

Linear inverse Gaussian theory and geostatistics

Thomas Mejer Hansen¹, Andre G. Journel², Albert Tarantola³, and Klaus Mosegaard¹

ABSTRACT

Inverse problems in geophysics require the introduction of complex a priori information and are solved using computationally expensive Monte Carlo techniques (where large portions of the model space are explored). The geostatistical method allows for fast integration of complex a priori information in the form of covariance functions and training images. We combine geostatistical methods and inverse problem theory to generate realizations of the posterior probability density function of any Gaussian linear inverse problem, honoring a priori information in the form of a covariance function describing the spatial connectivity of the model space parameters. This is achieved using sequential Gaussian simulation, a well-known, noniterative geostatistical

method for generating samples of a Gaussian random field with a given covariance function.

This work is a contribution to both linear inverse problem theory and geostatistics. Our main result is an efficient method to generate realizations, actual solutions rather than the conventional least-squares-based approach, to any Gaussian linear inverse problem using a noniterative method. The sequential approach to solving linear and weakly nonlinear problems is computationally efficient compared with traditional least-squares-based inversion. The sequential approach also allows one to solve the inverse problem in only a small part of the model space while conditioned to all available data. From a geostatistical point of view, the method can be used to condition realizations of Gaussian random fields to the possibly noisy linear average observations of the model space.

INTRODUCTION

Consider the expression

$$\mathbf{d} = \mathbf{G}\mathbf{m}, \quad (1)$$

where \mathbf{G} is a forward mapping operator (linear or nonlinear) that maps the model parameters \mathbf{m} into observations \mathbf{d} . Estimation of \mathbf{m} , using the forward mapping operator \mathbf{G} and the observed data \mathbf{d} , is referred to as solving the inverse problem.

Linear inversion, where the mapping operator \mathbf{G} is linear, generally provides an overly smooth estimate of the properties being estimated. For instance, in seismic tomography, using least-squares linearized inversion with observed traveltimes \mathbf{d} of seismic waves through a medium \mathbf{m} produces overly smooth estimates of the underlying velocity model \mathbf{m}_{est} (Phillips and Fehler, 1991). Also, complex a priori information cannot be readily included in the inversion.

Monte Carlo methods can be applied to such linear (and nonlinear) problems, allowing complex a priori information to be included (Mosegaard and Tarantola, 1995). In a tomography setup, a priori information could be the knowledge about the spatial correlation of the model space parameters, i.e., the spatial correlation of the subsurface velocity field, hereafter denoted as \mathbf{C}_M . However, the use of a standard Metropolis sampler to generate models that honor traveltimes data and a priori information is extremely inefficient. Such algorithms are asymptotic and short-memory samplers that do not take advantage of the linear and Gaussian nature of the problem.

Kriging is a geostatistical technique to interpolate observed data values in space, given a covariance model specifying the spatial correlation of data \mathbf{C}_M . The simplest form of kriging, called simple kriging, is in fact identical to a linear Gaussian inverse problem (equation 1) with direct observations of the model space $\mathbf{d}_{\text{obs}}(\mathbf{x}_i) = \mathbf{m}(\mathbf{x}_i)$ at locations \mathbf{x}_i . Interpolation based on kriging will produce maps of \mathbf{m} that are overly smooth, as compared to the assumed a priori cova-

Manuscript received by the Editor August 3, 2005; revised manuscript received May 5, 2006; published online November 3, 2006.

¹University of Copenhagen, Niels Bohr Institute, Juliane Mariesvej 28, DK-2100 Copenhagen East, Denmark. E-mail: tmh@gfy.ku.dk.

²Stanford University, Stanford Center for Reservoir Forecasting, Department of Geological and Environmental Sciences, Stanford, California 94305. E-mail: journel@pangea.stanford.edu.

³Institut de Physique du Globe de Paris, 4 place Jussieu, 75005 Paris, France. E-mail: tarantola@mac.com.

© 2006 Society of Exploration Geophysicists. All rights reserved.

riance model \mathbf{C}_M . Journel and Huijbregts (1978) give a detailed description of the kriging approach.

Geostatisticians use the concept of sequential simulation to correct the smoothing effect of kriging by simulating a number of realizations of the posterior probability function $\mathbf{d}_{\text{est}}(\mathbf{x})$ honoring both the a priori information \mathbf{C}_M and the observed values $\mathbf{d}_{\text{obs}}(\mathbf{x})$. Unlike Metropolis-based Monte Carlo methods, sequential simulation is not an iterative, asymptotic method. It has maximum computational efficiency in the sense that it produces one sample that honors the a priori information in each iteration. Goovaerts (1997) gives an introduction to sequential simulation.

In this paper, we present a methodology that applies the concept of sequential simulation to any linear Gaussian inverse problem with prior information given by a covariance function. This allows the generation of multiple realizations of the model space consistent with observations \mathbf{d} and the a priori model \mathbf{C}_M . As an example, we apply the method to a linear Gaussian tomographic inverse problem based on synthetic data.

We use notations from linear inverse theory and introduce geostatistical terms where needed. However, the choice of notation should not distract the reader from the fact that the methodology could be presented equally well using geostatistical notation. In this paper, we apply sequential simulation to solve the Gaussian linear inverse problem and extend conventional simple kriging and sequential simulation to include measurements of linear averages of the model space, in addition to point measurements.

SEQUENTIAL SIMULATION AND LINEAR GAUSSIAN INVERSE THEORY

The concept of sequential simulation, as used in geostatistics, can be applied to any Gaussian linear inverse problem to produce a number of realizations of a random field with a specified a priori covariance model \mathbf{C}_M and honoring observations \mathbf{d} that can be seen as linear averages of the model space.

The a priori information on the model parameters \mathbf{m} is represented using a Gaussian probability density with mean \mathbf{m}_0 and covariance \mathbf{C}_M . For the remainder of this text, we will not consider how to choose a priori \mathbf{C}_M but refer the interested reader to Goovaerts (1997) for a thorough discussion on the inference of \mathbf{C}_M .

Data measurements \mathbf{d} are represented using a Gaussian probability density with mean \mathbf{d}_0 and covariance \mathbf{C}_D . Further, we assume that the relation between data and model parameters is linear.

The Gaussian linear inverse problem — One data set

We start by introducing the Gaussian linear inverse problem for two types of observed data: type A direct measurements of some model parameters, and type B linear average measurements of model parameters. This distinction between data types may seem trivial, but it allows relatively straightforward implementation of sequential simulation.

Type A data — Direct measurements of model parameters

First, consider a situation where observations \mathbf{a}_0 (hereafter referred to as data of type A) are available and that they are direct measurements of model parameters at some locations \mathbf{x}_i in that $\mathbf{a}_0(\mathbf{x}_i) = \mathbf{m}(\mathbf{x}_i)$. The linear mapping function that transforms model parameters into observations is denoted as \mathbf{A} . This can be described as

$$\mathbf{a}_0 = \mathbf{A}\mathbf{m}. \quad (2)$$

The least-squares solution to equation 2 can be described by a Gaussian probability density (see Tarantola, 2005) with mean

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_M \mathbf{A}^t (\mathbf{A} \mathbf{C}_M \mathbf{A}^t + \mathbf{C}_D)^{-1} (\mathbf{a}_0 - \mathbf{A} \mathbf{m}_0) \quad (3)$$

and with covariance

$$\tilde{\mathbf{C}}_M = \mathbf{C}_M - \mathbf{C}_M \mathbf{A}^t (\mathbf{A} \mathbf{C}_M \mathbf{A}^t + \mathbf{C}_D)^{-1} \mathbf{A} \mathbf{C}_M. \quad (4)$$

A forward mapping operator like \mathbf{A} is relatively unconventional for inverse theory applied to the physical observations because \mathbf{A} is independent of any physical law and simply gives the value of the function $\mathbf{m}(\mathbf{x})$ at given locations \mathbf{x} . Equations 3 and 4 are identical to the solution to a simple kriging system, with $\tilde{\mathbf{m}}$ being the kriging mean and $\tilde{\mathbf{C}}_M$ the kriging variance.

Type B data — Linear average measurements of model parameters

Consider now a situation where we are able to measure some linear averages over the model space \mathbf{b}_0 (hereafter referred to as data of type B):

$$\mathbf{b}_0 = \mathbf{B}\mathbf{m}. \quad (5)$$

Here, \mathbf{B} could relate to the raypaths of a tomographic experiment. The least-squares solution to equation 5 can be described by a Gaussian probability density (see Tarantola, 2005) with mean and covariance as given in equations 3 and 4, substituting \mathbf{A} with \mathbf{B} and \mathbf{a}_0 with \mathbf{b}_0 .

Though it is obvious that data of type A can be seen as a special case of data of type B (when the volume average is measured at a point), splitting the observed data into types A and B makes utilization of sequential simulation straightforward.

The Gaussian linear inverse problem — Two data sets

Assume data of types A and B are available. Direct measurements of the model parameters \mathbf{a}_0 (type A) and linear averages of the model parameters \mathbf{b}_0 (type B) are such that

$$\mathbf{d}_0 = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix}, \quad \mathbf{C}_D = \begin{bmatrix} \mathbf{C}_{aa} & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^t & \mathbf{C}_{bb} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix}, \quad (6)$$

where \mathbf{C}_{aa} and \mathbf{C}_{bb} are data covariances for the observed data \mathbf{a} and \mathbf{b} and \mathbf{C}_{ab} is the cross data covariance between the two data types. The observations are linked to the model through the linear operator \mathbf{G} :

$$\mathbf{d}_0 = \mathbf{G}\mathbf{m}. \quad (7)$$

The least-squares solution to equation 7 then becomes a Gaussian probability density function with mean

$$\tilde{\mathbf{m}} = \mathbf{m}_0 + \mathbf{C}_M [\mathbf{A}^t \quad \mathbf{B}^t] \begin{bmatrix} \mathbf{T}_{aa} & \mathbf{T}_{ab} \\ \mathbf{T}_{ab}^t & \mathbf{T}_{bb} \end{bmatrix} \left(\begin{bmatrix} \mathbf{a}_0 \\ \mathbf{b}_0 \end{bmatrix} - \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{m}_0 \right) \quad (8)$$

and with covariance

$$\tilde{\mathbf{C}}_M = \mathbf{C}_M - \mathbf{C}_M [\mathbf{A}^t \quad \mathbf{B}^t] \begin{bmatrix} \mathbf{T}_{aa} & \mathbf{T}_{ab} \\ \mathbf{T}_{ab}^t & \mathbf{T}_{bb} \end{bmatrix} \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{C}_M, \quad (9)$$

where

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{aa} & \mathbf{S}_{ab} \\ \mathbf{S}_{ab}^t & \mathbf{S}_{bb} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_{aa} & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^t & \mathbf{C}_{bb} \end{bmatrix} + \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \mathbf{C}_M [\mathbf{A}^t & \mathbf{B}^t] \quad (10)$$

and

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{aa} & \mathbf{T}_{ab} \\ \mathbf{T}_{ab}^t & \mathbf{T}_{bb} \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{aa} & \mathbf{S}_{ab} \\ \mathbf{S}_{ab}^t & \mathbf{S}_{bb} \end{bmatrix}^{-1} = \mathbf{S}^{-1}. \quad (11)$$

Equations 8 and 9 can thus be seen as the solution to a linear inverse problem, with direct observations of the model space through \mathbf{a}_0 and linear averages of the model space through \mathbf{b}_0 .

Estimation of the posterior Gaussian probability at one point only

Equations 8 and 9 provide the mean and covariance of the posterior Gaussian probability density function at all locations in space, as defined by the \mathbf{G} and \mathbf{m} matrices.

In some situations, specifically when applying sequential simulation, it is convenient to estimate the posterior mean and covariance for only one point in the model space. The solution to the least-squares inverse problem in equation 7 at any location \mathbf{x} is given by

$$\begin{aligned} \tilde{m}(\mathbf{x}) &= m_0(\mathbf{x}) + \sum_{i=1}^n \sum_{j=1}^n C_M(\mathbf{x}, \mathbf{x}_i) (\mathbf{T}_{aa})_{ij} ((\mathbf{a}_0)_j - m_0(\mathbf{x}_j)) \\ &+ \sum_{l=1}^N \sum_{j=1}^n \int_{\mathcal{V}_l} dx C_M(\mathbf{x}, \mathbf{x}_l) (\mathbf{T}_{ab}^t)_{lj} ((\mathbf{a}_0)_j - m_0(\mathbf{x}_j)) \\ &+ \sum_{i=1}^n \sum_{J=1}^N C_M(\mathbf{x}, \mathbf{x}_i) (\mathbf{T}_{ab})_{iJ} \left((\mathbf{b}_0)_J - \int_{\mathcal{V}_J} dx m_0(\mathbf{x}_J) \right) \\ &+ \sum_{l=1}^N \sum_{J=1}^N \int_{\mathcal{V}_l} dx C_M(\mathbf{x}, \mathbf{x}_l) (\mathbf{T}_{bb})_{lJ} \\ &\times \left((\mathbf{b}_0)_J - \int_{\mathcal{V}_J} dx m_0(\mathbf{x}_J) \right). \end{aligned} \quad (12)$$

The posterior covariance being given by

$$\begin{aligned} \tilde{C}_M(\mathbf{x}, \mathbf{x}') &= C_M(\mathbf{x}, \mathbf{x}') - \sum_{i=1}^n \sum_{j=1}^n C_M(\mathbf{x}, \mathbf{x}_i) (\mathbf{T}_{aa})_{ij} C_M(\mathbf{x}_j, \mathbf{x}') \\ &- \sum_{l=1}^N \sum_{j=1}^n \int_{\mathcal{V}_l} dx C_M(\mathbf{x}, \mathbf{x}_l) (\mathbf{T}_{ab}^t)_{lj} C_M(\mathbf{x}_j, \mathbf{x}') \\ &- \sum_{i=1}^n \sum_{J=1}^N \int_{\mathcal{V}_J} dx C_M(\mathbf{x}, \mathbf{x}_i) (\mathbf{T}_{ab})_{iJ} C_M(\mathbf{x}_J, \mathbf{x}') \\ &- \sum_{l=1}^N \sum_{J=1}^N \int_{\mathcal{V}_l} dx \int_{\mathcal{V}_J} dx C_M(\mathbf{x}, \mathbf{x}_l) (\mathbf{T}_{bb})_{lJ} C_M(\mathbf{x}_J, \mathbf{x}'), \end{aligned} \quad (13)$$

where n is the number of available point data (type A) and N is the number of volume average data (type B). The values \mathcal{V}_l and \mathcal{V}_J are the integrals over the volume spanned by data of type B indicated by indices J, \dots , whereas j, \dots , are indices of data of type A.

Sequential simulation

Sequential simulation is a technique used to generate independent realizations of a random field describing a parameter \mathbf{m} . It requires that at any location \mathbf{x}_i , one can determine the probability distribution of $m(\mathbf{x}_i)$ conditional to the known model parameter values and the observed data. With respect to the theory described above, one must be able to estimate, at any location, the probability density function of the model parameter $m(\mathbf{x}_i)$ conditional to the observed data of types A and B.

As shown, $\tilde{m}(\mathbf{x})$ and $\tilde{C}_M(\mathbf{x}, \mathbf{x}')$ (equations 12 and 13) can be interpreted as the mean and covariance of the posterior Gaussian random field, conditioned to data of types A and B. Sequential simulation can now be performed as follows.

- 1) Randomly visit a point location in the model space, e.g., \mathbf{x}_i .
- 2) Compute the probability distribution conditional to the known and previously simulated nodes (\mathbf{a}_0 , data of type A) and the observed linear averages (\mathbf{b}_0 , data of type B). This can be done efficiently using equations 12 and 13. [Depending on the distribution of the data to simulate, either sequential Gaussian simulation (SGS) or direct sequential simulation (DSS) can be used. Using SGS, one can effectively skip this step and move on to step 3, as the conditional probability function is exactly given by a Gaussian with the calculated mean and variance. Using DSS, one can simulate any data distribution that can be described by a mean and a variance. To ensure histogram reproduction, one can reshape the conditional probability distribution, as proposed by Oz et al. (2003), prior to drawing a value, as in step 3. The application to non-Gaussian data distributions is discussed below.]
- 3) Draw a (random) value from the found conditional probability density function, e.g., $m_{\text{draw}}(\mathbf{x}_i)$. This is a sample of the posterior Gaussian probability function at location \mathbf{x}_i .
- 4) The simulated value $m_{\text{draw}}(\mathbf{x}_i)$ is now treated as a known datum of type A, $\mathbf{a}(\mathbf{x}_i)$, and added to the list of type A data such that $\mathbf{a}_{0,i+1} = [\mathbf{a}_{0,i}, \mathbf{a}(\mathbf{x}_i)]$.
- 5) Repeat steps 1–4 until all desired locations of the model space have been visited.

Note that for each visited point \mathbf{x}_i , a sample of the corresponding posterior Gaussian probability function is found and added to the conditioning data, i.e., data of type A. The estimated probability density function is conditioned to data of types A and B as opposed to only data of type A, as in traditional geostatistical SGS (Deutsch and Journel, 1998).

The collection of all these samples, one per location \mathbf{x}_i , is referred to as one realization of the posterior Gaussian random field. Additional independent realizations of the posterior random Gaussian field can be obtained by rerunning the algorithm with a different random seed. In practice, when using data neighborhoods of limited extent (as addressed later), a new random path must be chosen for each realization to ensure independent realizations.

The limited data neighborhood reduces computation time by reducing the number of conditioning data. Data of both types A and B that have small a priori covariance to the point being simulated \mathbf{x}_i , are removed because their contribution would be small to insignificant. The choice of such neighborhood is examined in more detail below.

Other methods

Gómez-Hernández and Cassiraga (2000) suggest using sequential cosimulation with linear constraints. However, they do not consider a formal link to linear inverse theory and consider only noise-free data.

Journal and Huijbregts (1978, p. 495) suggest an approach to conditional Gaussian simulation using a combination of unconditional simulation and kriging for data conditioning. Using this approach, Carr and Myers (1985) give a practical implementation of co-conditional simulation. Gloaguen et al. (2005) use cokriging of arrival times as measured from a cross-borehole survey to estimate a smooth-slowness field. Using the approach given by Journal and Huijbregts (1978) and Carr and Myers (1985), they combine unconditional Gaussian simulations with the previous cokriging result to produce stochastic realizations of the slowness field distribution conditioned to the observed traveltimes. The methodology of Gloaguen et al. (2005) is different from using SGS/DSS, particularly in that it requires a multivariate Gaussian distribution and hence has no flexibility to handle non-Gaussian conditioning data.

Monte Carlo-based algorithms can be used to generate realizations of a random field $m(\mathbf{x})$ consistent with prior information and observed data. Several iterative approaches exist. They share the same basic structure: An initial random image is generated, the model is randomly perturbed, and the perturbed model is either rejected or accepted according to some objective function. There are many ways to perturb the model, but one option is to switch the values of two randomly chosen locations. This ensures that the histogram of the initial random field is preserved. In theory, this approach is guaranteed to provide a solution (one realization of the posterior probability density function) consistent with a priori information and observed data. Unfortunately it is impossible to determine a priori when the solution will converge [see Goovaerts (1997) for more details].

GEOSTATISTICAL TOMOGRAPHY AS AN INVERSE PROBLEM: A CROSS-BOREHOLE EXAMPLE

As an example of an application of the developed theory, a synthetic cross-borehole tomographic example is considered.

A reference model with a 2D grid of 60 (horizontal) by 80 (vertical) nodes, with 250-m spacing between grid points is chosen. Using the SGSIM program from the GSLIB software (Deutsch and Journel, 1998), one realization of a Gaussian random field with a global mean and variance of $V_{\text{mean}} = 5 \text{ km/s}$ and $\sigma = 0.1 \text{ km/s}$ is selected as a reference velocity field (see Figure 1). An isotropic spherical covariance model with a range of 400 m (Figure 1c) is used to generate the reference velocity field. The range is the distance beyond which the covariance is zero, i.e., data separated by a distance larger than the range are uncorrelated.

In geostatistics, the equivalent semivariogram is typically used. The semivariogram is a measure of variability between two points separated by an offset of \mathbf{h} , linked to the covariance through $\gamma(\mathbf{h}) = C_M(0) - C_M(\mathbf{h})$. Deutsch and Journel (1998) provide a list of analytical covariance models. In this paper, the prior choice of covariance model is termed the true semivariogram/covariance model. Figure 1 also shows the covariance calculated from the reference model for checking purposes.

The covariance model can be made direction dependent and be anisotropic (different correlation length for different directions). It is only for the sake of simplicity that we use an isotropic covariance model [see Goovaerts (1997) for more details about covariance modeling].

Conventional SGS

The values in the left- and rightmost column at the boundaries of the reference field are assumed known values, mimicking well-log data. These data are referred to as well data. When all conditioning data are type A, traditional SGS algorithms can be used.

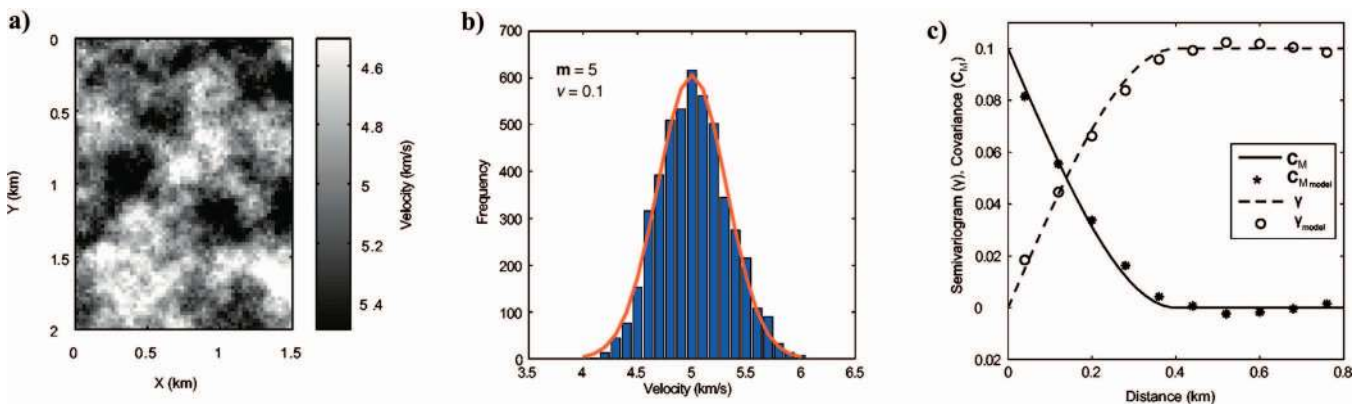


Figure 1. Reference model. (a) Reference velocity field. A realization of a Gaussian random field. (b) Histogram of the distribution of velocities in the reference model. (c) Synthetic covariance used to generate the reference model (solid line), and the covariance obtained from the reference model (stars). For comparison, the corresponding semivariogram is shown.

Figure 2 shows three conditional simulations, using the well data as conditioning data, and the true covariance model as the prior covariance function C_M . As can be seen (and expected), the three realizations vary significantly in the center of the model, where grid nodes are uncorrelated to the conditional data. All three simulations honor the basic statistics of the reference model in terms of mean, variance, covariance model, and histogram (not shown here).

Geostatistical tomography — Noise free

In addition to the borehole data, a number of rays traveling from one borehole to the other is now considered, mimicking a cross-borehole tomography setup. Thus, in this case, data of type B are also available. Sources are evenly spread along the left border of the left-most column, and receivers are evenly spread at the rightmost border of the rightmost column. Several models with varying ray density are examined by considering a range of selected sources and receivers. Figure 3 shows the ray geometry using two, five, and eight sources and receivers, respectively (the number of sources equals the number of receivers). All receivers are assumed to record signal from all sources; therefore, the number of rays is the square of the number of sources/receivers, i.e., 4, 25, and 64 rays. In this paper, we refer to the three types of data sets as a number of rays in square brackets, such as [4,25,64]. The data set with [0] rays refers to the case where only well data are available (no ray data). Ray tracing is used to calculate, from the reference field, the traveltime for each ray and the sensitivity of each cell in the model grid to each ray. From these data, the average velocity along the raypath is computed and is referred to as the observed data b_0 . These observed data can be seen as linear averages of the model space; hence, the developed theory can be applied. The observed data b_0 are for now considered noise free. Note that straight rays are considered for simplicity only. Any linear sensitivity kernel obtained using, for example, curvilinear rays or Fresnel zone-based sensitivity can be used.

Figure 4 shows three (out of 100) simulated fields for the three models with [4,25,64] rays. Figures 5 and 6 show vertical profiles of every fifth generated simulated fields, for [0,4,25,64] rays, at $x = 250$ m and $x = 750$ m (middle of x -axis range), respectively. Using four rays (of which only two actually cross the study area), no significant effect can be seen as compared to using no rays (Figures 4a, 5b, and 6b).

Using 25 rays, some features are resolved. The high-velocity feature (black) located around $x \approx 300$ m and $z \approx 1000$ m (Figure 1) can be identified on all three realizations of Figure 4b. This is even clearer from the series of vertical profiles at $x = 250$ m of Figure 5c, where most of the realizations include the high-velocity zone from $z \approx 0.8$ –1.3 km depth.

The vertical profiles using 64 rays (Figures 5d and 6d) show realizations that vary more tightly around the true profile.

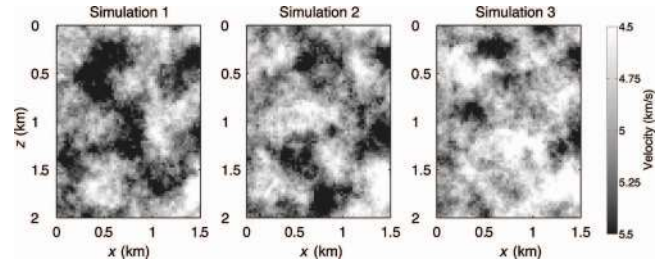


Figure 2. Three simulated fields, using the true covariance model as the a priori covariance model, and the well data (type A data) as conditioning data. No ray data are considered.

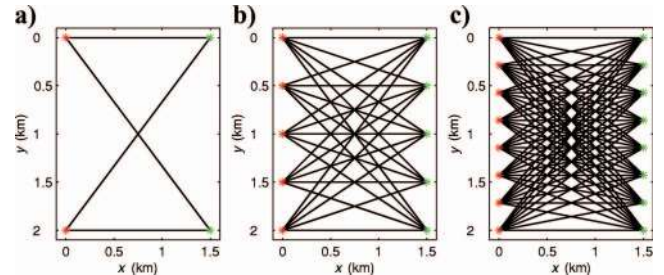


Figure 3. Ray coverage (lines) using (a) 4, (b) 25, and (c) 64 rays.

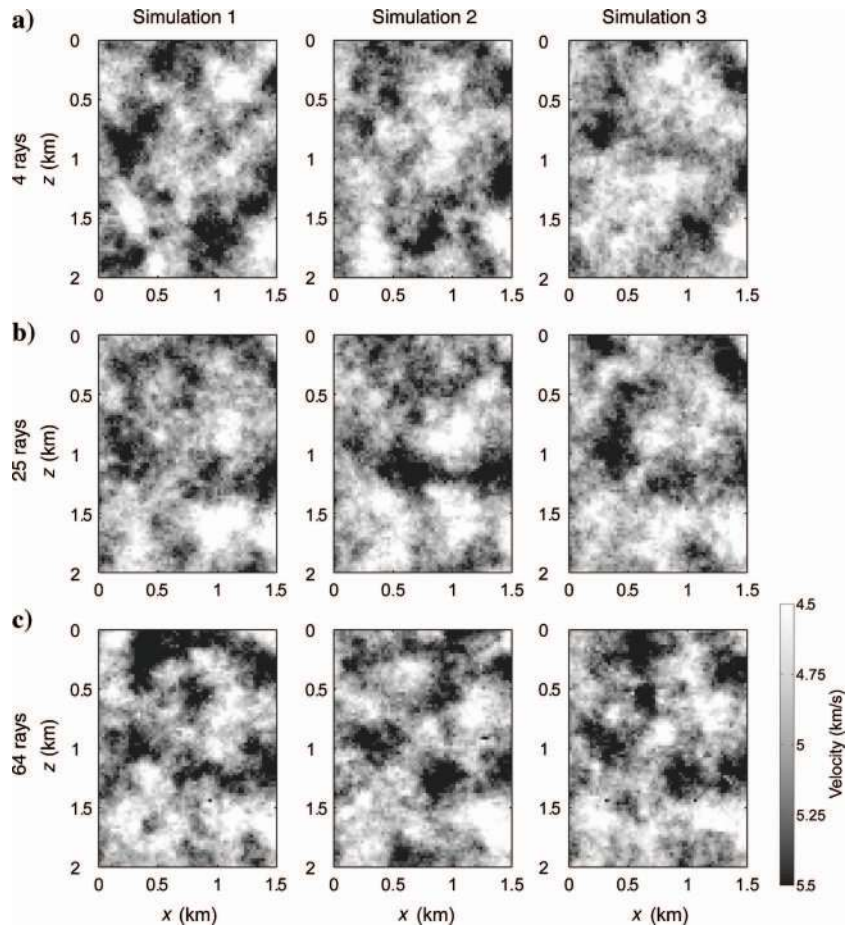


Figure 4. Three simulated fields conditioned to the well data (type A) and observed mean velocity along raypaths (type B) using (a) 4 rays, (b) 25 rays, and (c) 64 rays. The true covariance model was used as the a priori covariance to create all realizations.

Figure 7 shows some statistics of the generated simulated fields. The covariance calculated from the simulated fields match the reference model nicely (Figure 7a). The simulated fields show a Gaussian probability density distribution with the correct mean and variance (Figure 7b). Finally, all simulated fields result in apparent velocities along the raypaths that are close to the true values. The error is less than 0.1% and is from numerical round-off errors only.

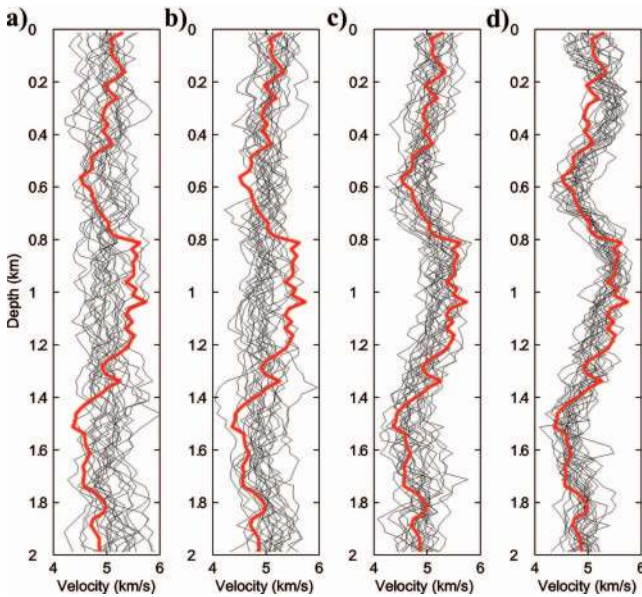


Figure 5. Vertical velocity profile at $x = 250$ m for every fifth generated realization using (a) no rays, (b) 4 rays, (c) 25 rays, and (d) 64 rays. The solid thick red line corresponds to the velocity profile of the reference model at $x = 250$ m.

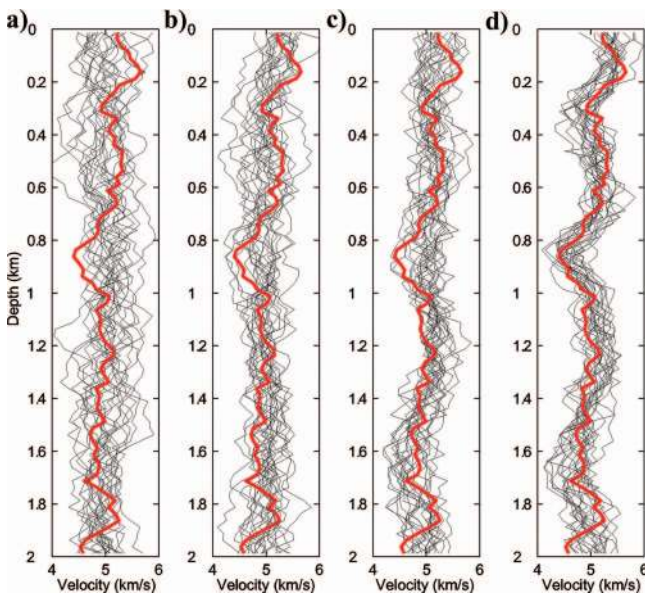


Figure 6. Vertical velocity profile at $x = 750$ m (the middle of the x -axis range) for every fifth generated realization using (a) no rays, (b) 4 rays, (c) 25 rays, and (d) 64 rays. The solid thick red line corresponds to the velocity profile of the reference model at $x = 750$ m.

Comparison to least-squares linear inversion

Figure 8 shows the E-type (the mean value of the conditional distribution at each location, calculated as the pointwise mean of all 100 realizations) for $[0, 4, 25, 64]$ rays. Figure 9 shows the least-squares estimation using equations 8 and 9 using the same true covariance.

The E-type and the least-squares result (Figures 8 and 9) should, in theory, give the same result if averaging is done on an infinite number of simulated realizations. As can be seen, the results of the two approaches are close to identical using only 100 simulated realizations.

Figure 10 shows the covariance function computed for the least-squares result conditioned to four rays, compared to the a priori assumed C_M and the mean covariance calculated from the 100 simulated realizations. The covariance estimated from the smooth least-squares result does not match the a priori chosen C_M lacking most of the higher-frequency variability. This is as expected from any least-squares estimated map and is known as the smoothing effect of estimation.

Figures 8–10 illustrate a very important point that is rarely recognized. The least-squares result (Figure 9) is a very unlikely sample of the posterior Gaussian random function. It is only an average of all solutions. A solution is one realization of the posterior Gaussian probability density field, as given by the sequential simulation approach shown in Figure 4.

In the case where a medium can be described by a Gaussian random field, as in the previous cross-borehole tomography example, sequential simulation can be used to analyze a general linear inverse problem through generation of multiple random field realizations that honor data **a** and **b** and prior information given by the covariance function C_M .

Effect of choice of prior covariance model

Figure 11 shows simulated realizations, conditioned to $[0, 4, 25, 64]$ rays and using a priori covariance models with a range of 200, 400, and 800 m. Because 400 m corresponds to the true range, ranges of 200 and 800 m correspond to situations where the selected covariance model differs from the true covariance model. Selecting a range of 200 m implies shorter spatial wavelengths, whereas a range of 800 m assumes longer spatial wavelengths, i.e., a smoother random field.

With no ray data included, the simulated fields directly reflect the choice of the prior covariance model. This is also the case when conditioning to only four ray observations. As more rays are used for conditioning, the change in the apparent correlation length (the average dominant distance between high and low features) of the simulated fields becomes less evident. The correlation length is consistently smaller for a range of 200 m as opposed to 400 m. Using an a priori range of 800 m shows another pattern. When few or no rays are included, the simulated field is relatively smooth (zero and four rays), indicating a long range. However, as more ray data are used for conditioning, the correlation length of the simulated fields diminishes, reflecting the actual and shorter correlation range (400 m) of the reference field.

This is verified by the experimental covariance calculated from the simulated fields. Experimental covariances for ranges of 200 and 400 m (Figures 7a and 12a) show that the experimental covariance model does not change as the amount of ray data increases. In fact, the experimental covariance fits the synthetic covariance of the a priori covariance model quite well.

The experimental covariance for an a priori range of 800 m (Figure 12b) shows that the range actually decreases toward the true range, as the number of rays increases. As expected, this shows that, with little or no ray data, the a priori information determines the unconstrained part of the model space. Increasing the number of conditioning data leads to increasing use of the data as opposed to the a priori information. The a priori information is honored only when it is consistent with the conditioning data.

In Figure 12c and d, the models with varying a priori range all showed a Gaussian-distributed velocity field with mean 5 km/s and variance 0.1 km/s. These results are similar to Figure 7b and c.

Noisy data

The simulations above assume noise-free data. In geostatistics, this is the usual way of dealing with so-called hard data. When considering geophysical data, one cannot avoid dealing with uncertainty of measurements. In inverse theory, it is well known that ignoring noise when trying to solve an inverse problem produces unreliable results because the data observations become inconsistent; no solution can be established that matches the observed data. In the present context, the sequential simulation process will fail. It produces simulations with extreme values and causes numerical matrix instability if forced to honor noisy data exactly.

As the theory is presented here, it is trivial to introduce data uncertainty through the data covariance matrix C_D in equation 6. Figure 13 shows the summary statistics of 100 simulations where Gaussian noise (zero mean and variance are 0.1 km/s) is added to the observed velocity average data d_{obs} . Uncorrelated data uncertainty is set to $C_{bb} = 0.1$ (off diagonal elements are set to zero), and well data are assumed noise free, $C_{aa} = 0$.

Comparing Figure 13 to the results of noise-free data (Figure 7), it is apparent that data d are reproduced by the simulated realizations within the uncertainty range specified by C_D because the histogram of $d_{obs} - d_{calculated}$ displays a Gaussian distribution with mean 0 and variance 0.1 km/s. The histogram of the simulated values at data locations can be approximated by a Gaussian distribution with mean 5.0 km/s and variance 0.1 km/s. Also, the experimental covariance of the simulated data matches the reference covariance reasonably well.

By allowing data uncertainty through C_D , we ensure that noisy data observations are fitted within their uncertainties, while honoring the a priori covariance information. Adding data uncertainty through C_D is referred to in geostati-

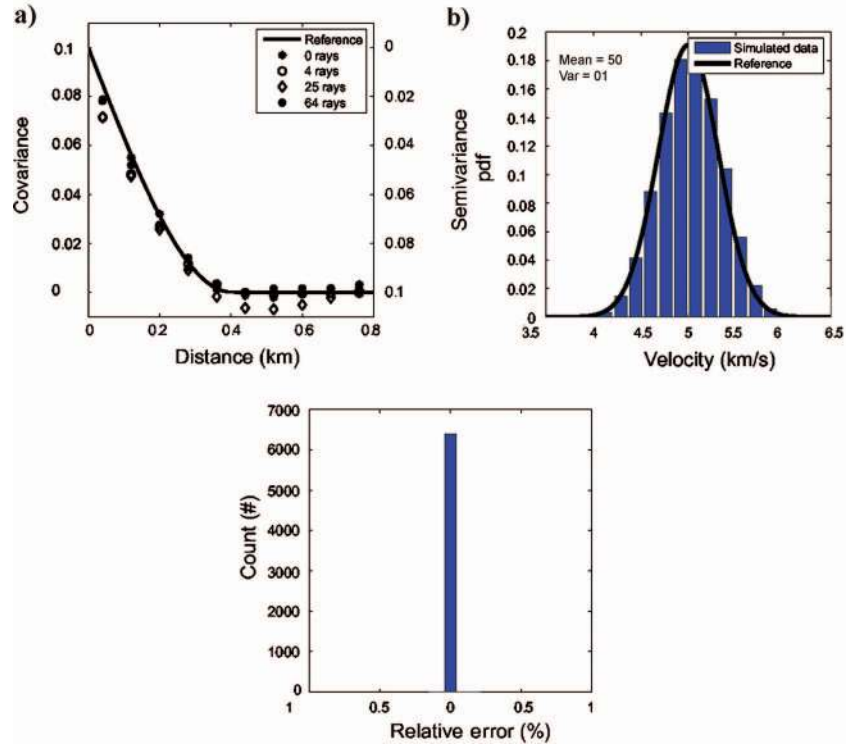


Figure 7. (a) True covariance, $C_M = 0.1$ Sph(0.4 km), and the mean of 100 experimental covariance models obtained from the conditional simulation shown in Figure 4. (b) Distribution of the velocity field for 100 simulations (using 64 rays) and the reference distribution used to set up the reference model. (c) Distribution of the relative error between the mean velocity along raypaths of 100 simulations (using 64 rays) and the true mean velocity along these raypaths.

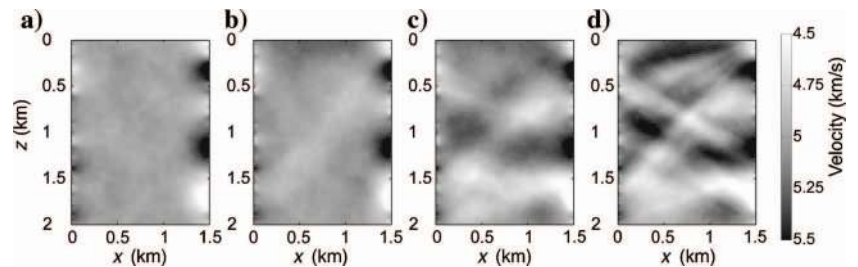


Figure 8. E-type (pointwise mean of 100 simulations) using (a) 0, (b) 4, (c) 25, and (d) 64 rays.

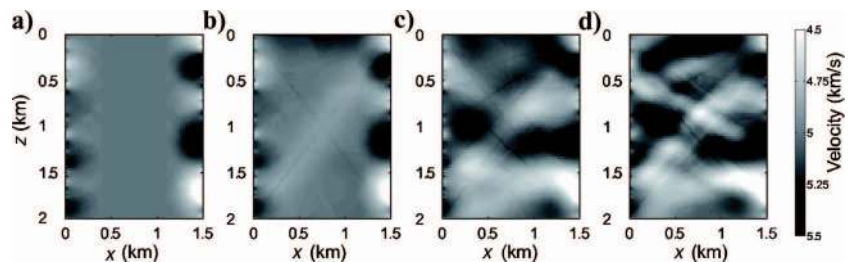


Figure 9. Least-squares estimates for (a) 0, (b) 4, (c) 25, and (d) 64 rays, using equations 8 and 9. Compare to the E-type results of the 100 simulated fields in Figure 8.

stics as specifying location-specific measurement error (Delhomme, 1978). This is identical to adding noise through C_D as suggested above.

SEQUENTIAL ESTIMATION

In case one is interested only in the smooth least-squares estimate, as in equations 8 and 9, the sequential approach also can be applied effectively. The approach to perform sequential estimation is only a slight variant to that of sequential simulation:

- 1) Visit a point in the model space, e.g., \mathbf{x}_i .
- 2) Compute the mean and covariance of the posterior Gaussian

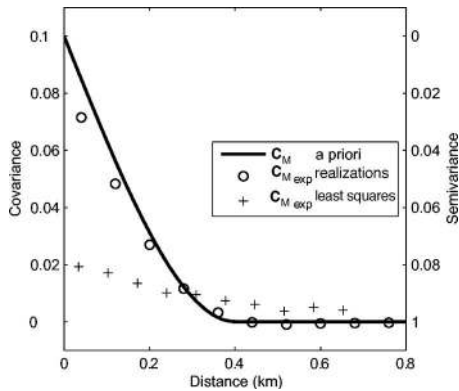


Figure 10. A priori covariance model and the covariance calculated from all realizations conditioned to four rays, as in Figure 4a, and the covariance calculated from the corresponding least-squares solution, i.e., the model shown in Figure 9b.

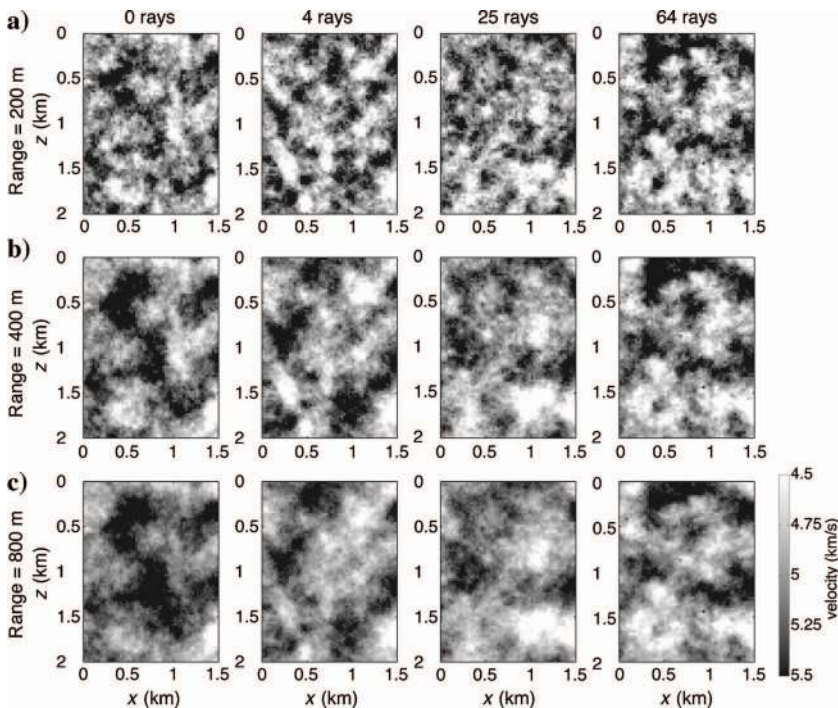


Figure 11. Simulated velocity field for a prior range of the covariance function of (a) 200 m, (b) 400 m, and (c) 800 m. For each row, the four simulated fields are the result of conditioning to [0, 4, 25, 64] rays, respectively.

probability of the model parameters $\tilde{m}(\mathbf{x}_i)$ and $\tilde{C}_M(\mathbf{x}, \mathbf{x}')$, which are given by equations 12 and 13.

- 3) Do not draw a value from the posterior Gaussian probability function. Do not add any data to the set of conditioning data (data of type A).
- 4) Repeat steps 1–3 until enough points have been estimated.

Thus, instead of drawing from a conditional probability function, as is the case in sequential simulation, one chooses the most likely model (the maximum likelihood model) at each location, through the means of the posterior probability function. This is also exactly what direct simple kriging would provide.

This approach will produce results identical to solving equations 8 and 9, which is the usual smooth least-squares estimate of a linear inverse problem. Solving equations 8 and 9 for all data locations at once requires the solution of one large linear system, whereas using the sequential approach requires the solution of many, but small, linear systems. For some linear problems, this sequential approach may prove computationally efficient.

We refer to this method as sequential least-squares estimation. It should not be confused with the sequential kriging method proposed by Vargas-Guzman and Yeh (1999), which is a kriging- (that is, least-squares) based estimation algorithm that allows data to be incorporated into the estimation process in an iterative manner as they become available while obtaining the same result as using all data at once.

APPLICATION TO WEAKLY NONLINEAR PROBLEMS

Instead of the linear problem of equation 7, consider

$$\mathbf{d} = \mathbf{G}(\mathbf{m}), \quad (14)$$

where \mathbf{G} is a weakly nonlinear function that maps model parameters \mathbf{m} into observations \mathbf{d} . By weakly nonlinear, we mean nonlinear problems where $\mathbf{G}(\mathbf{m})$ can be linearized around the prior model of \mathbf{m} , $\mathbf{m}_{\text{prior}}$. Such weakly nonlinear problems can be solved by a number of iterative techniques. Each step in the iterative process requires one to solve a linear system equivalent to equation 1 [see chapter 3.2.3 in Tarantola (2005)]. This can be done by a conventional least-squares inversion algorithm or by sequential least-squares estimation, both described above. The iteration continues until the estimated maximum likelihood model $\mathbf{m}_{\text{ML}}(\mathbf{x})$ has converged to a local optimal point.

The weakly nonlinear problem is now linearized around $\mathbf{G}(\mathbf{m}_{\text{ML}})$. Using the linearized kernel \mathbf{G} , sequential simulation can now be used to draw samples from the a posteriori distribution.

Thus, the sequential approach to linear inverse problems that we describe can be used to generate realizations of the posterior probability function of a Gaussian weakly nonlinear problem. In the iterative stage of the inversion, one applies sequential least-squares estimation until a local optimal

model has been found. Then sequential simulation is used to calculate samples of the posterior probability density function.

APPLICATION TO NON-GAUSSIAN DATA DISTRIBUTIONS

DSS is a generalization of SGS, which allows simulation of data with a non-Gaussian distribution. Any distribution that can be described by a mean and a variance can be simulated directly (Journel, 1993), using the exact same kriging system as shown here. The major limitation of DSS has been that, although the mean, variance, and covariance were reproduced by DSS, the histogram was not. This issue is addressed by Soares (2001) and Oz et al. (2003). They show a simple approach to ensure histogram reproduction. The only difference to SGS is that the shape of the conditional distribution from which a sample is drawn is modified from one simulated node to another to ensure histogram reproduction. Thus, the kriging system that is solved at each step of the simulation is identical to SGS. This generalization is important for the inclusion of linear average data, as we present. Indeed, a normal score transformation to make the data univariate Gaussian distributed would undo the linear average property and prevent their integration with a strict sequential Gaussian algorithm.

COMPUTATIONAL EFFICIENCY OF THE PROPOSED METHOD

Neighborhood

Using sequential simulation, the number of conditioning point data (data of type A) increases by one for each visited point in space. Therefore, the approach could be computationally inefficient because the linear system to be solved will keep increasing. However, as we previously noted, not all the conditioning data need to be used to solve the linear inverse problem at all points in space.

It is well known from geostatistics (Deutsch and Journel, 1998) that data much beyond the range of the semivariogram model/covariance model have little to no effect on the estimation of the mean and covariance of the posterior Gaussian probability. Typically only a small subset of available conditioning data are used.

Because we deal with linear inverse theory, only observations close, in terms of the covariance, to the point being simulated need to be considered. Thus, only those data of type B (and A) relatively close (e.g., two times to covariance range) need to be considered. We will refer to such data of types A and B as belonging to neighborhood A and B, respectively.

In addition, to ensure that observed data of type B are reproduced correctly, all previously simulated point data within the volume spanned by any type B data used must also be added to the conditioning data (Journel, 1999).

Comparison to least squares

Typically only a small fraction of the available data is used to calculate the posterior Gaussian probability at any given point in space. Therefore, sequential simulation can be seen as a process of solving many, but small, linear matrix equations, as compared to solving one large linear matrix equation when using conventional least-squares estimation (equations 8 and 9).

The size of the linear matrix equation to solve is important because it involves a matrix inversion. The time it takes to invert an $[N \times N]$ square matrix is approximately proportional to N^2 . Thus, solving very large linear inverse problems using equations 8 and 9 can become extremely inefficient, or impossible, not just because of the long computation time but also because of large physical memory requirements. In contrast, it may be feasible, for the same linear inverse problem, to use the sequential approach for solving inverse problems because it requires many more, but much smaller, matrices to be inverted.

Simulating only parts of the model space

The method proposed can be used to simulate regions smaller than that spanned by the linear average data.

Figure 14 shows the result of applying sequential simulation — simulation only at 1 km depth using 25 rays. All 25-ray data and all

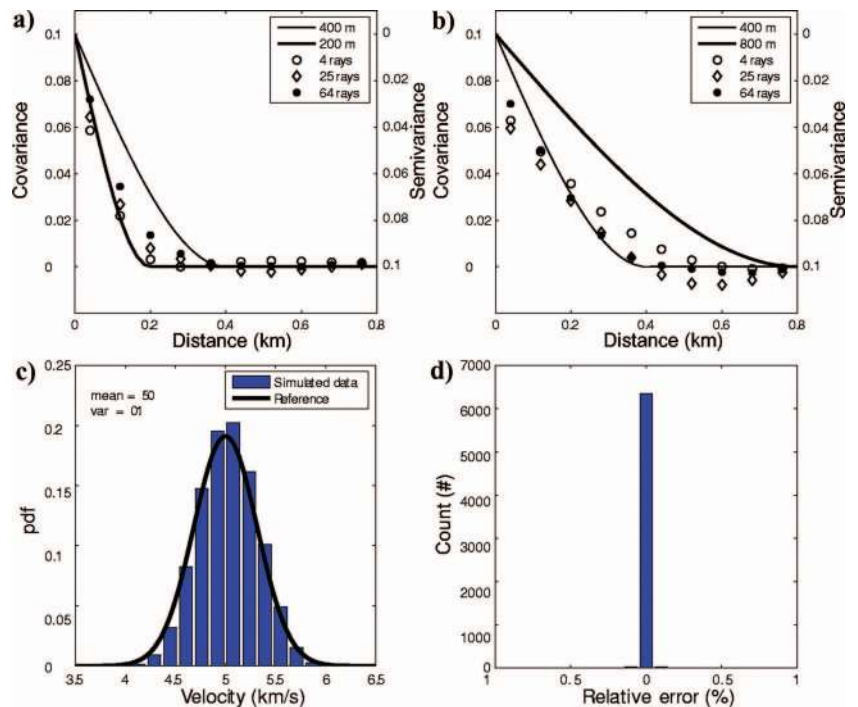


Figure 12. Similar to Figure 7a, but using a range of 800 m as opposed to the true range of 400 m. (a) Covariance models for $C_M = 0.1$ Sph(0.4 km) and $C_M = 0.1$ Sph(0.2 km), and the mean of 100 experimental covariance models obtained from conditional simulation using a prior range of 200 m. (b) Covariance models $C_M = 0.1$ Sph(0.4 km) and $C_M = 0.1$ Sph(0.8 km), and the mean of 100 experimental covariance models obtained from conditional simulation using a prior range of 800 m. (c) Distribution of the velocity field for 100 simulations (using 64 rays) and the reference distribution. (d) Distribution of relative error between the mean velocity along raypaths of 100 simulations (using 64 rays) and the true mean velocity along these raypaths.

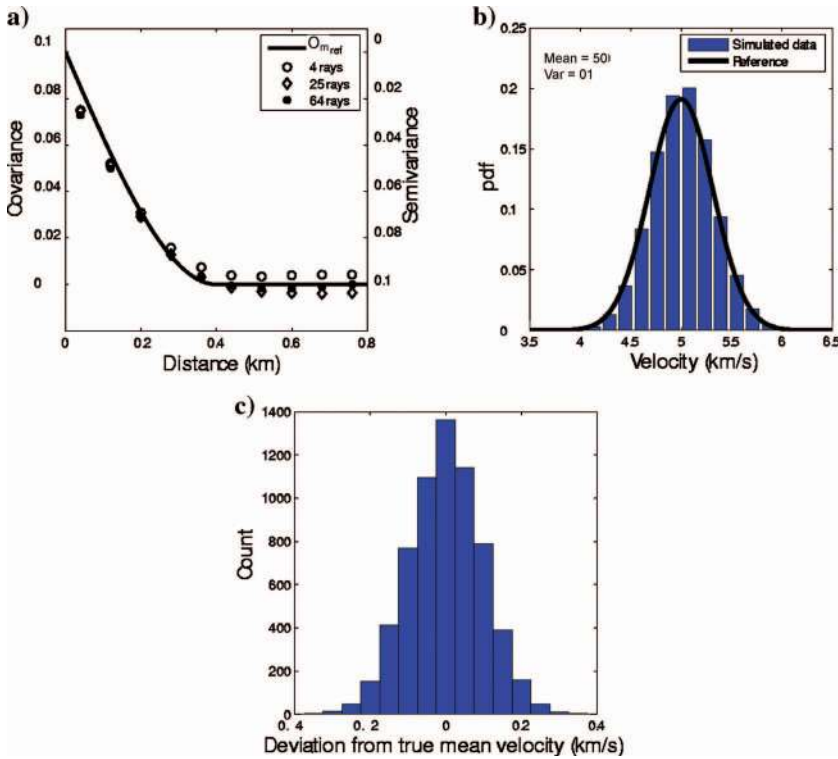


Figure 13. (a) Covariance model for $C_M = 0.1$ Sph(0.4 cm), and the mean of 100 experimental covariances calculated from conditional simulation using 4, 25, and 64 rays, contaminated with Gaussian noise with a variation of ± 0.1 km/s. (b) Distribution of the velocity field for 100 simulations (using 64 rays) and the reference distribution. (c) Relative variation of mean velocity along raypaths using 100 simulations (using 64 rays) compared to the true mean velocity along these raypaths. The variance is 0.0093 km/s, close to the specified data variance of 0.01 km/s.

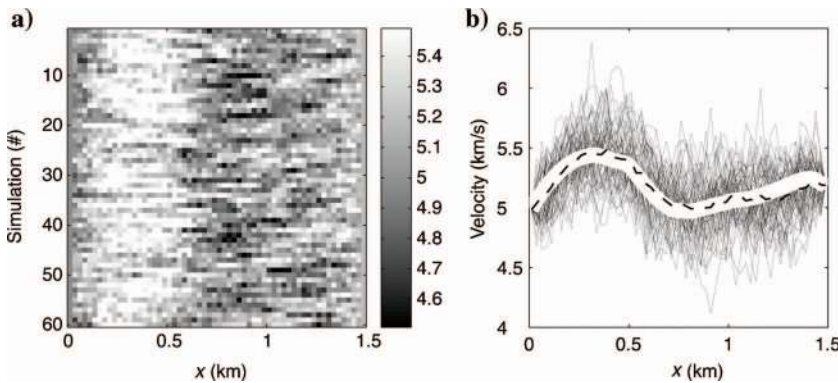


Figure 14. Result of simulating only at data locations at 1 km depth. (a) Sixty simulated rows at 1 km depth. (b) Black lines are 60 simulated velocity profiles. The thick white line is the result of a full least-squares inversion (all points in the model and all rays are considered) using equations 8 and 9, and the dashed black line is the E-type (pointwise mean) of the 60 simulations.

conditional well data are included in the simulation. As can be seen, the E-type matches the least-squares result quite well (Figure 14b).

Simulating 60 realizations at all points takes 529 s but only 18 s when visiting only the row at 1 km depth. The speedup of the code is approximately 30 times. Consider, for instance, a large 3D tomographic data set as a linear inverse problem, where one is only interested in a specific small part of the model space. Using conventional least-squares estimation, a very large matrix system must be inverted,

whereas using the sequential approach, one simply visits the points in space where one is interested in the model parameters, and simulates/estimates values based on relatively small (thus fast to solve) linear inverse problem.

CONCLUSION

We present the methodology to sequentially solve any Gaussian linear inverse problem, using the concept of sequential simulation. The major application is noniterative sequential simulation that will generate actual samples from the posterior probability density function, consistent with data observations and a priori information in form of a prior covariance model describing the spatial variability.

Compared to Monte Carlo-based approaches, the proposed approach is noniterative and hence computationally very efficient. The conventional least-squares inversion result gives the mean of all possible solutions, which makes it useful to infer only the posterior Gaussian probability density of one model parameter independent from other model parameters but useless to infer the joint posterior probability of the joint outcome of two or more model parameters. Using the sequential approach we present, one can generate a (large) number of samples from the posterior probability density function, which in fact are all solutions to the Gaussian linear problem. One can then calculate posterior statistics for any spatial model space feature. For example, one can get the answer to a question, such as, What is the probability that model parameters at locations x_1 and x_2 are connected by model parameters all larger than 8.1? Such questions and answers are crucial to, for example, flow-related property modeling.

We show that the sequential approach also can be used to efficiently solve any Gaussian linear inverse problem using sequential estimation. The combination of sequential estimation and simulation can be used to solve Gaussian weakly nonlinear inverse problems.

Linear inverse theory and geostatistics are closely related research fields. We have combined the sequential approach from geostatistics and data conditioning from inverse theory to bring new theory and application to both research fields. The theory presented in this paper is closely related to the geostatistical method of SGS, and can be seen as an extension of SGS to conditioning by linear averages.

Today geostatisticians use training images as a priori information of the spatial distribution, as opposed to the mere covariance model. Future work will emphasize bringing geostatistics and inverse theory closer together, where (1) inversionists will use realistic and complex a priori information used by geostatisticians in the form of training images and (2) geostatisticians will condition their methods di-

rectly to physical measurements, as opposed to considering such physical data as soft data.

The sequential simulation approach to solving linear inverse problems is applied to a ground-penetrating-radar cross-borehole tomographic problem.

ACKNOWLEDGMENTS

The authors would like to thank Yongshe Liu, Stanford University, for illuminating discussions and technical assistance. The authors also thank the associate editor and the reviewers for constructive comments and suggestions helping to clarify and finalize the paper.

REFERENCES

- Carr, J. R., and D. E. Myers, 1985, Cosim: A FORTRAN IV program for conditional simulation: *Computers and Geosciences*, **11**, 675–705.
- Delhomme, J. P., 1978, Kriging in the hydrosciences: *Advances in Water Resources*, **1**, 251–256.
- Deutsch, C. V., and A. G. Journel, 1998, *GSLIB, Geostatistical Software Library and User's Guide*, 2nd ed., Oxford University Press.
- Gloaguen, E., D. Marcotte, M. Chouteau, and H. Perroud, 2005, Borehole radar velocity inversion using cokriging and cosimulation: *Journal of Applied Geophysics*, **57**, 242–259.
- Gómez-Hernández, J., and E. F. Cassiraga, 2000, Sequential conditional simulations with linear constraints, in P. Monestiez, D. Allard, and R. Froidevaux, eds, *Geostatistics 2000* Cape Town: Geostatistical Society of Southern Africa, 999–1005.
- Goovaerts, P., 1997, *Geostatistics for natural resources evaluation*: Oxford University Press.
- Journel, A., 1993, Modeling uncertainty: Some conceptual thoughts, in R. Dimitrakopoulos, ed, *Geostatistics for the next century*: Kluwer, 30–43.
- Journel, A. G., 1999, Conditioning geostatistical operations to nonlinear volume averages: *Mathematical Geology*, **31**, 931–953.
- Journel, A. G., and C. J. Huijbregts, 1978, *Mining geostatistics*: Academic Press.
- Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: *Journal of Geophysical Research*, **100**, 12431–12447.
- Oz, B., C. V. Deutsch, T. T. Tran, and Y. Xie, 2003, DSSIM-HR: A FORTRAN 90 program for direct sequential simulation with histogram reproduction: *Computers and Geoscience*, **29**, 39–51.
- Phillips, W. S., and M. C. Fehler, 1991, Traveltime tomography: A comparison of popular methods: *Geophysics*, **56**, 1639–1649.
- Soares, A., 2001, Direct sequential simulation and cosimulation: *Mathematical Geology*, **33**, 911–926.
- Tarantola, A., 2005, *Inverse problem theory and methods for model parameter estimation*: Society for Industrial and Applied Mathematics.
- Vargas-Guzman, J. A., and T.-C. J. Yeh, 1999, Sequential kriging and cokriging: Two powerful geostatistical approaches: *Stochastic Environmental Research and Risk Assessment*, **13**, 416–435.