



Australian Government
Geoscience Australia



Monte Carlo (stochastic) inversions - similarities and differences to gradient based methods

Ross C Brodie, Geoscience Australia



Or should that have been?

Probabilistic framework - similarities and differences to the deterministic framework

Direct search methods - similarities and differences to gradient based methods

Outline

- Frameworks: Deterministic and probabilistic
- Methods: Gradient based and direct search
- Compare and contrast
- Example

Declaration

- Many of the illustrations in this presentation are supplied with permission from, and courtesy of, Malcolm Sambridge, Research School of Earth Sciences, Australian National University.

Inversion frameworks

- Deterministic framework
 - often called an optimization approach or problem
 - concerned with finding a single solution
 - usually the global extremum (max or min) of an objective function
- Probabilistic framework
 - often called probabilistic inference, ensemble inference, or Bayesian inference approaches
 - concerned with estimating the statistical distribution of feasible solutions
 - usually finds solutions in the vicinity of the global extremum plus several local extremum of an objective function

Crux of the difference between the frameworks

Deterministic framework

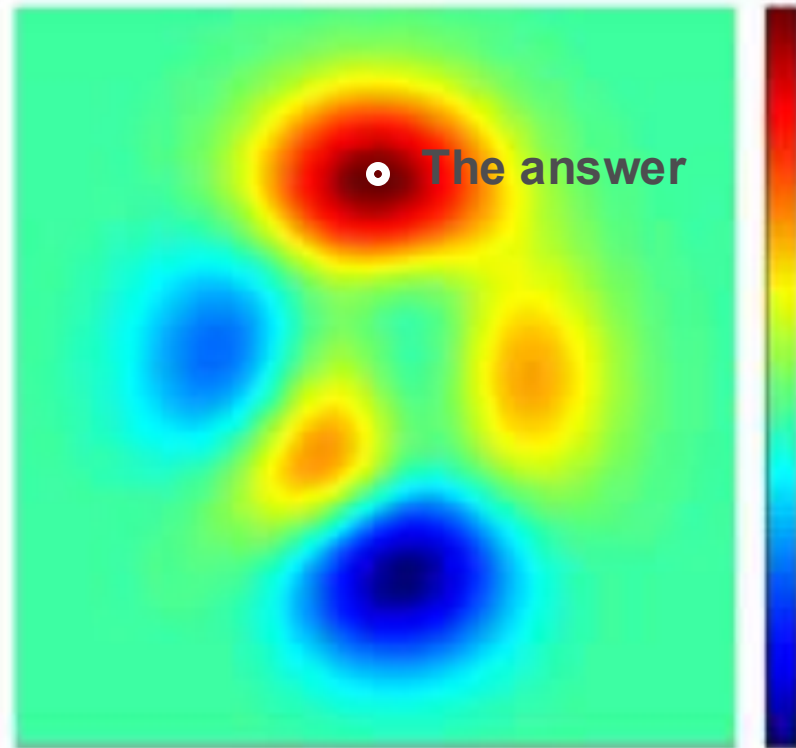
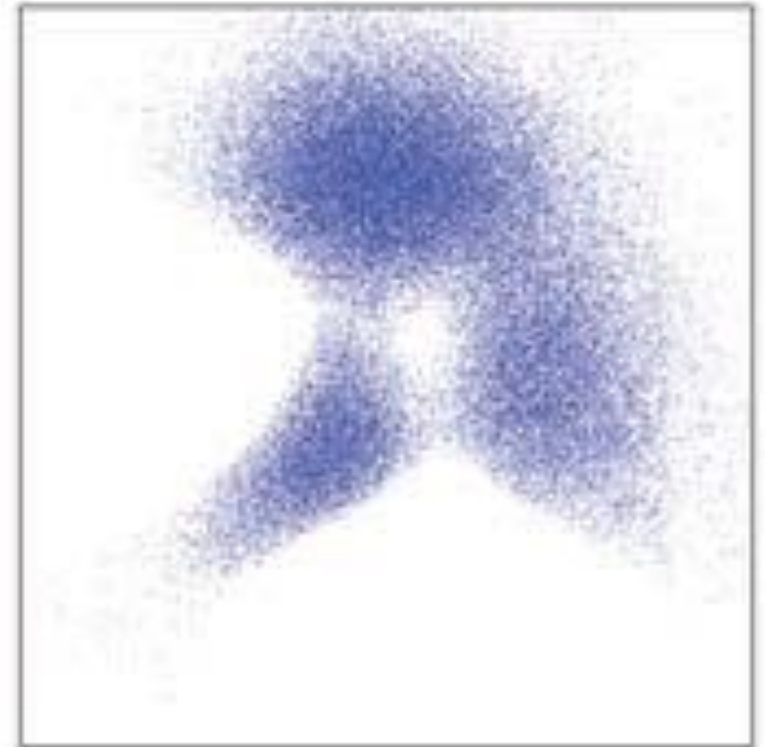


Image of Log Likelihood or
Negative of data misfit

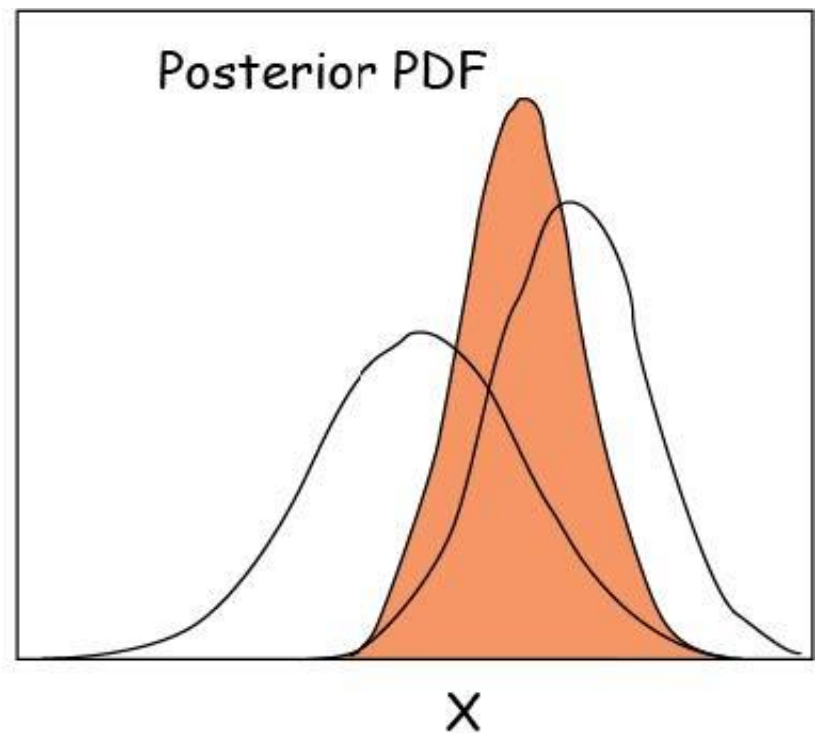
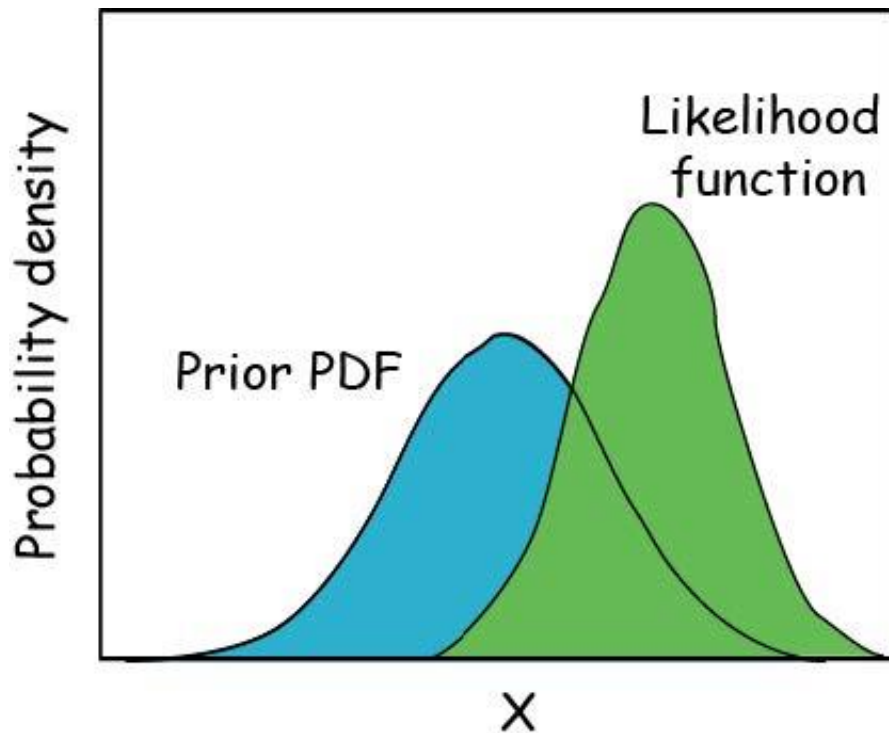
Probabilistic inference



Samples from the Posterior
Probability Density Function

Posterior Probability Density Function (PDF)

- Deterministic framework is concerned with locating, for example, the peak of the posterior PDF histogram
- Probabilistic framework seeks to estimate the whole shape of the posterior PDF histogram



Methods of finding the solution(s)

- Gradient based methods
- Direct search methods
 - Non Monte Carlo
 - Monte Carlo

Gradient based methods

- Basic idea is to traverse down the objective function surface using the slope direction (vector gradient) as guidance on the direction in which to go
- Data forward model must be differentiable (analytically or numerically) w.r.t. the unknown model parameters
- Works for many unknowns (dimensions)
- Only used in the deterministic framework
- Ideal for linear model-data relationships
- Work for weakly non-linear problems by iterative solution and linearizing about the current model... but can fail
- Fails for strongly non-linear problems
- e.g. steepest descent, conjugate gradients, (quasi) Newton methods

Monte Carlo methods

- Involves the use of random numbers and trials
- Not typically suitable for large number of unknowns
- Favoured method of sampling an unknown Posterior PDF
- Used in the deterministic framework
 - Uniform random search, Simulated annealing, Evolutionary and Genetic algorithms (GA)
- Also used in the probabilistic inference framework
 - Neighbour algorithm (NA)
 - Markov Chain Monte Carlo (McMC)
 - Reversible-jump Markov Chain Monte Carlo (rj- McMC)

Crux of the difference between methods

Gradient based method

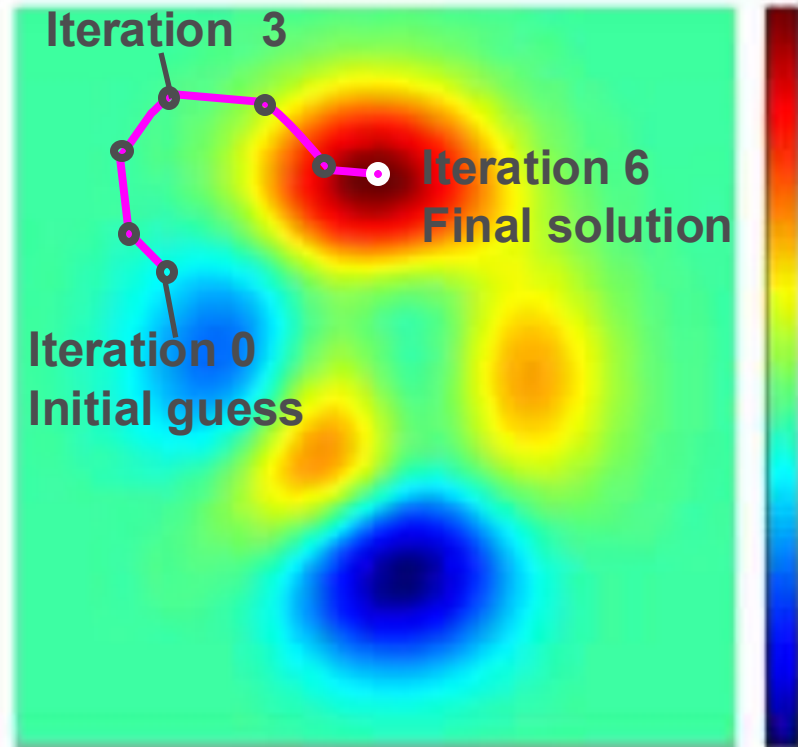


Image of Log Likelihood or Negative of data misfit

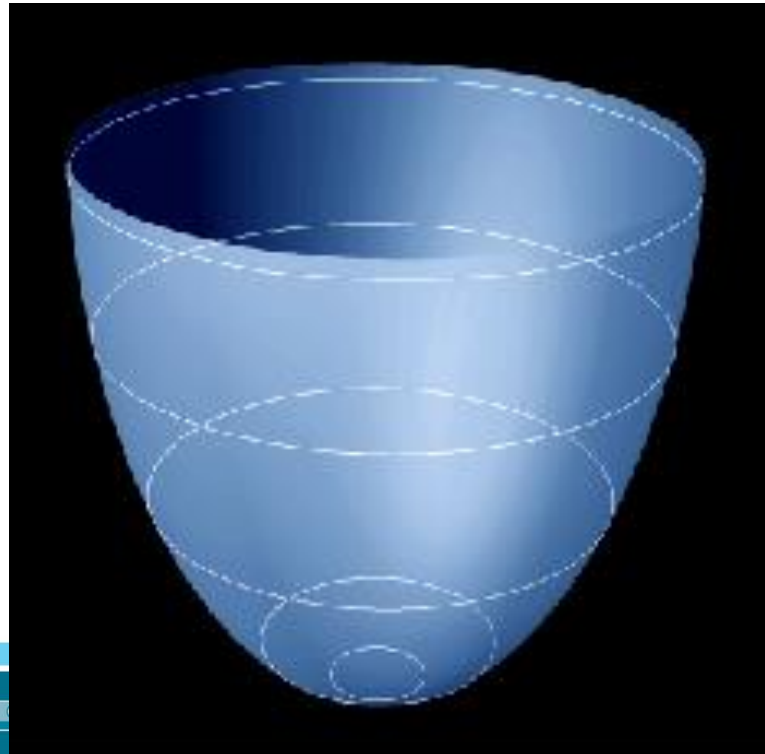
Monte Carlo sampling



Samples from the Posterior Probability Density Function

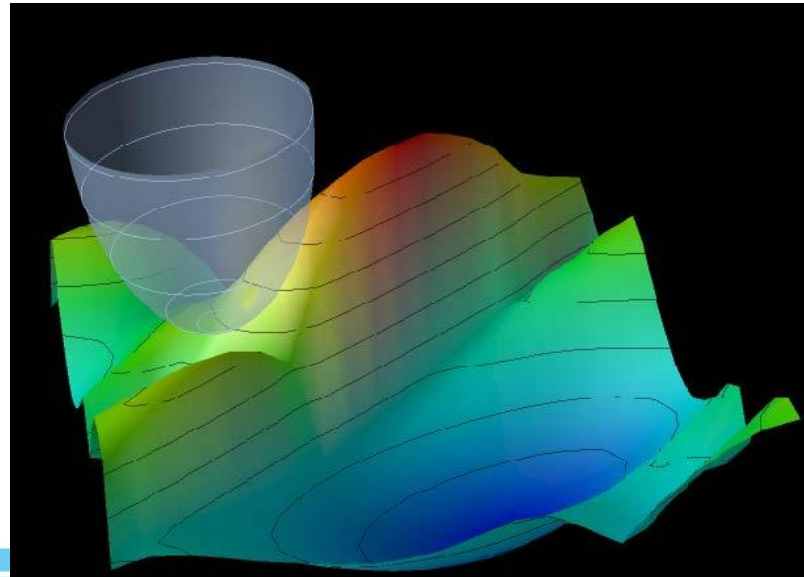
Linear data-model relationship

- Quadratic objective function surface
- Single minimum
- Gradient based methods are the most suitable solution method
- Monte Carlo methods are not justified



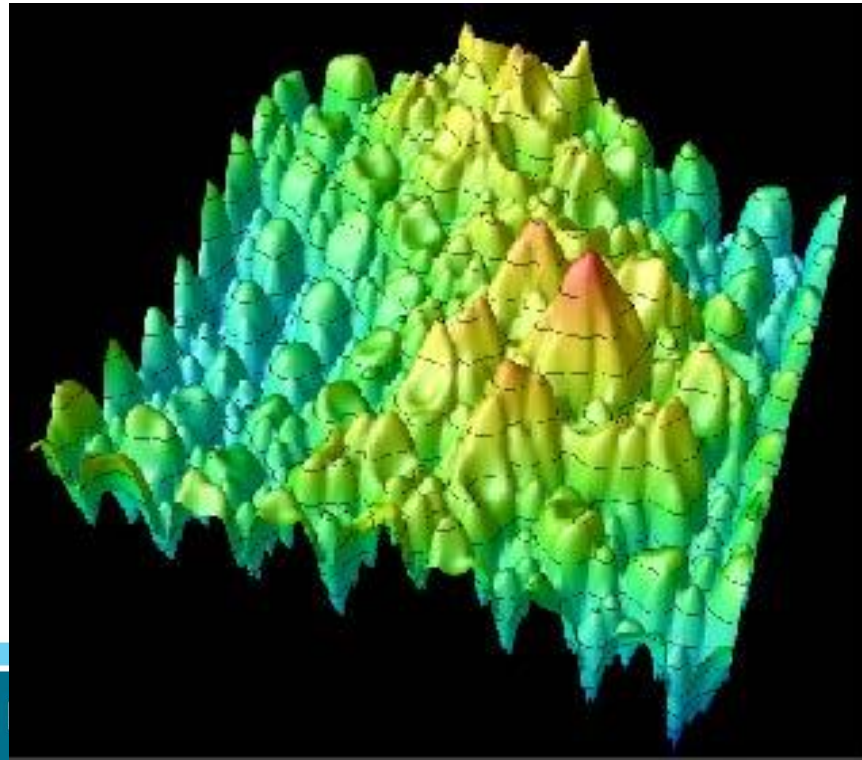
Weakly non-linear data-model relationship

- More complicated objective function surface
- Multiple minima
- Gradient methods are possible by linearizing the problem
- Gradient methods can get stuck in local minima, whereas Monte Carlo methods can climb out of local minima and explore multiple feasible regions



Strongly non-linear data-model relationship

- Extremely complicated objective function surface
- Multiple minima
- Gradient based methods are useless
- Monte Carlo methods can give good results

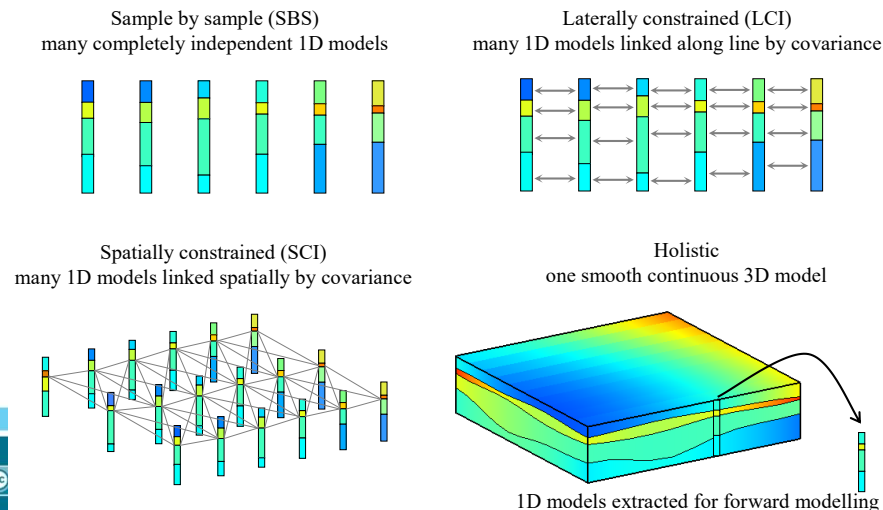


Derivatives

- Gradient based methods require that the data forward model must be differentiable (analytically or numerically) with respect to the unknown model parameters.
- Monte Carlo methods do not in general require derivatives
- Monte Carlo methods can thus usually be applied to more complex data-model relationships and it is easier to use more sophisticated objective functions
- Monte Carlo do not require large memory matrix inversions or solvers
- Monte Carlo methods are often easier to code up

Number of unknowns

- Gradient based methods can handle millions of unknown parameters to be solved for
- The amount of work to be done in Monte Carlo methods increases with the power of the number unknowns
- Monte Carlo methods are thus usually only suited to problems with less than 10s or perhaps 100s of unknowns
- What about software like GeoModeller? ... it is more a model refinement code than a full parameter search algorithm



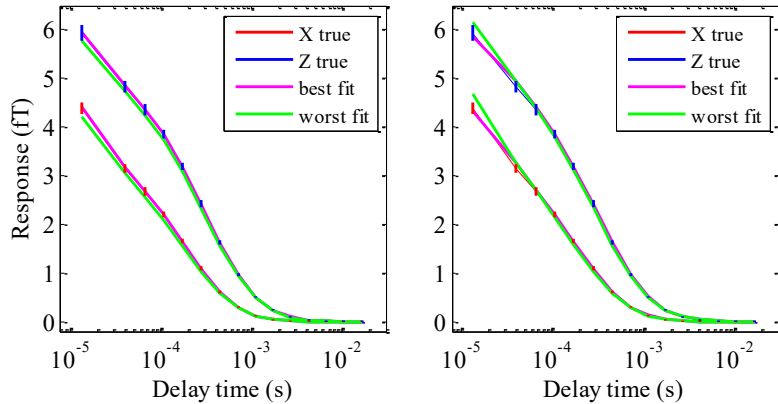
Computation time

- Monte Carlo methods require vastly more computation time than gradient based methods for optimization problems
- This is even more pronounced for the probabilistic inference framework where we wish to properly sample the posterior PDF
- For example, the deterministic gradient based inversion of a single 1D time domain AEM sounding may take ~3 seconds but may take about 30 minutes or more to perform, a full probabilistic inference rj-McMC inversion

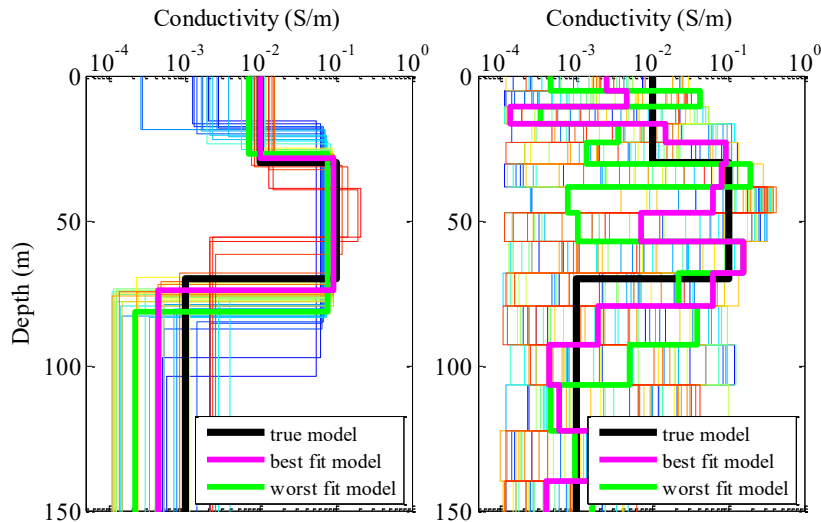
Uncertainty

- There are good methods for estimating uncertainty, trade-offs and resolution for gradient based linear problems - e.g., from the posterior model covariance and resolution matrices
- These can also be applied to weakly non-linear problems but the uncertainty estimates will not necessarily be accurate
- Uncertainty estimates are biased by the regularization that is applied and are likely to be overly optimistic
- Probabilistic inference methods provide a more robust means of estimating uncertainty and non-uniqueness than deterministic methods

Non-uniqueness



A small subset of 3 (left) and 15 layer models (right) that all fit the synthetic data within the prescribed noise levels



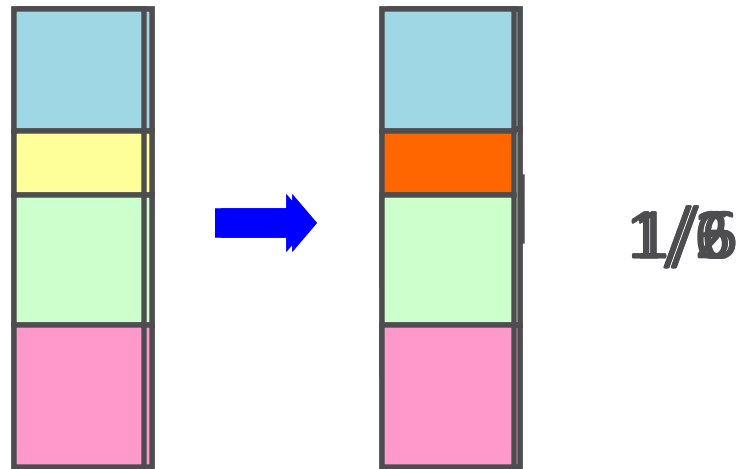
rj-McMC AEM inversion outline

- Reversible-jump Markov chain Monte Carlo (rj-McMC)
- Generate an ensemble of 100,000s of models that fit the data - i.e., not just a single model as in deterministic inversion
- Reversible-jump (or transdimensional) means that the algorithm automatically changes the number of layers
- The ensemble is generated via Markov chain importance sampling
- New models are proposed by randomly perturbing the last link (model) in the chain

rw-McMC inversion - proposal types

- Four types of model update may be proposed at each link in the chain

combinatorial change

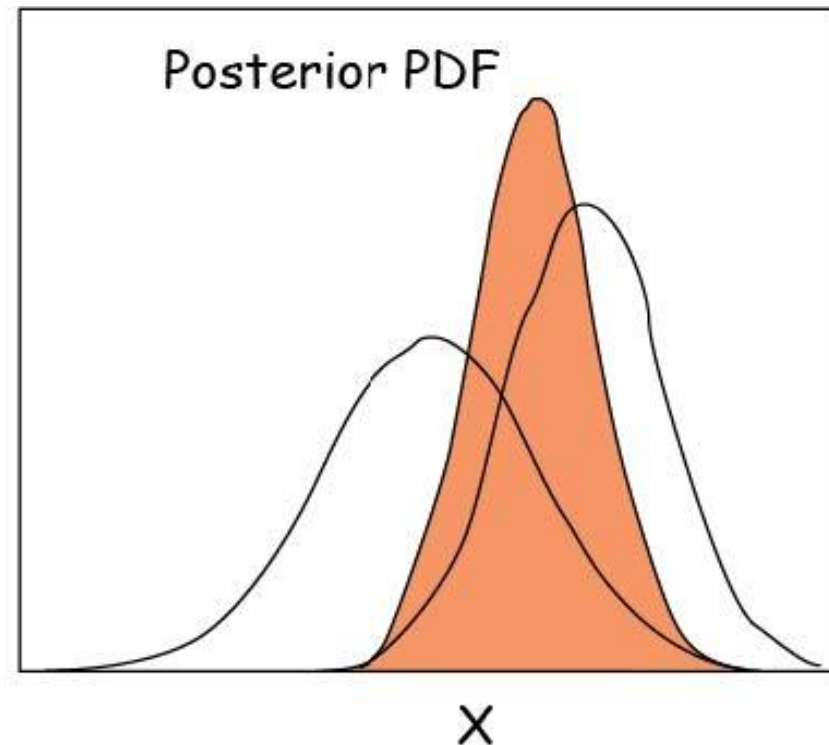
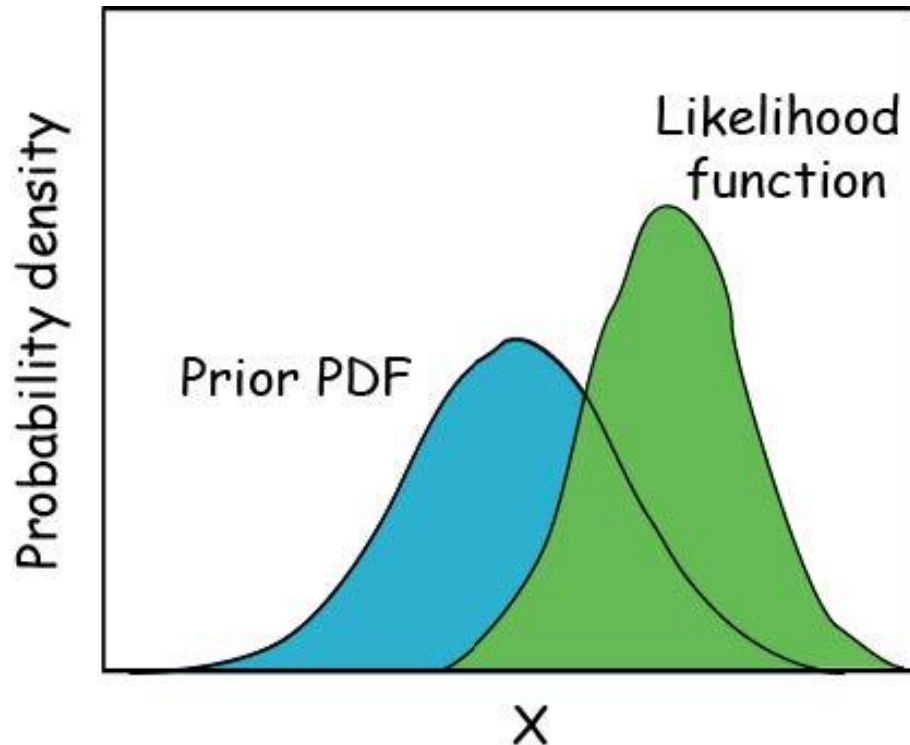


rj-McMC inversion - importance sampling

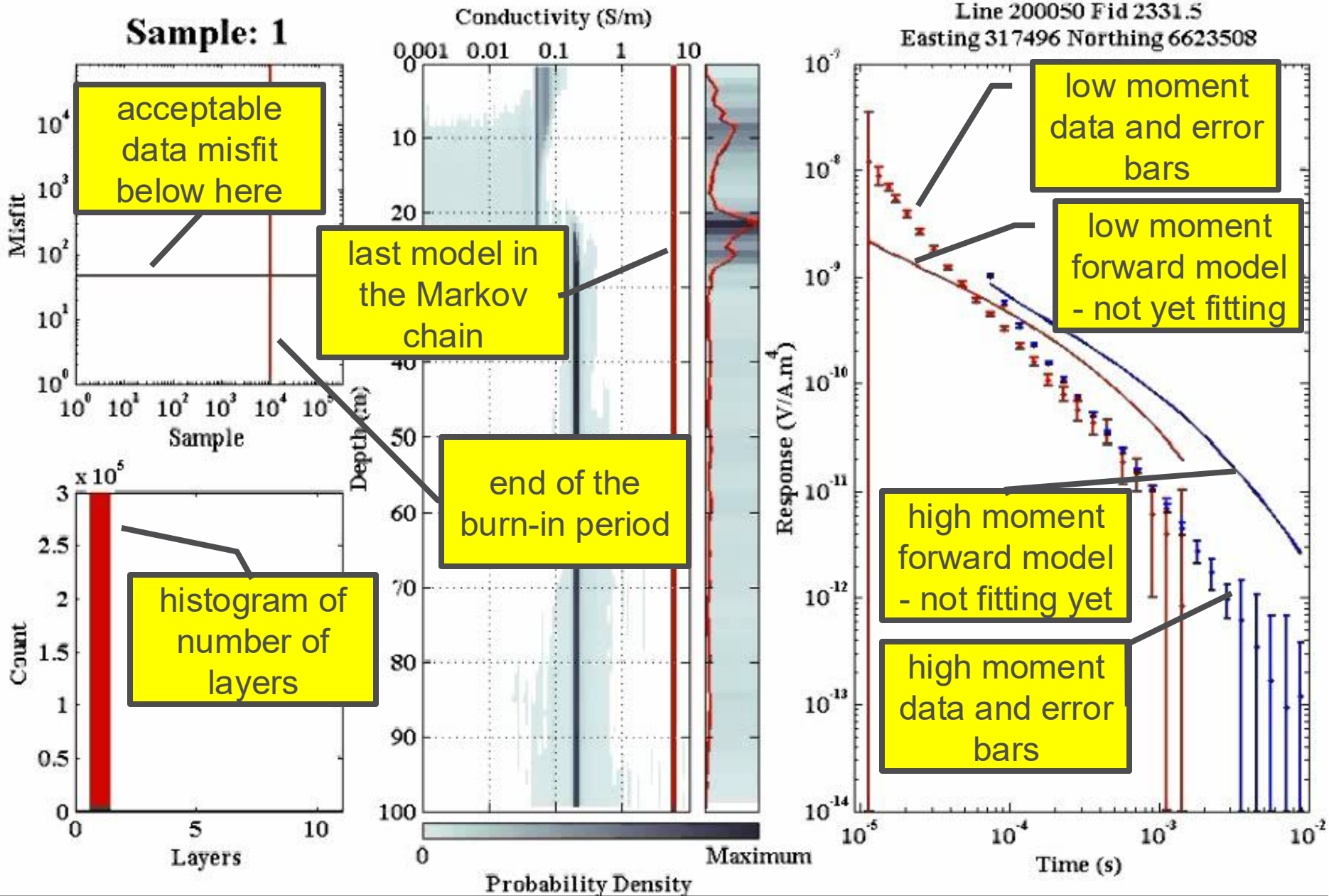
- We don't just accept any old proposed model, they are rejected from or accepted onto the end of the Markov chain based on an acceptance probability
- All other things being equal, this means proposed models are more likely to be accepted if they
 - fit the data better than the current model
 - are more different from the current model
 - are more similar to the *a priori* information
 - have fewer dimensions (layers)
- We end up with an ensemble of simple models that fit the data as well as being consistent with the prior information

rw-McMC inversion - importance sampling

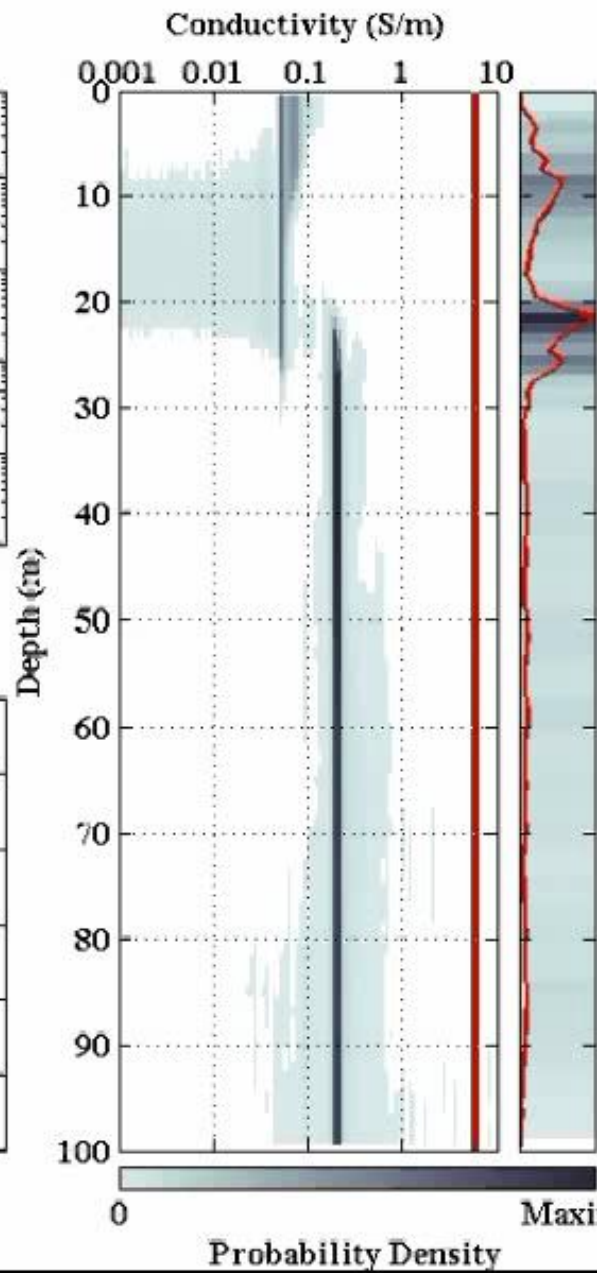
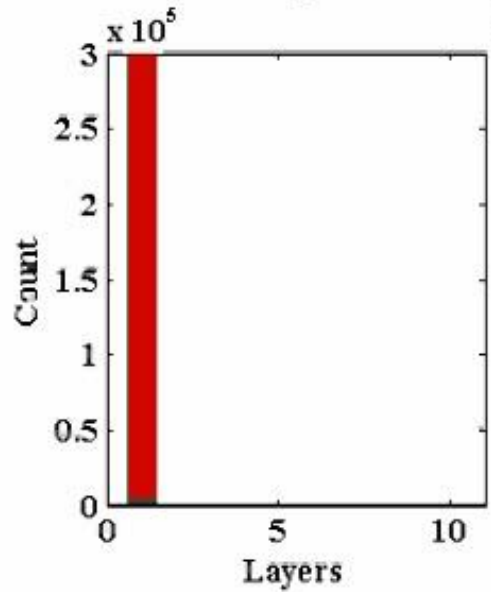
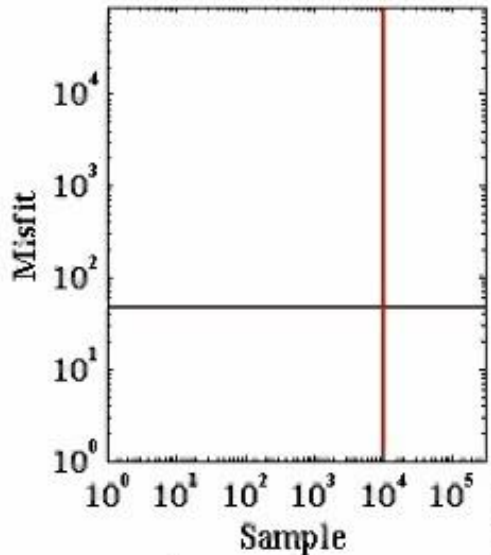
- And theoretically (cross your fingers and with any luck) the ensemble will eventually converge to a statistical distribution that matches the true posterior PDF of the model given the prior information and the new data



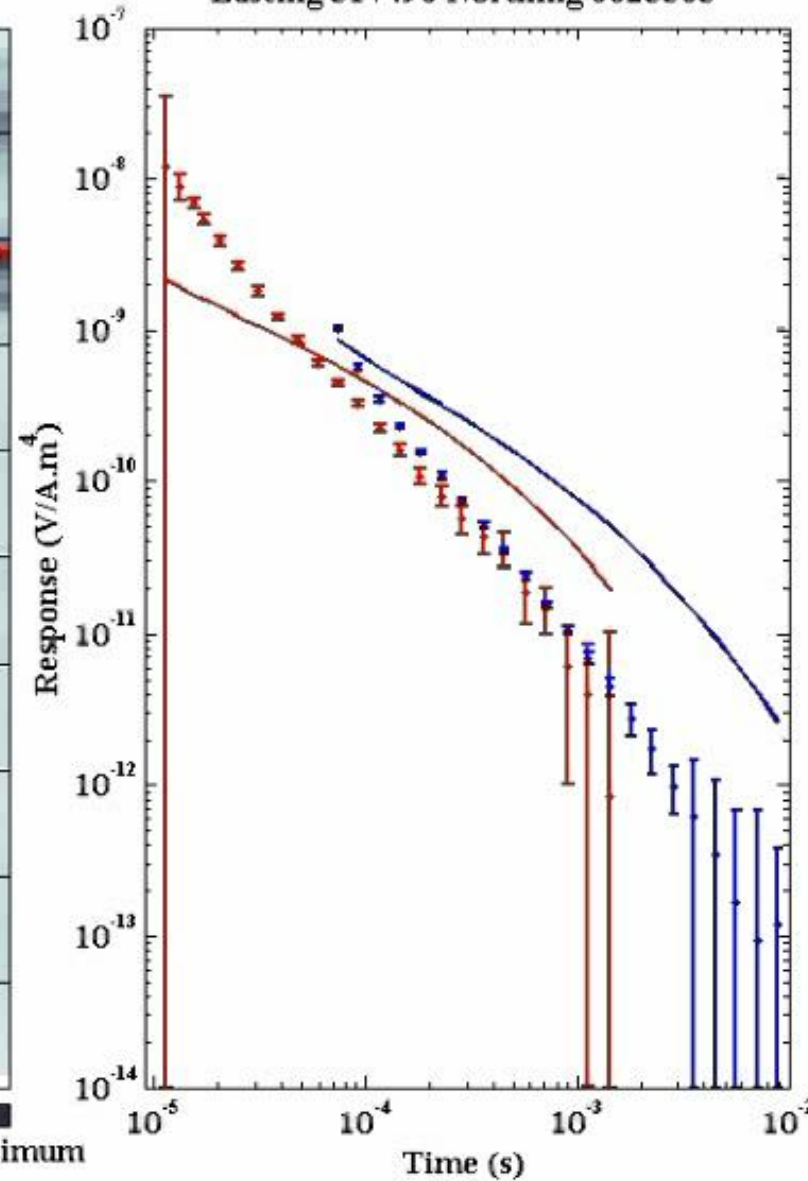
Sample: 1



Sample: 1

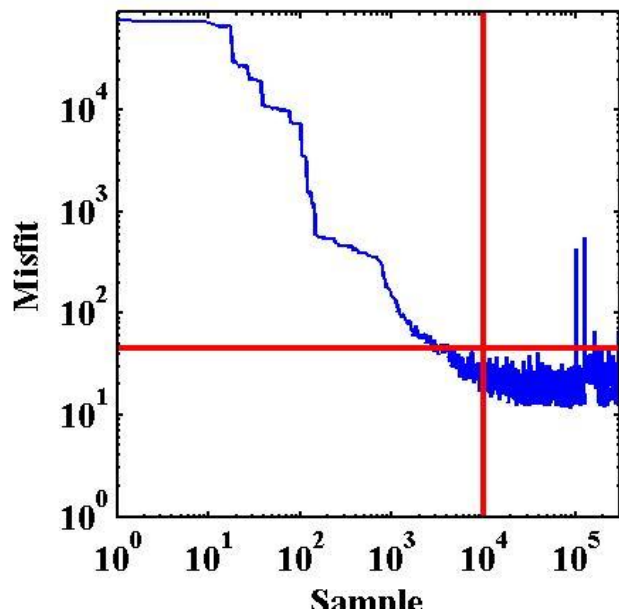


Line 200050 Fid 2331.5
Easting 317496 Northing 6623508

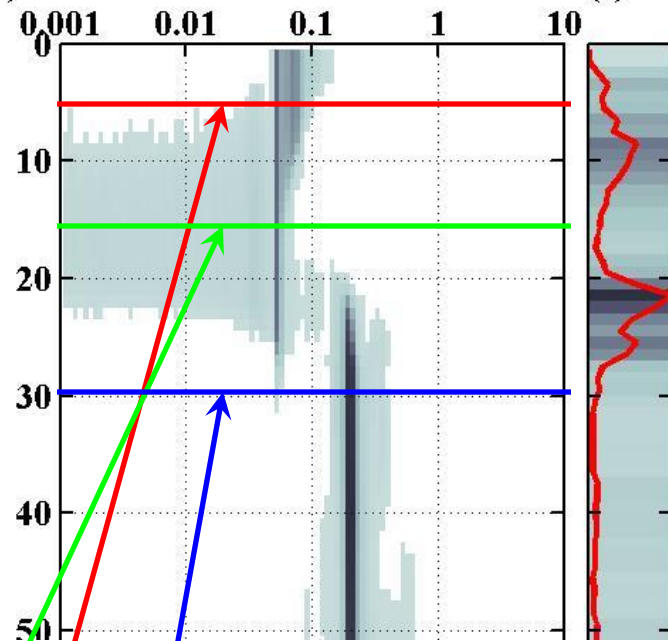


Line 200050 Fid 2331.5

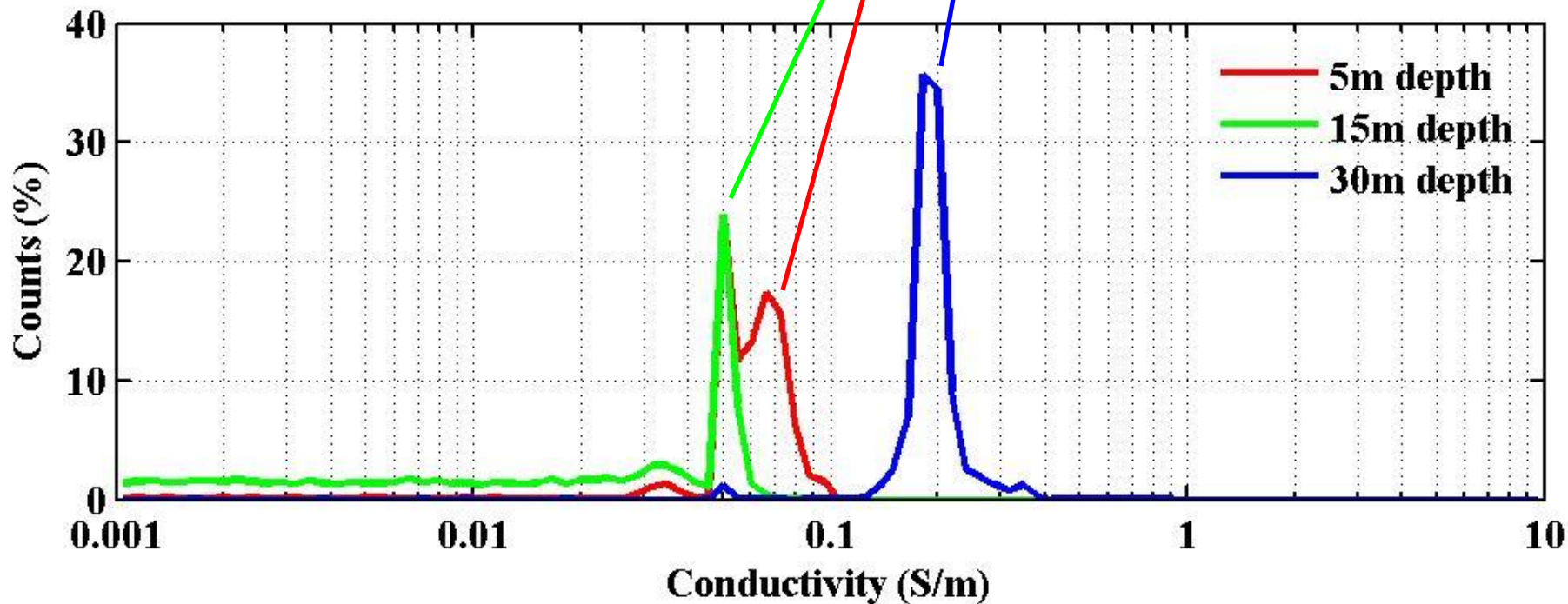
(a) Easting 317496 Northing 6623508



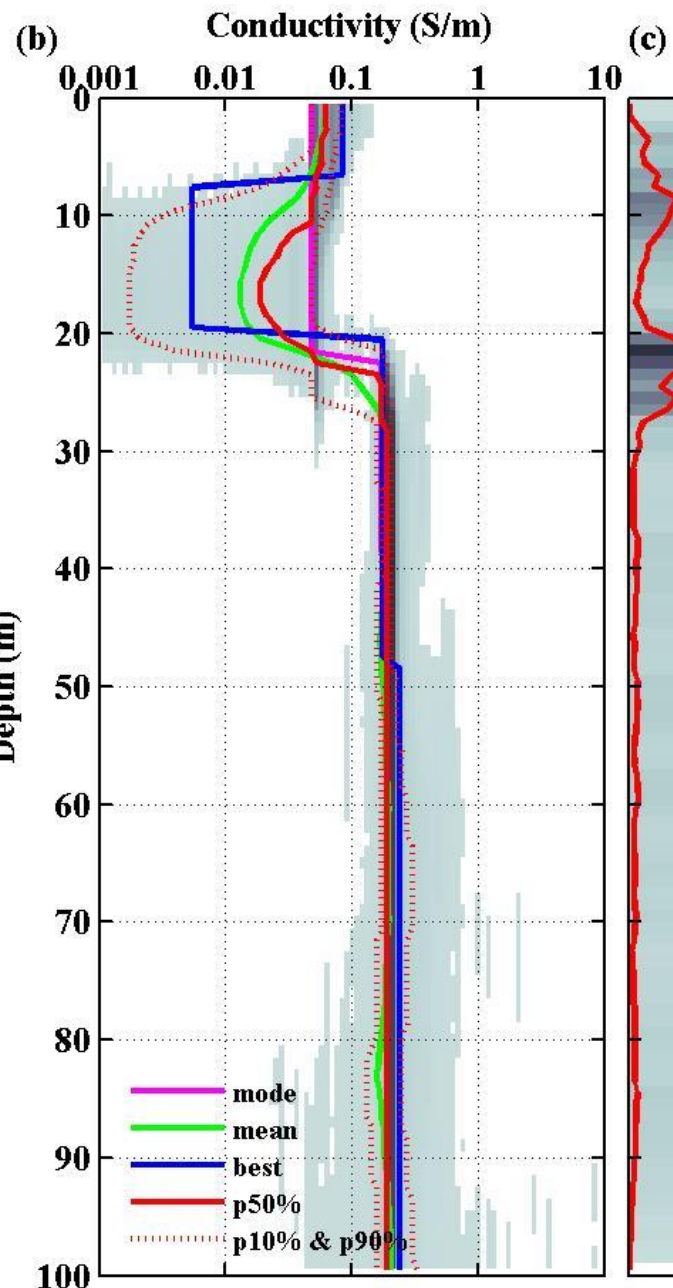
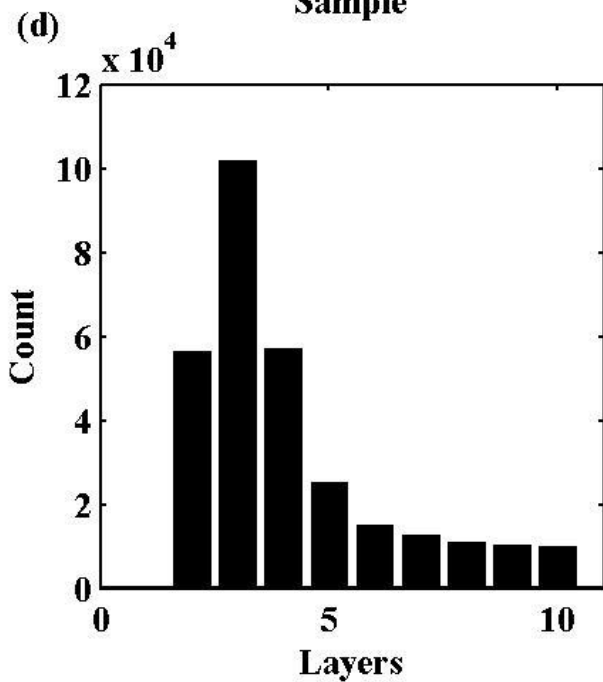
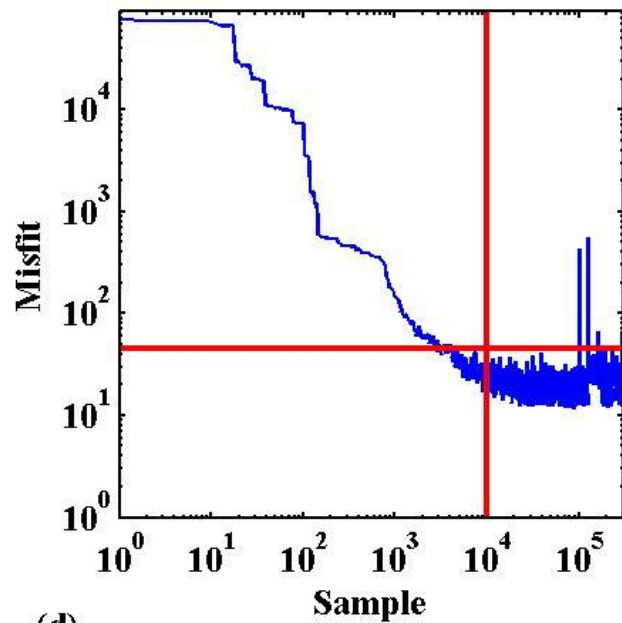
(b)



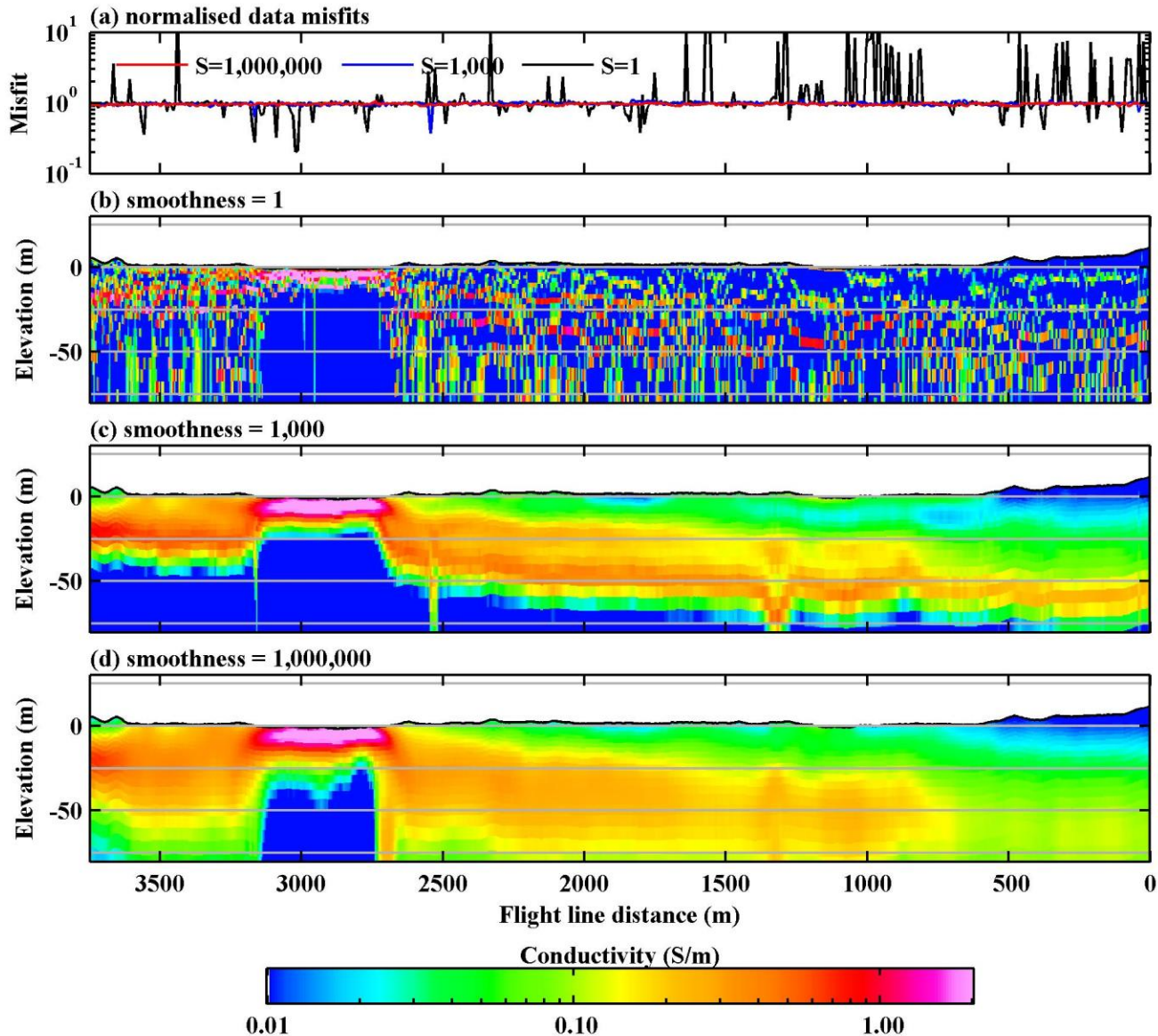
(c)



(a) Line 200050 Fid 2331.5
Easting 317496 Northing 6623508



Deterministic gradient based - smoothness

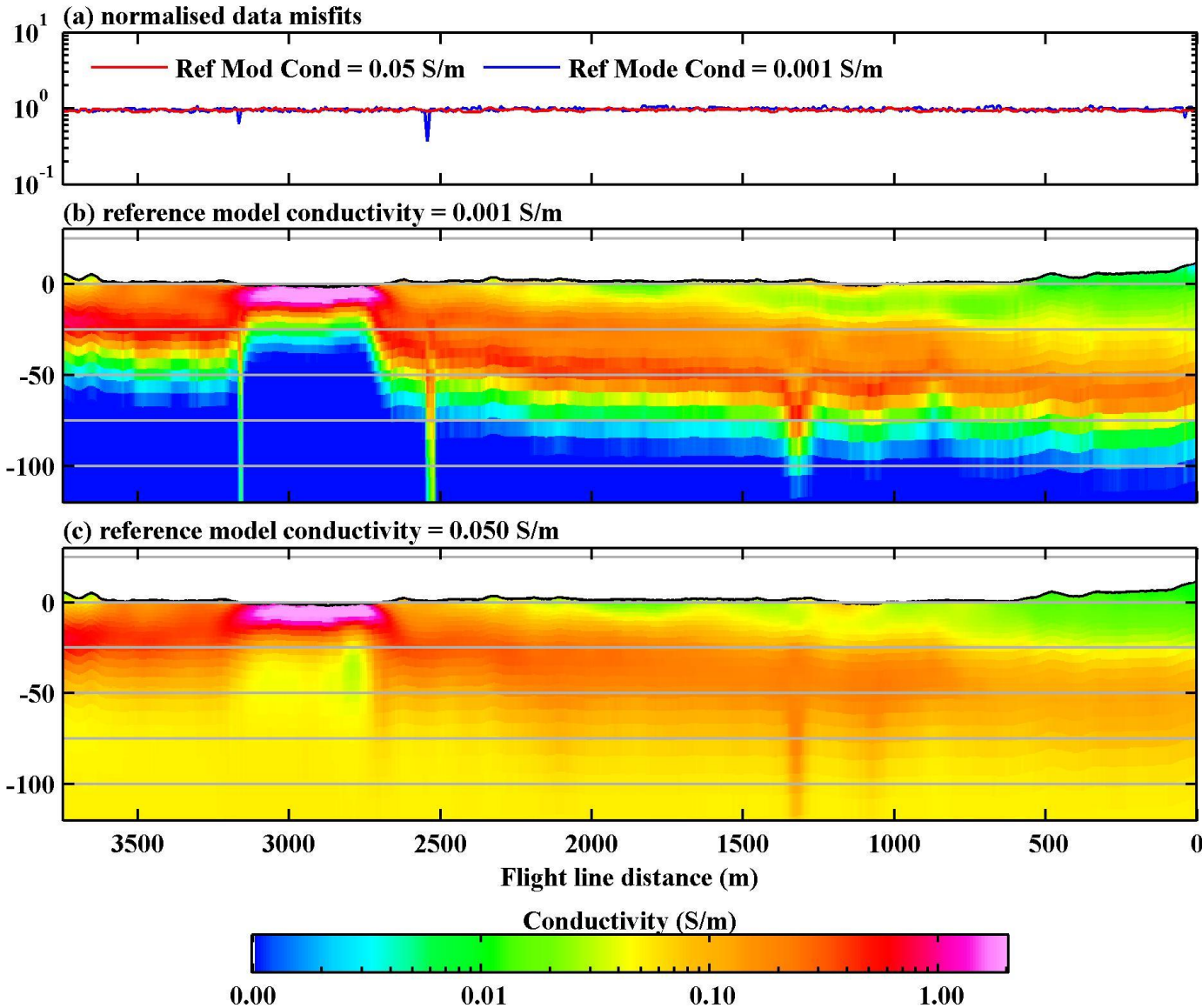


low smoothness

medium
smoothness

high smoothness

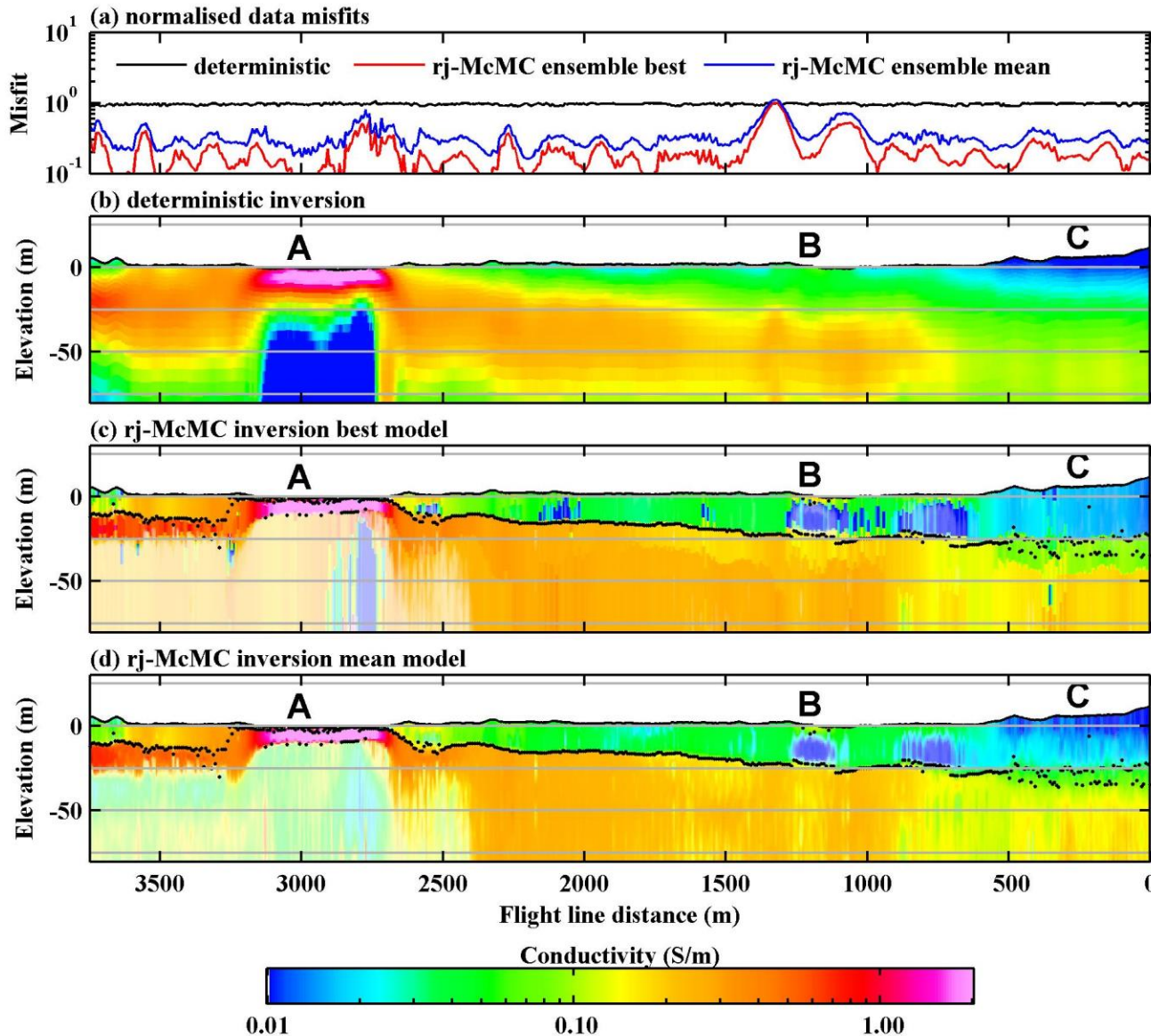
Deterministic gradient based - reference model



reference
conductivity
0.001 S/m

reference
conductivity
0.050 S/m

Deterministic grad based and rj-McMC comparison

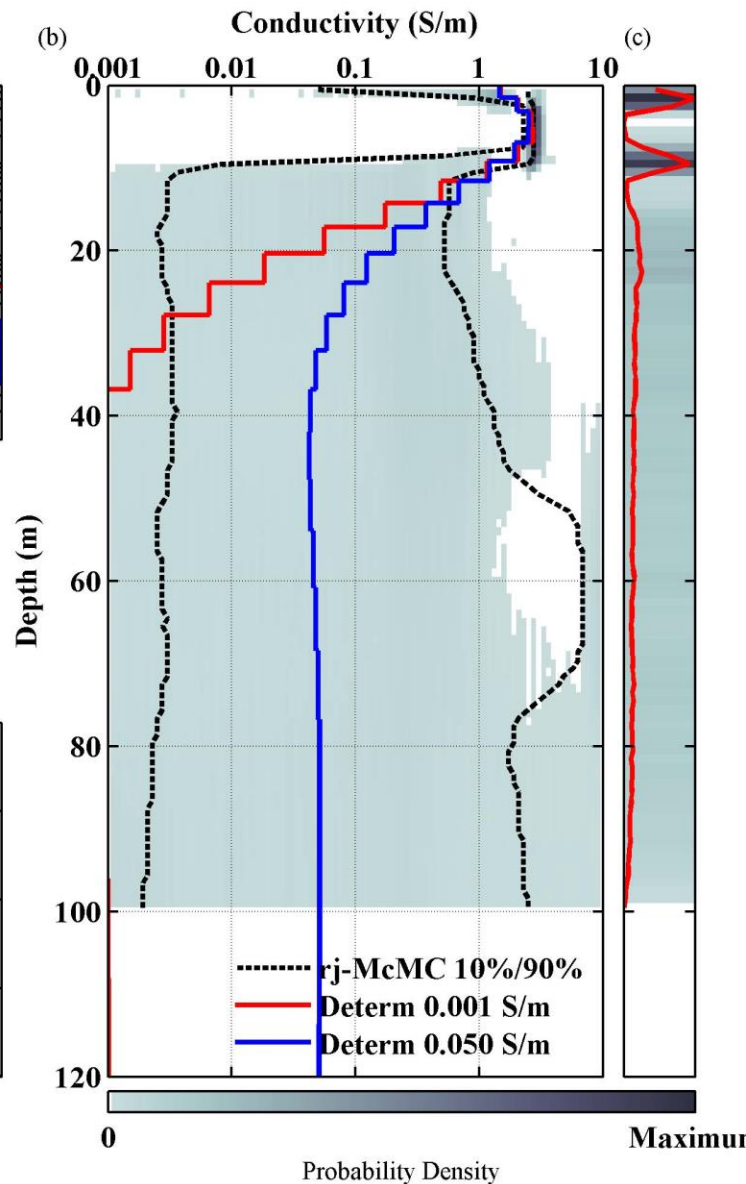
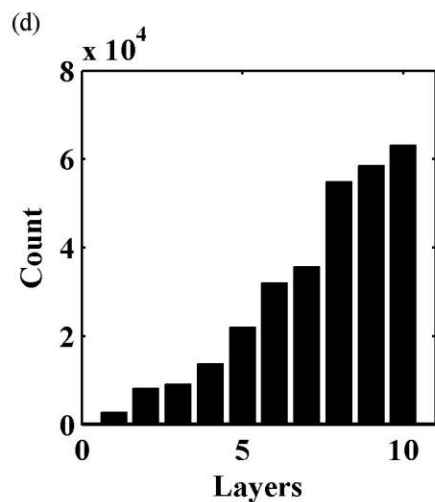
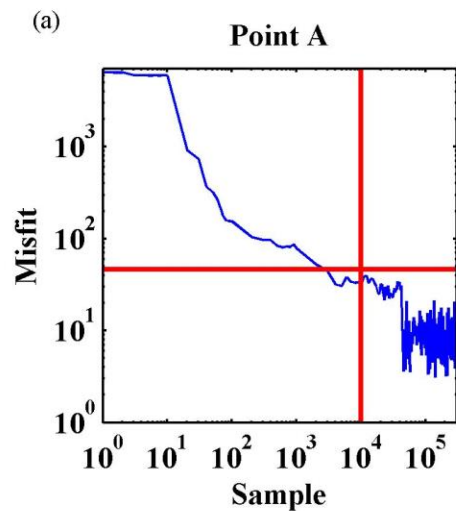
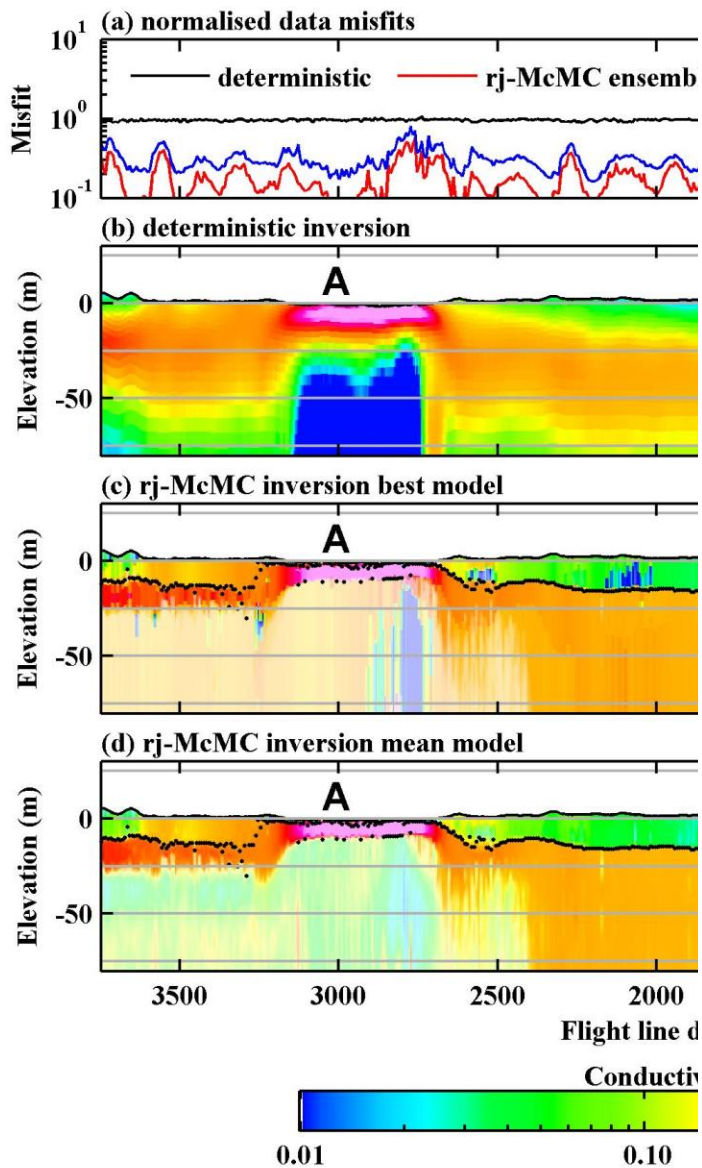


deterministic -
medium
smoothness

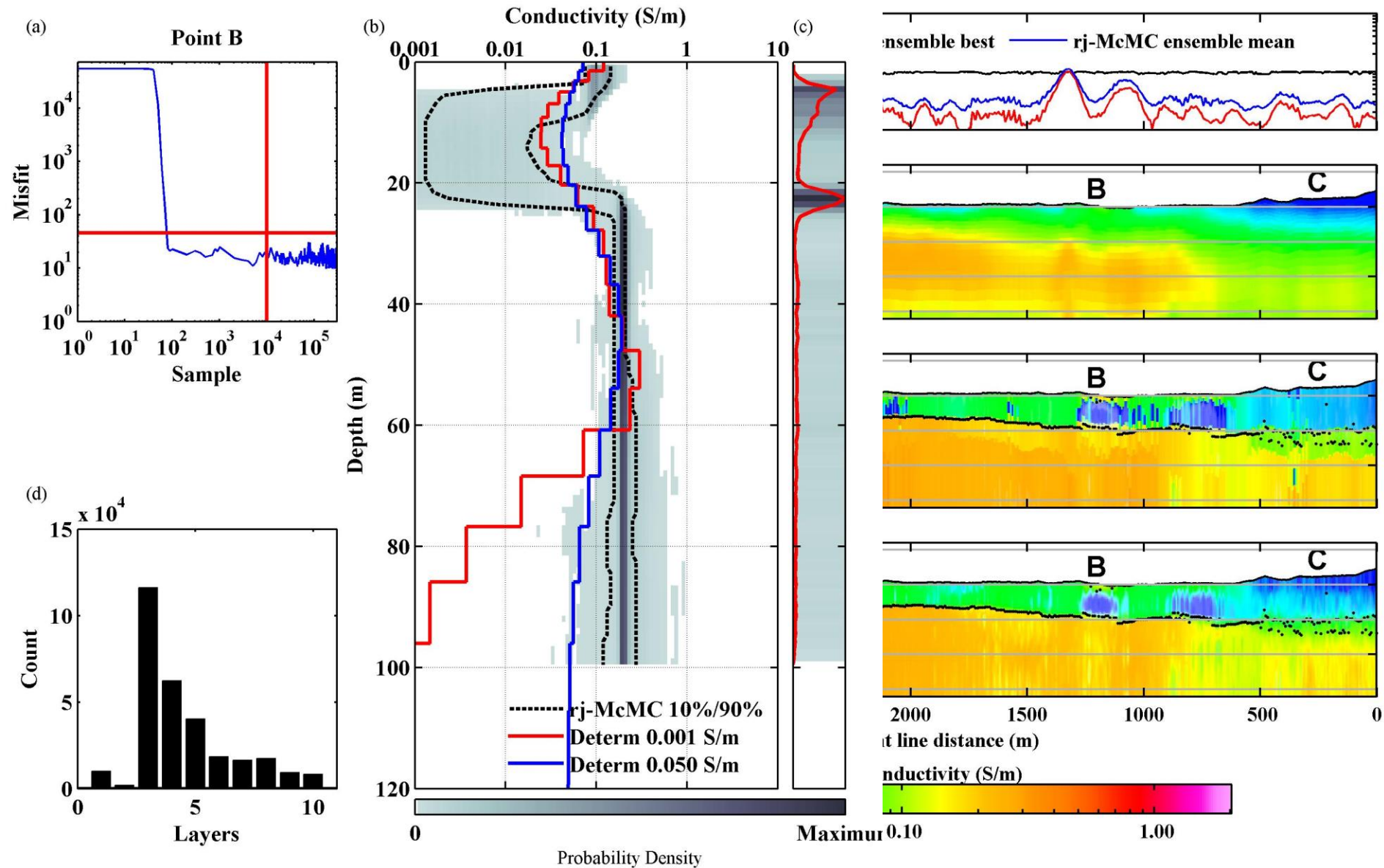
rj-McMC -
most probable
(best) model

rj-McMC -
mean model

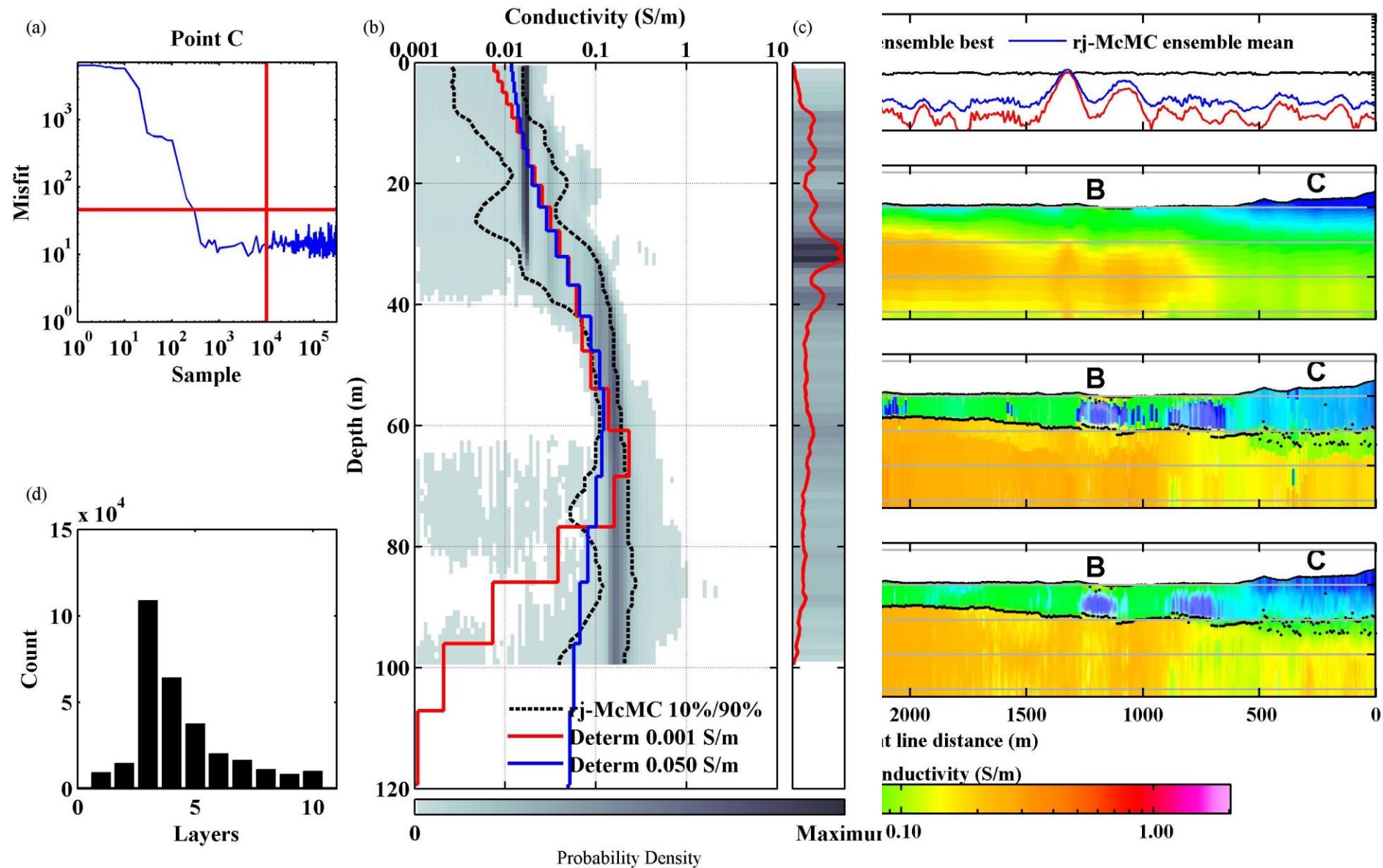
Location A



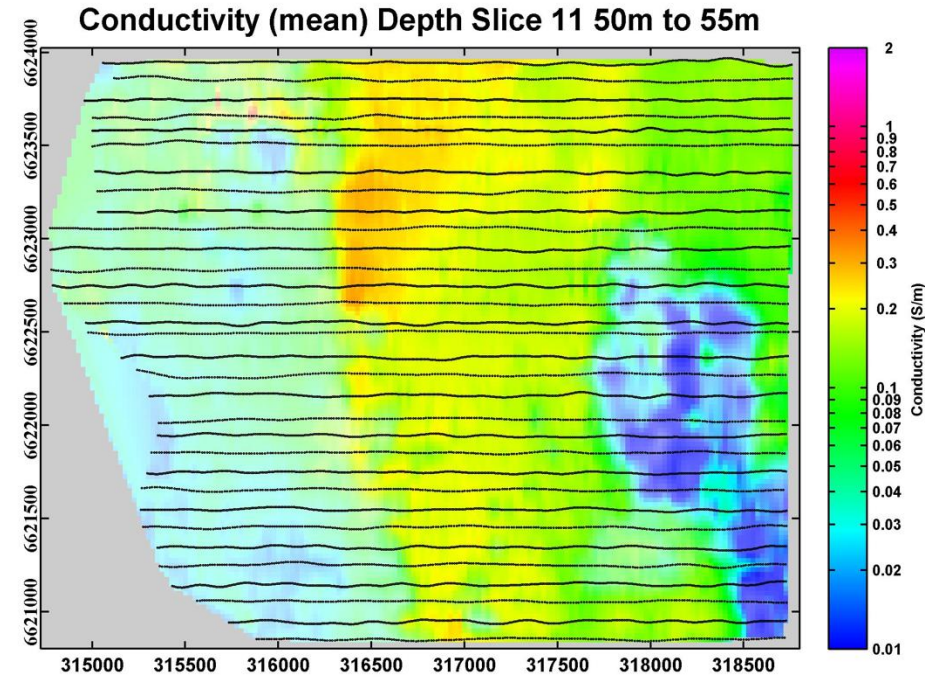
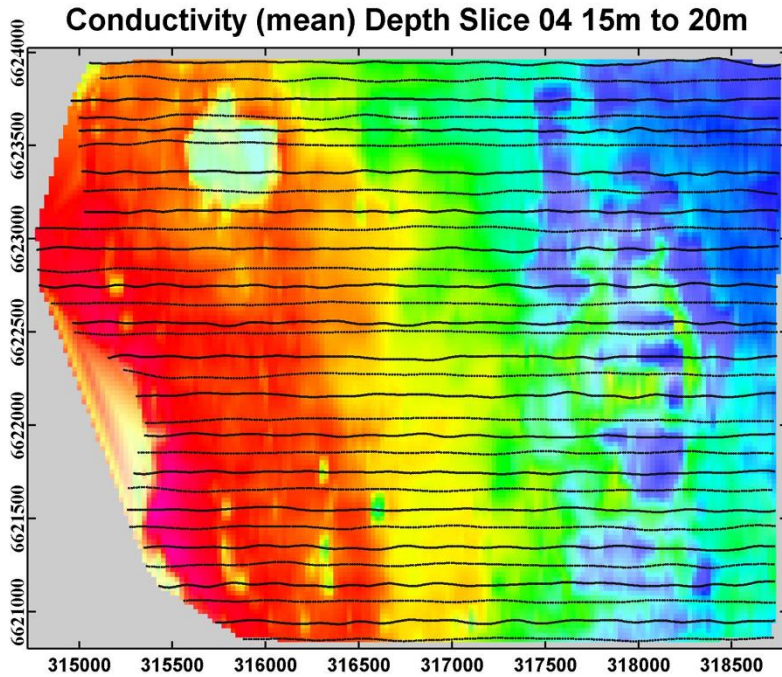
Location B



Location C



Conveying uncertainty on depth slice images



Acknowledgements

Malcolm Sambridge, Research School of Earth Sciences, Australian National University, for guidance on the Monte Carlo inversion algorithm.

National Computational Infrastructure (<http://www.nci.org.au>) which we used for the computational work.

Western Australia Department of Water for providing the Lake Thetis example dataset.



Australian Government
Geoscience Australia



Monte Carlo (stochastic) inversions - similarities and differences to gradient based methods

Ross Brodie, Geoscience Australia

Phone: +61 2 6249 9607

Web: www.ga.gov.au

Email: ross.c.brodie@ga.gov.au

Address: Cnr Jerrabomberra Avenue and Hindmarsh Drive, Symonston ACT 2609

Postal Address: GPO Box 378, Canberra ACT 2601