



Geochemistry, Geophysics, Geosystems

RESEARCH ARTICLE

10.1029/2019GC008465

Key Points:

- We enhance the Bayeslands framework for estimation of free parameters in landscape evolution models
- We use high-performance computing with parallel tempering MCMC for enhancing Bayeslands
- The enhanced Bayeslands framework not only reduces the computation time, but also provides better exploration of multimodal distributions

Correspondence to:

R. Chandra,
rohitash.chandra@sydney.edu.au

Citation:

Chandra, R., Muller, R. D., Azam, D., Deo, R., Butterworth, N., Salles, T., & Cripps, S. (2019). Multicore parallel tempering Bayeslands for basin and landscape evolution. *Geochemistry, Geophysics, Geosystems*, 20, 5082–5104. <https://doi.org/10.1029/2019GC008465>

Received 24 MAY 2019

Accepted 23 AUG 2019

Accepted article online 23 AUG 2019

Published online 16 NOV 2019

Multicore Parallel Tempering Bayeslands for Basin and Landscape Evolution

Rohitash Chandra^{1,2} , R. Dietmar Müller¹ , Danial Azam¹, Ratneel Deo¹, Nathaniel Butterworth³ , Tristan Salles¹ , and Sally Cripps⁴

¹EarthByte Group, School of Geosciences, University of Sydney, Sydney, New South Wales, Australia, ²Centre for Translational Data Science, University of Sydney, Sydney, New South Wales, Australia, ³Sydney Informatics Hub, University of Sydney, Sydney, New South Wales, Australia, ⁴School of Mathematics and Statistics, University of Sydney, Sydney, New South Wales, Australia

Abstract The Bayesian paradigm is becoming an increasingly popular framework for estimation and uncertainty quantification of unknown parameters in geophysical inversion problems. Badlands is a *landscape evolution model* for simulating topography evolution at a broad range of spatial and temporal scales. Our previous work presented Bayeslands that used the Bayesian inference for estimating unknown parameters in the Badlands model using Markov chain Monte Carlo sampling. Bayeslands faced challenges in terms of computational issues and convergence due to multimodal posterior distributions. Parallel tempering is an advanced Markov chain Monte Carlo method suited for irregular and multimodal posterior distributions. In this paper, we extend Bayeslands using parallel tempering with high-performance computing to address previous limitations in Bayeslands. Our results show that parallel tempering Bayeslands not only reduces the computation time, but also provides an improvement in sampling multimodal posterior distributions, which motivates future application to continental scale landscape evolution models.

1. Introduction

Understanding landscape evolution, and the associated accumulation of sediments in basins, is limited by increasingly sparse data back through geological time. Recent developments in landscape evolution models (LEMs) (Coulthard, 2001), such as the basin and landscape dynamics (Badlands) model, feature responses to surface uplift and subsidence over a large range of spatial scales and track sediments from source to sink (Salles & Hardiman, 2016; Salles et al., 2017, 2018). Badlands also has the capability to create synthetic basin stratigraphies. LEMs depend on uncertain initial and boundary conditions, and unknown parameters (Scott & Zhang, 1990), such as the initial topography, sea level, precipitation, and rock lithology and erodibility (Godard et al., 2006; Rejman et al., 1998). Estimating and quantifying uncertainty surrounding them given the topography changes over time is an incredibly difficult task.

To make inference about the parameters that govern landscape evolution, we need to constrain the model by fusing information from various sources in a probabilistic model. The sources of information include underlying equations that govern LEMs to represent geophysical processes, observed data such as present day topography, stratigraphy in sedimentary basins, and previous research that could be denoted as expert knowledge (Salles et al., 2018).

The Bayesian framework provides a logically consistent mechanism for fusing information from various sources to provide meaningful inference for unknown parameters in models. The prior distribution is a way to incorporate information from expert knowledge, and the likelihood expresses how likely the proposed values of the parameters are, given data and the model (Tarantola, 2006). The ability to fuse information from many sources in a principled fashion has made Bayesian inference an increasingly popular choice for the estimation and uncertainty quantification of parameters in complex models (Gallagher et al., 2009; Mosegaard & Vestergaard, 1991; Robert & Casella, 2011; Rocca et al., 2009; Sen & Stoffa, 2013). More specifically, there is some work that employs Bayesian inversion methodologies for landscape evolution-based forward models for specific applications. Initial work was done by Roberts and White (2010), who used Monte Carlo inverse modeling of river profiles to estimate uplift rate histories using African examples. Fox et al. (2014) reviewed uplift and erosion rate history of the Bergell intrusion from the inversion of low

temperature thermochronometric data, and Goren et al. (2014) focused on distinguishing tectonics from fluvial topography using formal linear inversion with applications to the Inyo Mountains, California. Fox et al. (2015) investigated the rate of Andean Plateau uplift using reversible jump Markov chain Monte Carlo (MCMC) inversion of river profiles.

Although Bayesian inversion has become popular in geophysics in the past few decades (Grandis et al., 1999; Malinverno, 2002; Mosegaard & Tarantola, 1995; Sambridge & Mosegaard, 2002; Sambridge, 1999), estimating the posterior distribution is often nontrivial. The posterior distributions can be multimodal, exhibit discontinuities, and are usually not available in closed form. MCMC sampling methods are used for estimation of such posterior distributions and crafting proposal distributions to explore them is very difficult (Gallagher et al., 2009).

The Metropolis-Hastings (MH) algorithm (Chib & Greenberg, 1995; Hastings, 1970; Metropolis et al., 1953) is a popular MCMC method to obtain iterations from a distribution that cannot be sampled directly. In the MH algorithm, the current state in the Markov chain moves to a new state via a transition kernel. This transition kernel consists of a proposal distribution and an acceptance probability. Given some regularity condition, this acceptance probability is constructed so that the Markov chain will converge to the posterior (stationary) distribution, irrespective of the proposal distribution. However, the proposal distribution does affect the rate at which the chain converges. A poorly chosen proposal distribution can result in an acceptance probability close to zero, and the chain will not converge in a timely fashion. Information about the gradient of the posterior may help with the construction of better proposal distribution (Girolami & Calderhead, 2011; Hoffman & Gelman, 2014; Neal, 2011); however, in geophysical forward models, the gradients are often unavailable or computationally expensive to obtain. In situations such as these, parallel tempering Markov Chain Monte Carlo (PT-MCMC)(Geyer & Thompson, 1995; Marinari & Parisi, 1992) offers a better chance of convergence.

PT-MCMC is well suited for exploring multimodal distributions by running replicas of the Markov chains in parallel. In PT-MCMC, the replicas converge to different stationary distributions, which effectively smooth out the local modes that exist in the posterior. The parameters in replicas are allowed to swap with each other, ensuring that the target chain will contain draws from other local maxima. In addition, parallel computing-based implementation of PT-MCMC can reduce the computational burden associated with complex forward models such as Badlands (Mills et al., 1992; Vrugt et al., 2006; Zhang et al., 2007). The potential for PT-MCMC in geoscience has been demonstrated for complex multimodal problems (Maraschini & Foti, 2010; Sambridge, 2013; Scalzo et al., 2019; Sen & Stoffa, 1996; Sen & Stoffa, 2013). Although the recent implementation of PT-MCMC with gradient-based proposals has shown promising performance (Chandra, Jain, et al., 2019), it is not feasible for the Badlands model because gradients are unavailable from the model.

Although MCMC methods with randomwalk proposal distributions are feasible for models that do not provide gradients, they are inefficient and likely to get stuck in local modes (Neal, 1996). In our previous work (Chandra, Azam, et al., 2019), we presented the Bayeslands framework that used single-chain MCMC with random walk proposal distribution. The method did not use parallel computing and hence referred to as single-chain Bayeslands (SC-Bayeslands), hereafter. SC-Bayeslands demonstrated that even in low-dimensional settings, the posterior surfaces of parameters exhibited highly irregular features, such as multimodality and discontinuities, making sampling difficult.

In this paper, we present a multicore parallel tempering Bayeslands (PT-Bayeslands) that features uncertainty quantification and estimation of selected parameters for basin and landscape evolution. We investigate several issues such as sampling multimodal posterior distribution, proposal distribution, computational time, and prediction quality of Badlands. We apply our technique to simulated and real-world topographies.

2. Background and Related Work

2.1. Badlands

Over the last decades, many numerical models have been proposed to simulate how the Earth surface has evolved over geological timescales in response to different driving forces such as tectonics or climatic variability (Adams et al., 2017; Campforts et al., 2017; Salles & Duclaux, 2015; Tucker & Hancock, 2010; Whipple & Tucker, 2002). These models combine empirical data and conceptual methods into a set of

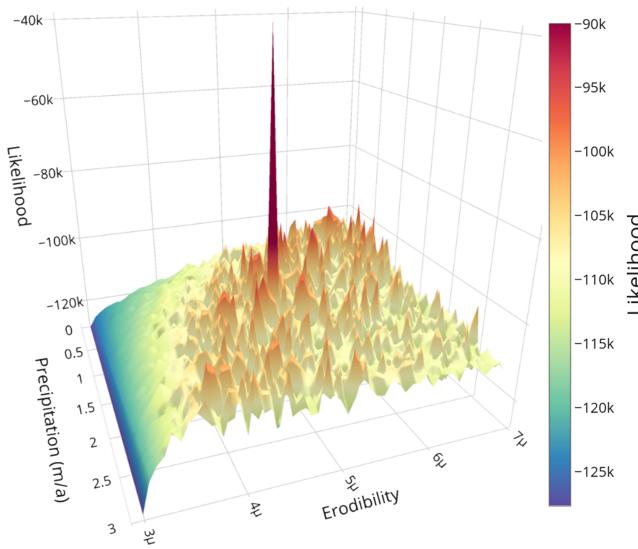


Figure 1. Log posterior (likelihood) surface of the continental margin problem as function of precipitation, ρ , and erodibility, ϵ . The true values have a clear peak in the log-posterior (precipitation = 1.5 m/a, erodibility = $5e^{-6}$); however, many suboptimal peaks (marked in red) for this parameter pair exist (Chandra et al., 2019).

mathematical equations that can be used to reconstruct landscape evolution and associated sediment fluxes (Hobley et al., 2011; Howard et al., 1994). They are currently used in many research fields such as hydrology, soil erosion, hillslope stability, and general landscape studies.

We use Badlands (Salles & Hardiman, 2016; Salles, 2016; Salles et al., 2018) to simulate regional to continental sediment deposition and associated sedimentary basin architecture (Salles et al., 2017, 2018). In its basic formulation, the Earth surface elevation changes, denoted by $\frac{\partial z}{\partial t}$, where z is the elevation and t refers to time in years related to the interaction of three types of processes: the tectonic uplift rate (U), the incision rate by rivers (I), and the hillslope processes (D).

$$\frac{\partial z}{\partial t} = U - I + D \quad (1)$$

In this study, fluvial incision rates and predicted sediment transport in rivers are solved using the stream power law (SPL; Harel et al., 2016; Stock & Montgomery, 1999). SPL relates the erosion rate I to the product of mean annual net precipitation rate (\bar{P}), drainage area (A), and local river gradient (S) and takes the form

$$I = k(\bar{P}A)^\gamma S^\lambda, \quad (2)$$

where k is an erodibility coefficient that depends on lithology and climate, while γ (<2) and λ (<4) are positive exponents (Chen et al., 2014) that mostly depend on catchment hydrology and the nature of the dominant erosional mechanism (Whipple & Tucker, 2002). Despite its simplicity, equation (2) reproduces many of the characteristics of mountainous landscapes, where detachment-limited erosion regime dominates (Tucker & Hancock, 2010). κ varies by several orders of magnitude not only based on lithology, climate, sedimentary flux, or river channel width but also with the chosen values of γ and λ .

In addition to overland flow, semicontinuous processes of soil displacement are accounted for using a linear diffusion law commonly referred to as soil creep (Tucker & Hancock, 2010):

$$D = \delta \nabla^2 z, \quad (3)$$

where δ is the diffusion coefficient. This transport rate depends linearly on topographic gradient and encapsulates in a simple formulation the processes operating on superficial sedimentary layers.

2.2. SC-Bayeslands

The posterior distributions of parameters in geophysical inversion problems are notoriously difficult to sample (Sambridge, 1999). They are high dimensional and multimodal and sometimes exhibit discontinuities. SC-Bayeslands provides a framework for traversing the posterior distribution of selected parameters of the Badlands model. In our previous work (Chandra, Azam, et al., 2019), we found that SC-Bayeslands had difficulty in convergence due to multimodal posterior distributions. In the experiments, we used two problems, namely, the synthetic crater (Cr) and the continental margin (CM) problem to demonstrate the effectiveness of Bayeslands. We considered estimation of *precipitation* and *erodibility*, which are the critical parameters in the Badlands model.

Figure 1 shows an example of the posterior distributions for precipitation and erodibility, while fixing all other parameters for the CM problem. We note that our previous work (Chandra et al., 2019) showed that different combinations of precipitation and erodibility gave rise to visually indistinguishable topography for both the Cr and CM problem. Figure 1 has a definite global maximum, but also several local maxima. Unless the chain has a starting value near this global maximum, it is unlikely that SC-Bayeslands would find it. However, it was shown that convergence to some of the local maxima gave topography prediction that remarkably resembled the ground-truth topography.

Table 1
Landscape Evolution Problems

Topography	T_{max} (years)	Length [km, pts]	Width [km, pts]	Resolution	Run time (s)
Cr	50,000	[0.24, 123]	[0.24, 123]	0.002	2.0
Mt	1,000,000	[80,202]	[40,102]	1.000	10.0
CM	1,000,000	[136.0, 136]	[123.0, 123]	1.000	7.5

Note. Cr = synthetic crater; Mt = synthetic mountain; CM = continental margin. The resolution factor (Res. factor) indicates the distance in kilometers (km) between two neighboring points (pts) along x or y axis of the topography grid. T_{max} denotes the maximum evolution time in years. The run time given in seconds (s) represent the approximate length of simulation time for one sample by Badlands.

Furthermore, the problems (Cr and CM) used in SC-Bayeslands took the computational time of a few seconds, and hence, it was possible to take thousands of samples for convergence. SC-Bayeslands is impractical for landscape evolution problems that have a computational time of several minutes to a few hours; hence, parallel computing is needed.

2.3. Parallel Tempering MCMC

PT-MCMC features an ensemble of replicas (samplers) that are executed in parallel with the ability to explore multimodal posterior distributions (Geyer & Thompson, 1995). Parallel tempering carries out an exchange of parameters in neighboring replicas during sampling that helps escape local minima. In other words, the Markov chains in the ensemble of replicas have stationary distributions, which are equal to (up to a proportionality constant) $p(\theta|\mathbf{D})^\beta$, where $\beta \in [0,1]$ and is known as the *temperature* of the replica. A replica that has a temperature of $\beta=0$ has a stationary distribution that is uniform, while one that has a temperature $\beta=1$ corresponds to a stationary distribution, which is the posterior $p(\theta|\mathbf{D})$. The replicas with smaller values of β are less likely to get stuck in local minima and thus explore a larger region. The replicas with higher values of β typically explore local regions. The choice of the temperature ladder and number of replicas governs the computational efficiency of the PT-MCMC and is the subject of much research. For instance, Kone and Kofke (2005) and Patriksson and van der Spoel (2008) describe an efficient method of finding the temperature ladder, and Miasojedow et al. (2013) proposed a method for attentively adapting the temperature ladder.

The other feature of parallel tempering is their feasibility of implementation in multicore or parallel computing architectures. In multicore implementation, factors such as interprocess communications need to be considered during the exchange between the neighboring replicas (Lamport, 1986). Effective communication strategies between the processes that execute the respective replicas need to be made to reduce the overhead

of processes waiting for others. Hence, a decentralized implementation of parallel tempering has been presented that eliminates global synchronization and reduces the overhead caused by interprocess communication in exchange of solutions between the chains that run in parallel cores (Li et al., 2009). Moreover, parallel tempering has been implemented in a distributed volunteer computing network where computers belonging to the general public were used with the help of multithreading and graphic processing units (Karimi et al., 2011). The implementation with *field programmable gate arrays* that has massive parallelism capabilities resulted in much better performance than multicore and graphic processing unit implementations (Mingas et al., 2017). In terms of applications, other studies have efficiently implemented parallel tempering via multicore architectures for exploration of Earth's resources (Reid et al., 2013).

3. Methodology

3.1. Synthetic Topography Data

Badlands is a stratigraphic forward model that takes at time $t=0$, an initial topography denoted by $\mathbf{D}_0 = (D_{s_10}, \dots, D_{s_n0})$, where D_{s_i0} is the elevation at site s_i at time $t=0$, for $i=1, \dots, n$, given n years.

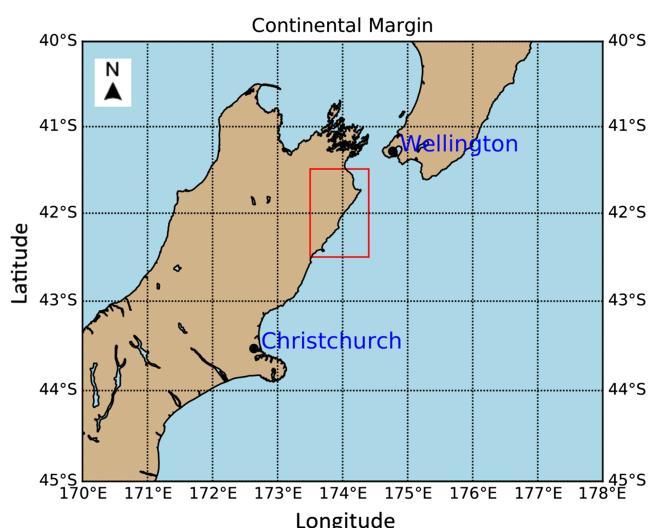


Figure 2. Continental margin problem selected from South Island of New Zealand, outlined by the red rectangle.

The parameters in the Badlands model are denoted by θ that are used to produce a series of topographies, $\mathbf{D}_0, \mathbf{D}_1, \dots, \mathbf{D}_T$. We assume that the final topography \mathbf{D}_T is the only topography to compare with the ground-truth topography. Therefore, the task of making inference about landscape evolution over time is challenging.

Different landscape evolution trajectories could lead to the same final topography, and to constrain the number of possible trajectories requires additional sources of information. One source of information is the history of sediment erosion/deposition at various locations. We denote the sediment erosion/deposition at time t by $\mathbf{z}_t = (z_{s_1,t}, \dots, z_{s_J,t})$, where $z_{s_j,t}$ is the sediment erosion/deposition at site s_j for $j=1, \dots, J$. We define another function that maps \mathbf{D}_0 and θ to sediment erosion/deposition by $\mathbf{g}_t(\mathbf{D}_0, \theta) = (g_{s_1,t}(\mathbf{D}_0, \theta), \dots, g_{s_m,t}(\mathbf{D}_0, \theta))$, for $t=1, \dots, T$.

In order to test the proposed methodology, we first use Badlands to create synthetic ground-truth topography data for two synthetic and a real-world landscape evolution problem. We use the present-day topography as the initial model topography. We use Badlands to model landscape and basin evolution given values for selected parameters (Table 2) and a specified time T_{max} (Table 1) given by the number of years to simulate the ground-truth data for topography and sediment erosion/deposition. We refer to the selected problems as synthetic crater (Cr), synthetic mountain (Mt) and continental margin (CM). The first two problems have an initial synthetic topography (hence the name), while the third uses a real landscape, in the northeastern region of the South Island in New Zealand as its initial topography (Figure 2). In all the problems, the Badlands model takes selected parameter values together with the initial topographies to produce final topographies. The initial and final topographies for all three problems are shown in Figure 3. We note that Cr and CM have been adapted from examples of the Badlands model (Salles & Hardiman, 2016; Salles, 2016; Salles et al., 2018) and also used in our previous work (Chandra et al., 2019).

The four key input parameters in Cr, Mt, and CM problems include precipitation (ρ), erodibility coefficient (k), and the contribution of precipitation and slope on erodibility denoted by m value (μ) and n value (ν) coefficients, respectively. Therefore, in the Cr problem, $\theta_{Cr}=(\rho, k, \mu, \nu)$. The Mt problem starts with a much simpler initial topography but also features the impact of tectonic uplift rate (u), so that $\theta_{Mt}=(\rho, k, \mu, \nu, u)$. The μ coefficient is particularly crucial in that it describes the sensitivity of a tectonically active mountain belt to changes in precipitation or tectonic accretion. It also defines how incision rates change as the discharge becomes apparent (Gasparini & Brandon, 2011). The CM problem does not feature tectonic uplift; however, since the location covers coastal region, it includes additional parameters that model the impact of the surface (γ) and marine (β) coefficients, so that $\theta_{CM}=(\rho, k, \mu, \nu, \beta, \lambda)$. The relationship between these parameters and landscape evolution is given by equations (1)–(3). The function that maps the initial topography and these parameters to the final topography (time $t=T$) is denoted by $\mathbf{f}_T(\mathbf{D}_0, \theta) = (f_{T,s_1}(\mathbf{D}_0, \theta), \dots, f_{T,s_n}(\mathbf{D}_0, \theta))$, where f_{T,s_i} is the elevation at time T for site s_i , for $i=1, \dots, n$.

Table 2 contains descriptions of the problem such as the evolution time, the area of the landscape given by its width and length, the amount of time taken to run the Badlands model, and the *resolution* factor, which indicates the distance in kilometers (km) between two neighboring points (pts) along x or y axis of the topography grid. Table 2 give details of the initial conditions used in the simulations for the various problem. Each of these problems features erosion/deposition of sediment over time. Figure 4 shows the sediment erosion/deposition synthetic ground-truth data at the final stage of evolution (present day) for Mt and CM problems. Note that the positive values indicate deposition and the negative values indicate erosion. The yellow dots indicate the locations for the ground-truth data for sediment erosion/deposition history. Hence, the likelihood function given in the following subsection takes both the landscape topography and erosion-deposition ground truth into account.

3.2. Model and Priors for PT-Bayeslands

We assume that topography at time t , and location s_i ($D_{s_i,t}|\theta, \mathbf{D}_0$) has normal distribution with an expected value equal to the Badlands model, given θ and a variance of τ^2 so that

$$D_{s_i,t} = f_{s_i,t}(\theta) + e_{s_i,t} \text{ with } e_{s_i,t} \sim N(0, \tau^2) \quad (4)$$

for $t=0, 1, \dots, T$ and $i=1, \dots, n$; where n refers to number of observations and T is the maximum simulation time in years. Note that we assume that the observations are independent because the correlation structure of the

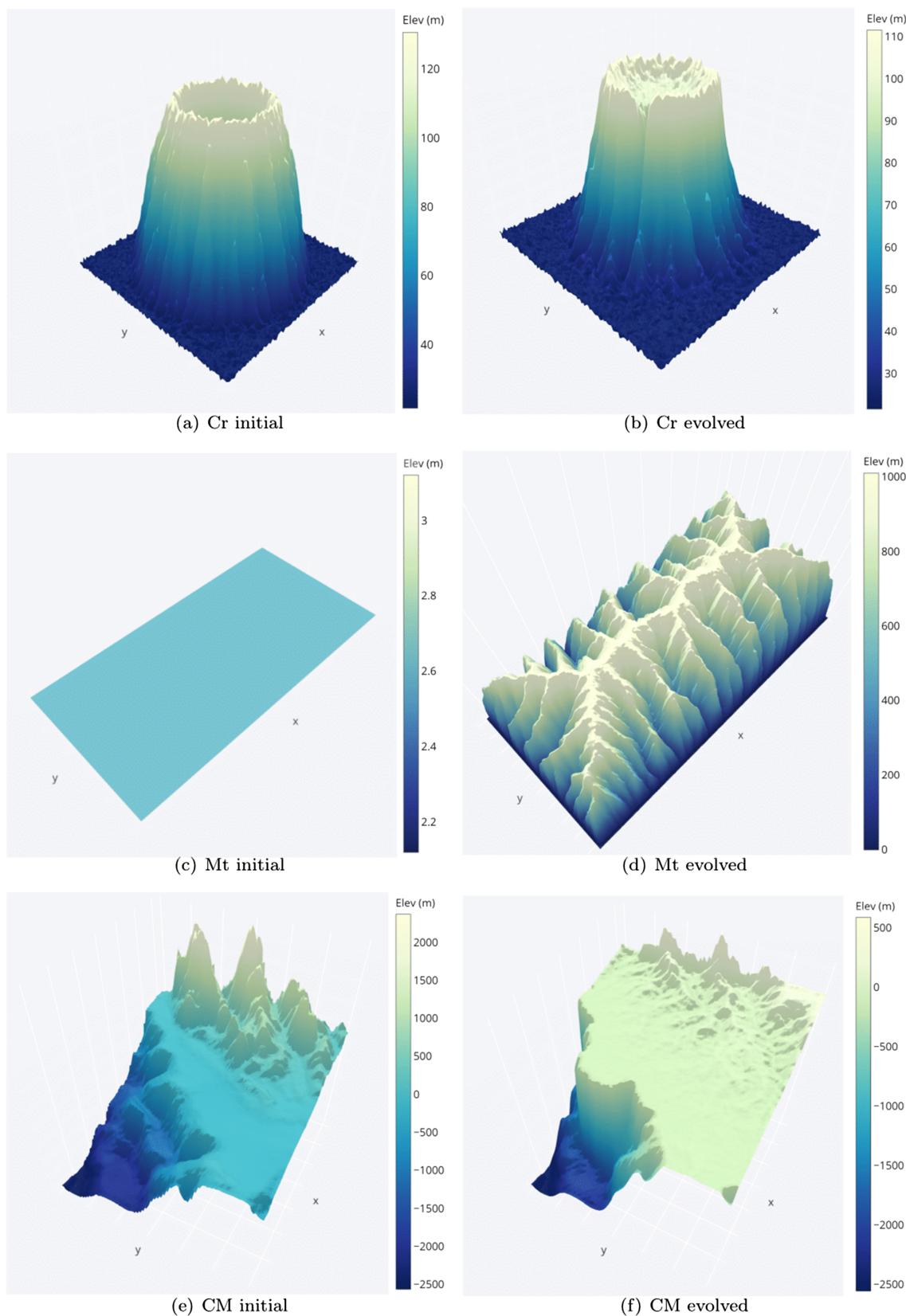


Figure 3. Initial (first column) and final or ground-truth (second column) topographies. Panels (a) and (b) refer to the synthetic-crater (Cr) landscape for $t = 0$ and $t = 50,000$ years, respectively. Panels (c) and (d) refer to the synthetic mountain (Mt) for $t = 0$ and $t = 1,000,000$ years, respectively, while Panels (e) and (f) refer to the continental margin (CM) landscape for $t = 0$ and $t = 1,000,000$ years, respectively.

Table 2
Parameter values for Synthetic Ground-Truth Topographies

Topography	ρ (m/year)	k	ν	μ	β	γ	u
Cr	1.5	5.0-e05	1.0	0.5	—	—	—
Mt	1.5	5.0-e06	1.0	0.5	—	—	1.0
CM	1.5	5.0-e06	1.0	0.5	0.5	0.8	—

Note. Cr = synthetic crater; Mt = synthetic mountain; CM = continental margin. The parameters include precipitation (ρ) in meters per year (m/year), erodibility (k), m value (μ), n value (ν), marine (β), surface (γ), and uplift (u) in millimeters per year (mm/year).

topography is embedded in the Badlands model. Although the choice of a normal distribution was made for convenience, we note that there could have been several other choices. For example, we could assume that the log of the topography has a normal distribution to allow for negative predicted values, which the model given by equation (4) is capable of producing. We could also assume that the errors are distributed according to a t distribution, or we could model the distribution of errors nonparametrically.

In the model given by equation (4), we assume an inverse gamma (IG) prior $\tau^2 \sim IG(\nu/2, 2/\nu)$. We integrate it so that the likelihood for the topography at time $t=T$ is $L_l(\theta)$ is

$$L_l(\theta) \propto \prod_{i=1}^n \left(1 + \frac{(D_{s_i,T} - f_{s_i,T}(\theta))^2}{\nu} \right)^{-\frac{\nu+1}{2}}, \quad (5)$$

where the subscript l in $L_l(\theta)$ denotes that it is the landscape likelihood in order to distinguish it from the sediment likelihood.

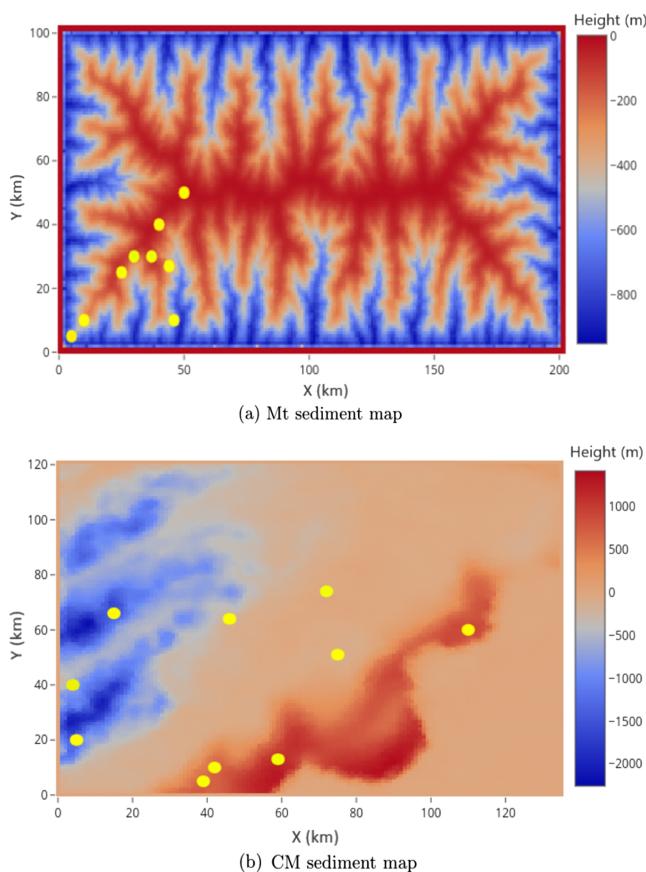


Figure 4. Sediment erosion/deposition present-day ground-truth data for synthetic mountain (Mt) and continental margin (CM) problems shown in Panel (a) and Panel (b), respectively. The selected locations for likelihood evaluation are overlaid as yellow dots. The color bar indicates elevation in meters (m), while the x and y axis give the distance in kilometers (km). Note that positive values indicate deposition and the negative values indicate erosion.

As noted in section 3.1, the history of sedimentary deposits is available at some locations, and we use this information to further constrain the number of possible landscape evolution trajectories. Again, we assume that observed sediment erosion/deposition values at time t , $z_{s,t}$, are a function of the Badlands model, given θ plus some Gaussian noise:

$$z_{s_j,t} = g_{s_j,t}(\theta) + \eta_{s_j,t} \text{ with } \eta_{s_j,t} \sim (0, \chi^2). \quad (6)$$

Analogous to the likelihood function for the topography (equation 5), the sediment likelihood $L_s(\theta)$, after integrating out χ^2 is

$$L_s(\theta) \propto \prod_{t=1}^T \prod_{j=1}^J \left(1 + \frac{(z_{s_j,t} - g_{s_j,t}(\theta))^2}{\nu} \right)^{-\frac{\nu+1}{2}}. \quad (7)$$

We assume that the elevation observations are independent of the sediment observations, so that

$$L(\theta) = L_s(\theta) \times L_l(\theta). \quad (8)$$

While this assumption may not hold for the simple Cr example, it is not an unreasonable assumption for the CM problem. The complicated topology of the CM problem, together with its coastal location, indicates that the sediment deposition is dispersed rather than stationary as in the Cr problem.

We use uniform priors for each of the parameters with the lower and upper limits given in Table 3, which reflects on the actual environmental and geological processes. For instance, logically, the values of precipitation and erodibility cannot be below zero. Trial experiments and the literature determine the upper limit (Salinger, 1980; Ummenhofer & England, 2007) to have a realistic assumption for the given problems. For instance, we reviewed the precipitation for South Island in New Zealand over the last few decades and gathered that the upper limit of 3 meters/year (m/a) would be a realistic assumption.

Table 3*Prior Distributions of Badlands Parameters*

Topography	ρ (m/year)	k	ν	μ	β	γ	λ
Cr	[0, 3.0]	[3.0e-05, 7.0e-05]	[0, 2.0]	[0, 2.0]	—	—	
Mt	[0, 3.0]	[3.0e-06, 7.0e-06]	[0, 2.0]	[0, 2.0]	—	—	[0.1, 5.0]
CM	[0, 3.0]	[3.0e-06, 7.0e-06]	[0, 2.0]	[0, 2.0]	[0.3, 0.7]	[0.6, 1.0]	—

Note. Cr = synthetic crater; Mt = synthetic mountain; CM = continental margin. Precipitation (ρ) in meters per year (m/yr), erodibility (k), m value (μ), n value (ν), marine (β), surface (γ), and uplift (λ) in millimeters per year (mm/year).

3.3. Sampling Scheme

3.3.1. Within-Replica Sampling: Proposal Distributions

PT-Bayeslands employs a randomwalk proposal distribution for each replica, denoted generically by $q(\theta|\theta^c)$, which is a multivariate normal distribution with mean vector μ and covariance matrix Σ . We set the mean vector μ to the current value in the chain θ^c and investigate the effect of two different covariance matrices; hereafter, known as the standard randomwalk (SRW) and adaptive randomwalk (ARW) proposal distribution.

In the SRW proposal distribution, Σ is fixed to be diagonal, so that $\Sigma = \text{diag}(\Sigma_1^2, \dots, \Sigma_P^2)$; where, Σ_j is the step size of the j th element of the parameter vector θ . The step size for parameter θ_j is chosen to be a combination of a fixed stepsize ϕ , which is common to all parameters, multiplied by the range of possible values for parameter θ_j , so that

$$\Sigma_j = (a_j - b_j) \times \phi, \quad (9)$$

where, a_j and b_j represent the maximum and minimum limits of the priors given in Table 1.

In the ARW proposal distribution, Σ is adapted every M intervals of within-replica sampling. Σ allows for the dependency between elements of θ and changes throughout the within-replica proposals sampler (Haario et al., 2001). The elements of Σ are adapted to the posterior using the sample covariance of the current chain history: $\Sigma = \text{cov}(\{\theta^{[0]}, \dots, \theta^{[i-1]}\}) + \text{diag}(\lambda_1^2, \dots, \lambda_P^2)$; where, $\theta^{[i]}$ is the i th iterate of θ in the chain and λ_j is the minimum stepsize.

Given the current value θ^c , the proposed value of θ^p is accepted with probability α , which with uniform priors and symmetric proposal distribution, reduces to

$$\alpha = \min\left(1, \frac{L(\mathbf{D}_T, \mathbf{Z}|\theta^p)}{L(\mathbf{D}_T, \mathbf{Z}|\theta^c)}\right). \quad (10)$$

3.3.2. Between-Replica Sampling

The *replica transition* procedure considers the exchange of two neighboring replicas. Suppose that there are M replicas indexed by m , with corresponding stationary distributions $p_m(\theta|\mathbf{D}_T) = p(\theta|\mathbf{D}_T)^{\beta_m}$, for $m=1,\dots,M$, with $\beta_1=1$ and $\beta_M < \beta_{M-1} < \dots < \beta_1$, and the temperature ladder $\beta=(\beta_M, \dots, \beta_1)$. The total parameter space consists of the chain m and the parameters within those replicas θ_m , and then the pair (m, θ) are jointly proposed and accepted/rejected according to the MH algorithm. The stationary distribution of chain m is $p(\theta_m|\mathbf{D}_T, \mathbf{Z})^{\beta_m}$. Suppose chains m and $m+1$ are at iteration k , with parameter values $\theta_m^{[k]}$ and $\theta_{m+1}^{[k]}$, we denote as θ_m^c and θ_{m+1}^c as the current, where the superscript c represents the current value. We propose a swap between these chains so that $\theta_m^p = \theta_{m+1}^c$ and $\theta_{m+1}^p = \theta_m^c$, where the superscript p represents the proposed value. We accept the proposed values with probability

$$\alpha = \min\left\{1, \left[\frac{p(\theta_m^c|\mathbf{D}_T, \mathbf{Z})}{p(\theta_{m+1}^c|\mathbf{D}_T, \mathbf{Z})} \right]^{\beta_{m+1}-\beta_m}\right\}. \quad (11)$$

If accepted, $\theta_{m+1}^{[k+1]} = \theta_{m+1}^p$ and $\theta_m^{[k+1]} = \theta_m^p$; otherwise, the chain remains where it is $\theta_{m+1}^{[k+1]} = \theta_{m+1}^{[k]}$ and $\theta_m^{[k+1]} = \theta_m^{[k]}$. We assign each of the replicas a temperature ladder that relaxes the likelihood, which affects the MH acceptance probability. Essentially, the replicas with higher temperature values have more probability of within-replica proposal acceptance, which can help in escaping from suboptimal modes. Given that the

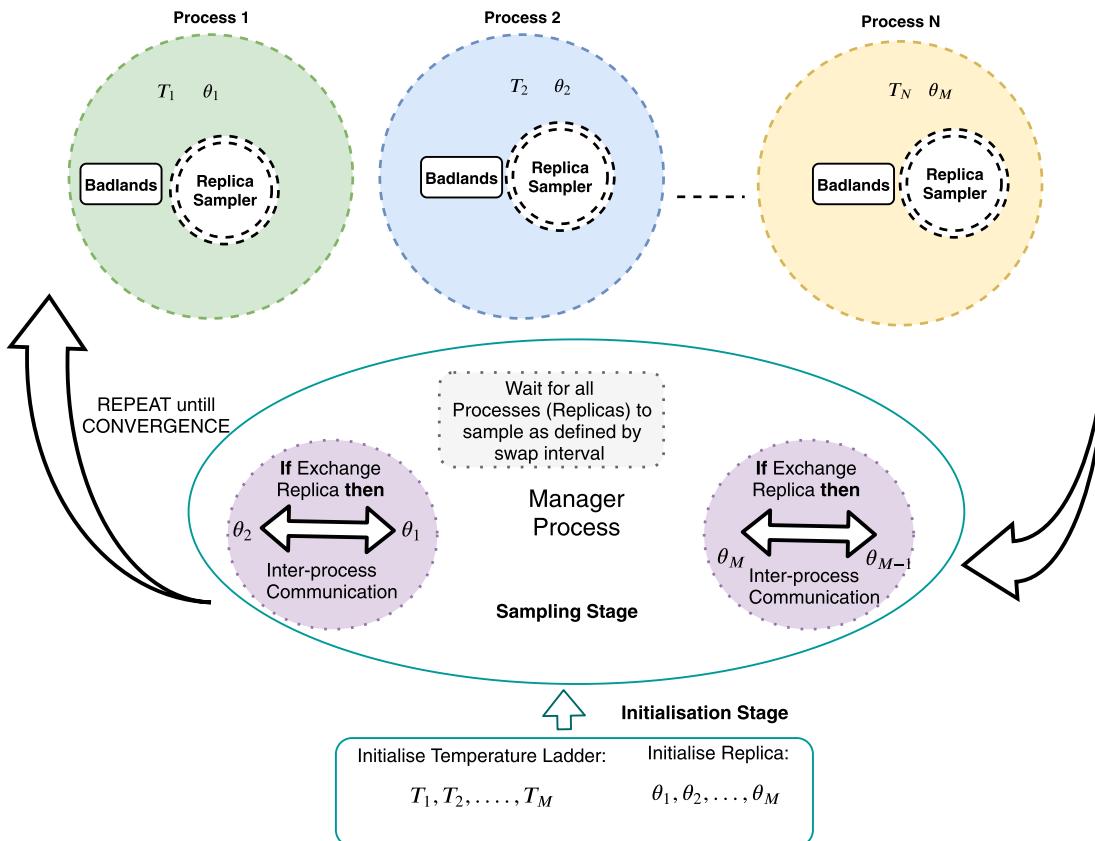


Figure 5. An overview of the replica ensemble executed on a multiprocessing architecture for parallel tempering Bayeslands. Note that the manager process controls the ensemble and enables exchange of the neighboring replicas.

temperature ladder is user defined, it is dependent upon the nature of the posterior. We use a geometric temperature ladder with

$$\beta_m = 1/\tau_m, \beta_M = 1/M \text{ and therefore, } \beta_m = \beta_{m-1} M^{-1/(M-1)} \quad (12)$$

for $m=2, \dots, M$, where M is the number of replicas and τ is the maximum temperature, which is user defined.

3.4. Multicore PT-Bayeslands

PT-Bayeslands features the Badlands model embedded in a PT-MCMC sampling scheme, both of which are computationally expensive, and thus, the combination of the two presents several challenges. Hence, we need an efficient sampling scheme that traverses the posterior effectively within a limited time. We note that in PT-MCMC, the replica with the temperature level of 1 only becomes part of the posterior. The rest of the replicas are used to enhance exploration with a mechanism to escape from local minima when needed. It is via the replica exchange that the configuration from the rest of the replicas becomes part of the posterior, provided that they move to replica with a temperature of 1.

In our implementation, we split the sampling scheme into two phases. The first phase uses parallel tempering with a temperature ladder defined by equation (12). The second phase sets the temperature ladder to be a vector of ones. Essentially, the first phase is used to ensure that the starting values of the second phase cover most of the parameter space. The second phase carries on with exchanges between adjacent chains to avoid them getting stuck in local modes. Figure 5 gives an overview of the different replicas that are executed on a multiprocessing architecture. The task of the *manager process* in Figure 5 is to manage the ensemble of replicas, where each replica runs on a separate processing core. Hence, given M replicas running in parallel, there are $M+1$ processes in total. The Badlands model is executed in the same processing core as the within-replica sampling, and interprocess communication is used to exchange neighboring replicas (Lamport, 1986). To

enhance computational efficiency, we minimize interprocess communication by only allowing replica exchange at user defined fixed intervals. The manager process waits for all replicas to complete sampling until the swap interval is reached in order to attempt replica exchange. Then, the manager process notifies the replicas to resume sampling with latest configurations in the chain for each replica.

Algorithm 1 outlines the sampling scheme that defines the number of replicas (M), the maximum number of iterations (Iter_{\max}), the replica swap or exchange interval (Swap_{int}), and the temperature ladder $\beta=(\beta_M, \dots, \beta_1)$ before drawing the initial values of θ_m for $m=1, \dots, M$ from the prior $p(\theta)$. Then, Algorithm 1 executes each of the replicas in parallel as shown in Figure 5. Each replica θ_m is updated when the respective proposal is accepted/rejected using the MH acceptance criterion given by Step 1.2, and the new value of θ is added to the posterior distribution. This procedure repeats until the replica reaches the swap interval (Swap_{int}), which determines how often the algorithm pauses and checks if neighboring replicas can be swapped using MH criterion (Step 2.2). The procedure repeats until the termination condition is satisfied as given by the maximum number of iterations.

Generally, it is expected that increasing the number of replicas in PT-Bayelands shortens the computation time; however, this is not necessarily true due to additional time taken with interprocess communication. The effect of increasing the number of replicas on the performance accuracy is investigated further in the following section.

Algorithm 1: Multicore Parallel Tempering MCMC

Result: Posterior distribution

- i. Set maximum number of iterations (Iter_{\max}), the swap interval (Swap_{int}), the number of replicas (M), and the temperature ladder $\beta = (\beta_M, \dots, \beta_1)$.
- ii. Initialise replica $\theta_m = \theta_{m^{[0]}}^{[0]}$, for $m = 1, \dots, M$.
- iii. Set the current values; $\theta_m^c = \theta_{m^{[0]}}^{[0]}$, and $m^c = m^{[0]}$

while Iter_{\max} **do**

for $m = 1, \dots, M$ **do**

for $k = 1, \dots, \text{Swap}_{\text{int}}$ **do**

Step 1: Replica sampling

1.1 Propose θ^p

1.2 Compute acceptance probability (equation (10))

Draw u from uniform distribution, $u \sim U[0, 1]$

if $u < \alpha$ **then**

| update, $\theta^{[k]} = \theta^p$

else

| retain, $\theta^{[k]} = \theta^c$

end

end

end

for $m = 1, \dots, M - 1$ **do**

Step 2: Replica Exchange

2.1 Select a pair of neighbouring replica, $[\theta_m, \theta_{m+1}]$

2.2 Compute acceptance probability (equation (11))

Draw u from uniform distribution, $u \sim U[0, 1]$

if $u < \alpha$ **then**

| exchange neighbouring replica, $\theta_m \leftrightarrow \theta_{m+1}$

else

end

end

end

3.5. Evaluation

We evaluate the performance of PT-Bayelands in three ways. The first is by comparing the posterior distribution of the parameters with those used to generate the ground-truth data. The second is by comparing the

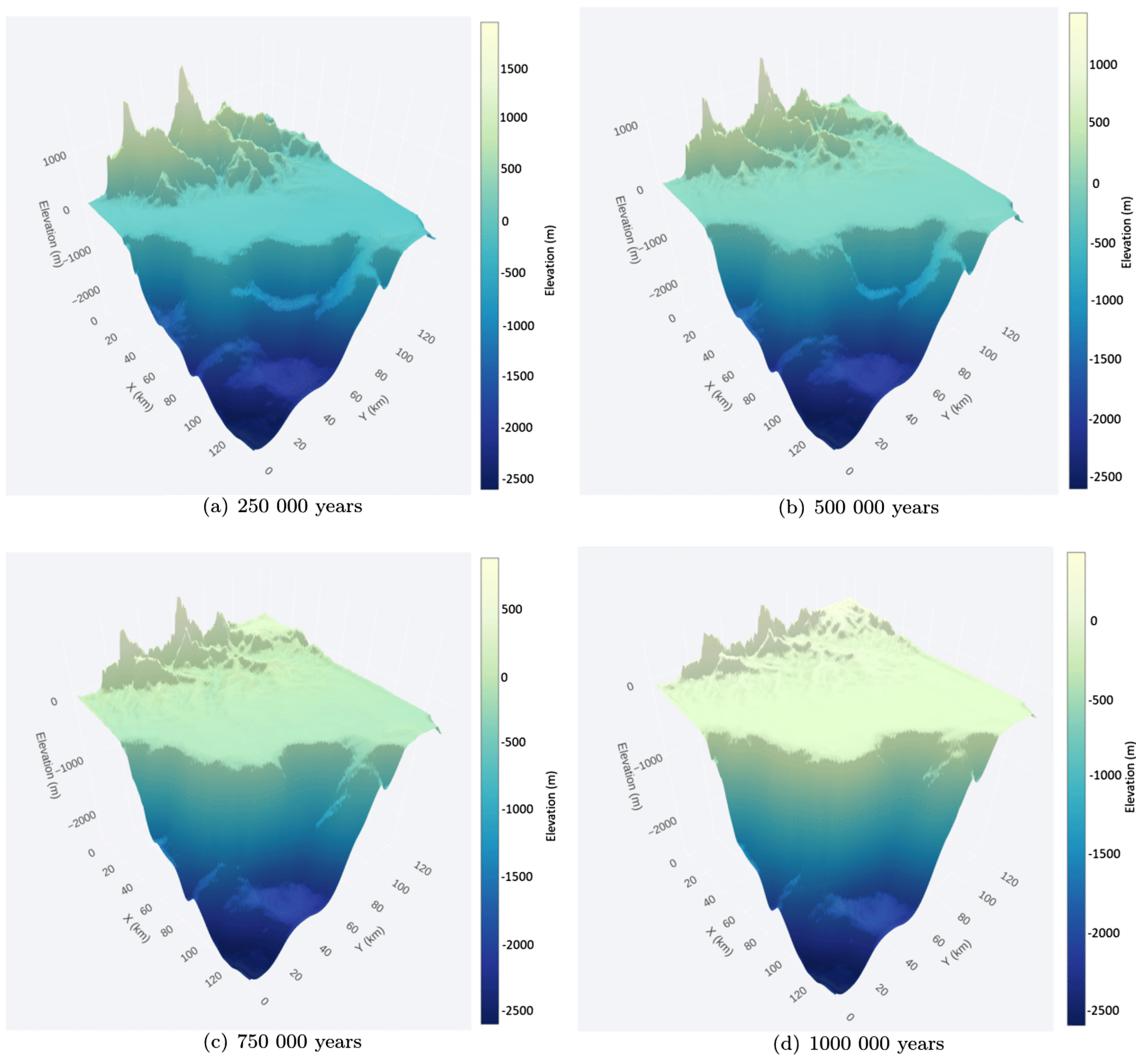


Figure 6. Badlands landscape evolution for four different time slices (250,000 years - Panel (a); 500,000 years - Panel (b); 750,000 years - Panel (c); 1000,000 years - Panel (d)) for the continental margin problem showing the elevation topography for 10,000 iterates with 10 replicas, using simple random walk as a proposal distribution from results shown in Table 8.

computational time for the PT-Bayeslands versus SC-Bayeslands from earlier work (Chandra et al., 2019). The third is by comparing the predicted/simulated Badlands landscape with the ground-truth using the root-mean-square error (RMSE). The RMSE for the elevation (elev) and sediment erosion/deposition (sed) is computed at each iteration of the sampling scheme and are given by

$$\text{RMSE}_{\text{elev}} = \sqrt{\frac{1}{n \times m} \sum_{i=1}^n \sum_{j=1}^m (g(\hat{\theta}_{T,i,j}) - g_{T,i,j}(\theta))^2},$$
$$\text{RMSE}_{\text{sed}} = \sqrt{\frac{1}{n_i \times v} \sum_{t=1}^{n_i} \sum_{j=1}^m (f(\hat{\theta}_{t,j}) - f(\theta_{t,j}))^2};$$

where, $\hat{\theta}$ is an estimated value chosen according to the proposal distribution and θ is the true value on which the ground truth topographies and sediment thickness were based. $f(\cdot)$ and $g(\cdot)$ represent the outputs of the Badlands model, as defined earlier, while m and n represent the size of the selected topography. v is the total number of selected points from sediment erosion/deposition as shown in Figure 4 over the selected time

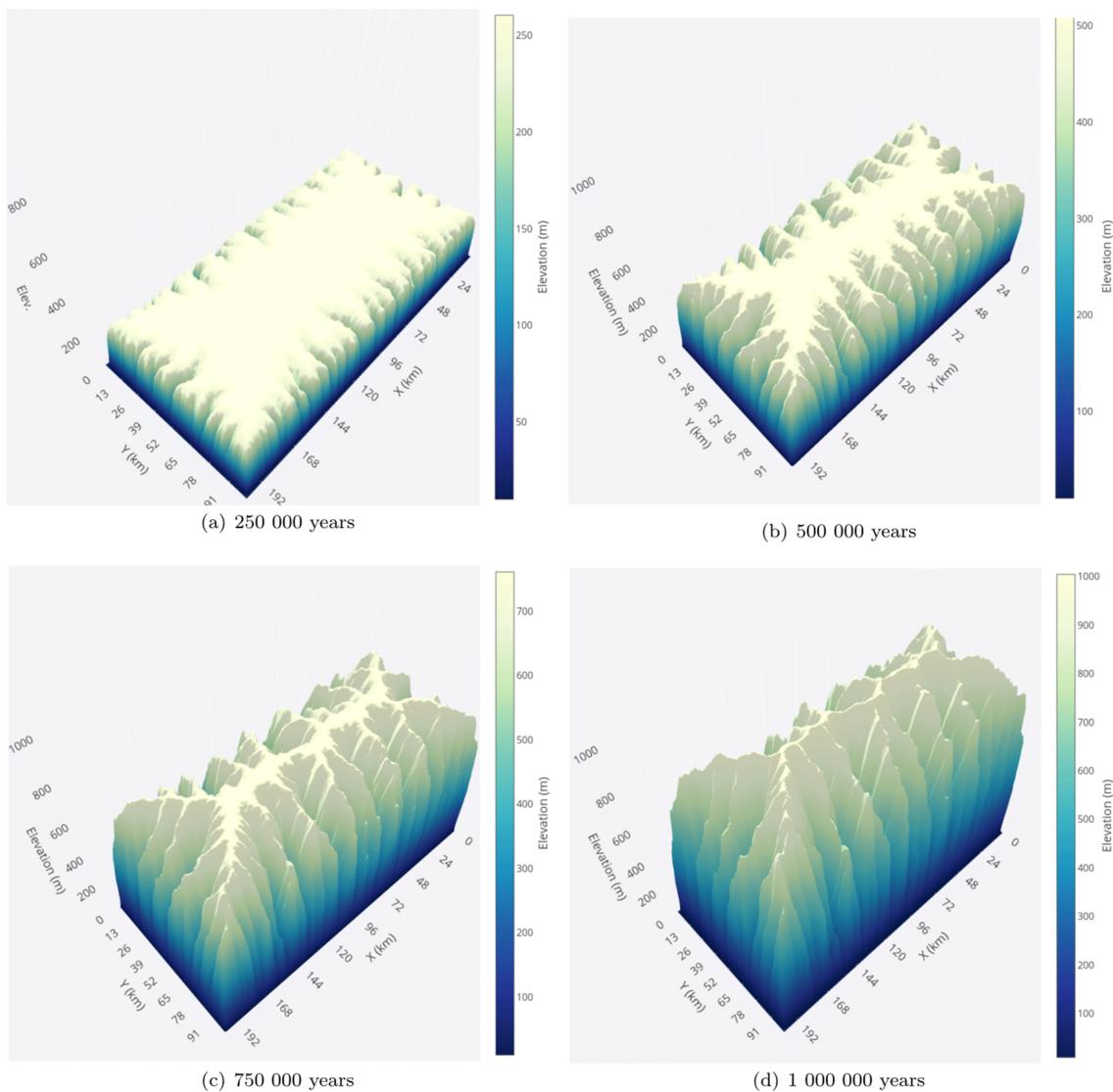


Figure 7. Badlands landscape evolution for four different time slices (250,000 years - Panel (a); 500,000 years - Panel (b); 750,000 years - Panel (c); 1000,000 years - Panel (d)) for the synthetic mountain problem showing the elevation topography for 10,000 iterates with 10 replicas, using simple random walk as a proposal distribution from results shown in Table 8.

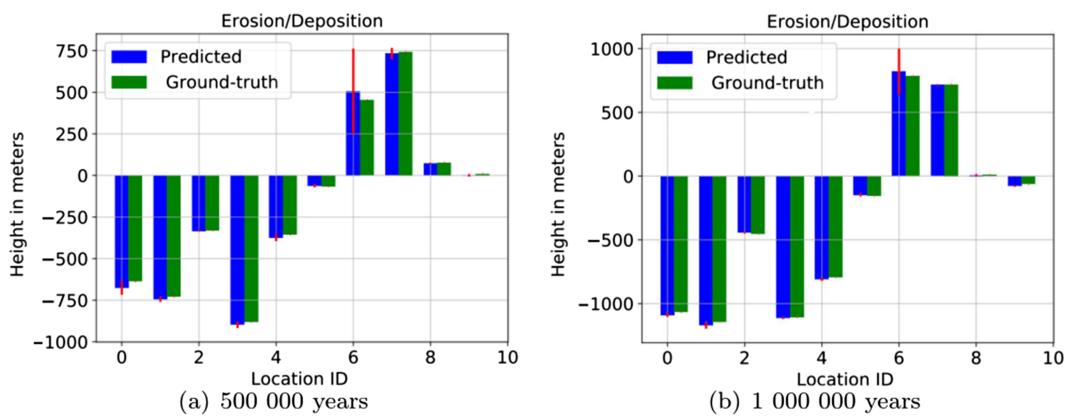


Figure 8. Badlands landscape evolution for two selected time slices (500,000 years - Panel (a); 1000,000 years - Panel (b)) corresponding to Figure 6 for the continental margin problem showing the sedimentary deposition for 10 selected points. The red lines show the uncertainty in prediction given by 95 % credible interval.

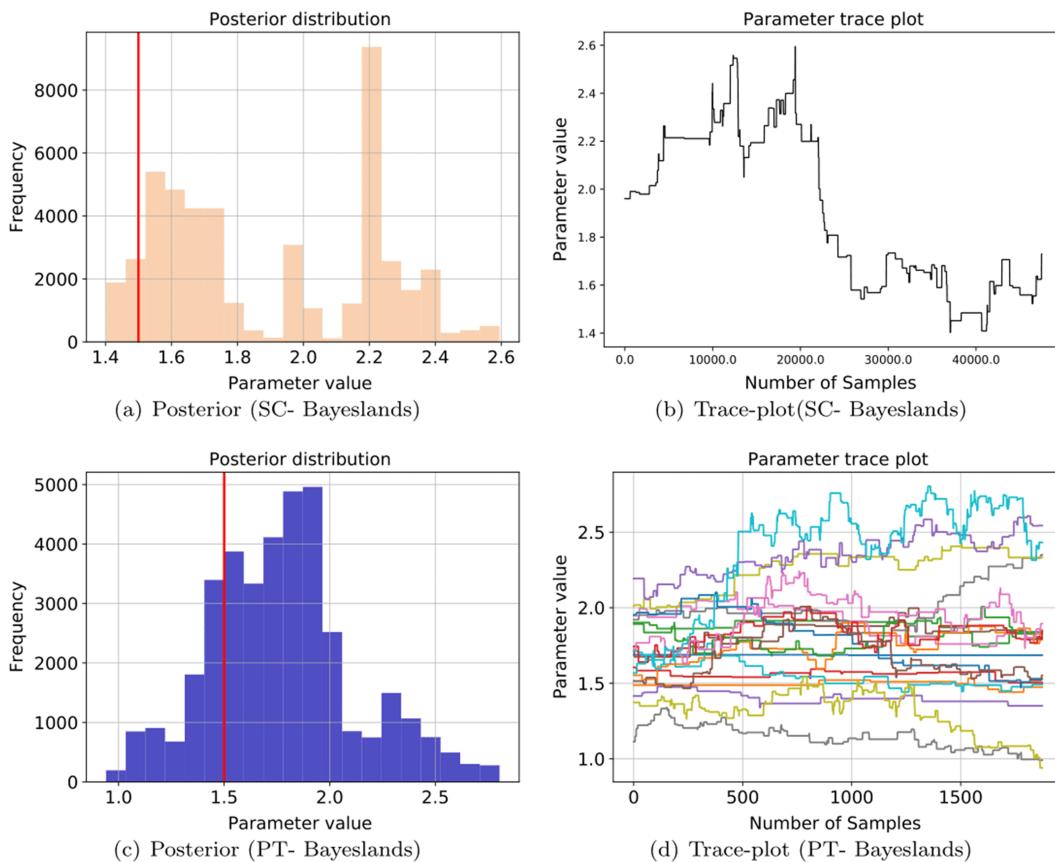


Figure 9. Comparison of parallel tempering (PT)-Bayeslands with single-chain (SC)-Bayeslands for the synthetic crater problem for 50,000 iterations. The panels show estimates of the posterior distribution and trace plot of the precipitation (ρ) for the SC-Bayeslands (Panel (a) and Panel (b)) and PT-Bayeslands (Panel (c) and Panel (d)), respectively. The trace plots show the accepted or current values of the chain for the given samples. Note that the results for SC-Bayeslands are taken from Chandra et al. (2019) with 5% burn-in, while PT-Bayeslands features 25% burn-in.

frame, n_t . Apart from the three application problems (Cr, Mt, and CM), we also test the performance of PT-Bayeslands to the different user settings, such as exchange rate Swap_{int} , and the maximum number of iterations Iter_{max} .

4. Results

4.1. Estimated Topography Elevation and Sediment Erosion/Deposition

We show the estimates of the posterior mean elevation and the sedimentary erosion/depositions produced by PT-Bayeslands for all three landscape problems and compare these with the ground truth. For the respective problems, the number of iterations was fixed at 10,000 with 10 replicas, and SRW was used as a proposal distribution. PT-Bayeslands produces posterior means of the elevation and sedimentary deposits each year, but due to limited space, we show these figures for selected time slices.

Figures 6 and 7 show the elevation of the CM and Mt problems at four selected time slices, respectively. Figure 8 shows the sedimentary erosion/deposition at the same two selected time slices. We select 10 points (Figure 4) and show how well PT-Bayeslands approximates the sedimentary erosion/disposition when compared to the ground truth at selected time slices (Figure 8).

4.2. Performance Comparison for Simpler Problems

To further investigate the issue of multimodality, we compare PT-Bayeslands with SC-Bayeslands for the respective problems in different settings. In the first setting, we allow only one parameter at a time to

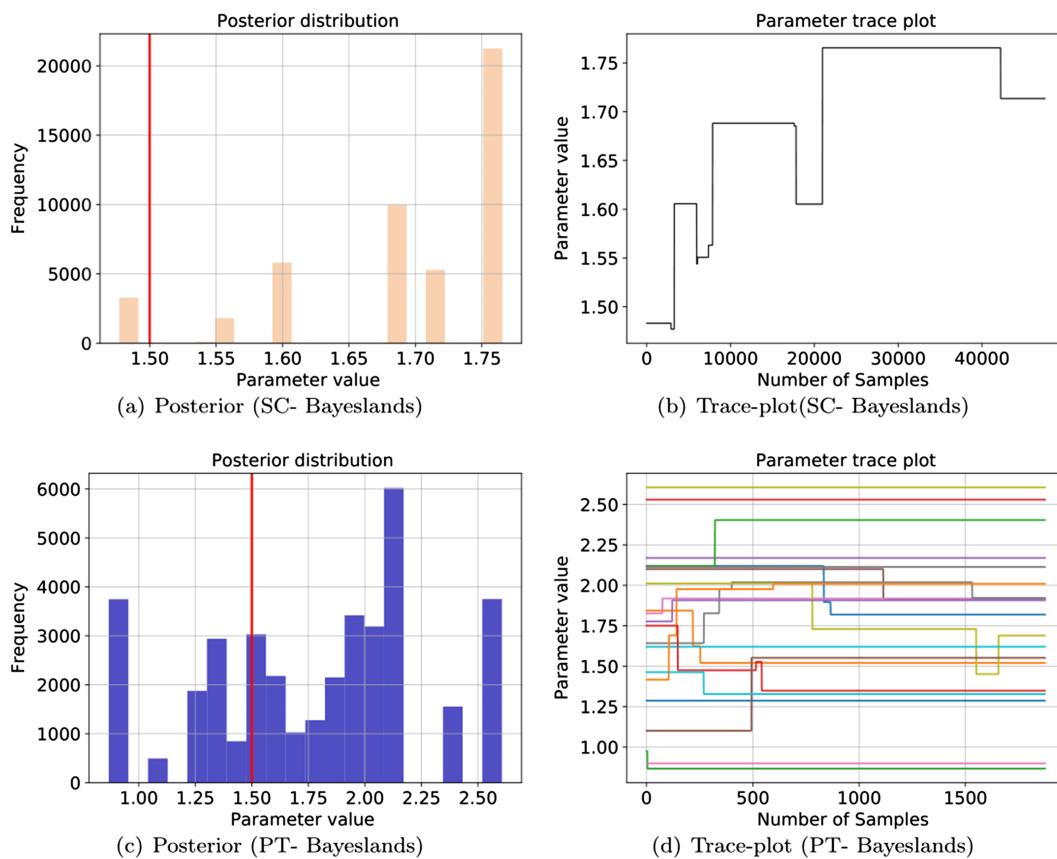


Figure 10. Comparison of parallel tempering (PT) Bayeslands with single-chain (SC) Bayeslands for the continental margin problem for 50,000 iterations. The panels show estimates of the posterior distribution and trace plot of the precipitation (ρ) for the SC-Bayeslands (Panel (a) and Panel (b)) and PT-Bayeslands (Panel (c) and Panel (d)), respectively. The trace plots show the accepted or current values of the chain for the given samples. Note that the results for SC-Bayeslands are taken from Chandra et al. (2019) with 5% burn-in, while PT-Bayeslands features 25% burn-in.

Table 4
Comparison of SC-Bayeslands with PT-Bayeslands

Problem	Method	Parameters allowed to vary	Time (min)	$RMSE_{elev}$ pos. mean	$RMSE_{sed}$ pos. mean
Cr	SC-Bayeslands	ρ	355.92	1.04	0.18
	SC-Bayeslands	k	356.18	1.05	0.18
	SC-Bayeslands	ρ, k	355.39	1.05	0.17
	PT-Bayeslands	ρ	44.05	1.05	0.18
	PT-Bayeslands	k	42.31	1.05	0.18
	PT-Bayeslands	ρ, k	39.91	1.05	0.18
	SC-Bayeslands	ρ	418.00	52.40	34.10
	SC-Bayeslands	k	423.70	51.50	43.30
	SC-Bayeslands	ρ, k	415.80	44.00	48.90
	PT-Bayeslands	ρ	59.40	58.30	48.10
CM	PT-Bayeslands	k	64.70	59.60	46.90
	PT-Bayeslands	ρ, k	60.60	60.30	50.60

Note. SC = single chain; PT = parallel tempering; CM = continental margin; Cr = synthetic crater; RMSE = root-mean-square error. The time taken for SC-Bayeslands and PT-Bayeslands is reported along with the accuracy given as posterior mean (pos. mean) of $RMSE_{elev}$ and $RMSE_{sed}$.

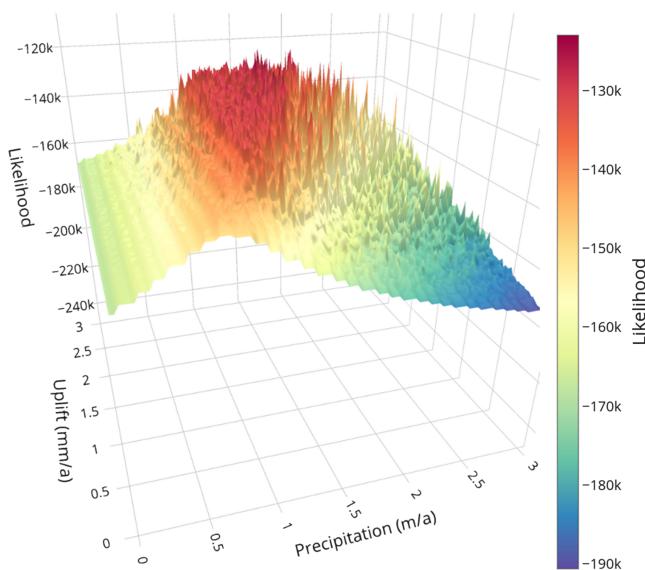


Figure 11. Logposterior surface of the Mountain problem as function of precipitation (ρ), and erodibility (k).

vary, while fixing the rest to their true value. The parameters that successively vary are precipitation (ρ) and erodibility (k) for the Cr and CM problem, while the rest are fixed to true values shown in Table 1. In the second setting, ρ and k vary simultaneously while fixing the remaining parameters.

Figure 9 displays results for the Cr problem in the second setting. Figures 9a and 9c show histogram estimates of the posterior distribution of ρ , when both ρ and k are allowed to vary. Figure 9a displays the estimated histogram using SC-Bayeslands, while Figure 9c displays the estimated histogram using PT-Bayeslands, and Figures 9b and 9d are the corresponding trace plots. These plots show how PT-Bayeslands vastly improves the exploration of the parameter space. SC-Bayeslands explores one mode for the first 20,000 iterations and then appears to shift to another mode for the remaining iterations.

Figure 10 displays results for the CM problem, which shows a similar trend. Although PT-Bayeslands performs better than SC-Bayeslands (Figure 10), the mixing between replicas is weak due to a large number of local modes in the log-posterior surface shown in Figure 1, which earlier highlighted the challenges of sampling. Figure 1 shows that the number of modes is much higher than the number of replicas, and a possible solution to this is to add more replicas, but then the sampler is not computationally feasible.

Table 4 shows the results from the experiments for the different combination of parameters for the two methods. Note that the total prediction accuracy is given for model elevation topography and sedimentary thickness in terms of accuracy presented as the posterior mean of $\text{RMSE}_{\text{elev}}$ and RMSE_{sed} . The sediment RMSE considers the ground truth, and final sediment distribution averaged over selected points in the model domain (see Figure 4). The results show that both methods achieve equally consistent prediction accuracy. However, PT-Bayeslands significantly reduces the computational time taken while also having better exploration features, as shown in Figure 9.

4.3. Multimodality

Geophysical inversion problems often exhibit posterior distributions, which are multimodal. Figure 11 shows the log-posterior (likelihood) surface for precipitation and uplift while fixing all other parameters in the Mt problem. We again note that our previous work (Chandra et al., 2019) shows that different combinations of precipitation and erodibility gave rise to visually indistinguishable topography for both the Cr and CM problem. Figure 11 shows that the posterior surface exhibits so many discontinuities that resemble CM posterior surface in Figure 1. SC-Bayeslands has difficulty to uncover or efficiently sample the surface structure. In this case, even PT-Bayeslands faces difficulties; hence, carefully crafted proposal distribution and temperature ladder is required to efficiently explore the space.

4.4. Effect of Simulation Setting

In this section, we evaluate the effect of the number of replicas, the number of iterations, and the swap rate using the CM problem given six free parameters as shown in Table 3. This problem was selected for these experiments because it has a difficult log-posterior surface with a single optimal mode and many suboptimal modes. Table 5 provides results for each of the settings, with a total number of 10,000 iterations. For example, the first setting has 10,000 iterations spread across four replicas with 2,500 iterations for each replica. The last setting has 48 replicas, and therefore only 208 iterations in each replica. We observe that

Table 5
Effect of Increasing the Number Replicas for the CM Topography With 10,000 Total Iterations

Replicas (cores)	Time (min)	$\text{RMSE}_{\text{elev}}$ (pos. mean)	RMSE_{sed} (pos. mean)
4	153.2	58.6	43.3
8	80.5	56.9	48.7
10	63.4	60.4	49.1
12	60.0	56.7	47.9
16	43.4	67.2	52.3
20	35.3	67.6	52.8
24	30.2	68.6	51.2
30	26.1	69.0	51.5
36	22.7	71.7	53.7
42	20.1	71.1	53.0
48	19.1	75.7	52.9

Note. CM = continental margin; RMSE = root-mean-square error. The prediction accuracy is given as posterior mean (pos. mean) of $\text{RMSE}_{\text{elev}}$ and RMSE_{sed} .

Table 6
Effect of the Number of Iterations for the CM Problem With 10 Replicas

Iterations	Time (min)	RMSE _{elev} (pos. mean)	RMSE _{sed} (pos. mean)
1,000	6.5	80.6	64.1
5,000	29.4	67.5	49.4
10,000	54.1	61.5	52.8
20,000	130.5	57.3	51.2
50,000	279.3	49.0	45.7
100,000	518.3	48.6	43.2

Note. CM = continental margin; RMSE = root-mean-square error. The prediction accuracy is given as posterior mean (pos. mean) of elevation RMSE_{elev} and sediment RMSE_{sed}.

increasing the number of replicas reduces the overall computation time. However, the number of replicas does not appear to have much effect on the elevation and sediment erosion/deposition prediction accuracy as measured by the RMSE_{elev} and RMSE_{sed}.

We then evaluate the effect that an increase in the number of iterations has on the prediction accuracy of PT-Bayeslands, for a given number of replicas. Table 6 presents summary of the results for the CM topography with 10 replicas. Increasing the number of iterations increases the the computational time taken in a linear fashion. The relationship between the number of iterations and the prediction accuracy is nonlinear given diminishing improvements to prediction accuracy as the number of iterations is increased. The optimal setting depends upon the trade-off between computational speed and prediction accuracy. We find that the interval 20,000 to 50,000 iterations are reasonable.

Table 7 contains the results of an experiment to investigate the effect of the replica-exchange rate (swap rate) on the prediction accuracy and the percentage of swaps accepted for the CM problem. Here, we used 10,000 iterations with 10 replicas. The results show that the swap rate has a negligible effect on the prediction accuracy.

4.5. Effect of the Proposal Distribution

In this section, we allow the number of iterations and the number of replicas to jointly vary, as well as allowing the proposal distribution (SRW and ARW). We show the effect of the different settings on the prediction accuracy for all three problems shown in Table 8. A significant observation in Table 8 is that the proposal distribution does not affect the prediction accuracy.

The Gelman-Rubin diagnostic evaluates MCMC convergence by analyzing the behavior of multiple Markov chains. Given multiple chains from different experimental runs, we provide an assessment by comparing the estimated between-chains and within-chain variances for each parameter, where large differences between the variances indicate nonconvergence. We provide convergence diagnosis for the two different proposal distributions (SRW and ARW) by running multiple chains using different initial values for 10,000 iterations for each topography problem. We calculate the *potential scale reduction factor* for each parameter, which provides the ratio of the parameter's sampling variance across all chains to its variance within any single chain. Table 9 shows convergence results for the two different proposal distributions as measured by the Gelman-Rubin diagnostic (Gelman & Rubin, 1992). The values for the potential scale reduction factor near 1 indicates convergence. Table 9 shows that the AWR proposals, in general, provide much better convergence when compared to SRW proposals. The only case where ARW is not near 1 is for two parameters (ρ and k).

Figure 12 show the posterior mean and associated uncertainty of a crosssection through the Y axis, with the X axis held constant at its middle value for both the SRW proposal and the ARW proposal for the respective problems. The figure shows that the estimated cross-sectional elevation is visually identical for both types of proposals, as expected theoretically.

We present further details for selected parameters for the given problems. We show details of the precipitation posterior for Cr and CM problems. We then show the Mt model uplift posterior as this is a unique parameter not present in the other problems. Figures 13 and 14 show the posterior distribution and trace plot for selected parameters for the respective problems. Note that the most extensive parameter space exploration by PT-Bayeslands occurs for the Cr problem. The exploration ability of the replicas is visible, given the trace

Table 7
Effect of the Swap Rate for the CM Problem With 10 Replicas (10,000 Iterations in Total)

Swap Rate	Time (min)	RMSE _{elev} (pos. mean)	RMSE _{sed} (pos. mean)	Swap %
0.01	88.4	69.9	51.2	22.2
0.02	94.6	72.8	58.6	23.0
0.03	100.2	76.7	58.2	21.8
0.04	91.5	66.1	46.5	22.4
0.05	90.1	63.9	45.2	22.9
0.06	94.5	74.5	51.9	23.4
0.07	90.1	69.7	50.7	22.0
0.08	92.5	71.1	51.7	22.2
0.09	88.5	67.5	44.3	25.7
0.10	95.2	69.8	58.2	22.8

Note. CM = continental margin; RMSE = root-mean-square error. Note that the posterior mean (pos. mean) for accuracy in terms of RMSE_{elev} and RMSE_{sed} are given in meters. The percentage swap (swap %) indicates the number of proposals successfully exchanged. The prediction accuracy is shown as posterior mean (pos. mean) of elevation RMSE_{elev} and sediment RMSE_{sed}.

Table 8

Typical Results in PT-Bayeslands Using SRW and ARW Proposal Distributions for the Three Problems (Cr, Mt, and CM)

Topography	Parameters	Proposals	Replica	T_{max}	$RMSE_{elev}$ (mean, std)	$RMSE_{sed}$ (mean, std)	Swap %
Cr	4	SRW	10	10,000	(1.05, 0.01)	(0.18, 0.01)	22.15
			20	50,000	(1.05, 0.01)	(0.18, 0.01)	13.09
	5	ARW	10	10,000	(1.05, 0.01)	(0.18, 0.01)	33.66
			20	50,000	(1.05, 0.01)	(0.18, 0.01)	26.33
Mt	5	SRW	10	10,000	(6.30, 3.11)	(0.87, 1.18)	22.31
			20	50,000	(5.48, 1.36)	(0.54, 0.51)	13.26
	6	ARW	10	10,000	(6.69, 2.55)	(0.69, 0.60)	22.30
			20	50,000	(5.44, 1.72)	(0.48, 0.37)	13.26
CM	6	SRW	10	10,000	(61.68, 17.31)	(45.74, 25.31)	23.62
			20	50,000	(61.86, 8.64)	(48.38, 17.97)	14.32
	7	ARW	10	50,000	(70.56, 8.38)	(50.80, 12.50)	31.87
			10	50,000	(63.02, 11.48)	(52.53, 20.81)	27.22

Note. SRW = simple random walk; ARW = adaptive random walk; CM = continental margin; Cr = synthetic crater; Mt = synthetic mountain; RMSE = root-mean-square error. T_{max} denotes the maximum number of iterations and the mean and standard deviation (std) of topography elevation ($RMSE_{elev}$) and sediment erosion/deposition ($RMSE_{sed}$) prediction accuracy is shown.

plot in Figure 13. This is in contrast with the exploration shown in trace plot given in the CM problem (Figure 14), since they have a highly irregular log-likelihood surface that has a number of suboptimal peaks. The trace plot shows that the replicas are essentially trapped in these suboptimal peaks; increasing the number of replicas (from 10 to 20) and sampling time (from 10,000 to 50,000) shows no major difference. However, even these suboptimal peaks give an acceptable prediction, as shown in Figure 12.

5. Discussion

The results presented in the previous section is summarized as follows:

1. For multimodal problems, PT-Bayeslands provides a better exploration of the posterior than SC-Bayeslands. This improvement in performance is particularly visible when the modes are not connected (as in the CM and Mt problem), and as the dimension of the problem increases. Despite the variability in results between SC-Bayeslands and PT-Bayeslands, the predictive performance for both the methods is similar. This suggests that SC-Bayeslands can converge in a suboptimal mode, even if it does not explore the multimodal posterior fully. These are significant results because multimodal posterior distributions are typically present in geological and geophysical inversion problems (Beskos et al., 2017; Dalla Mura et al., 2015).
2. Sampling multimodal distribution is a challenge; however, our results show that convergence on suboptimal modes gives similar topography evolution when compared with true values. The use of better proposal distributions can help in future work, such as using gradient-free meta-heuristics from the field of evolutionary algorithms (Drugan & Thierens, 2003; Ter Braak, 2006). Moreover, we also need to incorporate additional data, such as erosion rate obtained over different time intervals or estimates of sediment

Table 9

Convergence Diagnosis Showing the PSRF Score

Topography	Proposal	ρ	k	γ	λ	β	ϕ	u
Cr	SRW	4.71	5.19	2.63	11.49	—	—	—
	ARW	4.39	3.79	1.25	1.13	—	—	—
Mt	SRW	5.99	8.57	10.29	13.62	12.55	7.27	—
	ARW	1.78	1.77	1.17	1.33	1.12	1.24	—
CM	SRW	4.48	3.11	2.63	6.14	—	—	2.81
	ARW	1.73	1.89	1.16	1.34	—	—	1.21

Note. PSRF = potential scale reduction factor; SWR = simple random walk; ARW = adaptive random walk; CM = continental margin; Cr = synthetic crater; Mt = synthetic mountain. The respective parameters include; precipitation rate (ρ), coefficients for erodibility (k), m-value (γ), n-value (λ), marine (β), surface (ϕ), and uplift rate (u).

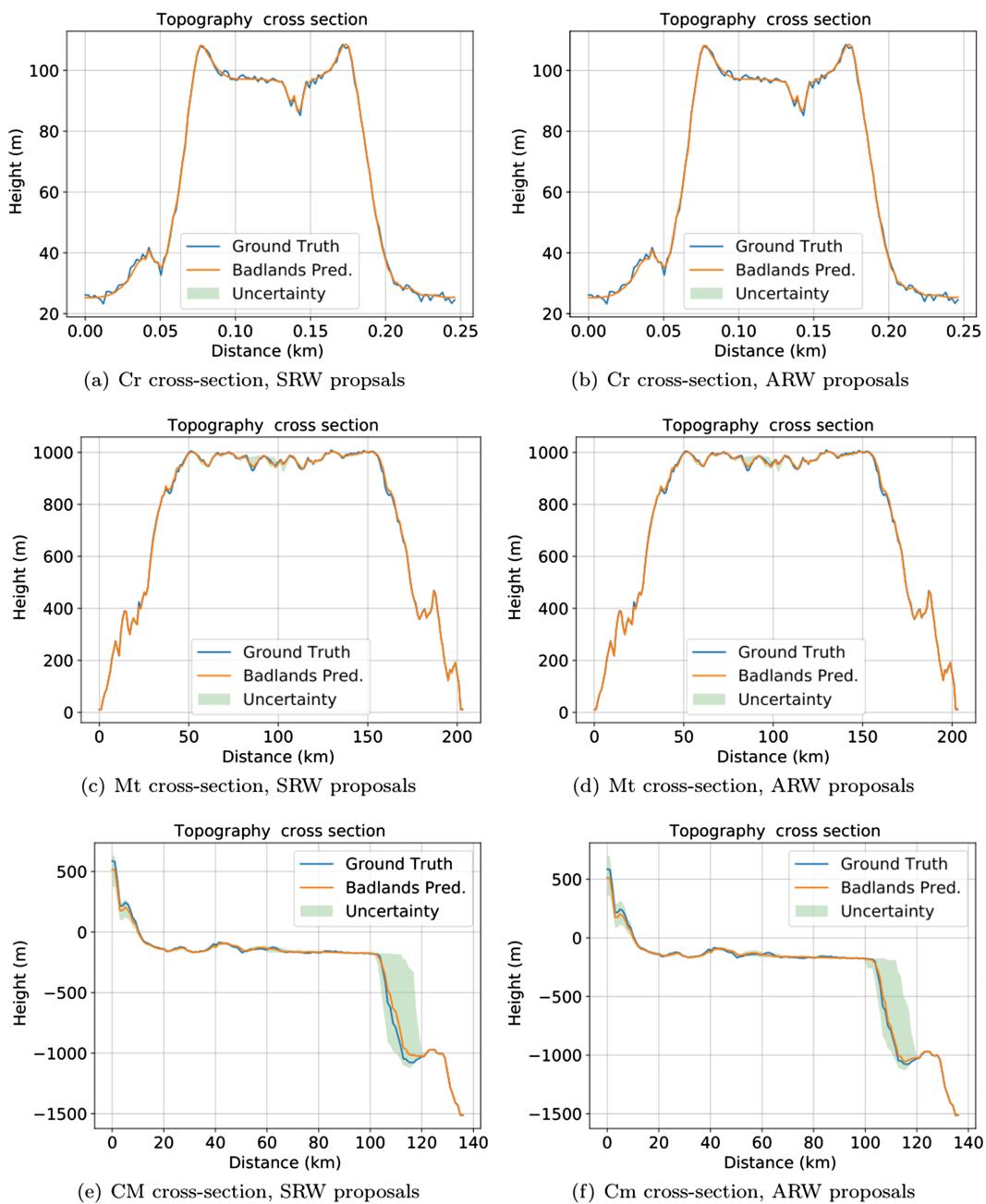


Figure 12. Crosssection comparing the synthetic ground-truth with PT-Bayeslands via Badlands predictions for the respective problems using simple randomwalk (SRW) and adaptive random walk (ARW) proposals. CM = continental margin (Panel (e) and Panel (f)); Cr = synthetic crater (Panel (a) and Panel (b)); Mt = synthetic mountain (Panel (c) and Panel (d)).

flux from surrounding sedimentary basins. The additional data can further improve the likelihood function and help to achieve better convergence.

3. The computation time for a given number of iterations does not scale linearly with the number of replicas. There is a trade-off between proposing to swap chains and the rate of convergence of the PT-Bayeslands. Swapping proposals between neighboring replicas give better mixing, but introduces computational overhead in a parallel computing environment since each replica needs to wait for all to complete sampling until their swap interval, in order to compute the neighbor swap probability by the managing process. This computational overhead can be higher if the number of replicas is large. We

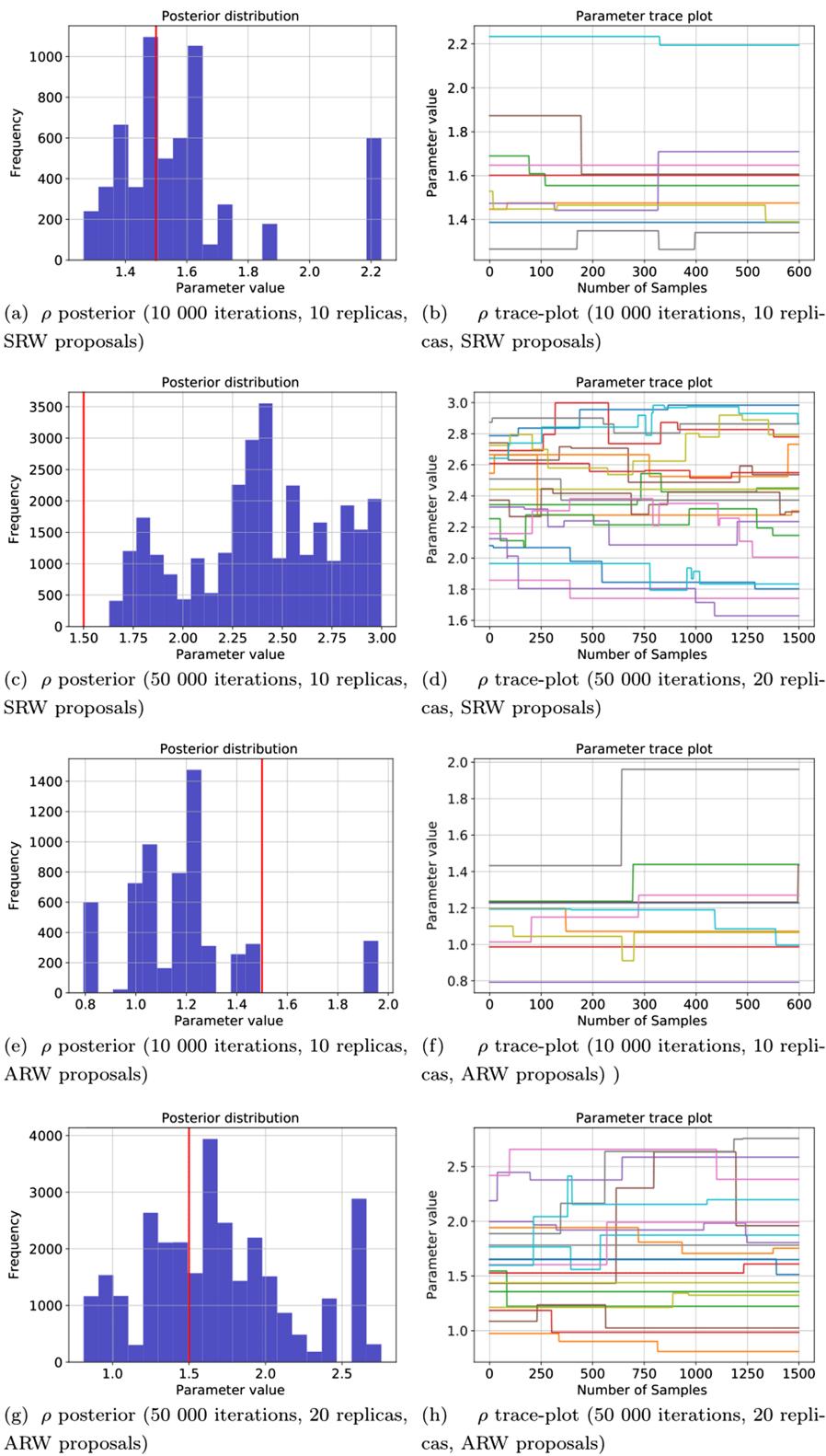


Figure 13. Posterior distribution and trace plot during sampling for precipitation (ρ) parameter for different sampling times for the synthetic crater problem using simple randomwalk (SRW) and adaptive randomwalk (ARW) proposal distributions in PT-Bayeslands. Panels (a) and (b) correspond to 10,000 iterations with a SRW proposal. Panels (c) and (d) correspond to 50,000 iterations with a SRW proposal. Panels (e) and (f) correspond to 10,000 iterations with an ARW proposal, and Panels (g) and (h) correspond to 50,000 iterations with an ARW proposal.

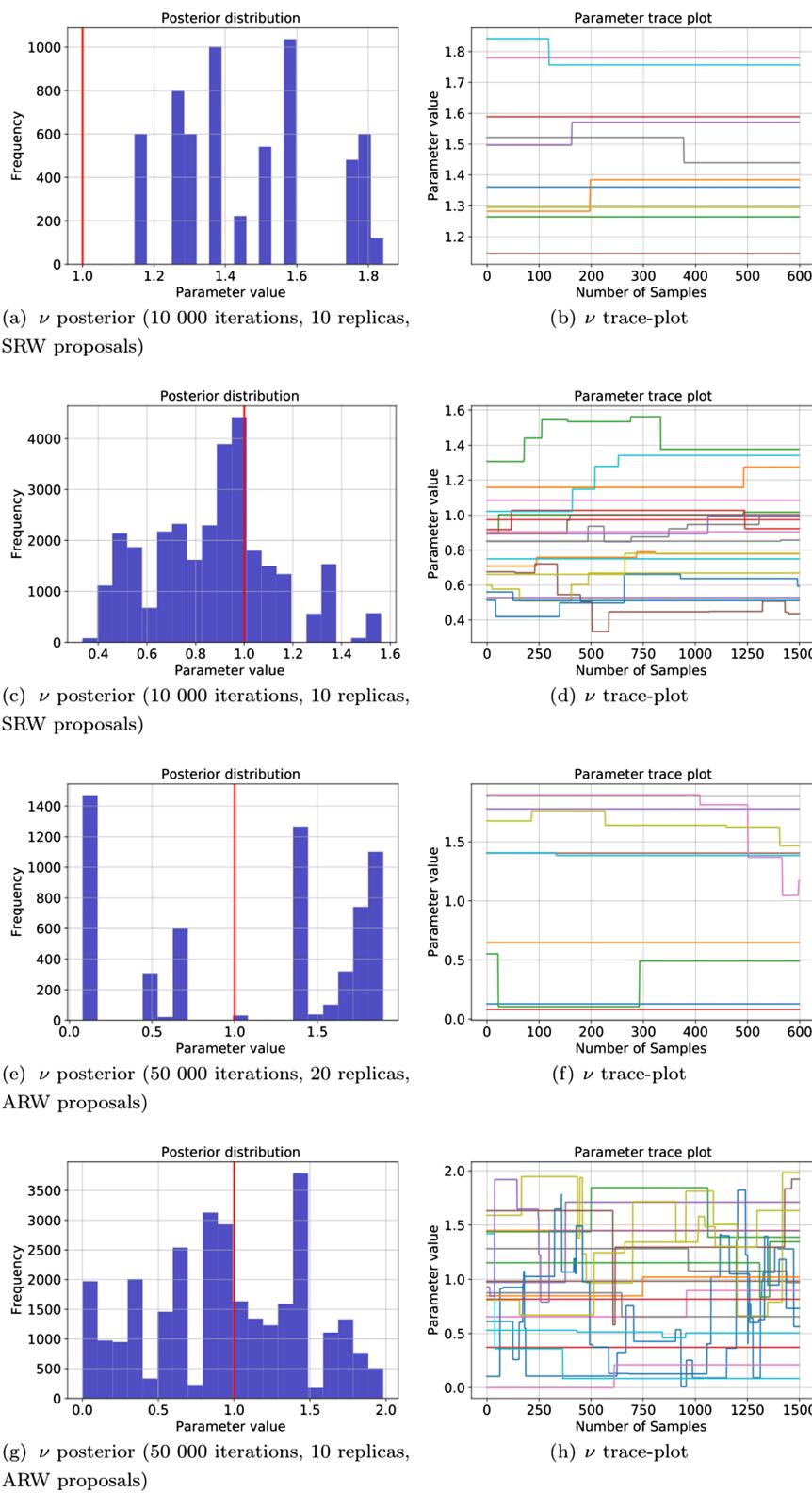


Figure 14. Posterior distribution and trace plot during sampling for n value (ν) parameter for different sampling times for the synthetic crater problem using simple randomwalk (SRW) and adaptive randomwalk (ARW) proposal distributions in PT-Bayeslands. Panels (a) and (b) correspond to 10,000 iterations with a SRW proposal. Panels (c) and (d) correspond to 50,000 iterations with a SRW proposal. Panels (e) and (f) correspond to 10,000 iterations with an ARW proposal, and Panels (g) and (h) correspond to 50,000 iterations with an ARW proposal.

note that the synthetic problems considered in this paper only required a few seconds of simulation time for Badlands, whereas in real-world problems, each simulation of Badlands could take several minutes to hours. In such cases, the trade-off between the frequency of swapping and prediction accuracy needs to be evaluated.

4. Increasing the number of cores does not necessarily mean that the prediction accuracy will get better. In particular, it is advantageous to use a larger number of cores for large-scale problems where the Badlands model takes hours to evaluate a single proposal.
5. The ARW and SRW proposal distributions have the same predictive accuracy; however, the ARW within chain proposal distribution explores the posterior better than the SRW proposal, which is evident by the convergence diagnosis (Table 9).
6. We highlight that a major limitation in the experiments is the assumption that initial topographies (paleo-elevation) is readily available. We simulated the initial topography and demonstrated the framework with small-scale synthetic problems. We need to note that initial topography is difficult to obtain given limited observations and, hence, it is a challenge to reconstruct initial topography. The framework can be used to fuse multiple sources of information about initial topography, which includes, expert knowledge, data, and models for paleo-elevation.
7. Multimodality arises given the combination of the different factors that have a similar effect on the topography development over time. Further work could also consider innovative ways to capture these processes in the likelihood function, such as paleo ground-truth data about ancient streams and rivers. Such data can help further constrain the sampling process and address problems that arise from multimodality.
8. In Badlands, similar to the model from Croissant and Braun (2014), we compute the erosion using stream power law that links the erosion rate to both the erodibility coefficient of soil, the drainage area, and the slope. In future work, the Bayeslands framework could consider other related models. However, changes to the LEM would feature additional parameters and could have a slightly different likelihood function to incorporate the streams and drainage rate that would affect sedimentary deposits.

6. Conclusions and Future Work

The PT-Bayeslands framework provides a rigorous approach for estimation and uncertainty quantification of free parameters in the Badlands model using multicore parallel tempering. We provide a comprehensive experimental evaluation of the method given different combinations of user settings that include sampling time, number of replicas, and replica swap rate for different types of problems. We also provided a comparison of two different proposal distributions with convergence diagnosis. In general, the results show that the method not only reduces the computation time but also provides a means to explore the parameter space in highly irregular multimodal distributions. Compared to previous work (SC-Bayeslands), the results show a better sampling of the parameters, along with better prediction accuracy of topography and sediment erosion/deposition. We further note that the proposed framework is general and, hence, apart from Badlands, it can accommodate other LEMs (Coulthard, 2001), such as the Landlab model (Hobley et al., 2017).

Future work can address several issues to overcome limitations. PT-Bayeslands needs to consider parameters such as precipitation, to vary across space and time. Such a model would allow us to measure the impact of climate change on the environment in terms of sediment erosion/deposition and elevation. Moreover, we can further address computational issues by surrogate-assisted models where a surrogate of Badlands would be implemented via machine learning, evaluating the proposals in a fraction of the time taken by Badlands. Such enhancements would benefit continental-scale LEMs over longer geological time (100 million years and longer). Furthermore, we can develop efficient gradient-free proposals as the number of parameters and the complexity of the model increases.

Data and Supplementary Material

In the results section, we only presented selected details of the posterior distribution, successive topography, and successive erosion-deposition predictions. Further results for the respective problems are provided as supplementary results, along with open source software (Python) and data for the test problems (at <https://doi.org/10.5281/zenodo.3368055>).

Acknowledgments

We acknowledge the Strategic Research Excellence Initiative (SREI 2017) Grant from the University of Sydney. We also acknowledge the Artemis high performance computing infrastructure provided by Sydney Informatics Hub of the University of Sydney. This work was also supported by Australian Research Council grant IH130200012. We sincerely thank Dr. Richard Scalzo for valuable discussions and Ashray Aman for technical support. Furthermore, we would like to sincerely thank the anonymous reviewers and the handling editor for helping us improve the paper.

References

- Adams, J. M., Gasparini, N. M., Hobley, D. E. J., Tucker, G. E., Hutton, E. W. H., Nudurupati, S. S., & Istanbulluoglu, E. (2017). The Landlab v1.0 OverlandFlow component: A Python tool for computing shallow-water flow across watersheds. *Geoscientific Model Development*, 10(4), 1645–1663.
- Beskos, A., Girolami, M., Lan, S., Farrell, P. E., & Stuart, A. M. (2017). Geometric MCMC for infinite-dimensional inverse problems. *Journal of Computational Physics*, 335, 327–351.
- Campforts, B., Schwanghart, W., & Govers, G. (2017). Accurate simulation of transient landscape evolution by eliminating numerical diffusion: The tlem 1.0 model. *Earth Surface Dynamics*, 5(1), 47–66.
- Chandra, R., Azam, D., Miller, R. D., Salles, T., & Cripps, S. (2019). Bayeslands: A Bayesian inference approach for parameter uncertainty quantification in badlands. *Computers & Geosciences*, 131, 89–101. <https://doi.org/10.1016/j.cageo.2019.06.012>
- Chandra, R., Jain, K., Deo, R. V., & Cripps, S. (2019). Langevin-gradient parallel tempering for Bayesian neural learning. *Neurocomputing*, 359, 315–326. <https://doi.org/10.1016/j.neucom.2019.05.082>
- Chen, A., Darbon, J., & Morel, J.-M. (2014). Landscape evolution models: A review of their fundamental equations. *Geomorphology*, 219, 68–86.
- Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327–335.
- Coulthard, T. J. (2001). Landscape evolution models: A software review. *Hydrological processes*, 15(1), 165–173.
- Croissant, T., & Braun, J. (2014). Constraining the stream power law: A novel approach combining a landscape evolution model and an inversion method. *Earth surface dynamics*, 2(1), 155–166.
- Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., & Benediktsson, J. A. (2015). Challenges and opportunities of multi-modality and data fusion in remote sensing. *Proceedings of the IEEE*, 103(9), 1585–1601.
- Drugan, M. M., & Thierens, D. (2003). Evolutionary Markov chain Monte Carlo, *International Conference on Artificial Evolution (Evolution Artificielle)* (pp. 63–76). Marseille, France: Springer.
- Fox, M., Bodin, T., & Shuster, D. L. (2015). Abrupt changes in the rate of Andean Plateau uplift from reversible jump Markov chain Monte Carlo inversion of river profiles. *Geomorphology*, 238, 1–14.
- Fox, M., Reverman, R., Herman, F., Fellin, M. G., Sternai, P., & Willett, S. D. (2014). Rock uplift and erosion rate history of the Bergell intrusion from the inversion of low temperature thermochronometric data. *Geochemistry, Geophysics, Geosystems*, 15, 1235–1257. <https://doi.org/10.1002/2013gc005224>
- Gallagher, K., Charvin, K., Nielsen, S., Cambridge, M., & Stephenson, J. (2009). Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for earth science problems. *Marine and Petroleum Geology*, 26(4), 525–535.
- Gasparini, N. M., & Brandon, M. T. (2011). A generalized power law approximation for fluvial incision of bedrock channels. *Journal of Geophysical Research*, 116, F02020. <https://doi.org/10.1029/2009jf001655>
- Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4), 457–472.
- Geyer, C. J., & Thompson, E. A. (1995). Annealing Markov chain Monte Carlo with applications to ancestral inference. *Journal of the American Statistical Association*, 90(431), 909–920.
- Girolami, M., & Calderhead, B. (2011). Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2), 123–214.
- Godard, V., Lavé, J., & Cattin, R. (2006). Numerical modelling of erosion processes in the Himalayas of Nepal: Effects of spatial variations of rock strength and precipitation. *Geological Society, London, Special Publications*, 253(1), 341–358.
- Goren, L., Fox, M., & Willett, S. D. (2014). Tectonics from fluvial topography using formal linear inversion: Theory and applications to the Inyo Mountains, California. *Journal of Geophysical Research: Earth Surface*, 119, 1651–1681. <https://doi.org/10.1002/2014jf003079>
- Grandis, H., Menivelle, M., & Roussignol, M. (1999). Bayesian inversion with Markov chains—I. The magnetotelluric one-dimensional case. *Geophysical Journal International*, 138(3), 757–768.
- Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2), 223–242.
- Harel, M.-A., Mudd, S., & Attal, M. (2016). Global analysis of the stream power law parameters based on worldwide 10be denudation rates. *Geomorphology*, 268, 184–196.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1), 97–109.
- Hobley, D. E., Adams, J. M., Nudurupati, S. S., Hutton, E. W., Gasparini, N. M., Istanbulluoglu, E., & Tucker, G. E. (2017). Creative computing with landlab: An open-source toolkit for building, coupling, and exploring two-dimensional numerical models of Earth-surface dynamics. *Earth Surface Dynamics*, 5(1), 21.
- Hobley, D. E. J., Sinclair, H. D., Mudd, S. M., & Cowie, P. A. (2011). Field calibration of sediment flux dependent river incision. *Journal of Geophysical Research*, 116, F04017. <https://doi.org/10.1029/2010jf001935>
- Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1), 1593–1623.
- Howard, A. D., Dietrich, W. E., & Seidl, M. A. (1994). Modeling fluvial erosion on regional to continental scales. *Journal of Geophysical Research*, 99, 13,971–13,986. <https://doi.org/10.1029/94JB00744>
- Karimi, K., Dickson, N., & Hamze, F. (2011). High-performance physics simulations using multi-core CPUs and GPGPUs in a volunteer computing context. *The International Journal of High Performance Computing Applications*, 25(1), 61–69.
- Kone, A., & Kofke, D. A. (2005). Selection of temperature intervals for parallel-tempering simulations. *The Journal of Chemical Physics*, 122(20), 206101.
- Lamport, L. (1986). On interprocess communication. *Distributed Computing*, 1(2), 86–101.
- Li, Y., Mascagni, M., & Gorin, A. (2009). A decentralized parallel implementation for parallel tempering algorithm. *Parallel Computing*, 35(5), 269–283.
- Malinverno, A. (2002). Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International*, 151(3), 675–688.
- Maraschini, M., & Foti, S. (2010). A Monte Carlo multimodal inversion of surface waves. *Geophysical Journal International*, 182(3), 1557–1566.
- Marinari, E., & Parisi, G. (1992). Simulated tempering: A new Monte Carlo scheme. *EPL (Europhysics Letters)*, 19(6), 451.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6), 1087–1092.

- Miasojedow, B., Moulines, E., & Vihola, M. (2013). An adaptive parallel tempering algorithm. *Journal of Computational and Graphical Statistics*, 22(3), 649–664.
- Mills, K., Fox, G., & Heimbach, R. (1992). Implementing an intervisibility analysis model on a parallel computing system. *Computers & Geosciences*, 18(8), 1047–1054.
- Mingas, G., Bottolo, L., & Bouganis, C.-S. (2017). Particle MCMC algorithms and architectures for accelerating inference in state-space models. *International Journal of Approximate Reasoning*, 83, 413–433.
- Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100, 12,431–12,447.
- Mosegaard, K., & Vestergaard, P. D. (1991). A simulated annealing approach to seismic model optimization with sparse prior information. *Geophysical Prospecting*, 39(5), 599–611.
- Neal, R. M. (1996). Sampling from multimodal distributions using tempered transitions. *Statistics and computing*, 6(4), 353–366.
- Neal, R. M. (2011). MCMC using Hamiltonian dynamics, *Handbook of Markov Chain Monte Carlo* (Vol. 2, pp. 113–162). NY: CRC Press New York.
- Patriksson, A., & van der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15), 2073–2077.
- Reid, A., Bonilla, E. V., McCalman, L., Rawling, T., & Ramos, F. (2013). Bayesian joint inversions for the exploration of Earth resources, *IJCAI* (pp. 2877–2884). Beijing, China: AAAI Press.
- Rejman, J., Turski, R., & Paluszak, J. (1998). Spatial and temporal variations in erodibility of loess soil. *Soil and Tillage Research*, 46(1-2), 61–68.
- Robert, C., & Casella, G. (2011). A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science*, 26, 102–115.
- Roberts, G. G., & White, N. (2010). Estimating uplift rate histories from river profiles using african examples. *Journal of Geophysical Research*, 115, B02406. <https://doi.org/10.1029/2009JB006692>
- Rocca, P., Benedetti, M., Donelli, M., Franceschini, D., & Massa, A. (2009). Evolutionary optimization as applied to inverse scattering problems. *Inverse Problems*, 25(12), 123003.
- Salinger, M. (1980). New zealand climate: I. Precipitation patterns. *Monthly Weather Review*, 108(11), 1892–1904.
- Salles, T. (2016). Badlands: A parallel basin and landscape dynamics model. *SoftwareX*, 5, 195–202.
- Salles, T., Ding, X., & Brocard, G. (2018). pyBadlands: A framework to simulate sediment transport, landscape dynamics and basin stratigraphic evolution through space and time. *PLoS one*, 13(4), e0195557.
- Salles, T., Ding, X., Webster, J. M., Vila-Concejo, A., Brocard, G., & Pall, J. (2018). A unified framework for modelling sediment fate from source to sink and its interactions with reef systems over geological times. *Scientific Reports*, 8(1), 5252. <https://doi.org/10.1038/s41598-018-23519-8>
- Salles, T., & Duclaux, G. (2015). Combined hillslope diffusion and sediment transport simulation applied to landscape dynamics modelling. *Earth Surface Processes and Landforms*, 40(6), 82339.
- Salles, T., Flament, N., & Müller, D. (2017). Influence of mantle flow on the drainage of eastern Australia since the jurassic period. *Geochemistry, Geophysics Geosystems*, 18, 280–305. <https://doi.org/10.1002/2016gc006617>
- Salles, T., & Hardiman, L. (2016). Badlands: An open-source, flexible and parallel framework to study landscape dynamics. *Computers & Geosciences*, 91, 77–89.
- Sambridge, M. (1999). Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble. *Geophysical Journal International*, 138(3), 727–746.
- Sambridge, M. (2013). A parallel tempering algorithm for probabilistic sampling and multimodal optimization. *Geophysical Journal International*, 196(1), 357–374.
- Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40(3), 1009. <https://doi.org/10.1029/2000rg000089>
- Scalzo, R., Kohn, D., Olieroock, H., Houseman, G., Chandra, R., Girolami, M., & Cripps, S. (2019). Efficiency and robustness in Monte Carlo sampling for 3-D geophysical inversions with obsidian v0.1.2: Setting up for success. *Geoscientific Model Development*, 12(7), 2941–2960. <https://doi.org/10.5194/gmd-12-2941-2019>
- Scott, L. R., & Zhang, S. (1990). Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Mathematics of Computation*, 54(190), 483–493.
- Sen, M. K., & Stoffa, P. L. (1996). Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophysical Prospecting*, 44(2), 313–350.
- Sen, M. K., & Stoffa, P. L. (2013). *Global Optimization Methods in Geophysical Inversion*: Cambridge University Press.
- Stock, J. D., & Montgomery, D. R. (1999). Geologic constraints on bedrock river incision using the stream power law. *Journal of Geophysical Research*, 104, 4983–4993. <https://doi.org/10.1029/98jb02139>
- Tarantola, A. (2006). Popper, Bayes and the inverse problem. *Nature physics*, 2(8), 492.
- Ter Braak, C. J. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3), 239–249.
- Tucker, G. E., & Hancock, G. R. (2010). Modelling landscape evolution. *Earth Surface Processes and Landforms*, 35(1), 28–50.
- Ummenhofer, C. C., & England, M. H. (2007). Interannual extremes in New Zealand precipitation linked to modes of southern hemisphere climate variability. *Journal of Climate*, 20(21), 5418–5440.
- Vrugt, J. A., Nualláin, B. O., Robinson, B. A., Bouten, W., Dekker, S. C., & Sloot, P. M. (2006). Application of parallel computing to stochastic parameter estimation in environmental models. *Computers & Geosciences*, 32(8), 1139–1155.
- Whipple, K. X., & Tucker, G. E. (2002). Implications of sediment-flux-dependent river incision models for landscape evolution. *Journal of Geophysical Research*, 107(B2), 2039. <https://doi.org/10.1029/2000jb000044>
- Zhang, H., Liu, M., Shi, Y., Yuen, D. A., Yan, Z., & Liang, G. (2007). Toward an automated parallel computing environment for geosciences. *Physics of the Earth and Planetary Interiors*, 163(1-4), 2–22.