Bayeslands: Framework featuring Bayesian inference for uncertainaity quanitification in basin and landscape evolution models (Badlands)

Bayeslands Workshop

Rohitash Chandra

EarthByte Group and Centre for Translational Data Science
The University of Sydney, Sydney, Australia

July 2, 2018

Introduction

- Models require methods that uncover the free parameters that best describes the data.
- Data in many cases is sparse, limited, or incomplete.
- The search for the free parameters in models: optimisation methods, eg. gradient based methods, simplex search, genetic or evolutionary algorithms, meta-heuristics [in case when no gradient information from model is available].

Introduction

- Challenges in optimisation methods given large scale of parameters - when gradient information is not available. The limitations are in terms of uncertainty quantification. The need to run multiple experiments with different initial set of parameters to check robustness for convergence. Limitations of p values for statistical tests.
- Bayesian inference methods - probability distributions instead of single point estimates. There is no need to run multiple given that the inference algorithm has converged to a distribution.

Bayesian inference

- Bayesian inference provides a principled approach towards uncertainty quantification of free parameters in geophysical forward models.
- The use of MCMC methods in the geosciences have been well established, with applications spanning from modelling geochronological ages, inferring sea-level and sediment supply from the stratigraphic record, and inferring groundwater contamination sources.

Bayesian inference

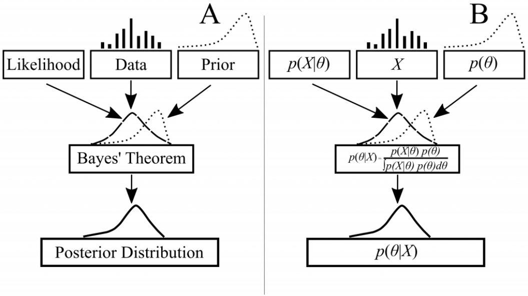
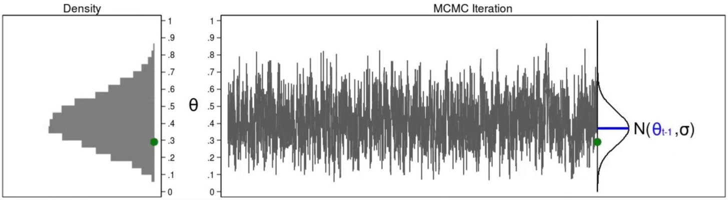


Figure 1: Bayesian inference overview

Markov Chain Monte Carlo sampling methods (MCMC) implement Bayesian inference that sample from a probability distribution. This is based on constructing a Markov chain after a number of steps that has the desired distribution as its equilibrium distribution.

MCMC framework



Step 1: $r(\theta_{new}, \theta_{t-1}) = \dfrac{\text{Posterior}(\theta_{new})}{\text{Posterior}(\theta_{t-1})} = \dfrac{\text{Beta}(1,1,0.290) \times \text{Binomial}(10,4,0.290)}{\text{Beta}(1,1,0.371) \times \text{Binomial}(10,4,0.371)} = 0.773$

Step 2: Acceptance probability $\alpha(\theta_{new}, \theta_{t-1}) = \min\{r(\theta_{new}, \theta_{t-1}), 1\} = \min\{0.773, 1\} = 0.773$

Step 3: Draw $u \sim \text{Uniform}(0,1) = 0.420$

Step 4: If $u < \alpha(\theta_{new}, \theta_{t-1}) \rightarrow$ If $0.420 < 0.773$ Then $\theta_t = \theta_{new} = 0.290$

Otherwise $\theta_t = \theta_{t-1} = 0.371$

Figure 2: MCMC sampling

Basin and landscape dynamics via Bayeslands

- Bayeslands is a framework for inference and uncertainty quantification in the Badlands model for basin and landscape evolution.
- Bayeslands extends *Badlands* by placing probability distributions over the free parameters such as rainfal and erobility - thereby turning a deterministic model into a probabilistic one.

Bayeslands framework

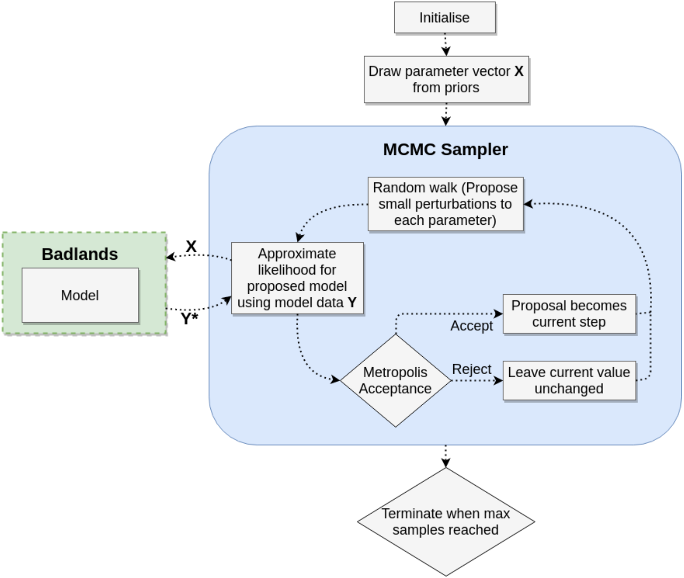
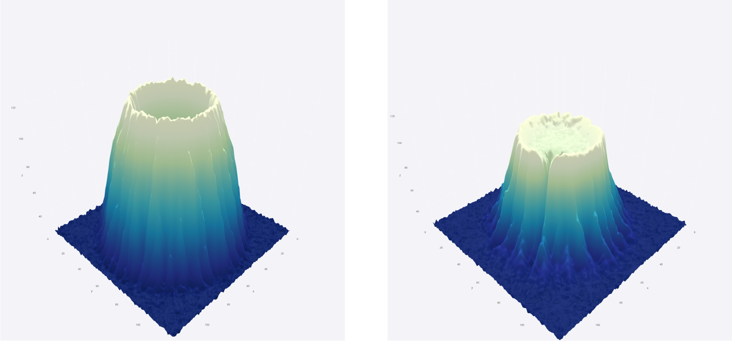


Figure 3: *Bayeslands* framework for Badlands.

Example 1: Crater

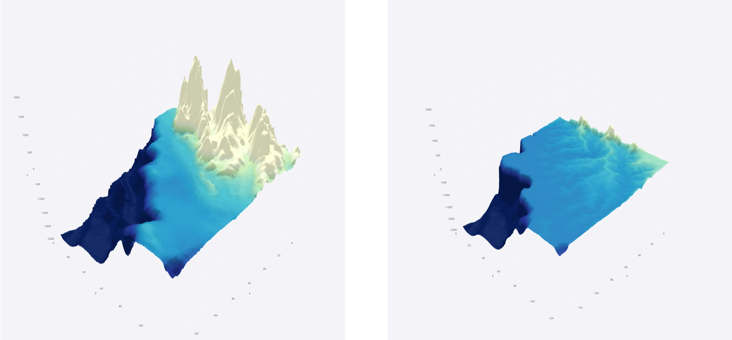


(a) Crater-extended initial topography

(b) Crater-extended synthetic ground-truth topography

Figure 4: Crater-extended: Initial and eroded ground-truth topography and sediment deposition after 50 000 years.

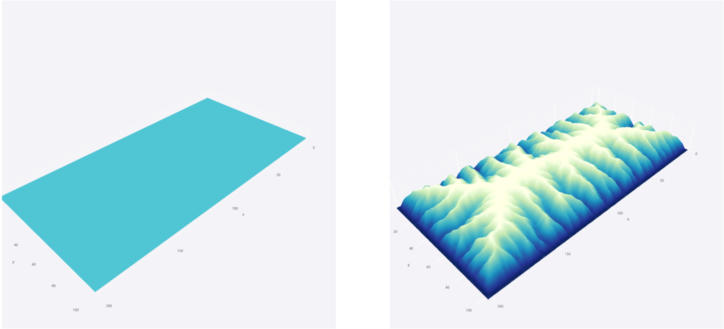
Example 1: Continental Margin



(a) CM-extended initial topography (b) CM-extended synthetic ground-truth topography

Figure 5: Continental Margin (CM)-extended: Initial and eroded ground-truth topography and sediment after 1 000 000 years.

Example 1: Mountain



(a) Mountain initial topography

(b) Mountain synthetic ground truth topography

Figure 6: Mountain: Initial and eroded ground-truth topography after 1 000 000 years evolution.

Parallel tempering

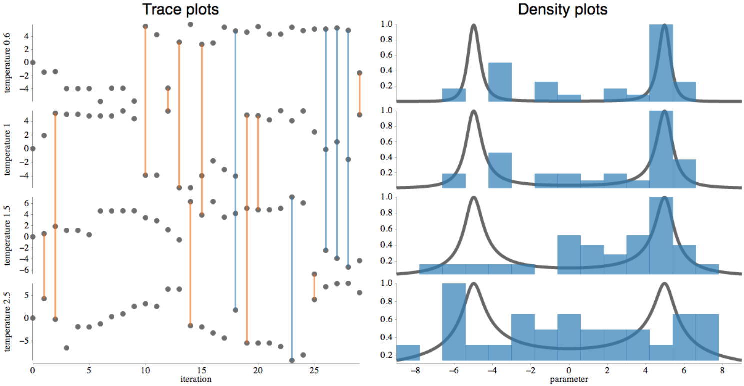


Figure 7: Parallel tempering - sampling

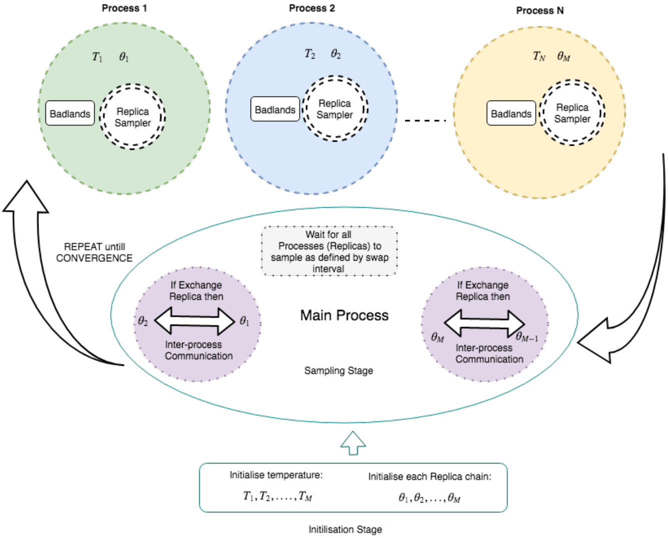


Figure 8: An overview of the different replicas that are executed on a multi-processing architecture. Note that the main process controls the given replicas and enables them to exchange the neighboring replicass given the swap time and probability of exchange is satisfied.

Model: Likelihood

$$L_l(\boldsymbol{\theta}) = \frac{1}{(2\pi\tau^2)^{n/2}} \exp\left\{ -\frac{1}{2} \frac{\sum_{t=1}^{T} \sum_{i=1}^{n} (D_{t,s_i} - f_{t,s_i}(\boldsymbol{\theta}))^2}{\tau^2} \right\}$$

where the subscript l, in $L_l(\theta)$, denotes that it is the landscape likelihood.

Model: Likelihood

The sediment likelihood, $L_s(\theta)$ is

$$L_s(\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{mT/2}} \exp\left\{ -\frac{1}{2} \sum_{t=1}^{T} \sum_{j=1}^{m} \frac{(Z_{t,s_j} - g_{t,s_j}(\boldsymbol{\theta}))^2}{\sigma^2} \right\}$$

giving a total likelihood as a function of theta $L(\boldsymbol{\theta})$ to be

$$L(\boldsymbol{\theta}) = L_s(\boldsymbol{\theta}) \times L_l(\boldsymbol{\theta}).$$

Design of Experiments

- Step 1: Investigate the effects on computational time and accuracy when increasing the number of replicas;
- Step 2: Evaluate the number of samples required for convergence defined in terms of prediction accuracy;
- Step 3: Using knowledge from above experiments, apply PT-Bayeslands to all the given problems and report the posterior distributions, computational time, uncertainty and accuracy in topography and sediment predictions.

Results

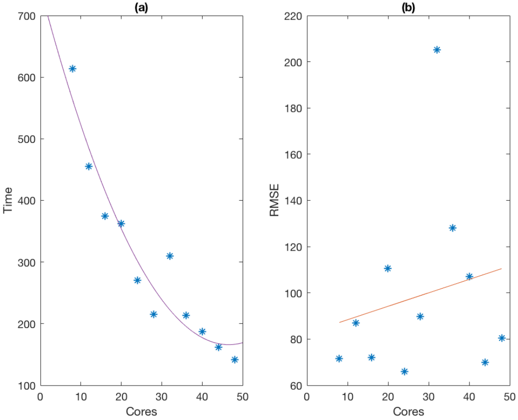


Figure 9: The effect of the number of cores on time, panel (a) and on the RMSE, panel (b). Note that a 95% confidence interval for the slope of the regression line in panel (b), is [-14.2,2.6], showing that the slope is not significantly different from zero.

Results

Replica (cores)	Time (minutes)	Pred. RMSE	Accepted %
2	3540	78.8	0.1
4	1384	62.7	0.1
8	614	71.5	0.4
12	455	87.0	0.5
16	375	72.0	0.6
20	362	110.6	0.3
24	270	65.9	0.4
28	215	89.7	0.5
32	310	205.1	0.4
36	213	128.0	0.6
40	187	107.0	0.6
44	162	70.0	0.6
48	142	80.5	0.8

Table 1: Effect of number replicas/cores for the CM-extended topography with 100 000 samples.

Results

Num. Samples	Time (mins.)	Pred. RMSE	Accepted %
1000	3	120.6	14.0
2000	6	222.5	15.3
4000	10	200.8	9.4
6000	18	98.3	3.7
8000	18	213.1	8.3
10000	27	77.3	2.8
50000	131	77.7	1.3
100000	258	95.1	1.0
150000	436	69.3	0.2

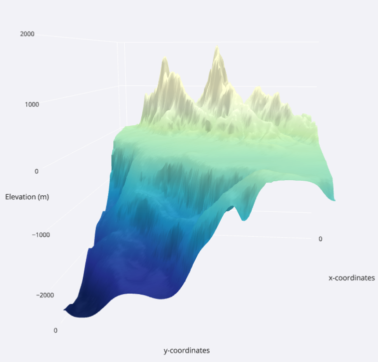
Table 2: Effect of the number of samples for the CM-extended topography running with 24 replicas on 24 cpus.

Results

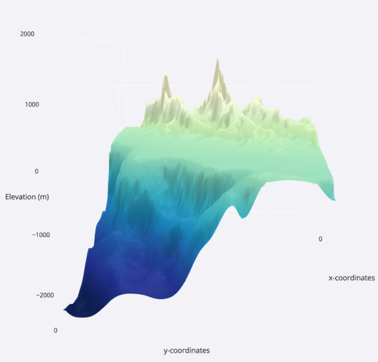
Topography	Time (min)	Sed. RMSE	Elev. RMSE	Pred. RMSE	Acc. %
Crater	80	4.2	1.1	5.3	12.2
Crater-ext	229	0.2	1.0	1.2	1.9
CM	50	2.6	19.9	22.5	1.0
CM-ext	257	50.2	49.0	99.2	0.6
Mountain	375	-	617.0	617.0	0.63

Table 3: Typical results for respective problems (24 replicas/cpus and 100 000 samples). Note that '-' in case of Mountain Sed. RMSE indicates that it was not part of the likelihood function

Results



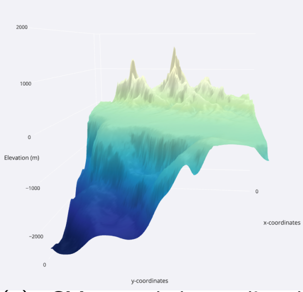
(a) CM-extended predicted topography after 25 % evolution time



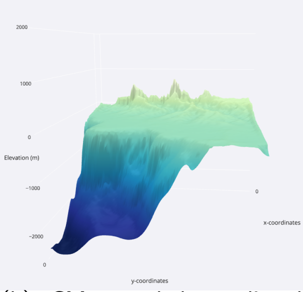
(b) CM-extended predicted topography after 50 % evolution time

Figure 10: CM-extended evolution over 1 000 000 years.

Results



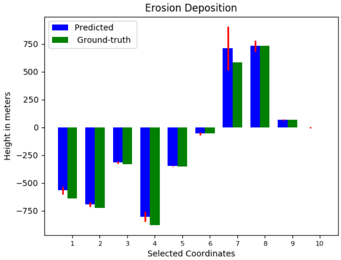
(a) CM-extended predicted topography after 50 % evolution time



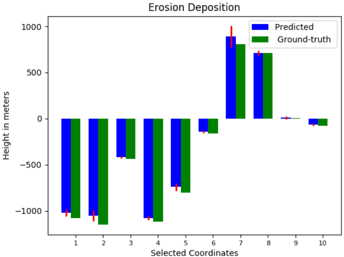
(b) CM-extended predicted topography after 100 % evolution time

Figure 11: CM-extended: Topography and erosion-deposition development for selected time frames. Note that erosion (positive) and deposition (negative) values given by the height in meters

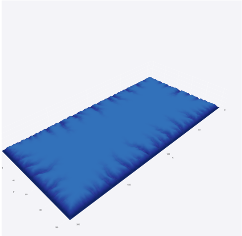
Results



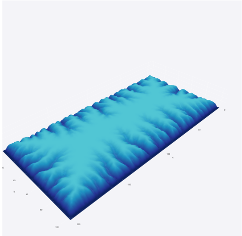
(a) CM-extended erosion-deposition after 50 % evolution time



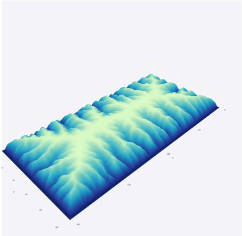
Results

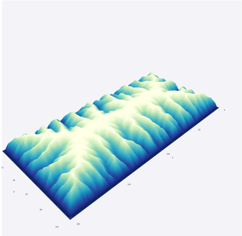


(a) Mountain predicted topography after 25 % evolution time



(b) Mountain predicted topography after 50 % evolution time





Results

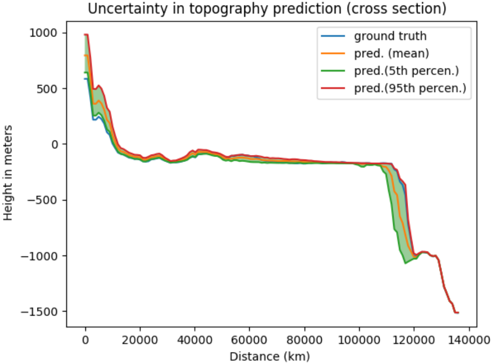


Figure 14: CM-extended: Cross-section of the elevation along the x-axis comparing the ground-truth of the Badlands-model with the PT-Bayelands predictions.

Results

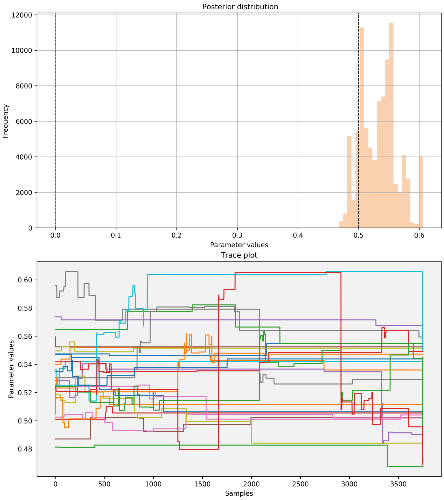


Figure 15: Typical posterior distribution (top panel) and replica traces (bottom panel) exploring the marine diffusion parameter-space for the CM-extended topography. In this model run 24 replicas are represented by the coloured lines. Parameter, c-marine, units are m^2a^{-1}

Discussion

- The convergence challenges in MCMC due computational time taken to run a single model. Challenge for Badlands especially for real world applications
- The need for simulated problems that better represent real-world topographies

Surface

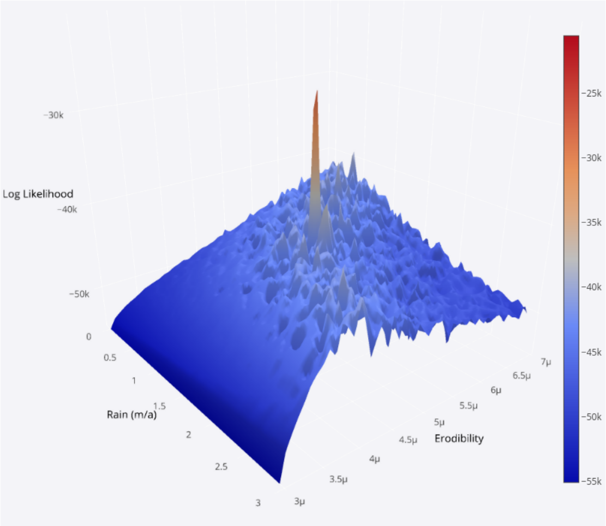


Figure 16: Likelihood surfaces of the Continental Margin topography for the rainfall and erodibility parameters only

Future work

- Surrogate assisted likelihood functions
- Robust sampling techniques such as parallel tempering that explore multi-modality.

Ahead

- Way ahead - real world applications is to employ multi-core parallel tempering with Surrogates for Likelihood evaluations.
- Also exploring GPU based acceleration for Badlands. Challenge will be to harness the power of GPU with multi-core parallel tempering implementations.

Implications

- In future, it would be helpful to model region-based or time-varying rainfall distributions in Badlands.
- The distributions could be used to generate more information about the effects of climate change in geological timescales.

Fusion with optimisation methods

- Scope for optimisation methods such as evolutionary algorithms for models with hundreds of parameters
- Scope for fusion of evolutionary algorithms with MCMC methods such as Parallel Tempering for better proposals.
- Can be extended to other geophysical inversion problems.
- Gradient information from Forward Models can be useful for Bayesian Inference and Optimisation
- Library - software package.

Thanks

Project members: Danial Azam, Dietmar Muller, Sally Cripps, Tristan Salles, Ratneel Deo, and Nathaniel Butterworth

- Bayeslands [1]
- PT-Bayeslands [2]

[1] https://github.com/badlands-model/BayesLands
[2] https://github.com/badlands-model/paralleltemp_Bayeslands