Geoderma 163 (2011) 150–162

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Confronting uncertainty in model-based geostatistics using Markov Chain Monte Carlo simulation

Budiman Minasny [a,*], Jasper A. Vrugt [b,c,d], Alex B. McBratney [a]

[a] Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW 2006, Australia
[b] Department of Civil and Environmental Engineering, University of California, Irvine, 92697 CA, USA
[c] Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, The Netherlands
[d] Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, 87545 NM, USA

ARTICLE INFO

Article history:
Received 27 September 2010
Received in revised form 27 February 2011
Accepted 23 March 2011
Available online 25 May 2011

Keywords:
Generalized linear mixed model
Bayesian inference
Bayesian kriging
Regression-kriging
Best linear unbiased predictor
Pedometrics

ABSTRACT

This paper demonstrates for the first time the use of Markov Chain Monte Carlo (MCMC) simulation for parameter inference in model-based soil geostatistics. We implemented the recently developed DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm to jointly summarize the posterior distribution of variogram parameters and the coefficients of a linear spatial model, and derive estimates of predictive uncertainty. The DREAM method runs multiple different Markov chains in parallel and jumps in each chain are generated from a discrete proposal distribution containing a fixed multiple of the difference of the states of randomly chosen pairs of other chains. This approach automatically scales the orientation and scale of the proposal distribution, and is especially designed to maintain detailed balance and ergodicity, thereby generating an exact approximation of the posterior probability density function (pdf) of the parameters of the linear model and variogram. This approach is tested using three different data sets from Australia involving variogram estimation of soil thickness, kriging of soil pH, and spatial prediction of soil organic carbon content. The results showed some advantages of MCMC over the conventional method of moments and residual maximum likelihood (REML) estimation. The posterior pdf derived with MCMC conveys important information about parameter uncertainty, multi-dimensional parameter correlation, and thus how many significant parameters are warranted by the calibration data. Parameter uncertainty constitutes only a small part of total prediction uncertainty for the case studies considered here. The prediction accuracies using MCMC and REML are similar. The variogram estimated using conventional approaches (method of moments, and without simulation) lies within the 95% prediction uncertainty interval of the posterior distribution derived with DREAM. Altogether our results show that conventional kriging and regression-kriging still remain a viable option for production mapping.

1. Introduction

The past decades have seen a growing interest in soil characterization, not only because many highly interrelated water, carbon, and vegetation processes take place in this part of our ecosystem, but also because the subsurface plays a major role in carbon storage, exchange of water and soil vapor with the ambient atmosphere, and contaminant transport to the groundwater. Many soil properties not only exhibit significant spatial and temporal variability, but are also difficult to measure directly. Geostatistics is primarily concerned with the mathematical description, prediction and uncertainty analysis of these spatially varying variables.

The basis of geostatistical estimation is the assumption that an observation of a variable z at location **s** is a realization of a random field $Z(\mathbf{s})$. The random function is assumed to be intrinsically stationary (Webster and Oliver, 2007):

$$E[Z(\mathbf{s})-Z(\mathbf{s}+\mathbf{h})]=0,$$

and

$$2\gamma(\mathbf{h})=E\left[|Z(\mathbf{s})-Z(\mathbf{s}+\mathbf{h})|^2\right] \quad (1)$$

where **h** is a lag vector, and $\gamma(\mathbf{h})$ is the variogram. The method of moments is usually used to estimate the so-called empirical variogram, and parameters of a variogram model are then estimated by fitting a model to the empirical variogram. Thus, the accuracy of spatial prediction is determined by how the variogram is estimated.

Methods that are commonly used in pedometrics: ordinary kriging, universal kriging (Matheron, 1969), kriging with external drift (Hudson and Wackernagel, 1994), and regression-kriging (Odeh et al., 1995), are increasingly being criticized, particularly for the *ad-hoc* approach in

* Corresponding author. Tel.: +61 2 90369043; fax: +61 2 93513706.
E-mail address: budiman.minasny@sydney.edu.au (B. Minasny).

0016-7061/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2011.03.011

which variogram parameters are estimated. The prior assumptions about the statistical properties of the error residuals are often violated, questioning the validity of the optimized variogram parameters and model predictions. These approaches are therefore nowadays referred to as classical methods by Diggle et al. (1998).

The paper of Diggle et al. (1998) introduced an alternative approach to geostatistical inference that is based on the application of formal statistical methods. This so-called model-based geostatistics relies on general principles of statistical inference and use an explicit stochastic error model of the data. The use of a proper likelihood function and *a posteriori* check of the error residual distribution not only constitutes good statistical practice, but also facilitates the application of a powerful array of multivariate statistical analyses to explore the marginal and joint posterior probability density functions of the nugget, sill and range. Other formal statistical approaches include the generalized linear-mixed model and empirical best linear unbiased prediction (EBLUP) (Zimmerman and Cressie, 1992).

Despite this progress made, it remains typically difficult to estimate the parameters of a variogram model and successfully solve the geostatistical inverse problem. That is, the covariance structure of the residuals is required for variogram fitting, yet in most cases is unknown *a priori*. The spatial covariance structure can only be obtained after fitting the trend. The residual maximum likelihood (REML) approach is therefore advocated (Lark et al., 2006), because it allows for independent estimation of the covariance structure by transforming the data into stationary increments that filtered out the trend. Nevertheless, REML or related methods only attempt to find the maximum likelihood values of the variogram parameters without recourse to estimating their underlying posterior uncertainty. This posterior probability distribution function (pdf) contains important information about the uncertainty and correlation of the various geostatistical parameters and can be used to derive meaningful estimates of spatial prediction uncertainty.

In recent years, much progress has been made toward the application of the REML approach for variogram fitting, and increasingly sophisticated inversion methods are being used for statistical inference of the geostatistical parameters. For example, Minasny and McBratney (2007) used the profile log-likelihood method, while Marchant and Lark (2007) used simulated annealing. Unfortunately, these methods only estimate the maximum likelihood parameter values without appropriately summarizing their underlying uncertainty. In the face of measurement error, incomplete process knowledge, and theoretical (model structural) inadequacies, it would be naive to rely on a single realization of the variogram and/or other geostatistical model parameters to quantify or predict the variables or processes observed in the field. Without a realistic assessment of parameter uncertainty, it is not possible to undertake with any confidence tasks such as quantifying spatial prediction error, assessing the information content of the measurement data, and inference of parameter correlation. We therefore argue that, it would be more productive to work with an ensemble of parameter values that are each within a desired interval of the maximum likelihood parameters given the calibration data at hand. An easy way to quantify parameter uncertainty is to use a traditional first-order statistical approach at the maximum likelihood variogram parameter estimates. This classical method assumes that the posterior pdf of the variogram parameters can be adequately approximated with a multivariate-normal (Gaussian) distribution. The variance of the prediction error, and the Fisher information (sensitivity) matrix contain the desired information about the scale and orientation of the parameter uncertainty (Marchant and Lark, 2004; Minasny and McBratney, 2005). However, many published contributions have shown that such linear confidence intervals typically underestimate the actual parameter uncertainty, particularly in the presence of high-dimensionality and significant model nonlinearity (*e.g.* Vrugt and Bouten, 2002).

A more general and often more reliable way to estimate parameter uncertainty is by implementing a Bayesian approach (Diggle et al., 1998) using Markov Chain Monte Carlo (MCMC) simulation. Unlike traditional first-order theory, this approach can readily cope with model nonlinearity and non-Gaussianity of the parameter uncertainty. Handcock and Stein (1993) were among the first to use Bayesian analysis to consider uncertainty in the covariance function and its effect on prediction quality. The outcome of such analysis is a posterior parameter distribution that is centered on the maximum likelihood values, and describes the underlying parameter uncertainty. Christensen (2004) demonstrated the use of MCMC for parameter estimation in model-based geostatistics, whereas Zhang (2002) used a MCMC approach to infer parameter distributions for spatial prediction of incidence rates of root disease. More recently, Aelion et al. (2009) used a Bayesian kriging approach for estimating spatial distributions of heavy metals in soil.

In this paper, we explored the usefulness and applicability of MCMC simulation for parameter inference in model-based geostatistics. We used the recently developed DiffeRential Evolution Adaptive Metropolis (DREAM) MCMC algorithm, which uses differential evolution as a genetic algorithm for parameter population evolution with a Metropolis selection rule. We tested this approach on three different data sets from Australia involving estimation of soil thickness variogram, spatial prediction of soil pH, and soil organic carbon content. We then compared the results of the MCMC analysis against the outcome of the standard method-of-moments and REML approach.

2. Methods

2.1. The linear spatial model

The model for prediction of variable $\mathbf{z}$ at location $\mathbf{x}$ can be written as:

$$\mathbf{z}(\mathbf{x}) = \mathbf{m}(\mathbf{x})^T\boldsymbol{\beta} + \boldsymbol{\varepsilon}(\mathbf{x}) \tag{2}$$

where $\mathbf{x}$ is the vector of spatial coordinates, $\mathbf{m}(\mathbf{x})$ is the design matrix of the trend function $[m_1, m_2, \ldots, m_p]$, $\boldsymbol{\beta}$ is the parameter vector (size $p \times 1$) for the trend, and $\boldsymbol{\varepsilon}$ is the vector of residuals with a mean of zero and covariance structure $\mathbf{K}$. The covariance function can be modelled using the Matérn function:

$$K_{ij} = c_0\delta_{ij} + c_1\left[\frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{r}\right)^{\nu}K_{\nu}\left(\frac{h}{r}\right)\right] \tag{3}$$

where K_{ij} is the covariance between observation i and j, h represents the separation distance between i and j, δ_{ij} denotes the Kronecker delta ($\delta_{ij}=1$ if $i=j$ and $\delta_{ij}=0$ when $i\neq j$), c_0 is the nugget variance, and c_0+c_1 signifies the sill variance, K_ν is a modified Bessel function of the second kind of order ν, Γ is the gamma function, r denotes the distance or 'range' parameter and ν is the 'smoothness'. This latter parameter allows for great flexibility in modelling the local spatial covariance. The corresponding semivariance of the Matérn function is:

$$\gamma(h) = c_0 + c_1\left(1-\frac{1}{2^{\nu-1}\Gamma(\nu)}\left(\frac{h}{r}\right)^{\nu}K_{\nu}\left(\frac{h}{r}\right)\right). \tag{4}$$

The parameters of this function can be summarized in a single vector, $\boldsymbol{\phi}=[c_0, c_1, r, \text{and } \nu]$.

The value of z at the unsampled location $\mathbf{x}_0$ can be computed from:

$$z(\mathbf{x}_0) = \mathbf{m}(\mathbf{x}_0)^T\hat{\boldsymbol{\beta}} + \mathbf{k}^T\mathbf{K}^{-1}\left(z-\mathbf{M}\hat{\beta}\right), \tag{5}$$

where $z=[z(x_1),\ldots,z(x_n)]^T$, $\mathbf{M}$ is the $(n\times p)$ design matrix $\mathbf{M}=[\mathbf{m}(x_1)\ldots\mathbf{m}(x_n)]^T$, $\mathbf{K}=\text{cov}(\mathbf{z},\mathbf{z}^T)$, and $\mathbf{k}=\text{cov}\{\mathbf{z},z(\mathbf{x}_0)\}$. The associated variance of the prediction at location z is:

$$\sigma_z^2 = k_0-\mathbf{k}^T\mathbf{K}^{-1}\mathbf{k} + \mathbf{d}^T\left(\mathbf{M}^T\mathbf{K}^{-1}\mathbf{M}\right)^{-1}\mathbf{d} \tag{6}$$

where $k_0 = \mathbf{K}(x_0, x_0)$ and $\mathbf{d} = \mathbf{m}(\mathbf{x}_0) - \mathbf{M}^T\mathbf{K}^{-1}\mathbf{k}$. The unknown parameters in Eqs. (4)–(5) are symbolized as $\boldsymbol{\theta} = [\boldsymbol{\beta}, \boldsymbol{\phi}]$.

The classical statistical or frequentist approach considers the model parameters $\boldsymbol{\theta}$ in Eqs. (4)–(5) to be fixed but unknown. Optimization methods can then be used to find the best parameter values that provide the closest fit to a set of measurements $\mathbf{z}$. Usually this involves minimization of some least squares objective function. Relatively little attention has been given to a realistic assessment of parameter uncertainty. As argued before, it would seem more productive to work with a distribution of parameter values rather than a single realization of parameter values.

2.2. Bayesian statistics and Markov Chain Monte Carlo simulation

While classical statistics considers the model parameters to be fixed and unknown, Bayesian statistics treat them as probabilistic variables with a joint posterior pdf which captures the probabilistic beliefs about $\boldsymbol{\theta}$ in light of the observed data. The prior pdf, $p(\boldsymbol{\theta})$, summarizes information about $\boldsymbol{\theta}$ before any data are collected. This usually consists of some lower and upper bounds of each of the parameters. When data become available, the parameter distributions are updated using Bayes theorem. This generally reduces the uncertainty of the different parameters. Most early approaches consider the model parameters to be the main source of uncertainty, yet recent contributions also simultaneously treat other sources of uncertainty including model, forcing data, and calibration data error (Kavetski et al., 2006; Vrugt et al., 2008).

The Bayesian approach solves many of the problem in standard statistics (model averaging and model selection), and provides a natural way of incorporating external information such as soft data outside the current data set (Gelman et al., 2004). In model-based geostatistics, the Bayesian approach is suited to address and quantify the different sources of uncertainty within a single consistent, integrated, and hierarchical framework.

To successfully implement the Bayesian paradigm, efficient sampling methods are required to summarize the posterior pdf. This distribution combines the data likelihood with a prior distribution, and contains all the desired information to make statistically sound inferences about the uncertainty of the individual components in the model. One of the advantages of the Bayesian approach over the conventional (frequentist) approach is that uncertainty in variance components is more easily taken into account using the posterior pdf (Diggle et al., 1998). Unfortunately, for most practical geostatistical problems this posterior distribution cannot be obtained through analytical means or by analytical approximation. We therefore resort to iterative approximation methods such as MCMC simulation to generate a sample from the posterior pdf.

To the best of our knowledge, Moyeed and Papritz (2002) are among the first to use MCMC simulation in model-based geostatistics for estimating topsoil concentrations of cobalt and copper. Their initial results showed that MCMC sampling with RWM had no immediate advantages over other classical kriging methods, but required significantly more computational resources, and manual interaction to appropriately sample the posterior distribution. Their results further demonstrated difficulties with diagnosing convergence of the RWM sampler. Meanwhile, Christensen (2004) used a more complicated Langevin–Metropolis approach. This approach not only requires the gradient of the likelihood function to be evaluated, but also needs manual trial-and-error adjustment of the scale of the proposal distribution (Ribeiro et al., 2003). This is rather inefficient and time consuming. Problems with the choice of the proposal distribution, the esoteric nature of the probability manipulations, computer automation, and premature convergence have limited the widespread application and use of MCMC simulation in model-based geostatistics.

Similar problems in another field of study (hydrology), have stimulated Vrugt and co-workers to develop adaptive MCMC methods that tune themselves en route to the posterior target distribution using information exchange between different trajectories running in parallel. Examples of this include the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm (Vrugt et al., 2003), and the DREAM algorithm (Vrugt et al., 2008, 2009). The DREAM algorithm is an adaptation of the SCEM-UA algorithm but maintains detailed balance and ergodicity while showing excellent efficiencies on complex, high-dimensional and multimodal posterior distributions (Vrugt et al., 2009). The DREAM algorithm (Vrugt et al., 2008, 2009) automatically tunes the scale and orientation of the proposal distribution during the search. No human intervention is required during the run. The only information that needs to be specified by the user is the prior parameter distribution (usually a uniform distribution with some parameter ranges), and a likelihood function that measures the closeness between the model predictions and the observational data. Based on this, DREAM returns the posterior pdf of the sampled parameters. Note that DREAM uses different statistical tests to diagnose whether convergence has been achieved. These tests include the Gelman and Rubin scale reduction factor, $\hat{R}$ (Gelman and Rubin, 1992), and the Raftery–Lewis burn-in length (Raftery and Lewis, 1992).

We now provide a short introduction to MCMC sampling, followed by a condensed description of the DREAM algorithm. This MCMC method is especially designed to be user-friendly, and automatically adapts the scale and orientation of the proposal distribution en route to the posterior distribution. The development presented below closely follows Vrugt et al. (2008, 2009) and details can be found there.

The basis of the MCMC method is a Markov chain that generates a random walk through the parameter space and successively visits solutions with stable frequency stemming from a fixed probability distribution. To visit configurations with a stable frequency, an MCMC algorithm generates trial moves from the current ("*old*") position of the Markov chain $\boldsymbol{\theta}_{t-1}$ to a new state $\boldsymbol{\theta}_t$. The earliest and most general MCMC approach is the random walk Metropolis (RWM) algorithm. Assuming that a random walk has already sampled points $\{\boldsymbol{\theta}_0,\ldots,\boldsymbol{\theta}_{t-1}\}$, this algorithm proceeds in the following three steps. First, a candidate point $\boldsymbol{\vartheta}$ is sampled from a proposal distribution q that is symmetric, $q(\boldsymbol{\theta}_{t-1}, \boldsymbol{\vartheta}) = q(\boldsymbol{\vartheta}, \boldsymbol{\theta}_{t-1})$ and may depend on the present location, $\boldsymbol{\theta}_{t-1}$. Next, the candidate point is either accepted or rejected using the Metropolis acceptance probability:

$$\alpha = \begin{cases} \min\left[\dfrac{\pi(\boldsymbol{\vartheta})q(\boldsymbol{\theta}_{t-1},\boldsymbol{\vartheta})}{\pi\left(\boldsymbol{\theta}_{t-1}\right)q\left(\boldsymbol{\vartheta},\boldsymbol{\theta}_{t-1}\right)},1\right] & \text{if } \pi\left(\boldsymbol{\theta}_{t-1}\right) > 0 \\ 1 & \text{if } \pi\left(\boldsymbol{\theta}_{t-1}\right) = 0 \end{cases} \tag{7}$$

where $\pi(\cdot)$ denotes the density of the target distribution. Finally, if the proposal is accepted, the chain moves to $\boldsymbol{\vartheta}$ otherwise the chain remains at its current location $\boldsymbol{\theta}_{t-1}$.

The original RWM scheme is constructed to maintain detailed balance with respect to $\pi(\cdot)$ at each step in the chain:

$$\pi(\boldsymbol{\theta}_{t-1})p(\boldsymbol{\theta}_{t-1}\rightarrow\boldsymbol{\vartheta}) = \pi(\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}\rightarrow\boldsymbol{\theta}_{t-1}) \tag{8}$$

where $\pi(\boldsymbol{\theta}_{t-1})$ or $\pi(\boldsymbol{\vartheta})$ denotes the probability of finding the system in the state $\boldsymbol{\theta}_{t-1}$ or $\boldsymbol{\vartheta}$, and $p(\boldsymbol{\theta}_{t-1}\rightarrow\boldsymbol{\vartheta})$ ($p(\boldsymbol{\vartheta}\rightarrow\boldsymbol{\theta}_{t-1})$) denotes the conditional probability to perform a trial move from $\boldsymbol{\theta}_{t-1}$ to $\boldsymbol{\vartheta}$ (or $\boldsymbol{\vartheta}$ to $\boldsymbol{\theta}_{t-1}$). The result is a Markov chain which under some regularity conditions has a unique stationary distribution with pdf $\pi(\cdot)$. Hastings (1970) extended Eq. (7) to include non-symmetrical proposal distributions, *i.e.* $q(\boldsymbol{\theta}_{t-1},\boldsymbol{\vartheta}) \neq q(\boldsymbol{\vartheta}, \boldsymbol{\theta}_{t-1})$ in which a proposal jump to ϑ and the reverse jump do not have equal probability. This extension is called the Metropolis Hastings algorithm (MH), and has

become the basic building block of many existing MCMC sampling schemes.

The simplicity of the original MH algorithm and theoretically sound statistical basis of the method has lead to widespread implementation and use. However, the MH algorithm requires tuning of some internal variables before the MCMC simulator works effectively. The efficiency of the method is affected by the scale and orientation of the proposal distribution, $q(\boldsymbol{\theta}_{t-1},\cdot)$ used to generate trial moves (transitions) in the Markov Chain. When the proposal distribution is too wide, too many candidate points are rejected, and therefore the chain will not mix efficiently and converge only slowly to the target distribution. On the other hand, when the proposal distribution is too narrow, nearly all candidate points are accepted, but the distance moved is so small that it will take a prohibitively large number of updates before the sampler has converged to the target distribution. The choice of the proposal distribution is therefore crucial and determines the practical applicability of MCMC simulation in many fields of study.

Automatic tuning of the proposal distribution is a powerful remedy to overcome the above difficulties. This approach uses the information from the sampling history to continuously adapt the shape and size of the proposal distribution and evolve the sampler efficiently towards a limiting distribution. Examples of such methods are the Adaptive Proposal (AP) and Adaptive Metropolis (AM) schemes presented in (Haario et al., 1999; Haario et al., 2001), respectively. These algorithms utilize a single Markov chain and continuously adapt the covariance, $\mathbf{C}_t$ of a Gaussian proposal distribution, $q_t(\boldsymbol{\theta}_{t-1},\cdot)=N_d(\boldsymbol{\theta}_{t-1},\boldsymbol{\Sigma}_t)$ using the information contained in the sample path of the chain, $\boldsymbol{\Sigma}_t=s_d\mathrm{Cov}(\boldsymbol{\theta}_0,\ldots,\boldsymbol{\theta}_{t-1})+s_d\varepsilon\mathbf{I}_d$. Here, s_d represents a scaling factor that depends only on the dimensionality d of the problem, $\mathbf{I}_d$ signifies the d-dimensional identity matrix, and ε is a constant with a small value compared to that of the target so that the whole parameter space can be reached. As a basic choice, the scaling factor is chosen to be $s_d=2.4^2/d$ which is optimal for Gaussian target and proposal distributions (Gelman and Rubin, 1992).

Vrugt et al. (2008) and Vrugt et al. (2009) have recently presented a novel adaptive MCMC algorithm to efficiently estimate the posterior parameter pdf. This method, entitled DiffeRential Evolution Adaptive Metropolis (DREAM), runs multiple chains simultaneously for global exploration of the parameter space, and automatically tunes the scale and orientation of the proposal distribution during the search. This scheme is an adaptation of the DE-MC (Differential Evolution Markov Chain, Ter Braak, 2006; Ter Braak and Vrugt, 2008) and SCEM-UA (Shuffled Complex Evolutionary Metropolis) algorithms (Vrugt et al., 2003) and produces a Markov Chain that is ergodic with a unique stationary distribution, the proof of which is given in Vrugt et al. (2009). Various case studies have demonstrated that DREAM is generally superior to other adaptive MCMC sampling approaches, particularly when dealing with complex, multi-modal search problems (Huisman et al., 2010; Vrugt et al., 2008). In summary, DREAM runs N different Markov Chains simultaneously in parallel. Let the states of the N chains be denoted by the d-dimensional parameter vectors $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_N$ where d is dimensionality of the posterior distribution. At each time, the locations of the individual chains form a population, conveniently stored as an $N\times d$ matrix $\boldsymbol{\Theta}_t$. Jumps in each chain, $i=1,\ldots,N$ are created using a discrete proposal distribution:

$$\boldsymbol{\vartheta}^i=\boldsymbol{\theta}^i_{t-1}+(1_d+\mathbf{e}_d)\eta\left(\tau,d_{eff}\right)\left[\sum_{j=1}^{\tau}\boldsymbol{\theta}^{r_1(j)}_{t-1}-\sum_{n=1}^{\tau}\boldsymbol{\theta}^{r_2(n)}_{t-1}\right]+\boldsymbol{\omega}_d \qquad (9)$$

where τ is the number of pairs used to generate the proposal, and $r_1(j)$, $r_2(n)\in\{1,\ldots,N\}$; $r_1(j)\neq r_2(n)\neq i$ for $j=1,\ldots,\tau$ and $n=1,\ldots,\tau$. Symbol $\mathbf{1}_d$ represents a d-dimensional vector of unity, η symbolizes the jump size. The values of $\mathbf{e}_d$ are drawn from a uniform distribution $U_d(-b,b)$ with $|b|<1$, and $\boldsymbol{\omega}_d$ is generated from a normal distribution $N_d(0,b^*)$, with b^* small compared to the width of the target distribution. The jump-size η depends on τ and d_{eff}, the number of dimensions that will be updated jointly (see next paragraph). The Metropolis ratio, $\alpha(\boldsymbol{\vartheta}^i,\boldsymbol{\theta}^i_{t-1})$ is used to decide whether to accept the candidate points or not. If the proposal point is accepted the chain moves to this new point, $\boldsymbol{\theta}^i_t=\boldsymbol{\vartheta}^i$, otherwise the chain remains at its current location, $\boldsymbol{\theta}^i_t=\boldsymbol{\theta}^i_{t-1}$. Because the joint pdf of the N chains factorizes to $\pi(\boldsymbol{\theta}_1)\times\ldots\times\pi(\boldsymbol{\theta}_N)$, the states $\boldsymbol{\theta}_1,\ldots\boldsymbol{\theta}_N$ of the individual chains are independent at any generation after DREAM has become independent of its initial value. After this so-called burn-in period, the convergence of a DREAM run can thus be monitored with the $\hat{R}$-statistic of Gelman and Rubin (1992). The $\hat{R}$-statistic is an estimate of the factor by which the current distribution might be reduced if the simulations were continued in the limit towards ∞. The $\hat{R}$-statistic compares the within-chain and between-chain variances of each of the d parameters, and convergence is diagnosed when $\hat{R}_j\leq 1.2$; for $j=1,\ldots,d$.

Our experience with complicated and high-dimensional target distributions suggests that convergence of MCMC sampling schemes can be enhanced if fewer than d dimensions are updated in each proposal point at a time. In DREAM we implement this idea as follows. After the $\boldsymbol{\vartheta}^i$, $i=1,\ldots,N$ are created for each of the N chains using Eq. (9), each element, $j=1,\ldots,d$ of the proposal ϑ^i_j is replaced with θ^i_j using a binomial scheme with crossover probability CR,

$$\vartheta^i_j=\begin{cases}\theta^i_j & \text{if } U\leq 1-CR, \quad d_{eff}=d_{eff}-1\\ \vartheta^i_j & \text{otherwise}\end{cases} \qquad j=1,\ldots,d \qquad (10)$$

where $U\in[0,1]$ is a draw from a uniform distribution. In practice, we use an algorithmic parameter n_{CR} that defines how many different CR values we are using, $\{m/n_{CR}\,|\,m=1,\ldots,n_{CR}\}$ during offspring creation. A detailed algorithmic description of DREAM appears in Vrugt et al. (2008) and Vrugt et al. (2009) and so will not be repeated here. These two papers also demonstrate the superiority of DREAM over other adaptive MCMC samplers.

The DREAM algorithm has a few algorithmic variables that need to be specified by the user before the code can be used for posterior inference. Based on recommendations made in Vrugt et al. (2009), we fix $\delta=\{1,2,3\}$, set $n_{CR}=3$, and use $N=2d$. By comparison with RWM, a good choice for $\eta=2.38/\sqrt{2\tau d_{eff}}$. This choice is expected to yield an acceptance probability of 0.44 for $d=1$, 0.28 for $d=5$ and 0.23 for large d. Every 5th generation, $\eta=1.0$ to allow direct jumps between disconnected posterior modes (Ter Braak, 2006). This concludes the settings of the DREAM algorithm.

What we are now left with is a discussion of the details of the respective case study at hand. To successfully implement DREAM we need to define the parameter ranges, an initial sampling distribution, and a likelihood function to compare model predictions with observational data. In each case study, we list the upper and lower bounds of the parameters, use Latin Hypercube sampling over this d-dimensional hypercube to initialize the starting points of the N different Markov chains, and use the following log-likelihood function (Mardia and Marshall, 1984) to estimate the parameters:

$$log(\pi)=L(\boldsymbol{\theta}\,|\,\mathbf{z})=\text{constant}-\frac{1}{2}\,log|\mathbf{K}|-\frac{1}{2}(\mathbf{z}-\mathbf{M}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{K}^{-1}(\mathbf{z}-\mathbf{M}\boldsymbol{\beta}) \qquad (11)$$

where $\boldsymbol{\theta}=[\boldsymbol{\beta}\;|\;\boldsymbol{\phi}]$ is the vector of parameters that needs to be estimated for the linear spatial model; $\boldsymbol{\beta}$ defines the linear trend, and $\boldsymbol{\phi}=[c_0, c_1, r, \text{ and } \nu]$ signifies the covariance function. Note that the constant simply (linearly) scales the log-likelihood function and does not affect the actual posterior distribution. We therefore ignore the constant in our runs with DREAM.

In each case study we assume a uniform prior distribution of the parameters with ranges listed in the corresponding tables. Thus prior to fitting the observational data, each parameter combination is assumed equally likely. Such non-informative prior merely scales the posterior density (product of prior and likelihood; or sum of log-prior and $L(\boldsymbol{\theta}|\mathbf{z})$), and similar to the constant in Eq. (11) can be ignored in our calculations with DREAM. The posterior pdf of the parameters is thus solely determined by the information contained in the experimental data. Practical experience suggests that prior distributions are only useful in the presence of limited data or poor sensitivity to the parameters. The more data the larger the value of $L(\boldsymbol{\theta}|\mathbf{z})$ becomes, and hence the less important the prior distribution becomes. The DREAM package can be obtained from the second author upon request.

2.3. Evaluation of different parameter estimation algorithms

We now compare MCMC simulation with DREAM against the classical and REML approaches. The classical or conventional approach first calculated the empirical variogram using the method of moments. A variogram model was then fitted to the empirical data using weighted nonlinear least squares (Minasny and McBratney, 2005). This approach is given in full detail in Webster and Oliver (2007), and so will not be discussed herein.

The second approach uses REML with details given in Lark and Cullis (2004) and Minasny and McBratney (2007). The simplex optimization algorithm was used to find the parameter values that maximize the REML log-likelihood function:

$$L(\theta|\mathbf{z}) = -\frac{n-p}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{K}| - \frac{1}{2}\log|\mathbf{W}| - \frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{Q}\,\mathbf{y} \qquad (12)$$

where $\mathbf{W}=\mathbf{M}^{\mathrm{T}}\mathbf{K}^{-1}\mathbf{M}$, $\mathbf{Q}=\mathbf{I}-\mathbf{M}\mathbf{W}^{-1}\mathbf{M}^{\mathrm{T}}\mathbf{K}^{-1}$ and $\mathbf{y}=\mathbf{T}\mathbf{z}$, denote a stationary data increments transformation of $\mathbf{z}$ with transformation matrix $\mathbf{T}=\mathbf{I}-\mathbf{M}(\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}$.

These two approaches were coded in MATLAB and can be obtained from the first author upon request. All our calculations reported herein were performed using MATLAB 7.10 (32 bit) on a Dell™ Precision workstation with a Dual-Core Intel® Core™ Extreme Processor (2.99 GHz) and 3.25 GB of RAM. A single MCMC trial with DREAM required about 50 min to complete 100,000 functions evaluations for the Matérn covariance model and Samford pH data set (see below). For the Edgeroi carbon data set and exponential covariance model, DREAM took approximately half an hour to perform 200,000 function evaluations, and explore the posterior target distribution.

2.4. Datasets

To test the usefulness and applicability of the DREAM algorithm, we use three different data sets from Australia. These data consider different soil variables and were derived from the following soil survey studies.

- Thickness of soil A horizon in cm (Pettitt and McBratney, 1993). These data are from a field near Forbes, New South Wales. The sampling was based on a nested sampling design where the field was divided into 18 equal-sized blocks of 160 m^2. In each block there were six sampling points along transects of 156 m long separated by 125 m, 25 m, 5 m, 1 m and 25 cm. This study illustrates the estimation of variogram parameters.
- The second dataset illustrates application of kriging using measurements of soil pH in $CaCl_2$ at a research station in Samford, Queensland with an area of 1 ha (Laslett et al., 1987). The samples were collected as follows: 121 grid samples were taken from the nodes of a square grid of 11×11 with a spacing of 10 m, eleven samples were made adjacent (20 cm separation) to those at the grid

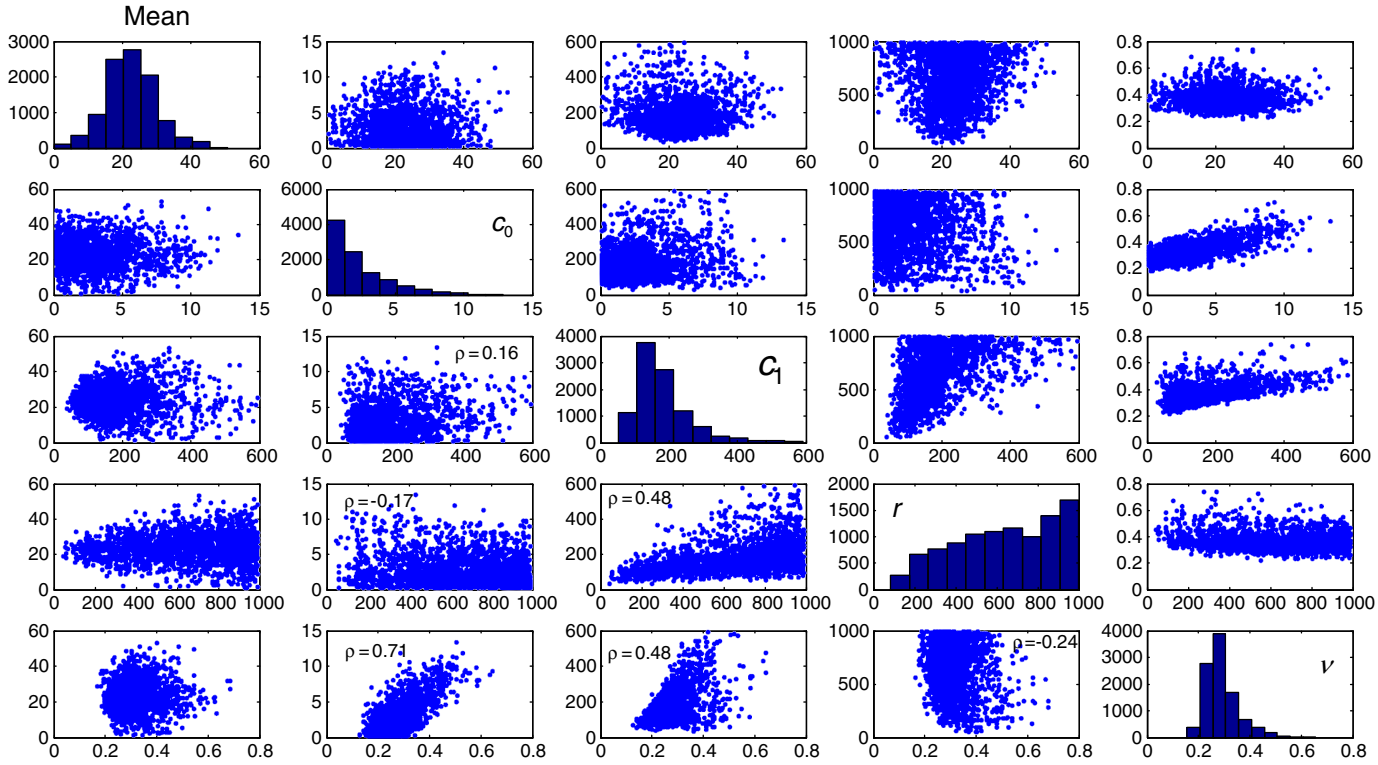

Fig. 1. Distribution and relationship between parameters of the variogram for soil thickness at Forbes. The symbol ρ signifies the linear correlation coefficient.

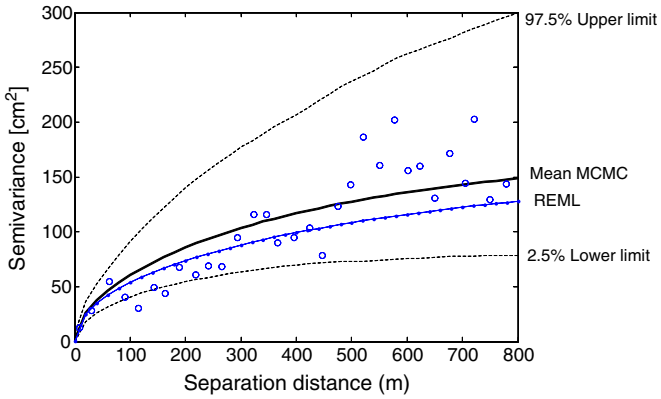

Fig. 2. Estimated variogram for soil thickness at Forbes. The circles represent the empirical variogram calculated using the method of moments.

Table 1
Posterior moments of the parameters for soil thickness at Forbes (in cm) using REML and MCMC simulation with DREAM.

	Mean [cm]	c_0 [cm^2]	c_1 [cm^2]	r [m]	ν	L
2.5 percentile	8.00	0.05	79.92	165.00	0.198	
Median	22.23	1.62	162.28	655.38	0.277	
97.5 percentile	39.33	8.17	414.42	985.49	0.457	
MCMC maximum L	22.28	0.15	118.14	449.79	0.244	−229.10
REML	22.58	4E-14	167.828	971.835	0.239	−229.30

nodes, and eleven samples were located half way between randomly selected pairs of grid sites. Additionally, there is an 8×8 square grid of data sites with 10 m spacing, located at intermediate points within the 11×11 grid. This grid of 8×8 points is not used for model calibration, but used as a validation set to test the spatial prediction methods.

- The third dataset involves the use of environmental covariates for soil mapping. The target variable is soil organic carbon (SOC) content of the surface soil (0–10 cm) in kg m^{-2}, from the survey in Edgeroi, New South Wales (Malone et al., 2009). Samples were obtained from 210 different sites on a systematic equilateral triangular grid with an approximate distance of 2.8 km between sites. An additional 104 sites were collected within the area. This area also has a complete coverage of environmental covariates (digital elevation model, Landsat TM images, gamma radiometrics) at a resolution of 90 m × 90 m. The aim is to map SOC content on a point support at a 90 m × 90 m grid spacing. We randomly selected 214 sites for model calibration, and 100 different sites were used as validation set to test the predictive ability of the different parameter estimation methods.

3. Results

3.1. Soil thickness at Forbes

Fig. 1 shows marginal distributions of the Matérn variogram parameters, including two-dimensional scatterplots of the posterior samples to illustrate pairwise correlation between the different parameters. The histograms and scatter plots were constructed from the last 10,000 samples (from a total of 100,000 iterations) of the posterior distribution generated with DREAM. The variogram parameters exhibit significant uncertainty, and the smoothness parameter ν exhibits a positive correlation with c_0 and c_1. The parameter r is not well defined with a marginal pdf that extends over the entire prior defined ranges. The positive correlation of parameter r with c_1 contributes to this significant uncertainty, and impairs parameter identifiability.

Fig. 2 displays the posterior mean variogram and associated 95% confidence interval (2.5 and 97.5 percentiles). These estimates are

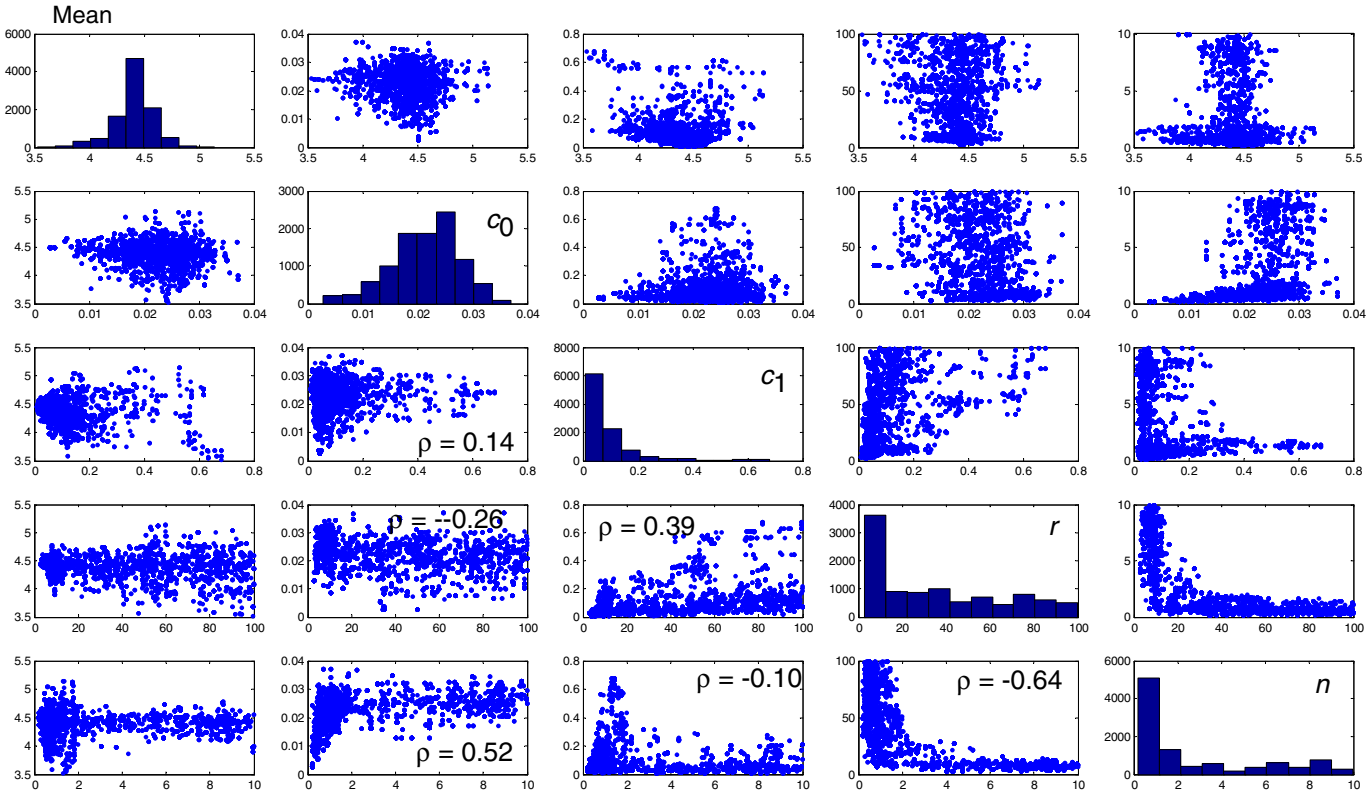

Fig. 3. Distribution and relationship between parameters of the variogram for soil pH at Samford.

Table 2
Posterior moments of parameters for soil pH at Samford using the classical REML approach and MCMC simulation with DREAM.

	Mean	c_0	c_1	r	ν	L
2.5 percentile	3.953	0.008	0.021	4.165	0.242	
Median	4.426	0.022	0.058	27.921	1.114	
97.5 percentile	4.759	0.032	0.393	94.580	9.070	
MCMC maximum L	4.433	0.015	0.029	12.848	0.657	174.38
REML	4.441	0.015	0.033	19.472	0.519	174.41

derived from the 10,000 posterior samples. To benchmark the MCMC results, we also plot the variogram derived with REML and the empirical variogram calculated using the method of moments. The posterior mean variogram derived with MCMC simulation deviates significantly from the variogram estimated with REML, yet the 95% prediction limits estimated with MCMC encompass the REML estimate and empirical variogram.

Table 1 summarizes the posterior values of the variogram model parameters estimated with MCMC simulation and REML. To enable a fair comparison between MCMC and REML, the variogram parameters of REML are first estimated using the log-likelihood function in Eq. (12), and these results are subsequently substituted in Eq. (11). Although REML and MCMC find approximately the same values of the maximum of the log-likelihood function, the corresponding variogram parameter values are quite different. Based on the tabulated statistics, it is particularly difficult to judge which method is more reliable, but the REML optimized value of the nugget variance, c_0 of virtually zero is considerably lower than the short range variation (within 0.25 m) of 1.14 cm^2 as estimated from the nested sampling design by Pettitt and McBratney (1993). The posterior median c_0 value of 1.62 cm^2 derived with MCMC simulation is in much better correspondence with the short-range variation observed in the field.

3.2. Soil pH at Samford

Fig. 3 shows histograms of the marginal posterior probability density functions of the variogram parameters obtained from the last 10,000 samples generated with DREAM. It is interesting to note that the smoothness parameter now displays an apparent non-linear correlation with parameter r. An alternative parameterization of the Matérn function has been suggested by Handcock and Wallis (1994), which allows r to be less dependent on ν.

Table 2 lists the parameters estimated with MCMC simulation and REML. Also for this data set, the maximum log-likelihood value obtained with DREAM is quite close to its counterpart separately derived with REML. Fig. 4 plots the MCMC derived posterior mean variogram and associated 95% uncertainty ranges. The prediction intervals are quite tight at small separation distances, but increase sharply with larger separation distance. The semivariance ranges between 0.03 and 0.28 pH^2 at a separation distance of 120 m, which is about 8 times higher than the uncertainty in the semivariance of the nugget. Note that the posterior mean variogram estimated with MCMC is significantly higher than the empirical variogram and its counterpart estimated with REML.

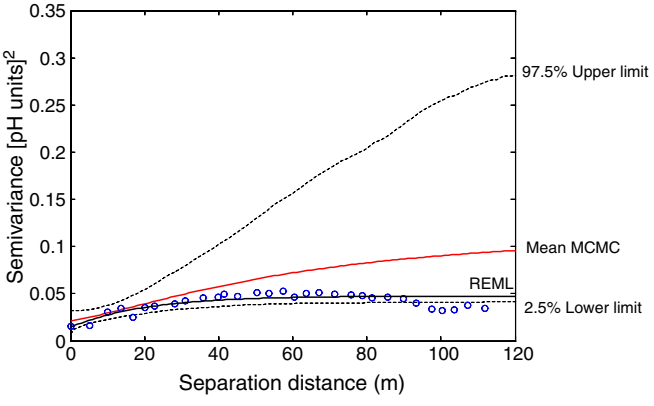

Fig. 4. Estimated variogram for soil pH at Samford. The circles represent the empirical variogram calculated using the method of moments.

We performed kriging on a regular grid of 2 m × 2 m using the posterior distribution of the variogram parameters derived with MCMC simulation using DREAM. For each location in the grid, the mean kriging prediction (Eq. 5) and its associated variance (Eq. 6) were calculated from the last 2000 posterior samples generated with DREAM. Fig. 5 shows the outcome of this analysis, and plots the mean posterior prediction (Fig. 5a), associated posterior variance (Fig. 5b), and mean kriging variance derived from Eq. (6). The effect of variogram parameter uncertainty on the final prediction is immediately apparent. The posterior prediction variance due to parameter uncertainty is highly related to the predicted value. Although the posterior uncertainty of the variogram parameters is quite large, its effect on prediction is quite small. The standard deviation constitutes only 0.05–1.6% of the actual pH prediction, while its variance is only about 0.5% of the actual kriging variance.

Table 3 lists the sum of squared error (*SSE*) and the median of the standardized squared deviation *SSD* values for the 64 independent validation sites. The *SSD* criterion measures the normalized distance to the theoretical estimates:

$$SSD = \frac{\{z(\mathbf{x}) - \hat{z}(\mathbf{x})\}^2}{\sigma_x^2} \tag{13}$$

where $z(\mathbf{x})$ ($\hat{z}(\mathbf{x})$) denotes the measured (predicted) value, and σ_x^2 signifies the kriging variance. A median value close to 0.455 indicates kriging with a correct variogram and data that represent a perfect multivariate-normal process. If the median value is smaller than 0.455 then kriging overestimates the actual variance, possibly due to the presence of outliers that have inflated the variogram. A value of the median larger than 0.455 is representative for an underestimation of the variance (Lark, 2000).

The MCMC approach results in the lowest value of the *SSE*, but the difference with REML is quite small (less than 2%). We also compared our results to a standard kriging method, which fits an exponential model to the empirical variogram. This approach results in a slightly higher *SSE* value. Indeed, for this particular data set the MCMC method receives superior performance, but the differences between the different methods appear relatively minor. All methods produce relatively high values of *SSD* indicating the presence of outliers that cannot be predicted with the different (geostatistical) parameter estimation methods. The independent validation data has a slightly larger variance ($\sigma = 0.257$) as compared to the prediction data ($\sigma = 0.210$). Nevertheless, results from cross validation (of the prediction data) show a similar median *SSD* value of about 0.378 for REML and 0.372 for MCMC with DREAM.

3.3. SOC content at Edgeroi

The final case study reported herein, involves analysis of the SOC data set collected at Edgeroi. The SOC calibration data, expressed in kg m^{-2}, are skewed, and we therefore used a square root transformation to ensure normality. Our preliminary analysis found the following covariates to be useful as predictors in the linear spatial model: Elevation, Multi-resolution Valley Bottom Flatness index (MrVBF, Gallant and Dowling, 2003), slope gradient, topographic wetness index (TWI), Landsat Band 3/2, Landsat Band 3/7 and gamma radiometrics K (Malone et al., 2009). The choice of covariates reflects a combination of topographic variables, soil variable (gamma

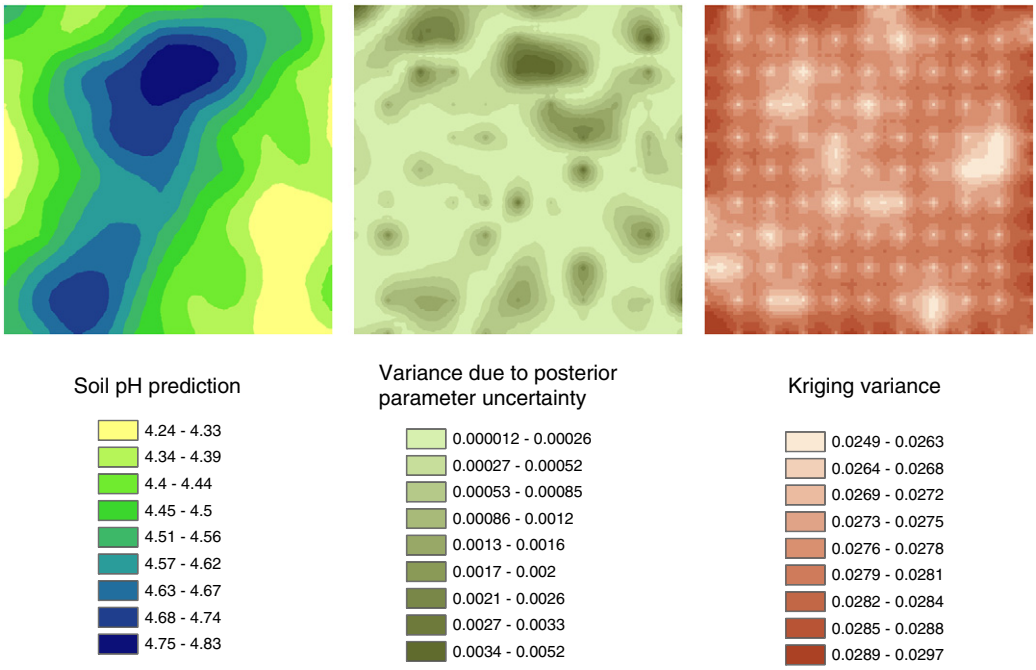

Fig. 5. (a) Posterior mean of the predicted soil pH at Samford using MCMC simulation with DREAM, (b) Variance of the pH prediction due to the posterior parameter uncertainty, (c) mean of the uncertainty in prediction (kriging variance).

radiometrics) and biophysical factors (Landsat TM bands). These different variables are generally useful in digital soil mapping.

We assumed an exponential covariance model, and set $\nu = 0.5$ in the Matérn function. Fig. 6 shows marginal posterior distributions, and scatter plots of the exponential model parameters. The parameter c_0 is relatively well identified, follows a normal distribution, and is appears uncorrelated with c_1 and r. The two other parameters are relatively poorly defined, with posterior ranges that extend a large part of the prior ranges. In particular, r cannot be constrained given the SOC data at hand. The inability to appropriately constrain c_1 and r is in part due to their correlation. This correlation also explains why the uncertainty of the variogram intensifies with increasing separation distance.

Table 4 lists the posterior parameters of the variogram and linear model derived with MCMC simulation, REML and regression-kriging. The MCMC approach results in posterior parameter pdfs that exhibit considerable uncertainty. The optimized parameters of the three different methods are difficult to compare, because each method has used a different log-likelihood function. Yet, it is quite obvious that the results of the methods are quite different, in particular for sill $(c_0 + c_1)$ of the variogram model. The MCMC estimates of this parameter are about twice higher than their counterparts separately derived with REML and regression kriging. Without a detailed assessment of the posterior uncertainty associated with the REML and regression kriging estimates it remains difficult to directly compare the parameter values. Fig. 7 plots histograms and pairwise scatter plots of the posterior parameters of the linear model. Most of the marginal posterior distributions exhibit significant uncertainty (intercept ranges between −8 and 4), but are well described with a normal distribution. It is interesting to observe that the posterior coefficients of the covariates (parameters) exhibit considerable scatter and correlation. For instance, consider the strong correlation between the intercept, and parameters of elevation and TWI. This suggests that a more parsimonious model could have been used, a conclusion that is more difficult to draw with REML and regression kriging when only information about the maximum likelihood parameter values is available.

Fig. 8 displays the mean variogram of the posterior distribution derived with MCMC simulation and the associated 95% prediction uncertainty intervals. We also include the results of the regression-kriging approach. This method treats the trend and residuals separately. First, a linear model is fitted to the data using ordinary least-squares. After this, the variogram of the residuals of the linear model is calculated using the method of moments. Finally, the exponential function is fitted to the variogram of the residuals. The empirical variogram of the residuals does not show spatial correlation at distances less than 10 km, whereas both MCMC and REML identify a spatial correlation at this distance.

Fig. 8 shows the MCMC derived posterior mean SOC prediction of the top 10 cm of the soil. To provide further insights into the performance of DREAM, we include marginal histograms of the SOC prediction at three different spatial locations on the map. On the plain (in the west), where SOC is low, the prediction is quite certain with a relatively tight distribution (3.2–3.6 $[\text{kg m}^{-2}]^{0.5}$). However in the high elevation area at the east, the distribution of SOC values is quite wide (5–17 $[\text{kg m}^{-2}]^{0.5}$). It is interesting to observe that in some areas, the marginal distributions of SOC deviate considerably from a normal distribution, demonstrating the advantages of MCMC for spatial prediction and uncertainty quantification. Ordinary kriging and related methods typically rely on the assumption of Gaussianity of the prediction error.

Fig. 9 compares the posterior variance of the SOC prediction (top) with that of the kriging variance (bottom). The variance of the

Table 3
Comparison between various parameter estimation techniques using the soil pH data set from Samford. The acronyms *SSE* and *SSD* stand for sum of squared error, and standardized squared deviation, respectively.

	SSE	Median *SSD*
MCMC — Matérn	3.02	0.829
REML — Matérn	3.07	0.823
MCMC — exponential	3.04	0.713
REML — exponential	3.06	0.788
Kriging — exponential	3.06	0.850

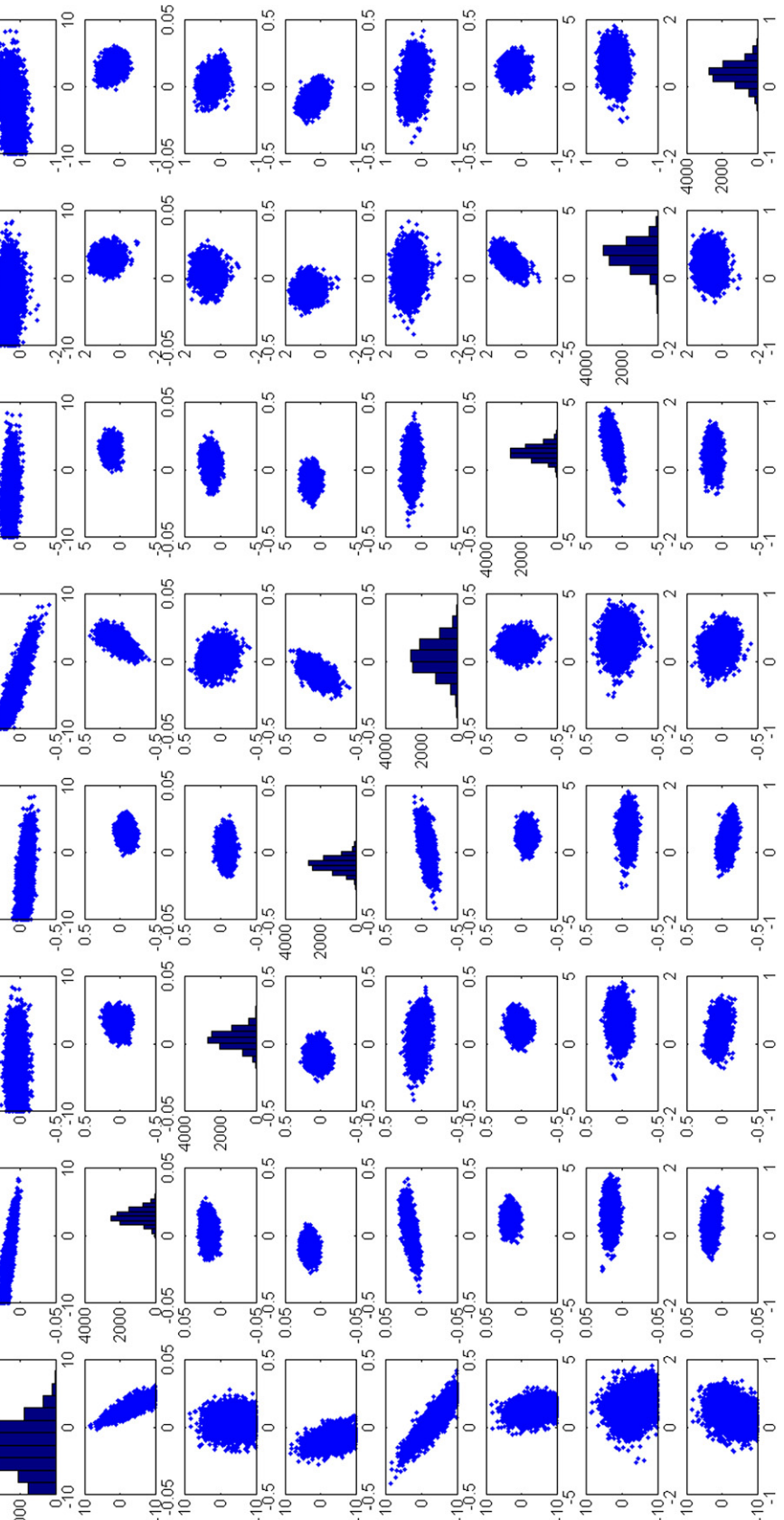

Fig. 6. Distribution and relationship between parameters of the variogram for topsoil organic carbon content at Edgeroi.

Table 4
Posterior moments of parameters for the prediction of soil organic carbon content (in $[kg\ m^{-2}]^{0.5}$) at Edgeroi using MCMC simulation with DREAM, REML and regression-kriging (RK). Maximum *L* refers to parameter values obtained at maximum likelihood using DREAM.

	c_0 $[kg\ m^{-2}]$	c_1 $[kg\ m^{-2}]$	r [m]	L
2.5 percentile	0.453	0.126	2058.2	
Median	0.601	0.491	14,986.2	
97.5 percentile	0.760	1.603	29,208.0	
MCMC maximum *L*	0.538	0.191	2891.9	−81.90
REML	0.562	0.249	3735.2	−80.56
RK	0.695	0.074	3063.3	−81.93

	Intercept	Elevation	MrVBF	Slope	TWI	Landsat B3/2	Landsat B3/7	gamma K
2.5 percentile	−7.992	0.001	−0.123	−0.226	−0.344	0.088	−0.134	−0.092
Median	0.294	0.011	0.028	−0.102	−0.071	1.225	0.635	0.260
97.5 percentile	8.193	0.023	0.171	0.018	0.200	2.208	1.426	0.590
MCMC maximum *L*	3.826	0.007	−0.023	−0.174	−0.168	1.321	0.546	0.260
REML	1.842	0.008	0.023	−0.107	−0.100	1.204	0.631	0.259
RK	4.443	0.004	−0.018	−0.119	−0.175	1.331	0.596	0.241

prediction of the linear model and variogram caused by the posterior parameter uncertainty is quite small compared to the kriging variance. This is consistent with the findings in our study, and in part due to the variogram model used. High elevation areas in the east display the largest SOC prediction uncertainties. This is caused by a lack of samples in this part of the domain, and significant spatial variations in elevation and terrain attributes. The kriging variance closely mimics the patterns of the observations. The variance is relatively small in the vicinity of individual observations and increases with separation distance from the different measurement points (Fig. 10). This is consistent with the functional shape of the variogram, and the increasing uncertainty of the semivariance with larger separation lags.

We compared the performance of MCMC simulation, REML and the regression-kriging approach for the 100 different measurement sites not included in the calibration. Table 5 summarizes the results of this analysis. The MCMC method generates the most consistent spatial predictions, with *SSE* value that is substantially smaller than the respective counterparts found with REML and regression kriging. Yet, the differences with REML can be considered relatively minor. Also with this data se the median *SSD* value is substantially smaller than 0.455, indicating the presence of several outliers that cannot be predicted with any of the different fitting techniques. The *SSD* statistic might therefore not be a reliable measure to evaluate the prediction at true validation sites.

4. Discussion

To the best of our knowledge this is the first study that has quantified, using Bayesian statistics and MCMC simulation, the posterior uncertainty of soil variogram model parameters, and their effect on spatial predictions. Previous attempts to use MCMC in geostatistical analysis have been plagued by problems with a correct choice of the proposal distribution, the esoteric nature of the probability manipulations, computer automation, and premature convergence. The DREAM code is an adaptive MCMC method that automatically tunes the scale and orientation of the proposal distribution during sampling, and returns the posterior pdf of the parameters without human intervention. The only information that need to be specified by the user are the initial parameter ranges, the prior distribution (can be flat — noninformative prior) and log-likelihood function used to compare the model predictions and data. The posterior parameter distribution contains desirable information about parameter uncertainty, correlation, and thus how many parameters are warranted by the calibration data. This distribution

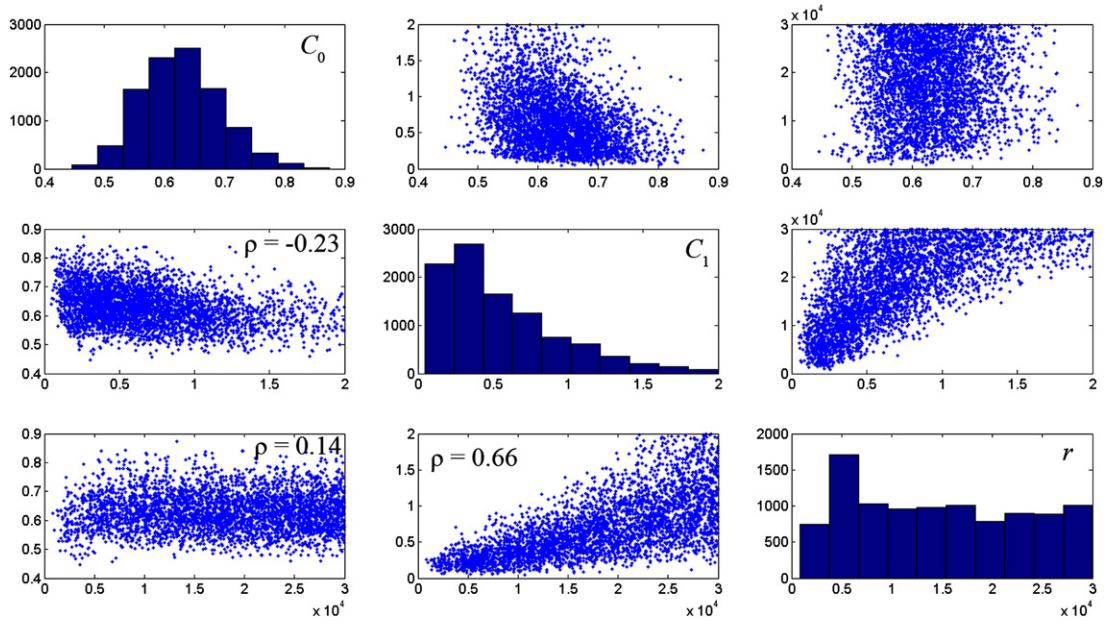

Fig. 7. Distribution and relationship between parameters of the linear model for topsoil organic carbon content at Edgeroi.

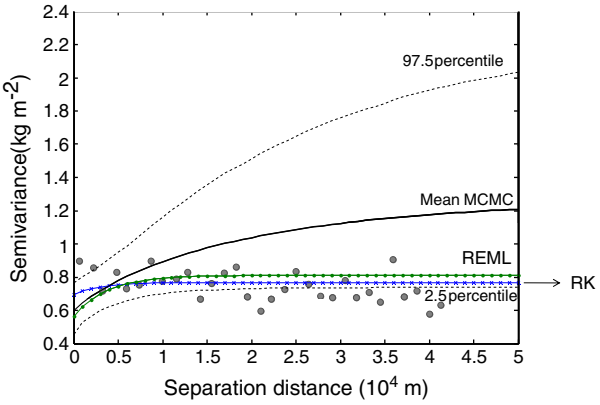

Fig. 8. Estimated variogram for topsoil organic carbon content at Edgeroi. The circles represent the empirical variogram calculated using the method of moments and the regression-kriging method.

can be used to assess model predictive uncertainty. This simply requires running the posterior samples through the model, and computing the posterior mean and variance of the prediction.

The posterior distribution typically deviates considerably from a multi-normal distribution. Furthermore there appears to be a high correlation between the parameters, especially the partial sill c_1 and range r. Lark et al. (2006) demonstrated with simulations that when the variogram of the residuals of the regression model were estimated using a regression-kriging approach, a bias would occur at long lags. This study, however, shows that prediction uncertainty at large separation distances can be quite high because of the high correlation between c_1 and r. Nevertheless, the estimates by REML and method of moments usually fall within the 95% uncertainty bounds of the estimates derived with MCMC simulation. Classical geostatistical methods thus remain useful, and exhibit no apparent bias when used for spatial prediction.

Handcock and Stein (1993) suggested that fitting empirical variograms may lead to predictive distributions that differ markedly from their respective counterparts obtained by using a formal Bayesian approach. While MCMC methods can estimate any arbitrary pdf, our case studies have shown that the (spatial) prediction uncertainty due to the posterior parameter uncertainty is quite small compared to the kriging variance. This conclusion is in part caused by the variogram models used, and underlying assumption of Gaussian distributed error residuals in our likelihood function. Explicit consideration of model structural error, and spatial correlation of the residuals in the formulation of the likelihood function is likely to significantly increase parameter uncertainty. This is our experience in time series analysis when fitting hydrologic models to time series of discharge observations (Schoups and Vrugt, 2010). To further improve the Bayesian analysis considered herein, we need to develop novel spatial prediction models, and likelihood functions that properly recognize the role of model error (structural inadequacies in the spatial prediction method), and calibration data error.

Here we summarize the main advantages of MCMC:

1. The likelihood or REML method, is designed to only infer the maximum likelihood values of the parameters without recourse to estimating the underlying posterior parameter uncertainty. On the contrary, MCMC methods estimate the entire posterior pdf of the parameters. This includes the maximum likelihood values of the parameters.
2. This posterior distribution conveys important information about parameter uncertainty, and multi-dimensional parameter correlation, and hence which parameters are well-identified given the calibration data set, and which parameters cannot be inferred. Parameters whose marginal distribution occupies only a small part of the prior ranges are well identifiable. Marginal distributions that extend the entire prior defined bounds are indicative for parameters that cannot be well identified given the calibration data at hand.

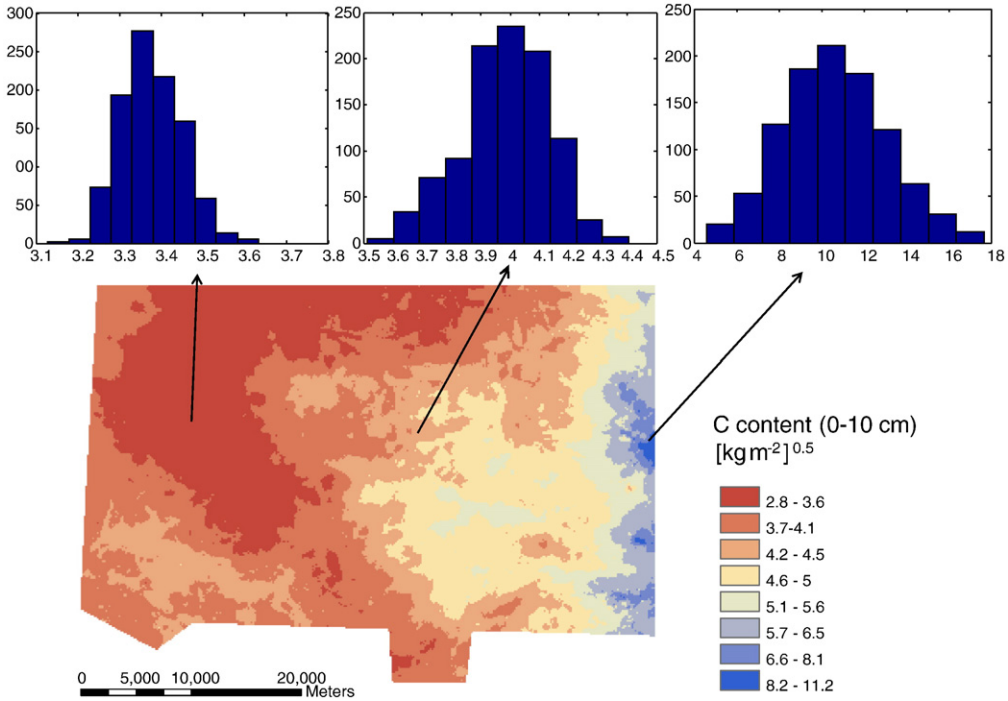

Fig. 9. Posterior mean of topsoil SOC predictions at Edgeroi using MCMC simulation with DREAM. To provide insights into the predictive distribution, we separately include histograms of the posterior prediction at three different spatial locations. The plotted values are square root transforms of the original SOC content in kg m^{-2}.

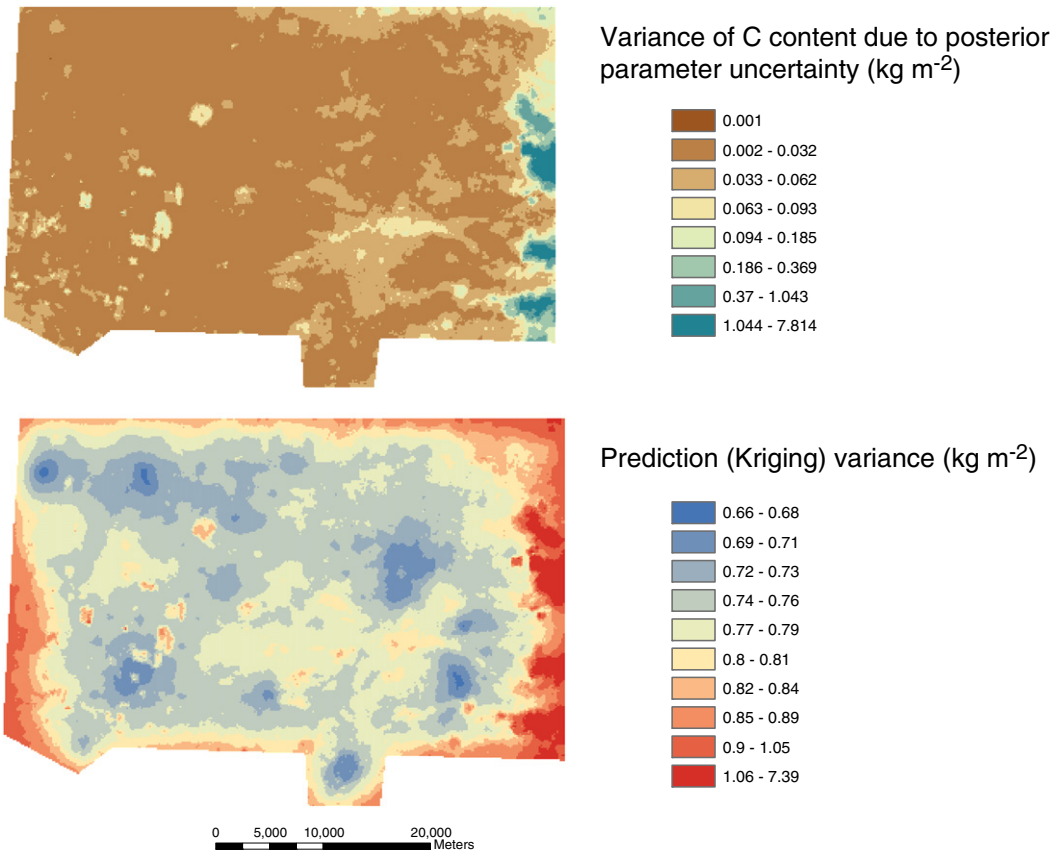

Fig. 10. (a) The variance of the SOC prediction due to posterior parameter uncertainty, (b) mean of the uncertainty in prediction (kriging variance). The values are square-root transforms of SOC content in kg m^{-2}.

3. The joint distribution also provides information on the correlation between the parameters. Two- and higher dimensional scatterplots between the different dimensions of the posterior distribution, reveal the presence of linear and nonlinear correlation among the parameters. This provides important insights into the uncertainty of the parameters and how many parameters are warranted by the calibration data.
4. The posterior distribution can be propagated through the model to get the prediction uncertainty at any defined interval, *e.g.* 99%, 95%, and 90%. This uncertainty envelope can then be compared against the data and visual inspection of this model — data mismatch can be used to explore structural inadequacies (also called model error) in the spatial model. Such analysis provides the necessary inspiration to improve our spatial prediction techniques.
5. MCMC is computationally intensive; however the REML approach can also be time consuming when prior information is vague or not available. Perhaps more important, REML is prone to getting stuck in a local minima. Repeated trials with different initial parameter values are needed to inspire confidence in the final optimized variogram. MCMC simulation with DREAM, on the contrary, consistently converges to the same approximate posterior distribution, irrespective of the initial starting points of the *N* chains. We therefore posit that DREAM is more efficient than classical geostatistical methods and REML.
6. The DREAM algorithm is especially designed to work with any likelihood function and simulation model, and can successfully solve high-dimensional inversion (posterior exploration) problems. Thus, it can be used for estimating parameters in a linear model of coregionalization (Lark and Papritz, 2003), and complex non-stationary models (Haskard et al., 2010). Moreover, Schoups and Vrugt (2010) recently presented a generalized likelihood function to deal with residual errors that are correlated, heteroskedastic, and non-Gaussian with varying degrees of kurtosis and skewness. Nonlinear models can also easily be incorporated. This will be difficult to implement in other likelihood approaches, such as REML.

Table 5
Comparison of the various variogram parameter estimation techniques using the observed SOC content of the top soil (units in $[kg\ m^{-2}]^{0.5}$) at 100 different sites at Edgeroi not included in the calibration. The *SSE* and *SSD* criteria have been previously defined in the main text, and caption of Table 3.

	SSE	Median *SSD*
MCMC	73.1	0.207
REML	78.1	0.266
Regression-kriging	80.5	0.252

For operational purposes, *e.g.* producing digital soil maps over large areas, current state-of-the-art MCMC approaches are still inefficient, and standard kriging and regression-kriging approach need to be used. Yet, recent advances in distributed computing using multi-try-DREAM sampling significantly increase the efficiency of MCMC simulation for high-dimensional posterior inference problems (Laloy and Vrugt, submitted for publication).

5. Conclusions

1. The DREAM algorithm, which automatically tunes the scale and orientation of the proposal distribution en route to the posterior target distribution, facilitates an easy implementation of MCMC simulation for geostatistical parameter inference. Based on

information about the parameter ranges, initial distribution, and log-likelihood function, DREAM returns the posterior pdf of the parameters.
2. The posterior parameter distribution derived with MCMC simulation can be used to summarize variogram and prediction uncertainty. Variogram uncertainty increases with separation distance, and is quite significant at large lags.
3. The variogram estimates by the conventional method of moments usually fall within the 95% uncertainty bounds derived with MCMC simulation. There thus remains significant scope for these original methods.
4. Posterior parameter uncertainty constitutes only a minor part of prediction uncertainty, which is dominated by the kriging variance. This conclusion is in part due to the variogram model and likelihood function used for spatial analysis. Parameter uncertainty is likely to increase if model error is explicitly considered.
5. Whereas REML is prone to getting stuck in a local basin of attraction, and thus requires repeated parameter estimation trials with different starting points, a single run with DREAM is sufficient to find the actual target distribution. MCMC simulation with DREAM therefore speeds up the efficiency of parameter estimation.

Acknowledgments

The DREAM software in Matlab can be obtained from the second author (jasper@uci.edu) upon request. The R code for DREAM is under development and available at http://dream.r-forge.r-project.org/. The data used in this study is available at http://www.pedometrics.org. Budiman Minasny acknowledges the QEII fellowship received from the Australian Research Council. The authors thank the two reviewers for their constructive suggestions.

References

Aelion, C.M., Davis, H.T., Liu, Y., Lawson, A.B., McDermott, S., 2009. Validation of Bayesian kriging of arsenic, chromium, lead, and mercury surface soil concentrations based on internode sampling. Environmental Science & Technology 43, 4432–4438.
Christensen, O.F., 2004. Monte Carlo maximum likelihood in model-based geostatistics. Journal of Computational and Graphical Statistics 13, 702–718.
Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. Journal of the Royal Statistical Society Series C-Applied Statistics 47, 299–326.
Gallant, J.C., Dowling, T.I., 2003. A multi resolution index of valley bottom flatness for mapping depositional areas. Water Resources Research 39, 1347–1360.
Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–511.
Gelman, A., Carlin, J.B., Stren, H.S., Rubin, D.B., 2004. Bayesian Data Analysis. Chapman & Hall, Boca Raton, Florida, USA.
Haario, H., Saksman, E., Tamminen, J., 1999. Adaptive proposal distribution for random walk Metropolis algorithm. Computational Statistics 14, 375–395.
Haario, H., Saksman, E., Tamminen, J., 2001. An adaptive Metropolis algorithm. Bernoulli 7, 223–242.
Handcock, M.S., Stein, M.L., 1993. A bayesian-analysis of kriging. Technometrics 35, 403–410.
Handcock, M.S., Wallis, J.R., 1994. An approach to statistical spatial–temporal modeling of meteorological fields (with discussion). Journal of the American Statistics Association 89, 368–390.
Haskard, K.A., Welham, S.J., Lark, R.M., 2010. Spectral tempering to model non-stationary covariance of nitrous oxide emissions from soil using continuous or categorical explanatory variables at a landscape scale. Geoderma 159, 358–370.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.
Hudson, G., Wackernagel, H., 1994. Mapping temperature using kriging with external drift: theory and an example from Scotland. International Journal of Climatology 14, 77–91.
Huisman, J.A., Rings, J., Vrugt, J.A., Sorg, J., Vereecken, H., 2010. Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion. Journal of Hydrology 380, 62–73.
Kavetski, D., Kuczera, G., Franks, S.W., 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research 42, W03407. doi:10.1029/2005WR004368.
Laloy, E., Vrugt, J.A., submitted for publication. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM$_{(ZS)}$ and high performance computing. Water Resources Research.
Lark, R.M., 2000. A comparison of some robust estimators of the variogram for use in soil survey. European Journal of Soil Science 51, 137–157.
Lark, R.M., Cullis, B.R., 2004. Model-based analysis using REML for inference from systematically sampled data on soil. European Journal of Soil Science 55, 799–813.
Lark, R.M., Papritz, A., 2003. Fitting a linear model of coregionalization for soil properties using simulated annealing. Geoderma 115, 1–16.
Lark, R.M., Cullis, B.R., Welham, S.J., 2006. On spatial prediction of soil properties in the presence of a spatial trend: the empirical best linear unbiased predictor (E-BLUP) with REML. European Journal of Soil Science 57, 787–799.
Laslett, G.M., McBratney, A.B., Pahl, P.J., Hutchinson, M.F., 1987. Comparison of several spatial prediction methods for soil pH. Journal of Soil Science 38, 325–341.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.
Marchant, B.P., Lark, R.M., 2004. Estimating variogram uncertainty. Mathematical Geology 36, 867–898.
Marchant, B.P., Lark, R.M., 2007. The Matérn variogram model: implications for uncertainty propagation and sampling in geostatistical surveys. Geoderma 140, 337–345.
Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 72, 135–146.
Matheron, G., 1969. Le Krigeage Universel. Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc.1. Ecole des Mines de Paris.
Minasny, B., McBratney, A.B., 2005. The Matern function as a general model for soil variograms. Geoderma 128, 192–207.
Minasny, B., McBratney, A.B., 2007. Spatial prediction of soil properties using EBLUP with the Matérn covariance function. Geoderma 140, 324–336.
Moyeed, R.A., Papritz, A., 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. Mathematical Geology 34, 365–386.
Odeh, I.O.A., McBratney, A.B., Chittleborough, D.J., 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67, 215–226.
Pettitt, A.N., McBratney, A.B., 1993. Sampling designs for estimating spatial variance-components. Applied Statistics-Journal of the Royal Statistical Society Series C 42, 185–209.
Raftery, A.E., Lewis, S., 1992. Bayesian Statistics 4, How many iterations in the Gibbs sampler? Oxford University Press, Oxford, U.K.
Ribeiro, P.J., Christensen, O.F., Diggle, P.J., 2003. geoR and geoRglm: software for model-based geostatistics. In: Hornik, K., Leisch, F., Zeileis, A. (Eds.), Proceedings of the 3 rd International Workshop on Distributed Statistical Computing (DSC 2003). Vienna, Austria.
Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic and non-Gaussian errors. Water Resources Research 46, W10531.
Ter Braak, C.J.F., 2006. A Markov Chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. Statistics and Computing 16, 239–249.
Ter Braak, C.J.F., Vrugt, J.A., 2008. Differential Evolution Markov Chain with snooker updater and fewer chains. Statistics and Computing 18, 435–446.
Vrugt, J.A., Bouten, W., 2002. Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. Soil Science Society of America Journal 66, 1740–1752.
Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resources Research 39 art. No. 1201.
Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A., 2008. Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44, W00B09.
Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Robinson, B.A., Hyman, J.M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences and Numerical Simulation 10, 273–290.
Webster, R., Oliver, M.A., 2007. Geostatistics for Environmental Scientists. Wiley, Chichester.
Zhang, H., 2002. On estimation and prediction for spatial generalized linear mixed models. Biometrics 58, 129–136.
Zimmerman, D.L., Cressie, N., 1992. Mean squared prediction error in the spatial linear model with estimated covariance parameters. Annals of the Institute of Statistical Mathematics 44, 27–43.