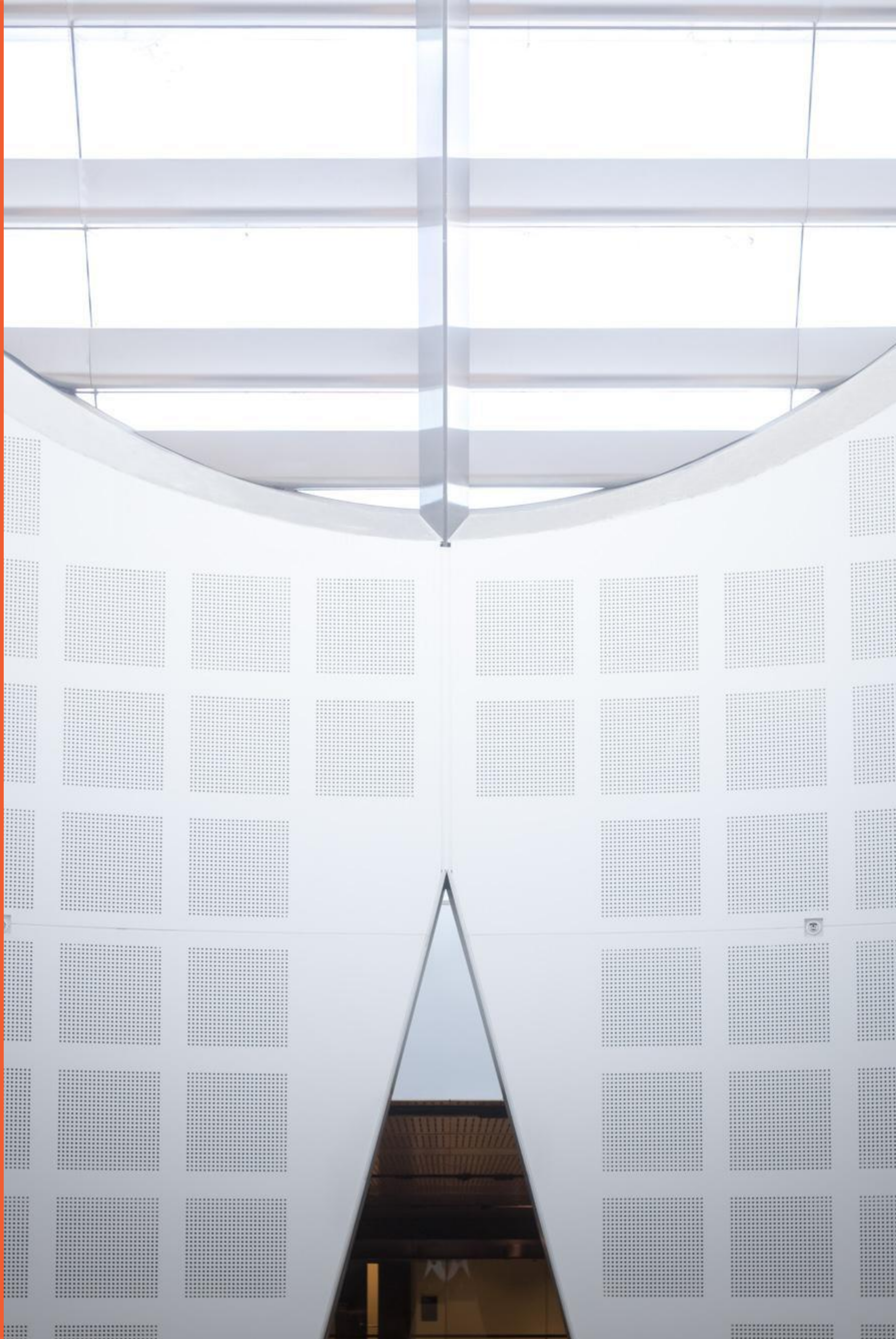
Unravelling Earth's geological history with geoscientific models powered by artificial intelligence

Dr. Rohitash Chandra

1. Centre for Translational Data Science
2. EarthByte Group, School of Geosciences

The University of Sydney

THE UNIVERSITY OF SYDNEY



Overview

- Bayesian inference and optimisation
- Landscape evolution models with application to Australian continent over 150 million years
- Reef evolution models with application to Great Barrier Reef over 10 thousand years
- Paleoclimate reconstruction over 500 million years
- Open Source Software

THE UNIVERSITY OF SYDNEY

Why Bayesian inference?

- Bayesian inference: principled approach towards uncertainty quantification of free parameters
- Calculate probability distributions of parameters instead of single point estimates
- Estimating the free parameters of a given model (posterior distribution) is often nontrivial
  - Challenges:
    - Effective prior information
    - multi-modal surfaces

Prof. Sally Cripps
Director, Centre for Translational Data Science
University of Sydney

MCMC/Parallel Tempering Framework

- Markov Chain Monte Carlo sampling methods (MCMC) sample from a probability distribution.
- Construct a Markov chain after several steps that reaches desired distribution as its equilibrium distribution.

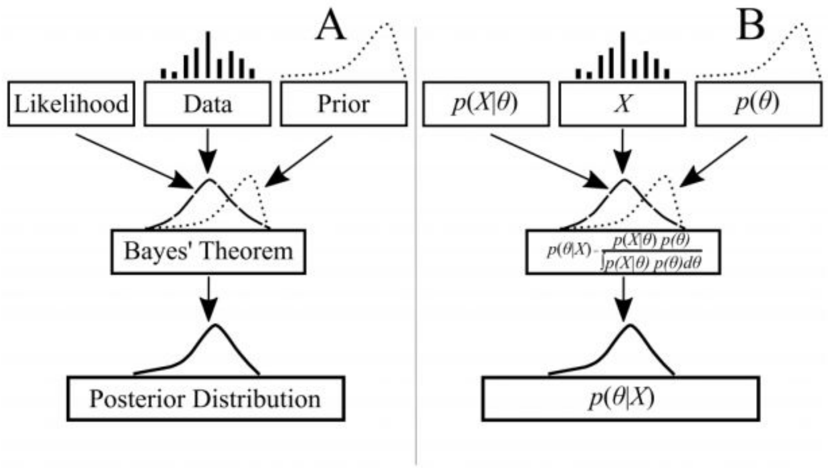
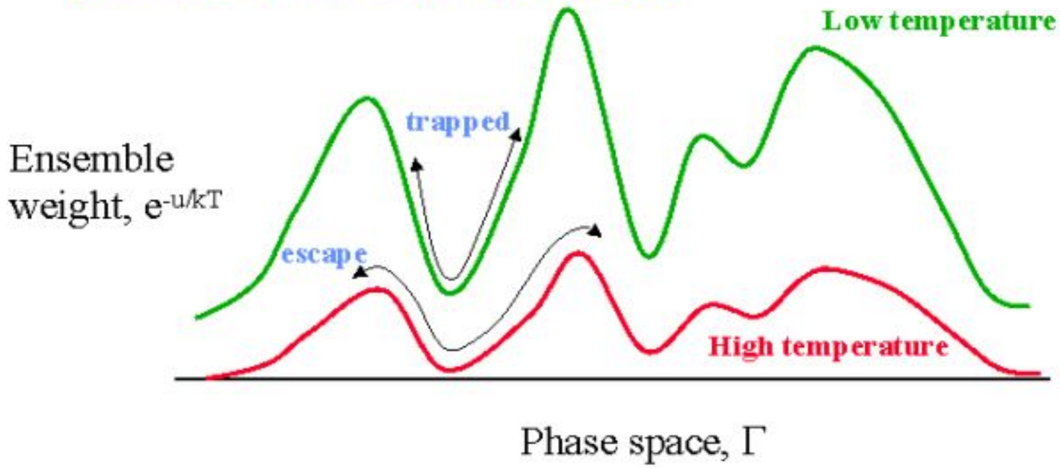


Figure 1: MCMC overview

Parallel Tempering MCMC Framework

- Parallel Tempering is a simulation method aimed at improving the dynamic properties of MCMC method
  - At high temperature a broader range of configurations is sampled
  - Barriers to transitions are lowered



  - How to simulate a low-temperature system with high-temperature barrier removal?

Figure 2: Parallel Tempering overview

MCMC/Parallel Tempering Framework

- Replica exchange demonstration

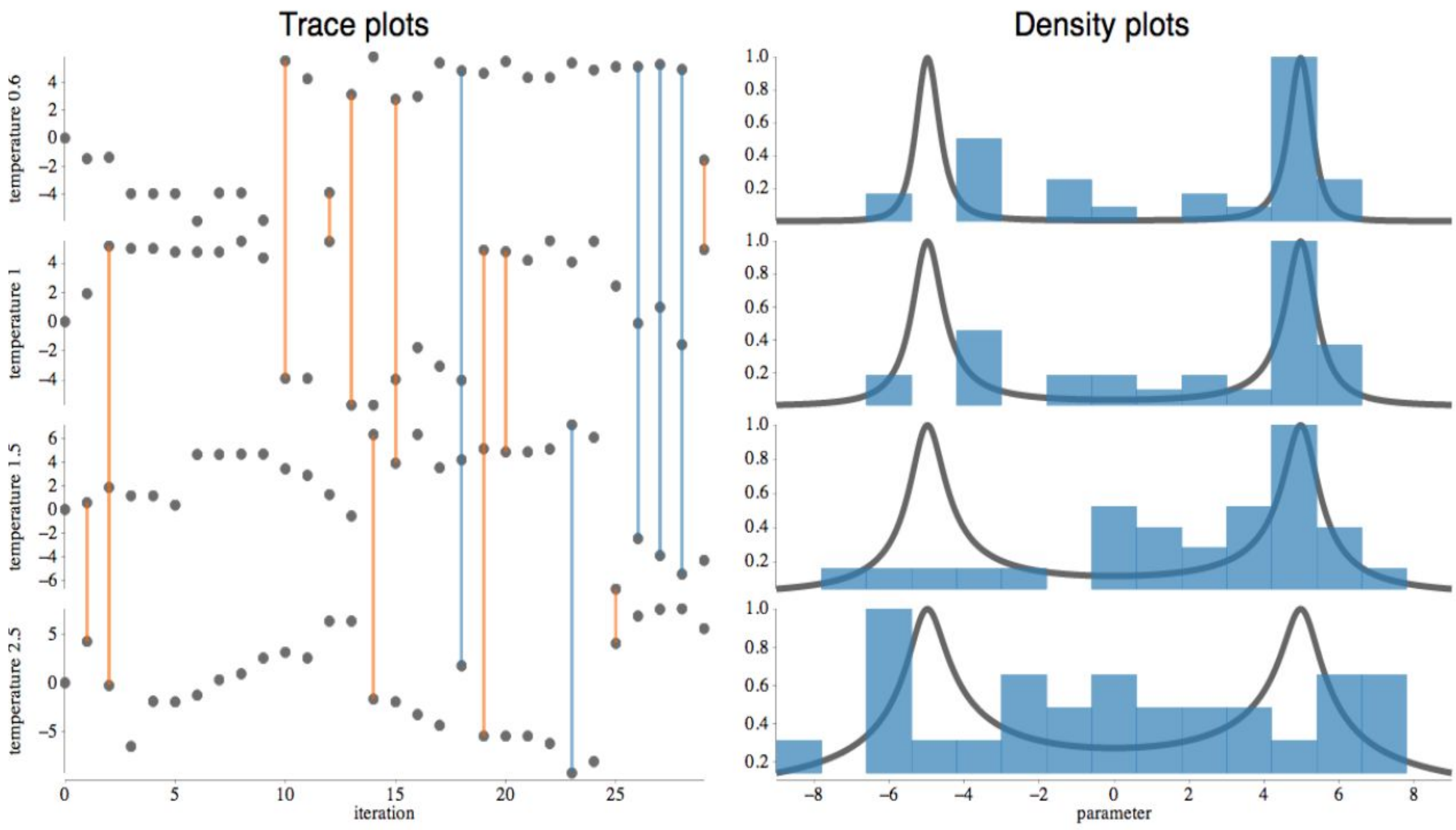


Figure 3: Parallel Tempering sampling

EarthByte

Building a Virtual Earth

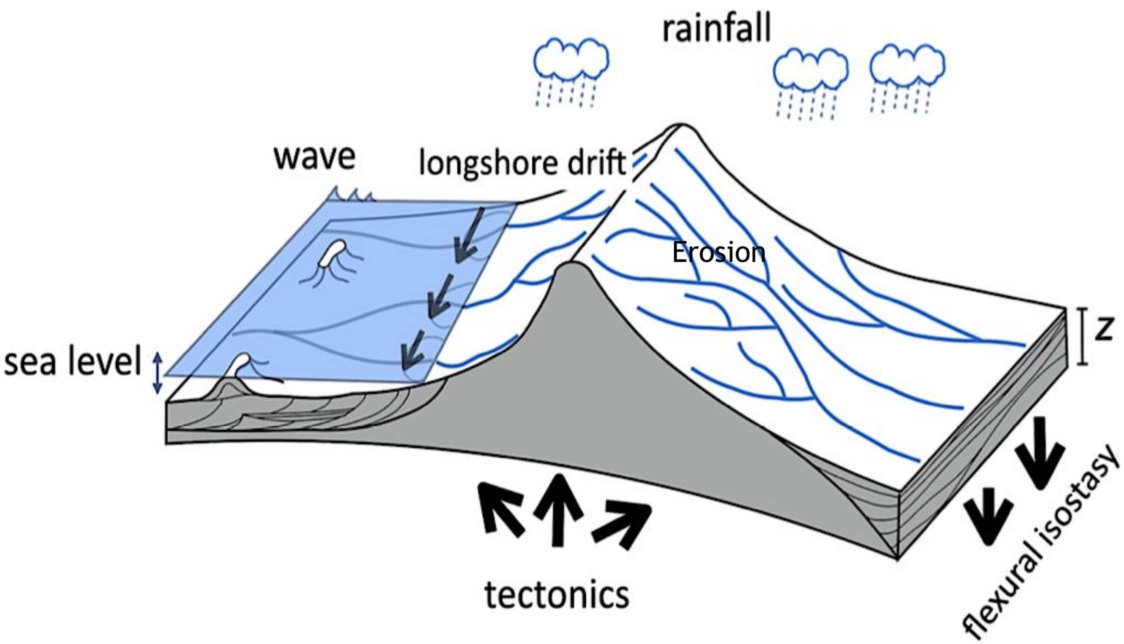
What is EarthByte?

EarthByte is an internationally leading eGeoscience collaboration between several Australian Universities, international centres of excellence and industry partners. One of the fundamental aims of the EarthByte Group is geodata synthesis through space and time, assimilating the wealth of disparate geological and geophysical data into a four-dimensional Earth model including tectonics, geodynamics and surface processes.

Prof. Dietmar Muller
Former ARC Laureate Fellow
Director, ARC Basin Genesis Hub
University of Sydney

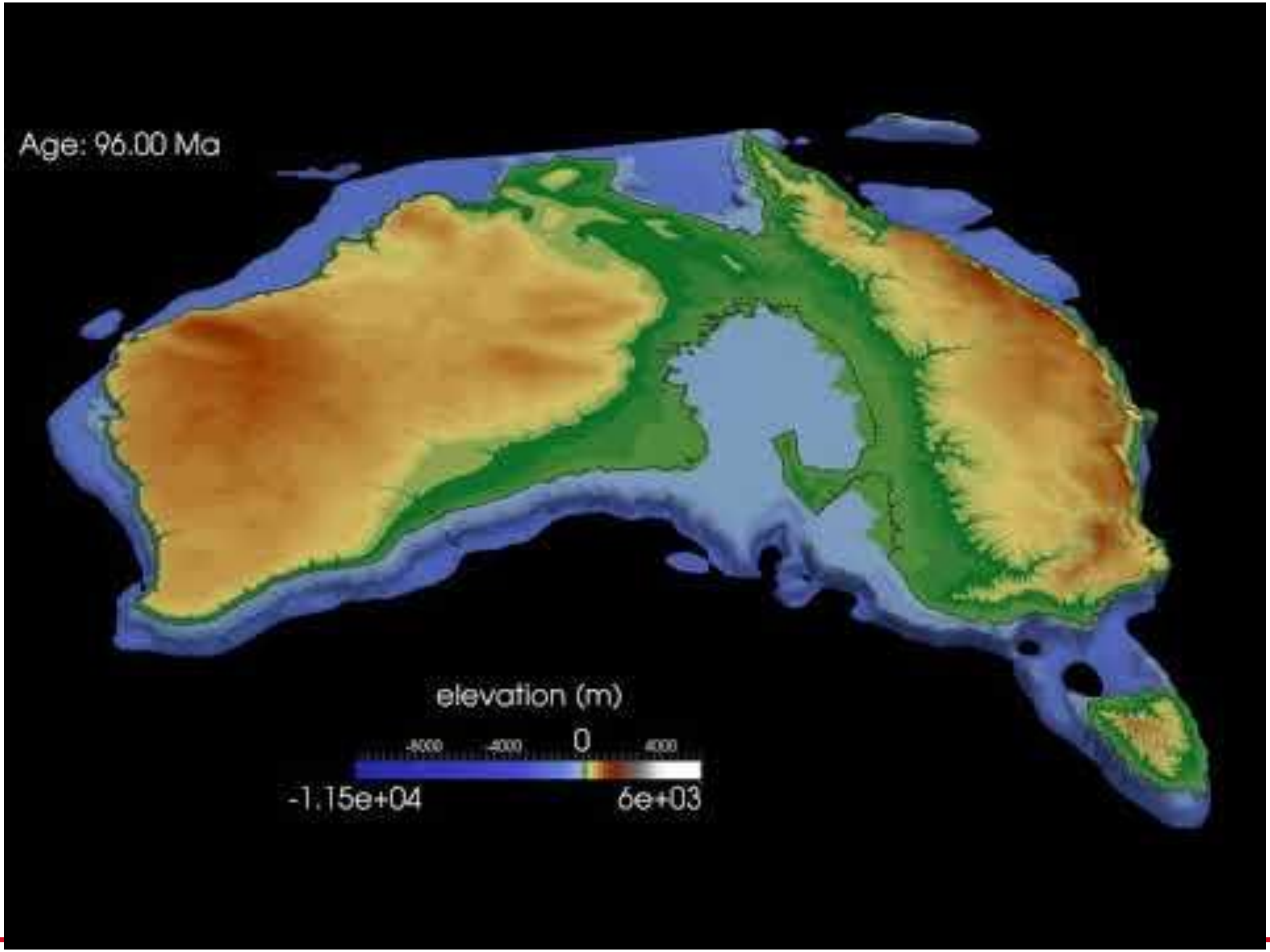
Badlands Model

Badlands: Modelling erosion and deposition



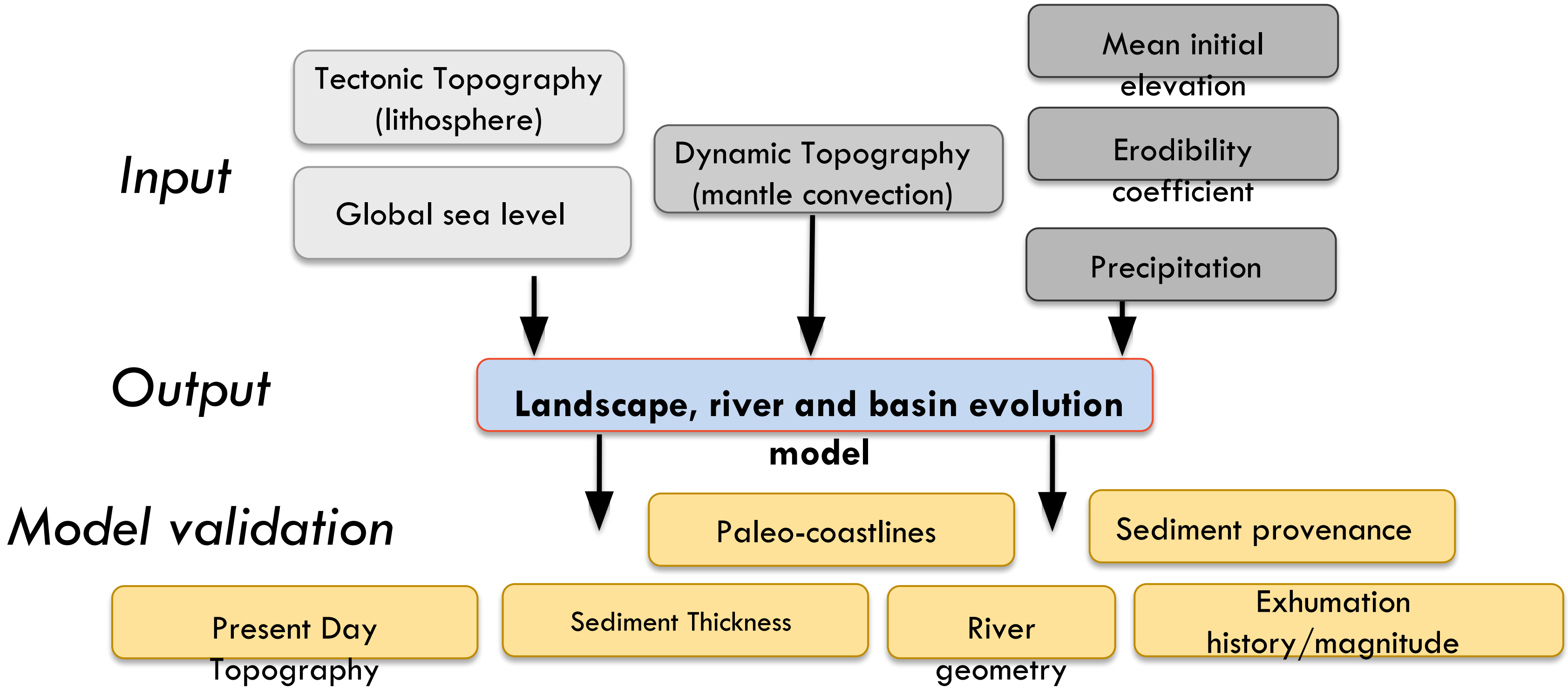
Salles and Hardiman, 2016

Dr. Tristan Salles
University of Sydney


Age: 96.00 Ma
elevation (m)
-8000
-4000
0
4000
-1.15e+04
6e+03

Badlands model components

Model inputs, output and validation



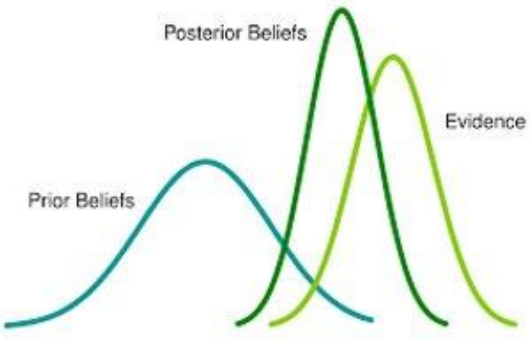
Bayeslands

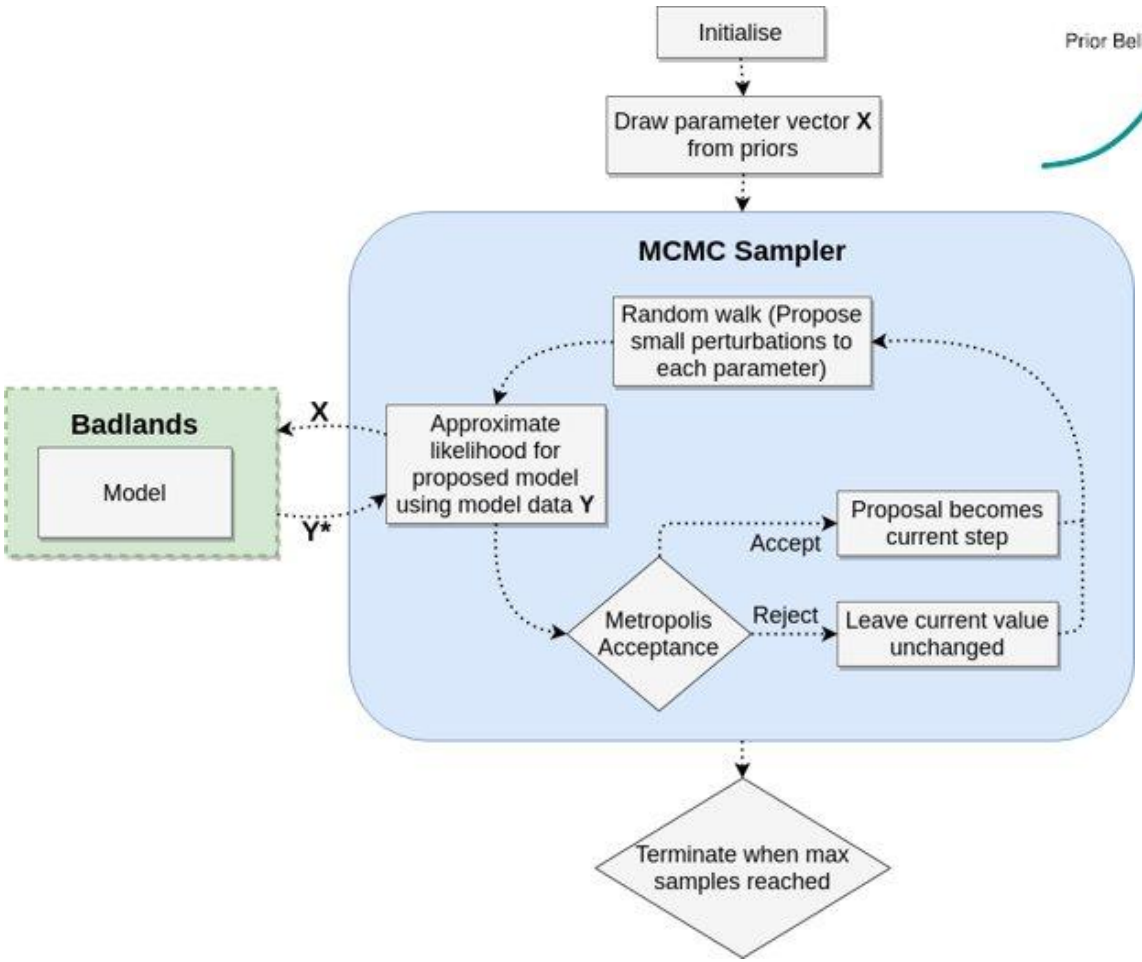
- Bayeslands, a framework for inference and uncertainty quantification in the Badlands
- Use the Bayesian paradigm to find the best-fit parameters driving basin evolution models using Badlands

Danial Azam
Research Engineer
EarthByte, University of Sydney

Dr. Nathaniel Butterworth
Senior Informatics Engineer
Sydney Informatics Hub
University of Sydney

Ratneel Deo
(External Collaborator)
University of the South Pacific


Posterior Beliefs
Evidence
Prior Beliefs


Initialise
Draw parameter vector X from priors
MCMC Sampler
Random walk (Propose small perturbations to each parameter)
Badlands
Model
X
Y*
Approximate likelihood for proposed model using model data Y
Proposal becomes current step
Accept
Metropolis Acceptance
Reject
Leave current value unchanged
Terminate when max samples reached

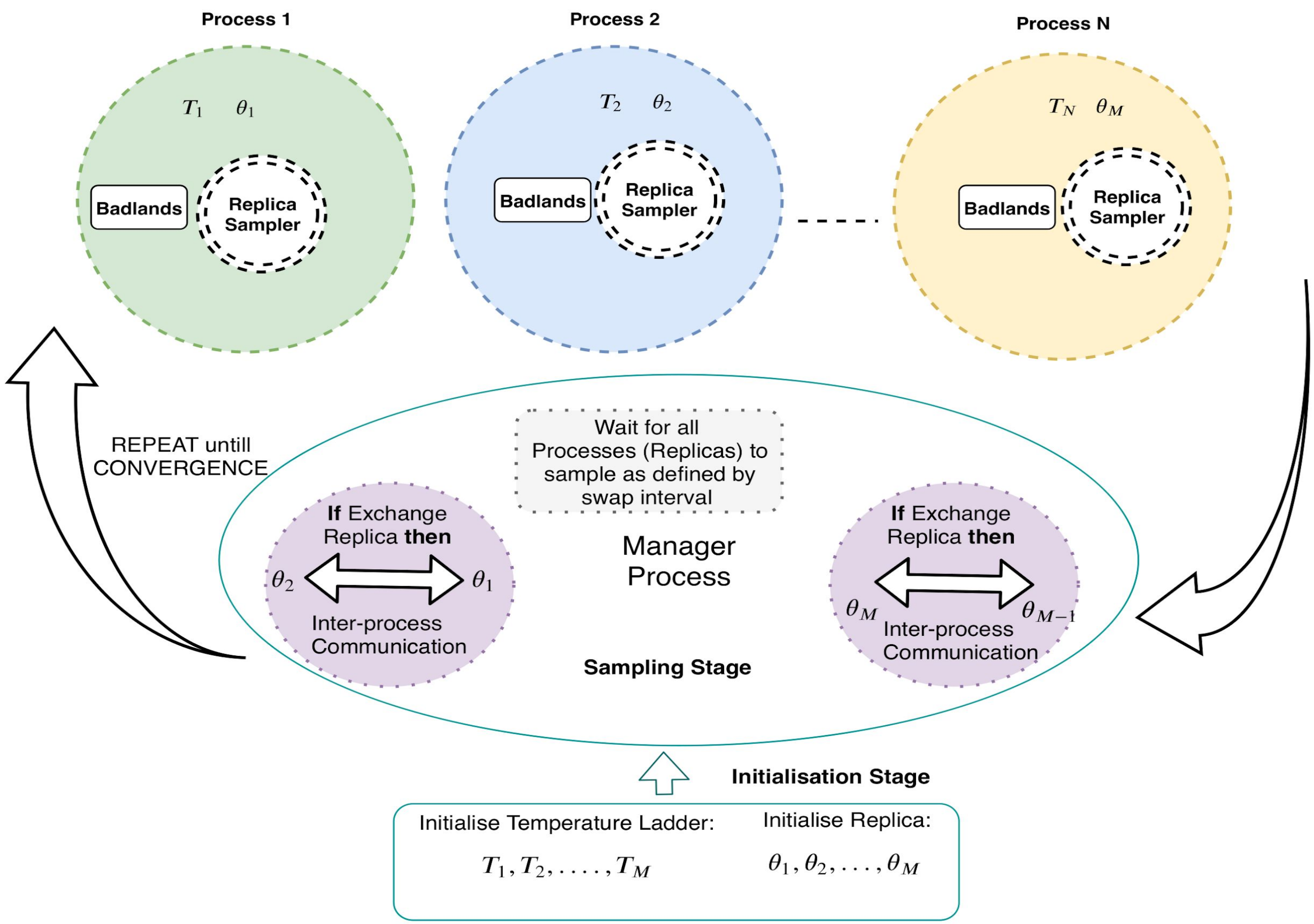
Bayeslands powered by parallel computing

- Badlands produces a series of consequent topographies at given time intervals
- We assume the final topography is the only observed topography
- Modelling requires
  - Initial topography
  - Set of input parameters

Input Parameters

- Input parameters include:
  - ***precipitation:*** temporal variations in precipitation as a constant value (metres per year)
  - ***erodibility:*** is scale-dependent coefficient and its value depends on lithology, channel width, flood frequency, channel hydraulics
  - ***m*** and ***n***: indicate how incision rate scales with bed shear stress for constant values of sediment flux and sediment transport capacity
  - ***caerial*** and ***cmarine***, the linear slope diffusion parameters

Framework



Process 1
T_1 θ_1
Badlands
Replica Sampler
Process 2
T_2 θ_2
Badlands
Replica Sampler
Process N
T_N θ_M
Badlands
Replica Sampler
REPEAT untill CONVERGENCE
Wait for all Processes (Replicas) to sample as defined by swap interval
If Exchange Replica then
θ_2 θ_1
Inter-process Communication
Manager Process
If Exchange Replica then
θ_M θ_{M-1}
Inter-process Communication
Sampling Stage
Initialisation Stage
Initialise Temperature Ladder:
$T_1, T_2, \ldots, T_M$
Initialise Replica:
$\theta_1, \theta_2, \ldots, \theta_M$

THE UNIVERSITY OF SYDNEY

Model Description

- Simulate the geomorphological evolution using parallel tempering Markov chain Monte Carlo (PT-MCMC)
  - Simulated time: **1** million years
  - Number of samples: **10000**
  - Number of replicas: **10**
- Likelihood function
  - elevation landscape
  - sediment erosion-deposition

Problem setup

- Selected continental margin problem taken from South Island, New Zealand

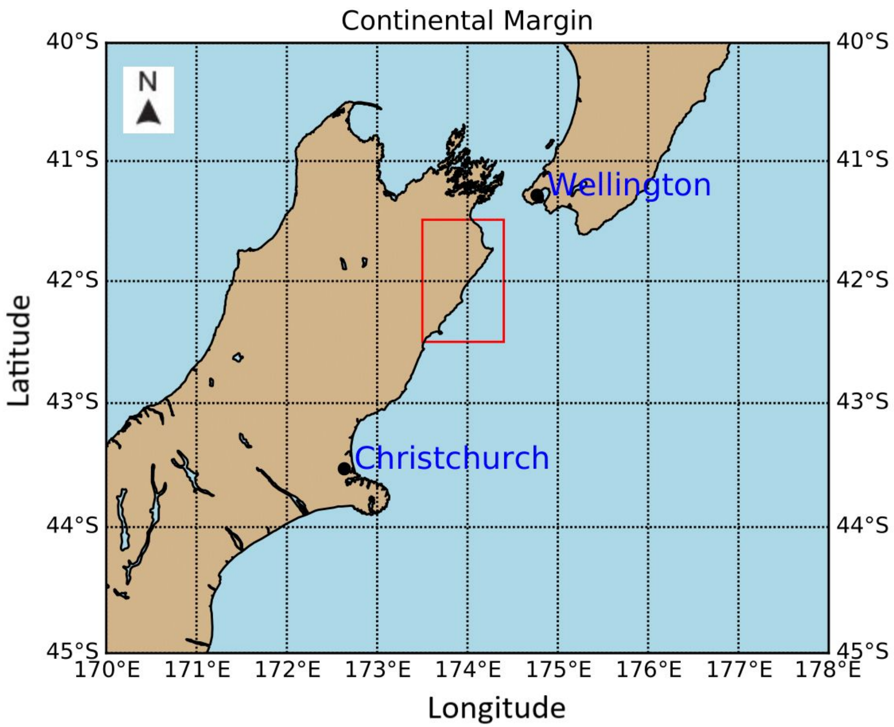
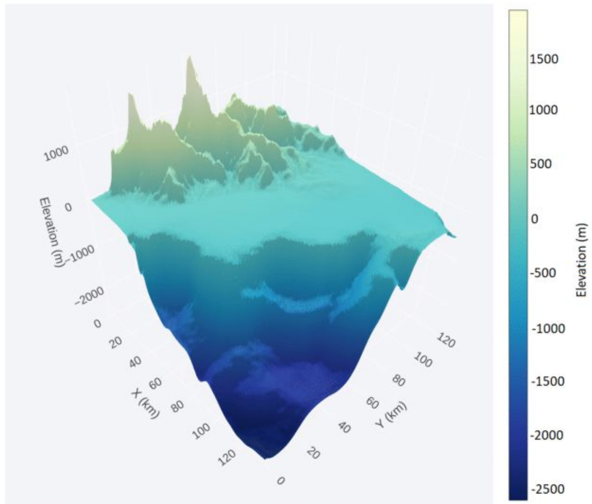
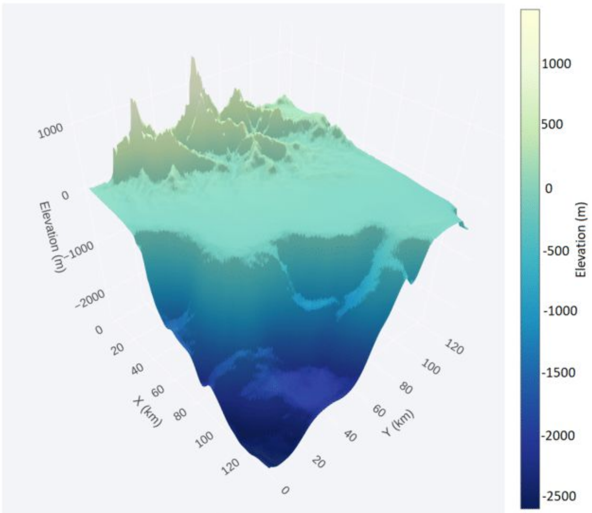


Figure 4: Selected continental margin problem from South Island, New Zealand outlined by the red rectangle.

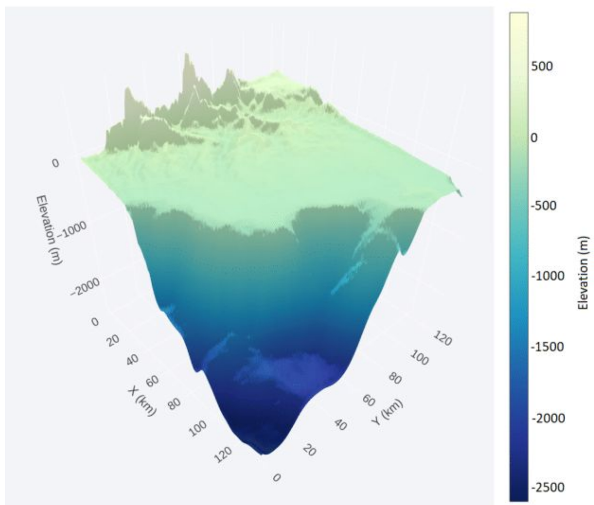
Landscape evolution over a million years



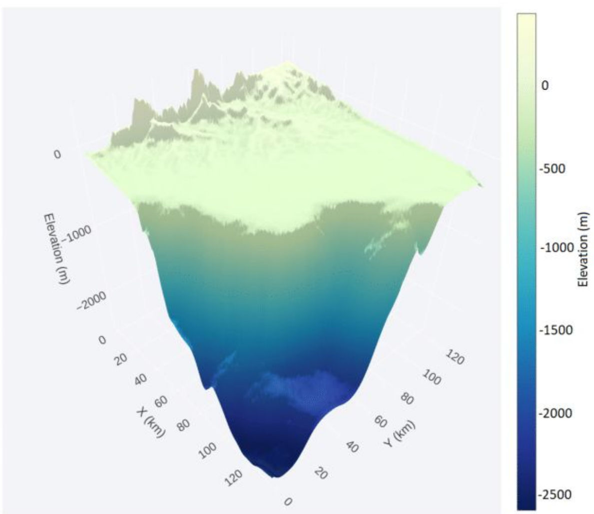
(a) 250 000 years



(b) 500 000 years



(c) 750 000 years



(d) 1000 000 years

Evaluation Criteria

- Performance is evaluated by:
  - Parameter estimates vs. parameters used in the generation of the simulated data
  - Quality of predictions
    - Sediment deposition in the basins
    - Elevation landscape

Solution space

- Solution space as function of *precipitation* and *erodibility*

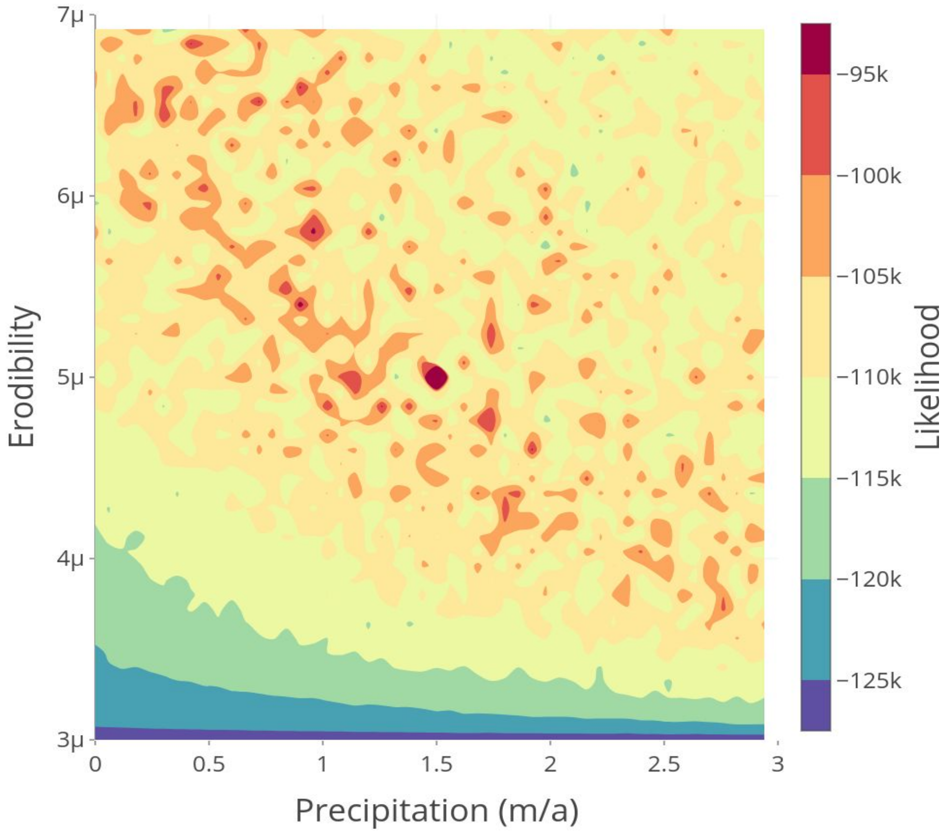


Figure 8 (a): Contour plot for Log solution space

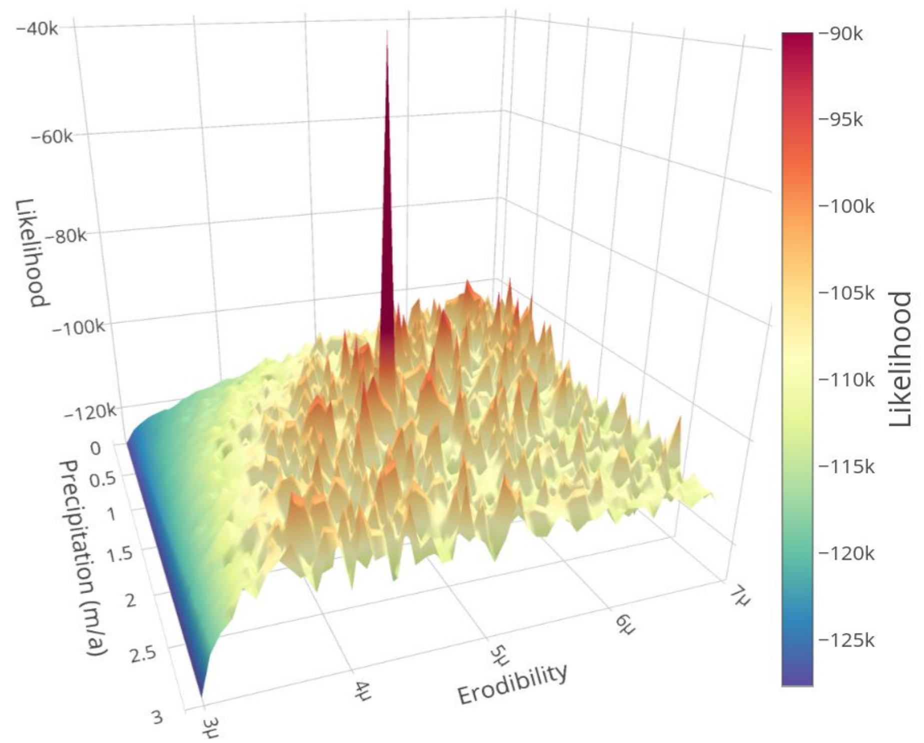


Figure 8 (b): Surface plot for Log solution space

Multimodality

- Geophysical inversion problems often exhibit multi-modal distributions
- Several possible input combinations give rise to identical elevation in the predicted topography

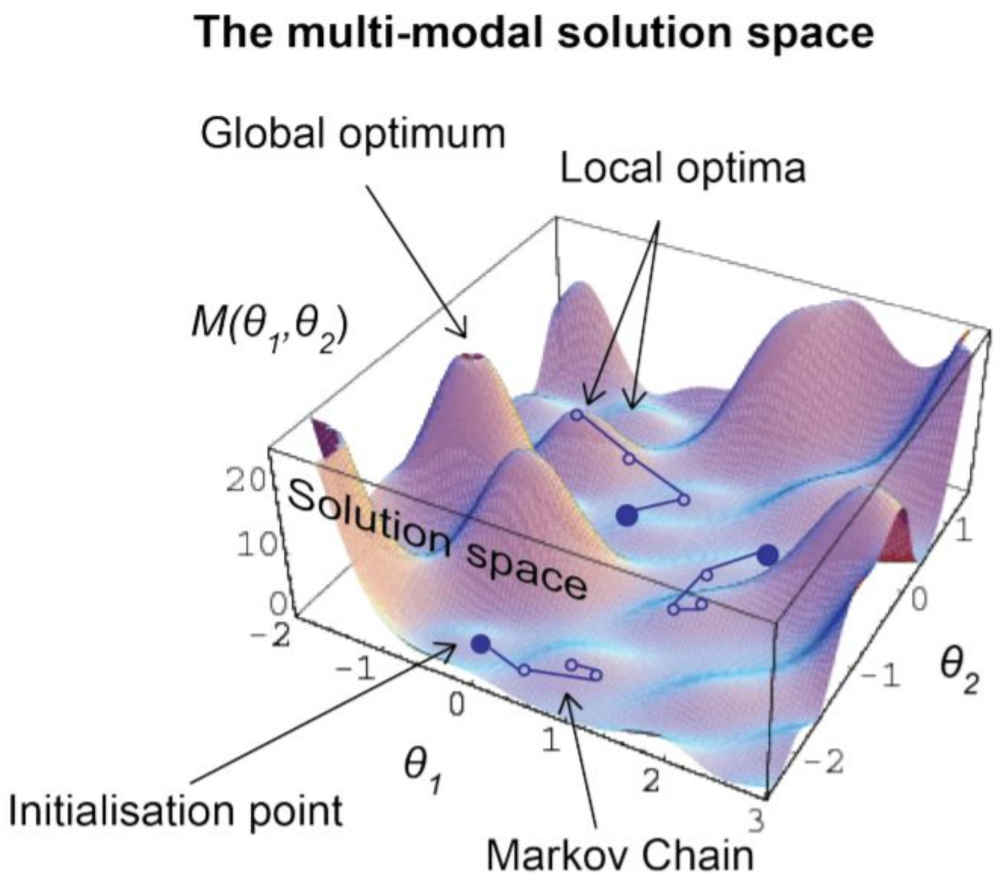


Figure 7: Multi-modal solution space

Multi-modal solution space

- Parallel Tempering helps efficiently explore the solution space

- Results show that convergence on sub-optimal modes gives similar topography evolution when compared with true values.

- The use of better proposal distributions can help in future work

Sediment Deposition

- To constrain multi-modality, use sediment erosion-deposition history along with elevation
- Figure heatmap shows the change in sediment thickness at the final time interval
- Yellow dots indicate 10 selected points

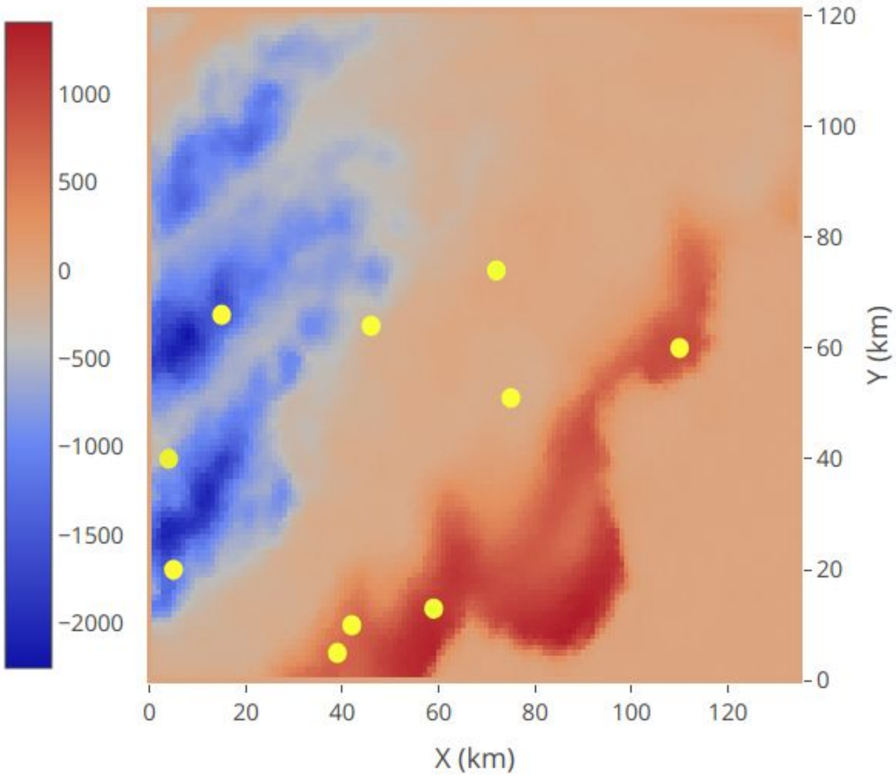


Figure 9 (a): Heatmap for evolved sediment deposition

Sediment Deposition

- Fig 9 (b) shows comparison of predicted vs ground truth at the chosen locations

- Note: positive values indicate deposition and the negative values indicate erosion

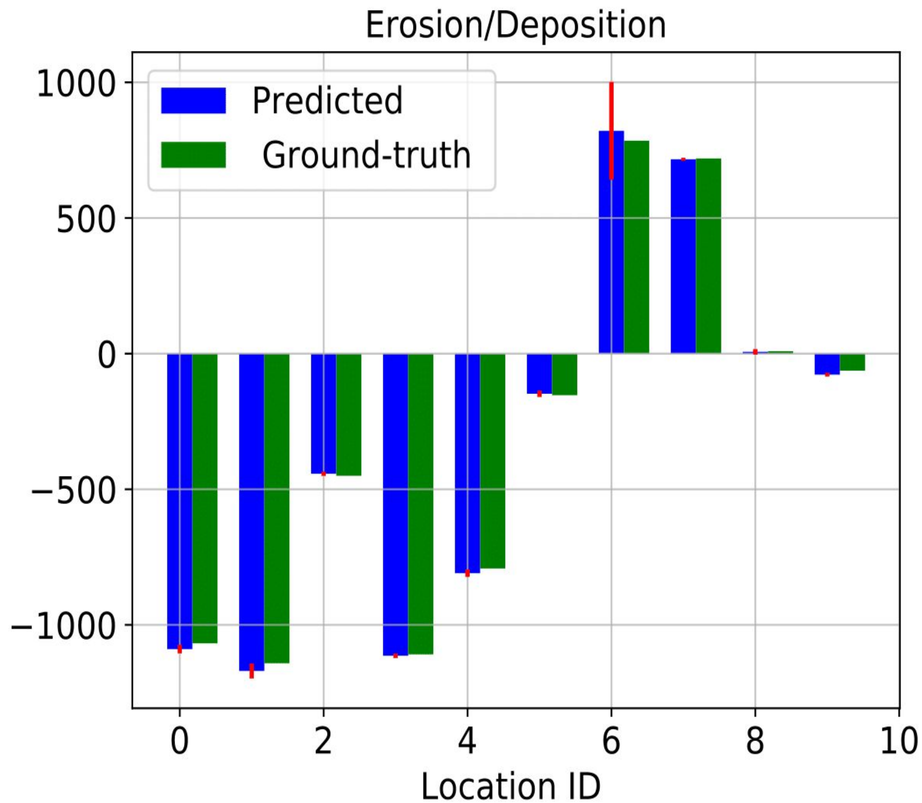


Figure 9 (b): Bar plot comparison of sediment deposition predicted vs ground-truth

Parameter estimate

- Estimates of posterior distribution and trace-plot of the precipitation parameter

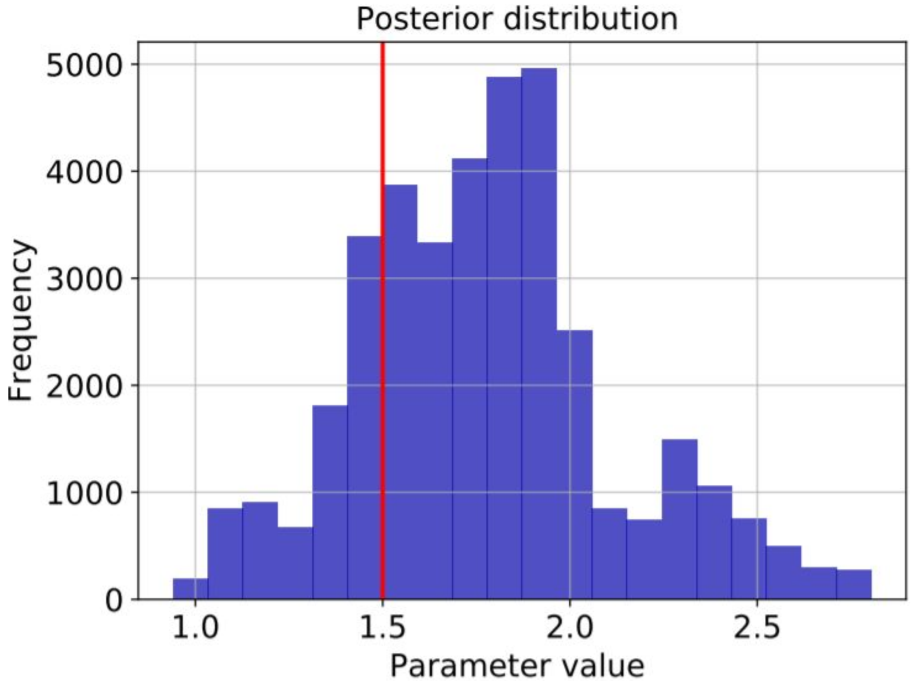


Figure 6 (a): Posterior distribution for precipitation (ρ) variable

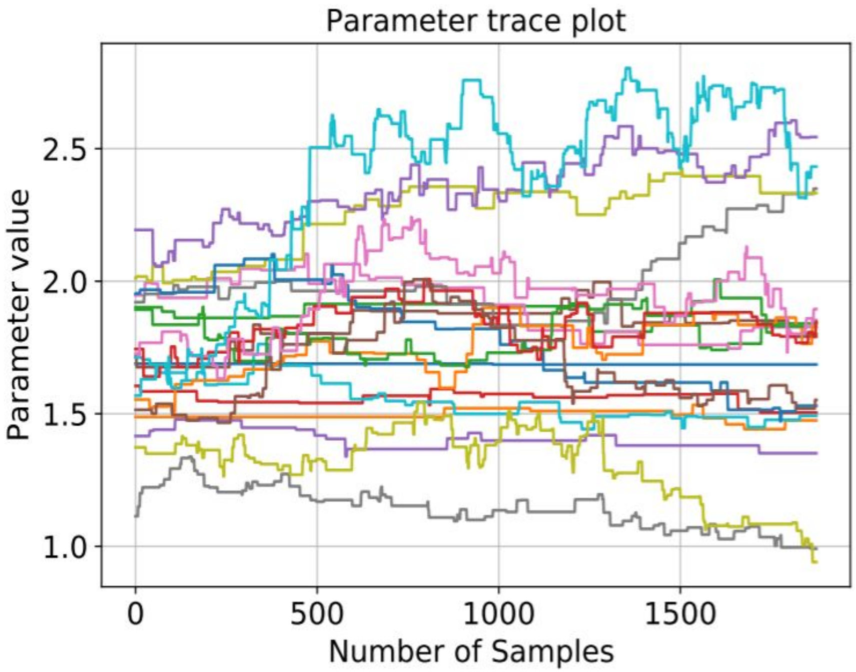
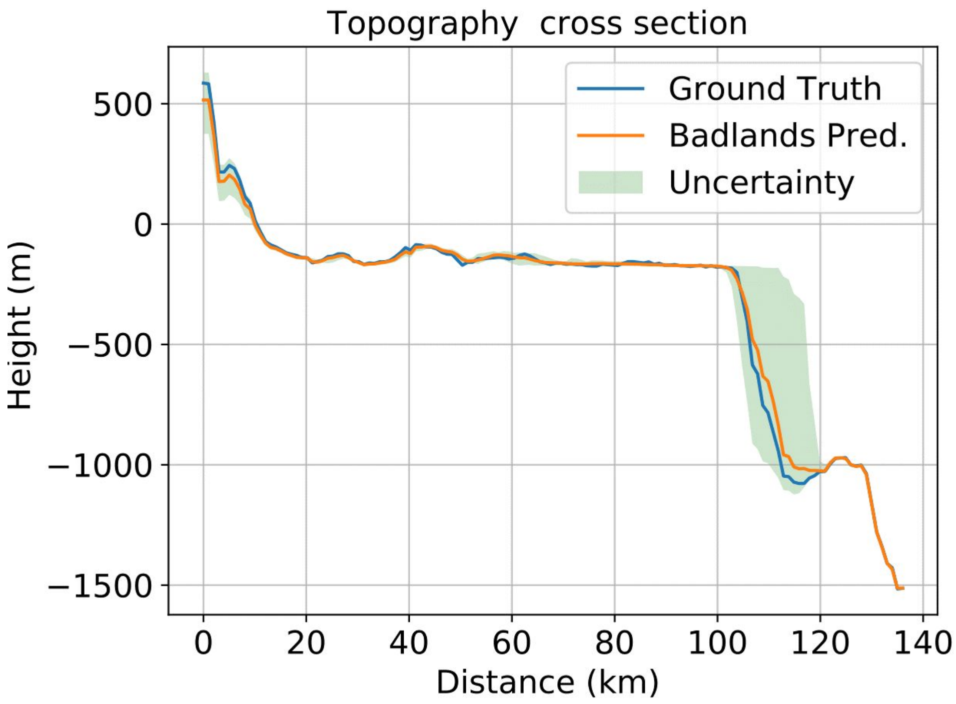


Figure 6 (b): Trace-plot for precipitation (ρ) variable

Elevation

- Cross-section of topography- Predicted vs. Ground-truth
- Uncertainty highlighted in green



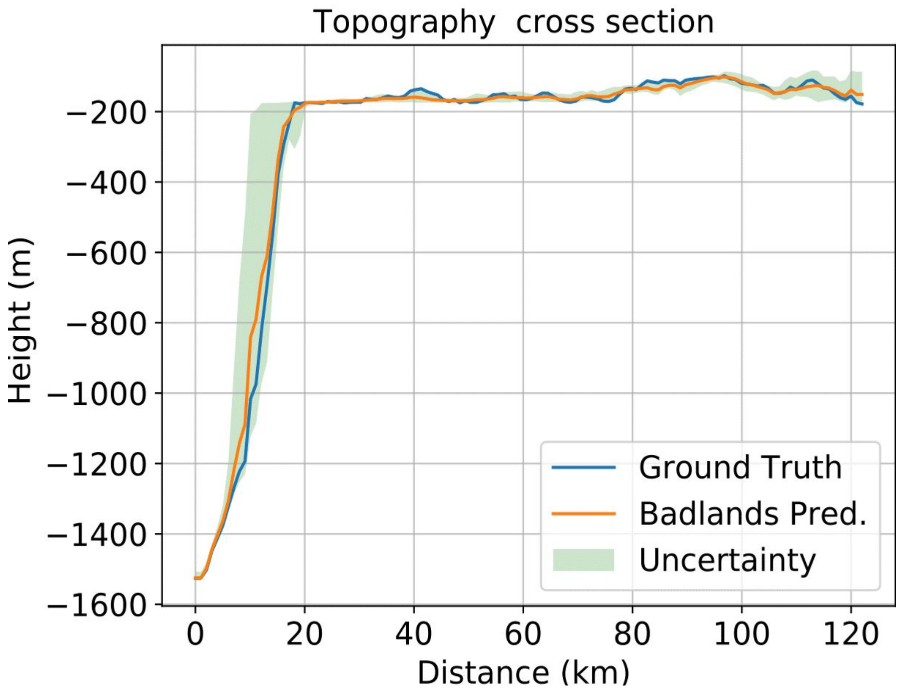
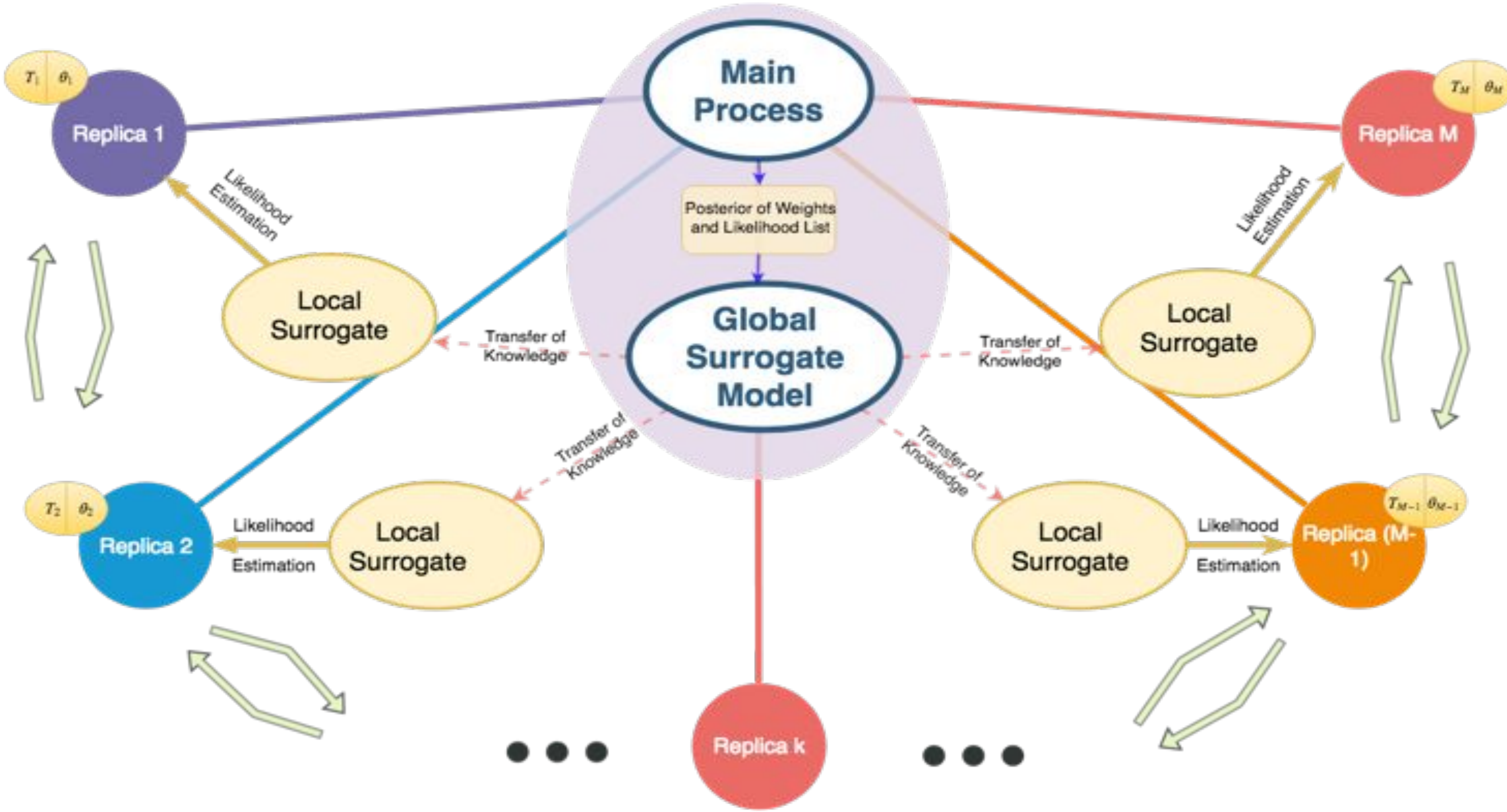
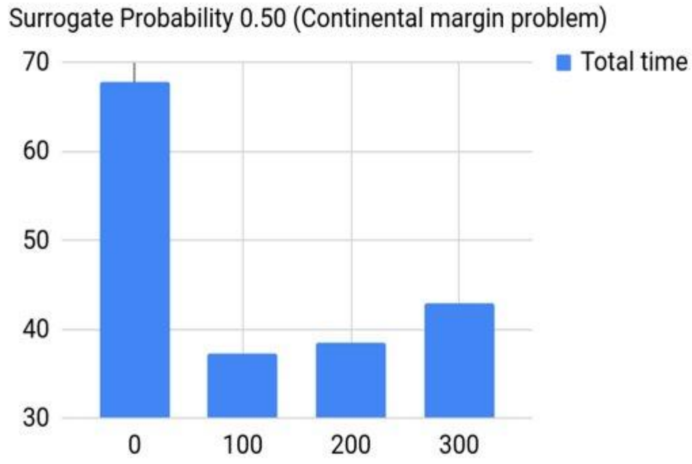
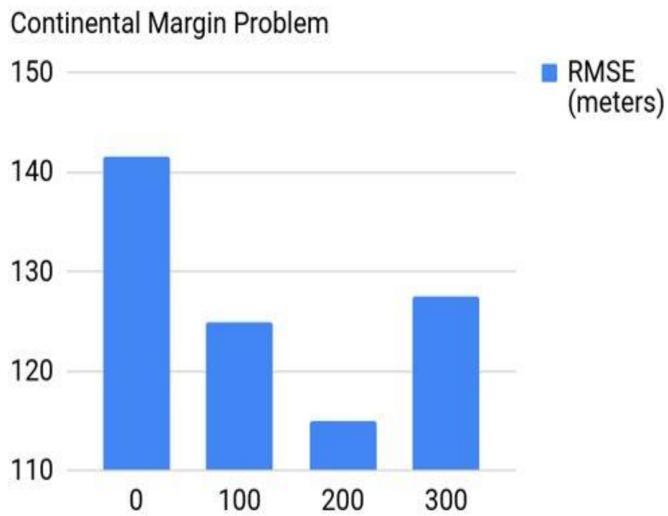


Figure 10: Cross-section comparison between predicted and ground truth. The uncertainty in prediction is highlighted in green.

Surrogate-assisted Bayeslands



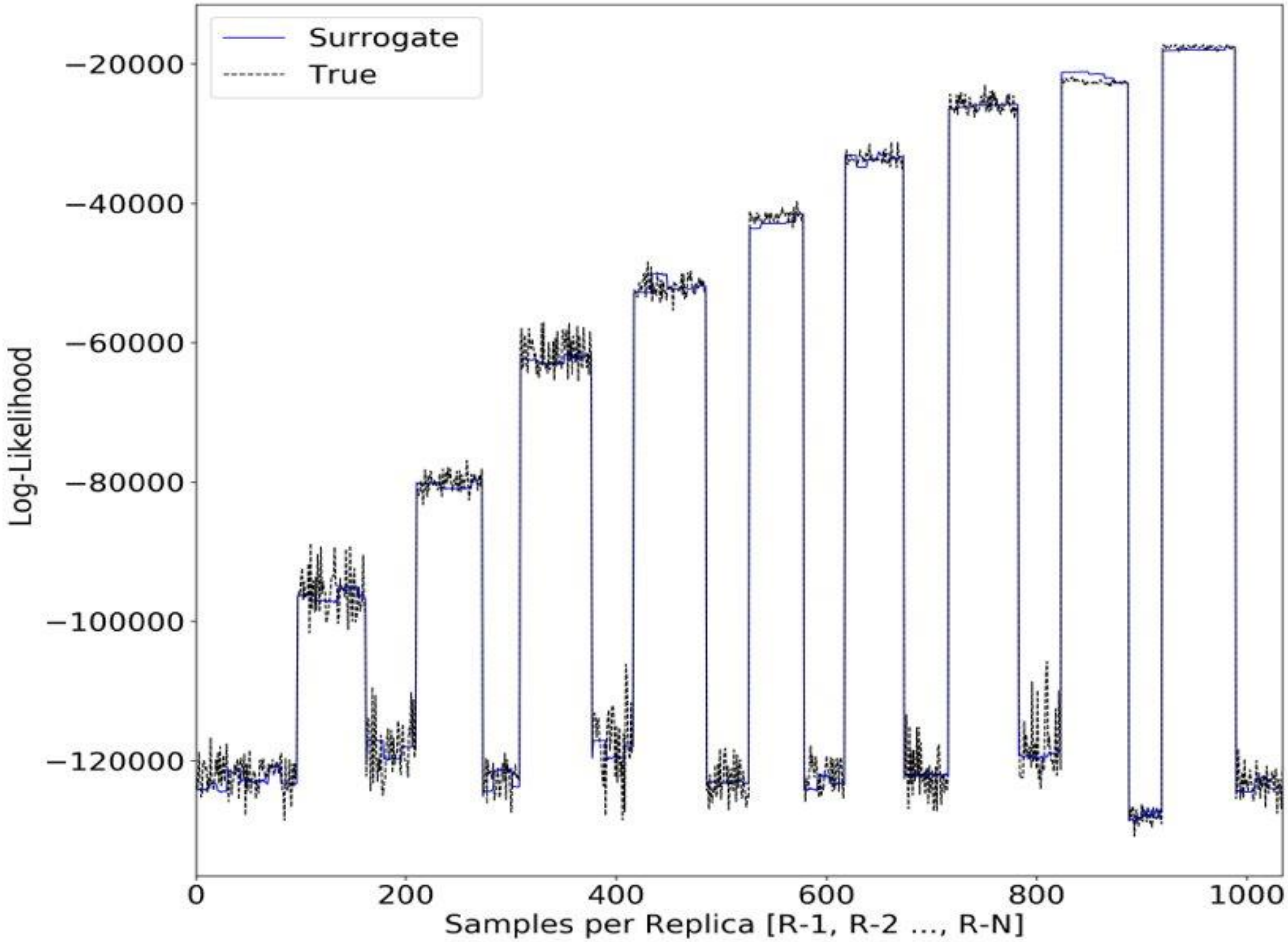


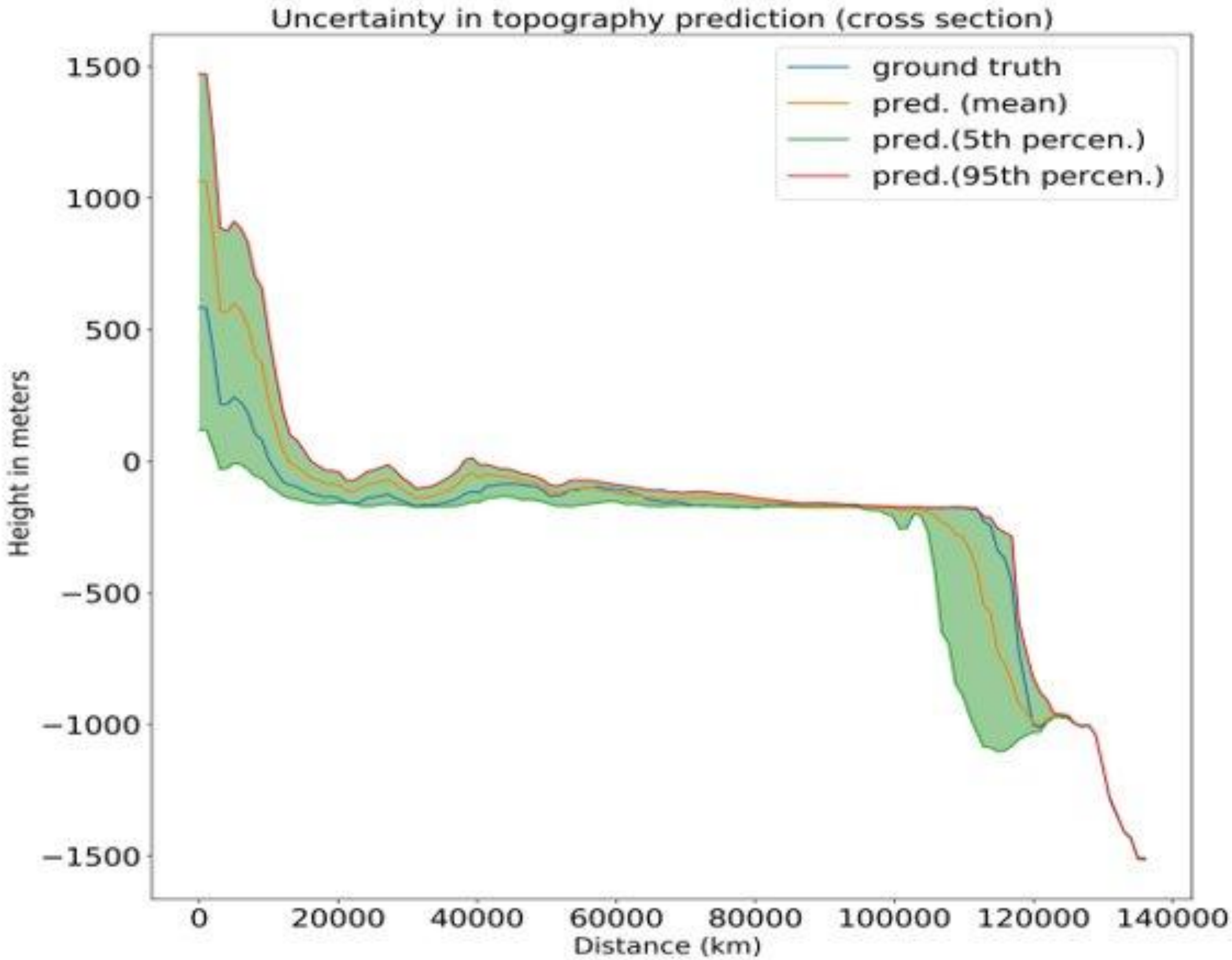


Konark Jain
(Intern, 2018)
Indian Institute of Technology
India

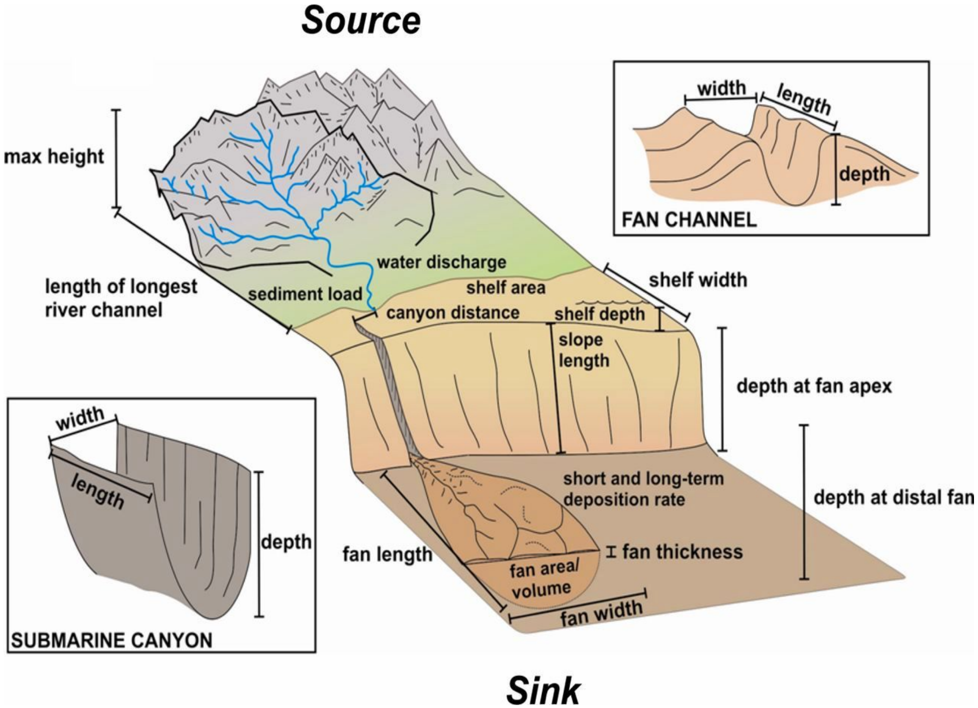
Arpit Kapoor
(Intern, 2018)
SRM Institute of Technology
India

Surrogate accuracy



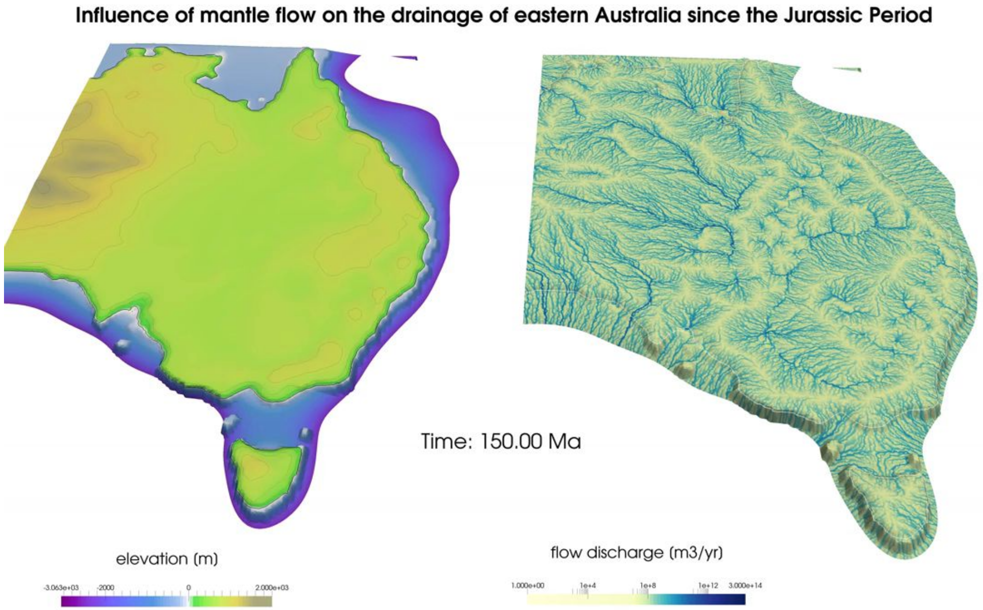

Uncertainty in topography prediction (cross section)
ground truth
pred. (mean)
pred.(5th percen.)
pred.(95th percen.)
1500
1000
500
0
−500
−1000
−1500
Height in meters
0
20000
40000
60000
80000
100000
120000
140000
Distance (km)

- Framework provides a rigorous approach for estimation and uncertainty quantification of key parameters in Badlands model
- Results show that the method provides a means to explore a highly irregular multi-modal parameter space
- Can be applied to large-scale source-to-sink models



Badlands: Large-scale source-to-sink models

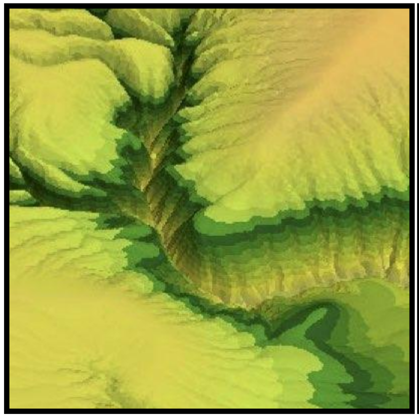
- Modelling of:
  - Sea level change
  - Role of deep Earth mantle convection
  - Coastline evolution
  - Erosion and sedimentation
  - River network evolution
  - Erodibility
  - Rainfall



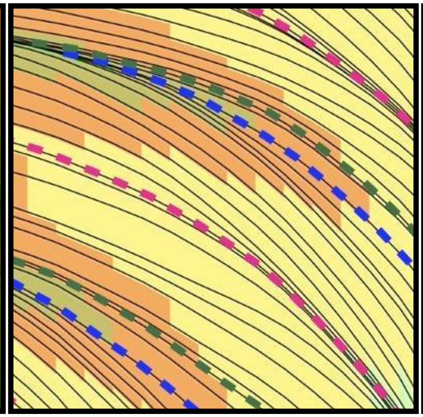
Models by Carmen Braz, Lauren Harrington

Applications to basin modelling

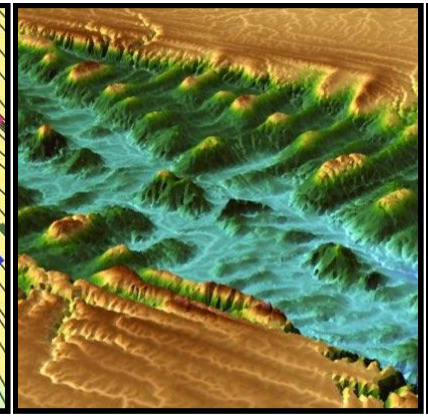
Geomorphology



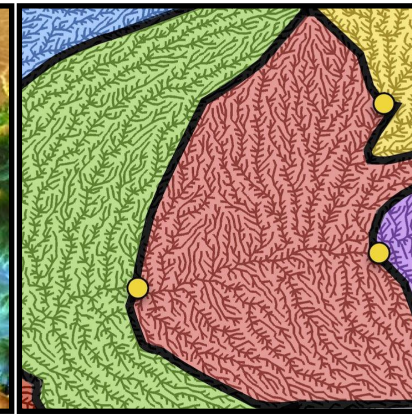
Stratigraphic modelling



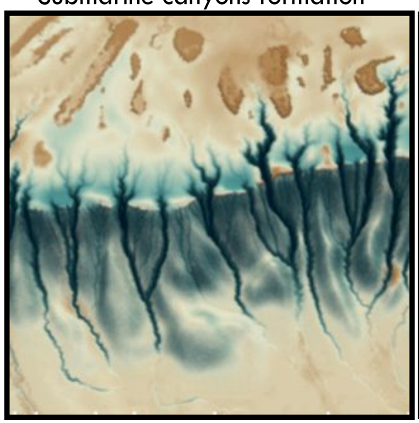
Tectonics & surface processes



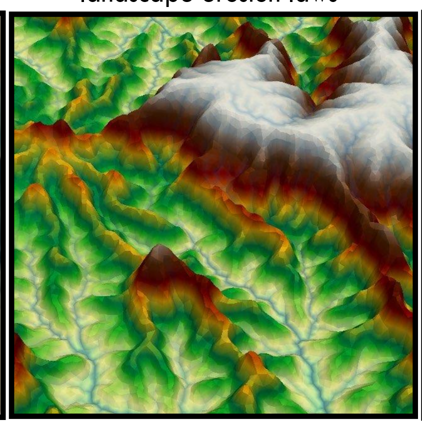
Catchment dynamics



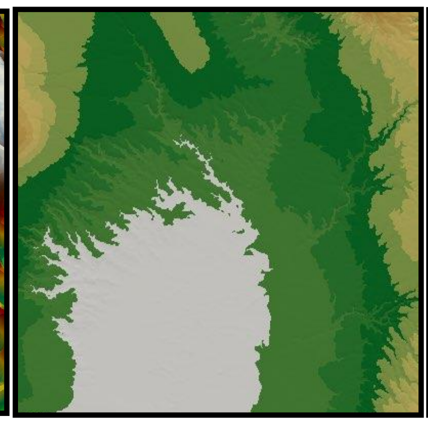
Submarine canyons formation



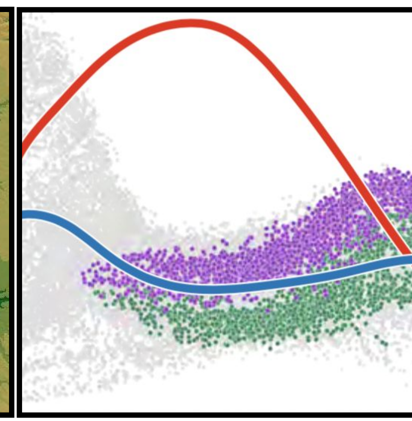
landscape erosion laws



Mantle flow & surface processes

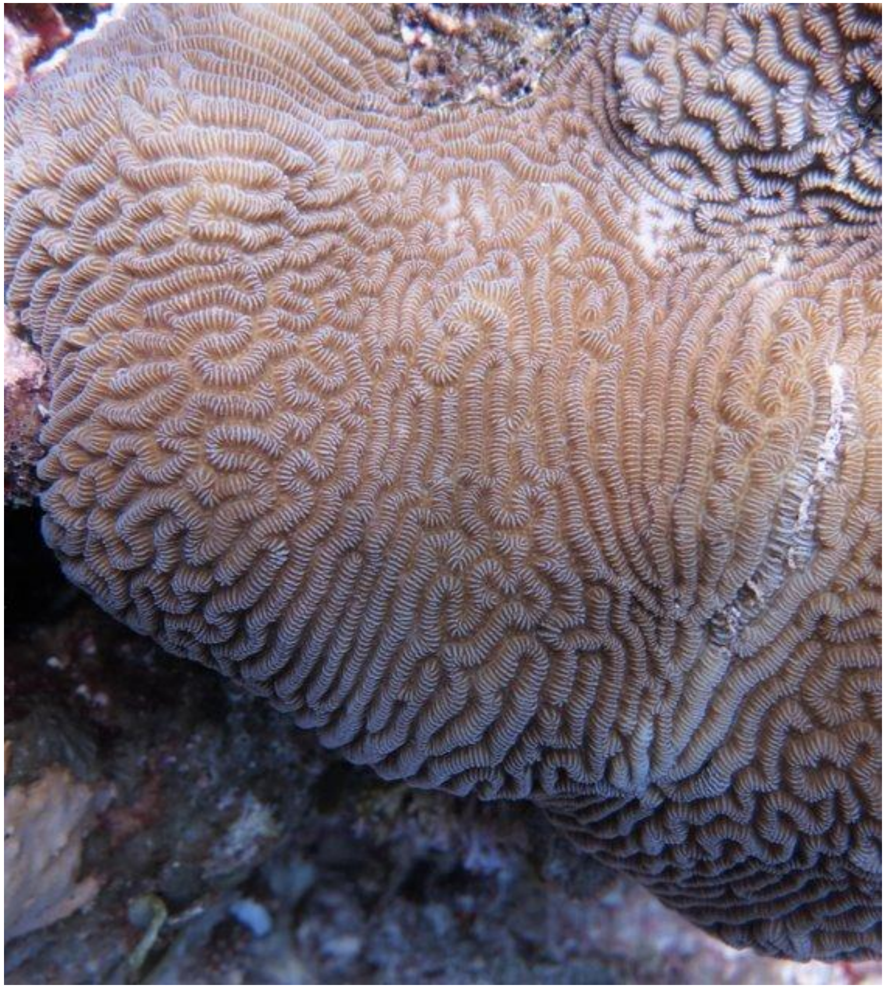


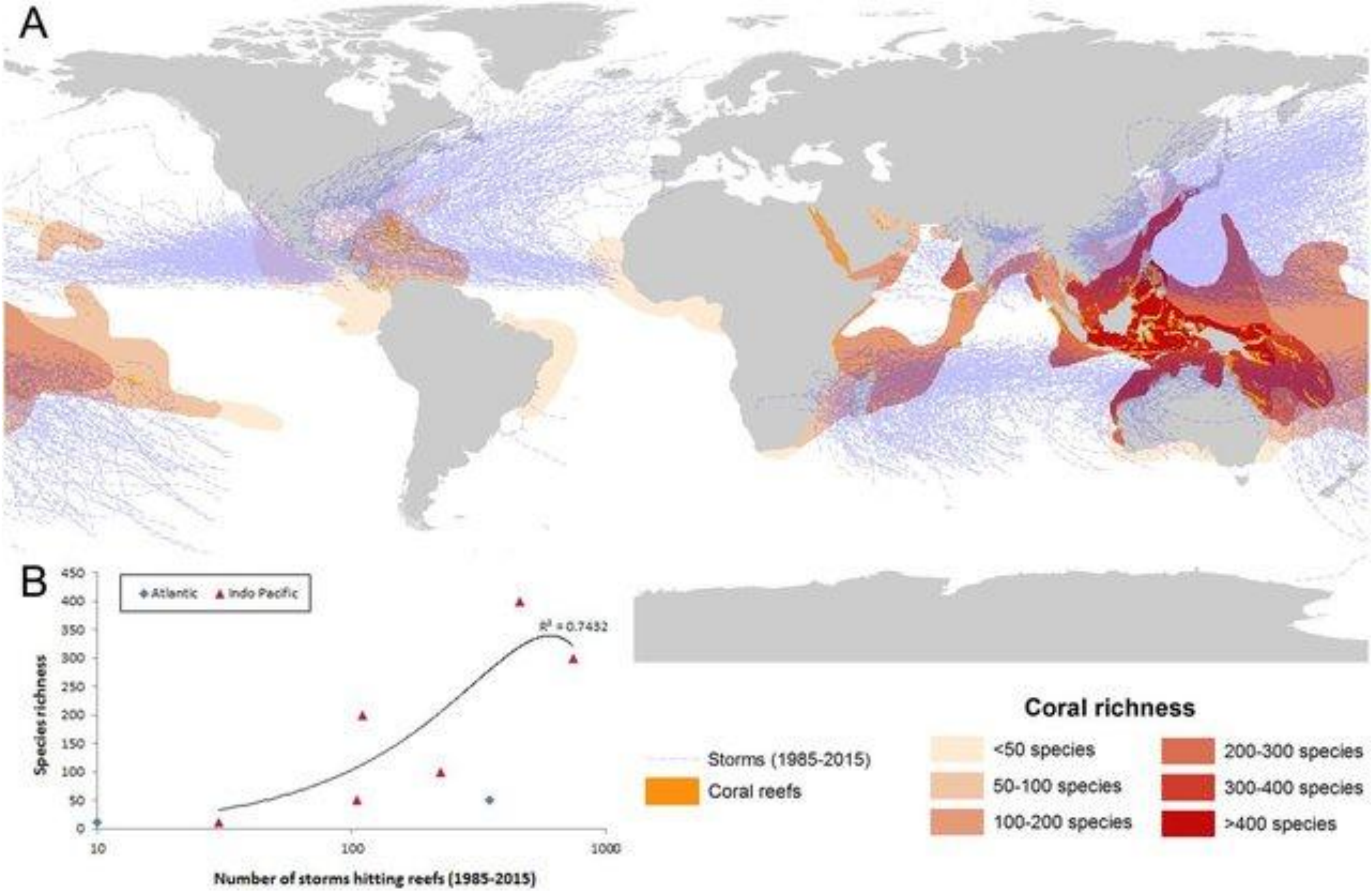
Species richness in mountain range



Reef modelling

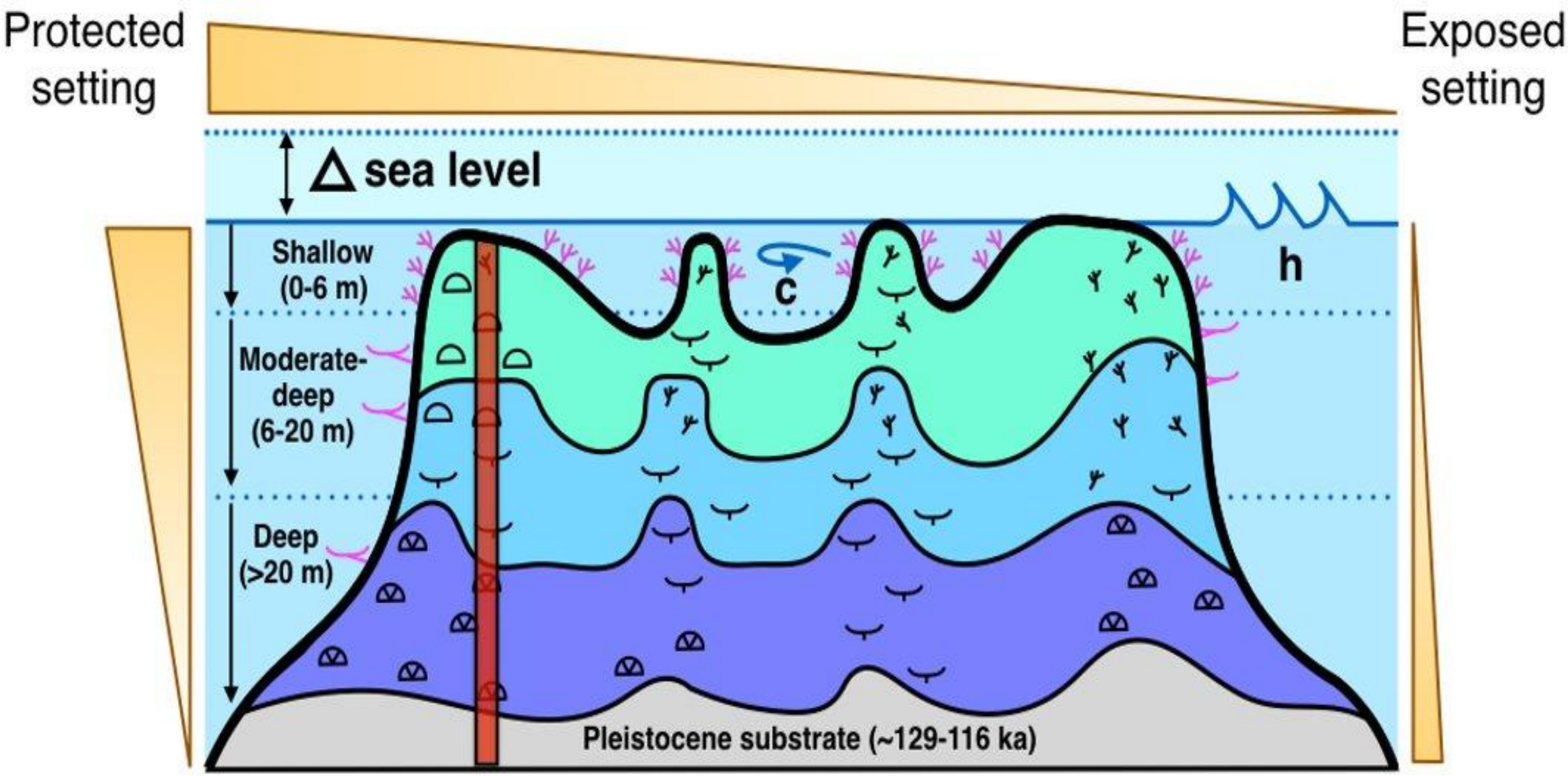
Bayesreef: Bayesian inference for estimation and uncertainty quantification of parameters in geological reef evolution model





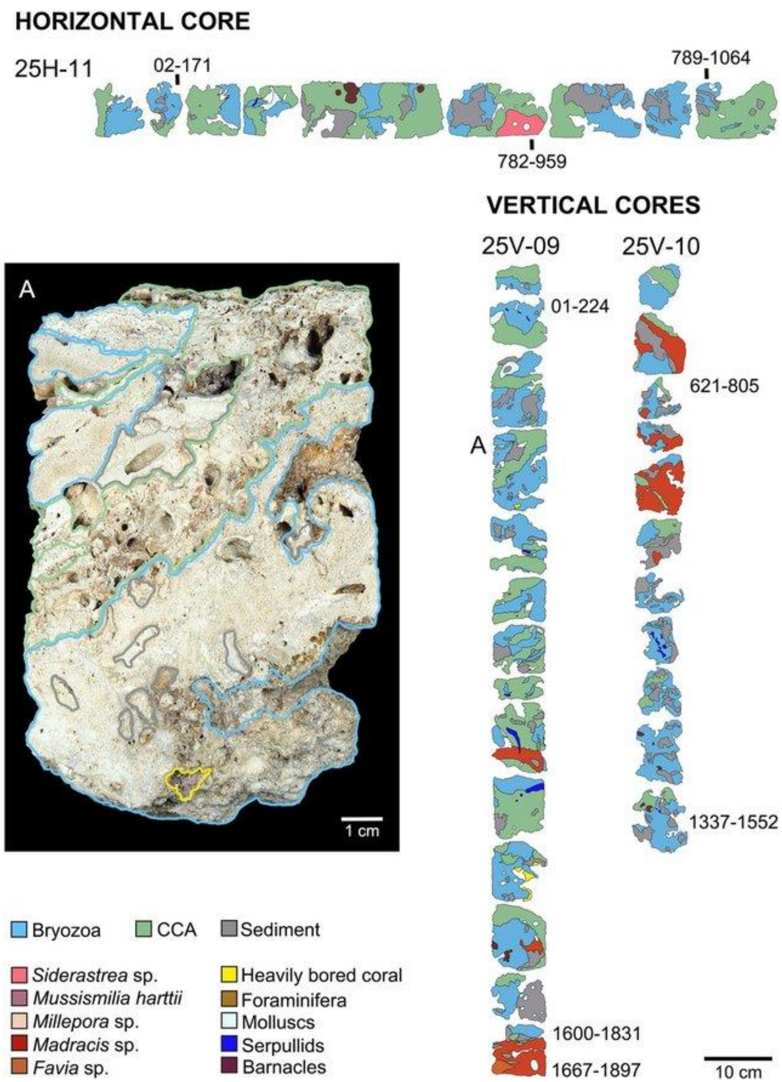
A
B
Atlantic
Indo Pacific
R^2 = 0.7432
Species richness
450
400
350
300
250
200
150
100
50
0
10
100
1000
Number of storms hitting reefs (1985-2015)
Storms (1985-2015)
Coral reefs
Coral richness
<50 species
50-100 species
100-200 species
200-300 species
300-400 species
>400 species

Environmental controls on reef growth



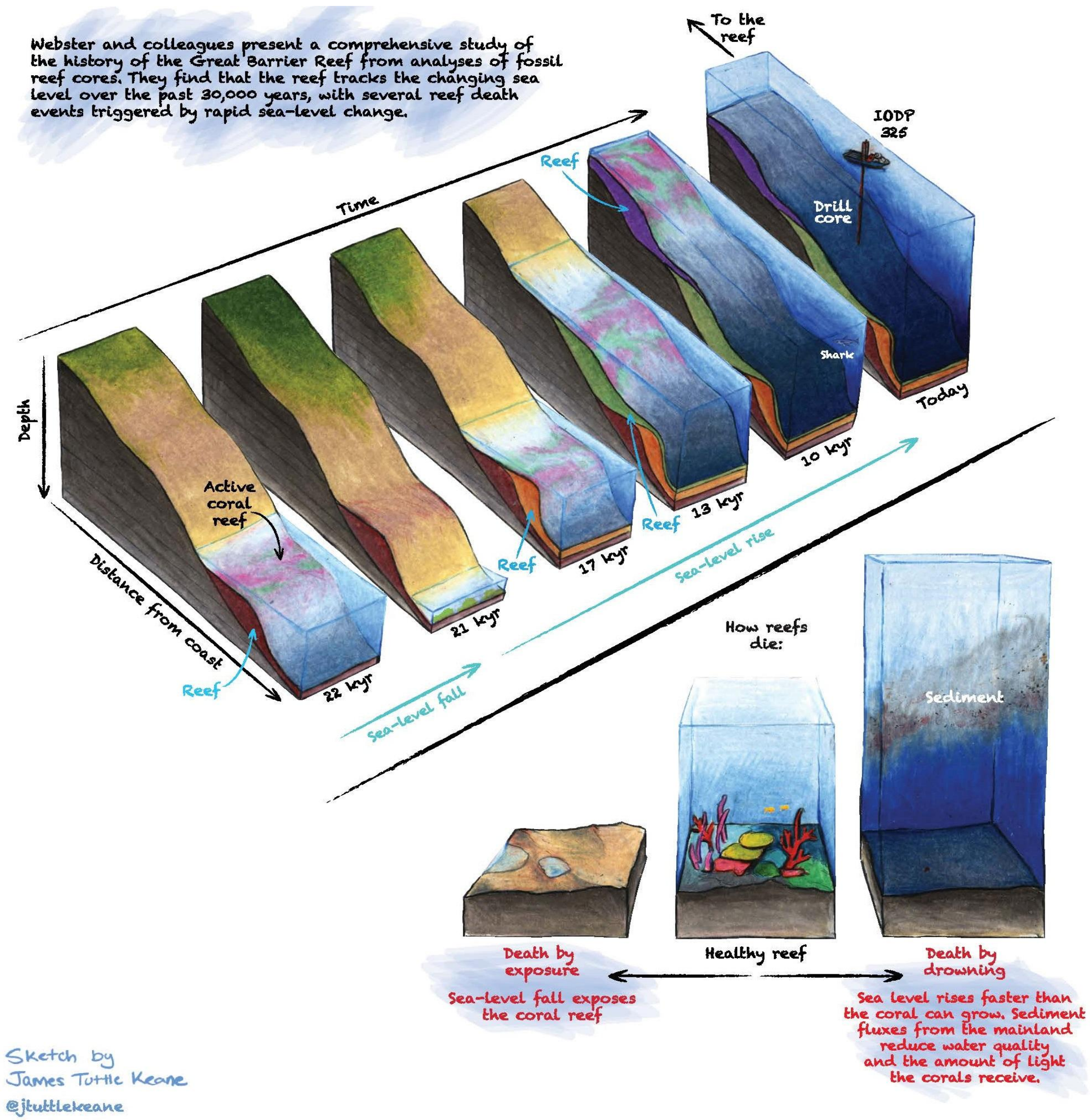

CENTIMETERS
Cat. No. 09-016
Fisher Scientific
INCHES
21'
16'
0'
23

Collaborators

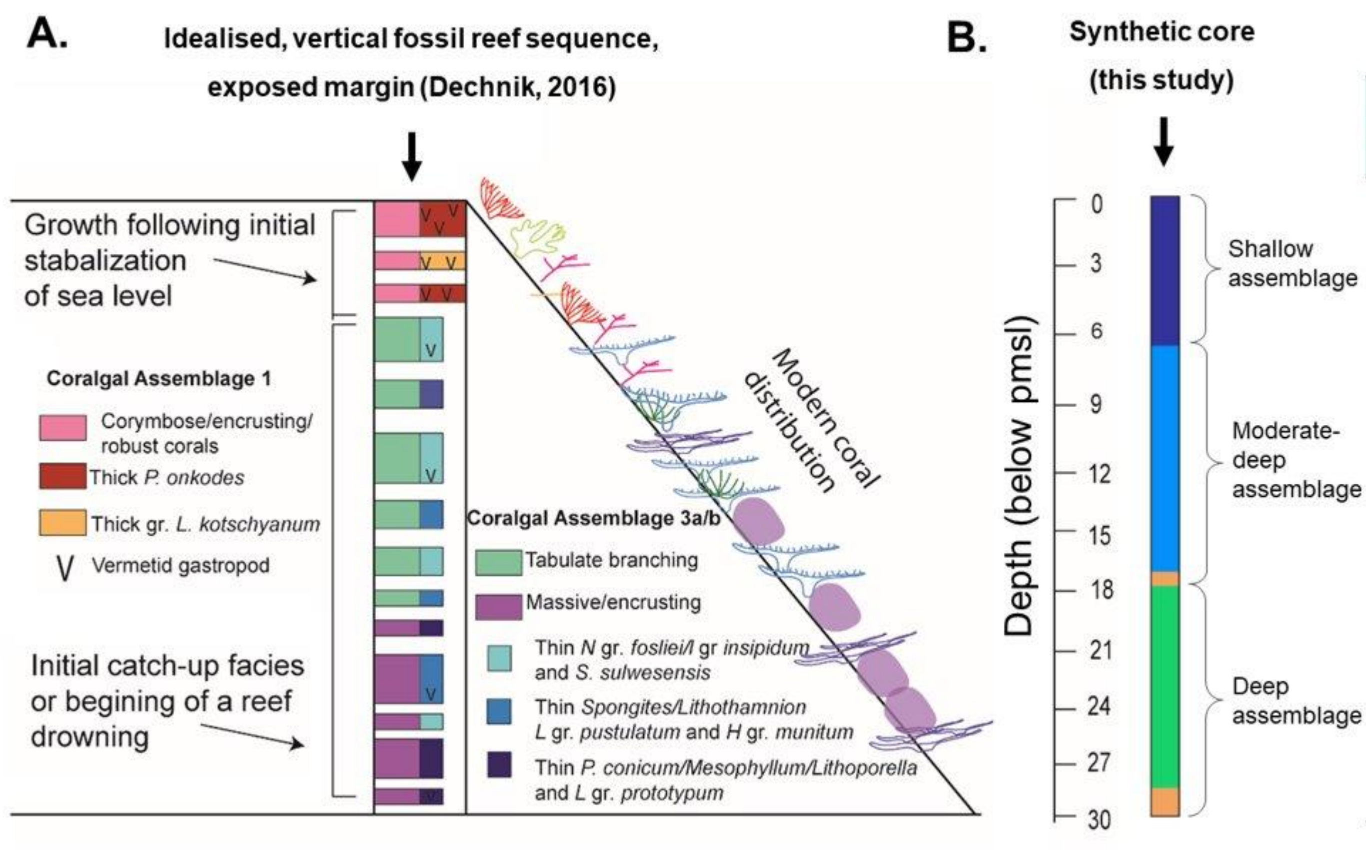


Asso. Prof. Jody Webster
The University of Sydney

Jodie Pall, USyd Medal (Honours)
University of Sydney
Engineer, Sydney Water



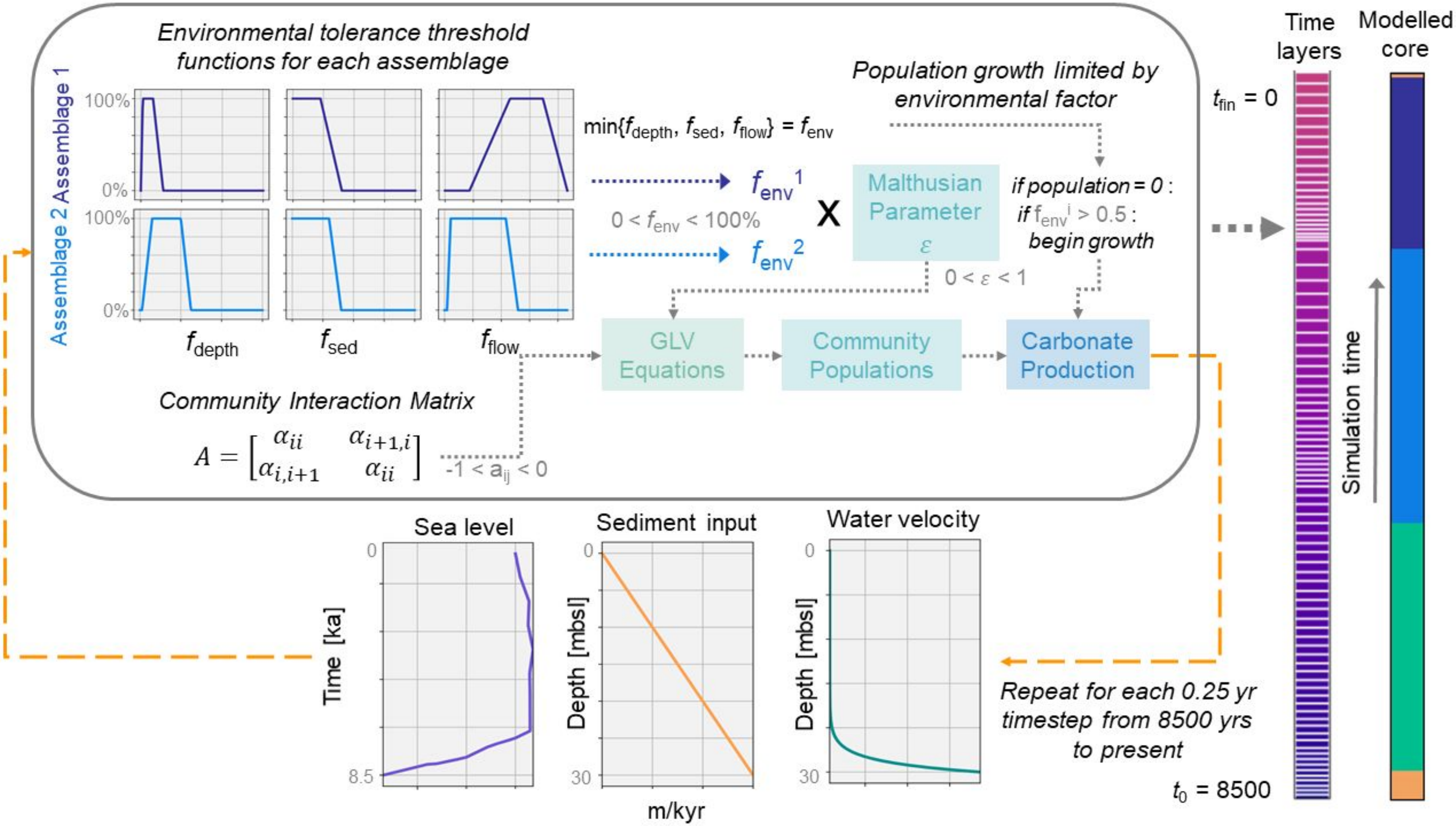
Webster and colleagues present a comprehensive study of the history of the Great Barrier Reef from analyses of fossil reef cores. They find that the reef tracks the changing sea level over the past 30,000 years, with several reef death events triggered by rapid sea-level change.
To the reef
IODP 325
Drill core
Time
Reef
Depth
Active coral reef
Distance from coast
Reef
22 kyr
21 kyr
Reef
17 kyr
Reef
13 kyr
Shark
10 kyr
Today
Sea-level fall
Sea-level rise
How reefs die:
Sediment
Death by exposure
Sea-level fall exposes the coral reef
Healthy reef
Death by drowning
Sea level rises faster than the coral can grow. Sediment fluxes from the mainland reduce water quality and the amount of light the corals receive.
Sketch by James Tuttle Keane
@jtuttlekeane



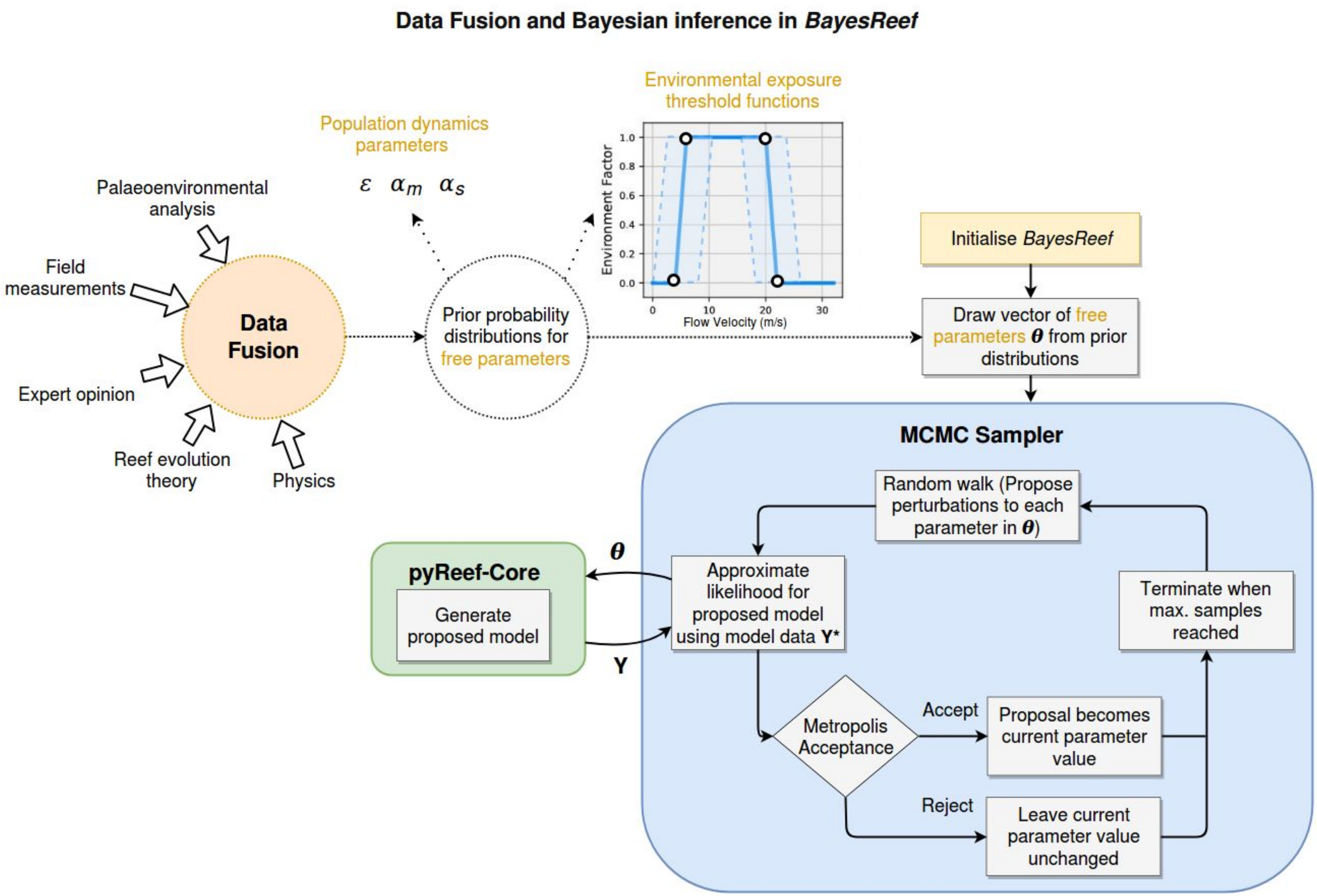
Assigned Numerical IDs

Depth interval [m]	Assemblage	Num. ID
0	Shallow, exposed	**0.143**
...	...	...
...	...	...
...	...	...
4	Shallow, exposed	**0.143**
6	Mod-deep, exposed	**0.286**
...	...	...
...	...	...
...	...	...
...	...	...
...	Mod-deep, exposed	**0.286**
16	Carbonate sediment	**0.581**
18	Deep, exposed	**0.429**
...	...	...
...	...	...
...	...	...
...	...	...
28	Deep, exposed	**0.429**
30	Carbonate sediment	**0.581**

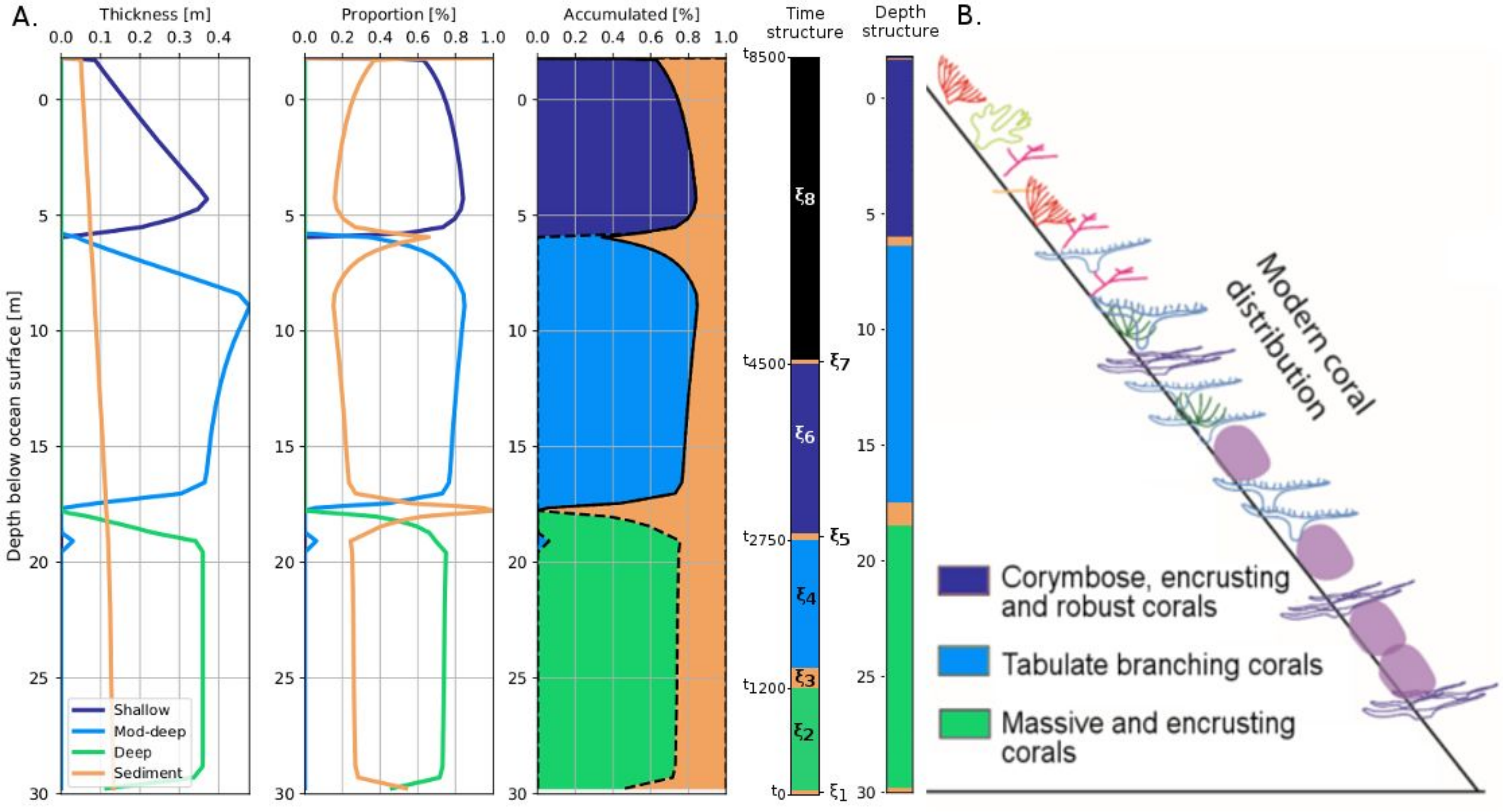
py-Reef-core Model

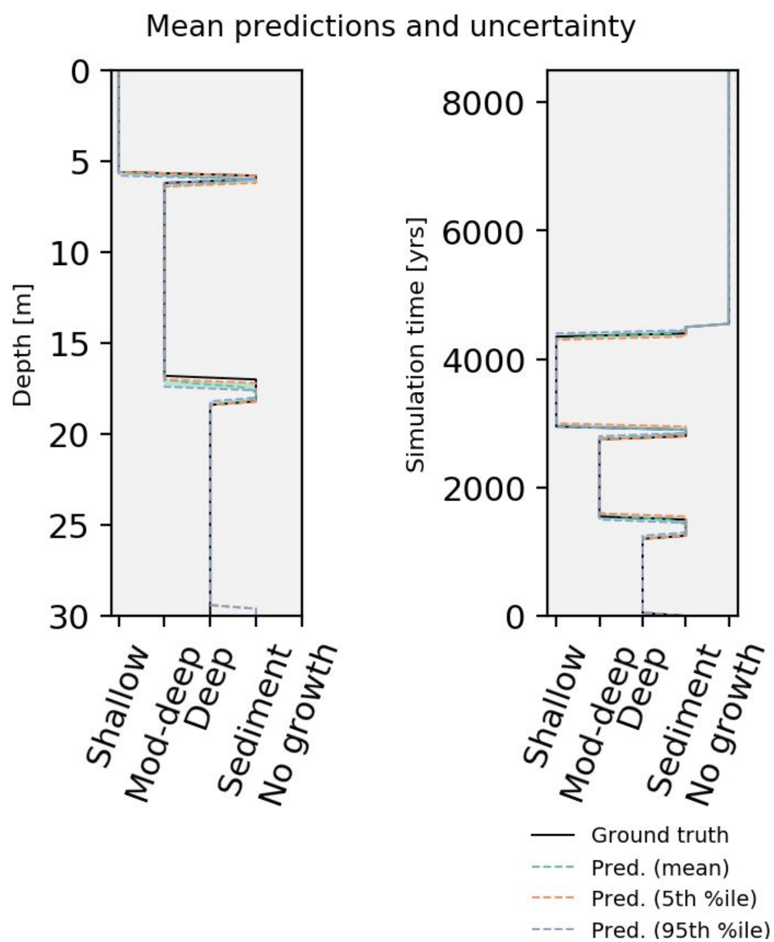


Bayesreef Framework



py-Reef-Core output

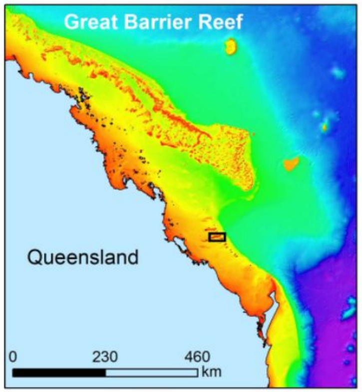


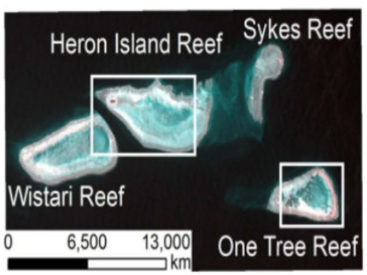


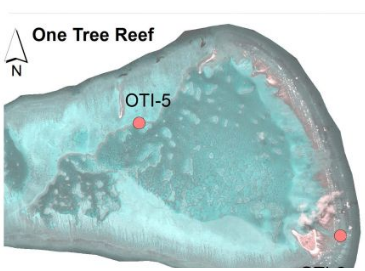
Mean predictions and uncertainty
Depth [m]
0
5
10
15
20
25
30
Shallow
Mod-deep
Deep
Sediment
No growth
Simulation time [yrs]
8000
6000
4000
2000
0
Shallow
Mod-deep
Deep
Sediment
No growth
Ground truth
Pred. (mean)
Pred. (5th %ile)
Pred. (95th %ile)

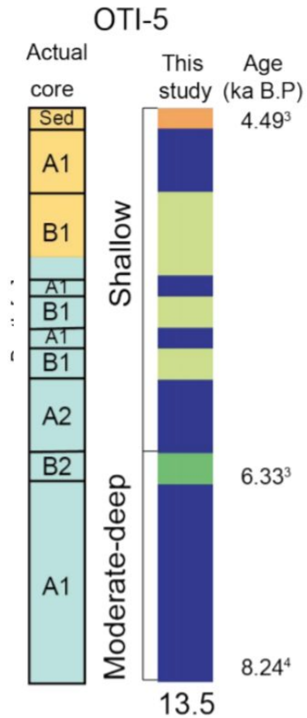
THE UNIVERSITY OF SYDNEY

One Tree Reef core







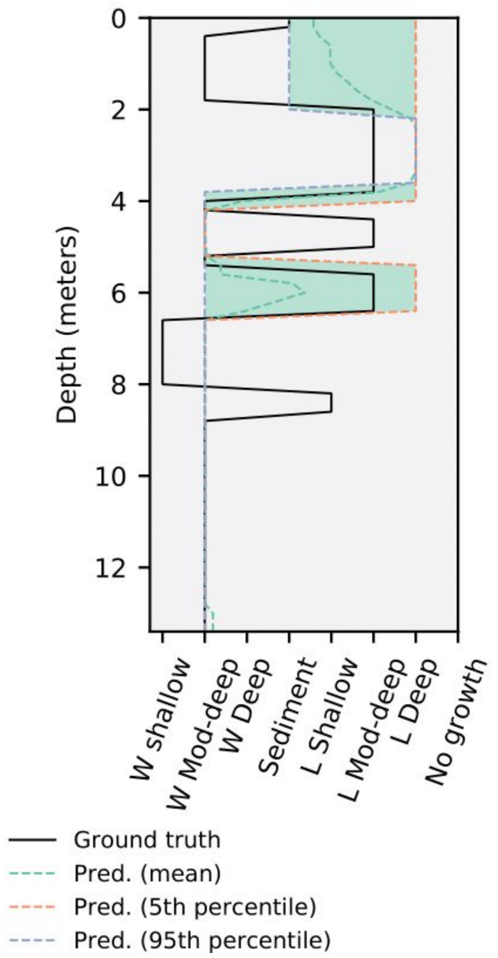


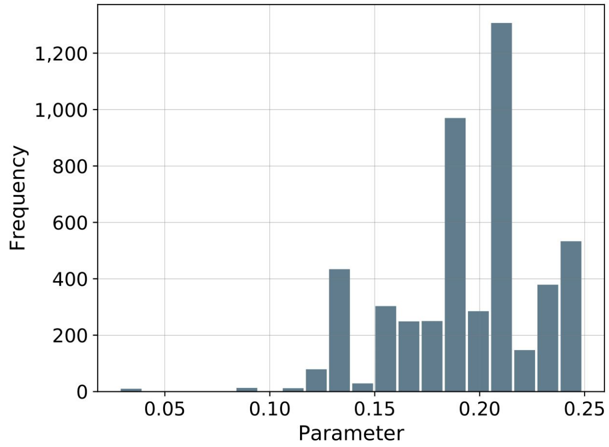
Legend

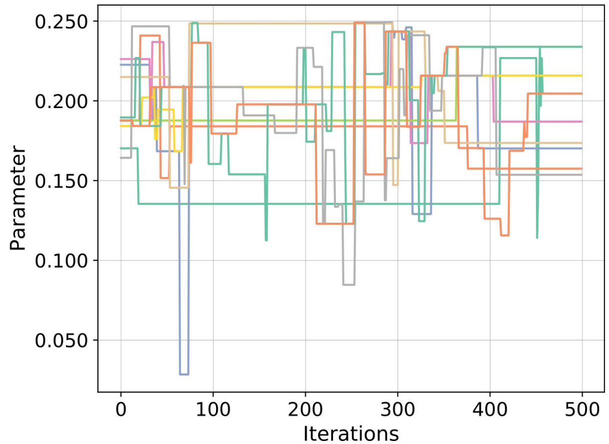
- Carbonate sediment
- Shallow, exposed assemblage
- Deep, exposed assemblage
- Moderate-deep, protected assemblage
- Deep, protected assemblage

- Shallow palaeo-water depths (0-4 m)
- Moderate palaeo-water depths (4-8 m)
- Deep palaeo-water depths (>8 m)

Results

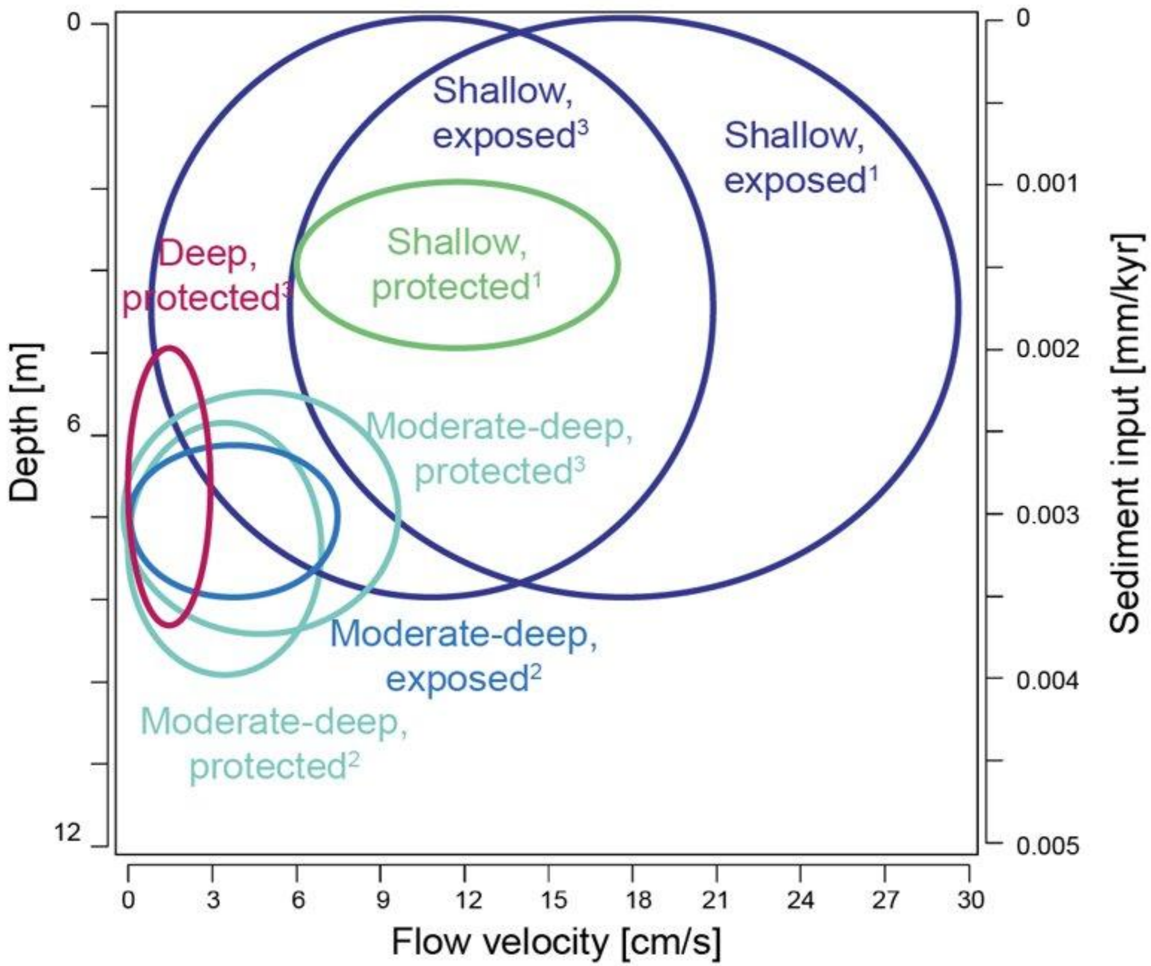






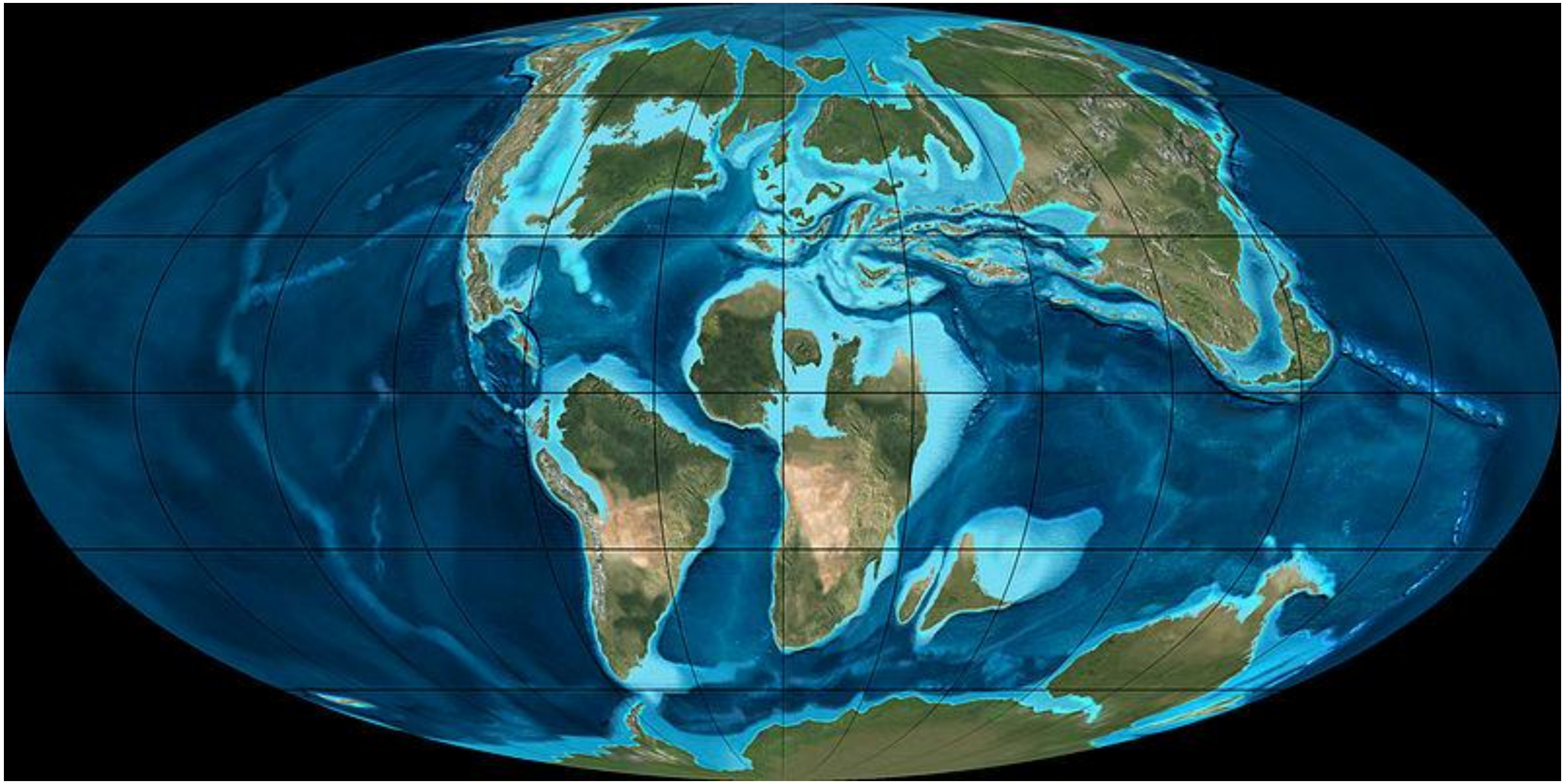
Malthusian parameter

Palaeo-environmental analysis

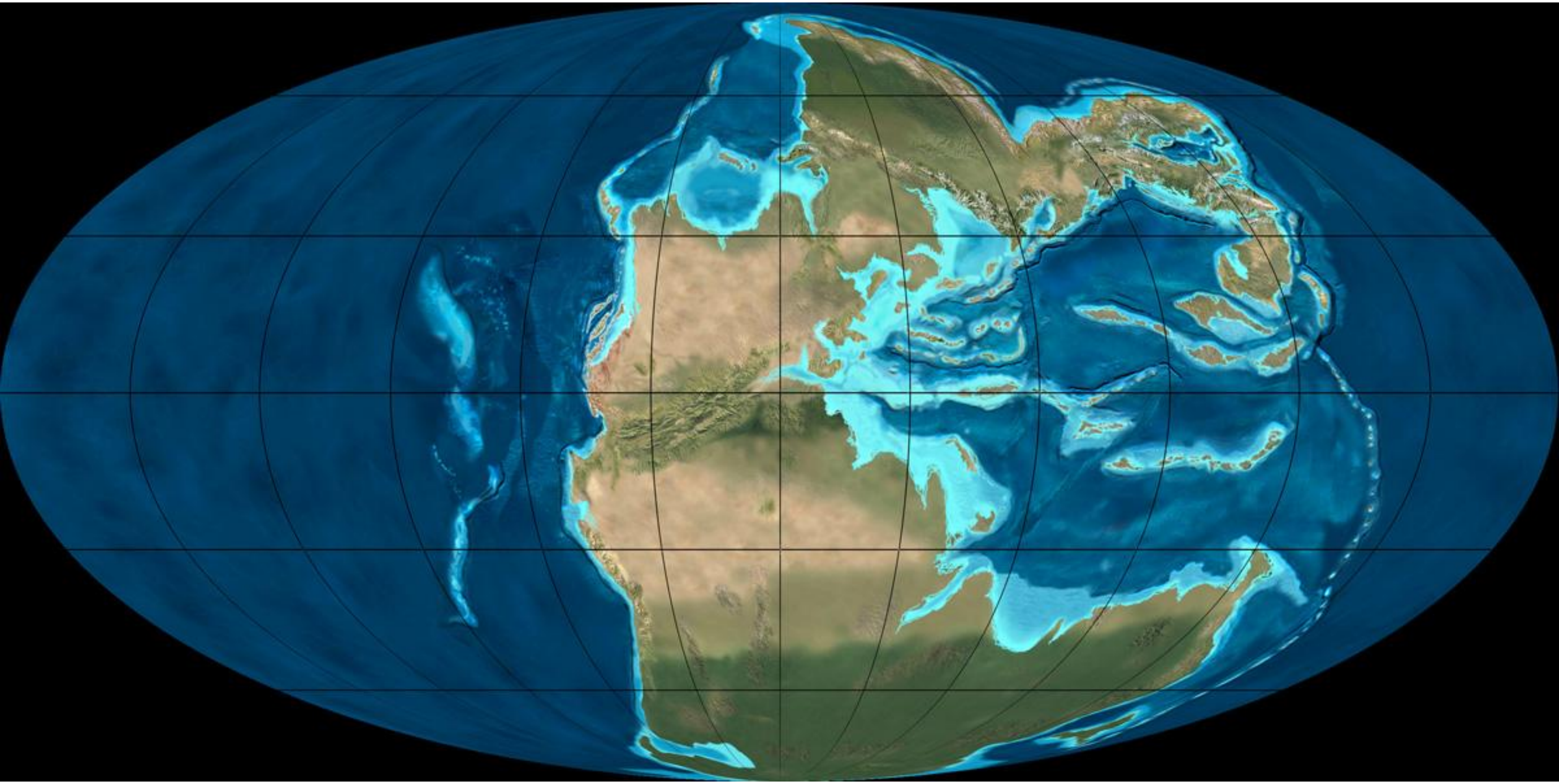


- Shallow, exposed assemblages are relatively insensitive to sediment input and flow velocity.
- Deeper assemblages have tighter environmental niches

Paleoclimate reconstruction

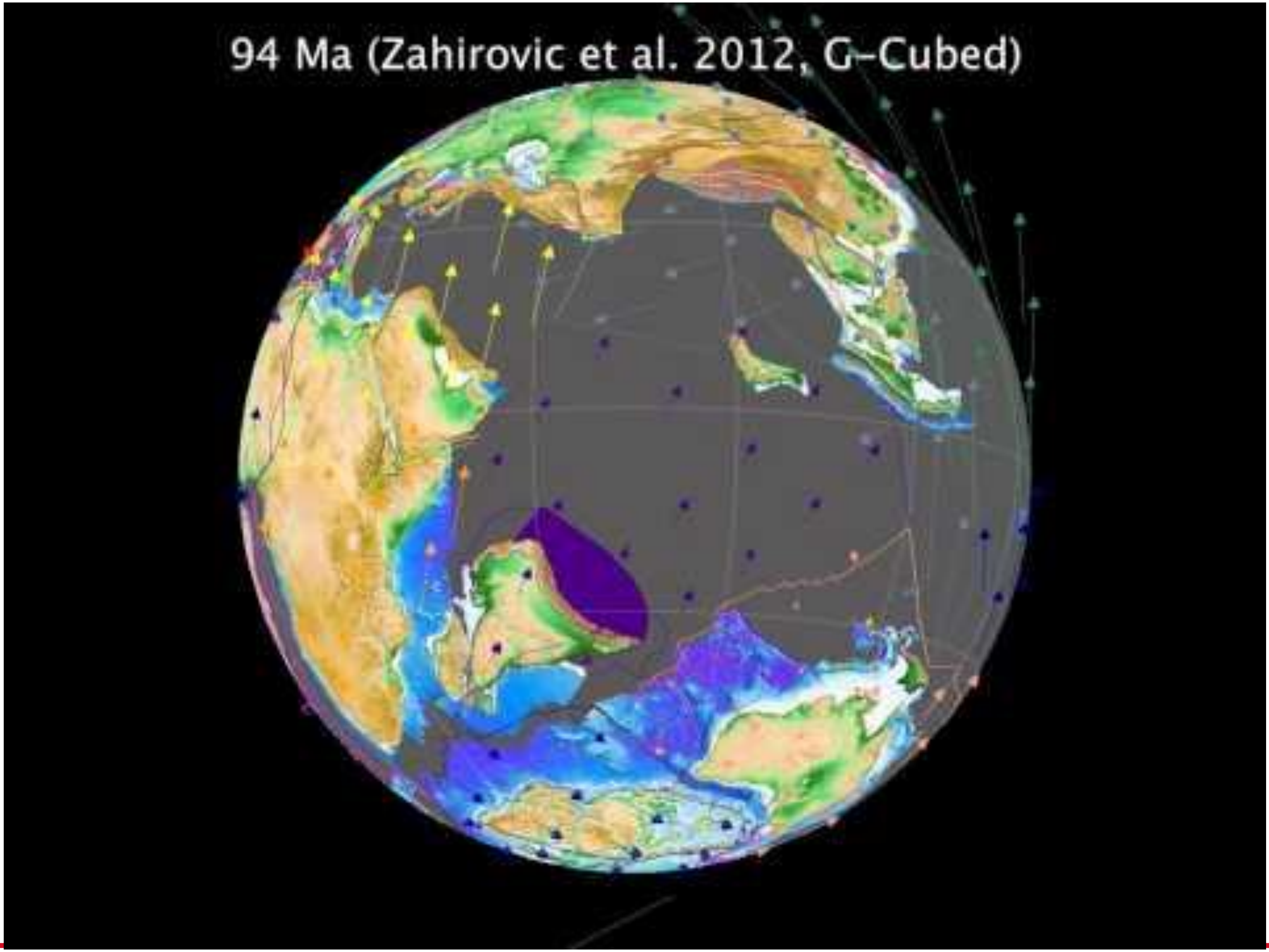


Late Cretaceous: 100 - 66 Million years ago

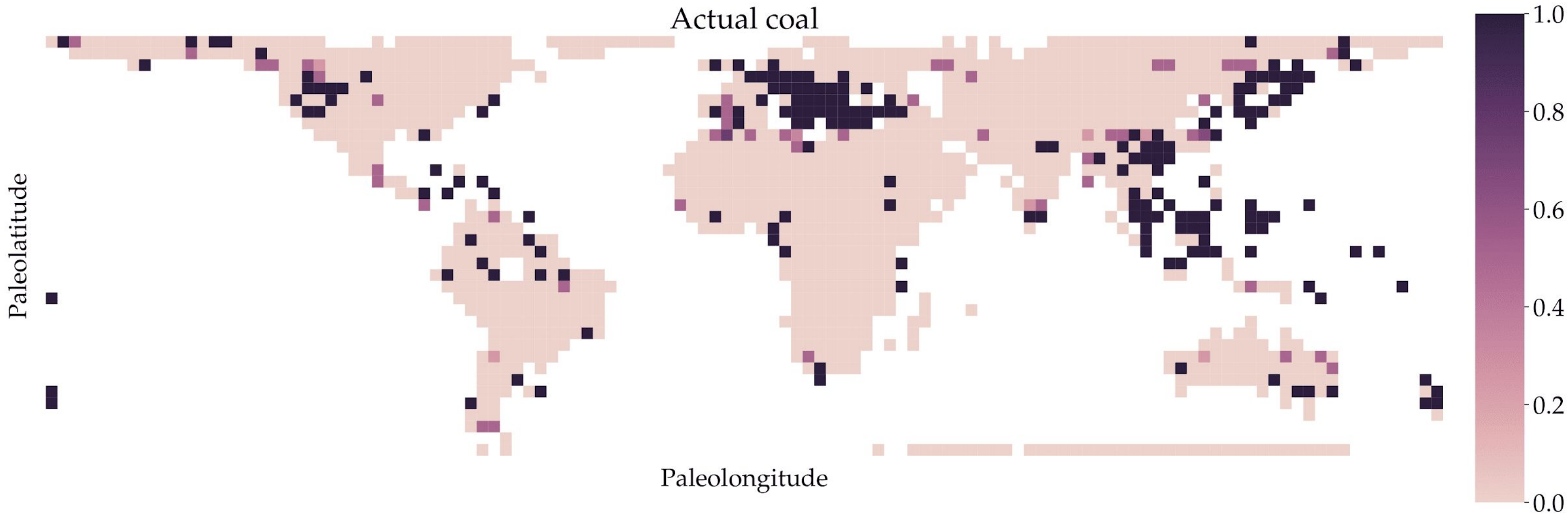


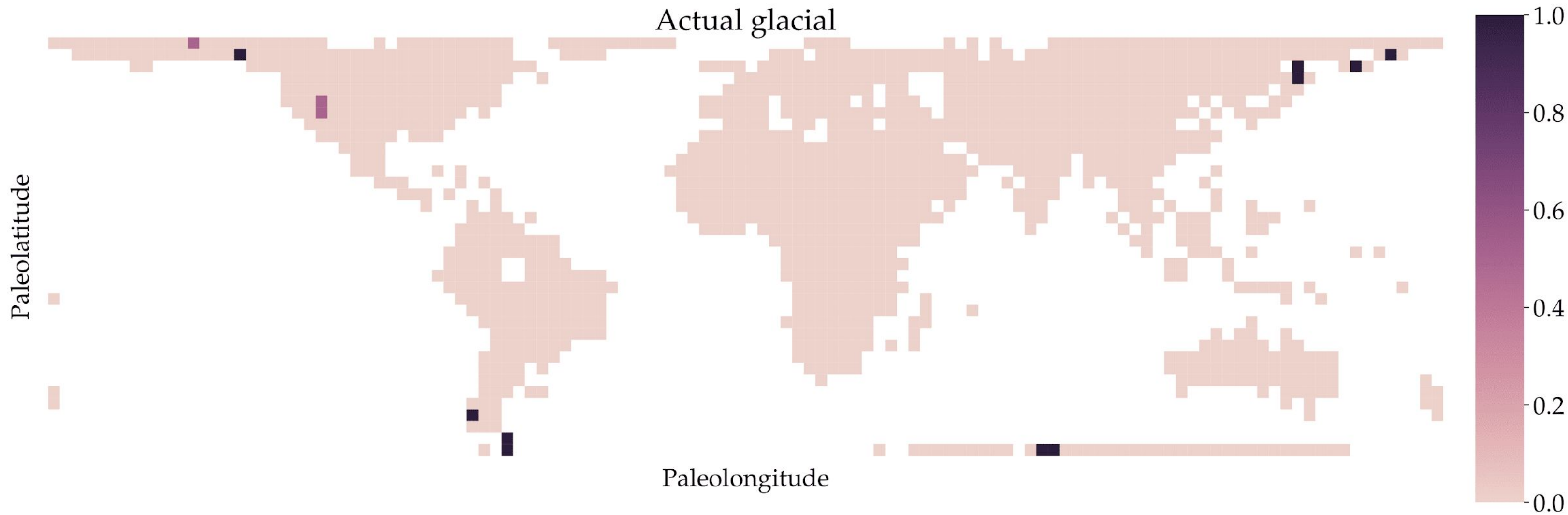
Approx. 240 million years ago

Paleoclimate reconstruction

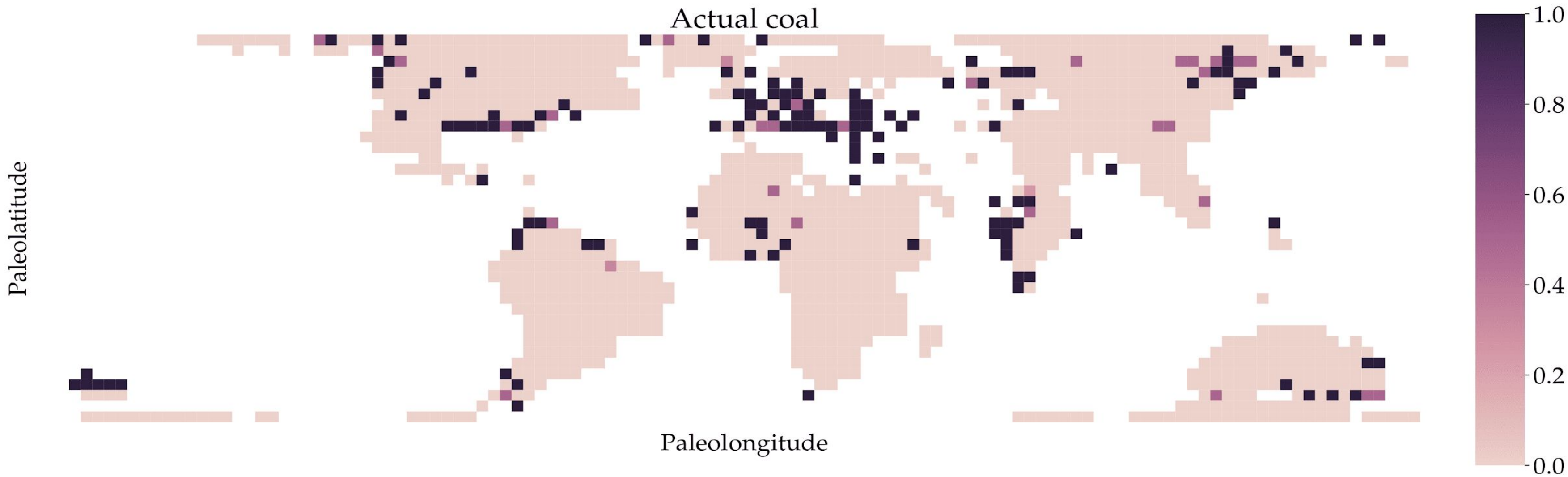


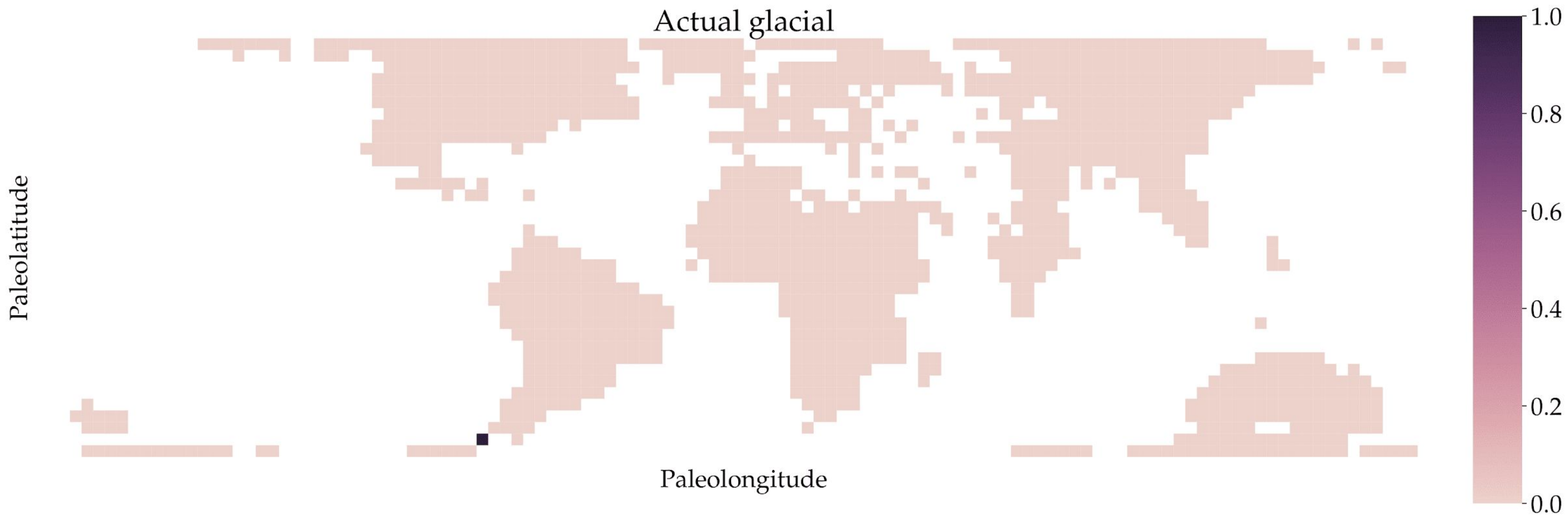
Data - Miocene 14 Ma





Data - 51 Ma


Actual coal
Paleolatitude
Paleolongitude
1.0
0.8
0.6
0.4
0.2
0.0


Actual glacial
Paleolatitude
Paleolongitude
1.0
0.8
0.6
0.4
0.2
0.0

ARC Training Centre for Data Analytics in Minerals and Resources

CTDS has been awarded a $4m ARC Industrial Transformation Training Centre grant to establish an $11m data science centre to support the management of Australia's natural resources.

The Data Analytics for Resources and Environments or DARE Centre will enable Australia's best up and coming researchers to apply their data science models against real-world challenges from water storage to biodiversity to mineral resources (2020-2025).

Resources and Contact

GitHub Organisation Repository:

- https://github.com/sydney-machine-learning
- https://github.com/intelligentEarth/
- https://github.com/EarthByte

Research Outputs:
https://sydney.edu.au/science/people/rohitash.chandra.php