# 6

# Probabilistic Integration of Geo-Information

**Thomas Mejer Hansen, Knud Skou Cordua, Andrea Zunino, and Klaus Mosegaard**


## ABSTRACT

The problem of inferring information about the Earth can be described as a data integration problem, where the solution is a probability distribution that combines all available information. The theory is conceptually simple, but application in practice can be challenging. Probabilistic data integration requires that the information at hand can be quantified in the form of a probability distribution, either (a) directly through specification of an analytical description of a probability distribution or (b) indirectly through algorithms that can sample an often unknown probability distribution. Once all information has been quantified, efficient numerical algorithms are needed for inferring information from the combined probability distribution. In the following, methods for probabilistic characterization of different kinds of geo-information are presented. Then a number of methods that allow inferring information from the probability distribution that combines all available information will be discussed. Straight forward application of classic sampling algorithms such as the rejection sampler and the Metropolis algorithm will in most cases lead to computationally intractable problems. However, a number of methods exist that can turn an otherwise intractable data integration problem into a manageable one.


## 6.1. INTRODUCTION

A fundamental problem in Earth sciences is how to combine available information about the Earth (geo-information) into one consistent model of the subsurface. One difficulty is that the available information is of very different nature. Examples of geo-information are geophysical measurements, well logs, remote sensing, knowledge about geological processes, and so on. One commonly used approach to solve this problem is to make use of inverse problem theory. An inverse problem is typically defined as a problem where information about unknown parameters of a physical system are inferred from indirect physical measurements (see, e.g., *Tarantola* [2005]; *Mosegaard and Hansen* [2015]). The term "inverse problem" implies that one seeks to invert a process. For example, in a forward process some physical response from the Earth is measured in the form of some data. In the inverse process, an Earth model (or a collection of Earth models) explaining the data is sought. Alternatively, inferring information about the Earth can be considered as a problem of integration of information, where information from indirect information may, or may not, be available.

Let $I_1, I_2, \ldots, I_N$ represent $N$ different types of sources of information available about the Earth. Let the Earth be described by a set of $M$ model parameters $\mathbf{m} = [m_1, m_2, \ldots, m_M]$. In a probabilistic formulation, the information about $\mathbf{m}$, from a specific type of information $I_i$, can be quantified by a probability distribution $f(\mathbf{m} \mid I_i)$, and hence $N$ probability distributions $f(\mathbf{m} \mid I_1), f(\mathbf{m} \mid I_2), \ldots, f(\mathbf{m} \mid I_N)$ describe all the information available about $\mathbf{m}$.

If the information is independent—that is, if $f(\mathbf{I}) = f(I_1, I_2, \ldots, I_N) = f(I_1)\, f(I_2), \ldots, f(I_N)$—then $f(\mathbf{m}|I_1)$,

*Solid Earth Physics, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark*

*Integrated Imaging of the Earth: Theory and Applications, Geophysical Monograph 218*, First Edition.
Edited by Max Moorkamp, Peter G. Lelièvre, Niklas Linde, and Amir Khan.

$f(\mathbf{m},|I_2,)$, …, $f(\mathbf{m}|I_N)$ can be considered statistically independent, and then the combined information from all sources of information is given by the probability distribution

$$\begin{aligned} f(\mathbf{m}\,|\,\mathbf{I}) &= f(\mathbf{m}\,|\,I_1,I_2,\ldots,I_N) \\ &= f(\mathbf{m}\,|\,I_1)\,f(\mathbf{m}\,|\,I_2)\ldots f(\mathbf{m}\,|\,I_N) \\ &= \prod_{i=1}^{N} f(\mathbf{m}\,|\,I_i). \end{aligned} \tag{6.1}$$

The probabilistic formulation in Eq. (6.1) is similar to the concept of "conjunction of states of information" proposed as an approach to solve inverse problems [*Tarantola and Valette*, 1982; *Tarantola*, 2005]. $f(\mathbf{m}|I_i)$ represents one state of information. *Tarantola*, [2005] considers the conjunction of two states of information: the *a priori* probability and the theoretical probability density (given by a *likelihood* function). The conjunction of these two states of information is referred to as the a posteriori probability distribution. Here we intentionally avoid using the terms a priori and *likelihood* for several reasons. First, we argue that the two states of information (the prior and likelihood) simply represent different types of information about the model parameters, as described by, in this case, two probability distributions $f(\mathbf{m}|I_1)$ and $f(\mathbf{m}|I_2)$, respectively. Second, traditionally most focus in inverse problems has been on the information available from indirect information, related to geophysical data, while the use of a priori information has historically been debated. Some support the argument that the a priori probability distribution should be chosen as noninformative as possible [*Scales and Sneider*, 1997; *Buland and Omre,* 2003]. Others have argued that direct information about the model parameters may be of more value than geophysical data [*Journel*, 1994]. Here, in line with *Jaynes* [1984], we suggest to make use of whatever information, $f(\mathbf{m}|I_i)$, that is available (be it more or less informative) about the model parameters. The importance of each type of information is independent of the source of the information and is solely quantified by $f(\mathbf{m}|I_i)$.

Probabilistic data integration using Eq. (6.1) is conceptually very simple, namely an application of statistical independence. In practice, however, inferring information about $f(\mathbf{m}|\mathbf{I})$ may not be trivial. First, the information available has to be quantified probabilistically. This can be either in the form of an analytical description of $f(\mathbf{m}|I_i)$ (e.g., a normal distribution) or in the form of an algorithm that samples $f(\mathbf{m}|I_i)$ (e.g., a pruned partially ordered Markov mixture model as sampled by the SNESIM algorithm [*Strebelle*, 2002; *Cordua et al.*, 2015]). Next, it may be a challenge to computationally efficiently infer information from $f(\mathbf{m}|\mathbf{I})$, even in cases where a mathematical expression for $f(\mathbf{m}|\mathbf{I})$ exist. The computational complexity is highly linked to the method that is applied for inferring such information.

In the following, we will discuss methods and algorithms that allow probabilistic integration of geo-information, as given in Eq. (6.1), such that inference from $f(\mathbf{m}|\mathbf{I})$ is possible.

First, methods for quantifying different types of geo-information (information about the Earth) through probability distributions will be reviewed. Geo-information differs in the form in which it is available and we argue that it can, crudely, be divided into two categories: "direct" and "indirect" information about $\mathbf{m}$. Direct information allows characterizing the model parameters directly, which can be done using a variety of methods based on, for example, geostatistics, Markov models, and parsimonious model assumptions. We show examples on how to infer direct information from a sample model, related to a variety of types of statistical models (Section 6.2). We also recall how indirect information (e.g., where geophysical data provides information about some property related to the model parameters) can be quantified by data, measuring uncertainty and modeling errors (Section 6.3).

Then we discuss and compare a number of widely used sampling methods for inference of information from the probability distribution representing the combined information $f(\mathbf{m}|\mathbf{I})$. Specifically, we discuss how the numerical efficiency of such methods is strongly related to the type and amount of information available (Sections 6.4 and 6.5), and demonstrate this in a case study (Section 6.6). Finally, we discuss how the entropy related to different types of information affect the complexity of the data integration problem (Section 6.7).

Any knowledge, direct or indirect, about the model parameters $\mathbf{m}$ is conditional to a specific type of information $I_i$. Hence, the use of the notation $f(\mathbf{m}|I_i)$. However, for brevity, we will occasionally make use of the shorter notation $f_{I_i}(\mathbf{m}) = f(\mathbf{m}\,|\,I_i)$, and $f_{\mathbf{I}}(\mathbf{m}) = f(\mathbf{m}|\mathbf{I})$ in parts of the remainder of the text.

## 6.2. QUANTIFYING DIRECT GEO-INFORMATION USING PROBABILITY DISTRIBUTIONS

Working with geo-data, model parameters $\mathbf{m}$ typically describe an earth model, where each model parameter $m_i$ refers to a physical property, or geological unit, of a point or volume located somewhere in a three-dimensional space. When information about the model parameters $\mathbf{m}$ is available, it will be referred to as "direct" information about the model parameters (as opposed to indirect information that provide information related to the model parameters through some function $g(\mathbf{m})$). This

type of information must be quantified through the probability distribution $f(\mathbf{m}|I_{direct})$.

Direct information can, for example, refer to knowledge about the value a model parameter can take. Physical laws may impose restrictions on the values that specific types of model parameters can attain. For example, a velocity cannot be negative and cannot exceed the speed of light. Other types of direct information are rooted in geological knowledge. For example, knowledge about how the Earth has evolved in time can lead to some information about what type of structures that can and cannot be expected in the Earth. It can also give rise to information about what kind of geology that can be expected and, hence, information about **m**. Such information can be rooted in both observations, theoretical studies, and numerical simulation. Sometimes, a "sample model" may be available. A sample model is an example of (perhaps a part of) a realization from the unknown probability distribution $f(\mathbf{m}|I_{direct})$. This is the case when, for example, outcrops available at one location can be considered representative at another location; that is, the same probability distribution is expected to represent the same subsurface variability at the location of the sample model and at the unknown location. See, for example, *Holliger and Levander* [1994] for an example of using a sample model.

In the following a wide variety of types of probability distributions will be considered that allow characterizing $f(\mathbf{m}|I_{direct})$. They differ in the type of assumptions that is made regarding the statistical properties of $f(\mathbf{m}|I_{direct})$. Each choice of type of probability distribution requires a specific set of statistical properties in order to define the probability distribution (such as, for example, the mean and covariance for a Gaussian probability distribution). Also, different methods exist for generating realizations for each type of probability distribution.

Whether one makes use of a simple, high-entropy probability distribution, or a more complex, low-entropy type of probability distribution, the workflow of quantifying the available information is the same: (1) Select a type of probability distribution and (2) infer the properties, for example, from a sample model that defines this probability distribution. In that sense the only difference between probability distributions based on one-point, two-point, and multiple-point statistics is related to what type of statistics is taken into account.

### 6.2.1. Quantifying $f(\mathbf{m} \mid I_{direct})$

In the following, a number of methods for characterization of direct information, $f(\mathbf{m}|I_{direct})$, will be given, both in case an analytical expression of $f(\mathbf{m}|I_{direct})$ is assumed and in case it is unknown, but numerical algorithms exist that allow sampling from $f(\mathbf{m}|I_{direct})$.

#### *6.2.1.1. Probability Distributions Based on One-Point Statistics*

If we assume that the information for each individual model parameters $m_i$ is independent on other model parameters, then

$$\begin{aligned} f_I(\mathbf{m}) &= f_I(m_1, m_2, \ldots, m_M) \\ &= f_I(m_1) f_I(m_2) \ldots f_I(m_M) \\ &= \prod_i^M f_I(m_i). \end{aligned} \tag{6.2}$$

When **m** describes an Earth model, where each model parameter is related to a location in space, we say that the model parameters are spatially uncorrelated.

***Uniform Distribution.*** The most simple model for direct characterization of **m** is the uncorrelated uniform model. Assuming that all model parameters are independent and are uniformly distributed between $m_{min}$ and $m_{max}$, then $f_I(\mathbf{m})$ can be described using Eq. (6.2), where each 1D marginal distribution is given by

$$f_{I,\mathcal{U}}(m_i) = \begin{cases} \dfrac{1}{m_{max} - m_{min}} & \text{for } m_{min} \le m_i \le m_{max} \\ 0 & \text{else.} \end{cases} \tag{6.3}$$

The spatially uncorrelated uniform distribution is the distribution that provides least information and maximum entropy (i.e., maximum disorder) given only an upper and lower limit (which in the limit may tend to $-\infty$ to $\infty$) [*Shannon*, 1948]. If one wants to assume as little as possible about **m**, then the spatially uncorrelated uniform distribution $f_{I,\mathcal{U}}(m_i)$ is often suggested [*Scales and Sneider*, 1997; *Sambridge and Mosegaard*, 2002].

***Univariate Normal Distribution.*** Another type of maximum entropy model (given a mean and a variance) is the uncorrelated Gaussian model. $f_I(\mathbf{m})$is then described by Eq. (6.2), where each 1D marginal distribution is given by

$$f_{I,\mathcal{N}}(m_i) = \left(\sigma\sqrt{2\pi}\right)^{-1} \exp\left(-\frac{1}{2}\frac{(m_i - \mu)^2}{\sigma^2}\right), \tag{6.4}$$

where $\mu$ and $\sigma$ represent the mean and the standard deviation of the univariate normal distribution.

The assumption of spatial independence may be convenient in that using any of $f_{I,\mathcal{U}}(m_i)$ or $f_{I,\mathcal{N}}(m_i)$ leads to a probability distribution for which the probability distribution value can be easily evaluated. However, such simple, spatially uncorrelated model parameters may not allow realistic characterization of actual available information. The assumption of spatial independence implies that two model parameters located infinitely close together is assumed to be independent—an assumption

that in general may not be consistent with most natural phenomena, as these may display highly correlated features. Fortunately, a number of probability distributions and methods exist that allow describing spatially dependent model parameters, along with characterization of more geologically realistic structures.

#### 6.2.1.2. Probability Distributions Based on Two-Point (Gaussian) Statistics

In the special case where $f(\mathrm{m}|I_{direct})$ can be described fully by the mean and covariance between pairs of model parameters $m_i$ and $m_j$ and where $m_i$ is normally distributed, then $f_I(\mathbf{m})$ is a Gaussian probability distribution with mean $\mathrm{m}_0$ and covariance $\mathbf{C}_\mathrm{m}$ ($\mathcal{N}(\mathrm{m}_0,\mathbf{C}_\mathrm{m})$), which is given analytically by

$$f_I\left(\mathbf{m}\,|\,\mathbf{m}_0,\mathbf{C}_\mathrm{m}\right)=\left((2\pi)^M\left|\mathbf{C}_\mathrm{m}\right|\right)^{-.5}\times\exp\left(-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}_0\right)^\top\mathbf{C}_\mathrm{m}^{-1}\left(\mathbf{m}-\mathbf{m}_0\right)\right). \quad (6.5)$$

The Gaussian description of $f_I(\mathbf{m})$ given in Eq. (6.5) is mathematically convenient. However, the Gaussian probability distribution is also the probability distribution with maximum entropy of all probability distributions with a given mean $\mathbf{m}_0$ and covariance $\mathbf{C}_\mathrm{m}$. The multivariate Gaussian distribution maximizes spatial disorder such that the Gaussian choice of probability distribution is not able to describe more structured features such as, for example, channel structures [*Journel and Deutsch*, 1993].

A rather rich family of probability distributions, reflecting quite different spatial structures, can be obtained from simple operations on realizations of a Gaussian probability distribution [*Emery*, 2007; *Armstrong et al.*, 2011]. In addition, numerical algorithms exist that, based on Gaussian statistics, can generate realizations that expose non-Gaussian spatial features, such as indicator simulation [*Journel and Isaaks*, 1984] and direct sequential simulation [*Soares*, 2001]. Note that in these cases no analytical description of the underlying probability distribution $f(\mathbf{m}|I_{direct})$ may exist, but numerical methods exist that allow sampling from $f(\mathbf{m}|I_{direct})$.

#### 6.2.1.3. Probability Distributions Based on Multiple-Point Statistics

An alternative to using the Gaussian framework is to consider a probability distribution over $\mathbf{m}$ based on statistics that describes the (co)relation between more than two model parameters at a time. This is known as probability distributions based on multiple-point statistics. In this case, $f(\mathbf{m}|I_{direct})$ cannot simply be described by the statistical variation between pairs of model parameters (as given in covariance-based probability distributions described above). Instead, the variation between multiple model parameters needs to be quantified. Usually, no parametric description exists to quantify such distributions. Instead, nonparametric distributions based on multiple-point statistics are obtained from sample models. For examples of methods that utilize multiple-point statistics see *Guardiano and Srivastava* [1993], *Tjelmeland and Besag* [1998], *Strebelle* [2002], *Mariethoz et al.* [2010], *Dimitrakopoulos et al.* [2010], *Lange et al.* [2012], *Mariethoz and Caers* [2014], and *Cordua et al.* [2015]

When $f_I(\mathbf{m})$ is based on multiple-point statistics, it is often a type of partially ordered Markov model (POMM) [*Cressie and Davidson*, 1998; *Cordua et al.*, 2015]:

$$f_{I,POMM}\left(\mathbf{m}\right)=\prod_{i=1}^{M}p\left(m_i\,|\,pa\left(m_i\right)\right). \quad (6.6)$$

$p(m_i|pa(m_i))$ is the conditional probability of $m_i$ given the so-called parents $pa(m_i)$ of the model parameter $m_i$, which are the model parameters that $m_i$ is conditional dependent on.

In practice, a realization of $f_{I,POMM}(\mathbf{m})$, Eq. (6.6), can be generated by sequentially visiting all model parameters (optionally in random order), while at each step generating a realization of $p(m_i|pa(m_i))$. Note, however, that in practice, $f_{I,POMM}(\mathbf{m})$ will change for different choices of simulation path [*Cordua et al.*, 2015]. Moreover, the distribution is also dependent on the individual outcome realizations because the algorithm that samples from this distribution will prune the number of parents [*Strebelle*, 2002]. If the random path used for simulation from a POMM Is chosen from a uniform distribution, the probability distribution being sampled is a so-called pruned mixture model (PMM) of partially ordered Markov models [*Cordua et al.*, 2015; *Daly*, 2005]:

$$f_{I,PMM}=\sum_{path}w^{path}f_{I,POMM}^{path}\left(\mathbf{m}\right), \quad (6.7)$$

where the weights are given as $w_{path}=\frac{1}{M!}$ for all paths and the sum is taken over all possible simulation paths ($M!$). $f_{I,PMM}(\mathbf{m})$ in Eq. (6.7) is, however, computationally intractable to obtain because the individual partial ordered Markov models depend on the pruning of the algorithm. This demands that the pruning related to all possible outcomes for all possible simulation paths have to be known in order to obtain an actual explicit mathematical expression of this probability distribution [*Cordua et al.*, 2015].

#### 6.2.1.4. Parsimonious/Trans-Dimensional Models

For the probability distributions considered previously, it has been assumed that the parameterization of $\mathbf{m}$ (i.e., the location and density of model parameters) has been chosen densely enough, as part of parameterization, to allow a realistic representation of spatial features [*Mosegaard and Hansen*, 2015].

However, one can choose to treat the number of model parameters as an unknown model parameter itself, which is referred to as using a trans-dimensional or parsimonious parameterization [*Constable et al*., 1987; *Malinverno*, 2002; *Bodin et al*., 2009].

*Malinverno* [2002] suggested a type of Monte Carlo-based inversion that allows, in their presented 1D case, the number of subsurface layers to vary. *Bodin et al*. [2009] explored this further and suggested a "self-parameterizing partition model" (trans-dimensional) approach that allows defining the subsurface using a number of basis functions, in this case exemplified using a number of Voronoi cells. In both studies, the number of layers/cells control the complexity of the subsurface. And, in both cases, algorithms are presented that allow randomly perturbing a subsurface model to update the number, location, and value of layers/cells.

A general formulation of a transdimensional description of the model parameters space, in form of $N_b$ basis functions can be given by (*Bodin et al*. [2009])

$$\mathbf{m} = \sum_{i=1}^{N_b} a_i B_i(\mathbf{x}), \tag{6.8}$$

where $B_i$ is a specific choice of kernel function, $\mathbf{x}$ is a location in space, and $a_i$ is its associated amplitude.

In the case where the basis function defines 2D Voronoi cells, then $\mathbf{m}$ can be completely characterized by the $3N_b$ parameters, $N_b$ values of each Voronoi cell, $\mathbf{v}_c$, as well as $N_b$ values for the $x$- and $y$-location for the center of each Voronoi cell, $\mathbf{x}_c$ and $\mathbf{y}_c$. A statistical model over these parameters can be given by $f_{I,V}(N_b, \mathbf{v}_c, \mathbf{x}_c, \mathbf{y}_c)$. For each realization of $f_{I,V}$, the value of any corresponding model parameter, $m_i$, regardless of the sampling density of the model parameters, can then be computed using Eq. (6.8).

Here we will simply consider the trans-dimensional model as a specific type of information about $\mathbf{m}$, for which no explicit description of $f(\mathbf{m}|I)$ may be given, but where algorithms exist to allow sampling $f(\mathbf{m}|I)$.

Note that in practice one will almost always implicitly make use of basis functions as part of parameterizing the model parameters. For example, when illustrating a set of model parameters, parameterized over a 2D grid, one tends to show this as an image of pixels, where each pixel reflect the value of one model parameter. This implies that each model parameter is assumed to reflect an average value within an area (as spanned by the pixel size) and not the value of a point. For a more detailed discussion on the implicit use of basis functions as part of parameterizing inverse problem, see *Mosegaard and Hansen* [2015].

### 6.2.2. Sampling from $f(\mathbf{m} \mid I_{direct})$

Many different methods exist that allow sampling (i.e., generating a sample of Earth models) from $f(\mathbf{m}|I_{direct})$. Here we pay special attention to sampling methods based on *sequential simulation*, as we shall later exploit some features of the sequential simulation approach that allow efficient sampling from $f(\mathbf{m} \mid I_1, I_2) \propto f(\mathbf{m} \mid I_1) f(\mathbf{m} \mid I_2)$ when either $f(\mathbf{m}|I_1)$ or $f(\mathbf{m}|I_2)$ can be sampled using sequential simulation.

#### *6.2.2.1. Sequential Simulation*

Sequential simulation is a method that can be used to generate a realization of a joint probability distribution $f(\mathbf{m}) = f(m_1, m_2, \ldots, m_M)$ in the case where the *conditional* distribution

$$f(m_i \mid \mathbf{m}_c) = f(m_i \mid m_1, m_2, \ldots, m_{i-1}) \tag{6.9}$$

can be evaluated for all sets of conditional model parameters $\mathbf{m}_c$. It is based on the product rule

$$f(\mathbf{m}) = f(m_1) f(m_2 \mid m_1) \prod_{k=3}^{M} f(m_k \mid m_1, m_2, \ldots, m_{k-1}). \tag{6.10}$$

A realization of $f(\mathbf{m})$ can be generated as $\mathbf{m}^*$ using the sequential simulation algorithm as follows:

**SEQUENTIAL SIMULATION**

Visit model parameter 1, $m_1$. Generate a realization $m_1^*$ of $f(m_1)$.

Visit model parameter 2, $m_2$. Generate a realization $m_2^*$ of $f(m_2 \mid m_1^*)$,

Visit model parameter 3, $m_3$. Generate a realization $m_3^*$ of $f(m_3 \mid m_1^*, m_2^*)$.

⋮

Visit model parameter $M$, $m_M$. Generate a realization $m_M^*$ of $f(m_M \mid m_1^*, m_2^*, \ldots, m_{M-1}^*)$.

Then $\mathbf{m}^* = \left[ m_1^*, m_2^*, \ldots, m_M^* \right]$ will be a realization of $f(\mathbf{m})$. The model parameters can be visited in any order, as long as all model parameters are eventually visited [*Gomez-Hernandez and Journel*, 1993].

At each step in the sequential simulation algorithm, one will typically compute the conditional distribution, $f(m_i|\mathbf{m}_c) = f(m_i|m_1, m_2, \ldots, m_{i-1})$, and then draw a realization of this distribution. Note, though, that in order to use sequential simulation, $f(m_i|\mathbf{m}_c)$ does not need to be explicitly computed. It is sufficient that a realization from $f(m_i|\mathbf{m}_c)$ can be generated.

In most practical applications of sequential simulation, it can be computationally difficult or impossible to

describe the full conditional distribution, Eq. (6.9). Instead, one can retain only a limited number of conditional model parameters—for example, based on proximity to the model parameter being simulated. This is referred to as using a "neighborhood" where the size of the neighborhood reflects the number of conditional model parameters. Such an application of sequential simulation will not sample the joint distribution exactly, but instead an approximation of it that can be described by a partially ordered Markov model (see *Cressie and Davidson* [1998]; *Cordua et al*. [2015]).

An early example of what can be seen as an application of sequential simulation using a neighborhood, is presented by *Shannon* [1948]. Here, sequential simulation is applied in order to simulate a sequence of English text character by character (based on a nonparametric probability distribution describing the occurrence of sets of characters inferred form on an English textbook used as a sample model). For each new character location a new character is simulated by generating a realization from a conditional distribution, conditional to a fixed number of preceding characters. The full conditional distribution is not computed. Instead, the first match, from a random starting point in the book (sample model) to the conditioning data, is chosen as a realization of the conditional distribution. This is equivalent to inferring the full conditional distribution from the sample model, followed by drawing a realization from the conditional distribution. More detailed descriptions of the theory and application of sequential simulation developed in geostatistical community can be found in, for example, *Gomez-Hernandez and Journel* [1993] and *Deutsch and Journel* [1998].

#### 6.2.2.2. Sequential Simulation of Probability Distributions Based on Two-Point (Gaussian) Statistics

If $f(\mathbf{m})$ is distributed according to a multivariate Gaussian model [see Eq. (6.5)], $f(\mathbf{m}) \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_m)$, then the conditional distribution, $f(m_i|\mathbf{m}_c)$, will be a 1D Gaussian distribution

$$f\left(m_i \mid m_1, m_2, \ldots, m_{i-1}\right) \sim \mathcal{N}\left(\mathbf{m}_0^*, \sigma^{2*}\right). \qquad (6.11)$$

The mean and the variance can be found by solving a simlpe kriging system, *Journel and Huijbregts* [1978], or equivalently by solving linear least squares system, *Hansen and Mosegaard* [2008]. Sequential simulation, based on Eq. (6.11) is also known as sequential Gaussian simulation, a widely used two-point statistical simulation algorithm, *Deutsch and Journel* [1998]; *Remy et al*. [2008].

Other variants of sequential Gaussian simulation are direct sequential simulation [*Soares*, 2001; *Oz et al*., 2003; *Hansen and Mosegaard*, 2008], sequential indicator simulation [*Caers*, 2000] and plurigaussian simulation [*Armstrong et al*., 2011].

#### 6.2.2.3. Sequential Simulation of Probability Distributions Based on Multiple-Point Statistics

Sequential simulation from a probability distribution based on multiple-point statistics such as the pruned mixture model based on partially ordered Markov models, Eq. (6.7) can be obtained through sequential simulation. The conditional distribution $f(m_i|\mathbf{m}_c)$ needed for sequential simulation is the term $p(m_i|pa(m_i))$ in Eq. (6.6). As noted previously, when $f(\mathbf{m})$ is based on multiple-point statistics, a parametric analytical description of both $f(\mathbf{m})$ and the conditional distribution $f(m_i|\mathbf{m}_c)$ are typically not provided and a nonparametric description of the joint distribution is computationally intractable to obtain [*Cordua et al*., 2015].

However a nonparametric formulation of the individual conditional distributions, $f(m_i|\mathbf{m}_c)$, needed for sequential simulation can be obtained directly from a sample model, most often in the form of a training image. A training image is a specific type of sample model (which in 2D is given by an image of pixels and in 3D by a cube of voxels) that represents realistic spatial variability. Such an image can, for example, be provided by a geological expert or from outcrops.

In the case where $f(\mathbf{m})$ represents a discrete probability distribution, *Guardiano and Srivastava* [1993] propose to scan the training image for a specific data event, as defined by the conditioning data, from which $f(m_i|\mathbf{m}_c)$ can be constructed. This is done at each step in the sequential simulation approach and, therefore, is computationally expensive.

*Strebelle* [2002] proposes to scan the training image only once for a large collection of data events and then store the result in a search tree. The conditional distribution $f(m_i|\mathbf{m}_c)$ can then be relatively efficiently obtained from the search tree during sequential simulation.

The direct sampling method [*Mariethoz et al*., 2010] essentially makes use of the approach proposed by *Shannon* [1948] described above. Here, the conditional distribution $f(m_i|\mathbf{m}_c)$ is never explicitly computed. Instead, a realization of $f(m_i|\mathbf{m}_c)$ is found by scanning the training image, from a random starting location, until the first matching data event is found (or within some tolerance).

These three methods represent different ways to generate a realization from $f(m_i|\mathbf{m}_c)$, and differ mostly in computational CPU and memory requirements. For an overview of related mulitple-point based sequential simulation sampling algorithms, see, for example, *Mariethoz and Caers* [2014].

#### 6.2.2.4. Sampling Methods not Based on Sequential Simulation

Many other types of methods, not based on sequential simulation, exist to generate realizations from probability distributions based on two-point or multiple-point statistics.

The fast Fourier transform moving average (FFT-MA) method is especially efficient for generating independent unconditional realizations of a stationary Gaussian distribution [*Le Ravalec et al.*, 2000]. Realizations can also be generated using LU decomposition of the covariance matrix. While this allows for a nonstationary covariance model, it is also computational inefficient for anything but very small models (see, e.g, *Deutsch and Journel* [1998]).

Realizations from probability distributions based on two-point or multiple-point statistics can also be obtained by locating models whose frequency distribution of patterns match the frequency distribution obtained from a sample model [*Peredo and Ortiz,* 2011; *Lange et al.*, 2012; *Cordua et al.*, 2015].

### 6.2.3. Quantifying $f(\mathbf{m}|I_{direct})$ from a Sample Model

A probability distribution describing direct information $f(\mathbf{m}|I_{direct})$ is almost never available directly. Instead the information may be available in form of a sample model, from which information about $f(\mathbf{m}|I_{direct})$ can be inferred.

Figure 6.1 shows an image of meandering sand channels (from *Strebelle* [2000]) that we will consider as an example of a sample model $\mathbf{m}_{sm}$. $\mathbf{m}_{sm}$ represents a 2D regular grid of electromagnetic wave velocity values, consisting of 125 × 125 cells with a cell distance of 0.15 m (the physical model is 18.75 m wide and deep). The sample model in Figure 6.1 only takes two values (0.11 m/ns and 0.13 m/ns). By assuming stationarity, the mean and standard deviation of all pixel values can be determined as $m_0 = 0.1155$ m/ns and $\sigma = 0.0089$ m/ns, respectively.

In the following, we will demonstrate how information from the sample model can be inferred and also used to characterize $f(\mathbf{m}|I_{direct})$ using the different type of probability distributions defined in the previous sections. For all considered cases, $f(\mathbf{m}|I_{direct})$ will describe a distribution over the model model parameters $\mathbf{m}$ which are spatially ordered in a 2D grid defined over 40 × 84 model parameters organized in a 2D grid, with a cell distance 0.15 m (5.85 m wide and 12.45 m deep). We shall later combine these different types of information with indirect information.

**$f(\mathbf{m}|I_{d1})$, Uncorrelated Gaussian.** A stationary probability distribution that describes uncorrelated Gaussian distributed model parameters is completely described by a mean and a variance, as given above. Figure 6.2a shows five realizations from such a Gaussian model where $f(m_i \mid I_{d1}) = \mathcal{N}(m_0, \sigma^2)$.

**$f(\mathbf{m} \mid I_{d2})$, Uncorrelated Binary Distribution.** The parameters of an uncorrelated probability distribution with a binary 1D marginal distribution can be inferred assuming stationarity, and considering the 1D marginal distribution of the training image as representative for all model parameters. This leads to $f(m_i = 0 \mid \mathbf{m}_{sm}) = 0.72$ and $f(m_i = 1 \mid \mathbf{m}_{sm}) = 0.28$. Figure 6.2b shows five realizations from such a skewed binary distribution, $f(\mathbf{m} \mid I_{d2})$.

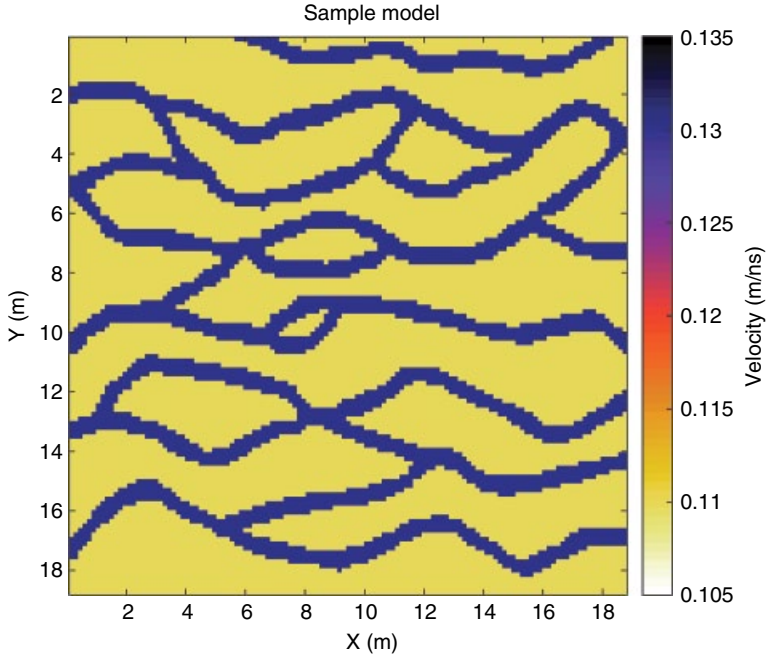


**Figure 6.1** Example of a sample model, $\mathbf{m}_{sm}$. From *Strebelle* [2000].

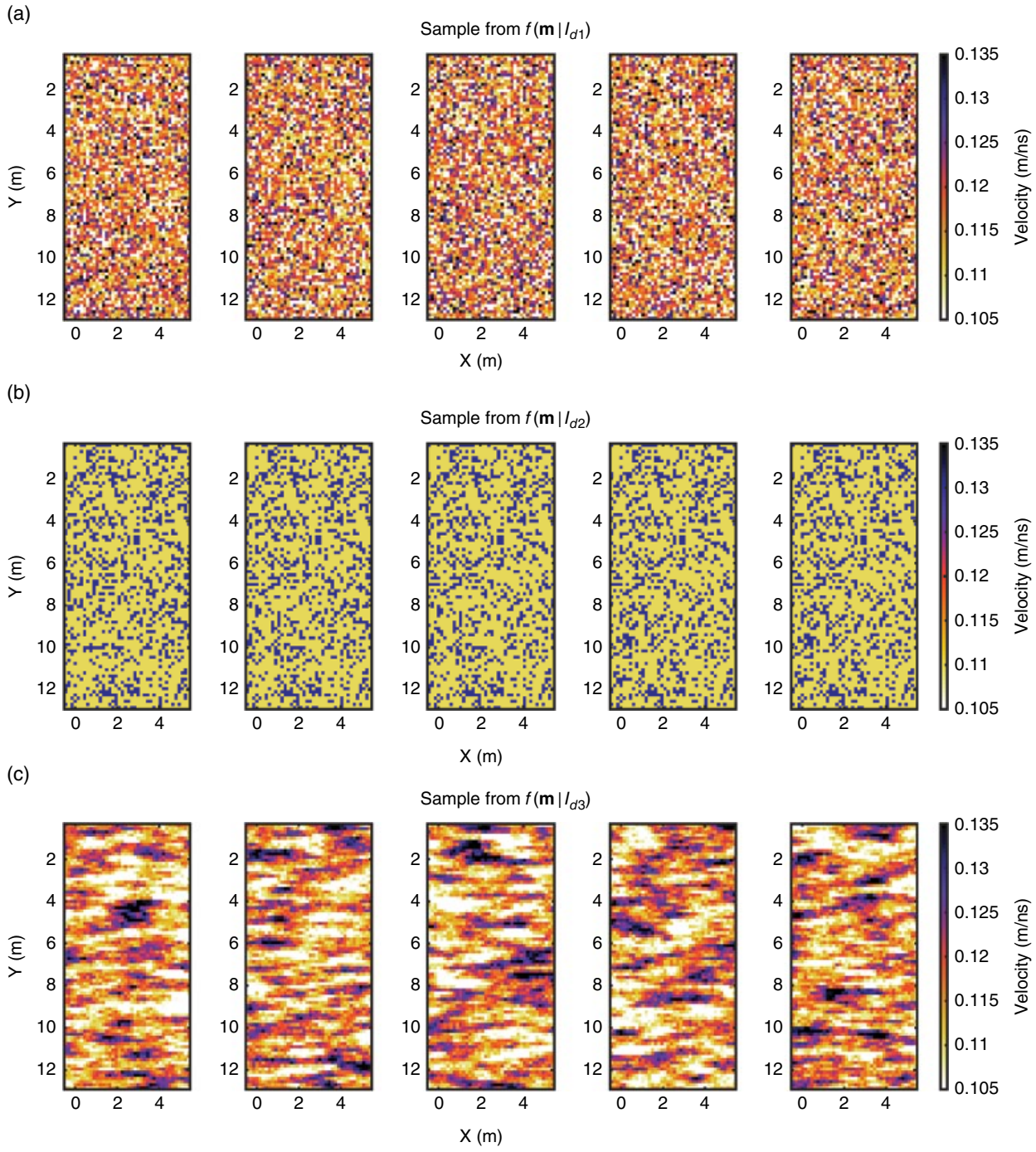


**Figure 6.2** Five realizations from **(a)** $f(\mathbf{m}\,|\,l_{d1})$, **(b)** $f(\mathbf{m}\,|\,l_{d2})$, and **(c)** $f(\mathbf{m}\,|\,l_{d3})$. See text for details.

$f(\mathbf{m}\,|\,l_{d3})$**, Correlated Gaussian Distribution.** A Gaussian probability distribution is completely described by a mean vector and a covariance matrix. Both the mean and the covariance (equivalent to a semi-variogram model) can be inferred from a sample model $\mathbf{m}_{sm}$. Figure 6.3 shows the experimental semi-variogram found from the sample model along the $x$- and $y$-axis, compared to the parametric semi-variogram model used to describe the experimental sem-variogram. From this model, a covariance matrix is constructed as $\mathbf{C}_{\mathrm{m}}$, such that $f(\mathbf{m}\,|\,l_{d3})$ can be described by the Gaussian distribution $\mathcal{N}(m_0,\mathbf{C}_{\mathrm{m}})$. Figure 6.2c shows five realizations from such a correlated Gaussian distribution.

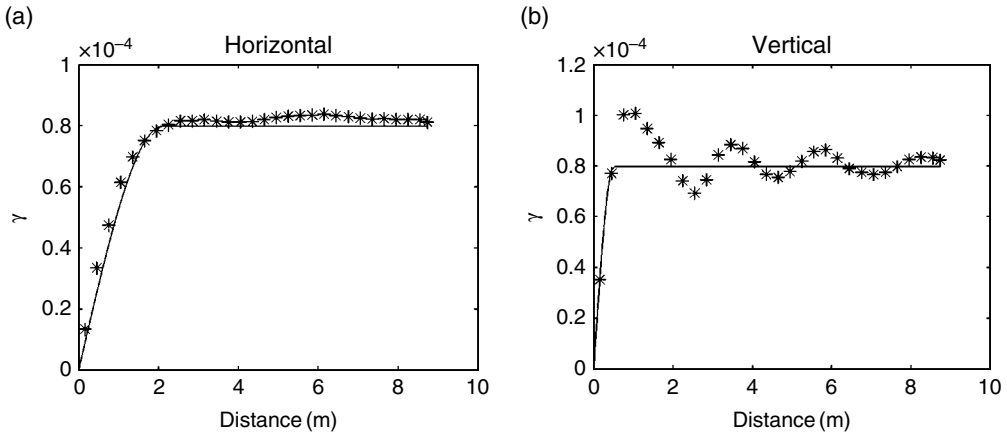


**Figure 6.3** Experimental semivariogram model inferred from the sample model in Figure 6.1 (*black asterisks*) compared to the semivariogram model chosen to represent the covariance model (*solid line*) of $f(\mathbf{m}\,|\,I_{d3})$ along the **(a)** horizontal axis and **(b)** vertical axis.

$f(\mathbf{m}\,|\,I_{d4})$**, Correlated Transformed Gaussian Distribution.** A simple way to simulate spatially correlated model parameters with an arbitrary non-Gaussian 1D marginal distribution is to apply an inverse normal score transformation to a realization from a Gaussian probability distribution. In the extreme case, a binary distribution, such as that of the distribution of the values in the sample model, can be assumed. This can also be obtained by truncating realizations from a Gaussian model. Figure 6.4a shows five realizations of such a probability distribution, $f(\mathbf{m}\,|\,I_{d4})$, reflecting the mean, covariance, and 1D marginal distribution obtained from $\mathbf{m}_{sm}$. The 2D covariance models used to describe the Gaussian distribution in the normal score space is chosen such that the experimental semi-variogram of the back-transformed realizations, along the $x$- and $y$-axis, reflects that of the sample model, as shown in Figure 6.5.

$f(\mathbf{m}\,|\,I_{d5})$**, Probability Distribution Based on Multiple-Point Statistics.** Fig. 6.4b shows five realizations generated using the SNESIM algorithm with $\mathbf{m}_{sm}$ as a training image [*Remy et al.*, 2008]. In this case, part of the multiple-point statistics of $\mathbf{m}_{sm}$ is consistent with the shown realizations. Strictly speaking, the SNESIM algorithm only samples from the same probability distribution if the same path is used and in the case where the neighborhood is kept constant [*Cordua et al.*, 2015]. But we will refer to these realizations as realizations from $f(\mathbf{m}\,|\,I_{d5})$.

$f(\mathbf{m}\,|\,I_{d6})$**, Voronoi Cells.** One could choose to describe information about $\mathbf{m}$ using the parsimonious approach and 2D Voronoi cells. Figure 6.4c shows a realization from, $f(\mathbf{m}\,|\,I_{d6})$, where the number of 2D Voronois cells $N_b$ is assumed uniformly distributed between 3 and 200. The $x$- and $y$-locations of the center of each cell is assumed to be located at a random location on the model parameter grid. The value of each Voronoi cell is assumed to be either 0.11 m/ns or 0.13 m/ns, with the same 1D marginal distribution as found in the sample model. Note that each realization in Figure 6.4c represents one set of parameters describing the Voronoi cells mapped into the exact same 2D $40 \times 84$ model parameter grid as for the other considered models of direct information. While the number of parameters that describe $f(\mathbf{m}\,|\,I_{d6})$ varies, the actual number of model parameters in $\mathbf{m}$ is fixed.

The probability distributions $f(\mathbf{m}\,|\,I_{d1}), \ldots, f(\mathbf{m}\,|\,I_{d5})$ are all consistent with the statistics from the sample model, in that $f(\mathbf{m}_{sm}\,|\,I) > 0$. In other words, part of the sample model the same size as $\mathbf{m}$ is possible as a realization of $f(\mathbf{m}\,|\,I_{d1}), \ldots, f(\mathbf{m}\,|\,I_{d5})$. This may not be the case for $f(\mathbf{m}\,|\,I_{d6})$. Further, $f(\mathbf{m}\,|\,I_{d1}) - f(\mathbf{m}\,|\,I_{d4})$ represent models with maximum disorder (maximum entropy) for the statistical correlations not specifically accounted for.

The main goal, when quantifying $f(\mathbf{m}\,|\,I_{direct})$, should be to define a probability distribution that has outcome realizations with the spatial (one-, two-, or multiple-point) statistics as obtained from the known sample model. Further, such a statistical model should represents the spatial structures as observed from an outcrop, or as known from geological expert knowledge. A simple way to validate the choice of probability distribution describing $\mathbf{m}_{sm}$ is to generate a set of independent realizations of $f(\mathbf{m}\,|\,I)$, as shown in Figures 6.2 and 6.4, and visually compare the realization to the sample model, Figure 6.1.

If the connectivity of the channel structures of the sample model, Figure 6.1, is an essential feature when characterizing the subsurface, then it may not be very useful to make use of the spatially uncorrelated models

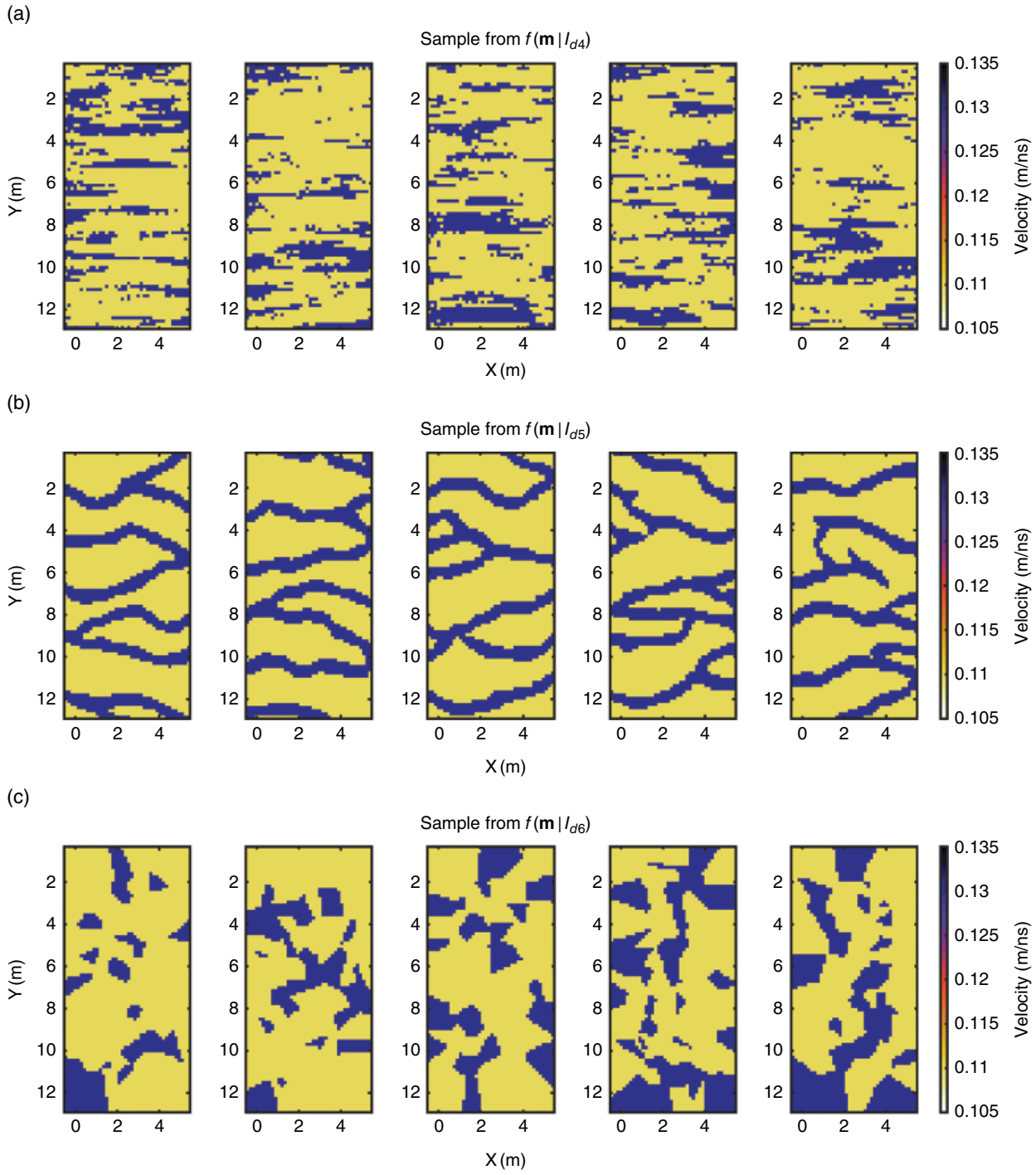


**Figure 6.4** Five realizations from **(a)** $f(\mathbf{m} \mid I_{d4})$, **(b)** $f(\mathbf{m} \mid I_{d5})$, and **(c)** $f(\mathbf{m} \mid I_{d6})$. See text for details.

$f(\mathbf{m} \mid I_{d1})$ and $f(\mathbf{m} \mid I_{d2})$ even if they may be consistent with some statistical properties of the sample model, as discussed by *Journel and Deutsch* [1993]. While the model based on Voronoi cells, $f(\mathbf{m} \mid I_{d6})$, may possess some features that can be mathematically useful, it is also evident from Figure 6.4c that such a model does not seem particularly useful to describe natural geological variability. Note that when the number of Voronoi cells becomes very high, $f(\mathbf{m} \mid I_{d6})$ may reflect the same kind of information as $f(\mathbf{m} \mid I_{d2})$.

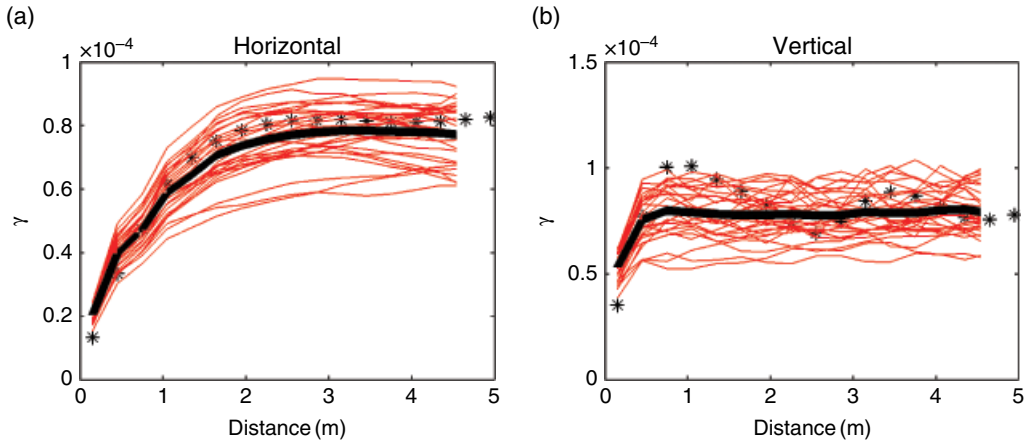


**Figure 6.5** Experimental semivariogram model inferred from the sample model in Figure 6.1 (*black asterisks*) compared to the experimental semivariogram of 10 realizations of $f(\mathbf{m} \mid I_{d4})$ (*red lines*) along the **(a)** horizontal axis and **(b)** vertical axis. *Black line* indicates the mean of the 10 semivariograms.

## 6.3. QUANTIFYING INDIRECT GEO-INFORMATION USING PROBABILITY DISTRIBUTIONS

As opposed to direct information, indirect information, $I_{indirect}$, is available in the form of data, $\mathbf{d}$, that is related to the model parameters $\mathbf{m}$ through a function $g$ as

$$\mathbf{d} = g(\mathbf{m}). \tag{6.12}$$

Evaluating Eq. (6.12) is often referred to as solving the forward problem. Examples of such indirect information are geophysical or remote sensing data.

A general probabilistic description of the relative probability of a certain model $\mathbf{m}$ given such indirectly observed data is the likelihood function [*Tarantola,* 2005]:

$$f(\mathbf{m} \mid I_{indirect}) \propto L(\mathbf{m}) = \int_{\mathcal{D}} \frac{\rho_D(g(\mathbf{m}))\theta(\mathbf{d} \mid \mathbf{m})}{\mu_D(\mathbf{d})}. \tag{6.13}$$

$\rho_D(\mathbf{d})$ describes *measurement uncertainties*, typically related to the instrument recording the data. $\theta(\mathbf{d} \mid \mathbf{m})$ is a probabilistic formulation of the forward modeling that describes the probability of a set of calculated data given a model $\mathbf{m}$. $\mu_D(\mathbf{d})$ is the homogeneous probability distribution (see *Tarantola* [2005] for more details).

The uncertainty related to the forward modeling may be significant and higher than the measurement uncertainty [*Hansen et al*., 2014]. However, in many cases the modeling error is ignored (i.e., described by a delta function) in which case Eq. (6.13) reduces to

$$L(\mathbf{m}) = \rho_D(g(\mathbf{m})) \tag{6.14}$$

In this case, evaluation of $f(\mathbf{m} \mid I)$ can be achieved as long as a probability distribution describing the measurement uncertainty can be evaluated. Very often the measurement errors are considered zero mean Gaussian distributed $\mathcal{N}(0, \mathbf{C}_d)$, in which case

$$\rho_D(g(\mathbf{m})) = \left((2\pi)^2 |\mathbf{C}_d|\right)^{-.5} \times \exp\left(-\frac{1}{2}\left(\mathbf{d}_{obs} - g(\mathbf{m})^\top\right)\mathbf{C}_d^{-1}\left(\mathbf{d}_{obs} - g(\mathbf{m})\right)\right). \tag{6.15}$$

If the modeling error is Gaussian, it can be described simply as an addition to the Gaussian measurement uncertainty and can then be accounted for through Eq. (6.15). More details on this topic can be found in *Hansen et al*. [2014].

Thus, in the latter simple case, the conditional probability $f(\mathbf{m} \mid I_{indirect})$ can be evaluated through Eq. (6.15) by solving the forward problem, Eq. (6.12), and evaluating the resulting data residual, $\mathbf{d}_{obs} - g(\mathbf{m})$.

The forward relation, Eq. (6.12), may be quite complex and involve mapping of the model parameters $\mathbf{m}$ into secondary parameters from which data can be computed. For example, seismic inversion can be formulated such that the primary model parameters reflect rock physical parameters. These must be transformed, for example, to elastic parameters in order to solve the forward problem in order to compute a seismic response. For a detailed discussion on complex forward models see *Bosch* [2015].

Note that the likelihood function, Eq. (6.13), is not strictly a probability distribution, as $\int L(\mathbf{m})\, d\mathbf{m}$ in general will not be 1. However, if the goal is to sample from $f(\mathbf{m} \mid I_1, I_2) = f(\mathbf{m} \mid I_1) f(\mathbf{m} \mid I_2)$, then a relative measure

proportional to $f(\mathbf{m} \mid I_{indirect})$, such as the likelihood, will suffice, and hence the normalization of the likelihood is not needed.

## 6.4. SAMPLING FROM A PROBABILITY DISTRIBUTION, $f(\mathbf{m} \mid \mathbf{I})$

In the ideal case, $f(\mathbf{m} \mid \mathbf{I})$—which describes all available information about $\mathbf{m}$—can be described analytically. In practice, however, this may not be possible, unless restrictions, such as Gaussian assumptions, are imposed on $f(\mathbf{m} \mid I_i)$.

A general approach for characterizing $f(\mathbf{m} \mid \mathbf{I})$ is by *sampling* it, which is done by generating a (representative) sample from $f(\mathbf{m} \mid \mathbf{I})$ that consists of a number of realizations that are distributed according to $f(\mathbf{m} \mid \mathbf{I})$. If this sample is large enough, any statistical measure or question related to $f(\mathbf{m} \mid \mathbf{I})$ can be probabilistically evaluated and answered.

In this section, a number of widely used methods for sampling from $f(\mathbf{m} \mid \mathbf{I})$ when a measure proportional to $f(\mathbf{m} \mid \mathbf{I})$ can be evaluated will be described. This implies that a measure proportional to the probability distribution value related to any of the independent types of information $f(\mathbf{m} \mid I_i)$ can be computed. This means that the previous defined models $f(\mathbf{m} \mid I_{d1})$, $f(\mathbf{m} \mid I_{d2})$, and $f(\mathbf{m} \mid I_{d3})$ can all readily be used. On the other hand, information available and quantified through numerical simulation algorithms, such as $f(\mathbf{m} \mid I_{d4})$, $f(\mathbf{m} \mid I_{d5})$, and $f(\mathbf{m} \mid I_{d6})$ where $f(\mathbf{m} \mid I_i)$ cannot readily be computed, cannot be considered by the algorithms discussed in this section.

### 6.4.1. Rejection Sampling

Any probability distribution for which $f(\mathbf{m})$ can be evaluated can in principle be sampled using rejection sampling. If $h(\mathbf{m})$ is a *proposal* distribution from which a realization can be generated (preferably in a computationally efficient manner) and for which $h(\mathbf{m}) \geq f(\mathbf{m})\ \forall \mathbf{m}$, then $f(\mathbf{m})$ can be sampled using the rejection sampling algorithm as follows:

**REJECTION SAMPLING ALGORITHM**

1. **Propose** a model $\mathbf{m}_{propose}$, as a realization of $h(\mathbf{m})$.
2. **Accept** this model with probability $P_{acc}$

$$P_{acc} = \frac{f\left(\mathbf{m}_{propose}\right)}{h\left(\mathbf{m}_{propose}\right)\max\left(f\left(\mathbf{m}\right)\right)}, \tag{6.16}$$

where $\max(f(\mathbf{m}))$ is the maximum value of $f(\mathbf{m})$.
Each accepted model will be a realization from $f(\mathbf{m})$, and the series of models accepted when the algorithm is run iteratively will be a representative sample from $f(\mathbf{m})$.

$h(\mathbf{m})$ is often chosen as the uniform distribution, in which case the acceptance probability becomes

$$P_{acc} = \frac{f\left(\mathbf{m}_{propose}\right)}{\max\left(f\left(\mathbf{m}\right)\right)}, \tag{6.17}$$

In many cases it may not be possible to estimate $\max(f(\mathbf{m}))$. Further, even in cases where $\max(f(\mathbf{m}))$ can be evaluated, the acceptance probability of the rejection sampler may be extremely low. Consider, for example, the Gaussian model [as in Eq. (6.5)] where the value

$$-2\ \log\left(f\left(\mathbf{m}^*\right)\right) \propto \left(\mathbf{m}^* - \mathbf{m}_0\right)^{\top}\ \mathbf{C}_{\mathrm{m}}^{-1}\ \left(\mathbf{m}^* - \mathbf{m}_0\right), \tag{6.18}$$

related to a realization $\mathbf{m}^*$, is distributed according to the $\chi^2$ distribution with $M$ degrees of freedom (where $M$ is the number of parameters of $\mathbf{m}$) [*Tarantola*, 2005]. For high values of $M$ the $\chi^2$ distribution will tend to be Gaussian distributed as $\mathcal{N}(M, 2M)$. This means that for high values of $M$, $\log(f(\mathbf{m}^*))$ will tend to be normally distributed as $\mathcal{N}\left(-M/2, \sqrt{M/2}\right)$. In other words, the most frequent probability value of a realization $\mathbf{m}^*$ of $f(\mathbf{m})$ will be $f(\mathbf{m}^*) \approx \exp(-M/2)$.

Considering $M = 10$, model parameters will lead to $f(\mathbf{m}^*) \approx \exp(-10/2) = 0.0067$. This means that in order to accept a typical realization $\mathbf{m}^*$ of $f(\mathbf{m})$ using the rejection sampler, it has to be proposed on average 1/0.0067 = 148 times. Considering $M = 20$ model parameters will lead to $f(\mathbf{m}^*) \approx \exp(-20/2) = 0.000045$, which means that in order to accept a typical realization $\mathbf{m}^*$ from $f(\mathbf{m})$ using the rejection sampler, it has to be proposed on average 22,026 times.

Thus, the rejection sampler, with a uniform proposal distribution, is extremely inefficient except for very-low (less than about five)-dimensional problems.

### 6.4.2. Metropolis–Hastings Algorithm

The Metropolis–Hastings algorithm is a Monte Carlo–based method for sampling a probability distribution $f(\mathbf{m})$ [*Metropolis et al.*, 1953]. At each step in a random walk the algorithm goes through two phases. In the "exploration" phase a new model is proposed in the vicinity of a current model. Then, in an "exploitation" phase the new model is either accepted or rejected as a realization from $f(\mathbf{m})$ as follows:

**THE METROPOLIS–HASTINGS ALGORITHM**

0. Generate a starting model, $\mathbf{m}_{current}$.
1. **Exploration.** Propose a new realization from $\mathbf{m}_{propose}$, in the vicinity of $\mathbf{m}_{current}$ by generating a realization from a transition probability $h(\mathbf{m}_{proposed} \mid \mathbf{m}_{current})$.

2. **Exploitation.** Accept the move to $\mathbf{m}_{propose}$ with the acceptance probability $P_{acc}$:

$$P_{acc} = \min\left(1, \frac{f\left(\mathbf{m}_{propose}\right) h\left(\mathbf{m}_{current} \mid \mathbf{m}_{propose}\right)}{f\left(\mathbf{m}_{current}\right) h\left(\mathbf{m}_{propose} \mid \mathbf{m}_{current}\right)}\right). \quad (6.19)$$

For simplicity it is often assumed that the proposal distribution is symmetrical such that $h(\mathbf{m}_{propose} \mid \mathbf{m}_{current}) = h(\mathbf{m}_{current} \mid \mathbf{m}_{propose})$. In this case, Eq. (6.19) reduces to

$$P_{acc} = \min\left(1, \frac{f\left(\mathbf{m}_{propose}\right)}{f\left(\mathbf{m}_{current}\right)}\right). \quad (6.20)$$

If the move is accepted, $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{current}$. Otherwise the random walk stays at the $\mathbf{m}_{current}$ location.

3. Goto 1.

It can be shown that $f(\mathbf{m})$ will be asymptotically sampled by running the Metropolis–Hasting algorithm.

Often a new model is proposed using a uniform or Gaussian transition distribution centered on $\mathbf{m}_{current}$. In such a case the exploration step simply consists of adding a realization of a uniform or Gaussian model, $\mathbf{m}_{\delta}$, with mean zero, to the current model, such that $\mathbf{m}_{proposed} = \mathbf{m}_{current} + \mathbf{m}_{\delta}$. The amplitude of $\mathbf{m}_{\delta}$ is referred to as the step-length.

One major advantage using the Metropolis–Hastings algorithm—as opposed to, for example, rejection sampling—is that this algorithm only relies on the relative change in probability value between the current and the proposed model for computing the acceptance probability, Eq. (6.19). Therefore the value of $\max(f(\mathbf{m}))$ does not need to be known as is the case using the rejection sampler.

A disadvantage is that the series of realizations generated by the Metropolis–Hastings algorithm are not independent. Thus, in order to obtain a statistical independent realization from $f(\mathbf{m})$, a number of iterations of the algorithm must be run. It may not be trivial to estimate how many iterations are needed in order to obtain an independent realization.

In addition, when the Metropolis–Hastings algorithm is started, it will, most often, not sample $f(\mathbf{m})$ immediately. Initially the algorithm will be in what is referred as the "burn-in" phase, in which state the algorithm searching for models that are consistent with $f(\mathbf{m})$. When the algorithm starts to sample $f(\mathbf{m})$, it is said to have reached burn-in.

The average distance between $\mathbf{m}_{propose}$ and $\mathbf{m}_{current}$ is called the exploration step-length. A large exploration step results in a more exploratory algorithm spanning relatively large volumes of probability at the expense of increasing computational demands. It is nontrivial to choose an exploration step-length that leads to maximum efficiency of the Metropolis sampling algorithm. It has been suggested that an exploration step-length leading to an accepted move in every third to fourth iteration provides a good compromise between exploration and computational efficiency [*Geman and Geman*, 1984]. In practice an optimal choice of exploration step-length is closely linked to the shape of the probability distribution being sampled.

The Metropolis–Hasting algorithm is guaranteed to asymptotically sample $f(\mathbf{m})$ in finite time. In practice, however, the Metropolis–Hastings algorithm can have difficulties sampling multimodal problems in high dimensions (i.e., problems where local areas of high probability exist, which are disconnected by areas of zero probability). In such cases, it may end up sampling a local area of high probability. There are no trivial tests to ensure that the full probability distribution is being sampled. A simple approach is to start more sampling algorithms (sometimes called chains) in parallel and then test whether they end up sampling the same distribution. A more formal approach is to make use of parallel tempering, where multiple chains run in parallel, where jumps between chains are allowed. Each chain is run with a different temperature, as known from simulated annealing. Parallel tempering is promising for lower-dimensional problems [*Sambridge*, 2013].

For all its shortcomings, the Metropolis–Hastings algorithm is computationally superior to rejection sampling, for sampling anything but very-low-dimensional probability distributions.

The computational efficiency of the Metropolis-Hastings algorithm is closely related to the choice of transition probability. The efficiency of the rejection sampler is linked to the choice of proposal distribution. Ideally such transition probabilities and proposal distributions should be chosen such that the acceptance rate is maximized. However, this is often not a trivial task. For example, a straightforward application of the Metropolis algorithm to sample from a multivariate Gaussian probability distribution with a Gaussian-type covariance using a symmetric proposal distribution will in practice be computationally extremely inefficient. This is due to the fact that any proposed model will lead to a discontinuity in the proposed model, which is inconsistent with the (spatial) smoothness implied by the Gaussian-type covariance model. The Hamiltonian Monte Carlo approach suggests to make use of the local gradient of the probability distribution being sampled, in order to allow faster mixing and higher acceptance probability of the Monte Carlo Chain [*Duane et al.*, 1987]. However, the Hamiltonian Monte Carlo requires that the gradient of the probability distribution being sampled can be evaluated. In the following, we will consider to sample probability distributions where the probability distribution value, and hence the gradient, may not be available.

We will, therefore, not consider the use of the Hamiltonian Monte Carlo any further.

The rejection sampler and Metropolis–Hasting algorithm as described above will, in the following, be referred to as the "classic" rejection sampler, and the "classic" Metropolis algorithm.

## 6.5. SAMPLING OF $f(\mathbf{m}\,|\,I_1, I_2) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$

Consider the case where $f(\mathbf{m}\,|\,\mathrm{I})$ is proportional to the product of two probability densities $f(\mathbf{m}\,|\,I_1)$ and $f(\mathbf{m}\,|\,I_2)$:

$$f(\mathbf{m}\,|\,\mathrm{I}) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2), \tag{6.21}$$

that is, a case identical to the information integration problem in Eq. (6.1), where information is available from two independent sources. One can choose to sample directly from $f(\mathbf{m}\,|\,\mathbf{I})$, using the methods described in the previous section, in case $f(\mathbf{m}\,|\,\mathbf{I})$ can be evaluated.

But, in many data integration problems one may not be able to evaluate $f(\mathbf{m}\,|\,\mathbf{I})$, as not all $f(\mathbf{m}\,|\,I_i)$ can be evaluated. For example, if $f(\mathbf{m}\,|\,I_i)$ describes the statistical information inferred from a training image, then, in most cases, the evaluation of $f(\mathbf{m}\,|\,I_i)$ is, until now, not possible.

It turns out that when $f(\mathbf{m}\,|\,I_1)$ and $f(\mathbf{m}\,|\,I_2)$ have certain properties, $f(\mathbf{m}\,|\,I_1, I_2)$ may be sampled even when either $f(\mathbf{m}\,|\,I_1)$ or $f(\mathbf{m}\,|\,I_2)$ cannot be evaluated.

Further, even when both $f(\mathbf{m}\,|\,I_1)$ or $f(\mathbf{m}\,|\,I_2)$ can be evaluated, simple alterations of the classical rejection sampler and Metropolis–Hastings, algorithm can lead to computationally much more efficient sampling methods.

### 6.5.1. Extended Rejection Sampling

Say that an algorithm exists that allows generation of independent realizations from $f(\mathbf{m}\,|\,I_1)$. Using $f(\mathbf{m}\,|\,I_1)$ as a proposal distribution for the rejection sampler results in a more efficient rejection sampler, specifically for the case of sampling the product $f(\mathbf{m}\,|\,\mathbf{I}) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$:

**EXTENDED REJECTION SAMPLING ALGORITHM OF $f(\mathbf{m}\,|\,\mathbf{I}) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$**

1. **Propose** a model $\mathbf{m}_{propose}$, as a realization of from $f(\mathbf{m}\,|\,I_1)$.

2. **Accept** this model with probability $P_{acc}$:

$$\begin{aligned} P_{acc} &= \frac{f\left(\mathbf{m}_{propose}\,|\,I_1, I_2\right)}{f\left(\mathbf{m}_{propose}\,|\,I_1\right)\max\left(f\left(\mathbf{m}\,|\,\mathbf{I}\right)\right)} \\ &= \frac{f\left(\mathbf{m}_{propose}\,|\,I_1\right)\, f\left(\mathbf{m}_{propose}\,|\,I_2\right)}{f\left(\mathbf{m}_{propose}\,|\,I_1\right)\max\left(f\left(\mathbf{m}\,|\,\mathbf{I}\right)\right)} \\ &\propto \frac{f\left(\mathbf{m}_{propose}\,|\,I_2\right)}{\max\left(f\left(\mathbf{m}\,|\,I_2\right)\right)}, \end{aligned} \tag{6.22, 6.23}$$

where $\max(f(\mathbf{m}\,|\,I_2))$ is the maximum probability distribution value of $f(\mathbf{m}\,|\,I_2)$.

Note that in this case the actual probability distribution value of $f(\mathbf{m}\,|\,I_1)$ or $f(\mathbf{m}\,|\,I_1, I_2)$ need never be evaluated, as long as an algorithm exists that generates realizations of $f(\mathbf{m}\,|\,I_1)$. If the algorithm that samples $f(\mathbf{m}\,|\,I_1)$ is reasonably efficient, the extended rejection sampling algorithm may be computationally much more efficient than the classic rejection sampler.

As demonstrated previously (Sections 6.2.1.2 and 6.2.1.3), a large collection of algorithms have been developed in recent decades, which are able to generate realizations of a (possibly unknown) probability distribution, such as, for example $f(\mathbf{m}\,|\,I_1)$, which can therefore be used as part of a rejection sampler to sample from $f(\mathbf{m}\,|\,I_1, I_2) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$, if only $f(\mathbf{m}\,|\,I_2)$ can be evaluated.

### 6.5.2. The Extended Metropolis Algorithm

The extended Metropolis algorithm [*Mosegaard and Tarantola*, 1995] is a modified version of the classic Metropolis–Hastings algorithm designed to sample the product of two probability distributions, $f(\mathbf{m}\,|\,I_1, I_2) = k\, f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$, in the specific case where an algorithm exists to iteratively sample $f(\mathbf{m}\,|\,I_1)$. It can be applied as follows:

0. **Init** Generate a starting model, $\mathbf{m}_{current}$, as a realization of $f(\mathbf{m}\,|\,I_1)$.

1. **Exploration.** Propose a new realization of $f(\mathbf{m}\,|\,I_1)$, $\mathbf{m}_{propose}$, in the vicinity of $\mathbf{m}_{current}$.

2. **Exploitation.** Accept the move to $\mathbf{m}_{propose}$ with the acceptance probability $P_{acc}$:

$$P_{acc} = \min\left(1, \frac{f\left(\mathbf{m}_{proposed}\,|\,I_2\right)}{f\left(\mathbf{m}_{current}\,|\,I_2\right)}\right). \tag{6.24}$$

If the move is accepted, $\mathbf{m}_{propose}$ becomes $\mathbf{m}_{current}$. Otherwise the random walk stays at the $\mathbf{m}_{current}$ location. Goto 1.

The exploration must be implemented in such way that when iterating, only the exploration step (i.e., accepting all model proposals) should lead to an algorithm sampling $f(\mathbf{m}\,|\,I_1)$.

To apply the extended Metropolis algorithm, one must (a) be able to compute a value proportional to $f(\mathbf{m}\,|\,I_2)$ for any proposed model $\mathbf{m}_{proposed}$ and (b) be able to perform a random walk that will sample $f(\mathbf{m}\,|\,I_1)$. There is no requirement to be able to evaluate neither $f\left(\mathbf{m}\,|\,I_1\right)$

nor the product $f(\mathbf{m}\,|\,I_1, I_2) \propto f(\mathbf{m}\,|\,I_1)\, f(\mathbf{m}\,|\,I_2)$. A "black box" algorithm that can perform a random walk which samples $f(\mathbf{m}\,|\,I_1)$ is sufficient [*Mosegaard and Tarantola*, 1995].

All the advantages and disadvantages of using the classic Metropolis–Hastings algorithm listed above also applies when using the extended Metropolis algorithm. However, if an algorithm exists that allows performing a random walk, such that $f(\mathbf{m}\,|\,I_1)$ is sampled, then the extended Metropolis algorithm may be orders of magnitude more efficient than using the Metropolis–Hastings algorithm.

#### *6.5.2.1. Sequential Gibbs Sampling*

The sequential simulation algorithm, Section 6.2.2, was originally developed to allow efficient simulation of independent realizations from probability distributions $f(\mathbf{m}\,|\,I_1)$ based on two- and multiple-point statistics, as demonstrated in Sections 6.2.1.1 and 6.2.1.3.

Therefore, any sampling algorithm based on sequential simulation can be used to perform a random walk, where each visited model is independent of its neighbors. This corresponds to a random walk with maximum exploration and hence maximum step-length. However, a crucial part of applying the extended Metropolis algorithm is the ability to control the step-length—that is, controlling the exploratory nature of the algorithm performing the random walk sampling $f(\mathbf{m}\,|\,I_1)$.

Sampling of $f(\mathbf{m}\,|\,I_1)$, using an arbitrary step-length, can be accomplished using sequential Gibbs sampling [*Hansen et al.*, 2008, 2012], for any probability distribution that can be sampled using sequential simulation. See also *Fu* and *Gómez-Hernández* [2008]; *Irving et al.* [2010] for related methods specific for Gaussian based models, and *Mariethoz et al.* [2010] for a method similar to *Hansen et al.* [2008].

Assume that $\mathbf{m}_1$ is a "current" model, which is a realization of $f(\mathbf{m}\,|\,I_1)$. Then one step of the sequential Gibbs sampling algorithm will generate a new realization $\mathbf{m}_2$ of $f(\mathbf{m}\,|\,I_1)$ in the vicinity of $\mathbf{m}_1$ using the following steps:

1. Select a subset $U$ of all $M$ model parameters, $\mathbf{m}_{1,i\in U}$.
2. Use sequential simulation to generate a realization $\mathbf{m}^*_{i\in U}$ of $f(\mathbf{m}_{i\in U}\,|\,\mathbf{m}_{i\notin U})$; that is, re-simulate the model parameters in $U$ conditional to the model parameters not in $U$.
3. Update the next model, $\mathbf{m}_2$, as $\mathbf{m}_{2,i\notin U} = \mathbf{m}_{1,i\notin U}$ and $\mathbf{m}_{2,i\in U} = \mathbf{m}^*_{i\in U}$.

Performing these steps iteratively will generate a series of models that will represent a random walk sampling $f(\mathbf{m}\,|\,I_1)$. This is exactly the requirements of the "black" box algorithm needed by the extended Metropolis algorithm to sample $f(\mathbf{m}\,|\,I_1)$.

The number of model parameters in the subset $U$ reflects the step-length. The longest step-length is when $U$ contains all the model parameters in which case $\mathbf{m}_2$ will be independent of $\mathbf{m}_1$.

The sequential Gibbs sampler can in principle be used to sample any of the probability distributions described in Sections 6.2.1.1–6.2.1.3 through a random walk with an arbitrary step size.

Note that a perfect application of sequential Gibbs sampling requires sampling from the full conditional distribution, Eq. (6.9), at each iteration. Most of the simulation algorithms based on sequential simulation described previously make use of a data neighborhood in which case the conditional distribution will only be approximately correct. If the neighborhood is chosen sufficiently large for probability distributions based on two-point statistics, this approximation does in practice provide the same results as when using a full neighborhood. *Cordua et al.* [2015] observed that using sequential Gibbs sampling with the multiple-point based SNESIM algorithm, *Strebelle* [2002], will render a sampling algorithm, where the sampled probability distribution depends on the step perturbation size of the sequential Gibbs perturbation. A correction using frequency matching *Lange et al.* [2012] is suggested to remedy the unwanted effect of perturbations size and in this way remain to sample from a probability distribution that satisfies the multiple-point statistics from the training image.

#### *6.5.2.2. Independent Extended Metropolis Algorithm*

A simple variant of the extended Metropolis algorithm is when the step-length is set to its maximum; that is, a new independent realization of the $f(\mathbf{m}\,|\,I_1)$ is proposed in the exploration step, similar to the metropolized independence sampler proposed by *Liu* [1996]. In this case, any probability distribution from which independent realizations can be generated can be used for probabilistic data integration. Thus, there is no need to use the sequential Gibbs sampler. This means that, in principle, most developed geostatistical algorithms can be used to describe information that can be used for data integration problems. Application of the independent extended Metropolis algorithm avoids the problem of estimating the normalization constant in the acceptance ratio, as is needed when applying the rejection sampler, which may lead to a computationally much faster algorithm. This algorithm is as simple to implement as the rejection sampler, which in practice render the rejection sampler obsolete. Compared to the extended Metropolis algorithm, the independent extended Metropolis algorithm is easier to implement, but also much less computationally efficient.

### 6.6. EXAMPLE OF SAMPLING $f(\mathbf{m}\,|\,I_1, I_2)$

To demonstrate different aspects of some of the presented algorithms, consider a crosshole tomographic inverse problem. 40 × 84 model parameters represent a 2D electromagnetic velocity field of size 5.85 m × 12.45 m. This is exactly the same model size as considered in Figures 6.2 and 6.4.

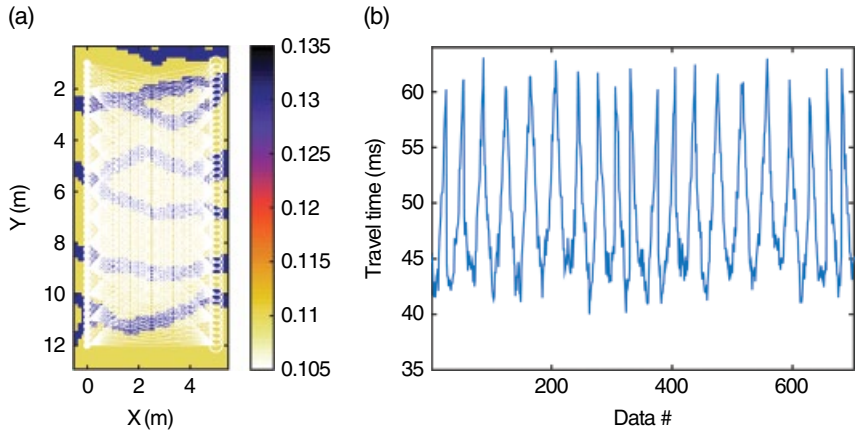


**Figure 6.6** **(a)** Reference velocity model (*white lines* connect source and receiver locations). **(b)** Reference travel-time dataset.

Figure 6.6 shows a reference model generated as a realization from $f(\mathbf{m} \mid I_{d5})$, which is based on the statistics inferred from the sample model in Figure 6.1.

Mimicking a cross borehole tomographic experiment, travel times of electromagnetic waves from 702 source locations to 702 receiver locations, as indicated in Figure 6.6a, are computed using finite frequency theory using *Hansen et al*. [2013a]. Then a realization of zero mean uncorrelated Gaussian noise with standard deviation of 0.8 ns, $\mathbf{C}_{\mathrm{d}} = 0.8^2\mathbf{I}$, is added to the travel-time data, which are then considered as "observed" data.

Thus one type of indirect information, which we will refer to as $I_{indirect}$, related to geophysical travel time measurement, is available. $I_{indirect}$ then specifies not only the travel-time data and the measurement uncertainty, $\mathcal{N}(0,\mathbf{C}_{\mathrm{d}})$, but also knowledge about how to solve the forward problem. This means that we are able to use the likelihood function in Eq. (6.13) to evaluate $f(\mathbf{m} \mid I_{indirect})$.

Any of the previously defined probability distributions describing different types of direct information $f(\mathbf{m} \mid I_{d1}),\ldots,f(\mathbf{m} \mid I_{d6})$ is also, in turn, considered as information available about the model parameters. Recall that Figures 6.2 and 6.4 show realizations from these probability distributions.

The problem is now to solve the data integration problem by generating a sample from $f(\mathbf{m} \mid I_{di}, I_{indirect}) = k\, f(\mathbf{m} \mid I_{di})\, f(\mathbf{m} \mid I_{indirect})$, where $i = 1,\ldots,6$.

### 6.6.1. Sampling $f(\mathbf{m} \mid I_{di}, I_{indirect})$ Using Rejection Sampling

The extended rejection sampler presented previously, making use of sequential simulation to generate realizations of the direct information, can in principle be used to sample the joint distribution $f(\mathbf{m} \mid I_{di}, I_{indirect})$. However, in practice the rejection sampler is only applicable to very-low-dimensional sampling problems and could not be applied for the current case.

### 6.6.2. Sampling $f(\mathbf{m} \mid I_{di}, I_{indirect})$ Using the Extended Metropolis Algorithm

For all the considered probability distributions based on direct information, $f(\mathbf{m} \mid I_{d1}),\ldots, f(\mathrm{m} \mid I_{d6})$, a random walk that samples the probability distribution, with arbitrary step-length, can be performed using sequential Gibbs sampling. Hence, the combined information $f(\mathbf{m} \mid I_{di}, I_{indirect})$ can be sampled using the extended Metropolis algorithm, without ever evaluating $f(\mathbf{m} \mid I_{di})$.

The extended Metropolis algorithm has been run for 100,000 iterations drawing realizations from $f(\mathbf{m} \mid I_{di}, I_{Indirect})$, for each of the six types of direct information. In all runs, the step-length is selected such that the acceptance rate of the algorithm is around 30%. For details about running the extended Metropolis algorithm, see, for example, *Cordua et al*. [2012] and *Hansen et al*. [2013a].

The extended Metropolis sampler was especially prone to be caught in local minima sampling $f(\mathbf{m} \mid I_{d6}, I_{indirect})$, and therefore the parallel tempering algorithm was used in this case [*Sambridge*, 2013].

Figures 6.7 and 6.8 show five realizations from the probability distribution describing the combined information of $f(\mathbf{m} \mid I_{d1}, I_{indirect}),\ldots, f(\mathbf{m} \mid I_{d6}, I_{indirect})$. These realizations should be compared to the realizations from the probability distribution based on direct information in Figures 6.2 and 6.4.

Comparing Figures 6.7 and 6.8 to Figures 6.2 and 6.4, it is obvious that the spatial variability from the direct information is preserved in the realizations from the combined probability distributions. If the direct information defines the subsurface as a set of Voronoi cells, as for

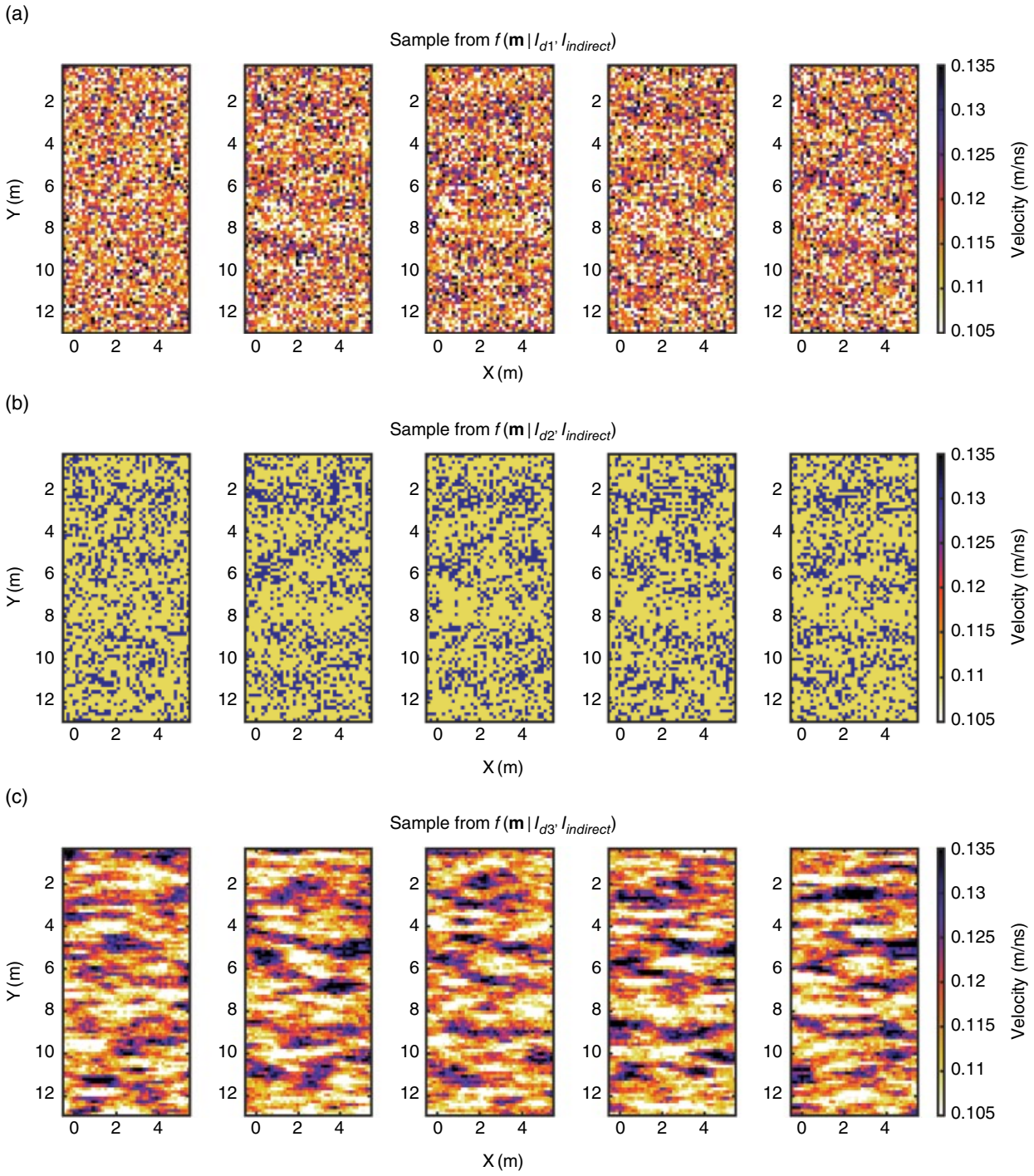


**Figure 6.7** Five realizations from **(a)** $f(\mathbf{m} \mid I_{d1}, I_{indirect})$, **(b)** $f(\mathbf{m} \mid I_{d2}, I_{indirect})$, and **(c)** $f(\mathbf{m} \mid I_{d3}, I_{indirect})$. See text for details.

$f(\mathbf{m} \mid I_{d6})$, then realizations from the combined probability distribution will consist of Voronoi cells (Figure 6.8c). Then one should off course consider whether a set of Voronoi cells provide a geologically reasonable description of Earth structures. In this case, the realizations of $f(\mathbf{m} \mid I_{d6}, I_{indirect})$ does not seem to resemble realistic geological variability.

The choice of a spatially uncorrelated probability distribution to describe direct information, such as $f(\mathbf{m} \mid I_{d1})$ and $f(\mathbf{m} \mid I_{d2})$, will also affect the combined information content of $f(\mathbf{m} \mid I_{d1}, I_{indirect})$ and $f(\mathbf{m} \mid I_{d1}, I_{indirect})$, which will also exhibit maximum spatial disorder in the outcome realizations. If more indirect information is available (e.g., less noise or more data),

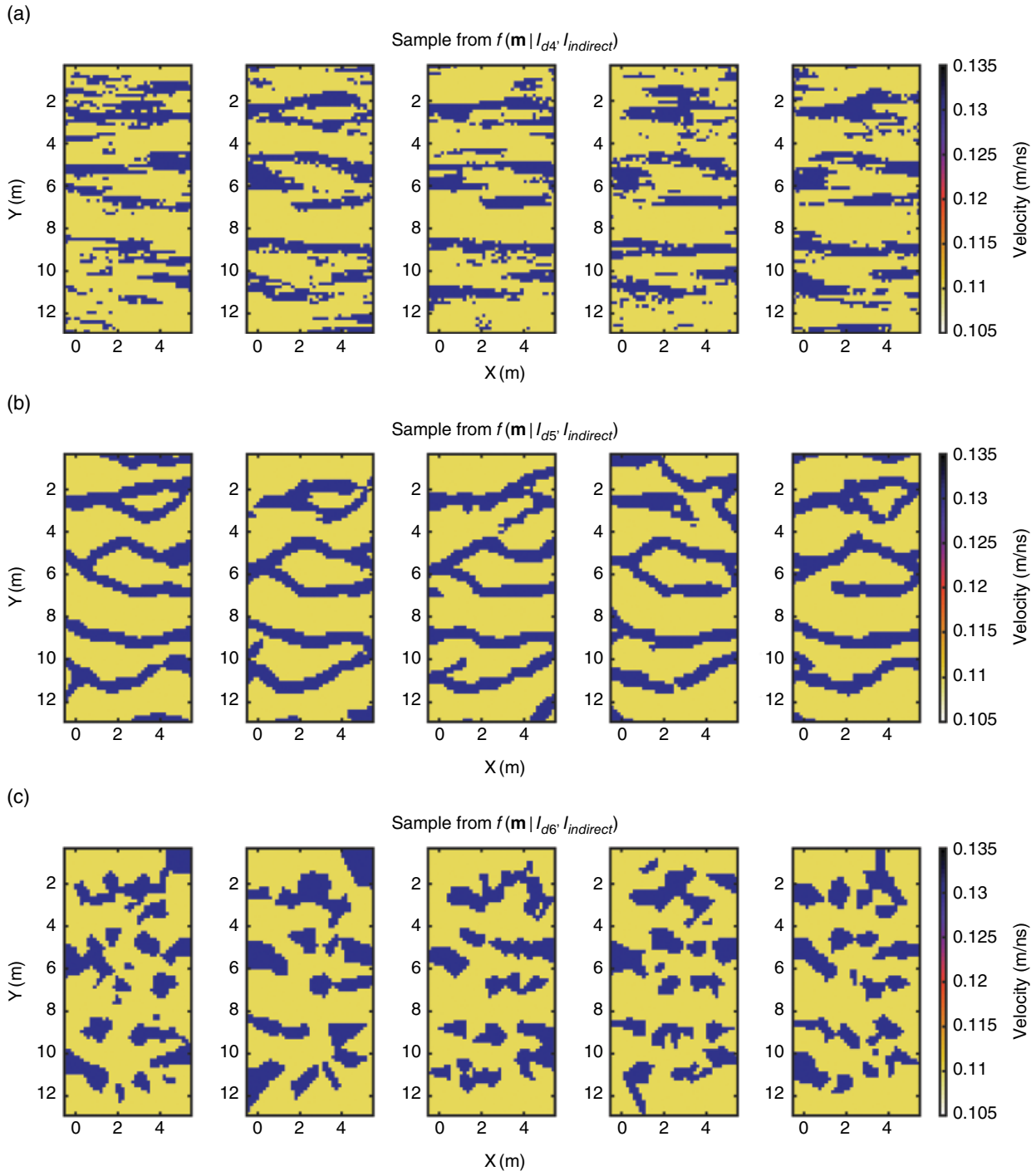


**Figure 6.8** Five realizations from **(a)** $f(\mathbf{m} \mid I_{d4}, I_{indirect})$, **(b)** $f(\mathbf{m} \mid I_{d5}, I_{indirect})$, and **(c)** $f(\mathbf{m} \mid I_{d6}, I_{indirect})$. See text for details.

such that $f(\mathbf{m} \mid I_{indirect})$ will be more informed, then realizations of the combined probability distribution $f(\mathbf{m} \mid I_{d1}, I_{Indirect})$ may expose more correlated features, corresponding the actual reference model. However, the information that cannot be resolved by the indirect information will stem from the probability distribution of direct information.

Figure 6.9 shows the pointwise mean (sometime called the etype mean) computed from all realizations. This indicates that, on average, the correct location of the

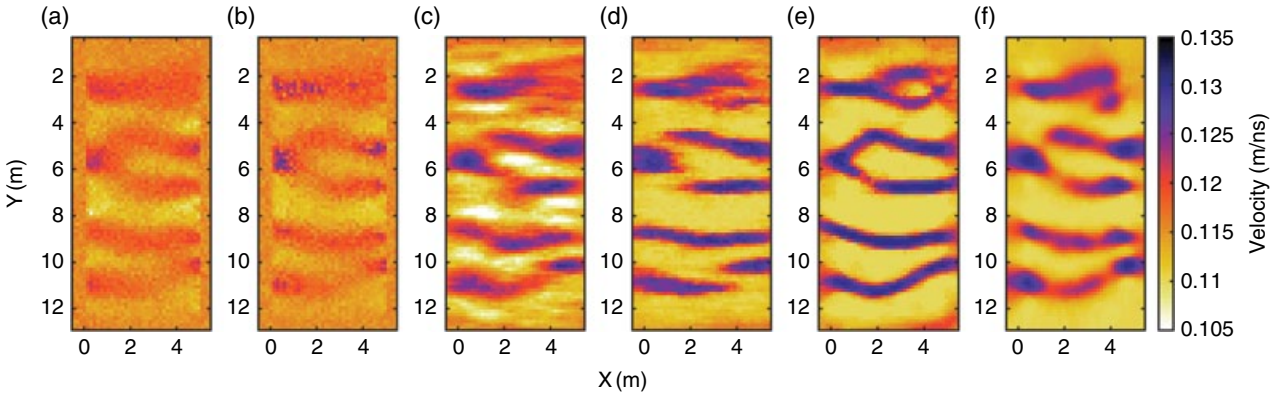


**Figure 6.9** Pixelwise mean model obtained from a sample of **(a)** $f(\mathbf{m} \mid I_{d1}, I_{indirect})$, **(b)** $f(\mathbf{m} \mid I_{d2}, I_{indirect})$, **(c)** $f(\mathbf{m} \mid I_{d3}, I_{indirect})$, **(d)** $f(\mathbf{m} \mid I_{d4}, I_{indirect})$, **(e)** $f(\mathbf{m} \mid I_{d5}, I_{indirect})$, and **(f)** $f(\mathbf{m} \mid I_{d6}, I_{indirect})$.

**Table 6.1** Correlation Coefficient Between Independent Realizations of $f(\mathbf{m} \mid I_{di}, I_{indirect})$

| | $I_{d1}$ | $I_{d2}$ | $I_{d3}$ | $I_{d4}$ | $I_{d5}$ | $I_{d6}$ |
|---|---|---|---|---|---|---|
| CC | 0.08 | 0.08 | 0.35 | 0.44 | 0.55 | 0.05 |

channel structures can be identified, even if they cannot be identified on the individual realizations. Specifically, using direct information probability distribution based on Voronoi cells results in individual realizations from $f(\mathbf{m} \mid I_{d6}, I_{Indirect})$ that are clearly geologically unrealistic (compared to the sample model) (Figure 6.8c), while realizations from $f(\mathbf{m} \mid I_{d5}, I_{Indirect})$ results in geologically highly realistic realizations (Figure 6.8b). On average, though, the pointwise mean is remarkable similar (Figures 6.9e and 6.9f). Note that such average models are, in general, not solutions to the data integration problem, as they may be inconsistent with both the direct and indirect information.

If the goal is to simulate geologically realistic features, then Figures 6.7 and 6.8 clearly show that, for this case, direct information describing geological realistic features are essential.

Table 6.1 provides the correlation coefficient between independent realizations of $f(\mathbf{m} \mid I_{di}, I_{indirect})$. A high number indicates that independent realizations are very similar and, hence, that the model parameters are well-resolved. Relying on the spatially uncorrelated models, $f(\mathbf{m} \mid I_{d1}, I_{indirect})$ and $f(\mathbf{m} \mid I_{d1}, I_{indirect})$ provides a very low correlation coefficient, which may suggest a poor resolution. The correlation coefficient increases as information about the model parameters, consistent with the reference model, increases. This indicates that as more information is available, consistent with the actual unknown subsurface, the resolution will increase.

### 6.6.3. Sampling $f(\mathbf{m} \mid I_{d3}, I_{indirect})$ Using the Classic Metropolis Algorithm

$f(\mathbf{m} \mid I_{d3})$ represents a Gaussian probability distribution and can be evaluated directly using Eq. (6.5). Therefore $f(\mathbf{m} \mid I_{d3}, I_{indirect})$ can be evaluated and, hence, sampled using the classic Metropolis algorithm.

Using a spatially uncorrelated uniform proposal distribution, with velocity values between 0.0755 m/ns and 0.1555 m/s, the classic Metropolis algorithm has been run for 4 million iterations in order to sample $f(\mathbf{m} \mid I_{d3}, I_{indirect})$.

Figure 6.10 shows three independent realizations from $f(\mathbf{m} \mid I_{d3}, I_{indirect})$ as well as the corresponding pointwise mean model. These results are comparable to the results obtained using the extended Metropolis sampler (Figures 6.7b and 6.9c).

Figure 6.11 shows the logarithm of the probability distribution values for $f(\mathbf{m} \mid I_{d3})$, $f(\mathbf{m} \mid I_{indirect})$, and $f(\mathbf{m} \mid I_{d3}, I_{indirect})$ as a function of iteration number using classic Metropolis algorithm, and $f(\mathbf{m} \mid I_{indirect})$ using the extended Metropolis algorithm. Both algorithms tend to sample models with comparable values for $f(\mathbf{m} \mid I_{indirect})$—that is, suggesting, as Figure 6.10, that the same probability distribution has been sampled.

However, it also highlights that the number of iterations needed to achieve burn-in—that is, where the algorithm starts to generate realizations of $f(\mathbf{m} \mid I_{d3}, I_{indirect})$—is very different. Using the extended Metropolis algorithm burn-in is reached after around $10^3$ iterations, whereas it takes about $10^6$ iterations to reach burn-in using the classic Metropolis algorithm with a uniform proposal distribution. Further, the number of iterations between independent realizations is about $4 \times 10^3$ using the extended Metropolis algorithm but about $1.5 \times 10^6$ using the classic Metropolis algorithm. Hence, the difference in computational requirements for sampling $f(\mathbf{m} \mid I_{d3}, I_{indirect})$

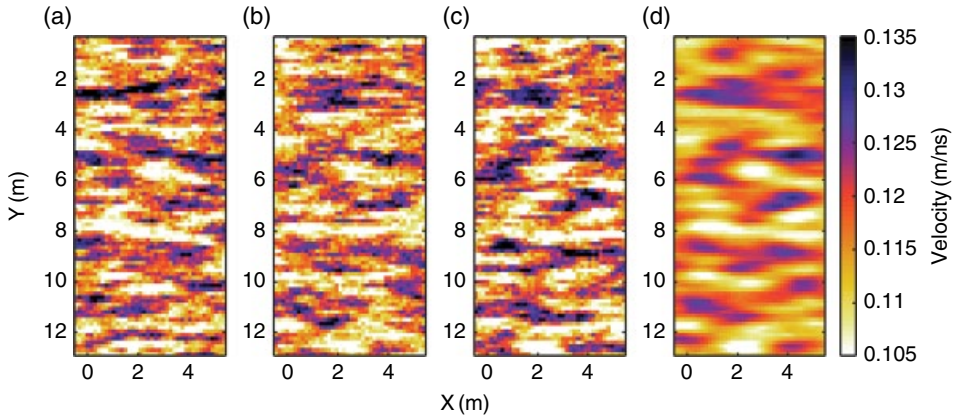


**Figure 6.10** Three realizations from $f(\mathbf{m} \mid I_{d3}, I_{indirect})$, **(a–c)** and the pointwise average **(d)** obtained using 4,000,000 iterations of the classic Metropolis alogrithm.

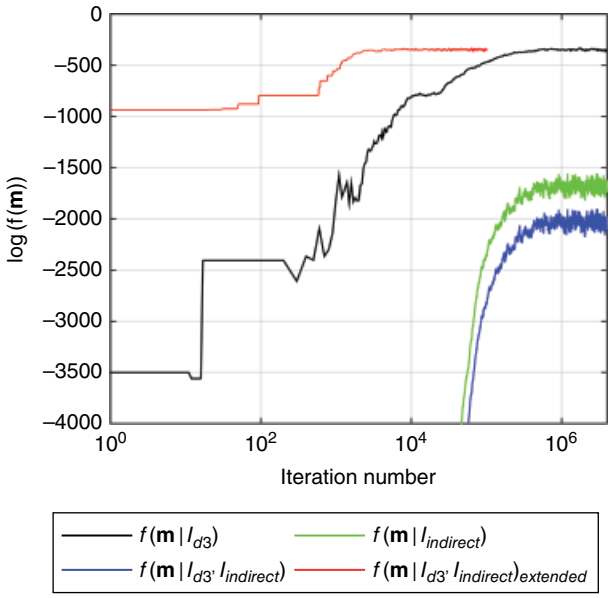


**Figure 6.11** $\log(f(\mathbf{m}))$ as a function of iteration number for $f(\mathbf{m}|I_{indirect})$ (*green*), $f(\mathbf{m}|I_{d3})$ (*black*), and $f(\mathbf{m}|I_{d3}, I_{indirect})$ (*blue*) using the classic Metropolis algorithm and for $f(\mathbf{m}|I_{d3}, I_{indirect})_{extended}$ using the extended Metropolis algorithm (*red*).

using the two types of Metropolis algorithms is close to a factor of 1000.

The main reason for this huge difference in computational efficiency is related to the fact that using the classic Metropolis algorithm, one must sample $f(\mathbf{m} \mid I_{d3})$ as part of sampling $f(\mathbf{m} \mid I_{d3}, I_{indirect})$. On the other hand, using the extended Metropolis algorithm, the use of sequential Gibbs sampling ensures that all proposed models are realizations of $f(\mathbf{m} \mid I_{d3})$, and hence the computational requirements are mostly related to evaluating $f(\mathbf{m} \mid I_{d3}, I_{indirect})$.

## 6.7. DISCUSSION

The example in the previous section demonstrates the benefits of being able to use information about for example geologically plausible structures. It also demonstrates a case where the information quantified by $f(\mathbf{m} \mid I_{d1})$, …, $f(\mathbf{m} \mid I_{d5})$ is consistent with the actual "Earth" as shown in Figure 6.6a. However, in practice the inference of statistical properties from a sample model, such as the one shown in Figure 6.1, may be associated with varying degrees of subjectivity. Also, an inferred statistical model may not be able to describe the actual spatial properties of the subsurface.

Consider the sample model in Figure 6.1. The width of the channels in this sample model is consistently around 0.6 m. The same is the case for the width of the channels in the realizations of $f(\mathbf{m} \mid I_{d5})$ shown in Figure 6.4b. In fact, the probability of locating a channel (that is, not intersecting other channels) with width $w > 1$ m or $w < 0.45$ m is zero. Further, in this sample model each model parameter can only take two values. This means that any other value will have a probability of zero of occurring. For any real case, the information exemplified in the sample model in Figure 6.1 will most likely exhibit too little variability. Hence, $f(\mathbf{m} \mid I_{d5})$ may be low in entropy and, in fact, inconsistent with the true Earth. Therefore it may be difficult, if not impossible, to integrate this information with other types of data. For an example on the use of inconsistent direct information, see, for example, *Hansen et al.* [2008]. The difficulty in quantifying direct information is that one should try to quantify as much direct information as possible, while at the same time allow realistic uncertainty [*Jaynes*, 1984; *Journel and Deutsch*, 1993].

**Extreme High-Entropy Uniform Model.** An extreme choice of an uninformed statistical model is the uniform model. Consider the integration of two types of

independent information, $f(\mathbf{m} \mid I_1)$ and $f(\mathbf{m} \mid I_2)$, where $f(\mathbf{m} \mid I_1)$ represents a uniform distribution $\mathcal{U}(-\infty, \infty)$. Then $f(\mathbf{m} \mid I_1, I_2)$ is given by

$$f(\mathbf{m} \mid I_1, I_2) = f(\mathbf{m} \mid I_1) f(\mathbf{m} \mid I_2) \tag{6.25}$$

$$\propto f(\mathbf{m} \mid I_2) \tag{6.26}$$

In other words, when $f(\mathbf{m} \mid I_1)$ is a uniform distribution, it adds no information about the model parameters, as $f(\mathbf{m} \mid I_1, I_2) \propto f(\mathbf{m} \mid I_2)$. In principle, any 40 × 84 pixel random cutout of the reference model in Figure 6.6a is as probable an outcome of the uniform model $f(\mathbf{m} \mid I_{d2})$ as any of the single realizations shown in Figures 6.2b, 6.4a, 6.4b, and 6.4c. In reality, though, the uniform model as a choice for a description of the distribution of $\mathbf{m}$ involves a rather extreme assumption about maximum entropy, or maximum disorder. Any typical realization of $f(\mathbf{m} \mid I_1)$ will expose high disorder. In other words, the probability of realizing a model with a high degree of disorder is very high. The probability of realizing a highly ordered model, such as the reference model with ordered channel like structures, is extremely low. This is exactly what is exemplified by the realizations from $f(\mathbf{m} \mid I_{d2})$ in Figure 6.2b. The high entropy assumptions of $f(\mathbf{m} \mid I_{d2})$ will also be associated to $f(\mathbf{m} \mid I_{d2}, I_{indirect})$ as shown in Figure 6.7b.

This poses a problem, not only related to visual plausibility. In real life, end users may not be interested in the model parameters $\mathbf{m}$ themselves, but in a variable $\mathbf{k}$ linked to the model parameters through some transfer function h as $\mathbf{k} = h(\mathbf{m})$. $\mathbf{k}$ may be very sensitive to the type of spatial variability. Consider, for example, flow modeling of groundwater reservoir or hydrocarbon reservoirs. Say the channel structures (blue in Figure 6.6a) represents highly permeable structures embedded in low permeability material. Then, flow modeling results will provide radically different results depending on which model is chosen to describe spatial variability. For illustrative examples see, for example, *Journal and Deutsch* [1993]; *Journel and Zhang* [2006].

Another property of spatially uncorrelated models, such as $f(\mathbf{m} \mid I_{d1})$ and $f(\mathbf{m} \mid I_{d2})$, is that the number of effective "free" model parameters $M_f$ is the same as the number of model parameters $M$, $M_f = M$. The number of "free" model parameters is the minimum number of model parameters needed to represent $\mathbf{m}$ [*Hansen et al.*, 2009]. When the number of model parameters increases, the data integration problem may become increasingly more difficult in terms of sampling from the distribution of combined information.

**Extreme-Low Entropy Models.** Other types of models represent cases of extreme low entropy. Consider, for example, a checkerboard model in a regular grid, where each model parameter (pixel) takes the value "black" or "white". The neighbor pixel up or down, left or right to one centered pixel has the opposite value as the center pixel. This also means that for such a model, the number of free parameters is $M_f = 1$, independent of the actual number of parameters. If such a checkerboard model is used to describe direct information, then an exhaustive search of all possible models can be undertaken simply by evaluating two models, one with a white pixel centered at a reference parameter and one with a black pixel centered at a reference parameter.

Another extreme type of low-entropy model is the multivariate Gaussian model, where all the model parameters are completely correlated. Again, this would indicate that one only needs to know the value of one model parameter in order to know the value of all model parameters ($M_f = 1$) independent of the number of model parameters.

**Intermediate Entropy Models.** In general, the number of free model parameters will depend on the chosen a priori model. For multivariate Gaussian models, *Hansen et al.* [2009] demonstrate that in general the number of effective free parameters is related to the correlation length. The longer the correlation length, the smaller the value of $M_f$. When the correlation length is zero, the model parameters become independent, and hence $M_f = M$.

**Low Entropy as the Source of Inconsistencies.** In order to avoid inconsistencies in data integration, careful consideration should be used when quantifying different types of information $f(\mathbf{m} \mid I_i)$. If only known information is quantified and all uncertainties are taken into account, inconsistencies should not arise.

However, sometimes data integration, in the form of sampling from $f(\mathbf{m} \mid I_1, I_2)$, can become unsolvable if there is inconsistency between the available information [*Hansen et al.*, 2008]. There are at least four explanations: (1) The direct information is specified such that the data cannot be matched within their uncertainty, (2) the modeling uncertainty related to the forward model is underestimated, (3) the measurement uncertainty is underestimated, and (4) the parameterization has been chosen too sparse to allow realistic representation of Earth structures [*Mosegaard and Hansen*, 2015]. In any case inconsistencies may arise when some of the information has been described with too little uncertainty.

**Sampling from $f(\mathbf{m} \mid I_{d5}, I_{indirect})$ Using Different Neighborhood.** The entropy, and the degree of spatial variability, is affected when the size of the neighborhood is changed (i.e., when changing the number of conditional data), which is used to compute/evaluate the conditional distribution as part of running sequential simulation. The smaller the amount of conditional data, the smaller the amount of information that is assumed (the entropy increases).

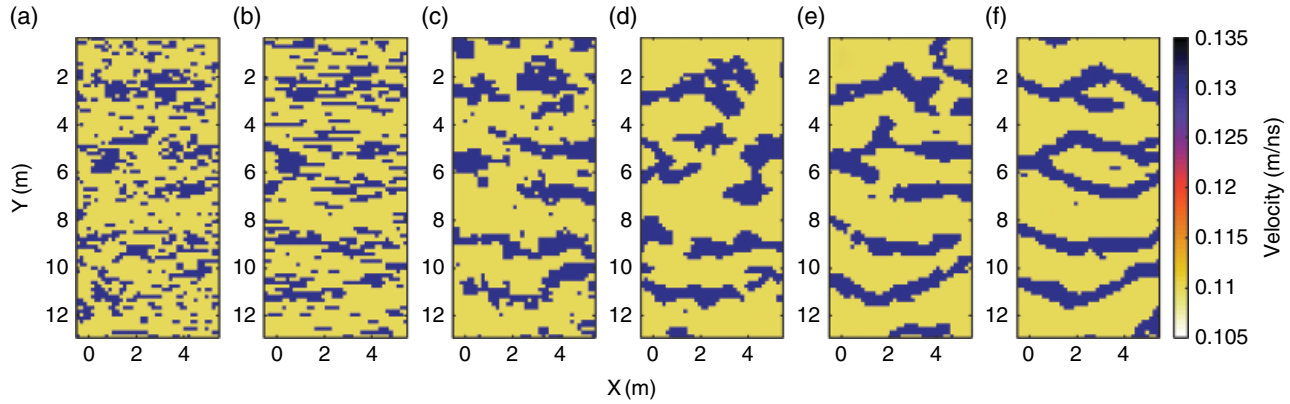


**Figure 6.12** One realization from $f(\mathrm{m}|I_{d5}, I_{indirect})$ using different number for conditioning data, $N_c$ (i.e., different size data neighborhood). **(a)** $N_c = 1$, **(b)** $N_c = 2$, **(c)** $N_c = 4$, **(d)** $N_c = 8$, **(e)** $N_c = 15$, **(f)** $N_c = 30$.

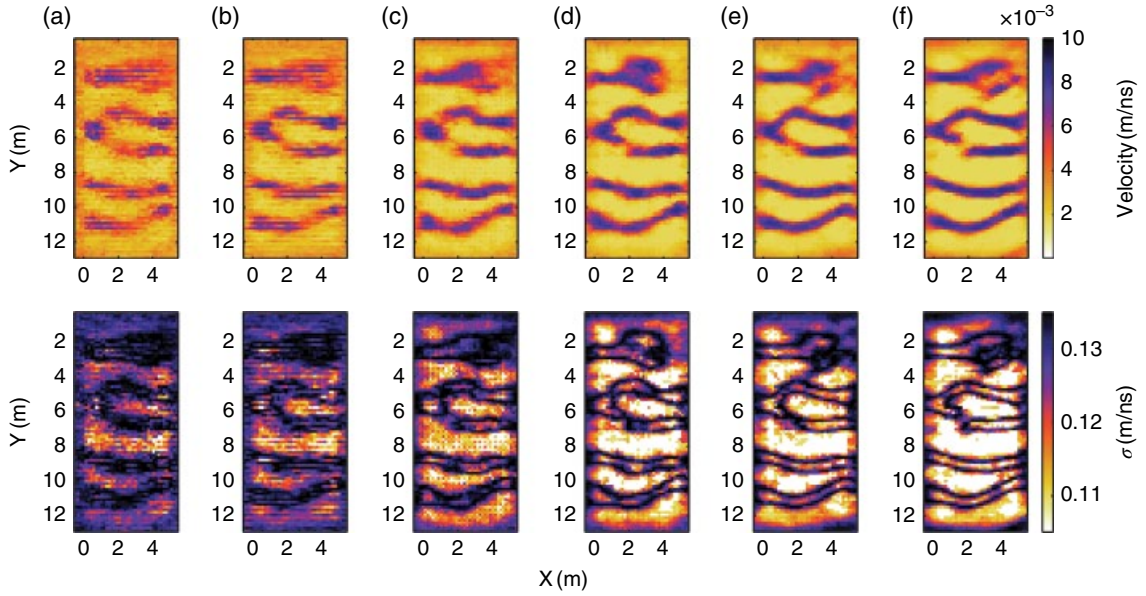


**Figure 6.13** The pointwise mean (*top*) and standard deviation (*bottom*) from a sample of $f(\mathbf{m}|I_{d5}, I_{indirect})$ using different number for conditioning data, $N_c$ (i.e., different size data neighborhood). **(a)** $N_c = 1$, **(b)** $N_c = 2$, **(c)** $N_c = 4$, **(d)** $N_c = 8$, **(e)** $N_c = 15$, **(f)** $N_c = 30$.

For the realizations generated from $f(\mathbf{m} \mid I_{d5})$ and $f(\mathbf{m} \mid I_{d5}, I_{indirect})$, shown in Figures 6.4b and 6.8b, the number of conditional points for the sequential simulation algorithm used is $N_c = 60$. Figure 6.12 shows one realization obtained from sampling $f(\mathbf{m} \mid I_{d5}, I_{indirect})$ using $N_c = [1, 2, 4, 8, 15, 30]$. The same type of extended Metropolis algorithm as described earlier is used. Figure 6.13 shows the corresponding pointwise mean (top row) and point wise standard deviation (bottom row) obtained from all generated realizations. Note how the variability is increasingly associated with the location of the channel edges as the number of conditional data increases. Note that $f(\mathbf{m} \mid I_{d5})$ corresponds to $f(\mathbf{m} \mid I_{d2})$ when no conditioning points data are used (i.e., assuming no spatial dependency). The spatial disorder is clearly seen to decrease as the number of conditioning points increase.

## 6.8. CONCLUSIONS

The goal of probabilistic data integration is to (1) integrate all available information $\mathbf{I} = [I_1, I_2, \ldots, I_N]$ related to model parameters $\mathbf{m}$ into one probability distribution $f(\mathbf{m} \mid \mathbf{I})$ and (2) generate a large sample from $f(\mathbf{m} \mid \mathrm{I})$ allowing detailed uncertainty analysis and propagation of uncertainty into other types of parameters (such as, for example, related to flow simulations).

In some rare cases, $f(\mathbf{m} \mid \mathbf{I})$ can be evaluated, in which case the "classic" Metropolis algorithm (or in principle

the rejection sampler) can be used to sample $f(\mathbf{m} \mid \mathbf{I})$ directly. However, the type of (usually simplistic) information that can be quantified and allows evaluation of $f(\mathbf{m} \mid \mathbf{I})$ is often not adequate to describe information at hand. Further, even when this is the case, such a sampling problem can become prohibitively computationally demanding, even for the relatively small 2D models considered here, which will be, in practice, intractable. Direct sampling of $f(\mathbf{m} \mid \mathbf{I})$ using the rejection sampler or the classic Metropolis algorithm will, in general, lead to a computationally intractable problem.

On the other hand, complex models of direct information can be quantified in a way that allows efficient sampling, based on sequential simulation, from these models, without the need to evaluate $f(\mathbf{m} \mid I_1)$ and, hence, $f(\mathbf{m} \mid \mathbf{I})$. When such information is available, together with other types of information such as indirect information from, for example, geophysical data $I_{indirect}$, where $f(\mathbf{m} \mid I_{indirect})$ can be evaluated, then $f(\mathbf{m} \mid I_1, I_{indirect})$ can be sampled efficiently using the extended Metropolis algorithm utilizing the sequential Gibbs sampler to sample $f(\mathbf{m} \mid I_1)$.

Compared to using direct sampling of $f(\mathbf{m} \mid \mathbf{I})$ using the classic Metropolis algorithm with a uniform proposal distributions, the use of extended Metropolis can lead to a sampling problem that is orders of magnitude more tractable.

A wide range of statistical methods, providing varying degrees of information content, are currently available that can be used with the extended Metropolis algorithm and that allow characterization of probability distributions describing quite complex and geologically realistic spatial features. These methods allow building statistical models that assume, in principle, a lot more than is typically known. Therefore, care should be taken when quantifying direct information, to avoid subjective information such that only information that is actually known is quantified and taken into account and such that all uncertainties are taken into account. If this is not the case, then the data integration problem may become either inconsistent and unsolvable, or solvable but providing biased results with too little associated uncertainty.

On the other hand, realistic description of direct information has several advantages: (1) Realizations from the probability distribution describing the combined information will be consistent with structural geological information. (2) Sampling from $f(\mathrm{m} \mid \mathrm{I})$ will be computationally more efficient. (3) The complexity of the inverse problem can be dramatically reduced due to the reduced number of effective free model parameters.

## ACKNOWLEDGMENTS

All computations have been performed using the SIPPI Matlab package [*Hansen et al.*, 2013b], and codes for reproducing the results can be found at http://sippi.sourceforge.net/. We thank Mats Lundh Gulbrandsen for discussions and editing. We thank two reviewers for their constructive and useful critique.

## REFERENCES

Armstrong, M., A. Galli, H. Beucher, G. Loc'h, D. Renard, B. Doligez, R. Eschard, and F. Geffroy (2011), *Plurigaussian Simulations in Geosciences*, Springer Science & Business Media, New York.

Bodin, T., M. Sambridge, and K. Gallagher (2009), A self-parametrizing partition model approach to tomographic inverse problems, *Inverse Problems*, *25*(5), 055,009.

Bosch, M. (2015), Inference networks in earth models with multiple components and data, in *Integrated Imaging in Earth Science*, M. Moorkamp, N. Linde, P. Lelievre, and A. Khan, eds., AGU, Washington, DC.

Buland, A., and H. Omre (2003), Bayesian linearized avo inversion, *Geophysics*, *68*(1), 185–198.

Caers, J. (2000), Direct sequential indicator simulation, in *Proceedings of the 6th International Geostatistics Congress, Cape Town, South Africa, April 10–14, 2000*, W. Kleingeld and D. Krige, eds., 12 pp.

Constable, S. C., R. L. Parker, and C. G. Constable (1987), Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*(3), 289–300.

Cordua, K. S., T. M. Hansen, and K. Mosegaard (2012), Monte Carlo full waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information, *Geophysics*, *77*, H19–H31, doi:10.1190/geo2011-0170.1.

Cordua, K. S., T. M. Hansen, and K. Mosegaard (2015), Improving the pattern reproducibility of multiple-point-based prior models using frequency matching, *Mathe. Geosci.*, *47*, 317–343.

Cressie, N., and J. L. Davidson (1998), Image analysis with partially ordered Markov models, *Comput. Stat. Data Anal.* *29*(1), 1–26.

Daly, C. (2005), Higher order models using entropy, Markov random fields and sequential simulation, Geostatistics Banff 2004, Springer Netherlands, pp. 215–224.

Deutsch, C. V., and A. G. Journel (1998), *GSLIB, Geostatistical Software Library and User's Guide, Applied Geostatistics*, 2nd ed., Oxford University Press, New York, 384 pp.

Dimitrakopoulos, R., H. Mustapha, and E. Gloaguen (2010), High-order statistics of spatial random fields: Exploring spatial cumulants for modeling complex non-Gaussian and non-linear phenomena, *Mathematical Geosciences*, *42*(1), 65–99.

Duane, S., A. D. Kennedy, B. J. Pendleton, and D. Roweth (1987), Hybrid monte carlo, *Physics Lett. B*, *195*(2), 216–222.

Emery, X. (2007), Using the gibbs sampler for conditional simulation of Gaussian-based random fields, *Comput. Geosci.*, *33*(4), 522–537.

Fu, J., and J. J. Gómez-Hernández (2008), Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method, *Inverse Probl. Sci. Eng.*, *16*(7), 865–884.

Geman, S., and D. Geman (1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Machine Intell.*, *6*, 721–741.

Gomez-Hernandez, J., and A. Journel (1993), Joint sequential simulation of multi-Gaussian fields, *Geostatistics Troia*, *92*, 85–94.

Guardiano, F., and R. Srivastava (1993), Multivariate geostatistics: Beyond bivariate moments, *Geostatistics-Troia*, *1*, 133–144.

Hansen, T. M., and K. Mosegaard (2008), VISIM: Sequential simulation for linear inverse problems, *Comput. Geosci., 34*(1), 53–76.

Hansen, T. M., K. Mosegaard, and K. C. Cordua (2008), Using geostatistics to describe complex a priori information for inverse problems, in *VIII International Geostatistics Congress*, Vol. 1, J. M. Ortiz and X. Emery, eds., Mining Engineering Department, University of Chile, pp. 329–338.

Hansen, T. M., K. S. Cordua, and K. Mosegaard (2009), Reducing complexity of inverse problems using geostatistical priors, in *Proceeding from IAMG 09, August 23–28, 2009, Stanford, CA*.

Hansen, T. M., K. C. Cordua, and K. Mosegaard (2012), Inverse problems with nontrivial priors—Efficient solution through sequential Gibbs sampling, *Computational Geosciences*, *16*(3), 593–611, doi:10.1007/s10596-011-9271-1.

Hansen, T., K. Cordua, M. Looms, and K. Mosegaard (2013a), 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, T., K. Cordua, M. Looms, and K. Mosegaard (2013b), 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, T. M., K. S. Cordua, B. H. Jacobsen, and K. Mosegaard (2014), Accounting for imperfect forward modeling in geophysical inverse problems exemplified for crosshole tomography, *Geophysics*, *79*(3), H1–H21.

Holliger, K., and A. Levander (1994), Lower crustal reflectivity modeled by rheological controls on mafic intrusions, *Geology*, *22*(4), 367–370.

Irving, J., and K. Singha (2010), Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resour Res.*, *46*, W11514.

Jaynes, E. T. (1984), Prior information and ambiguity in inverse problems, *Inverse Problems*, *14*, 151–166.

Journel, A., and E. Isaaks (1984), Conditional indicator simulation: Application to a saskatchewan uranium deposit, *J. Int. Associ. Math. Geol.*, *16*(7), 685–718.

Journel, A., and T. Zhang (2006), The necessity of a multiple-point prior model, *Math. Geol.*, *38*(5), 591–610.

Journel, A. G. (1994), Modeling uncertainty: Some conceptual thoughts, in *Geostatistics for the next Century*, Springer, New York, pp. 30–43.

Journel, A. G., and C. V. Deutsch (1993), Entropy and spatial disorder, *Math. Geol.*, *25*(3), 329–355.

Journel, A. G., and C. J. Huijbregts (1978), *Mining Geostatistics*, Academic Press, New York, 600 pp.

Lange, K., J. Frydendall, K. S. Cordua, T. M. Hansen, Y. Melnikova, and K. Mosegaard (2012), A frequency matching method: Solving inverse problems by use of geologically realistic prior information, *Math. Geosci.*, *44*(7), 783–803.

Le Ravalec, M., B. Noetinger, and L. Y. Hu (2000), The FFT moving average (FFT-MA) generator: An efficient numerical method for generating and conditioning Gaussian simulations, *Math. Geol.*, *32*(6), 701–723.

Liu, J. S. (1996), Metropolized independent sampling with comparisons to rejection sampling and importance sampling, *Stat. Comput.*, *6*(2), 113–119.

Malinverno, A. (2002), Parsimonious bayesian markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophysical Journal International*, *151*(3), 675–688.

Mariethoz, G., and J. Caers (2014), *Multiple-Point Geostatistics: Stochastic Modeling with Training Images*, Wiley-Blackwell, Hoboken, NJ, 376 pp.

Mariethoz, G., P. Renard, and J. Straubhaar (2010), The direct sampling method to perform multiple-point geostatistical simulations, *Water Resources Res.*, *46*(11).

Metropolis, N., M. Rosenbluth, A. Rosenbluth, A. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092.

Mariethoz G., P. Renard, and J. Caers (2010), Bayesian inverse problem and optimization with iterative spatial resampling, *Water Resour. Res.*, *46*, W11530, http://dx.doi.org/10.1029/2010WR009274.

Mosegaard, K., and T. M. Hansen (2015), Inverse methods: Problem formulation and probabilistic solutions, in *Integrated Imaging in Earth Science*, M. Moorkamp, N. Linde, P. Lelievre, and A. Khan, eds., AGU, Washington, DC.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophy. Res.*, *100*(B7), 12,431–12,447.

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, *Comput. Geosci.*, *29*(1), 39–51, doi:http://dx.doi.org/10.1016/S0098-3004(02)00071-7.

Peredo, O., and J. M. Ortiz (2011), Parallel implementation of simulated annealing to reproduce multiple-point statistics, *Comput. Geosci.*, *37*(8), 1110–1121.

Remy, N., A. Boucher, and J. Wu (2008), *Applied Geostatistics with SGeMS: A User's Guide*, Cambridge University Press, New York.

Sambridge, M. (2013), A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.*, p. ggt342.

Sambridge, M., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Rev. Geophy.*, *40*(3), 3–1.

Scales, J. A., and R. Sneider (1997), To Bayes or not to Bayes?, *Geophysics*, *62*(4), 1045–1046.

Shannon, C. E. (1948), A mathematical theory of communication, *ACM SIGMOBILE Mobile Comput Commun. Rev—reprint 2001*, *5*(1), 3–55.

Soares, A. (2001), Direct sequential simulation and cosimulation, *Math. Geol.*, *33*(8), 911–926.

Strebelle, S. (2000), Sequential simulation drawing structures from training images, Ph.D. thesis, Stanford University.

Strebelle, S. (2002), Conditional simulation of complex geological structures using multiple-point statistics, *Math. Geol.*, *34*(1), 1–20.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia.

Tarantola, A., and B. Valette (1982), Inverse problems—Quest for information, *J. Geophys. 50*(3), 150–170.

Tjelmeland, H., and J. Besag (1998), Markov random fields with higher-order interactions, *Scand. J. Stat.*, *25*(3), 415–433.