Case study

Improving in situ data acquisition using training images and a Bayesian mixture model

Mohammad Javad Abdollahifard [a],[*], Gregoire Mariethoz [b], Mohammadreza Pourfard [c]

[a] Electrical Engineering Department, Tafresh University, Tafresh, Iran
[b] School of Civil and Environmental Engineering, University of New South Wales, Sydney, Australia
[c] Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

ABSTRACT

Estimating the spatial distribution of physical processes using a minimum number of samples is of vital importance in earth science applications where sampling is costly. In recent years, training image-based methods have received a lot of attention for interpolation and simulation. However, training images have never been employed to optimize spatial sampling process. In this paper, a sequential compressive sampling method is presented which decides the location of new samples based on a training image. First, a Bayesian mixture model is developed based on the training patterns. Then, using this model, unknown values are estimated based on a limited number of random samples. Since the model is probabilistic, it allows estimating local uncertainty conditionally to the available samples. Based on this, new samples are sequentially extracted from the locations with maximum uncertainty. Experiments show that compared to a random sampling strategy, the proposed supervised sampling method significantly reduces the number of samples needed to achieve the same level of accuracy, even when the training image is not optimally chosen. The method has the potential to reduce the number of observations necessary for the characterization of environmental processes.

1. Introduction

The characterization of spatial phenomena is important for several applications in geosciences. In situ measurements provide critical information when processing remote sensing data and are the basis for interpolating regionalized variables. In many cases data acquisition is expensive. Therefore, an important research goal is to obtain the best possible characterization with the smallest possible number of samples. In this regard, prior knowledge of the spatial variability can be a key asset in designing an efficient sampling strategy, which so far has not yet been fully exploited (Stumpf et al., 2012).

In this paper a method is presented which exploits such prior knowledge in the sampling design. Incorporating spatial prior knowledge in the sampling design can be useful in a range of applications, such as to determine the measurement locations for the estimation of a spatially distributed variable (Wang et al., 2014), to make an optimal choice of training set for remote sensing image classification (Tuia et al., 2009), or for the characterization of subsurface structures based on sparse and expensive drillholes

(Van Groenigen et al., 1999).

The quality of maps obtained by interpolation of observations of a target environmental variable at a limited number of locations is partly determined by the location of the interpolated data (Brus and Heuvelink, 2007). Different sampling design techniques have been proposed attempting to optimize sample locations (Brus and Heuvelink, 2007; Brus and De Gruijter, 1997; Vasat et al., 2010; Delmelle and Goovaerts, 2009; Aerts et al., 2013b, 2013a). Most of these methods are based on minimizing a Kriging variance (Atkinson and Curran, 1995). Brus and Heuvelink (2007) have suggested minimizing the spatially averaged universal Kriging variance using simulated annealing. Vasat et al. (2010) have suggested using Average Kriging Variance for simultaneous optimization of the sampling design for multiple soil variables. Delmelle and Goovaerts (2009) proposed to weight the Kriging variance with another criterion giving greater sampling importance to locations exhibiting significant spatial roughness. A sequential Kriging-based method is proposed by Aerts et al. (2013a,b) to select the sampling pattern for assessment of electromagnetic field exposure. In all of these methods, simple spatial models such as variograms are employed to characterize the behavior of the target field.

On the other hand, multiple-point statistics approaches (MPS) have recently attracted significant attention for the characterization

* Corresponding author.
   E-mail address: mj_abdollahi@aut.ac.ir (M.J. Abdollahifard).

of complex spatial features, with applications in subsurface modeling, remote sensing, rainfall simulations and climate modeling (Tang et al., 2013; Boucher et al., 2008; Wojcik et al., 2009; Jha et al., 2013; Mariethoz et al., 2012; Abdollahifard and Faez, 2013a; Abdollahifard, 2016). In MPS, a training image (TI) is adopted, deemed to encapsulate the spatial characteristics of the target field. It is composed of patterns that are a priori likely to be found in the field. The training image is typically considered as a convenient way of formalizing expert knowledge on the expected spatial continuity, which comes in addition to the available data. Several algorithms exist to complete or interpolate the simulation grid (SG) based on the TI patterns. A classical way of proceeding involves first extracting a pattern database from the TI, and then using it for computing a local conditional probability at each node of the simulation grid. MPS methods are appropriate for generating complex spatial structures, where they usually outperform two-point simulation methods (Zinn and Harvey, 2003; Schlüter and Vogel, 2011).

The spatial model employed in MPS, namely the TI, contains much more information than histograms or variograms. To the best of our knowledge however, the valuable information contained in the TI has not been employed to improve sampling design. The aim of this paper is to fill this gap and use the rich information contained in the TI for an improved sampling process.

The newly developed Compressive Sensing (CS) theory emphasizes that a signal can be completely recovered from much fewer samples compared to the number of samples required in conventional reconstruction schemes (Donoho, 2006; Candès and Wakin, 2008). The idea is to include a prior knowledge of the signal (via a signal model) to obtain a similar reconstruction with fewer samples. The model is intended to capture the common aspects of a signal family in order to minimize the sampling information needed to reconstruct a specific signal from the same family. An evident parallel can be made between such prior knowledge and the TI used in multiple-point geostatistics. In the basic form of CS, the model is provided using a linear dictionary which supports sparse representation of desired signals (Donoho, 2006). In general the dictionary may be either a fixed dictionary (e.g. wavelets or discrete cosine transform (Candès and Wakin, 2008; Griffin and Tsakalides, 2007), or an adaptive linear/nonlinear dictionary tailored to a specific set of signals to allow higher sparsity levels (Olshausen and Field, 1997; Engan et al., 1999; Kreutz-Delgado et al., 2003; Elad and Aharon, 2006; Gangeh et al., 2013; Nguyen et al., 2013). Recently, an analytic manifold model is also developed for modeling image edges and employed for reconstruction of binary geological images with moderate success (Abdollahifard and Ahmadi, 2016).

Besides the sparsity level, the number of necessary samples is a function of the amount of incoherence between sampling and representation bases. Some researchers suggested selecting the sampling basis using the overall signal behavior to reduce the coherence between sampling and representation bases (Elad, 2007; Duarte-Carvajalino and Sapiro, 2009). In these methods, the sampling is decided at an initial stage, and therefore the sampling basis does not depend on the locations and values of previous samples. Furthermore, there exist a large amount of research on adaptive compressive sensing (Aldroubi et al., 2011; Ke et al., 2010; Soni and Haupt, 2012; Arias-Castro et al., 2013) and task-driven compressive sensing (Duarte-Carvajalino et al., 2013; Ashok et al., 2008; Baheti and Neifeld, 2009; Mahalanobis and Muise, 2009; Ashok et al., 2013). It is worth emphasizing that these methods are all specific to the field of signal processing and have never been applied to geostatistical problems. In geoscience applications, it is common to acquire samples directly and sequentially as a series of successive measurement campaigns. The direct sampling constraint prevents the application of any desired sampling basis, and just allows the selection of sampling locations.

Nonparametric Bayesian models are powerful probabilistic signal modeling tools that have been extensively employed in the context of CS (Abdollahifard and Faez, 2013b; Bishop and Winn, 2000; Paisley et al., 2010; Chen et al., 2010; Ji et al., 2008; Yu and Sapiro, 2011). Non-parametric Bayesian models have been employed to model nonlinear data manifolds in different applications (Abdollahifard and Faez, 2013b; Bishop and Winn, 2000; Chen et al., 2010). As will be discussed in Section 4, such manifold models, namely Mixture of Probabilistic Principal Component Analyzers (MPPCA) and Mixture of Factor Analyzers (MFA), are capable of representing complex data structures such as spatial heterogeneity. Since such manifold models are flexible enough to be fitted on nonlinear data spaces (such as defined by a TI) and to capture data space tightly, they provide important modeling information which allows reconstructing the complete signal from a tiny number of samples. However, the location of samples significantly affects the reconstruction performance. This sensitivity can be used to determine which sampling locations have the highest potential to bring useful spatial information. It should be noted that many other probabilistic modeling tools developed in image processing literature have the potential to be used to address sampling design problem based on training images.

The main contribution of our paper is to employ multiple-point statistics for sampling design. We believe that two important factors have to be considered to select the location of samples:

1. the signal behavior (which can be modeled using a manifold model based on the training patterns of a TI),
2. the locations and the values of the previously drawn samples.

As discussed above, none of the previous studies considered both these factors simultaneously. We propose a greedy sampling approach whereby in a first step, the TI patterns are used to train a MFA model. Then, based on this model and a limited number of initial samples, uncertainty values are estimated on the entire interpolation domain. Next, new samples are extracted from local maxima of uncertainty, and the operation is repeated sequentially. The structure of the proposed method is similar to that of Kriging-based methods of Aerts et al. (2013a,b). The MFA estimated value and the uncertainty are equivalent to the mean and variance of Kriging. However, unlike Aerts et al. (2013a,b), we have included concepts coming from multiple-point statistics in our model through the training image. As illustrated in Section 6, the improved model employed in our method leads to a significantly more efficient sampling design. It should be noted that the proposed sampling method is sequential, and hence difficult to apply or inapplicable to dynamic fields which show fast time variations, e.g. air quality monitoring (Romary et al., 2011, 2014).

The paper is organized as follows. In Section 2 a brief review of basic concepts of compressive sensing is presented. In Section 3, an overview of the proposed method is presented and motivated using a simple example. In Section 4, the typical locus of spatial patterns is discussed and the MFA model is introduced as a suitable tool for the modeling of such loci. In Section 5, the proposed algorithm is presented in detail. Finally, Section 6 evaluates the method on a series of test cases.

2. Background on compressive sampling

Although compressive sensing has received considerable attention in the signal processing community, geostatisticians and geoscientists are less familiar with basic concepts in this field. In this section, the CS basics are briefly introduced avoiding unnecessary formulations. CS is a technique for efficiently

reconstructing a d-dimensional signal t (a vector containing d components) from a set of observations $\chi \in \mathbb{R}^n$ where $d > n$. In general this leads to an underdetermined system with infinite number of solutions. By assuming sparseness or compressibility of the signal in some domain, a specific solution which is also sparse or compressible in that domain could be selected as final solution. Using this technique nearly all useful information contained in the signal can be retrieved using a small number of compressive measurements.

Assume that the information content of t is obtained through linear measurements so that $\chi = \Phi t$ ($\Phi \in \mathbb{R}^{n \times d}$). Furthermore, assume that t can be expressed in the representation basis W in the form $t = Wv$ where $v \in \mathbb{R}^l$ is sparse meaning that it has only a few nonzero elements ($l \geq d$). Combining the above equations leads to an underdetermined linear system in the form $\chi = \Phi Wv = Av$. The sparsest solution for this system can be obtained by solving the following optimization problem:

$$\hat{v} = \text{argmin}_v \| v \|_{\ell_0} \quad \text{subject to} \quad \chi = Av \tag{1}$$

where $\| v \|_{\ell_0}$ denotes the number of nonzero elements of v. The above optimization problem is a NP-hard one. An approximate solution is to replace the ℓ_0 norm with ℓ_1 norm. It has been shown that the ℓ_1 norm leads to sparse solutions as well (Donoho, 2006; Candès and Wakin, 2008).

Two important factors determine the effectiveness of CS: the representation basis, W, and the sampling basis, Φ. The sparser the signal in representation domain, the fewer samples are required for complete signal recovery. A common dictionary, like DCT, represents some signals in a sparse form. However, dictionary learning techniques attempt to obtain sparser representations through adapting the dictionary to a specific class of signals (Olshausen and Field, 1997; Engan et al., 1999; Kreutz-Delgado et al., 2003; Elad and Aharon, 2006). On the other hand, the sampling basis should be selected incoherent with representation basis as much as possible (Donoho, 2006). If the sampling and representation domains are the same (maximum coherence), each sample only motivates one location in representation domain and it is impossible to recover unsampled values. On the other hand, one sample in the sampling domain contains a lot of information in the representation domain if the sampling and representation domains are highly incoherent.

As indicated before, in geoscience applications the samples are usually extracted directly and only their locations can be decided. In other words, each row of Φ contains zeros except in the sampling location. Since the number of available samples is very limited in our application, it is necessary to select a suitable representation domain and to decide the location of samples carefully.

Linear transformations have limited potential for compressive representation of signals. Nonlinear transformations provide much more flexibility to capture complex signal spaces. Kernelized dictionary learning methods are available in the literature, which allow high degree of compression (Gangeh et al., 2013; Nguyen et al., 2013). However, it seems difficult to use them to estimate uncertainties. As an alternative, a mixture of linear models can be used to model a nonlinear space using a set of tangent hyperplanes. Convenient mathematical properties of such models make them more suited to estimate uncertainty values.

3. Algorithm overview

Consider an unknown scalar field $Z(x)$, where $x = [x_1, x_2]^T$ denotes a location. Suppose that a small number of samples are extracted from $Z(x)$ at random (Fig. 1(a)). The value of a field at a specific location is mainly determined by its neighboring pixels and is almost independent from far pixels. Therefore, in order to make the modeling problem tractable, the field can be divided into small patches. Ψ_x denotes a $k \times k$ patch centered at pixel x, called a *data-event* ($d = k^2$). In order to preserve pattern continuity, a minimum overlap of 50% is considered for neighboring patches in this paper. Considering overlapping patches to prevent edge effects is a common practice in image processing literature (see for example Peyré, 2009).

Vectorizing the data-event leads to a vector t ($\in \mathbb{R}^d$) with a known part denoted by t_{ob} ($\in \mathbb{R}^{d_{ob}}$) and a missing part denoted by t_{mis} ($\in \mathbb{R}^{d_{mis}}$). The known part is typically informed by data that are already acquired. If the observed samples provide enough information for reconstruction of missing values with desired accuracy, then no new samples will be necessary. However, initial samples are often so sparse that more samples are required to achieve the desired accuracy. The goal is to find the best locations for extraction of new samples from the field.
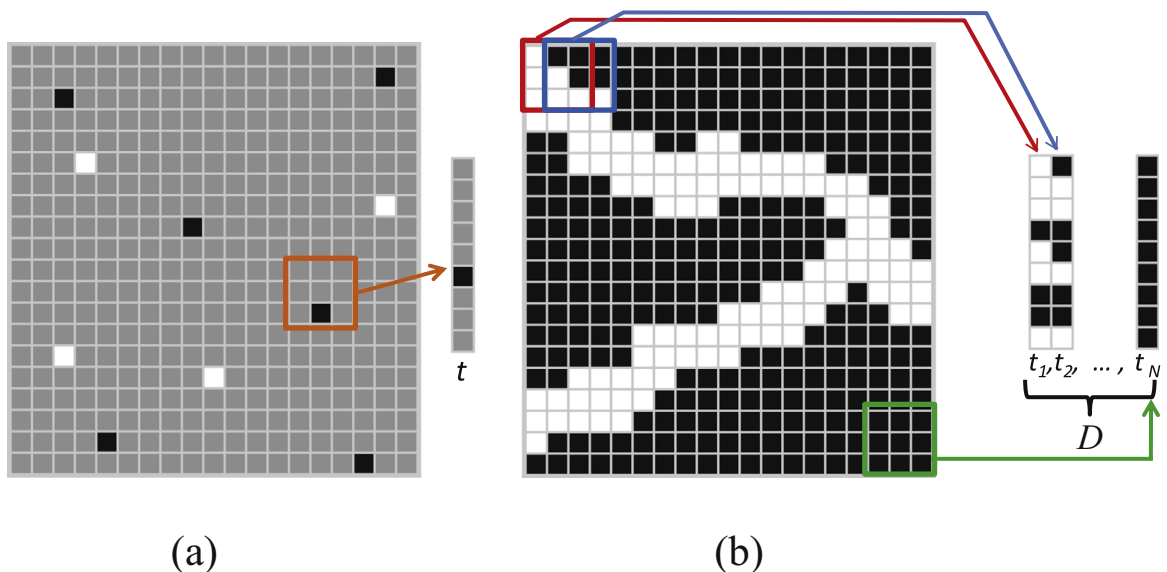

(a)   (b)

Fig. 1. (a) An unknown binary field $Z(x)$ with a small number of observed samples. Unknown values are depicted in gray. A typical data-event is depicted besides the image. (b) A training image $TI(y)$ ($y = [y_1, y_2]^T$). TI patterns are vectorized to form a data-base of training vectors, D.
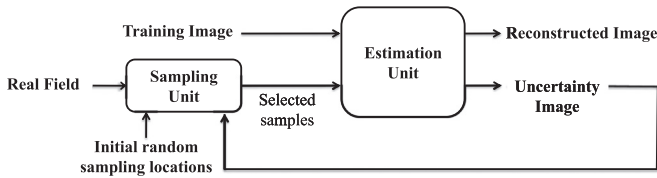
Fig. 2. Workflow of proposed sampling scheme.

As in multiple-point statistics simulation methods, it is assumed that there exists an expert-provided example field $TI(y)$ ($y \in \mathbb{R}^2$) which describes the spatial behavior of the field, i.e. the training image (see Fig. 1(b)). The TI contains patterns that are likely to be found in the real field. A large number of $k \times k$ training patterns (Ψ_y s for all y in the TI domain) are extracted from the TI, and then vectorized to form a data-base of training vectors $D = [t_1, …, t_N]$. The training vectors extracted from the TI are highly correlated and compressible. As will be discussed in Section 4, the training patterns are modeled using a probabilistic model, and then the model is used to estimate unknown values and uncertainty values all over the field.

In this paper a Bayesian approach is adopted. Given t_{ob}, the posterior probability distribution function $p(t_{mis}|t_{ob})$ is computed and then the mean and the variance of this distribution are taken as estimations for missing values and uncertainty values respectively. By putting the uncertainty values together, an uncertainty image $u(x)$ is formed and new samples are extracted from local maxima of $u(x)$. Fig. 2 illustrates the proposed workflow. Using new samples, the estimated variables are refined and the process continues until no more data need to be measured.

3.1. An illustrative example

Our sampling approach is founded on two important principles: first, the underlying field behavior is employed to conduct the sampling process through a training image. Second, the samples are taken sequentially and the locations and values of previous samples take part in the designation of new samples locations. In this section, the significance of such an approach is illustrated using a simple example. Consider the basic training image of Fig. 3(a). As indicated previously, it is assumed that the TI is composed of patterns that are likely to be found in the real field. The goal is to reconstruct a patch belonging to the family of 9×9 patterns of this image using a minimum number of samples.

Given the signal behavior, it is obvious that if two samples are located on a horizontal line their values are equal and one of them is practically useless. This example emphasizes that the proper location of the samples depends on the signal behavior. On the other hand, if two samples are drawn from top and bottom of a patch and both of them are white (or black) then the other points of the patch can be estimated precisely. But if one of the samples is white and the other one is black, then more samples are required to find the exact location of the border (Fig. 3(a)).

This example demonstrates that both the locations and the values of the previous samples can be employed for localization of
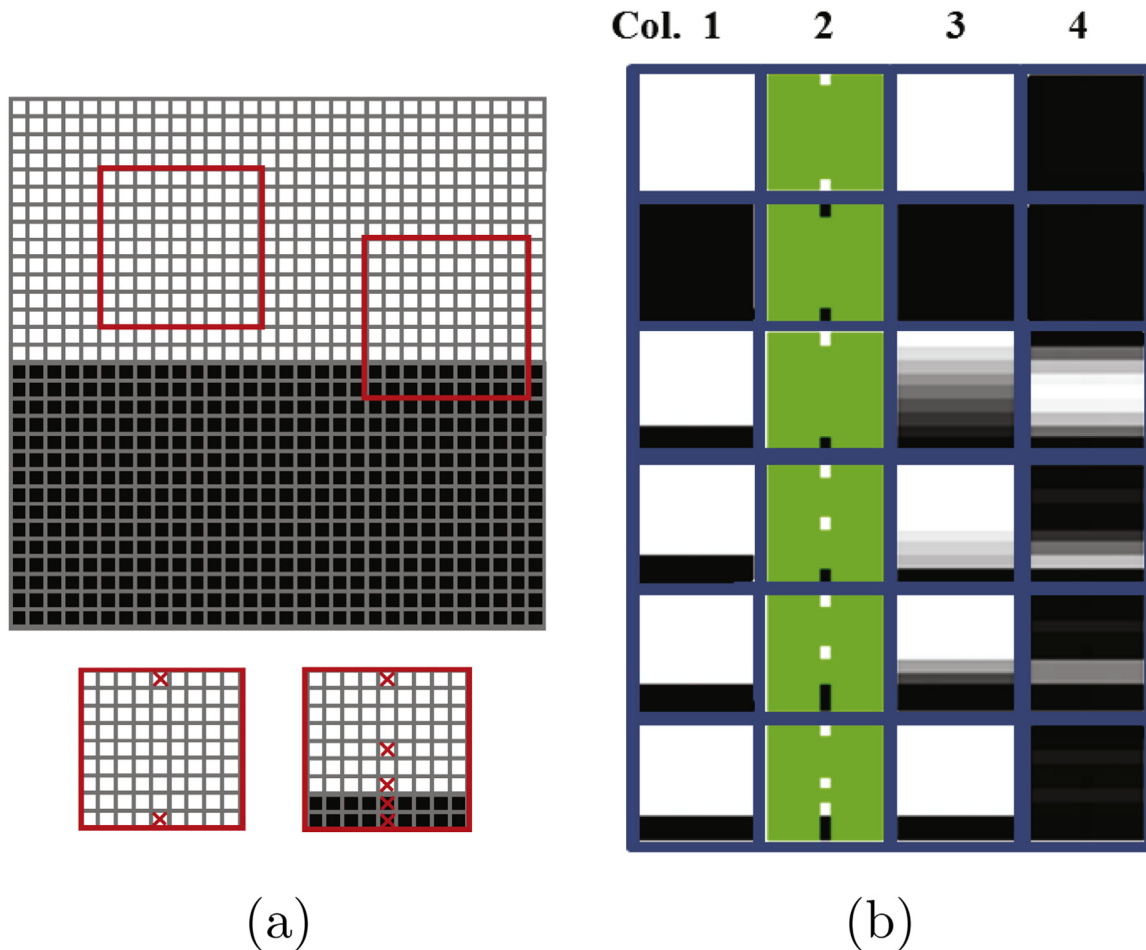

(a)   (b)

Fig. 3. Illustration of importance of image behavior, samples locations and values in sequential sampling. (a) A simple image and two extracted patches. (b) Sampling and reconstruction of 9×9 patches based on the proposed method (the details are given later in Section 5). Original patches, extracted samples, reconstructed patches, and scaled variance maps are depicted in columns 1–4 from left to right respectively.

new samples. It also shows that it is beneficial to select the samples in a sequential manner.

A brute-force way of choosing samples locations would be to generate, at each sampling stage, a large number (say K) of realizations conditioned to any previously available samples. Then, based on the variance map through the realizations, one could choose the best locations for further sampling. However, such an approach would require a large number of realizations and therefore would be very CPU intensive. Abdollahifard and Faez (2013b) showed that a Bayesian mixture model can be effectively employed for simulation, with computational cost comparable to state of the art simulation methods. In this paper, it is shown that the Bayesian mixture model can be a more viable alternative to such brute-force approach, because in its context computing an additional variance map does not have a significant effect on the computational cost. We therefore propose a method where the uncertainty map can be computed directly, while the mentioned brute-force method requires K simulations for computing the uncertainty map, where K would be in the order of 100.

4. Modeling

Different patterns belonging to a single TI are typically correlated and therefore compressible using an appropriate transform. Assuming a low-rank Gaussian distribution for t, it could be represented as a linear function of a low-dimensional feature vector v ($\in \mathbb{R}^q$) as follows:

$$t = Wv + \mu + \epsilon, \tag{2}$$

where v is a zero-mean Gaussian variable $v \sim \mathcal{N}(0, I_q)$ and $q \ll d$. μ ($\in \mathbb{R}^d$) is the mean vector, W ($\in \mathbb{R}^{d \times q}$) is the transformation matrix and ϵ is a Gaussian noise vector with precision (i.e. inverse variance) of τ, $\epsilon \sim \mathcal{N}(0, \tau^{-1} I_d)$ where $\tau \in \mathbb{R}$.

Based on this model, $p(t_{mis}|t_{ob})$ can be computed. Assuming a single Gaussian distribution, one can simply verify that the locations with maximum variance will depend on locations of previous samples and not on their values. This is contradictory with our observation in the example of Fig. 3 and hence this model is not sufficient for our modeling problem. In this section we first discuss the possible forms of TI patterns locus in d dimensional space. Then the proper model for modeling such locus is introduced.

4.1. TI patterns locus

The TI patterns could be considered as points in a d-dimensional space. However, their locus does not occupy the whole space and is restricted to a limited sub-space. The key observation is that the spatial behavior of underground fields is repeated with minor changes in different locations. Consider the binary image of Fig. 4 which is a simplification of a channelized depositional system with white as sand and black as shale (Strebelle, 2002). With few exceptions, all of 10×10 patches of the image can be constructed by translating and/or rotating the simple patch depicted besides the image. Similarly, many variables in geosciences have the property that their patches can be reconstructed by applying some simple wrapping on one or more simple patches (Yu et al. (2012) described natural image patches in the same way to solve inverse problems).

To further clarify and visualize the locus, let us consider a simpler example taken from Chen et al. (2010). Fig. 5(a) shows a 33-point discrete signal obtained from regular sampling of a Gaussian signal with mean a. This signal can be considered as a point in a 33-dimensional coordinate system. By changing a, a family of signals is obtained. For visualization, 3 random

Fig. 4. Nearly all patterns of this binary channelized image (251 by 251 pixels) can be reconstructed by applying simple mappings (like rotation and translation) to the elementary pattern depicted besides the image. Therefore the space of the patterns in this image has very low effective dimensionality.

components of this 33-element vector are selected and considered as a point in a three-dimensional space. For different values of a the locus of such points is depicted in Fig. 5(b). Note that in this plot the color spectrum, from blue to red, is intended to indicate the value of a. From this figure it can be seen that although the locus is placed in a 3-dimensional space, its effective dimensionality is one. The reason is that each signal can be obtained by applying a 1D translation to a reference signal.

Similarly, the 10×10 patches of the TI of Fig. 4 can be seen as distributed on a manifold with effective dimensionality of three (two for translation and one for rotation) in the 100-dimensional space. Therefore, the patches of such images are distributed on nonlinear manifolds with very low effective dimensionality. A modeling tool is needed to achieve such dimensionality reduction. This model should be capable of capturing the signal space as tightly as possible to provide maximum modeling information and therefore minimize the sampling needs.

4.2. MFA model

As indicated before, a single Gaussian distribution is not enough to model a nonlinear manifold. Bayesian hierarchical modeling, which can also be seen as a mixture of Gaussians, was previously employed in classical geostatistics for clustering geostatistical data (Allard and Guillot, 2000). Mixture models can also be employed to model nonlinear manifolds. Non-parametric Bayesian mixture models like MPPCA and MFA have been successfully employed to approximate different data manifolds using mixture of tangent hyper-planes (Abdollahifard and Faez, 2013b; Bishop and Winn, 2000; Chen et al., 2010). These models are flexible and can adapt to the signal space during a training phase by changing their parameters. Suppose that using a set of observed two-element signals we aim to model their space and then use the model for reconstruction of new incomplete signals of this family. The hypothetical scatter plot of such signals is depicted in Fig. 6, where each point denotes one observed two-dimensional signal. These observations can be used as training data to describe the signal space. In this example, the nonlinear signal space is approximated properly using a mixture model with 6 components.

In this paper, a non-parametric MFA model is employed for modeling of the signal space. Here we use the model of Chen et al. (2010), which we rapidly summarize in the following. The generative equation of a single factor model is given in Eq. (2). Before generalizing Eq. (2) to a mixture model, we address the problem of inferring the effective dimensionality (rank of Gaussians, i.e. q) of signal space. To this end, the feature vector $\hat{v}$ is permitted to have
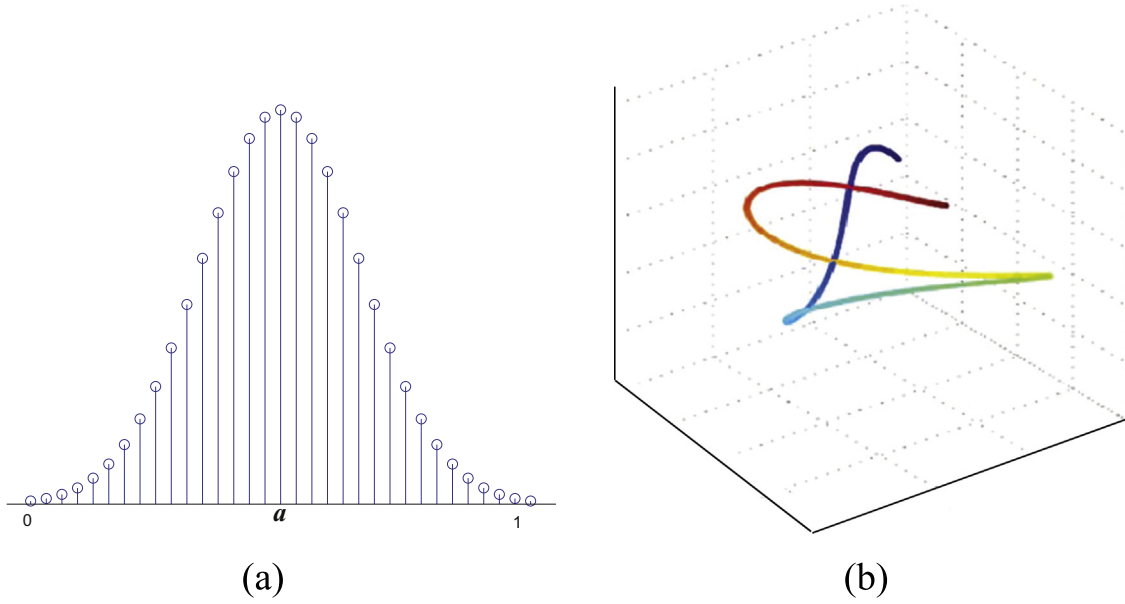
Fig. 5. (a) A 33-point discrete signal. Consider a family of such signals obtained by varying the mean a. Although such signals are 33 dimensional, their effective dimensionality is one because they only depend on a. (b) 3 random components are selected from 33 components of the signal and their locus is drawn for different values of a. As expected, the signal is distributed on a nonlinear curve of dimensionality one. This example is taken from Chen et al. (2010).
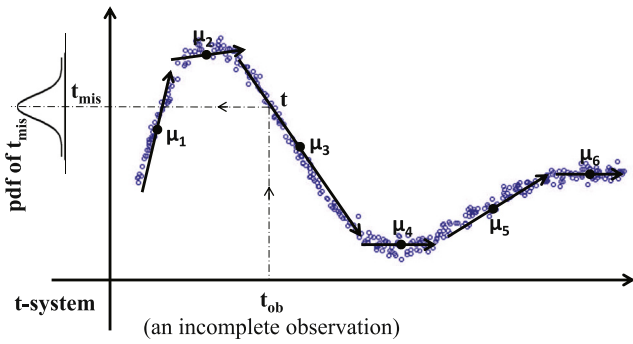

Fig. 6. Scatter plot of a set of 400 two-dimensional signals (points) distributed on a nonlinear manifold with effective dimensionality of one. The manifold is modeled using 6 one-dimensional subspaces, with origins $\mu_1, ..., \mu_6$. Using this model, new incomplete signals (like t_{ob}) can be reconstructed by projecting them onto the model. In addition to estimating missing values t_{mis}, the pdf of these missing parts can also be determined.
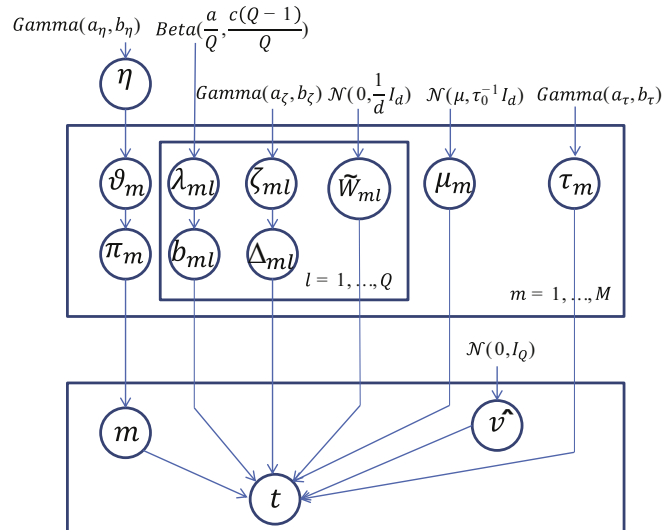

Fig. 7. Graphical model of the MFA model.

the dimension Q where $q \ll Q \leq d$, and then some of its components are deactivated as follows (Chen et al., 2010):

$$v = \hat{v} \circ b, \quad \hat{v} \sim \mathcal{N}(\mathbf{0}, I_Q), \quad b \sim \prod_{l=1}^{Q} Bernoulli(\lambda_l) \tag{3}$$

where $\circ$ denotes point-wise vector product. A Beta prior distribution is assigned to λ:

$$\lambda \sim \prod_{l=1}^{Q} Beta(a/Q, c(Q-1)/Q) \tag{4}$$

where the hyper-parameters a and c can be used to impose prior belief on the number of factors.

The factor model can be generalized to a mixture of M factors as follows:

$$t = W_m v + \mu_m + \epsilon, \tag{5}$$

where $m = 1, ..., M$ is the mixture index. In fact, t is considered to have a mixture distribution as follows:

$$p(t) = \sum_{m=1}^{M} \pi_m \mathcal{N}(t|\mu_m, \Sigma_m), \tag{6}$$

where π_m is the probability of mth mixture component and Σ_m is the covariance matrix of mth component. The graphical model of MFA is depicted in Fig. 7. The hierarchical model is represented as follows (Chen et al., 2010):

$$p(t|v, W_m, \mu_m, \tau_m) = \mathcal{N}(W_m v + \mu_m, \tau_m^{-1} I_d), \tag{7}$$

$$v = \hat{v} \circ b_m, \quad W_m = \bar{W}_m \Delta_m, \tag{8}$$

$$p(\hat{v}) = \mathcal{N}(\mathbf{0}, I_Q), \tag{9}$$

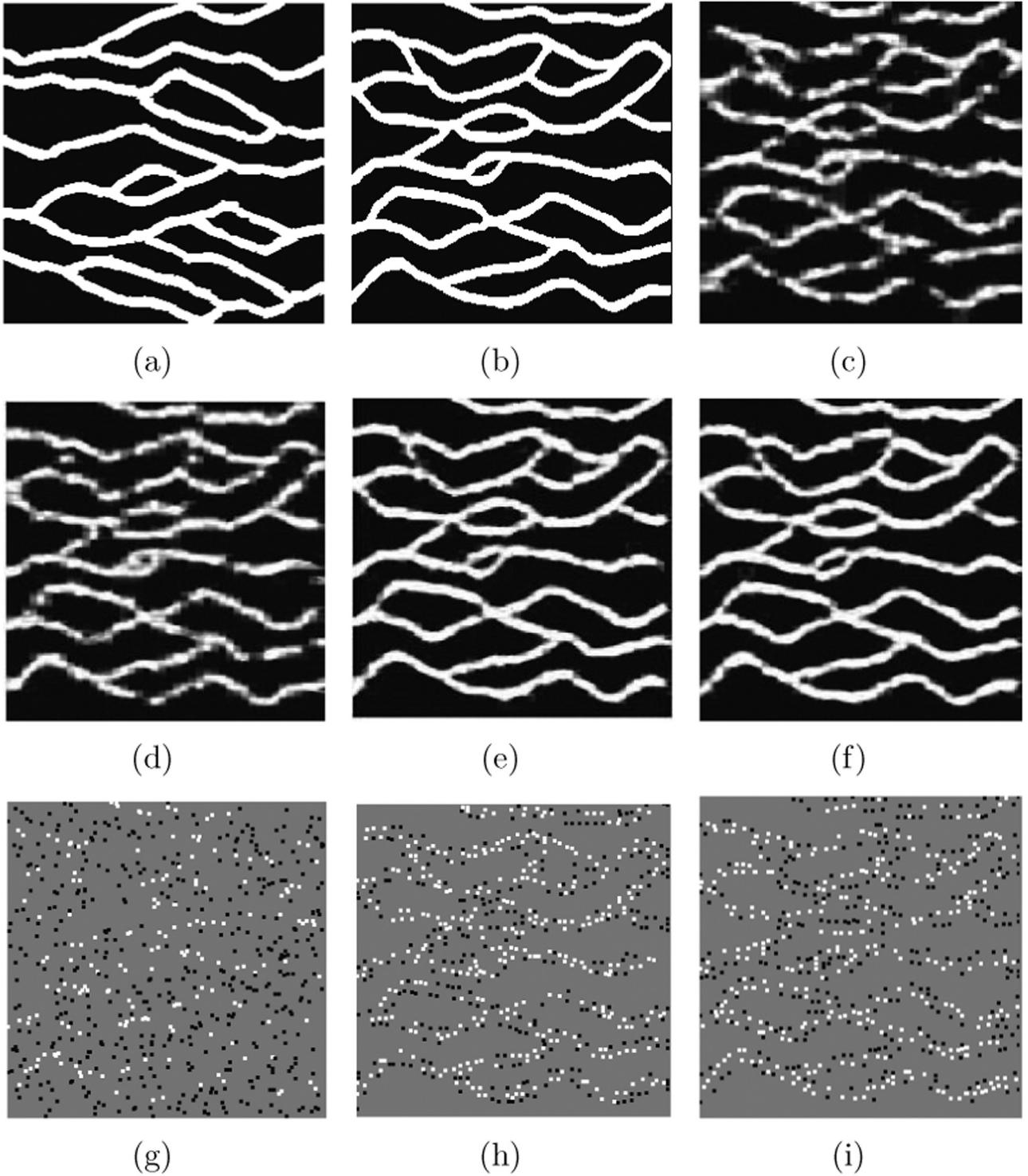$$p(m|\pi_1, ...\pi_M) = Mult(1; \pi_1, ...\pi_M), \tag{10}$$

Fig. 8. Comparison of reconstruction performance using different sampling strategies on a channelized image. (a) Training image *TI*. (b) Hypothetical real field *Z*. (c), (d), (e) and (f) An image reconstructed using 3% samples extracted randomly, based on Kriging variance, based on MPS simulation variance and based on MFA variance (proposed method) respectively. (g), (h) and (i) samples extracted at 3rd sampling step using Kriging-based, simulation-based and MFA based samplings respectively. Note that samples are depicted larger than one pixel to become clearly visible.

$$\pi_m = \vartheta_m \prod_{k=1}^{m-1} (1 - \vartheta_k), \quad \pi_M = 1 - \sum_{m=1}^{M-1} \pi_m, \tag{11}$$

$$p(\vartheta_m | \eta) = Beta(1, \eta), \tag{12}$$

$$p(\boldsymbol{b_m} | \lambda_{\boldsymbol{m}}) = \prod_{l=1}^{Q} Bernoulli(\lambda_{ml}), \tag{13}$$

$$p(\lambda_{\boldsymbol{m}}) = \prod_{l=1}^{Q} Beta(a/Q, c(Q-1)/Q), \tag{14}$$
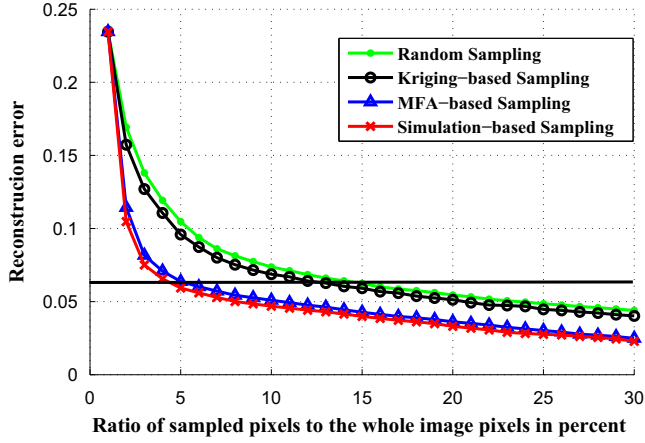
Fig. 9. Reconstruction error for random and supervised samplings for sampling rate of 1–30% for the experiment of Fig. 8.

$$p(\boldsymbol{\mu_m}|\tau_0) = \mathcal{N}(\mu, \tau_0^{-1}\boldsymbol{I_d}), \tag{15}$$

$$\tilde{\boldsymbol{W}}_{\boldsymbol{m}} = \prod_{l=1}^{Q} \mathcal{N}\left(\boldsymbol{0}, \frac{1}{d}\boldsymbol{I_d}\right), \tag{16}$$

$$p(\boldsymbol{\Delta_m}|\zeta_{m1}, ..., \zeta_{mQ}) = \prod_{l=1}^{Q} \mathcal{N}(0, \zeta_{ml}^{-1}), \tag{17}$$

Note that $\boldsymbol{\Delta_m}$ is a diagonal matrix whose lth diagonal element is drawn from $\mathcal{N}(0, \zeta_{ml}^{-1})$. The diagonal elements of $\boldsymbol{\Delta_m}$ encode the importance of each column of $\tilde{\boldsymbol{W}}_{\boldsymbol{m}}$. *Mult*() denotes the multinomial distribution and μ is the mean vector of all training vectors. Eq. (16) means that each column of $\tilde{\boldsymbol{W}}_{\boldsymbol{m}}$ is drawn independently from $\mathcal{N}(\boldsymbol{0}, (1/d)\boldsymbol{I_d})$. In this model a Dirichlet process is employed to automatically infer the number of necessary clusters, M, from the training data. The Dirichlet distribution is a conjugate prior for the parameters of a multinomial distribution. As depicted in Fig. 7, Gamma priors are considered for τ_ms, η, and ζ_{ml}s: $\tau_m \sim Gamma(a_\tau, b_\tau)$, $\eta \sim Gamma(a_\eta, b_\eta)$, $\zeta_{ml} \sim Gamma(a_\zeta, b_\zeta)$. The hyper-parameters τ_0, a_τ, b_τ, a_η, b_η, a_ζ and b_ζ are all set to 10^{-6}. Furthermore we set $a = c = 1$. Note that hyper-parameters are set to very small values to have non-informative priors (Chen et al., 2010).

Based on this model the pdf of t could be represented as a mixture of low-rank Gaussians with covariance matrix Σ_m:

$$\boldsymbol{\Sigma_m} = \boldsymbol{W_m}\boldsymbol{W_m^T} + \tau_m^{-1}\boldsymbol{I_d} = \tilde{\boldsymbol{W}}_{\boldsymbol{m}}\tilde{\boldsymbol{\Delta}}_{\boldsymbol{m}}^2\tilde{\boldsymbol{W}}_{\boldsymbol{m}}^T + \tau_m^{-1}\boldsymbol{I_d}, \tag{18}$$

where $\tilde{\boldsymbol{\Delta}}_{\boldsymbol{m}} = \boldsymbol{\Delta_m} \, diag\,(b_{m1}, ..., b_{mQ})$. Therefore, the rank of the mth Gaussian is approximately equal to the number of nonzero elements of $\boldsymbol{b_m}$, denoted by q_m. The employed MFA model is so complicated that the posterior distribution cannot be computed analytically. To handle this problem, approximate inference methods like Gibbs sampling are employed to infer the model parameters using the training data (Abdollahifard and Faez, 2013b; Chen et al., 2010).

5. Sequential compressive sampling

Suppose that a mixture of low-rank subspaces is fitted to the signal locus (e.g. consider the fitted model in Fig. 6), and the

missing part of an incomplete signal t needs to be completed using this model. Denote the largest subspace dimensionality by $q = \max_m\{q_m\}$. Intuitively, it seems that if the number of observed elements of t, i.e. d_{ob}, is greater than or equal to q, then t_{mis} could be reconstructed with acceptable certainty, e.g. consider the way in which the missing part of a two-dimensional signal is reconstructed in Fig. 6.

More formally, once the training phase is completed, the computed model parameters including π_ms, $\boldsymbol{\mu_m}$ s, $\boldsymbol{W_m}$ s, τ_ms, and $\boldsymbol{b_m}$ s are treated as fixed values in reconstruction and sampling phases. Now suppose that an observation vector $\chi \in \mathbb{R}^n$ is given in the following form:

$$\chi = \boldsymbol{\Phi}t, \tag{19}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{n \times d}$ and $n \ll d$. If all elements of each row of $\boldsymbol{\Phi}$ are zeros except for one, χ will be a small subset of t. The posterior distribution of t, given the observation vector χ, is a Gaussian mixture model as well (Chen et al., 2010):

$$p(\boldsymbol{t}|\chi) = \sum_{m=1}^{M} \tilde{\pi}_m \mathcal{N}(\boldsymbol{t}|\tilde{\boldsymbol{\mu}}_{\boldsymbol{m}}, \tilde{\boldsymbol{\Sigma}}_{\boldsymbol{m}}), \tag{20}$$

where

$$\tilde{\pi}_m = \frac{\pi_m \mathcal{N}(\chi|\boldsymbol{\Phi}\boldsymbol{\mu_m}, \boldsymbol{\Phi}\boldsymbol{\Sigma_m}\boldsymbol{\Phi^T})}{\sum_{l=1}^{M} \pi_l \mathcal{N}(\chi|\boldsymbol{\Phi}\boldsymbol{\mu_l}, \boldsymbol{\Phi}\boldsymbol{\Sigma_l}\boldsymbol{\Phi^T})}, \tag{21}$$

$$\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{m}} = \boldsymbol{\Sigma_m} - \boldsymbol{\Sigma_m}\boldsymbol{\Phi^T}(\boldsymbol{\Phi}\boldsymbol{\Sigma_m}\boldsymbol{\Phi^T})^{-1}\boldsymbol{\Phi}\boldsymbol{\Sigma_m}, \tag{22}$$

$$\tilde{\boldsymbol{\mu}}_{\boldsymbol{m}} = \boldsymbol{\mu_m} + \boldsymbol{\Sigma_m}\boldsymbol{\Phi^T}(\boldsymbol{\Phi}\boldsymbol{\Sigma_m}\boldsymbol{\Phi^T})^{-1}(\chi - \boldsymbol{\Phi}\boldsymbol{\mu_m}). \tag{23}$$

In a Bayesian approach, missing values of the signal t are estimated using the posterior distribution. The estimated signal, $\hat{t}$, can be computed as the expectation of t with respect to $p(t|\chi)$:

$$\hat{\boldsymbol{t}} = E_{t|\chi}(\boldsymbol{t}) = \int \boldsymbol{t}p(\boldsymbol{t}|\chi)\,d\boldsymbol{t} = \tilde{\mu} \tag{24}$$

It is easy to show that the mean vector and the covariance matrix of the Gaussian mixture model (20) can be expressed as follows:

$$\tilde{\mu} = \sum_{m=1}^{M} \tilde{\pi}_m \tilde{\boldsymbol{\mu}}_{\boldsymbol{m}}, \tag{25}$$

$$\tilde{\Sigma} = \sum_{m=1}^{M} \tilde{\pi}_m\{\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{m}} + (\tilde{\boldsymbol{\mu}}_{\boldsymbol{m}} - \tilde{\mu})(\tilde{\boldsymbol{\mu}}_{\boldsymbol{m}} - \tilde{\mu})^T\}. \tag{26}$$

$\tilde{\mu}$ can be regarded as an estimation for the signal t. The diagonal elements of the covariance vector $\tilde{\Sigma}$ are the estimated variance or uncertainty of signal elements. These uncertainty values are gathered in a vector $\boldsymbol{u} \in \mathbb{R}^d$. The mean (expectation) and the uncertainty of the posterior distribution could be compared with Kriging mean and variance respectively, as used for example by Diggle et al. (2007).

Consider again the image of Fig. 3(a). At the beginning, a MFA model is trained using all patches of this image. Then some incomplete patches are used to estimate the mean and variance of the whole patch using Eqs. (25) and (26), as shown in Fig. 3(b). As expected, in the first two rows of Fig. 3(b), where both top and bottom samples have the same color, the estimated variance is zero in the whole patch. Low variance values indicate a high level of certainty for the patch values in column 3 of Fig. 3(b). In other rows, more uncertainty is observed in specific locations of the patch. Consider for example the third row of Fig. 3(b): based on the training image, it is known that if the top sample is white and the bottom sample is black, a sharp border occurs somewhere in
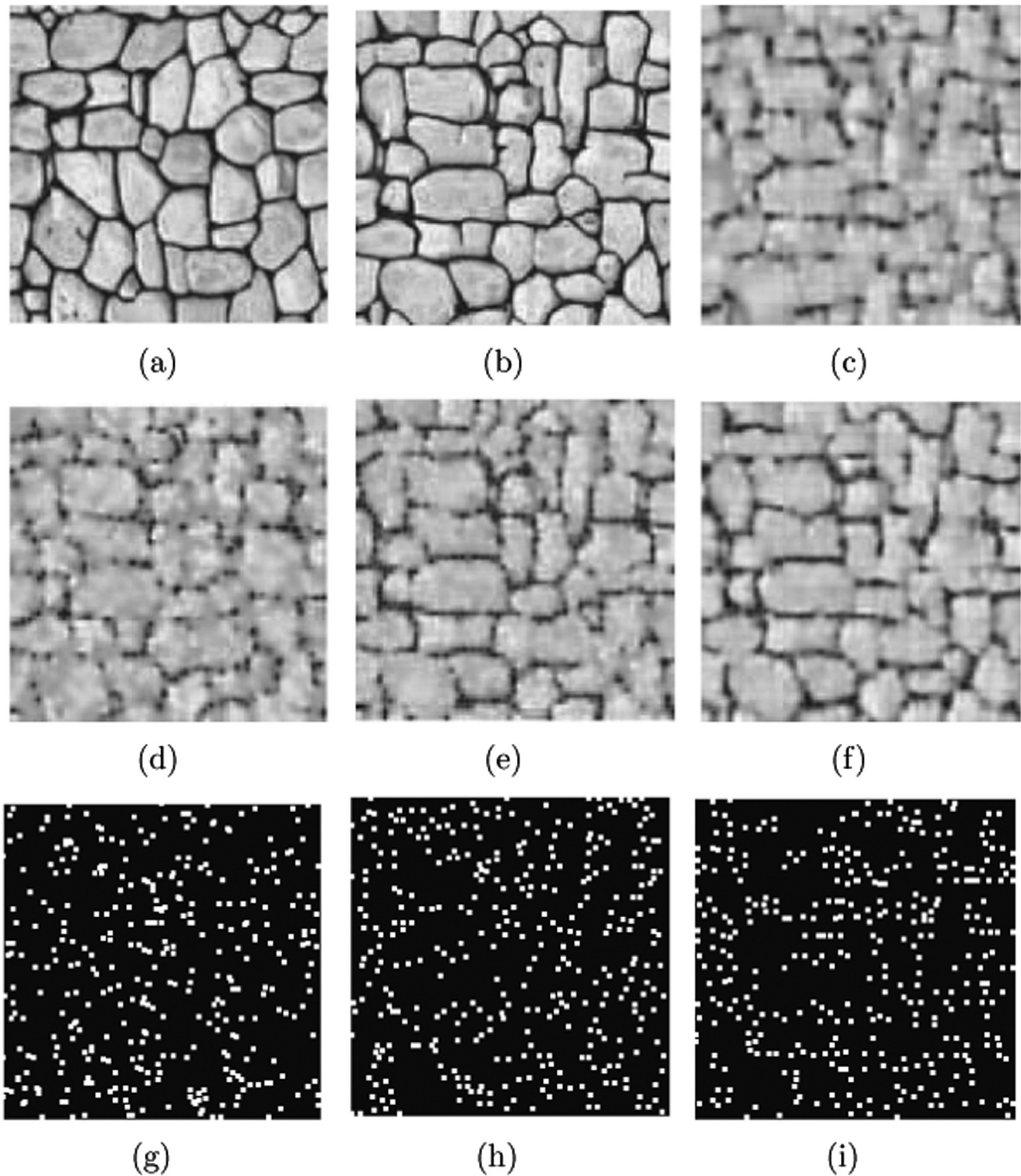
Fig. 10. Comparison of reconstruction performance using different sampling strategies on a continuous image. (a) Training image *TI*. (b) Hypothetical real field *Z*. (c), (d), (e) and (f) An image reconstructed using 3% samples extracted randomly, based on Kriging variance, based on MPS simulation variance and based on MFA variance (proposed method) respectively. (g), (h) and (i) samples extracted at 3rd sampling step using Kriging-based, simulation-based and MFA based samplings respectively. Unlike previous example all sample locations are depicted in white regardless of their values.

between, but the exact location of the border is unknown. Hence, the model is uncertain about estimated patch values. In such cases more samples must be drawn to explore this boundary, which is precisely what our model does. In each row, the point with maximum uncertainty is sampled and its value is used in the following row. Finally, the exact reconstruction can be achieved using 5 samples (last row). Interestingly, the location of selected samples is optimal from a subjective point of view in this simple example.

5.1. Algorithm

Our algorithm reconstructs a scalar field, $Z(x)$, using a limited amount of samples, where x is a 2- or 3-dimensional position vector. It has three steps:

Modeling step: A MFA model is trained using the TI patterns. A large number of overlapping $k \times k$ patterns are extracted from the TI (see Fig. 1). Then, the parameters of the MFA model are
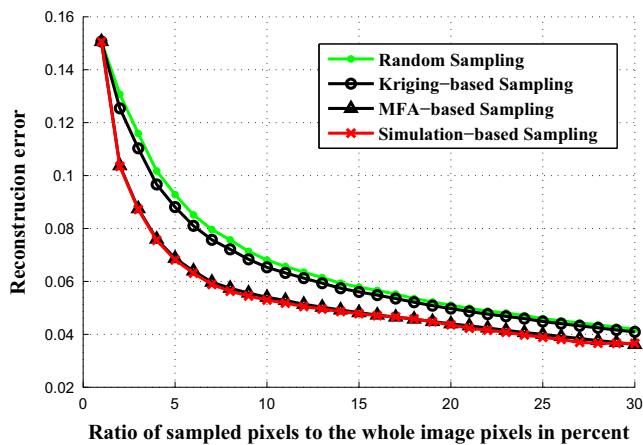
Fig. 11. Reconstruction error for random and supervised samplings for sampling rate of 1–30% for the experiment of Fig. 10.

estimated using a Gibbs Sampler based on training patterns (Abdollahifard and Faez, 2013b; Bishop and Nasrabadi, 2006).

Estimation step: The reconstructed image, $\hat{Z}(x)$, and the uncertainty estimate, $\hat{u}(x)$, are produced based on the parameters computed in the modeling step. $Z(x)$ is divided into $k \times k$ windows. Each window is composed of a very limited amount of observed samples and a majority of unknowns (see Fig. 1). For each window, a specific sampling matrix Φ can be found which specifies the location of the samples (see (19)). Using Eqs. (25) and (26), the mean and the variance are estimated at each pixel of each window. $\hat{Z}(x)$ and $\hat{u}(x)$ are obtained by putting the reconstructed windows together. Note that the windows can be selected with some overlap to prevent inconsistencies at window borders. In such cases the overlapping values are averaged to form $\hat{Z}(x)$ and $\hat{u}(x)$.

Sampling step: The location of new samples is suggested by analyzing the uncertainty results. New samples are extracted by selecting local maxima in uncertainty map $\hat{u}(x)$ with uncertainty values greater than a specific threshold th. An alternative, which is employed in our tests (Section 6), would be to select a given number of highest peaks among local maxima. Using the new samples, $Z(x)$ is updated and the algorithm goes back to the estimation step. It is also possible to go back to the modeling step and update the model using the information acquired from new samples. However, for now we only investigate the potential of the approach with a given TI as model, which is fixed.

6. Numerical experiments

As indicated in Section 1, traditional supervised sampling methods typically work based on Kriging (Brus and Heuvelink, 2007; Brus and De Gruijter, 1997; Vasat et al., 2010; Delmelle and Goovaerts, 2009; Aerts et al., 2013a,b). Aerts et al. (2013a,b) have suggested estimating the unknown filed values and uncertainties (Kriging-variance) using a variogram model. Then, new samples are extracted from highly uncertain locations in a sequential manner. The same approach is adopted in our paper except that we use a different model for estimation of unknown values and uncertainties which incorporates multiple-point statistics through a training image.

As a brute-force alternative for our method, given any set of observed samples MPS simulation methods can be employed to produce large number (e.g. $K = 100$) of realizations and compute its sample mean and variance in each point. Then, the variance can be considered as uncertainty and new samples can be drawn from

highly uncertain locations. In this section, our sampling method is tested versus random, Kriging-based and simulation-based samplings on different fields.

At the beginning, the $k \times k$ training patterns extracted from the TI are vectorized and used to train an MFA model (see Fig. 1). In principle, k should be selected based on the TI behavior so that the patterns include correlated values in the neighborhood. An entropy-based method for selecting the proper window size is proposed by Honarkhah and Caers (2010). This method suggests values close to 13 for all images used in this paper. Therefore k is set to 13 for all experiments. Then, another image with the same spatial behavior as the TI is considered as the real field $Z(x)$. At first, 1% of pixels of Z are sampled randomly and the field is reconstructed. Then, another 1% of Z is sampled and added to previous samples using four different strategies. The reconstruction is then performed by taking the new samples into account, and the process continues by going back to the sampling step. At each step, the reconstruction accuracy is evaluated by comparing the reconstructed field with the real field.

In the first sampling strategy, the new samples are selected randomly. Remaining strategies extract new samples from local maxima of an uncertainty map. The uncertainty map is formed using Kriging variance, simulation-based variance and MFA variance (the proposed method). Simulation-based variance is computed using Fast Direct Sampling (Abdollahifard and Faez, 2013a). Note that usually, the number of local maxima in the uncertainty map is larger than 1% of the image pixels. In such cases, the maxima with larger uncertainty are preferred.

In addition to the uncertainty map, the field values can also be estimated using methods like Kriging. However, to allow fair comparison of sampling methods the reconstruction is carried out using the MFA model for all sampling strategies.

Images of Fig. 8(a) and (b), which are used respectively as training image *TI* and real field *Z*, have been considered as simplifications of a channelized depositional systems (Strebelle, 2002). Fig. 8(c), (d), (e) and (f) show respectively the reconstruction results using 3% of samples extracted at random, based on Kriging variance, based on simulation variance and based on MFA variance (proposed method). It is clear that simulation-based and MFA-based samplings outperform random and Kriging-based samplings. Considering that simulation-based sampling requires 100 realizations at each step while the suggested method computes the uncertainty map directly, one can confirm the computational efficiency of our algorithm. Note that for real fields with very large sizes, the computational efficiency is a major concern.

The suggested samples at third sampling step using different sampling strategies are also depicted in Fig. 8(g), (h) and (i). Suggested samples in the simulation-based and MFA-based methods are mainly concentrated around image edges, whereas no specific structure can be detected in Kriging-based samples. As illustrated in Fig. 3 sharp changes of the field values at image edges cause uncertainties regarding exact location of the edge and necessitate further sampling. The suggested method extracts samples around borders and prevents redundant samples in textureless areas, therefore reducing the sampling cost.

The reconstruction accuracy for different sampling strategies is compared in Fig. 9 for sampling rates of 1–30%. Reconstruction error is computed using a normalized Manhattan distance. Note that Kriging-based sampling results in a very limited improvement in comparison to random sampling. Although Kriging-based sampling is recently employed for sampling design of smooth fields (e.g. electromagnetic field exposure) with moderate success (Aerts et al., 2013a,b), our experiments reveal that it does not provide good performance for highly structured and complex fields (e.g. geological fields).

Consider the horizontal line depicted in Fig. 9 which indicates a
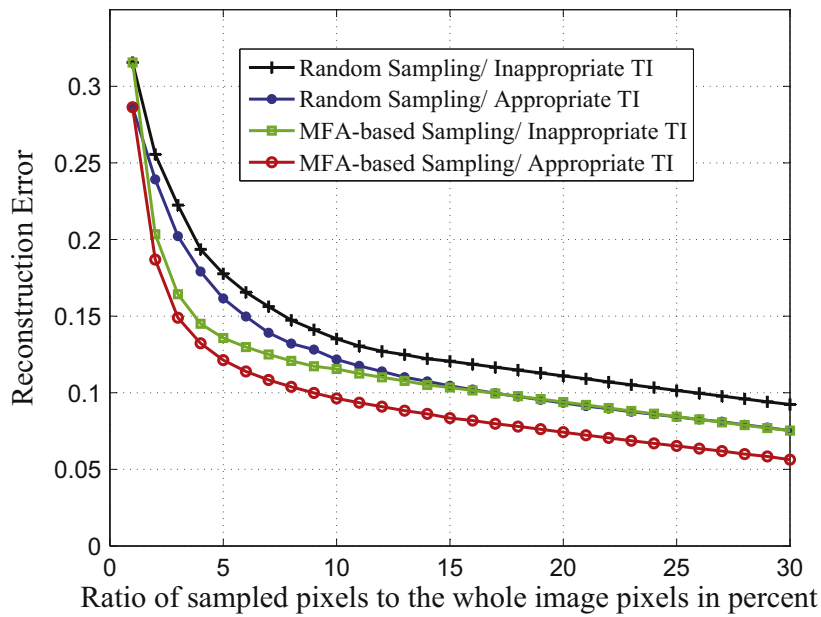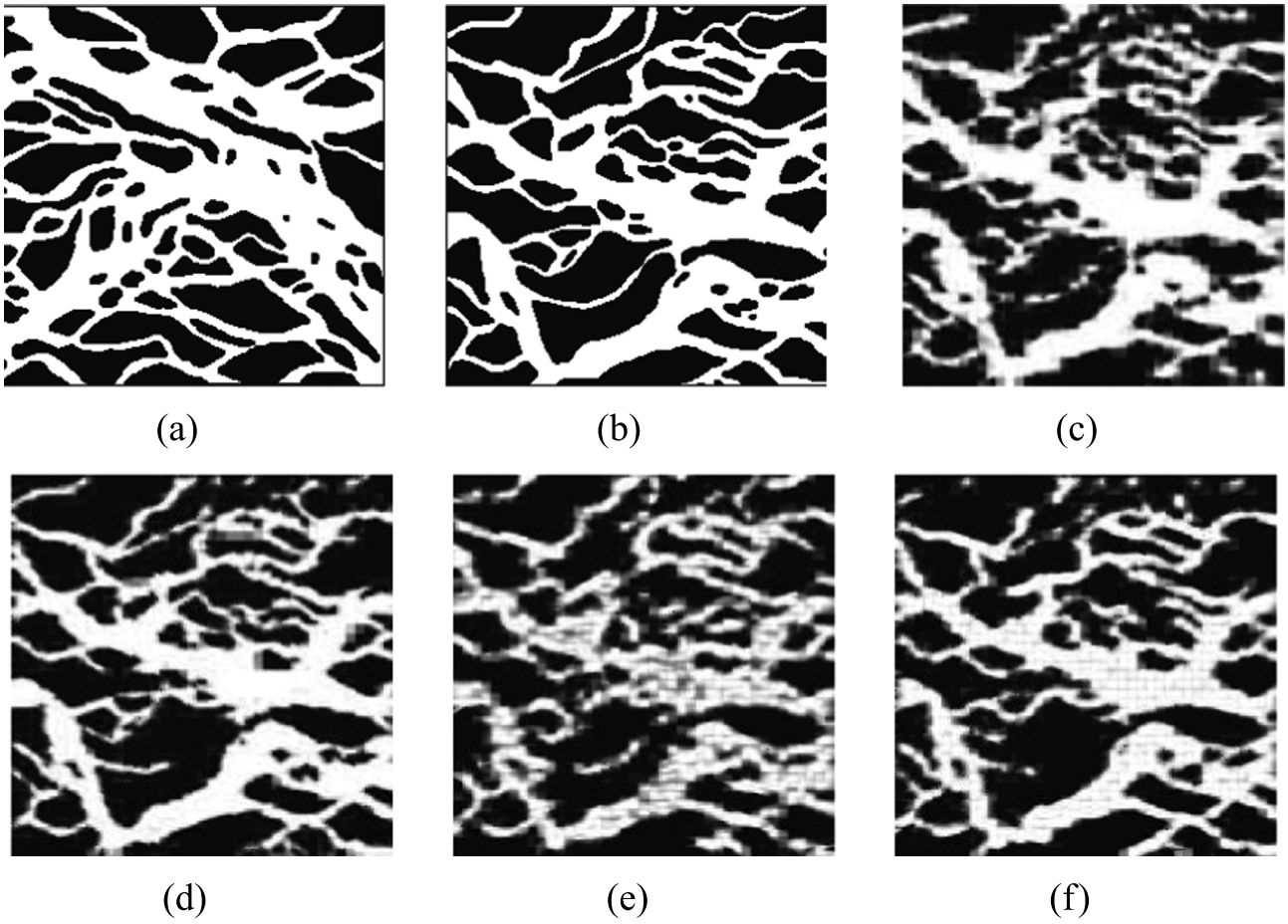
Fig. 12. Study of the effect of an inappropriate TI on the performance of our method. (a) An appropriate TI used to train the first MFA model, (b) real field, (c) and (d) reconstruction results for 5% random and MFA-based samples respectively, using the appropriate TI, (e) and (f) reconstruction results for 5% random and MFA-based samples respectively, using the inappropriate TI of Fig. 8a. (g) Reconstruction error for the four different strategies.

certain reconstruction error. This level of accuracy is achieved using 5% samples in our method while the same accuracy is achieved using 14% samples in random sampling. In other words,

our algorithm achieves the same accuracy using approximately 1/3 of samples. The results suggest that turning from two-point statistics to multiple-point statistics not only leads to improvements
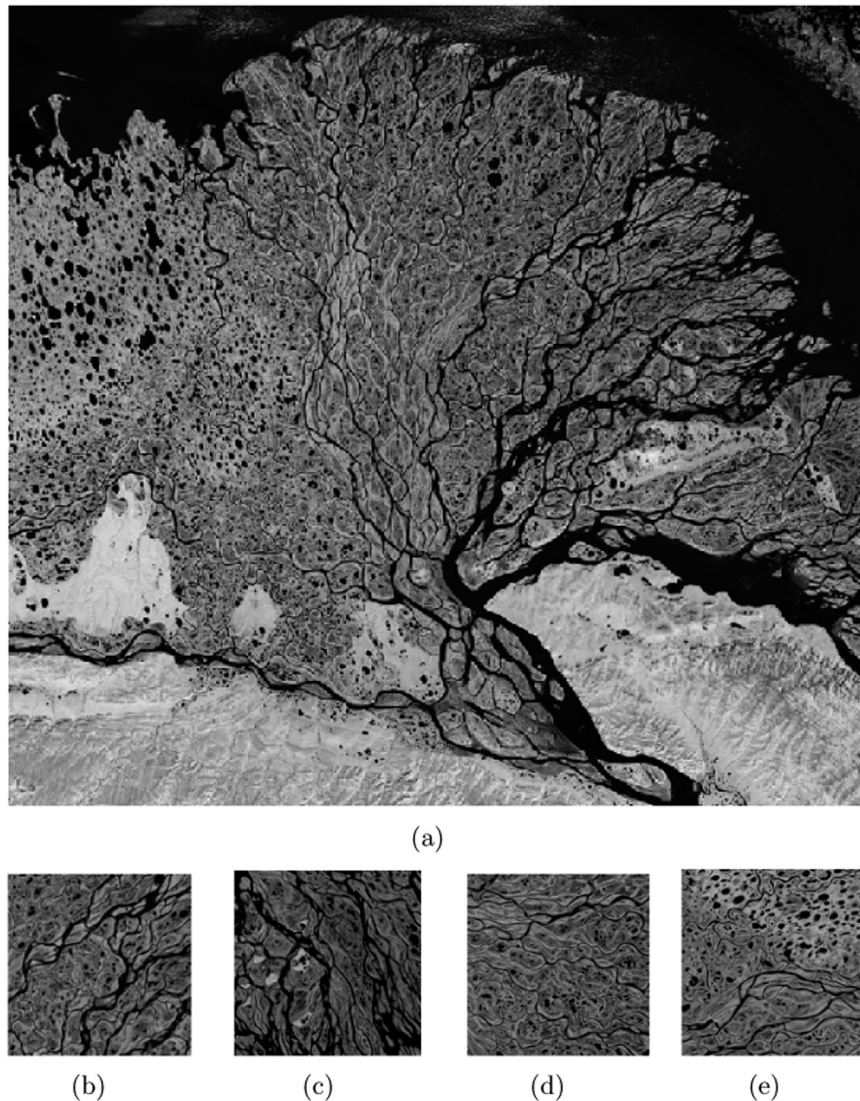
Fig. 13. (a) A satellite image of the Lena Delta in Russia (763 × 817), (b)–(e) Four TIs of size 150 × 150 describing the behavior of the delta in different regions.

in simulation methods (as widely studied by different researchers Strebelle, 2002; Arpat and Caers, 2007; Abdollahifard and Faez, 2013a,b), but also it can be efficiently employed in data acquisition design.

All experiments are carried out on a laptop computer with a 2.6 GHz processor. In these tests the required time for training the MFA model is 1323 s. Furthermore, the time required to produce an uncertainty map based on the MFA model is 127 s. Then, the total time required for 30 sampling stages is 5133 s (=1.43 h). On the other hand, the time required for producing one realization using FDS (Fast Direct Sampling) is 113 s. In 30 stages we have produced 3000 realizations requiring 339,000 s (=3.9 days). Therefore, the proposed algorithm is 66 times faster than the brute force sampling. Moreover, we believe that the MFA code used in this study has the potential to be optimized for faster implementation.

A similar experiment is carried out based on the continuous 250 × 250 TI of Fig. 10(a). Again, the first image on the top-left is considered as *TI*, and the second one is considered as real field *Z*. Training patterns of size 13 × 13 are extracted from the TI and used to train the MFA model. As depicted in Fig. 10(c), (d), (e), and (f), the reconstruction accuracy of the proposed method is again superior to random sampling and Kriging-based samplings and comparable to simulation-based sampling. Unlike the previous example, here the variable is continuous and therefore has also

textures far from objects borders. Therefore, the variations in the uncertainty image are not as sharp as previous experiment. Thus, more attention is needed to detect the structure of samples in simulation-based and MFA-based methods. Furthermore, the slope of reconstruction error plot, depicted in Fig. 11, is smaller in this example for low sampling rates, although the proposed sampling still performs far better than random and Kriging-based samplings. In this set of tests, the MFA method takes 1241 s for training and 3780 s for producing 30 uncertainty maps. The brute force method takes total of 335,100 s which is 66.7 times slower than the proposed method.

The above results assume that the chosen training image is representative of the real underlying field. Here we examine the robustness of our method in a case that the TI is not optimally chosen. The goal is to reconstruct the binary image of Fig. 12 (b) which is based on a satellite image of the Ganges delta (Bangladesh), with soil properties classified as channel (white) and alluvial bars (black). Two different TIs are used to train the MFA model. The first model is trained using the TI depicted in Fig. 12 (a) which is a suitable representative of the real field. Based on this model, Fig. 12c and d are reconstructed using 5% random and MFA-based samples, respectively.

The same exercise is carried out using Fig. 8(a) as TI, which is clearly suboptimal in this case. Using the model trained by this
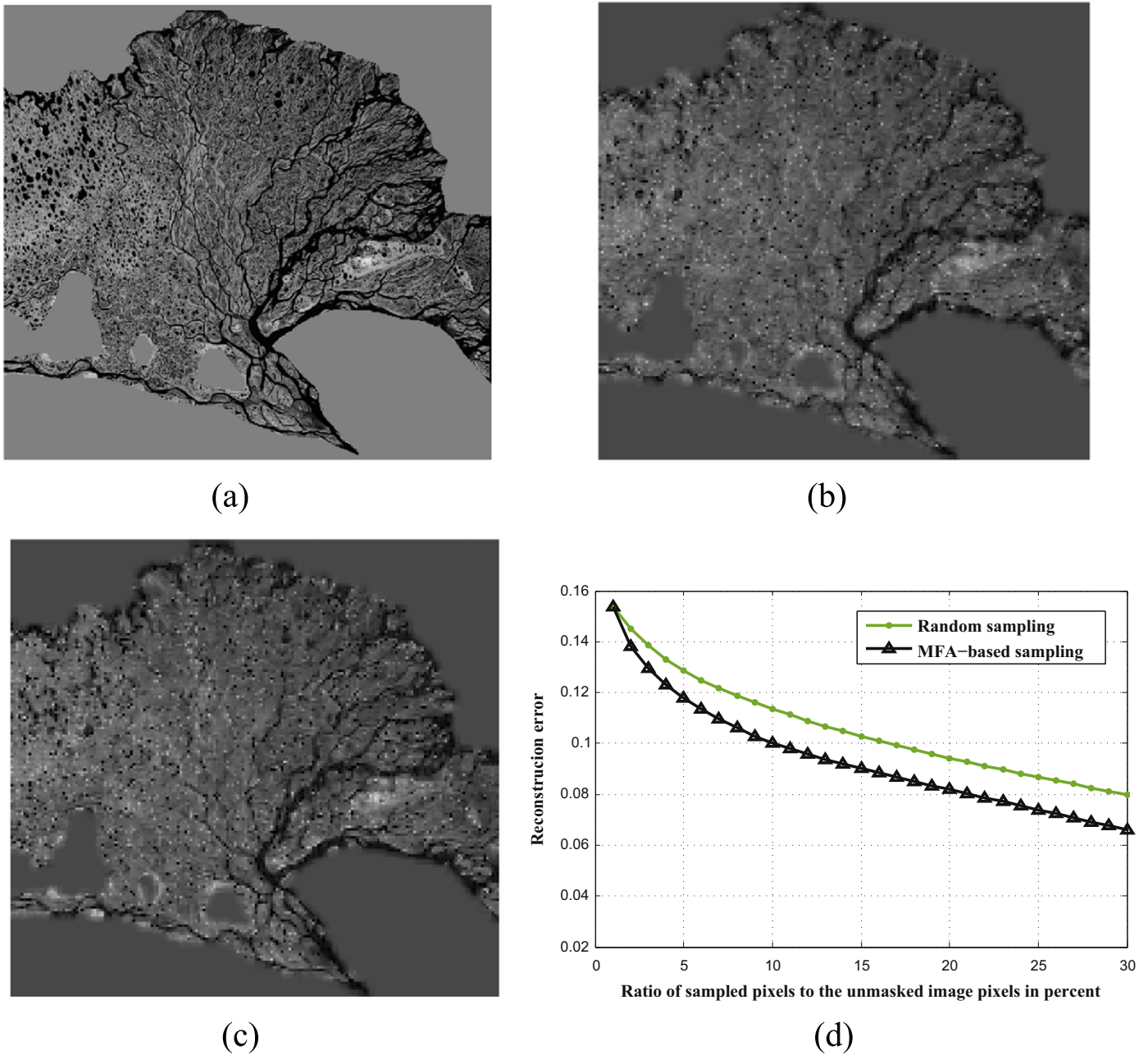
Fig. 14. (a) The original image is masked to include the delta only. No samples are allowed in the masked regions depicted in gray. (b) and (c) The reconstructed images using 7% random and MFA-based samples respectively. The coarse structures are better retrieved using MFA-based samples. (d) Reconstruction error for different sampling rates.

inappropriate TI, Fig. 12(e) and (f) are obtained using 5% random and MFA-based samples respectively. In all cases MFA-based samples outperform random samples. Furthermore, Fig. 12(f) is sharper than Fig. 12(c), especially near the borders, indicating that our method is robust even when the TI chosen is not completely adequate. In the body of broad channels of Fig. 12(e) and (f) a light checkerboards effect can be detected. The reason is the TI used to train the second model (Fig. 8(a)) does not contain such broad channels. Fig. 12(g) presents an overall comparison of aforementioned sampling and modeling strategies for different sampling rates. At low sampling rates, which are of high practical importance, MFA-based sampling based on the inappropriate model outperforms random sampling with appropriate model.

As assumed in the above experiments, in most MPS applications researchers use a single TI to describe a field (Strebelle, 2002;

Arpat and Caers, 2007; Abdollahifard and Faez, 2013a,b). The TI can be selected large enough to cover different field variations. However, the proposed method is not limited to use just a single TI and the patterns from multiple TIs could easily be included in the training data-base. Especially when working with very large unstationary fields, the model can be trained with multiple stationary images describing the behavior of different regions.

As an example, consider the Lena image (a satellite image of the Lena Delta in Russia), depicted in Fig. 13(a). The image is very structured containing some connected channels with different orientations, disconnected lakes and meandering patterns. Four training images depicted in Fig. 13 (b)–(e) are employed to model different variations of this field. A data-base composed of 13×13 patterns of these TIs is used to train the MFA model. Then, the model is used for reconstruction and supervised sampling of the real field of Fig. 13(a). Since our goal is to reconstruct the delta

patterns, samples are only extracted from the delta by masking other parts (see Fig. 14(a)).

Fig. 14 (b) and (c) show the reconstructed images using 7% random and MFA-based samples respectively. Due to the complex nature of the field, none of the reconstructed images represent the details of the real field appropriately. However, the coarse structures are better reconstructed using supervised samples.

The reconstruction error is compared in Fig. 14(d) for random and MFA-based sampling methods. In comparison with previous experiments, the reconstruction error curve decreases slowly by increasing the number of samples in this test. This could be justified by considering the presence of narrow, meandering and diverse patterns which make it difficult to predict the field values in a neighborhood by observing a few samples. As a result, observing a sample affects on uncertainty estimates in a very limited neighborhood and a large number of samples are required to achieve acceptable reconstruction accuracy. Nevertheless, the performance of our sampling method is still considerably superior to random sampling.

7. Conclusion

A sequential compressive sampling method is presented in this paper based on a MFA model. MFA models are capable of capturing nonlinear manifolds using a mixture of low-dimensional linear models. Using the trained model, the missing values are estimated along with their uncertainties, and then the sampling method extracts new samples from local maxima of the resulting uncertainty.

Using the patterns obtained from a training image, the behavior of the field is modeled. Then, based on the previously drawn samples and the modeled behavior, sampling locations are suggested. Compared to random and Kriging-based sampling approaches, this method extracts more informative samples and prevents redundant sampling. Samples are usually concentrated around objects borders where, intuitively, much uncertainty arises from the exact location of the borders. On the other hand, textureless areas are usually reconstructed accurately using a lesser amount of samples. The variable density of sampling leads to a high reconstruction accuracy with a given number of samples. Experiments reveal that the proposed method significantly reduces the number of samples necessary to achieve a given reconstruction accuracy.

Our method can be used to design data acquisition campaigns in geosciences, where measurements are expensive and often taken sequentially. The proposed sampling strategy allows obtaining a similar characterization than random sampling with only 30% to 50% of the data, which can lead to a significant reduction in data acquisition costs. In all examples shown the model is solely based on the training image. Future work will be devoted to refining the model at each sampling iteration using the newly acquired samples. Such model updates should bring significant improvement especially when the training image is not well representative of the behavior of the modeled domain.

References

Abdollahifard, M.J., 2016. Fast multiple-point simulation using a data-driven path and an efficient gradient-based search. Comput. Geosci. 86, 64–74.

Abdollahifard, M.J., Ahmadi, S., 2016. Reconstruction of binary geological images using analytical edge and object models. Comput. Geosci.

Abdollahifard, M.J., Faez, K., 2013a. Fast direct sampling for multiple-point stochastic simulation. Arab. J. Geosci., 1–13.

Abdollahifard, M.J., Faez, K., 2013b. Stochastic simulation of patterns using Bayesian pattern modeling. Computat. Geosci. 17 (1), 99–116.

Aerts, S., Deschrijver, D., Joseph, W., Verloock, L., Goeminne, F., Martens, L., Dhaene, T., 2013a. Exposure assessment of mobile phone base station radiation in an outdoor environment using sequential surrogate modeling. Bioelectromagnetics 34 (4), 300–311.

Aerts, S., Deschrijver, D., Verloock, L., Dhaene, T., Martens, L., Joseph, W., 2013b. Assessment of outdoor radiofrequency electromagnetic field exposure through hotspot localization using kriging-based sequential sampling. Environ. Res. 126, 184–191.

Aldroubi, A., Wang, H., Zarringhalam, K., 2011. Sequential adaptive compressed sampling via Huffman codes. Sampling Theory Signal Image Process. 10 (4).

Allard, D., Guillot, G., 2000. Clustering geostatistical data. In: Proceedings of the Sixth Geostatistical Conference.

Arias-Castro, E., Candes, E.J., Davenport, M.A., 2013. On the fundamental limits of adaptive sensing. IEEE Trans. Inf. Theory 59 (1), 472–481.

Arpat, G.B., Caers, J., 2007. Conditional simulation with patterns. Math. Geol. 39 (2), 177–203.

Ashok, A., Baheti, P.K., Neifeld, M.A., 2008. Compressive imaging system design using task-specific information. Appl. Opt. 47 (25), 4457–4471.

Ashok, A., Huang, L.-C., Neifeld, M.A., 2013. Information optimal compressive sensing: static measurement design. JOSA A 30 (5), 831–853.

Atkinson, P.M., Curran, P.J., 1995. Defining an optimal size of support for remote sensing investigations. IEEE Trans. Geosci. Remote Sens. 33 (3), 768–776.

Baheti, P.K., Neifeld, M.A., 2009. Recognition using information-optimal adaptive feature-specific imaging. JOSA A 26 (4), 1055–1070.

Bishop, C.M., Nasrabadi, N.M., 2006. Pattern Recognition and Machine Learning vol. 1. Springer, New York.

Bishop, C.M., Winn, J.M., 2000. Non-linear Bayesian image modelling. In: Computer Vision-ECCV 2000. Springer, Dublin, Ireland, pp. 3–17.

Boucher, A., Kyriakidis, P.C., Cronkite-Ratcliff, C., 2008. Geostatistical solutions for super-resolution land cover mapping. IEEE Trans. Geosci. Remote Sens. 46 (1), 272–283.

Brus, D., De Gruijter, J., 1997. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). Geoderma 80 (1), 1–44.

Brus, D.J., Heuvelink, G., 2007. Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138 (1), 86–95.

Candès, E.J., Wakin, M.B., 2008. An introduction to compressive sampling. IEEE Signal Process. Mag. 25 (2), 21–30.

Chen, M., Silva, J., Paisley, J., Wang, C., Dunson, D., Carin, L., 2010. Compressive sensing on manifolds using a nonparametric mixture of factor analyzers: algorithm and performance bounds. IEEE Trans. Signal Process. 58 (12), 6140–6155.

Delmelle, E.M., Goovaerts, P., 2009. Second-phase sampling designs for non-stationary spatial variables. Geoderma 153 (1), 205–216.

Diggle, P., Ribeiro P., 2007. Model-based geostatistics, Springer series in statistics. Springer, New York.

Donoho, D.L., 2006. Compressed sensing. IEEE Trans. Inf. Theory 52 (4), 1289–1306.

Duarte-Carvajalino, J.M., Sapiro, G., 2009. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. IEEE Trans. Image Process. 18 (7), 1395–1408.

Duarte-Carvajalino, J.M., Yu, G., Carin, L., Sapiro, G., 2013. Task-driven adaptive statistical compressive sensing of Gaussian mixture models. IEEE Trans. Signal Process. 61 (3), 585–600.

Elad, M., 2007. Optimized projections for compressed sensing. IEEE Trans. Signal Process. 55 (12), 5695–5702.

Elad, M., Aharon, M., 2006. Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. 15 (12), 3736–3745.

Engan, K., Aase, S.O., Hakon Husoy, J., 1999. Method of optimal directions for frame design. In: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings, vol. 5. IEEE, Phoenix, Arizona, USA, pp. 2443–2446.

Gangeh, M., Ghodsi, A., Kamel, M.S., 2013. Kernelized supervised dictionary learning. IEEE Trans. Signal Process. 61 (19), 4753–4767.

Griffin, A., Tsakalides, P., 2007. Compressed sensing of audio signals using multiple sensors. Reconstruction 3 (4), 5.

Honarkhah, M., Caers, J., 2010. Stochastic simulation of patterns using distance-based pattern modeling. Math. Geosci. 42 (5), 487–517.

Jha, S.K., Mariethoz, G., Evans, J.P., McCabe, M.F., 2013. Demonstration of a geostatistical approach to physically consistent downscaling of climate modeling simulations. Water Resourc. Res. 49 (1), 245–259.

Ji, S., Xue, Y., Carin, L., 2008. Bayesian compressive sensing. IEEE Trans. Signal Process. 56 (6), 2346–2356.

Ke, J., Ashok, A., Neifeld, M.A., 2010. Object reconstruction from adaptive compressive measurements in feature-specific imaging. Appl. Opt. 49 (34), H27–H39.

Kreutz-Delgado, K., Murray, J.F., Rao, B.D., Engan, K., Lee, T.-W., Sejnowski, T.J., 2003. Dictionary learning algorithms for sparse representation. Neural Comput. 15 (2), 349–396.

Mahalanobis, A., Muise, R., 2009. Object specific image reconstruction using a compressive sensing architecture for application in surveillance systems. IEEE Trans. Aerosp. Electron. Syst. 45 (3), 1167–1180.

Mariethoz, G., McCabe, M.F., Renard, P., 2012. Spatiotemporal reconstruction of gaps in multivariate fields using the direct sampling approach. Water Resourc. Res. 48, 10.

Nguyen, H., Patel, V., Nasrabadi, N., Chellappa, R., 2013. Design of non-linear kernel

dictionaries for object recognition. IEEE Trans. Image Process. 22 (12), 5123–5135.

Olshausen, B.A., Field, D.J., 1997. Sparse coding with an overcomplete basis set: a strategy employed by v1? Vis. Res. 37 (23), 3311–3325.

Peyré, G., 2009. Manifold models for signals and images. Comput. Vis. Image Underst. 113 (2), 249–260.

Paisley, J., Zhou, M., Sapiro, G., Carin, L., 2010. Nonparametric image interpolation and dictionary learning using spatially-dependent Dirichlet and beta process priors. In: 2010 17th IEEE International Conference on Image Processing (ICIP). IEEE, Hong Kong, pp. 1869–1872.

Romary, T., De Fouquet, C., Malherbe, L., 2011. Sampling design for air quality measurement surveys: an optimization approach. Atmosph. Environ. 45 (21), 3613–3620.

Romary, T., Malherbe, L., Fouquet, C., 2014. Optimal spatial design for air quality measurement surveys. Environmetrics 25 (1), 16–28.

Schlüter, S., Vogel, H.-J., 2011. On the reconstruction of structural and functional properties in random heterogeneous media. Adv. Water Resourc. 34 (2), 314–325.

Soni, A., Haupt, J., 2012. Learning sparse representations for adaptive compressive sensing. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, Kyoto, Japan, pp. 2097–2100.

Strebelle, S., 2002. Conditional simulation of complex geological structures using multiple-point statistics. Math. Geol. 34 (1), 1–21.

Stumpf, A., Lachiche, N., Kerle, N., Malet, J., Puissant, A., 2012. Adaptive spatial sampling with active random forest for object-oriented landslide mapping. In: Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International.

IEEE, Munich, Germany, pp. 87–90.

Tang, Y., Atkinson, P.M., Wardrop, N.A., Zhang, J., 2013. Multiple-point geostatistical simulation for post-processing a remotely sensed land cover classification. Spat. Stat. 5, 69–84.

Tuia, D., Ratle, F., Pacifici, F., Kanevski, M.F., Emery, W.J., 2009. Active learning methods for remote sensing image classification. IEEE Trans. Geosci. Remote Sens. 47 (7), 2218–2232.

Van Groenigen, J., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87 (3), 239–259.

Vasat, R., Heuvelink, G., Boruvka, L., 2010. Sampling design optimization for multivariate soil mapping. Geoderma 155 (3), 147–153.

Wang, J., Ge, Y., Heuvelink, G.B., Zhou, C., 2014. Spatial sampling design for estimating regional gpp with spatial heterogeneities. IEEE, Geosci. Remote Sens. Lett. 11 (2), 539–543.

Wojcik, R., McLaughlin, D., Konings, A.G., Entekhabi, D., 2009. Conditioning stochastic rainfall replicates on remote sensing data. IEEE Trans. Geosci. Remote Sens. 47 (8), 2436–2449.

Yu, G., Sapiro, G., 2011. Statistical compressed sensing of gaussian mixture models. IEEE Trans. Signal Process. 59 (12), 5842–5858.

Yu, G., Sapiro, G., Mallat, S., 2012. Solving inverse problems with piecewise linear estimators: from Gaussian mixture models to structured sparsity. IEEE Trans. Image Process. 21 (5), 2481–2499.

Zinn, B., Harvey, C.F., 2003. When good statistical models of aquifer heterogeneity go bad: a comparison of flow, dispersion, and mass transfer in connected and multivariate gaussian hydraulic conductivity fields. Water Resourc. Res. 39 (3), 1051.