Bayes, MCMC and Emulation

Richard Hobbs, Qunshu Tang &
Camila Caiado
Durham University

Alan Roberts
GRL

Bayes

- Useful References:

Tutorial:
http://www.sciencedirect.com/science/article/pii/S1007570415000428

MUCM toolkit:
http://mucm.aston.ac.uk/MUCM/MUCMToolkit/index.php?page=MetaHomePag

Goldstein book chapter on model adequacy:
http://onlinelibrary.wiley.com/doi/10.1002/9781118351475.ch26/summary

Goldstein and Wooff book:
http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470015624.html

A decent website with basic Bayesian theory:
http://www.bayesian-inference.com/index

Outline

- Introduction to Bayes
- Introduction to problem to be solved
- The Simulator
- MCMC
- Introduction to exercise
  Run exercise
- Metropolis-Hastings
- Analyse results

Introduction to Bayes

- Why Bayes
  - Uncertainty
    - Data (noise, resolution)
    - Simulator (physics, input/output spaces)
    - Model parameters and variables
  - Discrepancy
    - Internal/external

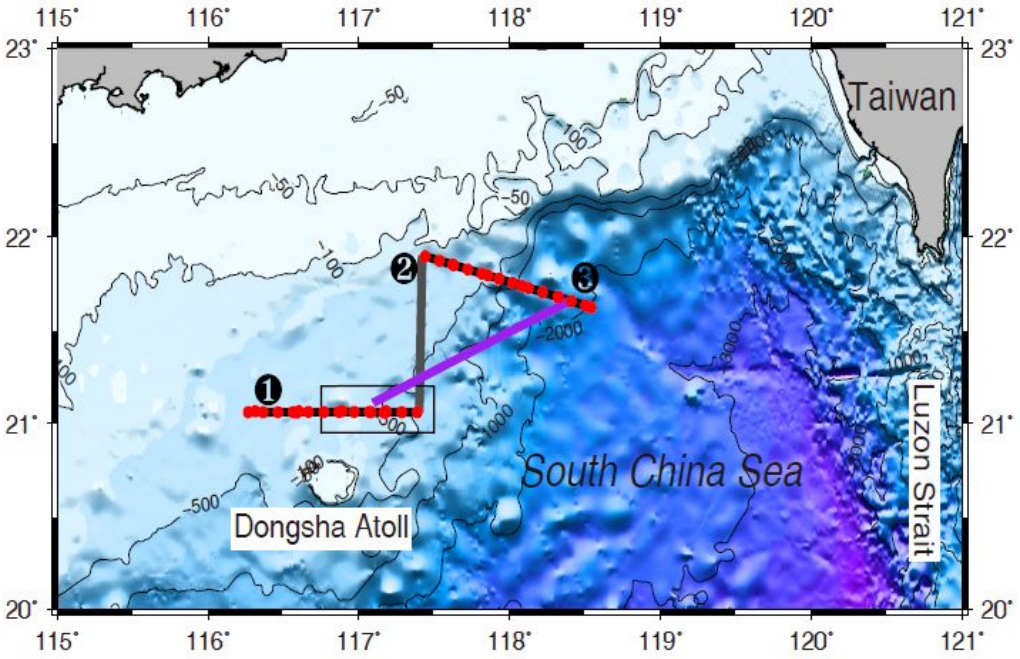
Classic inversion vs statistical inversion

- **Classic inversion**
  - Simulator coupled with a transfer function to update model to fit data in least number of steps
  - Fast
  - No uncertainty
  - Can get trapped in local minima
  - Requires computation of gradient of simulator output against model parameters (difficult and expensive)
  - Must start close to global solution

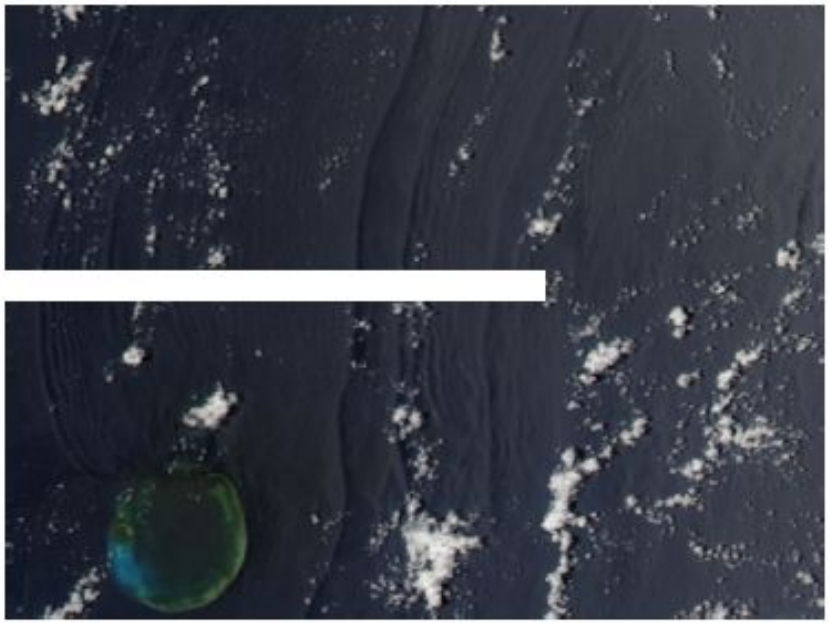
Classic inversion vs statistical inversion

- **Statistical inversion**
  - Requires only the simulator to sample model space
  - Allows expert input/judgements to provide prior info on distribution of all parameters
  - Uncertainty analysis
  - Can explore relevant part of model space
  - Converges faster if start is close to global solution but this is not a requirement
  - Less likely to be trapped by local minima

Introduction to problem to be solved

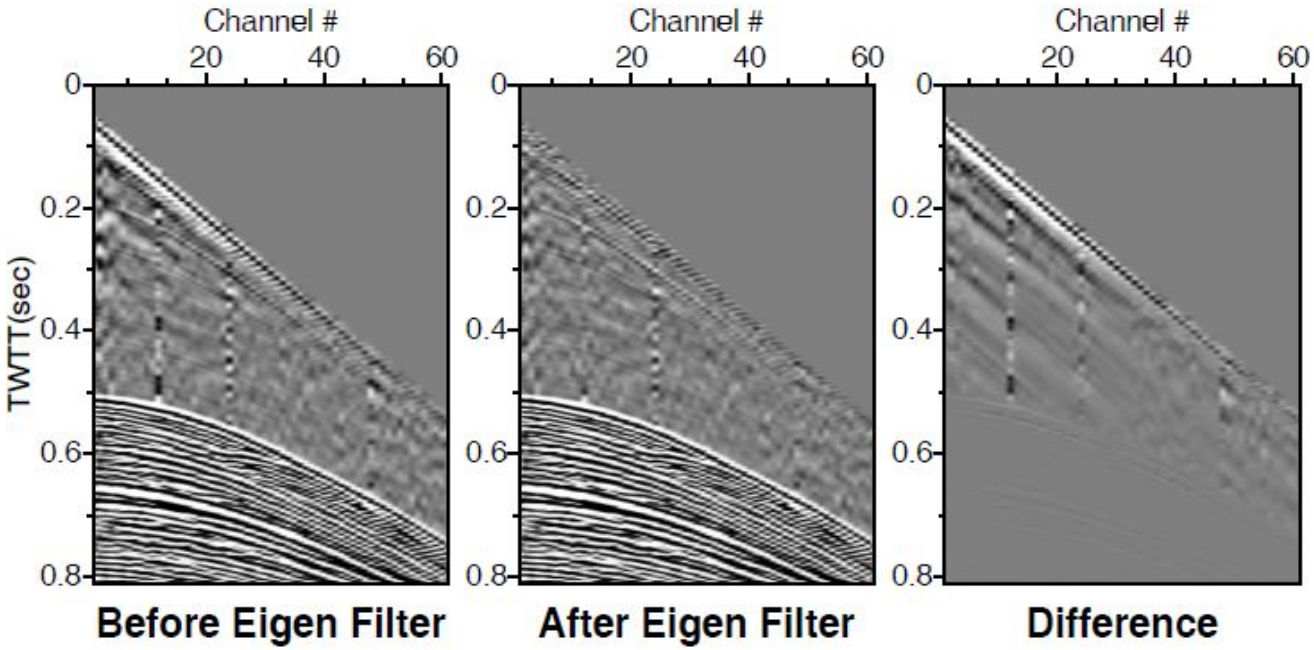


Internal waves are generated when currents or tides interact with topography. Here we examine 'solitons' created in the Luzon Strait as they interact with the China margin. They can be detected on satellite images



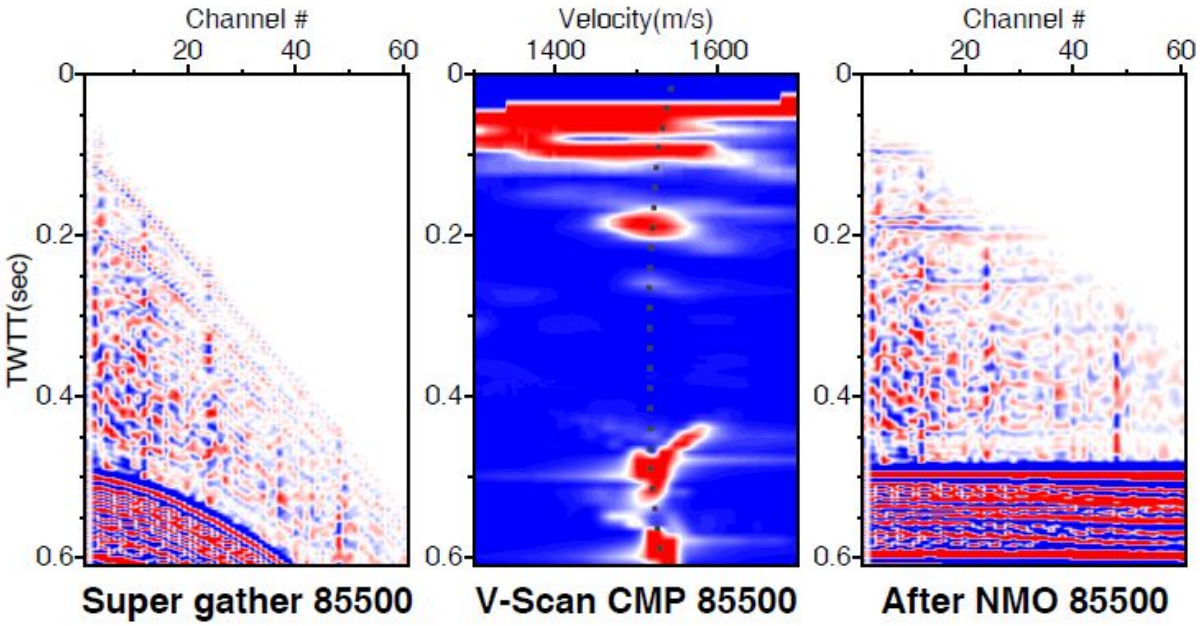
Seismic Oceanography

- Reflectivity is created from impedance boundaries in water caused by stratification.
- Water Impedance is a function of temperature, salinity and pressure (depth)

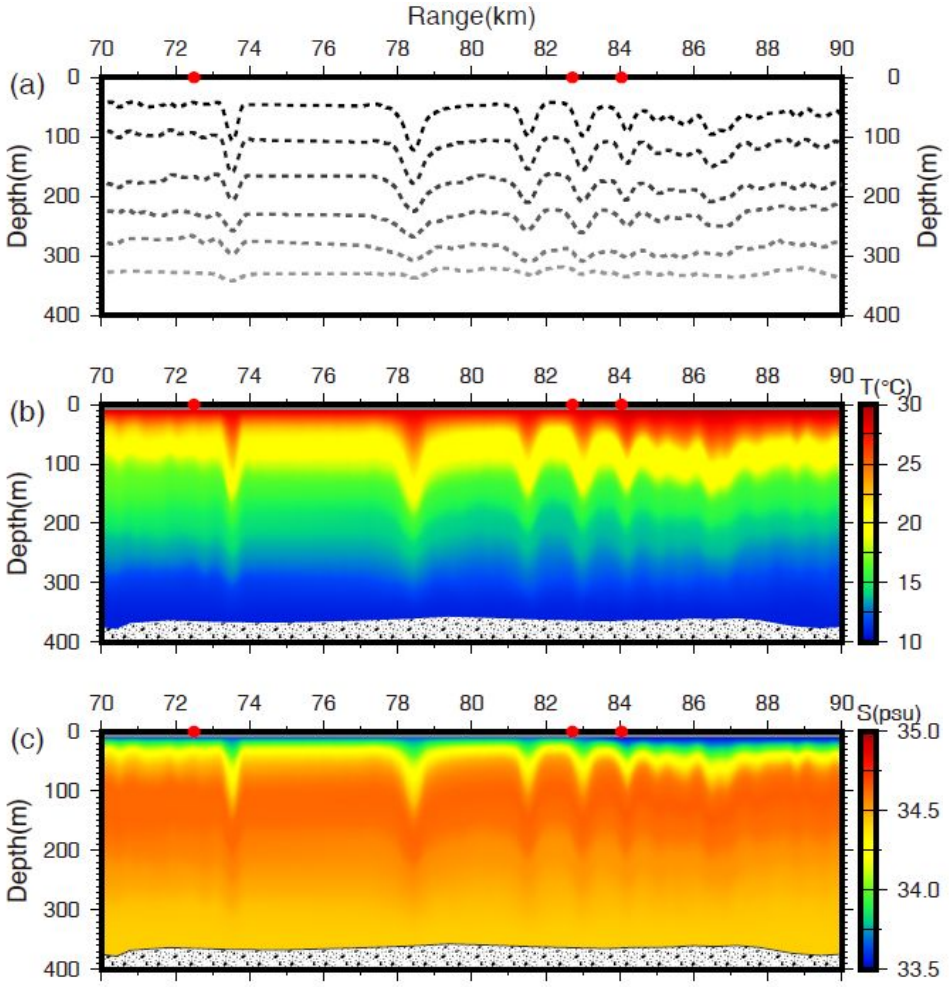


Seismic Oceanography

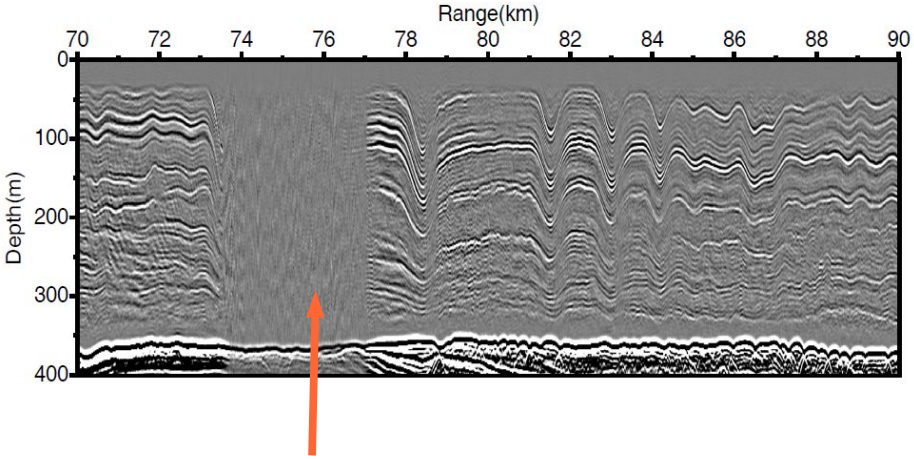
- Reflections can be processed as conventional seismic reflection data (though some complications as the reflectivity is dynamic)



Seismic Oceanography



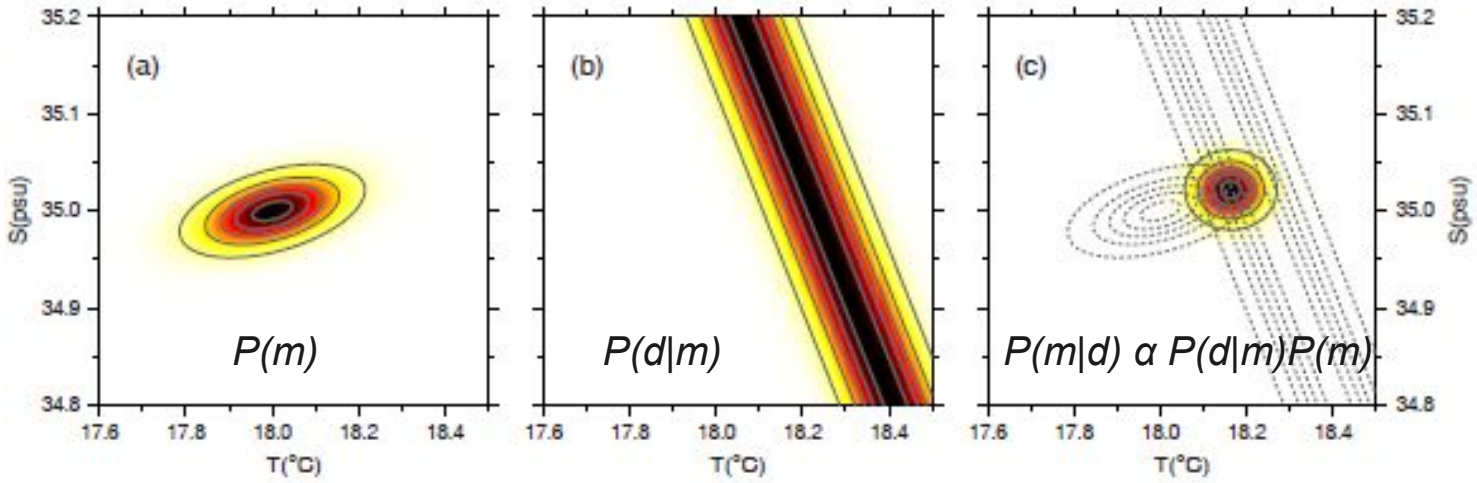
Probable temperature and salinity structure based on interpretation of seismic image



Zone of no data as receiver array was on the surface

Bayes

- Bayesian inference is a method of statistical inference in which Bayes' rule is used to update the probability for a hypothesis as evidence is acquired. (Wikipedia)



Initial probability of model | Likelihood | Updated probability of model

Markov Chain Monte Carlo

- In statistics, Markov Chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from a probability distribution based on constructing a Markov chain (a memoryless random walk) that has the desired distribution as its equilibrium distribution. The state of the chain after a number of steps is then used as a sample of the desired distribution. The quality of the sample improves as a function of the number of steps. (Wikipedia)

Exercise

- We will generate some seismic reflection data from known temperature and salinity measurements with added noise. We then use a Markov Chain Monte Carlo algorithm to use it to try to recover the original temperature, salinity models.

  We use synthetic data so we can assess how well the algorithm has performed

Exercise

1) Compute some seismic reflection data using a simulator:
   - Using data from a CTD cast (T & S vs depth)
2) Perturb the input data by adding noise to become our input data
3) Derive a starting model
4) Calculate covariance between T & S based on CTD data
5) Run MCMC routine to recover T & S from the noisy data

Run exercise

- Walk through code
- MH_simulated – forward model from ctd data
- MH_noise_prior – add noise and compute starting salinity/temperature curves
- MH_cov – compute covariance T,S at each depth
- MH_MCMC – met hastings calculation

The Simulator

- Requirement to compute reflectivity given input parameters of T, S & pressure.

- Open file **MCMC/MH_simulated .m**
  - At each depth, sample T & S
  - Use equations of state for sea-water (sw_....) to compute pressure, sound-speed and density
    - Simple polynomial functions so quick to compute
- Compute impedance differences
- Compute reflectivity -> what we measure

Noise, prior and starting model

- Open file **MCMC/MH_noise_prior.m**
  - The first part adds noise to our reflectivity data with a signal-to-noise ratio of 5 (*you can change SNR value to investigate how robust the result is to noise*)
  - The second part defines the starting model. To do this we fit a polynomial to the T & S data to start we use a 4^{th} order function (*you can change this value to investigate how robust the result is to starting model)*

Covariance

- Open file **MCMC/MH_cov.m**
  - Temperature and salinity are NOT independent and are linked through density (gravity rules OK).
  - To describe this we compute the covariance matrix over a sliding window

$$\sigma = \begin{pmatrix} \sum (T_i - \mu_T)^2 & \sum (T_j - \mu_t)(S_j - \mu_S) \\ \sum (S_j - \mu_S)(T_j - \mu_T) & \sum (S_j - \mu_S)^2 \end{pmatrix}$$

  - over a window length M (*you can change M value to investigate effect but MUST be odd*)
  - We extract the variance of T and S to guide candidate sampling

MCMC

- Open file **MCMC/MH_MCMC.m**
  - The key parameter here is *N* the number of iterations. It need to be suffice to let the chain converge (adding more samples does not change the shape of the posterior density function), but not too long otherwise you have to wait for ever.... (*you can change N suggest values of >1000*)
  - So lets get the code running then I will explain it function

Metropolis-Hastings

- Steps (i is iteration, j is depth level)
  - Independently draw a candidate set of parameters {T,S} based on computed variance of T and S which is set to be ¼ of width of prior. This ensures that any step is not too large which may result in too many rejected models
  - Compute the model probability ratio between the new and previous T,S pairs (this incorporates the previously computed covariance)

Metropolis-Hastings

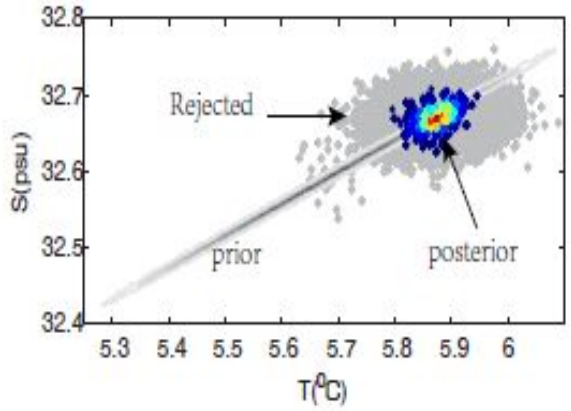
- Use the simulator to compute reflectivity of new and previous model
- Compute likelihoods, ie how likely would the observed data have been obtained from the new and previous model
- Compute ratio of likelihoods
- Multiply this by ratio of model probabilities to compute α

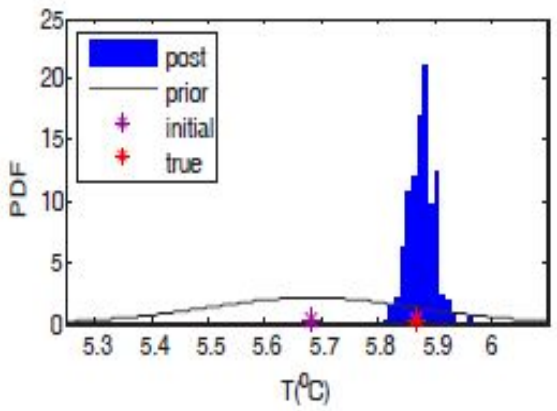
Metropolis-Hastings

- Accept or reject
  - If ratio is > 1 then this is a more likely model so always accept
  - If ratio is 0 always reject
  - If $0 < \alpha < 1$
    - Take a random number between 0 and 1 from uniform distribution u
    - If $u < \alpha$ accept this model even though it is not as good as the previous; otherwise reject
- Step to candidate model or repeat current model if candidate is rejected
- REPEAT

Plotting PDFs

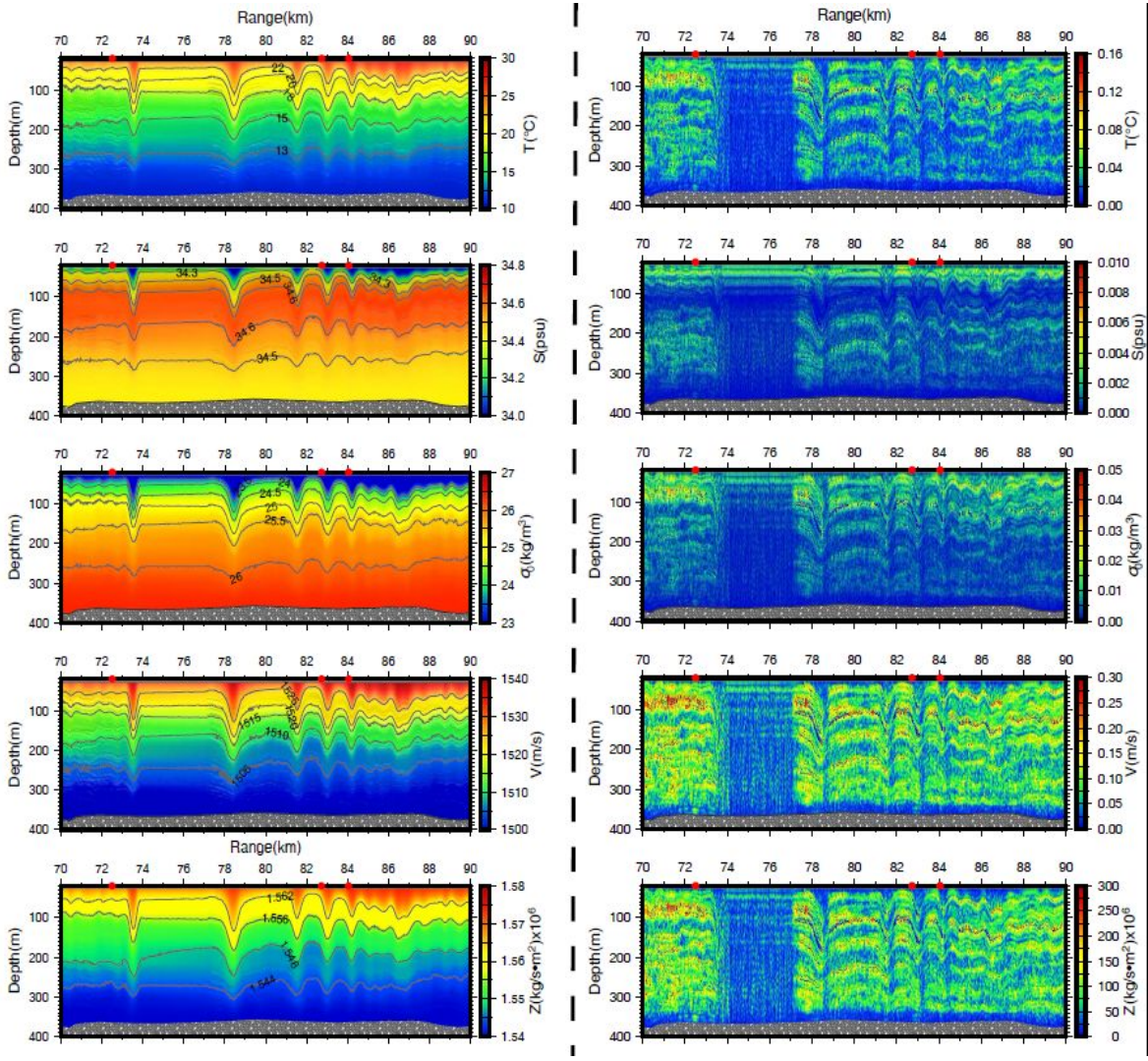
- Once chain is complete you can plot the posterior density function (PDF) for each parameter. This is a histogram of the number of times the 'walk' visited the value. The width is a measure of the uncertainty





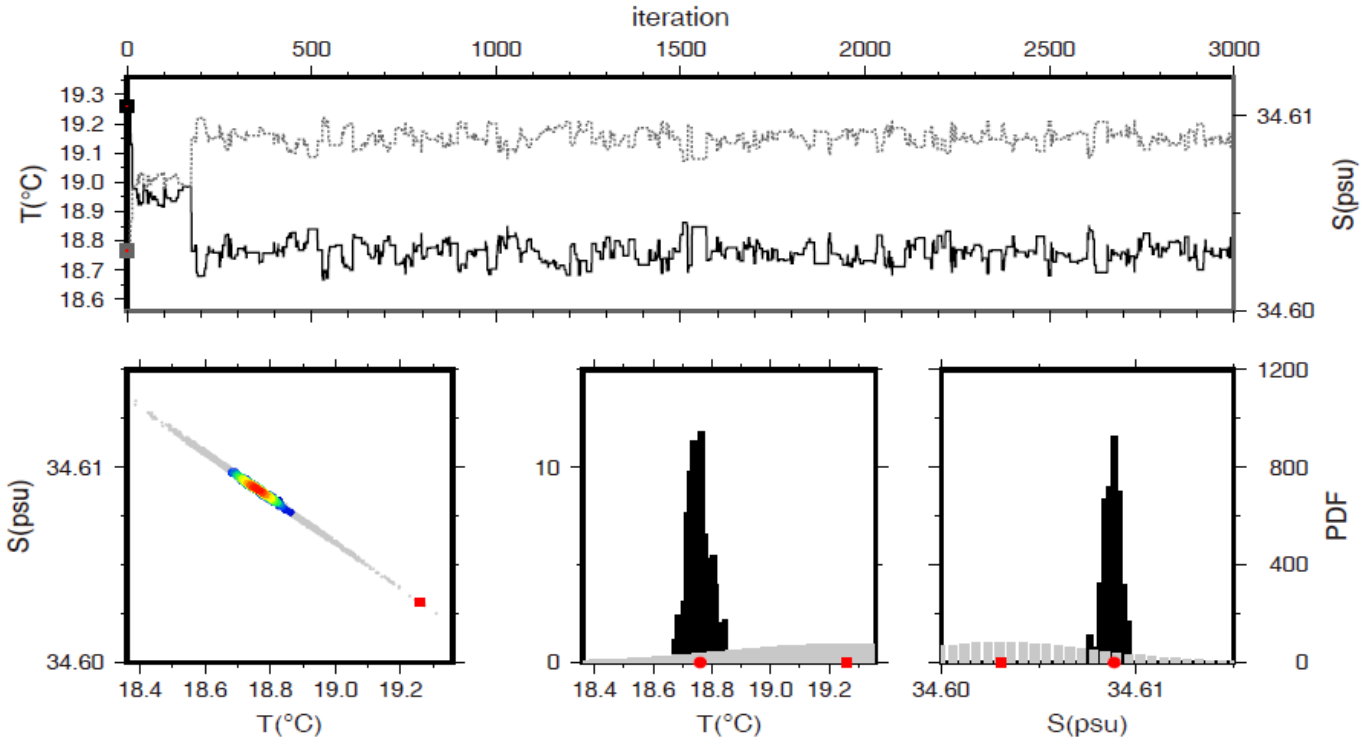
Conveying the message

- In this exercise you have modelled one trace.....and there could be thousands



Pitfalls?

- A major problem with MCMC is the number of runs of the simulator needed to gather enough information to provide a robust result – has the series converged?



Uncertainties

- We have only really explored one source of uncertainty in this exercise – that of noisy data
- What other source of uncertainty are there?

Uncertainties

- Noise
- Discretisation/sampling (Nyquist)
- Simulator error – we cheat on the physics to make the problem tractable
- Model parametrisation – does the real Earth look like your model?

  → **Model discrepancy**
- MCMC allows this to be explicity included based on expert judgement

Emulation

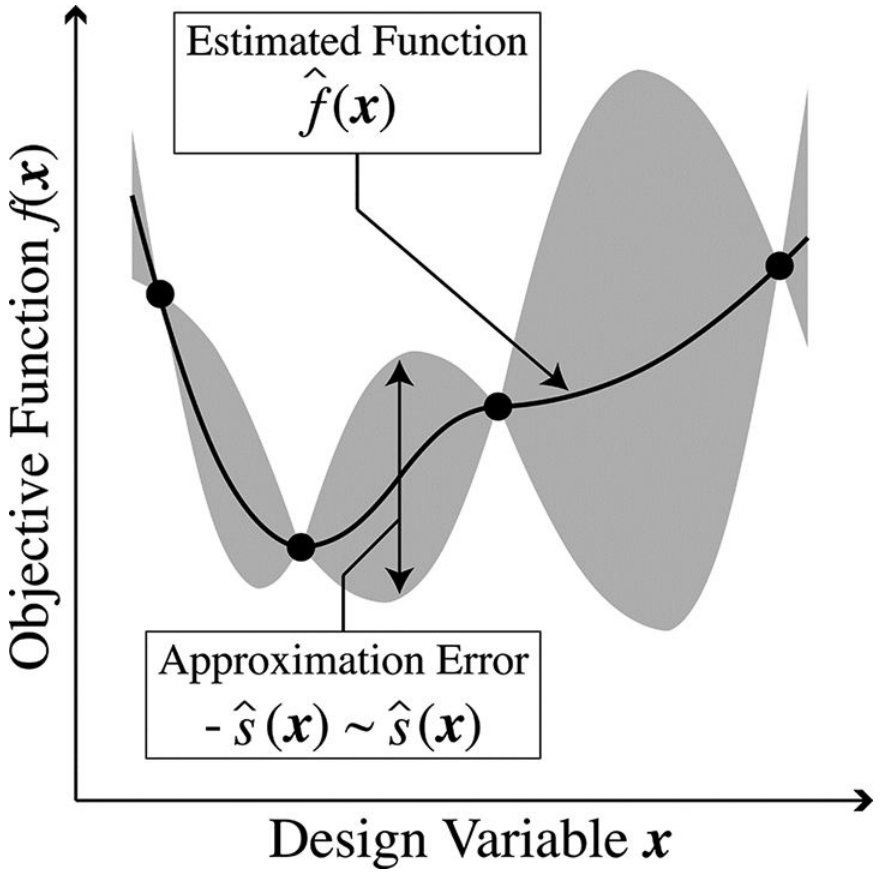
- Here the simulator is fast – what if you wanted to a raytraced model or 3D MT data using MCMC?
- How long to run 1000 models for each parameter along each ray?

  If the simulator takes 0.1 s per ray and there are 25 OBS with 100 time picks each, the model has 100 parameters, and you want to run 1000 iterations for each ray. How long would it take?

Emulation

- An emulator is a surrogate model with calibrated uncertainty.
  - Run the simulator a limited number of time that sample the whole model space
  - Create a surface through the results and construct a polynomial to describe the result
  - Run the simulator a second time using different inputs
  - Use the result to calibrate uncertainty in polynomial away from calibration points

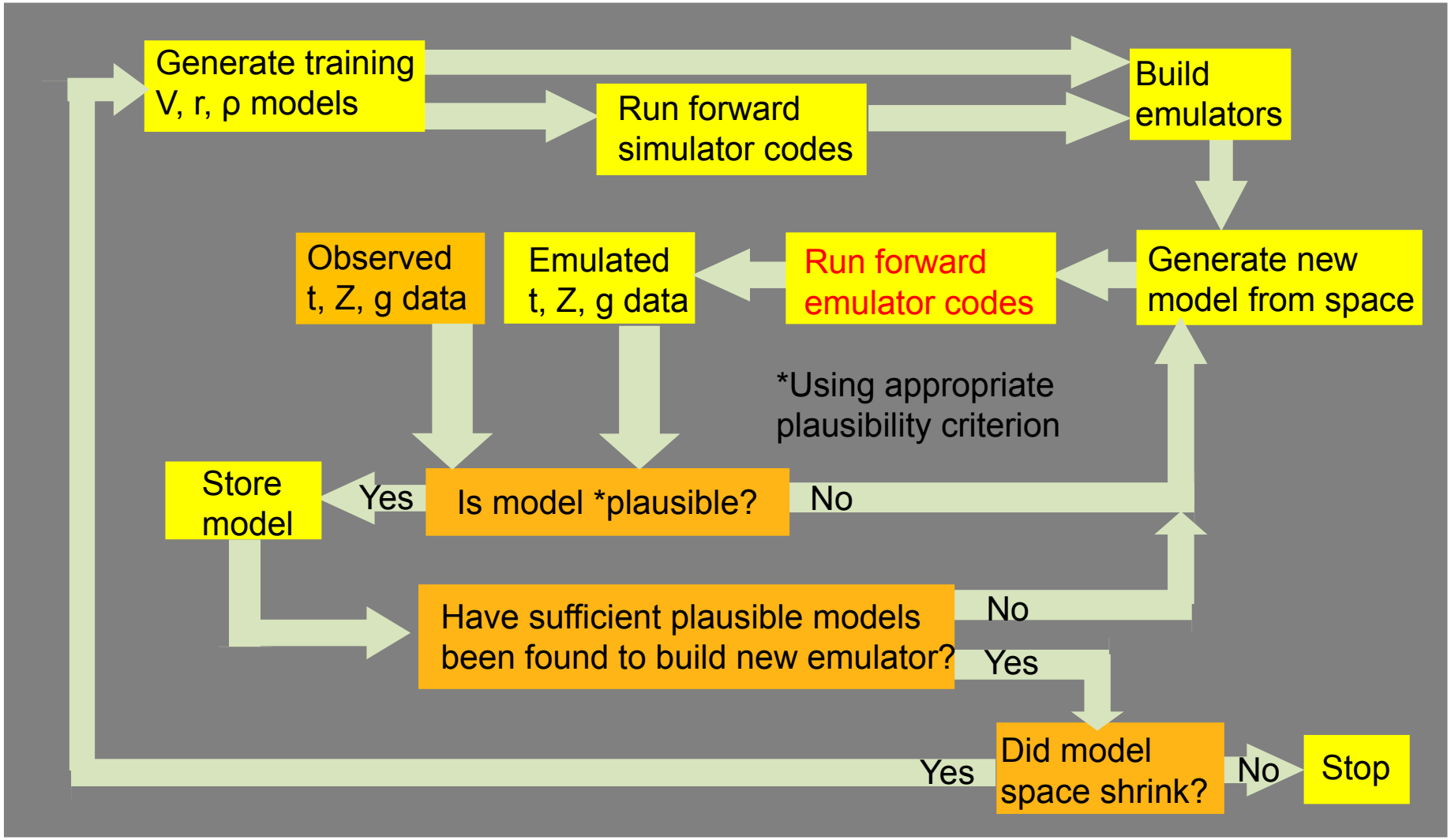
Emulation


Estimated Function
$\hat{f}(\boldsymbol{x})$
Objective Function $f(\boldsymbol{x})$
Approximation Error
$-\hat{s}(\boldsymbol{x}) \sim \hat{s}(\boldsymbol{x})$
Design Variable $\boldsymbol{x}$

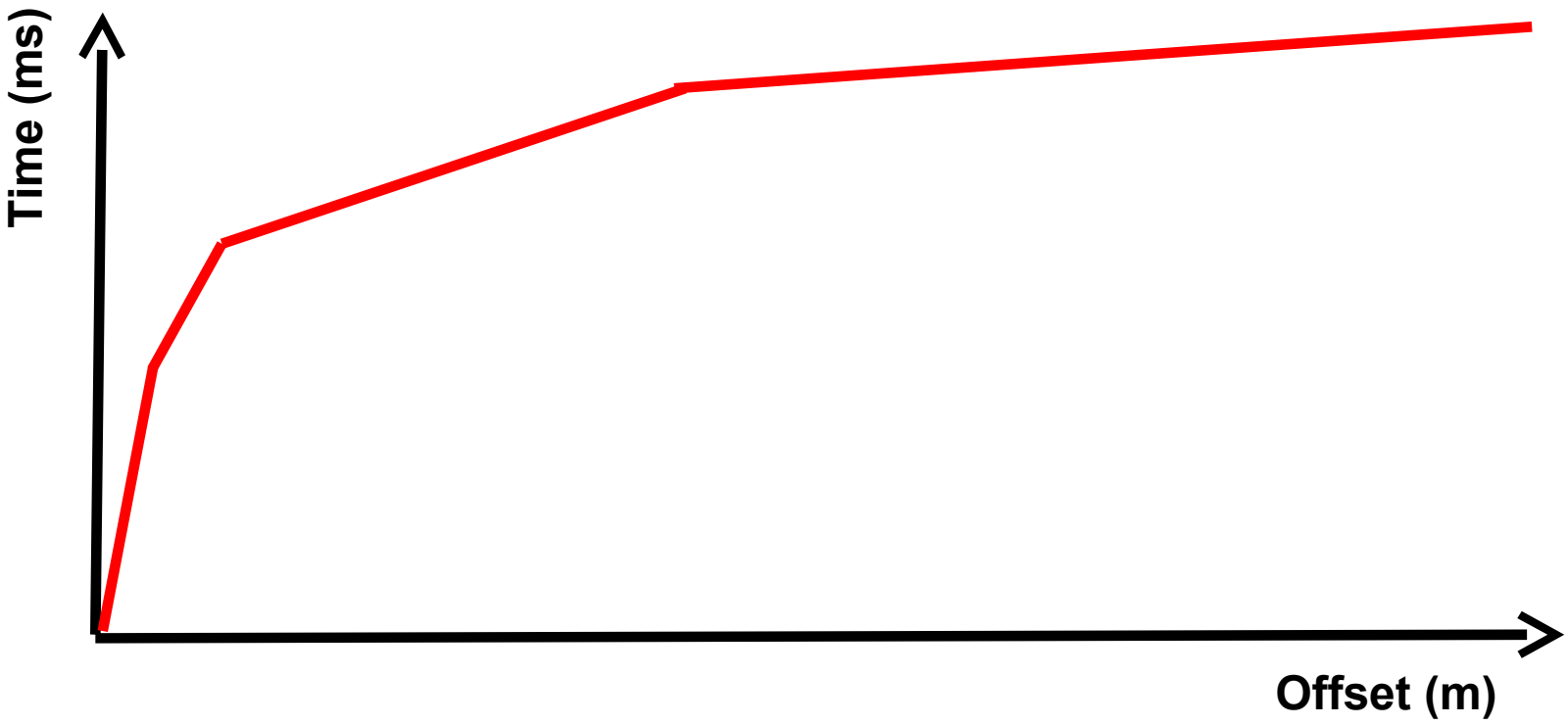
Emulation

- Run the MCMC (M-H) using the emulator as a proxy for the simulator
- Identify regions of model space that never fit the observations (implausibility test)
- Update model space by removal of implausible areas
- Re-compute emulator
- Converged when error in emulator is at the same level as the other uncertainty terms (noise, model discrepancy...)

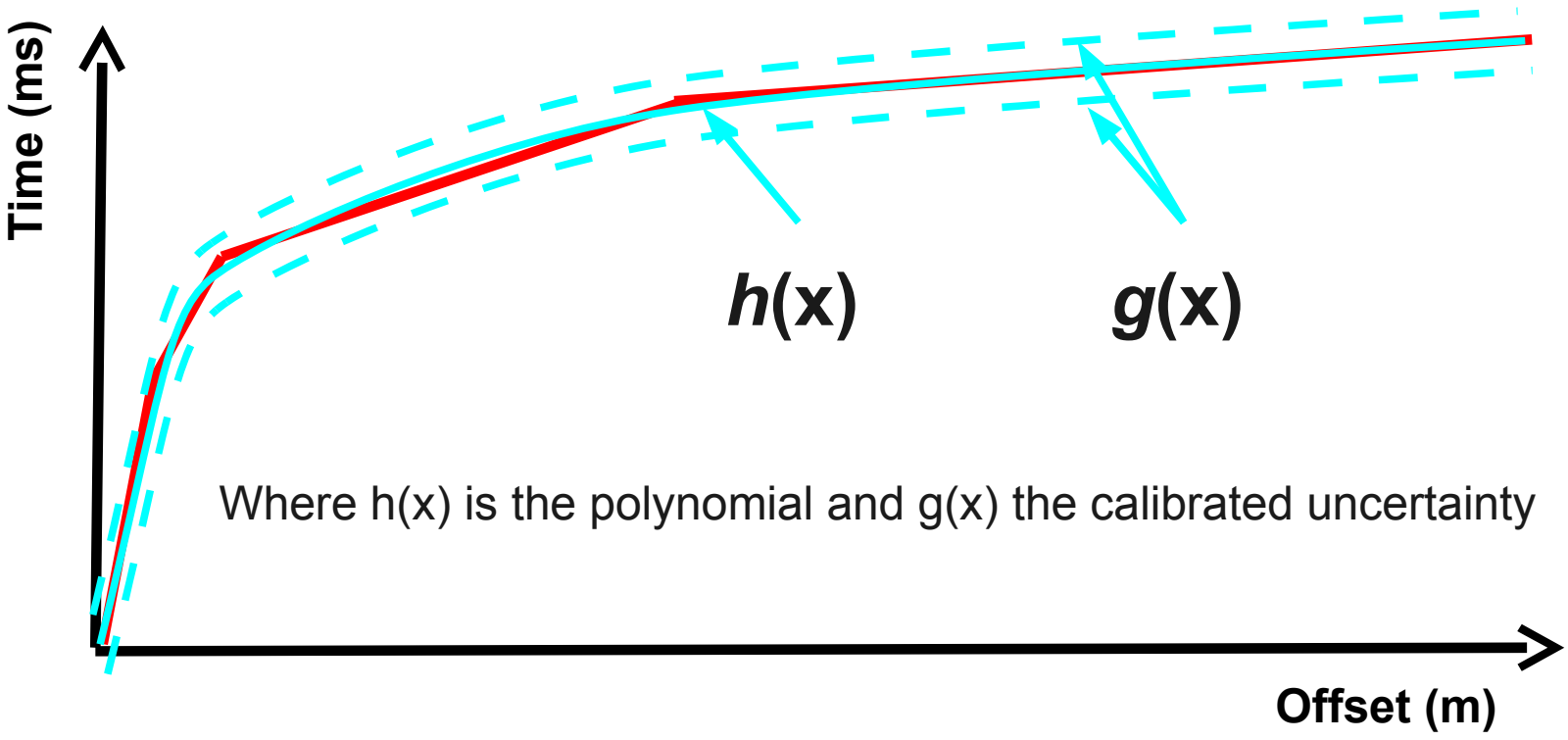
Emulators


Generate training
V, r, ρ models
Run forward
simulator codes
Build
emulators
Observed
t, Z, g data
Emulated
t, Z, g data
Run forward
emulator codes
Generate new
model from space
*Using appropriate
plausibility criterion
Store
model
Yes
Is model *plausible?
No
Have sufficient plausible models
been found to build new emulator?
No
Yes
Yes
Did model
space shrink?
No
Stop

Building an emulator

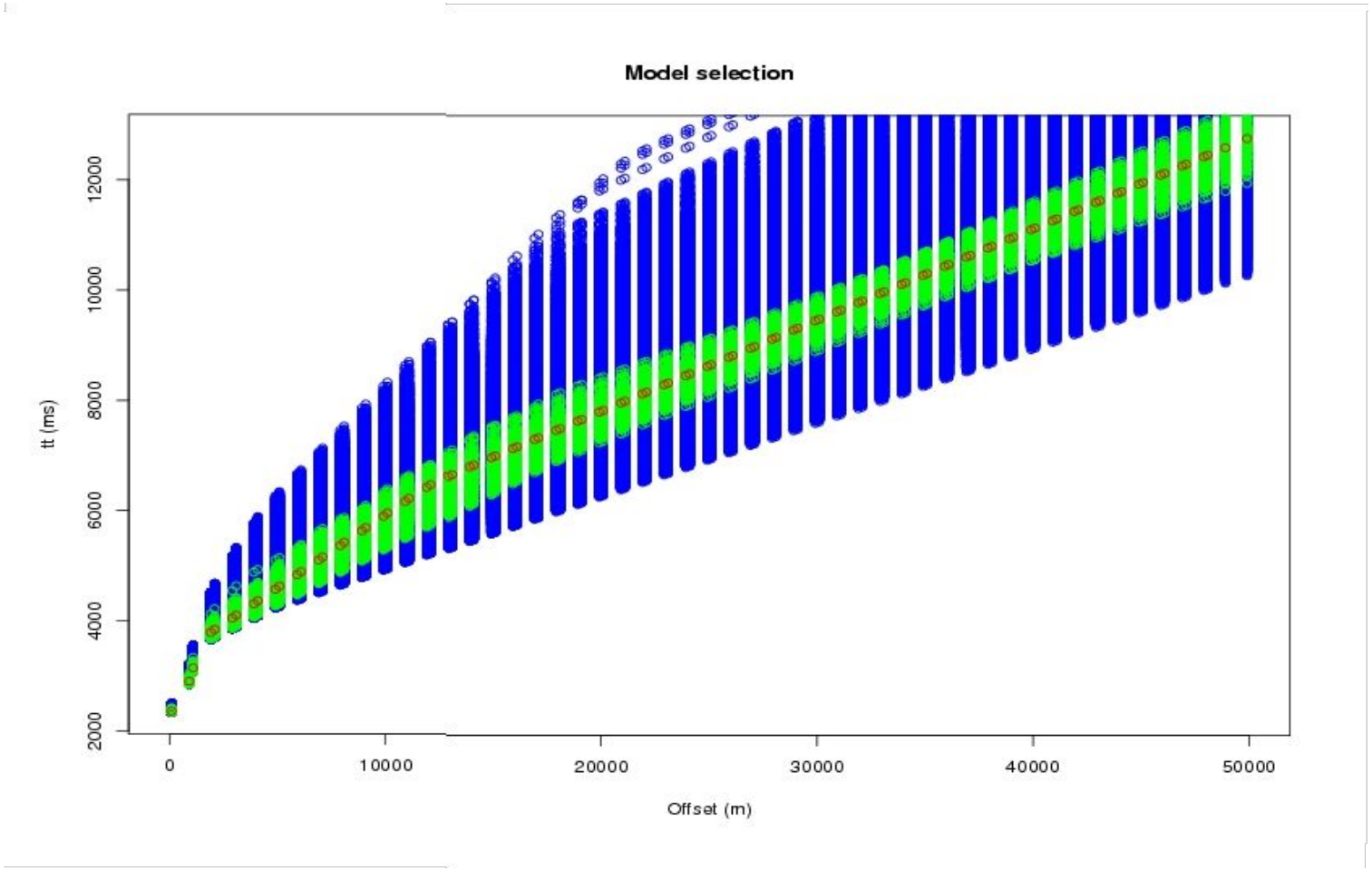


Building an emulator



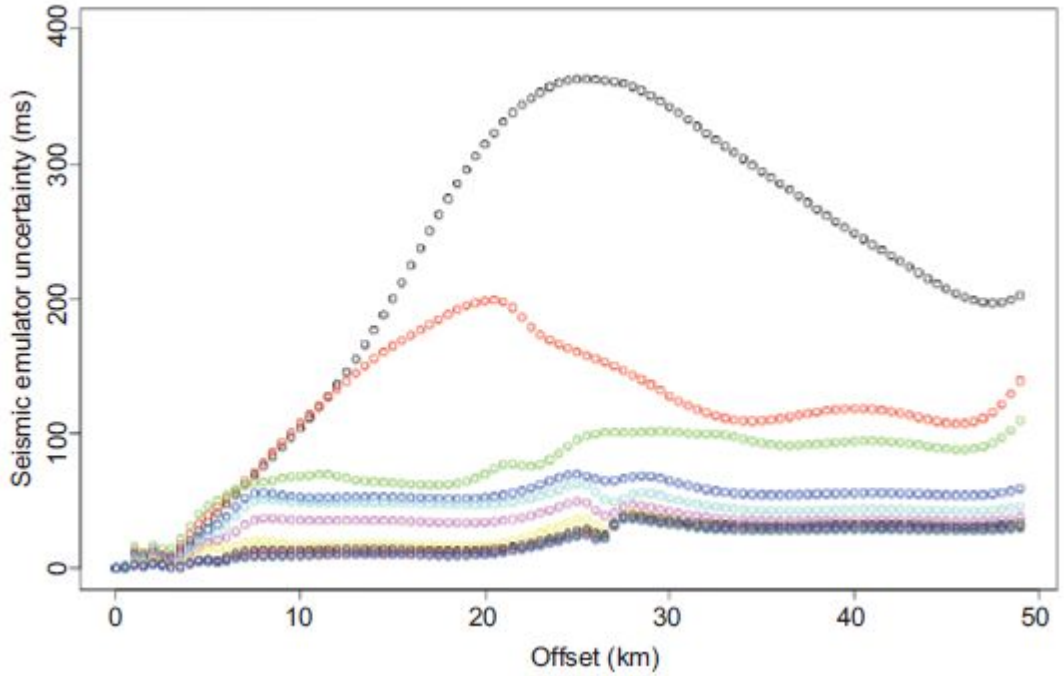
Where h(x) is the polynomial and g(x) the calibrated uncertainty

Using an emulator



Using an emulator

- The resultant error in the emulator is reduced on each iteration when implausible model space is removed



Roberts, A.W., et al., Crustal constraint through complete model space screening for diverse geophysical datasets facilitated by emulation, Tectonophysics (2012), doi:10.1016/j.tecto.2012.03.006

Bayes

- Why do classic inversion – now you know its limitations?

- Useful References:

Tutorial: http://www.sciencedirect.com/science/article/pii/S1007570415000428

MUCM toolkit:
http://mucm.aston.ac.uk/MUCM/MUCMToolkit/index.php?page=MetaHomePage.html

Goldstein book chapter on model adequacy:
http://onlinelibrary.wiley.com/doi/10.1002/9781118351475.ch26/summary

Goldstein and Wooff book:
http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0470015624.html

A decent website with basic Bayesian theory: http://www.bayesian-inference.com/index