"Bayesian inference for modelling geo-coastal, basin and landscape evolution"

Basin Genesis Hub - Workshop

Rohitash Chandra

Centre for Translational Data Science
The University of Sydney, Sydney, Australia

February 22, 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

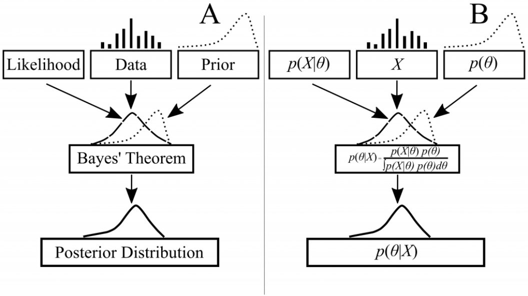
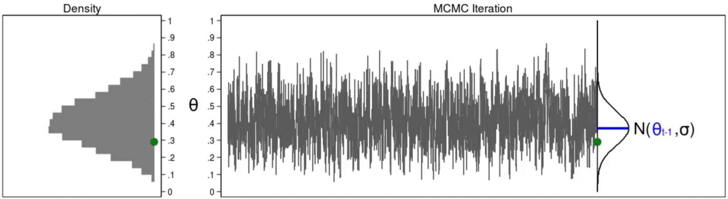


Figure 1: Bayesian inference overview

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

BayesReef: A Bayesian inference framework for modelling vertical coral reef growth and response to environmental changes

- Bayesian inference methods have rarely been applied to reef modelling, despite evidence of their usefulness when handling models with complex, interrelating parameters.
- BayesReef is a novel framework to explore data fusion with a MCMC approach to modelling long-term reef development.
- BayesReef extends the model in *PyReef-Core* by placing probability distributions over the initial conditions thereby turning a deterministic model into a probabilistic one.

Reef-Core Data

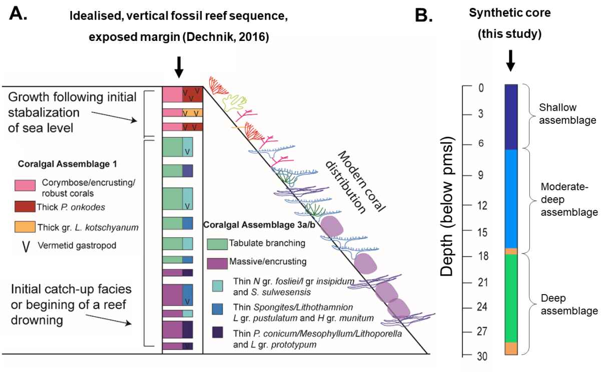


Figure 3: Schematic representation of synthetic data construction. (A) Ideal shallowing-up fossil reef sequence (B) Synthetic core representing ideal shallowing-upward, catch-up sequence, and detail of assemblages for a synthetic core.

Results

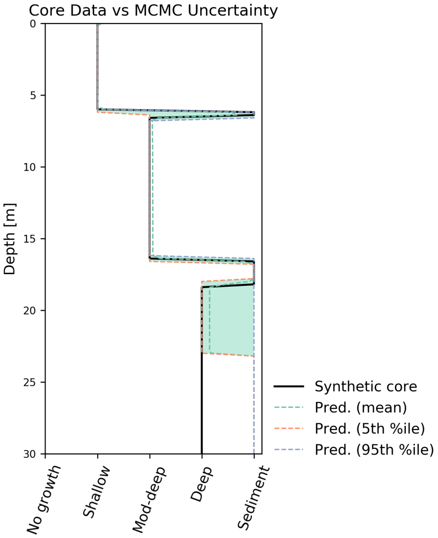


Figure 4: Reef-core prediction with uncertainty

Results

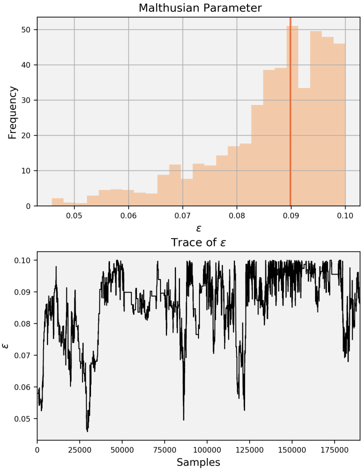


Figure 5: Posterior distribution for the Malthusian parameter ϵ. The solid line in red shows the true value (0.086).

BayesReef framework

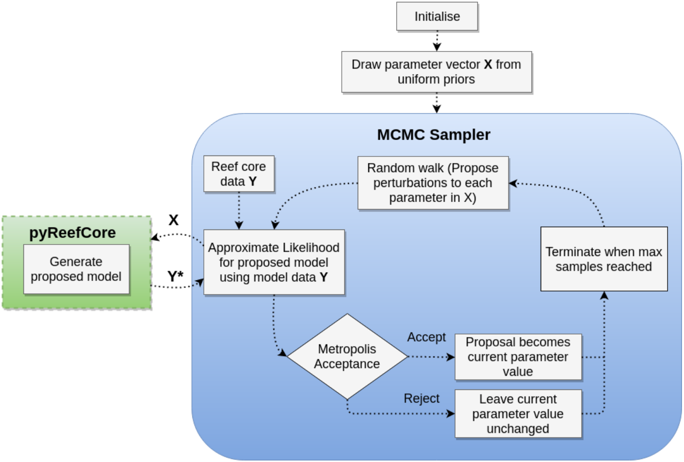


Figure 6: Workflow of *BayesReef* framework which uses MCMC sampling for inference of free parameters in pyReef-Core model.

Discussion

- Given a complex pyReef core model with 27 parameters, BayesReef provided accurate reef-core outputs when compared to reef-core data.
- In future work, other Bayesian inference methods such as parallel tempering and reversible jump MCMC can be used in the BayesReef framework.

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

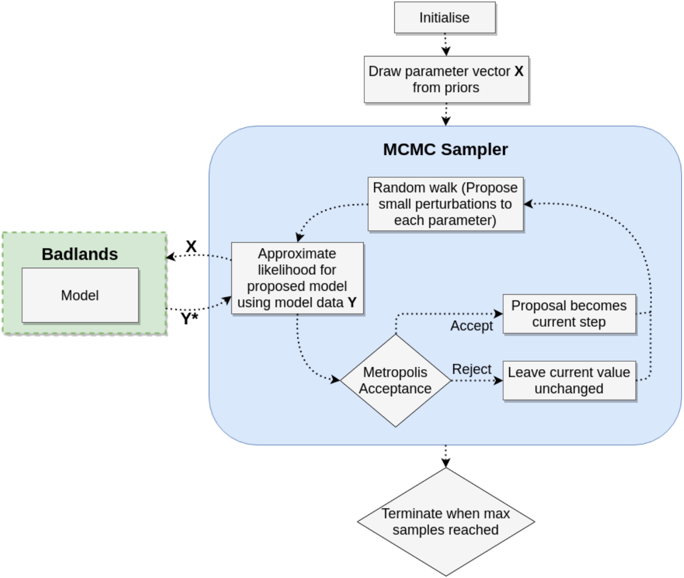


Figure 7: *Bayeslands* framework for Badlands.

Crater

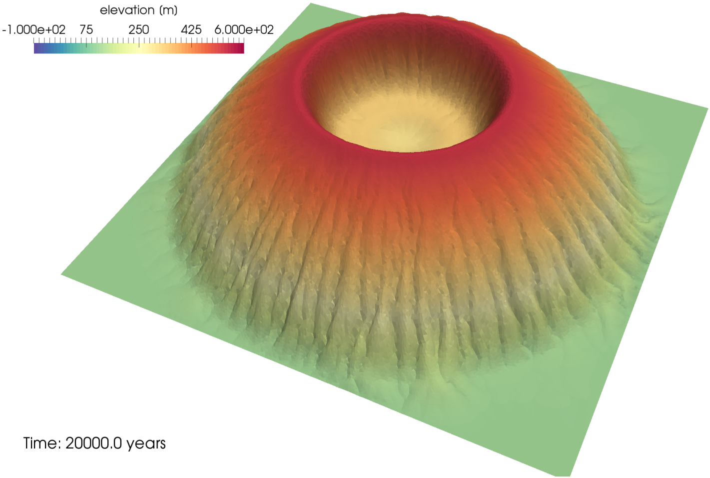


Figure 8: 20 000 years

Crater

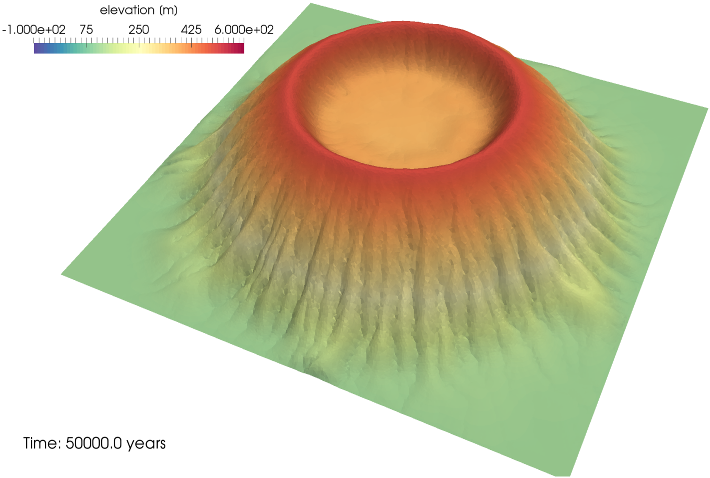


Figure 9: 50 000 years

Crater

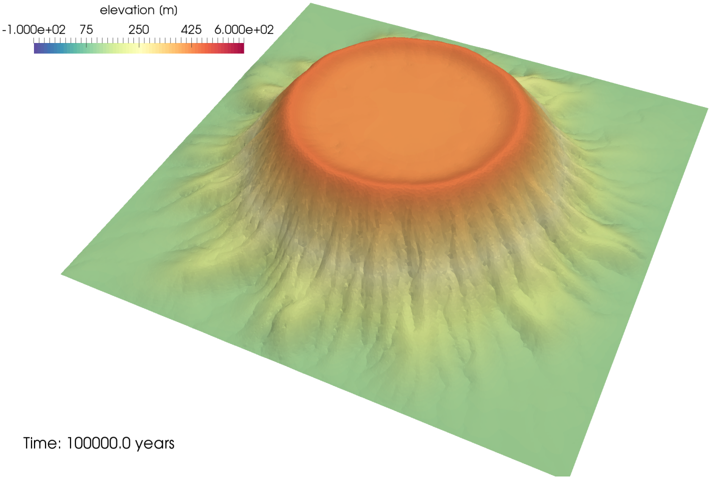


Figure 10: 100 000 years

Crater

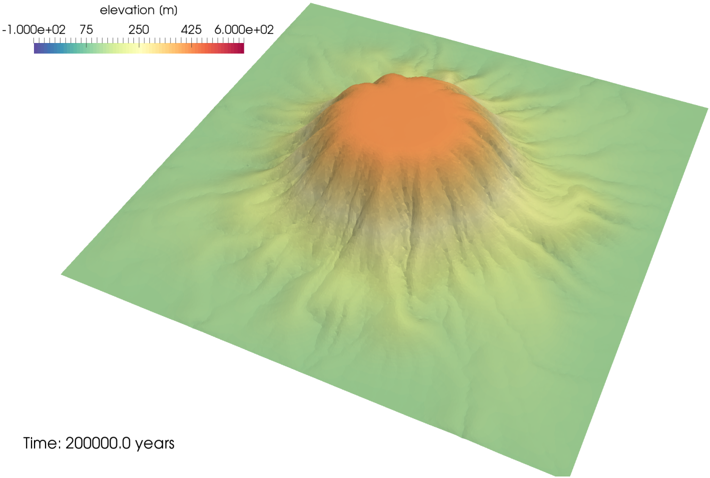
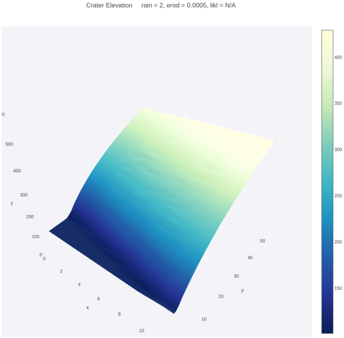
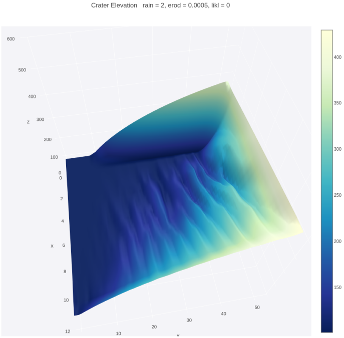


Figure 11: 200 000 years

Example 1: Hillside



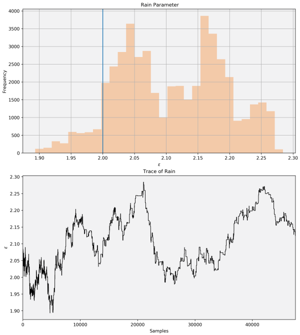
(a) Hillside predicted (mean)



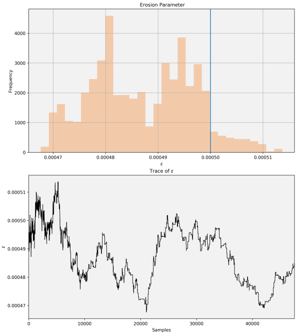
(b) Hillside eroded ground-truth topography

Figure 12: Hillside: initial and eroded after 5k years

Example 1: Posterior distributions



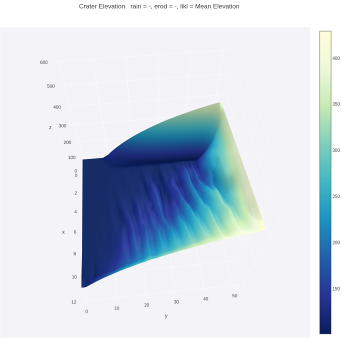
(a) Hill-slope rainfall distribution



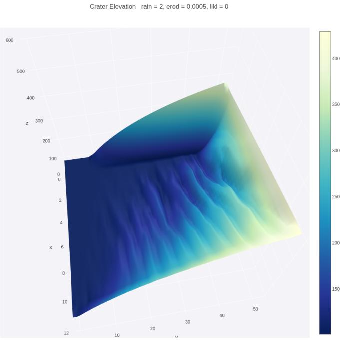
(b) Hill-slope erodibility distribution

Figure 13: Hill-slope: Posterior distribution of rainfall and erodibility

Example 1: Predicted topography via Bayeslands



(a) Hillside predicted via Bayeslands



(b) Hillside ground-truth topography

Figure 14: Hillside: Predicted via Bayeslands

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

Future work

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

Parallel tempering

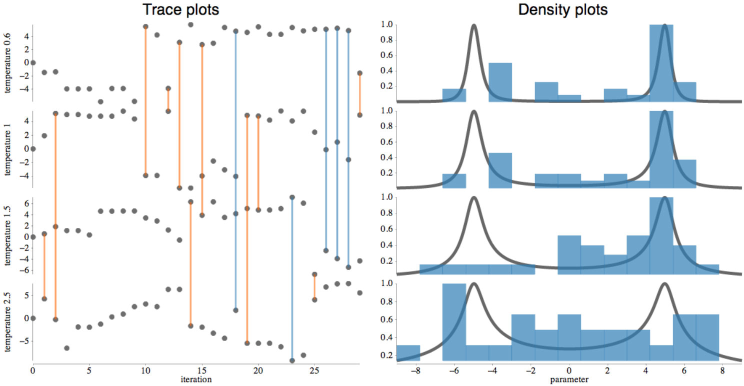


Figure 15: Parallel tempering - sampling

Multi-core and GPU implementations

- We have implemented in-house multi-core parallel tempering python package for Badlands.
- 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, Jody Webster, Jodie Pall, and Nathaniel Butterworth

- Bayeslands [1]
- BayesReef [2]
- Multi-core parallel tempering [3]

[1] https://github.com/badlands-model/BayesLands
[2] https://github.com/pyReef-model/BayesReef
[3] https://github.com/rohitash-chandra/multicore_parallel_temp_mixturemodel