{C}_\mathrm{m}\mathbf{G}' + \mathbf{C}_\mathrm{n} \tag{6}$$

such that the corresponding probability density, $f_f(\mathbf{d})$, takes the form

$$f_f(\mathbf{d}) = k_f \exp\left(-\frac{1}{2}(\mathbf{d} - (\mathbf{G}\mathbf{m}_0 + \mathbf{n}_0))^t \, (\mathbf{G}\,\mathbf{C}_\mathrm{m}\,\mathbf{G}' + \mathbf{C}_\mathrm{n})^{-1} \, (\mathbf{d} - S(\mathbf{G}\mathbf{m}_0 + \mathbf{n}_0))\right) \tag{7}$$

where

$$k_f = ((2\pi)^N \left|\mathbf{G}\,\mathbf{C}_\mathrm{m}\,\mathbf{G}' + \mathbf{C}_\mathrm{n}\right|)^{-0.5}.$$

In case measurement uncertainty is disregarded, and N observed data are direct observations of the $\mathrm{M} = N$ model parameters (in which case the forward operator $\mathbf{G}$ is an identity matrix such that $\mathbf{d} = \mathbf{m}$) Eq. (7) reduces to

$$f_f(\mathbf{d}) = ((2\pi)^N \, |\mathbf{C}_\mathrm{m}|)^{-0.5} \exp\left(-\frac{1}{2}(\mathbf{d} - \mathbf{m}_0)^t \, \mathbf{C}_\mathrm{m}^{-1} \, (\mathbf{d} - \mathbf{m}_0)\right) \tag{8}$$

which is equivalent to the likelihood function defined in Eq. (1) in Kitanidis and Lane (1985). Equation (7) can, therefore, be seen as a generalization of likelihood given by Kitanidis and Lane (1985) that allows accounting for data of volume support, and measurement uncertainties through a Gaussian noise model $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$.

2.1 Parameterizing $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$

Our focus is to infer information about $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$, given observed data $\mathbf{d}_\mathrm{obs}$, knowledge about the linear averaging through $\mathbf{G}$, and measurement uncertainties through $\mathcal{N}(\mathbf{n}_0, \mathbf{C}_\mathrm{n})$. To do this, $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ is parameterized using a set of parameters that defines a collection of stationary anisotropic multivariate Gaussian models. A parametric description of a covariance model is described by a linear combination of the so-called Nugget type covariance model and a covariance model of type T: Gaussian

(T = Gau), Exponential (T = Exp) or Spherical (T = Sph). The variance is referred to as σ^2. The nugget fraction, α, defines the fraction of the total variance due to the Nugget model. The isotropic covariance between two model parameters m_i and m_j, separated by distance h is given by

$$C(h) = \alpha\, \sigma^2 \mathrm{Nug}(h) \;+\; (1-\alpha)\, \sigma^2\, T(h,r) \tag{9}$$

where

$$\mathrm{Nug}(h) = \begin{cases} 1 \text{ if } h = 0 \\ 0 \text{ otherwise} \end{cases}$$

$$\mathrm{Sph}(h,r) = \begin{cases} 1 - 1.5\left(\frac{h}{r}\right) + 0.5\left(\frac{h}{r}\right)^3 \text{ if } h \le r \\ 0 \quad \text{otherwise} \end{cases}$$

$$\mathrm{Exp}(h,r) = \exp\frac{-3h}{r}$$

$$\mathrm{Gau}(h,r) = \exp\frac{-3h^2}{r^2}.$$

The isotropic formulation of Eq. (9) can be generalized to a three-dimensional anisotropic covariance model using scaling (up to three range parameters, r_1, r_2, r_3) and rotation (up to three angles of rotation $\theta_1, \theta_2, \theta_3$) of the coordinate system used to compute the distance between points. See, e.g., Goovaerts (1997) chapter 4, and Remy et al. (2008), page 26.

A specific choice of a set of parameters describing the Gaussian model $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ is defined by the vector Ψ, which we refer to as Gaussian model parameters. A one-dimensional Gaussian model with an isotropic covariance model is completely defined by $\Psi_{1d} = [\mathbf{m}_0, T, \alpha, \sigma^2, r_1]$, a two-dimensional Gaussian model with an anisotropic covariance model by $\Psi_{2d} = [\mathbf{m}_0, T, \alpha, \sigma^2, r_1, r_2, \theta_1]$ and a Gaussian model with a three-dimensional anisotropic covariance model by $\Psi_{3d} = [\mathbf{m}_0, T, \alpha, \sigma^2, r_1, r_2, r_3, \theta_1, \theta_2, \theta_3]$. In the following, $\mathbf{C}_\psi$ will refer to the model-covariance matrix computed based on all the covariance model parameters Ψ, except $\mathbf{m}_0$. This allows formulating a likelihood function $L(\mathbf{d})$, which provides the likelihood that the data $\mathbf{d}$ are a realization of the Gaussian probability distribution $f_f(\mathbf{d})$ as given in Eq. (7), conditional to the choice of Gaussian model parameters Ψ describing the distribution of the model parameters, as

$$L(\Psi) \;=\; k_f \exp\Bigg(-\frac{1}{2}(\mathbf{d}_\mathrm{obs} - (\mathbf{G}\mathbf{m}_0 + \mathbf{n}_0))^t \;(\mathbf{G}\, \mathbf{C}_\psi\, \mathbf{G}' + \mathbf{C}_\mathrm{n})^{-1} \;(\mathbf{d}_{obs} - (\mathbf{G}\mathbf{m}_0 + \mathbf{n}_0))\Bigg). \tag{10}$$

Equation (10) can be used seamlessly with any of the methods previously proposed for maximum-likelihood estimation of the parameters Ψ defining $\mathcal{N}(\mathbf{m}_0, \mathbf{C}_\mathrm{m})$ (Kitanidis and Lane 1985; Warnes and Ripley 1987; Pardo-Igúzquiza 1997, 1998; Pardo-Igúzquiza and Dowd 2005). Note though that Warnes and Ripley (1987) demonstrate

that the likelihood function, in case of noise-free data of point support, can be multimodal, and that the global maximum-likelihood may not correspond to a reasonable value of the covariance model parameters. Therefore, resolution analysis of the covariance model parameters is necessary to obtain a sufficient characterization of the model parameters.

Here, we will pose the problem of inference of Ψ as a probabilistic inverse problem. This allows the use of prior information of Ψ and full non-linear uncertainty and resolution analysis of the covariance properties, through analysis of the a posteriori distribution of Ψ.

2.2 Probabilistic Inference of Ψ

The problem of inferring information about a set of parameters, Ψ, that describe a Gaussian model of spatial variability of some model parameters, $\mathbf{m}$, given some observed data, $\mathbf{d}_{\text{obs}}$, related to the model parameters by the linear forward model defined in Eq. (1) can now be posed as a probabilistic inverse problem. Given the likelihood, $L(\Psi)$, and a probability density describing a priori information about the Gaussian model parameters $\rho(\Psi)$, the a posteriori probability distribution is given by

$$\sigma(\Psi) = k\ \rho(\Psi)\ L(\Psi) \tag{11}$$

where k is a normalization constant. $\sigma(\Psi)$ describes the combined information obtained from the data, $\mathbf{d}_{\text{obs}}$ and the independent prior information about Ψ (Tarantola and Valette 1982b; Tarantola 2005). In general, it is not possible to obtain an analytical description of $\sigma(\Psi)$. For small-scale problems (considering for example two parameters of Ψ) one can choose to exhaustively evaluate the a posteriori probability in a densely sampled grid. In such a case, one has to choose a sampling density so dense that no local minimum is excluded. As a general approach, appropriate for higher dimensional problems, we suggest to use sampling methods such as the rejection sampler, the Gibbs sampler, or the Metropolis algorithm to generate a representative sample from the a posteriori probability density. The solution to the inverse problem will then be a collection of sets of Gaussian model parameters distributed according to $\sigma(\Psi)$ (i.e., a sample from $\sigma(\Psi)$). Such a sample allows detailed resolution analysis of the model parameters Ψ (Mosegaard 1998).

3 Examples

To illustrate the methodology, we will consider a realization from a two-dimensional anisotropic Gaussian model Ψ_{ref}, with Gaussian model parameters given as T = 'Sph', $r_1 = 6$ m, $r_2 = 2$ m, $\alpha = 0$, $\theta_1 = 110$, $\sigma^2 = 0.0003\,\text{m}^2/\text{ns}^2$, and $m_0 = 0.14$ m/ns (i.e., $\mathbf{m}_0$ is constant), in a 7×13 m grid, with grid notes for every 0.25 m (1,537 grid notes). This realization will be used as reference model (see Fig. 1a). The Gaussian model parameters chosen to generate this realization is chosen, as an example, to reflect the speed of electromagnetic waves (i.e., GPR signals) in the Vadose zone, for a mix of unconsolidated clay and sand sediments, as in Hansen et al. (2013b).

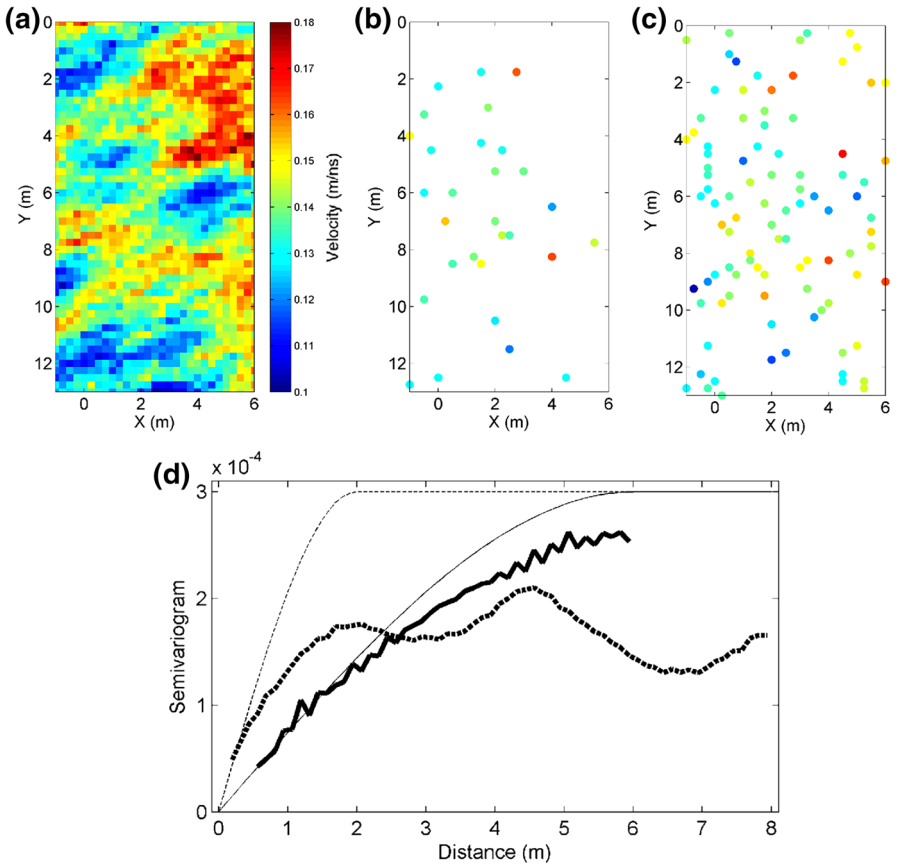

Fig. 1 **a** Reference model, **b** $\mathbf{d}_{p30}$, 30 randomly selected data of point support, **c** $\mathbf{d}_{p100}$, 100 randomly selected data of point support. A realization of Gaussian *white noise* with $\sigma = 0.01$ has been added to $\mathbf{d}_{p30}$ and $\mathbf{d}_{p100}$. **d** Experimental semivariogram model for the reference model along 110 (*thick solid*) and 20 (*thick dashed*) degrees from North, compared to the semivariogram defined by the chosen covariance model parameters (*thin lines*)

Figure 1d shows the experimental semivariogram $\gamma(h)$ (linked to the covariance model through $\gamma(h) = C(0) - C(h)$) obtained from the reference model along the direction of maximum ($\theta = 110°$) and minimum ($\theta = 110 - 90 = 20°$) correlation (thick lines) and compares these to the semivariogram defined by the chosen covariance model parameters along the same directions (thin lines). The variance of the reference model (i.e., the sill) is 0.00019 $\mathrm{m}^2/\mathrm{ns}^2$, which is somewhat smaller the reference value of 0.0003 $\mathrm{m}^2/\mathrm{ns}^2$, and the apparent range along direction $\theta = 110°$ seems to be slightly higher from the reference model, than the reference value of $r_1 = 6$ m. The differences between the experimental and synthetic semivariogram models are caused by ergodic variations due to the use of a finite size model parameter space.

A number of different types of data sets are considered obtained as direct observations (i.e., point support) and linear averages of from the 1,537 points in the reference velocity model. All of the data sets defined below, $\mathbf{d}_{p30}$, $\mathbf{d}_{p100}$, $\mathbf{d}_{2\times2}$, $\mathbf{d}_{6\times6}$, $\mathbf{d}_{\mathrm{tomo30}}$, $\mathbf{d}_{\mathrm{tomo100}}$, and $\mathbf{d}_{\mathrm{tomo702}}$ have been contaminated with noise to resemble measurement

uncertainty. This noise is generated as a realization of uncorrelated zero mean Gaussian noise with a standard deviation of 0.01 m/ns.

Reference data of point support $\mathbf{d}_{p30}$ and $\mathbf{d}_{p100}$ reflect two data sets obtained by selecting a subset of 30 and 100 randomly selected points as data, out of the 1,537 available data of points (Figs. 1b, c).

Reference data of volume support Two types of data with volume support are considered: simple average pixel values of images (related to, e.g., remote sensing) and more complex linear averages related to cross-hole tomography.

Some satellite images record the average property over an area, i.e., data of volume support as defined by the pixel size of the image observed by the satellite. If the relation between the data at volume and point support is linear, this relation can be expressed by Eq. (1). Figure 2a, b shows two data sets with different spatial resolution, generated as simple linear averages in sets of 2×2 and 6×6 pixels from the reference model in Fig. 1a.

The data sets $\mathbf{d}_{2\times2}$ and $\mathbf{d}_{6\times6}$ can be seen as simple upscaled (using linear averaging) versions of the reference model (i.e., reference image) consisting of 1,537 data of point support. Each linear average data is measured over the same support volume, and there is no overlap between the support of each linear average data. The averaging kernel related to some geophysical data can be much more complex, see, e.g., Fig. 1 in Hansen et al. (2014). Figure 3b illustrates the layout of a measurement geometry for a typical cross-hole tomographic experiment. The travel time delay for an electromagnetic wave initiated at different depths in a left borehole (at $x = 0$ m) is recorded at different depths in a right borehole (at $x = 5$ m). The red lines on Fig. 3 denote lines between such sources and receivers.

The calculated travel time is influenced by the choice of averaging kernel, $\mathbf{G}$, in Eq. (1). Here, we make use of the so-called linear fat ray theory Jensen et al. (2000),

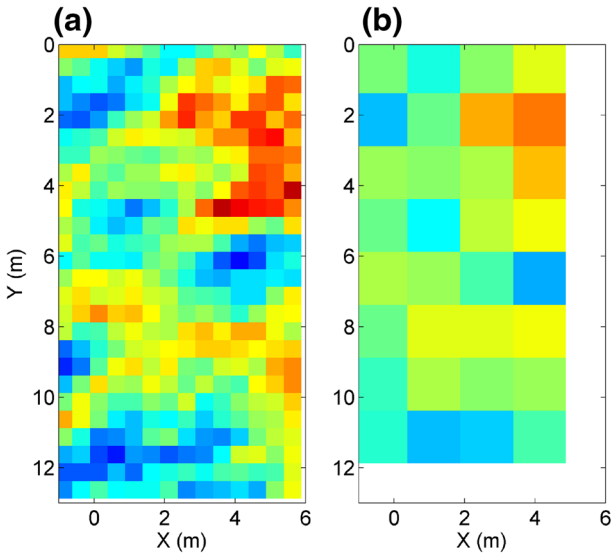

Fig. 2 Linear average data (remote sensing) **a** $\mathbf{d}_{2\times2}$, and **b** $\mathbf{d}_{6\times6}$

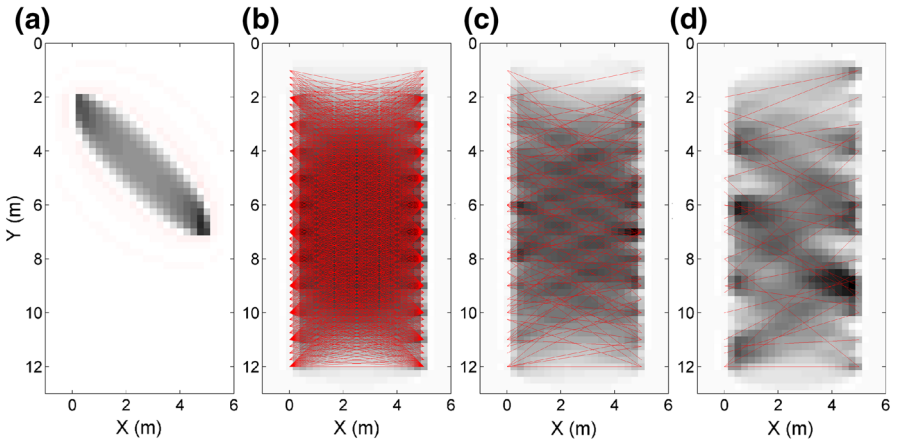

Fig. 3 Sensitivity kernel for **a** one, **b** 702, **c** 100, and **d** 30 randomly chosen set(s) of sources and receivers. The *red lines* connect sets of used sources and receivers

which means that the travel time is sensitive to the velocity field in a zone (the first Fresnel zone) around the straight line connecting the source and the receiver. Such a sensitivity kernel is shown in Fig. 3a for one row in $\mathbf{G}$ for a source located at ($x = 0$ m, $y = 2$ m) and a receiver located at ($x = 5$ m, $y = 7$ m).

The travel time can be translated into a measure of average velocity within the averaging kernel defined by the rows in $\mathbf{G}$. For more easy comparison with the other types of data that are considered here, we choose to use average velocities as observed data. Figure 3c, d shows 100 and 30 randomly selected ray paths from which linear average kernels have been computed such that two data sets of 100 and 30 average velocities, respectively, have been computed as $\mathbf{d}_{\text{tomo100}}$ and $\mathbf{d}_{\text{tomo30}}$.

3.1 Inference of α From Noisy Data of Point Support

First, the data sets of point support, $\mathbf{d}_{p30}$ and $\mathbf{d}_{p100}$ are considered. Initially, only the nugget fraction α is considered unknown, while all other parameters of Ψ are assumed known. The a priori distribution of the nugget fraction, $\rho(\alpha)$, is chosen as a uniform distribution between 0 and 1, which means that the a posteriori probability density is equivalent to the likelihood function in this interval. This also means that the prior probability that the nugget fraction is below 0.1 is given by $P_\rho(\alpha < 0.1) = 0.1$.

Initially, we will disregard the fact that data are contaminated with noise and incorrectly treat data as noise free. The corresponding a posteriori probability densities are shown on Fig. 4 as red curves for $\sigma(\alpha|\mathbf{d}_{p30})$ (thin dashed line) $\sigma(\alpha|\mathbf{d}_{p100})$ (thicker dashed line). Considering the small dataset $\mathbf{d}_{p30}$, it seems that one can conclude with confidence that the nugget fraction is above 0.1. For $\mathbf{d}_{p100}$ it seems that it can be concluded that the nugget fraction is between 0.13 and 0.7. In fact, we can evaluate $P(\alpha < 0.1 \mid \mathbf{d}_{p30}) = 0.005$, and $P(\alpha < 0.1 \mid \mathbf{d}_{p100}) = 0.0002$. This means that there is close to zero probability that the nugget fraction is below 0.1, which obviously leads to wrong conclusions. Recall that the actual value of the

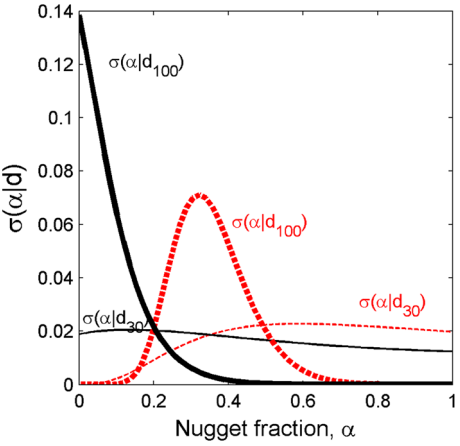

Fig. 4 A posteriori probability $\sigma(\alpha|\mathbf{d})$ in case accounting for (*black lines*) and disregarding (*red dotted lines*) measurement uncertainty, given $\mathbf{d}_{p30}$ (*thin lines*) and $\mathbf{d}_{p100}$ (*thick lines*)

nugget fraction is $\alpha = 0$. Thus, not only does ignoring the noise provide a biased result, it also provides an inversion result inconsistent with the actual known reference value.

A similar observation, that noise on data will lead to an apparent nugget component in an experimental semivariogram/covariance analysis, has been made by, e.g., Goovaerts (1997, page 31) and Chiles and Delfiner (2012, page 52). However, as we have just demonstrated, such an apparent nugget is an artifact, not reflecting the actual (true) underlying random field. One cannot model uncertainty of data by assuming a nugget model. Realizations from such a random field will exhibit much more small-scale variability than really exist, which could lead to wrong interpretations and conclusions. The reason an apparent nugget appears when computing an experimental semivariogram from noisy data is that classical experimental semivariogram analysis cannot account for known noise on data. The methodology presented here specifically accounts for Gaussian distributed noise on data.

If noise is accounted for in the likelihood as defined in Eq. (10), the corresponding a posteriori probability densities looks different. This result is shown in Fig. 4 as black curves for $\sigma(\alpha|\mathbf{d}_{p30})$ (thin line) $\sigma(\alpha|\mathbf{d}_{p100})$ (thicker line). Considering the small data set, the posterior probability of having a nugget fraction of less than 0.1 is $P(\alpha < 0.1 \mid \mathbf{d}_{p30}) = 0.14$, which is slightly higher than the prior probability. When considering the larger data set this probability becomes much higher , $P(\alpha < 0.1 \mid \mathbf{d}_{p100}) = 0.7$. Thus, by correctly accounting for the data noise, the inversion results demonstrate that a small nugget fraction has high probability, which is consistent with the (true) reference model.

This simple example clearly demonstrates the importance of being able to account for measurement uncertainty when it is present. If not accounted for, one may infer a significant nugget fraction that may not be real. Determination of a significant nugget fraction based on such an incorrect covariance model can have serious interpretation implications.

3.2 Inference of Ψ Through Sampling the a Posteriori Probability Density

In the following, we will infer information about the covariance model parameters defined by Ψ ($\mathbf{m}_0$ is considered known in all cases) by sampling the a posteriori probability density $\sigma(\Psi|\mathbf{d}_{\text{obs}})$ [Eq. (11)] using the extended Metropolis algorithm (Mosegaard and Tarantola 1995). In practice, we have implemented the methodology as an extension of the SIPPI software package (Hansen et al. 2013a).

The extended Metropolis algorithm is a Markov Chain Monte Carlo method that allows efficient sampling the a posteriori distribution for non-linear inverse problem with, in principle, complex a priori information, (Mosegaard and Tarantola 1995). This algorithm differs from the Metropolis–Hasting algorithm (Hastings 1970) in that the a priori probability does not need to be computed. It is sufficient that a (possibly black box) algorithm exists that can perform a random walk according to the a priori probability distribution. The a posteriori probability distribution can then be sampled by performing a random walk in the space of a priori acceptable models, where each move from one realizations of the prior, Ψ_{current}, to a new realization of the prior, Ψ_{proposed} , is accepted with probability $P_{\text{acc}} = \min[1 \frac{L(\Psi_{\text{proposed}})}{L(\Psi_{\text{current}})}]$. Details on the extended Metropolis algorithm can be found in Mosegaard and Tarantola (1995) and Hansen et al. (2013a).

The a priori information for each parameter in Ψ is chosen to be independent such that the full a priori distribution can be characterized by a set of independent one-dimensional marginal a priori probability distributions. The a priori distribution describing both ranges, r_1 and r_2, is a uniform distribution between 0 and 10 m. The prior distribution of the nugget fraction is a uniform distribution between 0 and 1. The prior distribution of the variance is a uniform distribution between 0.0001 and 0.0003. The prior distribution of the rotation angle, θ_1, is described by a generalized Gaussian distribution with mean of 90°, standard deviation of 30° and norm of 30.

In the following, we apply the extended Metropolis algorithm to sample the a posteriori probability distribution over the covariance model parameters given different types of observed data.

3.2.1 Inference of $\sigma(r_1, r_2 \mid \mathbf{d}_{\text{obs}})$

Initially, we consider all properties of Ψ known, except for the ranges, r_1 and r_2. Figure 5a–f shows the 10 % (light gray), 50 % (gray), and 90 % (black) highest posterior density region (Box and Tiao 1973) for the corresponding two-dimensional a posteriori probability distributions $\sigma(r_1, r_2|\mathbf{d}_{p100})$, $\sigma(r_1, r_2|\mathbf{d}_{p30})$, $\sigma(r_1, r_2|\mathbf{d}_{2\times2})$, $\sigma(r_1, r_2|\mathbf{d}_{6\times6})$, $\sigma(r_1, r_2|\mathbf{d}_{\text{tomo30}})$, and $\sigma(r_1, r_2|\mathbf{d}_{\text{tomo100}})$, for the six considered data sets, $\mathbf{d}_{p100}$, $\mathbf{d}_{p30}$, $\mathbf{d}_{2\times2}$, $\mathbf{d}_{6\times6}$, $\mathbf{d}_{\text{tomo100}}$, and $\mathbf{d}_{\text{tomo30}}$.

Figure 5d illustrates that the use of the data $\mathbf{d}_{6\times6}$ is the choice of data that results in the least resolved r_1 and r_2. An area of relatively high posterior probability is visible around, $r_1 = 4$ and $r_2 = 1$, which does not coincide with the actual known reference values of $r_1 = 6$ and $r_2 = 2$. At the same time, any value of r_1 and r_2, except $r_1 = 0$ and $r_2 = 0$, seems to be associated with significant a posteriori probability. This, in

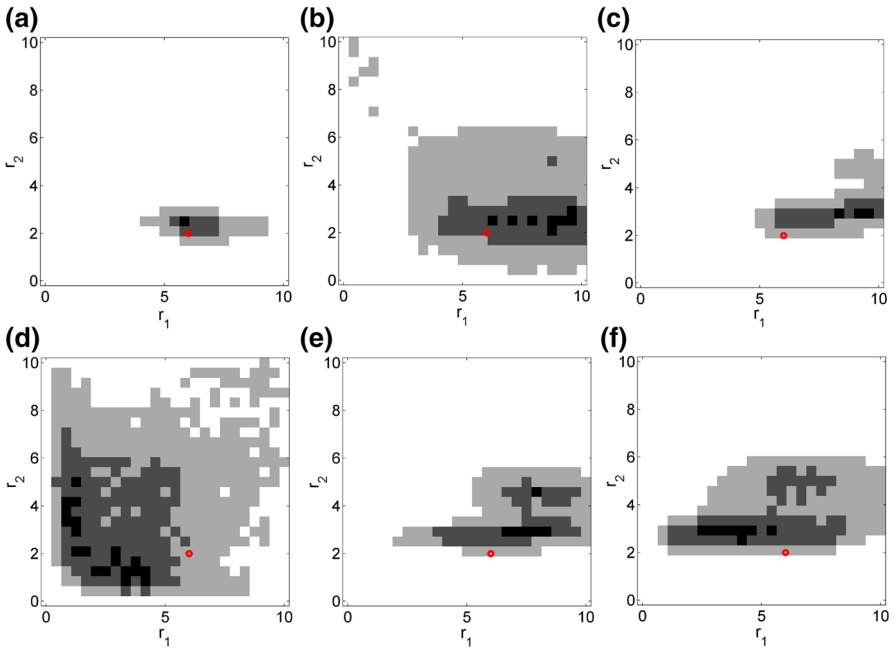

Fig. 5 Two-dimensional a posteriori probability density $\sigma(r_1, r_2|\mathbf{d})$ obtained from **a** $\mathbf{d}_{p100}$, **b** $\mathbf{d}_{p30}$, **c** $\mathbf{d}_{2\times2}$, **d** $\mathbf{d}_{6\times6}$, **e** $\mathbf{d}_{\text{tomo100}}$, and **f** $\mathbf{d}_{\text{tomo30}}$. *Red circle* indicates the value of the reference parameters ($r_1 = 6$ m, $r_2 = 2$ m). *Gray scales* indicates 10 % (*light gray*), 50 % (*gray*), and 90 % (*black*) highest posterior density regions

effect, suggests that using $\mathbf{d}_{6\times6}$ as data, r_1 and r_2 are close to unresolved, and that little to no information can be inferred from the data about r_1 and r_2.

On the other hand, using $\mathbf{d}_{p100}$ provides an a posteriori probability $\sigma(r_1, r_2|\mathbf{d}_{p100})$ (Fig. 5a) that resolves r_1 and r_2 very well. From the two-dimensional marginal posterior distribution one can see that the posterior uncertainty about the horizontal range r_1 is larger than the uncertainty related to the vertical range r_2. Figure 5a illustrates that the area of significant posterior probability is related to one coherent area centered around the reference values.

Considering the other data sets, $\mathbf{d}_{p30}$, $\mathbf{d}_{2\times2}$, $\mathbf{d}_{\text{tomo30}}$, and $\mathbf{d}_{\text{tomo100}}$, the resulting posterior distributions can all be characterized by bimodal distributions. That is, the area of high probability is split between two separate coherent areas.

3.2.2 Inference of $\sigma(r_1, r_2, \theta_1, \sigma^2, \alpha \mid \mathbf{d}_{obs})$

All properties of $\Psi = [r_1, r_2, \theta_1, \alpha, \sigma^2]$ are now considered to be unknown except for the type of covariance model (known as $T = \text{Sph}$ in the reference model) and the background velocity (known to be 0.14 m/ns). Then, the extended Metropolis algorithm is used to generate a large sample from the a posteriori probabilities, $\sigma(\Psi\ |\mathbf{d}_{p100})$, $\sigma(\Psi\ |\mathbf{d}_{6\times6})$, and $\sigma(\Psi\ |\mathbf{d}_{\text{tomo100}})$. Corresponding two-dimensional marginal a posteriori distributions is shown in Figs. 6, 7, 8.

As for the previous case, the data $\mathbf{d}_{p100}$ leads to the best resolved parameters of Ψ, Fig. 6. Most of the accepted realizations of Ψ is located in an area close to the reference values (denoted by a red dot). There is though a number of models located quite far away from the references value with significant posterior probability around $r_1 = 2$ m, $r_2 = 5$ m, and $\theta_1 = 50°$. Note how r_1 has higher posterior probability for values higher than the reference maximum range of 6 m. In addition, the posterior distribution of the sill/variance has its maximum below that of the reference value (of 0.0003 $\mathrm{m}^2/\mathrm{ns}^2$). These observations are consistent with experimental semivariogram analysis of the reference model discussed previously, see Fig. 1d.

Considering $\mathbf{d}_{6\times6}$ leads to the poorest resolution of Ψ. Smaller ranges have a relative higher a posteriori probability compared to higher ranges, as seen on the two-dimensional marginal of r_1 and r_2 on Fig. 7. At the same time, most all a priori acceptable values of Ψ have non-zero a posteriori probability, i.e., any value in-between 0 and 10 m can in principle be consistent with the observed data, $\mathbf{d}_{6\times6}$. However, some conclusions can be made from these two-dimensional marginals. If both ranges are very small (less than 1 m), then both the variance, σ^2, and the nugget fraction α must also be high. Vice versa, if the variance is relatively low, there is an increased posterior probability that the ranges will also be higher than 1 m. Figure 7 illustrates that it is difficult to infer detailed information about the covariance model parameters Ψ from a data set such as $\mathbf{d}_{6\times6}$.

Considering $\mathbf{d}_{\mathrm{tomo100}}$, Fig. 8, the resolution of the inferred parameters Ψ is somewhat in-between that of $\sigma(\Psi\ |\mathbf{d}_{p100})$ and $\sigma(\Psi\ |\mathbf{d}_{p6\times6})$. The two-dimensional marginal for r_1 and θ_1 shows two clearly separated areas of high posterior probability. One around the true reference parameters and one around $r_1 = 2$ m and $\theta_1 = 60°$. Again, this implies that local sampling and optimization techniques may not be useful because such methods will only find a local area of high posterior probability and other areas with significant posterior probability of Ψ values may be missed.

Figure 9 shows the corresponding 5 one-dimensional marginal a priori, $\rho(\Psi\ |\mathbf{d}_{\mathrm{tomo100}})$ (black), and a posteriori, $\sigma(\Psi\ |\mathbf{d}_{\mathrm{tomo100}})$ (red), distributions. Note how it is harder to locate separate areas of high local a posteriori probability from the one-dimensional marginal distributions as compared to the two-dimensional marginal distributions in Fig. 8. The relatively clear separation of areas of high posteriori probability from the two-dimensional marginal of r_1 and θ_1 is much less obvious from the one-dimensional marginal that suggests an apparently easier sampling problem.

3.2.3 Inference of $\sigma(r_1, r_2, T\,|\,\mathbf{d}_{\mathrm{obs}})$

In the previous examples, the type of covariance model T has been assumed known. In principle, information about T can be inferred from Eq. (11) using, e.g., the extended Metropolis algorithm. However, in practice this poses some problems.

First of all, the physical interpretation of the range of a Spherical type covariance model, $T = \mathrm{Sph}$, is very different from that of a Gaussian type, $T = \mathrm{Gau}$. Therefore, it may not make much sense to infer information about the covariance model type for a fixed range. However, it may be valid to infer information about the covariance model type if the ranges are inferred at the same time. Even in this case, this may pose some challenges with respect to the sampling problem.

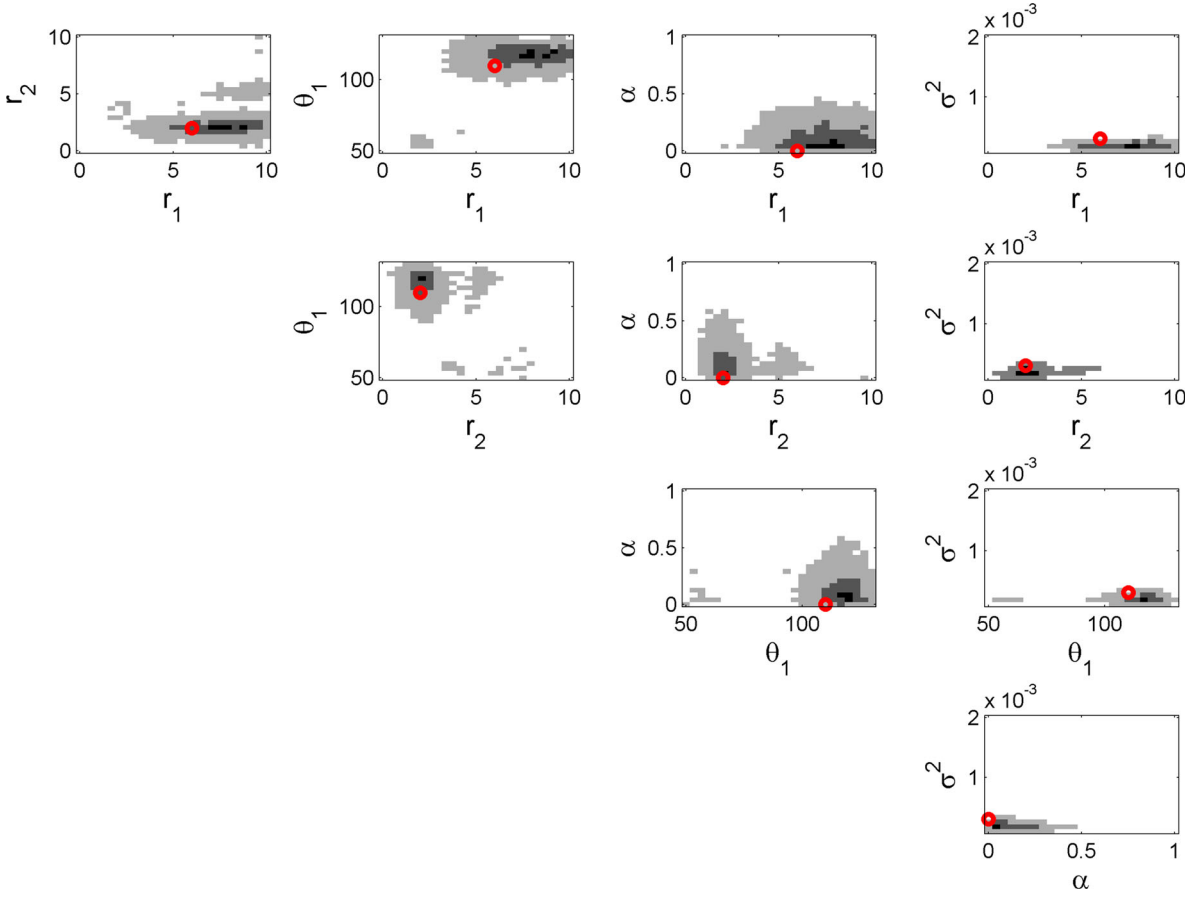

Fig. 6 Two-dimensional marginal a posteriori probability density from $\sigma(\Psi = [r_1, r_2, \theta_1, \sigma^2, \alpha] \mid \mathbf{d}_{p100})$. *Gray scales* indicates 10 % (*light gray*), 50 % (*gray*), and 90 % (*black*) highest posterior density regions. The *red circle* indicates the values of the reference parameters

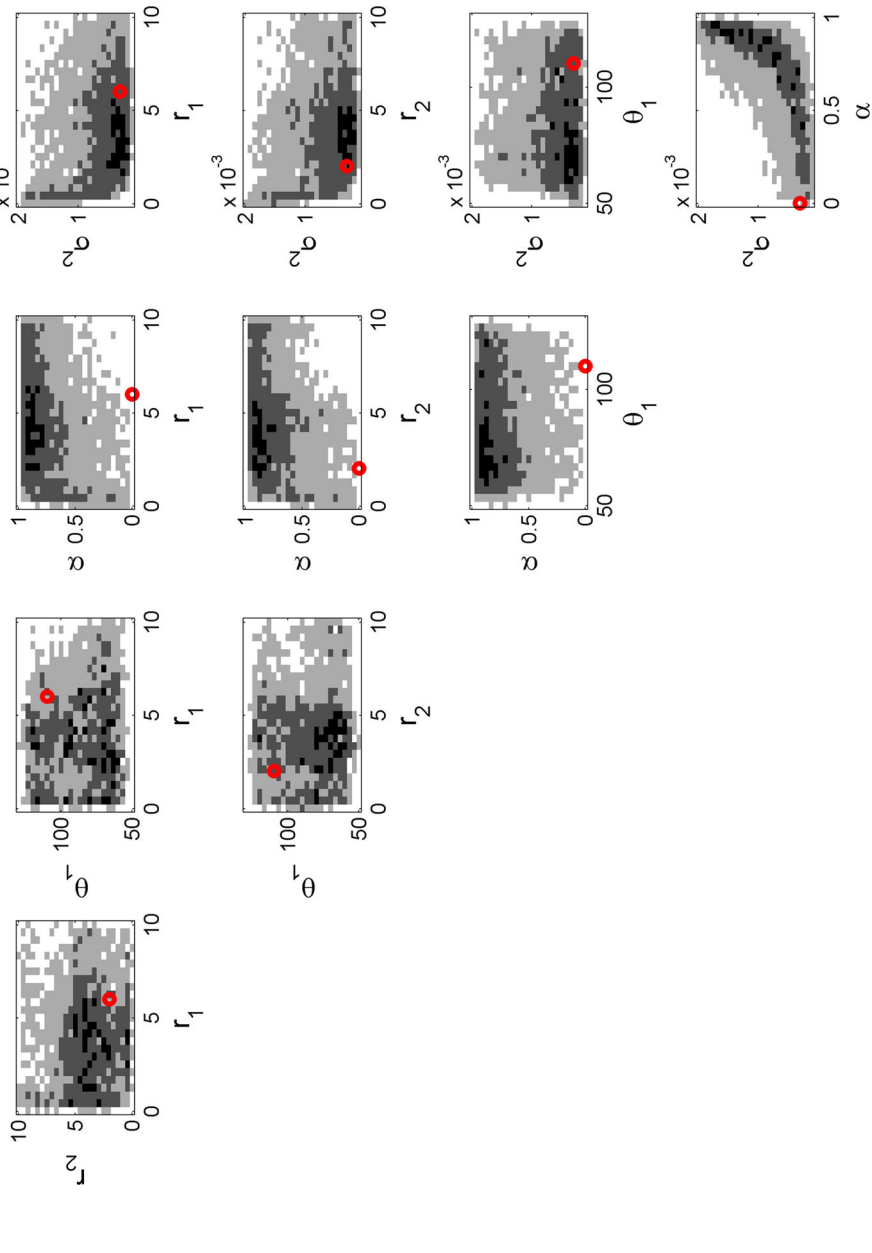

Fig. 7 Two-dimensional marginal a posteriori probability density from $\sigma(\Psi = [r_1, r_2, \theta_1, \sigma^2, \alpha] \mid \mathbf{d}_{6\times 6})$. *Grayscale* as in Fig. 6

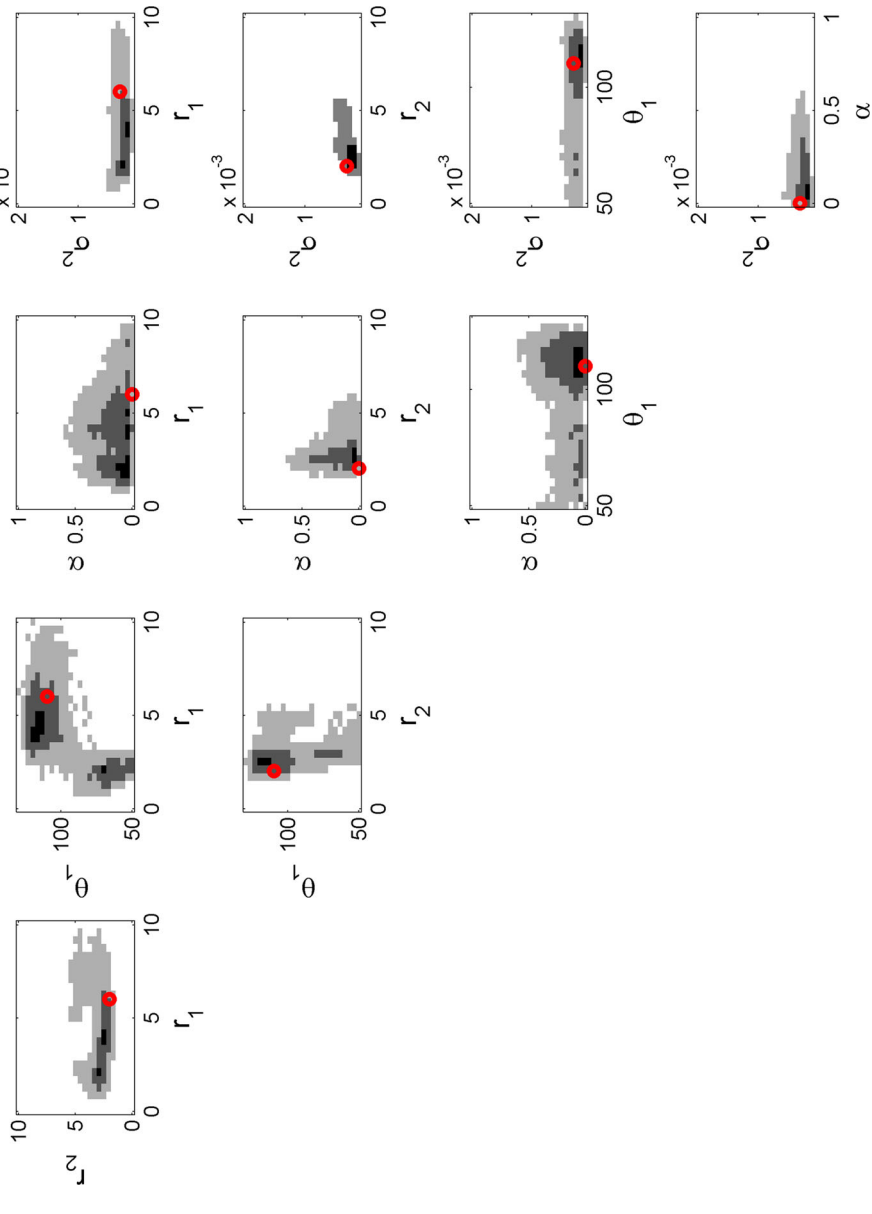

Fig. 8 Two-dimensional marginal a posteriori probability density from $\sigma(\boldsymbol{\Psi} = [r_1, r_2, \theta_1, \sigma^2, \alpha]|\mathbf{d}_{\text{tomo100}})$. *Grayscale* as in Fig. 6

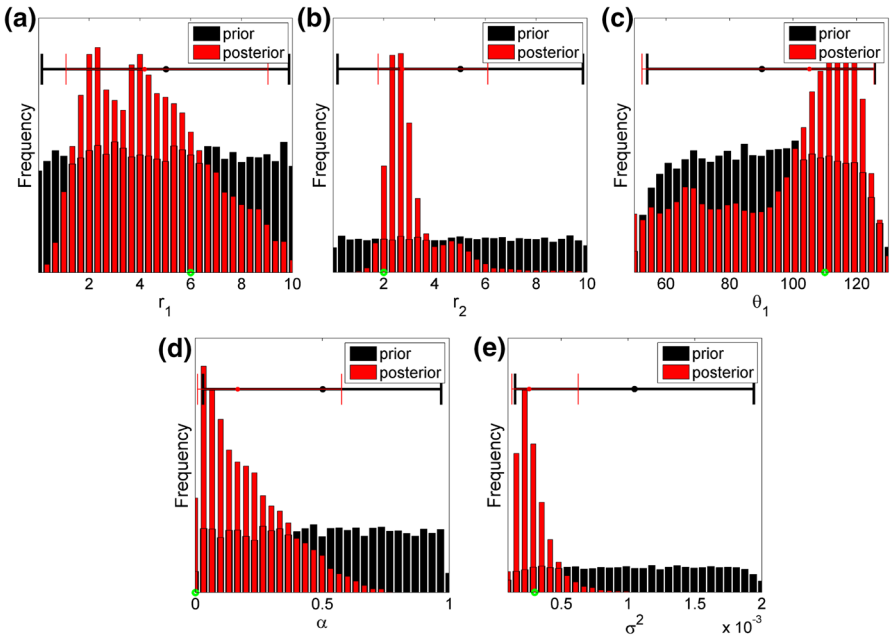

Fig. 9 One-dimensional marginal a posteriori probability density from $\sigma(\Psi = [r_1, r_2, \theta_1, \sigma^2, \alpha] | \mathbf{d}_{\text{tomo100}})$. The *green circle* indicate the reference value

The extended Metropolis algorithm relies on a method to perform a random walk in the space of possible a priori models. The efficiency of the algorithm is related to the step-length, i.e., the distance between successive models in the random walk. The step-length can easily be adjusted for the continuous parameters used above, but the covariance model type is a categorical variable that in this case can take three values, that are not ordered. There is no meaningful way to define the distance between T = Gau, T = Sph, and T = Exp. This means, there is no way to choose the step-length for such a specific type of variable. To avoid such problems one can choose to make use of the Metropolized independent sampler, which is a special case of the Metropolis algorithm, where each proposed model is independent from the previous considered model (Liu 1996). This is trivially extended to use with the extended Metropolis sampler. The computational efficiency of such a sampler resemble that of a rejection sampler, with the important benefit that the absolute level of the likelihood need not be known, as is necessary when using the rejection sampler. Even so, the Metropolized independence sampler is prone to the same problems related to the curse of dimensionality as the rejection sampler, and is thus only applicable to low-dimensional inverse problem. Inference of a few covariance model parameters can be seen as such a low-dimensional problem.

The extended Metropolized independence sampler is used to sample the a posteriori distributions $\sigma(\Psi \mid \mathbf{d}p100)$ and $\sigma(\Psi \mid \mathbf{d}\text{tomo}100)$, where $\Psi = [r_1, r_2, T]$. A priori it is assumed that the three types of covariance models are equally probable, $\rho(T = \text{'Sph'}) = \rho(T = \text{'Gau'}) = \rho(T = \text{'Exp'}) = 1/3$.

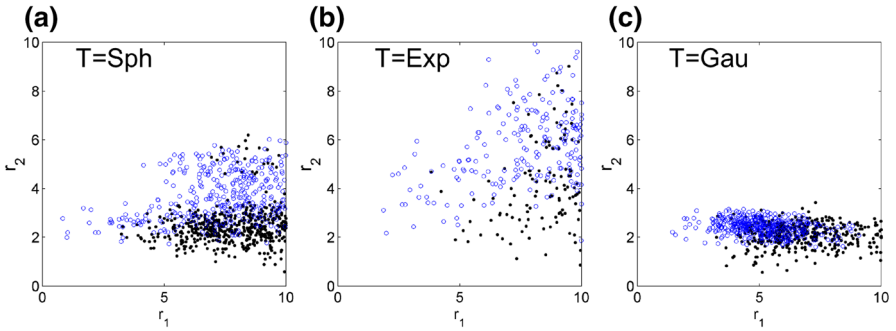

Fig. 10 Two-dimensional marginal a posteriori distributions $\sigma(r_1, r_2 \mid \mathbf{d}_{p100})$ (*black dots*) and $\sigma(r_1, r_2 \mid \mathbf{d}_{\text{tomo100}})$ (*blue circles*), in case **a** $T = \text{Sph}$, **b** $T = \text{Exp}$, and **c** $T = \text{Gau}$

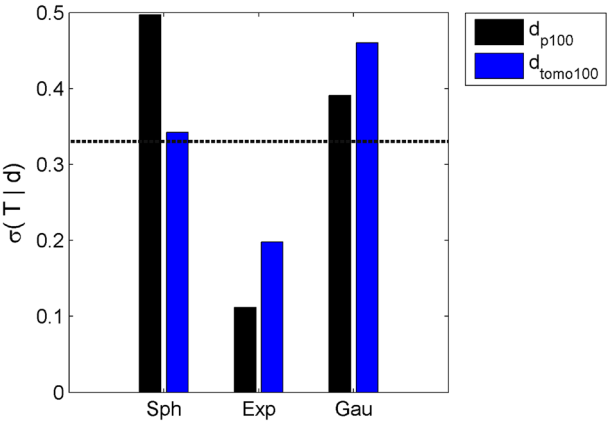

Fig. 11 One-dimensional marginal a posteriori distributions $\sigma(T \mid \mathbf{d}_{p100})$ (*black*) and $\sigma(T \mid \mathbf{d}_{\text{tomo100}})$ (*blue*). *Dashed line* a priori probability

Figure 10a–c show 800 realizations from two-dimensional marginals for $\sigma(r_1, r_2, T \mid \mathbf{d}_{p100})$ (black dots) and $\sigma(r_1, r_2, T \mid \mathbf{d}_{\text{tomo100}})$ (blue circles) for each type of covariance model. It is clear to see that each covariance type is associated with a specific posterior marginal distribution of range values. The Exponential type covariance model, $T = \text{Exp}$, is the one that portraits the largest variability in range values, Fig. 10b.

The actual frequency of occurrence of each type of covariance, T, corresponds to the posterior probability of the covariance type. Figure 11 shows the one-dimensional marginal probability of covariance type for $\sigma(T \mid \mathbf{d}_{p100})$ and $\sigma(T \mid \mathbf{d}_{\text{tomo100}})$. The most striking observation is perhaps that all the three considered covariance type models show significant posterior probability. Both the Spherical and Gaussian type covariance models show increased (and similar) posterior probability as opposed to prior probability, while the Exponential type covariance model is the least probable. The Gaussian type covariance model has slightly higher posterior probability than the Spherical type covariance model, given $\mathbf{d}_{\text{tomo100}}$, while the opposite can be observed for $\mathbf{d}_{p100}$. Figure 11 demonstrates that the covariance model type T cannot be inferred

well from the tomographic data set $\mathbf{d}_{\text{tomo100}}$. In general, the covariance type is not well resolved using any of the considered data.

Alternate covariance type models The use of different kinds of covariance models is a challenge both in terms of interpreting the physical meaning of the parameters of Ψ, but also in practical application of the sampling algorithm. Another option, not pursued here, is to make use of for example the Matérn covariance model as, e.g., Truong et al. (2014). It is defined using a modified Bessel function of order μ, which allow one to describe the type of spatial variability ranging from a smooth random field, when $\mu = \infty$, to random field with more small-scale variability, for smaller values of μ. Thus, instead of using three types of different possible covariance models, one could make use of the Matérn covariance function and treat μ as an unknown continuous parameter, that could be inferred efficiently using the approach presented in this paper.

4 Discussion

As discussed in the introduction, many methods have been suggested for inference of an assumed Gaussian a priori model describing as set of observed linear average data. The method we have suggested has some properties that separates it from these methods: any type of linear average data, from data of point to volume support (such as average values within arbitrarily shaped regions), can be handled with no difficulty. Noise related to linear average data is naturally handled as a Gaussian model of, possibly correlated, data errors. All information available, such as information about the observed data, and the a priori information, is quantified by probabilities. This allows the formulation of the problem of inference of information about Gaussian model parameter describing a Gaussian model as a probabilistic inverse problem. The solution to such an inverse problem is the a posteriori probability that basically present an integration of all the information at hand about the Gaussian model parameters. In this sense, we believe the proposed methodology is perhaps the most general way to deal with the problem of inference of covariance model parameters from any type of linear average data.

If the observed data cannot be seen as linear averages of data at point support, the presented methodology cannot be used. Instead one can make use of more computational demanding methods such as suggested in Hansen et al. (2013a) that allow inference of parameters of a Gaussian model, as well as the properties of the spatially distributed model parameters simultaneously, while data can be non-linearly related to the model parameters.

Inference of the nugget fraction from linear average data In the example shown above, we found that it was possible to infer information about all the considered covariance model parameters defined by Ψ, including the nugget fraction. This seems to be in conflict with Truong et al. (2014) who conclude that the nugget of point support variogram cannot be inferred from block support data. Such claims has also been made by, e.g., Lele and Das (2000) and Goovaerts (2008).

The data used by Truong et al. (2014) were similar to that of $\mathbf{d}_{2\times 2}$ and $\mathbf{d}_{6\times 6}$, i.e., average values measured within non-overlapping pixels. Figure 7 shows that

when considering $\mathbf{d}_{6\times 6}$, all values of nugget fraction between 0 and 1 have significant posterior probability. However, from the two-dimensional marginals it was evident that for example if the variance is relatively small, then then the nugget fraction must also be small, and vice versa. In this respect, we have inferred information (conditional to, e.g., the variance) about the nugget fraction.

Considering the data related to tomography, the volume average data consists of data with overlapping linear average kernels, see Fig. 3. In this case, comparison of the one-dimensional marginal prior and posterior distribution of the nugget fraction, Fig. 9e, clearly demonstrates that the information about the nugget fraction can be inferred. With about 90 % certainty it can be concluded that the nugget fraction is below 0.5, with most posterior probability assigned to the a nugget fraction close to 0.

5 Conclusions

A general approach for computing the likelihood that a set of, possibly noisy, data of point and/or volume support is a realization from a Gaussian probability distribution with specific Gaussian model parameters has been presented. This can be seen as a generalization of the classical likelihood approach used for inference of covariance model parameters given exact data of point support. This measure has been used to pose a problem of inference of Gaussian model parameters (given noisy data of point/volume support) as a probabilistically formulated inverse problem. This allows use of a priori information on the Gaussian model parameters. The solution to such an inverse problem is an a posteriori probability distribution over the Gaussian model parameters. Through numerous examples, using noisy data of both point and volume support, we have demonstrated how a sample from this a posteriori probability can be generated using for example the extended Metropolis algorithm. In this way, a thorough resolution analysis of the Gaussian model parameters can be provided. When disregarding the noise, i.e., treating noisy data as noise-free data, inference of Gaussian model parameters will, erroneously, suggest a nugget fraction in an attempt to account for the noise on data. It was demonstrated how the proposed method allows accounting for Gaussian distributed noise on data, such that data noise is not erroneously estimated as a nugget fraction.

Acknowledgments We thank Dong E&P for financial support. Matlab code for inferring properties of three-dimensional anisotropic covariance models from noisy linear average data (and data of point support) using the proposed method is available from http://sippi.sourceforge.net/. We thank professor emeritus Carl Christian Tscherning for illuminating discussion about collocated least-squares.

References

Arabelos D, Tscherning CC (2003) Globally covering a-priori regional gravity covariance models. Adv Geosci 1:143–147

Asli M, Marcotte D, Chouteau M (2000) Direct inversion of gravity data by cokriging. In: Kleingeld W, Krige D (eds) Proceedings of the 6th international geostatistics congress, Cape Town, South Africa, 10–14 April, pp 64–73

Atkinson PM (2013) Downscaling in remote sensing. Int J Appl Earth Obs Geoinf 22:106–114. doi:10.1016/j.jag.2012.04.012

Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Addison-Wesley, Reading, MA
Chiles J-P, Delfiner P (2012) Geostatistics: modeling spatial uncertainty. Wiley series in probability and statistics. Wiley, Hoboken, NJ
Cordua KS, Looms MC, Nielsen L (2008) Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data. Vadose Zone J 7(1):263
Cressie N (1985) Fitting variogram models by weighted least squares. Math Geol 17(5):563–586
Desassis N, Renard P (2013) Automatic variogram modeling by iterative least squares: univariate and multivariate cases. Math Geosci 34(4):453–470
Emery X (2010) Iterative algorithms for fitting a linear model of coregionalization. Comput Geosci 36(9):1150–1160
Frykman P, Deutsch C (1999) Geostatistical scaling laws applied to core and log data. In: Proceedings of SPE annual technical conference and exhibition, pp 887–898
Frykman P, Deutsch C (2002) Practical application of geostatistical scaling laws for data integration. Petrophys 43(3):153–171
Georgii H-O (2008) Stochastics: introduction to probability and statistics, 1st edn. Walter de Gruyter, Berlin. ISBN: 3110191458
Giroux B, Gloaguen E, Chouteau M (2007) bh_tomo: a Matlab borehole georadar 2d tomography package. Comput Geosci 33(1):126–137
Gloaguen E, Marcotte D, Chouteau M, Perroud H (2005) Borehole radar velocity inversion using cokriging and cosimulation. J Appl Geophys 57(4):242–259
Gloaguen E, Marcotte D, Giroux B, Dubreuil-Boisclair C, Chouteau M, Aubertin M (2007) Stochastic borehole radar velocity and attenuation tomographies using cokriging and cosimulation. J Appl Geophys 62(2):141–157
Goovaerts P (1997) Geostatistics for natural resources evalutaion. Applied geostatistics series. Oxford University Press, New York
Goovaerts P (2008) Kriging and semivariogram deconvolution in the presence of irregular geographical units. Math Geosc 40(1):101–128
Goovaerts P (2010) Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. Math Geosc 42(5):535–554
Hansen TM, Mosegaard K (2008) VISIM: sequential simulation for linear inverse problems. Comput Geosci 34(1):53–76
Hansen TM, Journel AG, Tarantola A, Mosegaard K (2006) Linear inverse Gaussian theory and geostatistics. Geophysics 71(6):101–111
Hansen TM, Looms MC, Nielsen L (2008) Inferring the subsurface structural covariance model using cross-borehole ground penetrating radar tomography. Vadose Zone J 7(1):249–262
Hansen TM, Cordua KS, Looms MC, Mosegaard K (2013a) SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: part 1, methodology. Comput Geosci 52:470–480. doi:10.1016/j.cageo.2012.09.004
Hansen TM, Cordua KS, Looms MC, Mosegaard K (2013b) SIPPI: a Matlab toolbox for sampling the solution to inverse problems with complex prior information: part 2, application to cross hole GPR tomography. Comput Geosci 52:481–492. doi:10.1016/j.cageo.2012.10.001
Hansen TM, Cordua KS, Jacobsen BH, Mosegaard K (2014) Accounting for imperfect forward modeling in geophysical inverse problems exemplified for crosshole tomography. Geophysics 79(3):1–21
Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57(1):97
Herzfeld UC (1992) Least-squares collocation, geophysical inverse theory and geostatistics: a bird's eye view. Geophys J Int 111(2):237–249
Isaaks EH, Srivastava RM (1989) Applied geostatistics. Oxford University Press, Oxford
Jarmołowski W, Bakuła M (2014) Precise estimation of covariance parameters in least-squares collocation by restricted maximum likelihood. Studia Geophysica et Geodaetica 58(2):171–189
Jensen JM, Jacobsen BH, Christensen-Dalsgaard J (2000) Sensitivity kernels for time-distance inversion. Solar Phys 192(1–2):231–239
Journel AG, Huijbregts CJ (1978) Mining geostatistics. Academic Press, New York
Kay SM (2006) Intuitive probability and random processes using matlab. Springer, New York
Kelsall J, Wakefield J (2002) Modeling spatial variation in disease risk. J Am Stat Assoc 97(459):692–701
Kitanidis PK, Lane RW (1985) Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss–Newton method. J Hydrol 79(1/2):53–71

Knudsen P (1987) Estimation and modelling of the local empirical covariance function using gravity and satellite altimeter data. Bull Geod 61(2):145–160
Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Meddelse no. 44, Geodaetisk Institut, Koebenhavn, p 80
Kupfersberger H, Deutsch CV, Journel AG (1998) Deriving constraints on small-scale variograms due to variograms of large-scale data. Math Geol 30(7):837–852
Kyriakidis PC (2004) A geostatistical framework for area-to-point spatial interpolation. Geogr Anal 36(3):259–289
Lark R, Papritz A (2003) Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115(3):245–260
Lele SR, Das A (2000) Elicited data and incorporation of expert opinion for statistical inference in spatial studies. Math Geol 32(4):465–487. doi:10.1023/A:1007525900030
Liu JS (1996) Metropolized independent sampling with comparisons to rejection sampling and importance sampling. Stat Comput 6(2):113–119
Liu Y, Journel AG (2009) A package for geostatistical integration of coarse and fine scale data. Comput Geosci 35(3):527–547
Looms MC, Hansen TM, Cordua KS, Nielsen L, Jensen KH, Binley A (2010) Geostatistical inference using crosshole ground-penetrating radar. Geophysics 75(6):29
Mosegaard K (1998) Resolution analysis of general inverse problems through inverse Monte Carlo sampling. Inverse Probl 14:405
Mosegaard K, Tarantola A (1995) Monte Carlo sampling of solutions to inverse problems. J Geophys Res 100(B7):12431–12447
Pardo-Igúzquiza E (1997) Mlreml: a computer program for the inference of spatial covariance parameters by maximum likelihood and restricted maximum likelihood. Comput Geosci 23(2):153–162
Pardo-Igúzquiza E (1998) Maximum likelihood estimation of spatial covariance parameters. Math Geol 30(1):95–108
Pardo-Igúzquiza E (1999) Varfit: a Fortran-77 program for fitting variogram models by weighted least squares. Comput Geosci 25(3):251–261
Pardo-Igúzquiza E, Dowd PA (2005) Emlk2d: a computer program for spatial estimation using empirical maximum likelihood kriging. Comput Geosci 31:361–370
Remy N, Boucher A, Wu J (2008) Applied geostatistics with SGeMS: a user's guide. Cambridge University Press, Cambridge
Shamsipour P, Marcotte D, Chouteau M, Keating P (2010) 3d stochastic inversion of gravity data using cokriging and cosimulation. Geophysics 75(1):1–10
Tarantola A (2005) Inverse problem theory and methods for model parameter estimation. Society of Industrialand Applies Mathematics, Philadelphia
Tarantola A, Valette B (1982a) Generalized nonlinear inverse problems solved using the least squares criterion. Rev Geophys Space Phys 20(2):219–232
Tarantola A, Valette B (1982b) Inverse problems = quest for information. J Geophys 50(3):150–170
Truong PN, Heuvelink GMB, Pebesma E (2014) Bayesian area-to-point kriging using expert knowledge as informative priors. Int J Appl Earth Obs Geoinf 30:128–138. doi:10.1016/j.jag.2014.01.019
Warnes J, Ripley B (1987) Problems with likelihood estimation of covariance functions of spatial gaussian processes. Biometrika 74(3):640–642