A Tutorial on inverse problems and model space search

Malcolm Sambridge

Research School of Earth Sciences

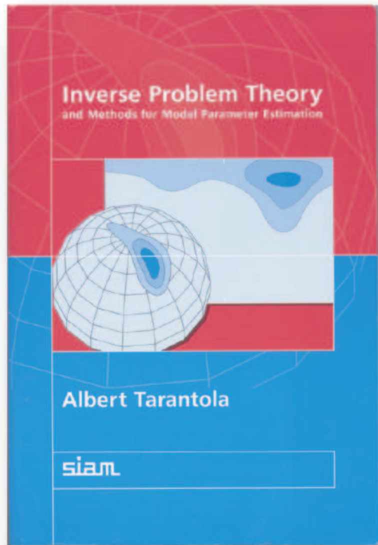
Australian National University

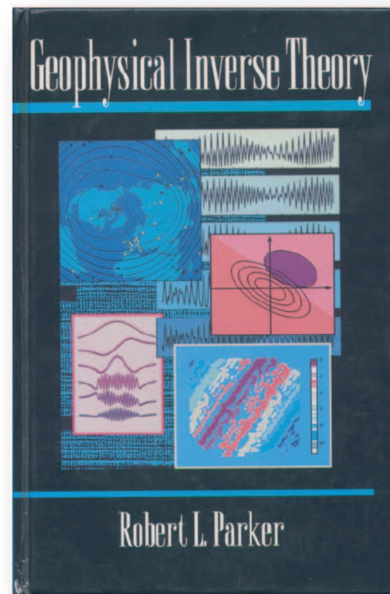
25th Course of the International School of Geophysics

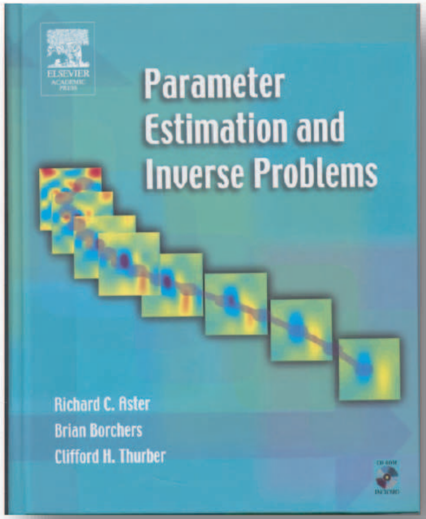
9th Int. Workshop on Numerical Modeling of Mantle Convection and Lithospheric Dynamics

8-14 September 2005, EMFCSC, Erice, Sicily.

books







See also Samizdat press (http://samizdat.mines.edu)

A tutorial on inverse problems

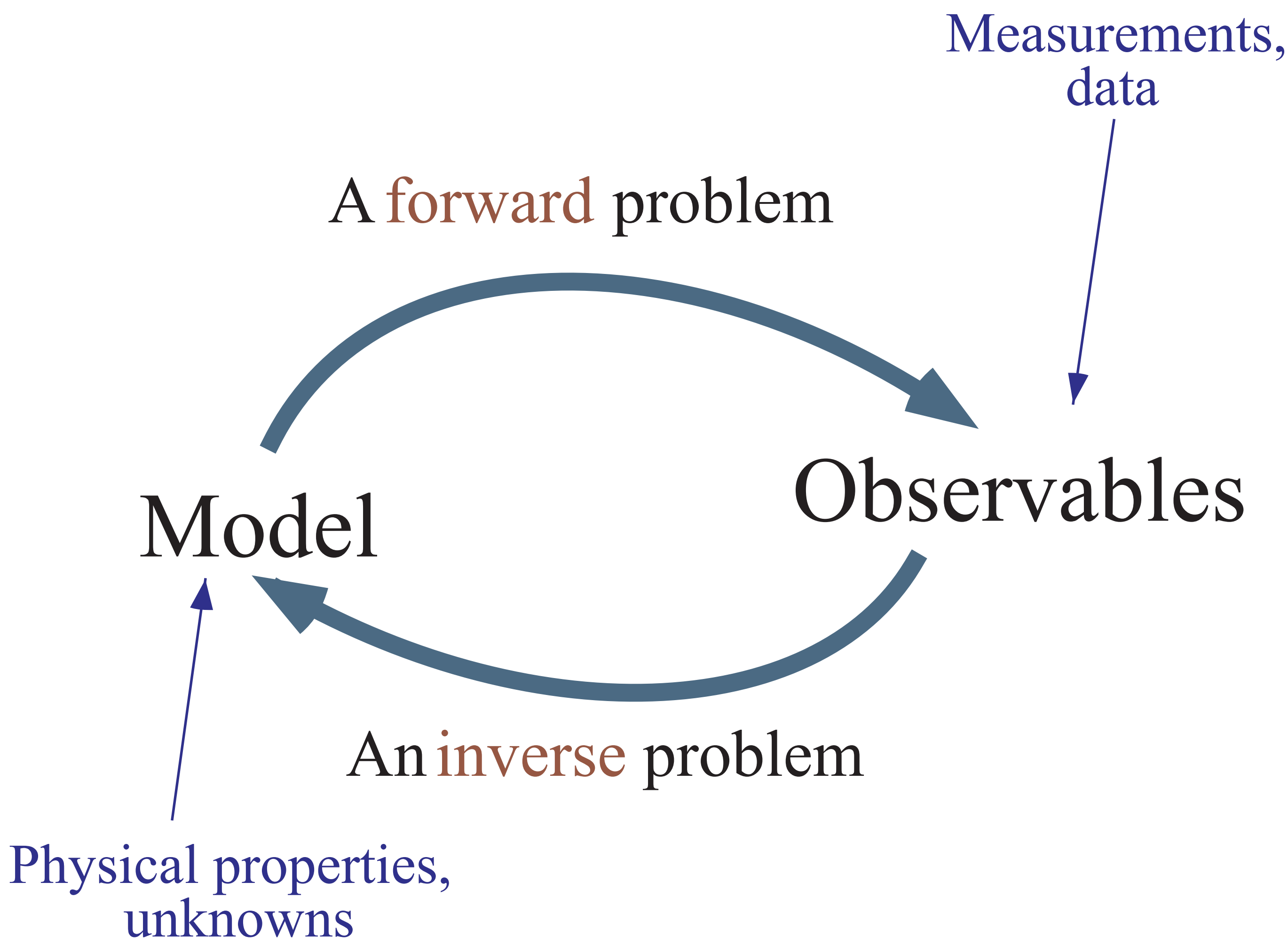
- Principles of inverse problems
- Fitting data and nonuniqueness
- Iterative methods
- Direct search methods and applications
- Uncertainty and Bayesian inference

What is an inverse problem ?

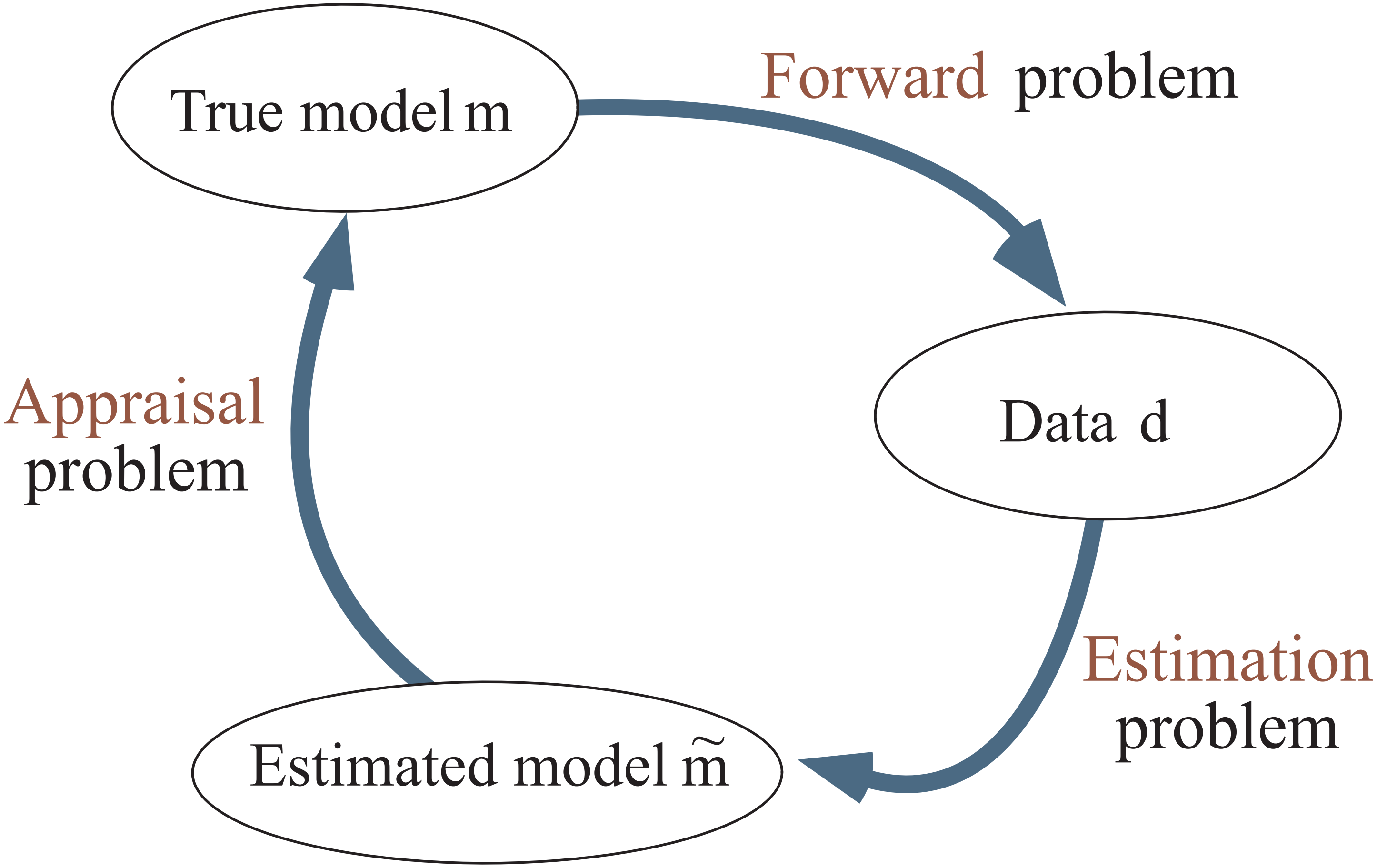
Most people, if you describe a train of events to them, will tell you what the result would be. There are few people, however, who, if you told them a result, would be able to evolve from their own inner consciousness what the steps were which led up to that result. This power is what I mean when I talk of reasoning backwards.

Sherlock Holmes *A Study in Scarlet* by Arthur Conan Doyle

Forward and inverse problems

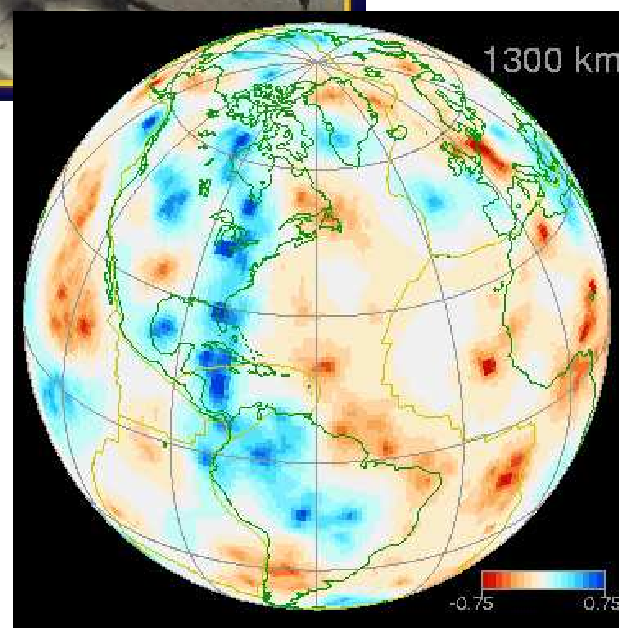


Estimation and Appraisal

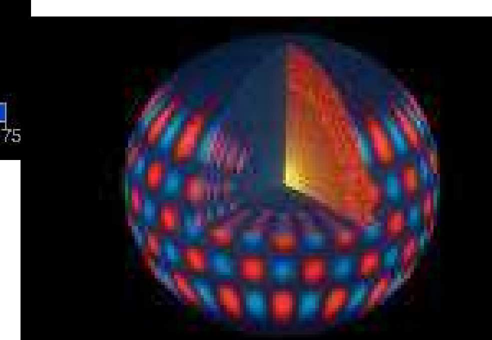


Many inverse problems

Medical tomography
1970s



Seismic tomography
1980s



Helioseismology
1990s

A philosophy for inverse problems

A way of asking questions of data !

The information you get back depends upon:

- *The question you pose,*
- *The data you have,*
- *How you define fit to data,*
- *Your parameterization of the unknowns,*

$$m(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \mathbf{B}_i(\mathbf{x})$$

- *Your definition of a solution.*

Types of forward problem

Linear,

$$\boldsymbol{d} = G\boldsymbol{m}$$

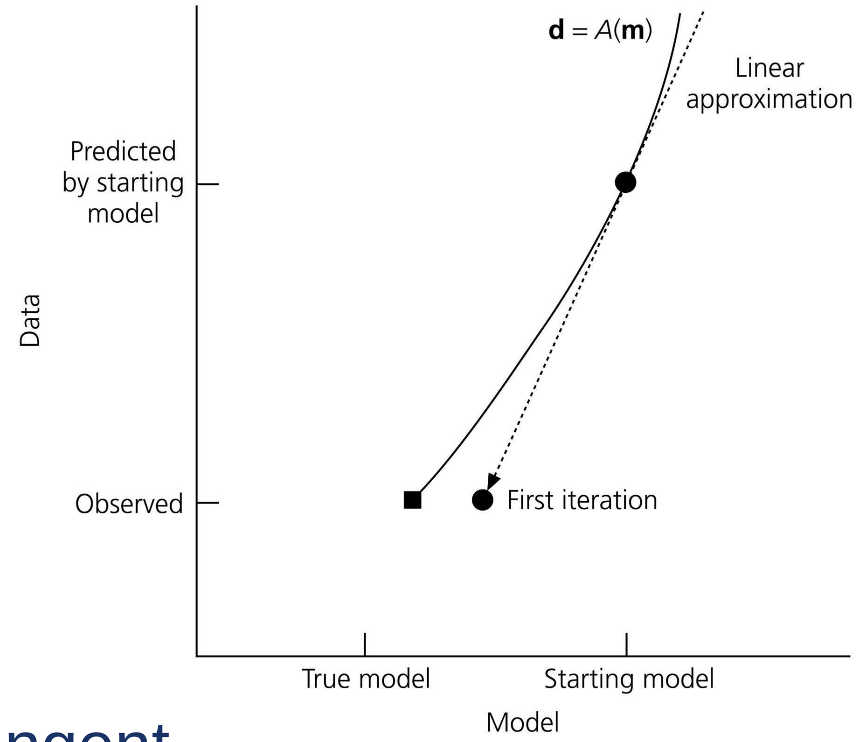
Nonlinear,

$$\boldsymbol{d} = g(\boldsymbol{m})$$

Linearized,

$$\delta\boldsymbol{d} = G\delta\boldsymbol{m}$$

Figure 7.2-2: Illustration of the the effect of linearizing about an inverse problem starting model.



Linearization is like taking a tangent.

Much of inverse theory is based on linearization...

...but its usually only an approximation !

Fitting the data

Data/model relationship,

$$\boldsymbol{d} = g(\boldsymbol{m})$$

To fit the data we need to measure data misfit,

$$\phi(\boldsymbol{d}, \boldsymbol{m}) = (\boldsymbol{d} - g(\boldsymbol{m}))^T C_D^{-1} (\boldsymbol{d} - g(\boldsymbol{m}))$$

For a linearized problem,

$$\phi(\delta \boldsymbol{d}, \delta \boldsymbol{m}) = (\delta \boldsymbol{d} - G \delta \boldsymbol{m})^T C_D^{-1} (\delta \boldsymbol{d} - G \delta \boldsymbol{m})$$

Should we just optimize $\phi(\boldsymbol{d}, \boldsymbol{m})$ with respect to $\boldsymbol{m}$?

A least squares solution

From

$$\delta \boldsymbol{d} = G \delta \boldsymbol{m}$$

we find $\delta \boldsymbol{m}$ which minimizes ϕ, $\ldots$ and get the normal equations

$$\delta \boldsymbol{m} = (G^T C_D^{-1} G)^{-1} G^T C_D^{-1} \delta \boldsymbol{d}$$

We introduce the generalized inverse as

$$\delta \boldsymbol{m} = G^{-g} \delta \boldsymbol{d}$$

Note that if data covariance matrix has the form

$$C_D^{-1} = \sigma^{-2} I$$

the estimated model is independent of the data errors !

Travel time tomography

Travel time equation

$$t = \int_{R_o} \frac{1}{v(\boldsymbol{x})} dl = \int_{R_o} s(\boldsymbol{x}) dl$$

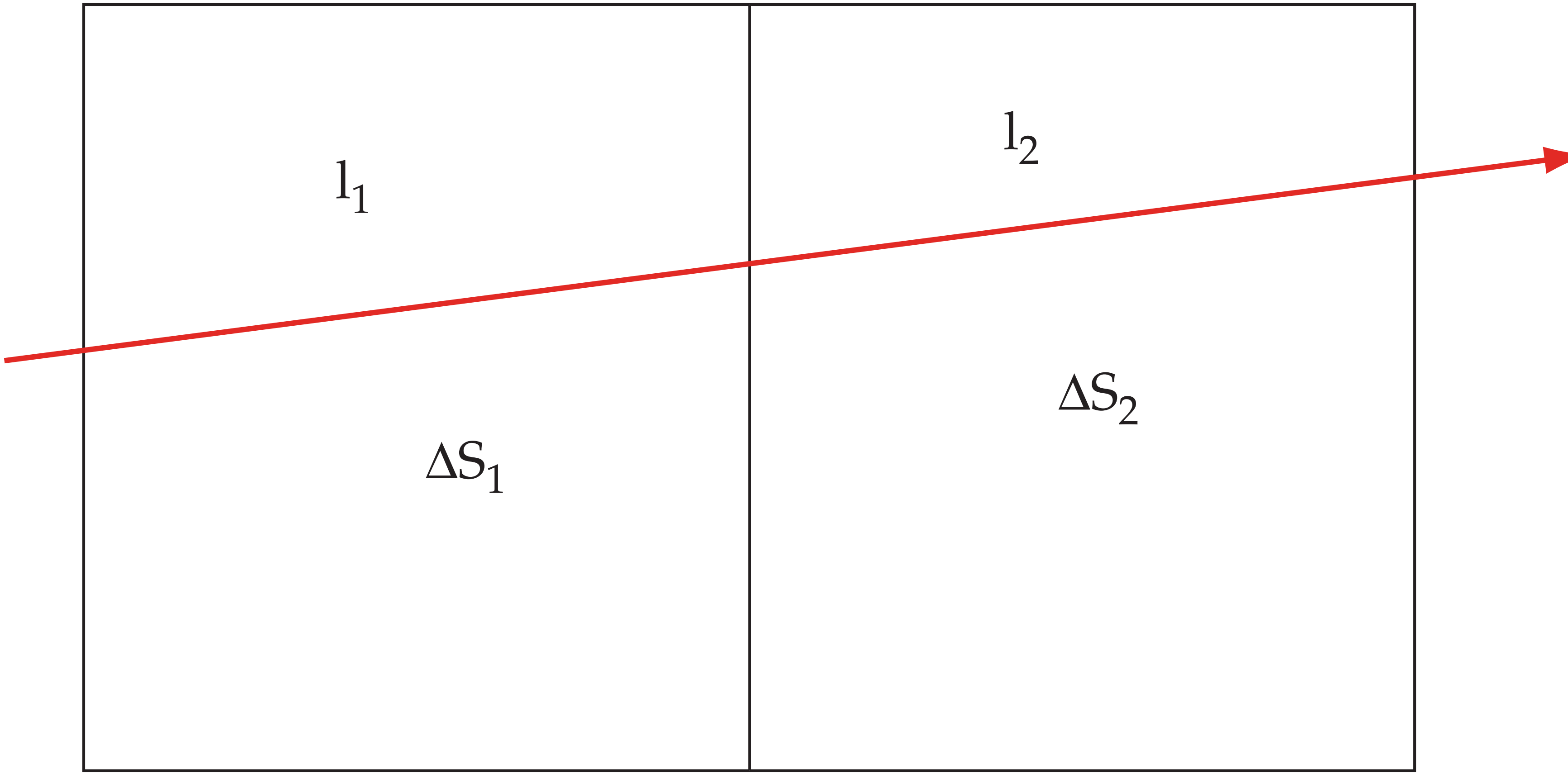
If we choose a reference slowness field $s_o(\boldsymbol{x})$ and linearize the relationship about it, we get

$$\delta t = \int_{R_o} \delta s(\boldsymbol{x}) dl$$

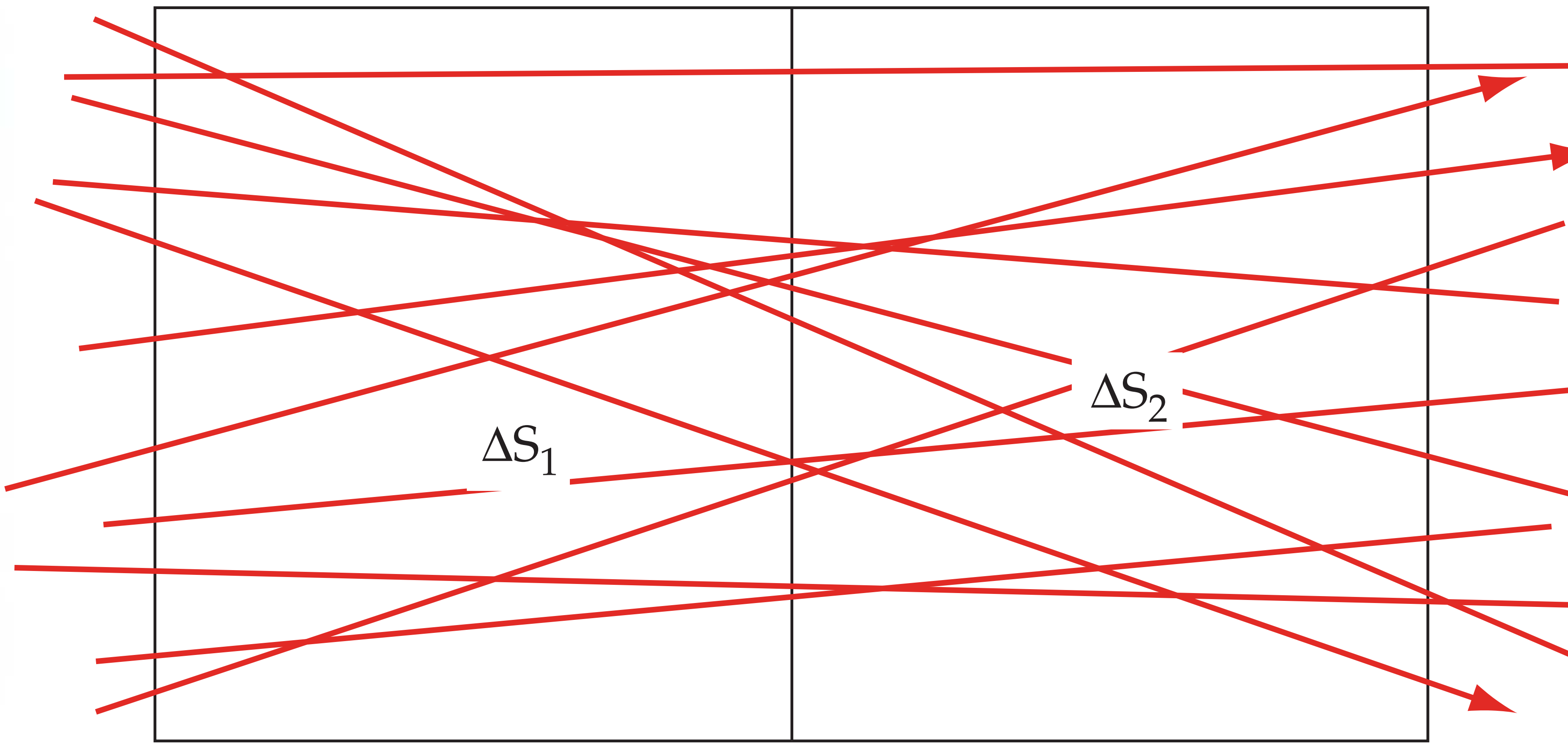
The basis of all travel time tomography.
Discretization: Choose a set of basis functions

$$\delta s(\boldsymbol{x}) = \sum_{j=1}^{M} m_j \phi_j(\boldsymbol{x}) \quad \Rightarrow \quad \boxed{\delta \boldsymbol{d} = G \delta \boldsymbol{m}}$$

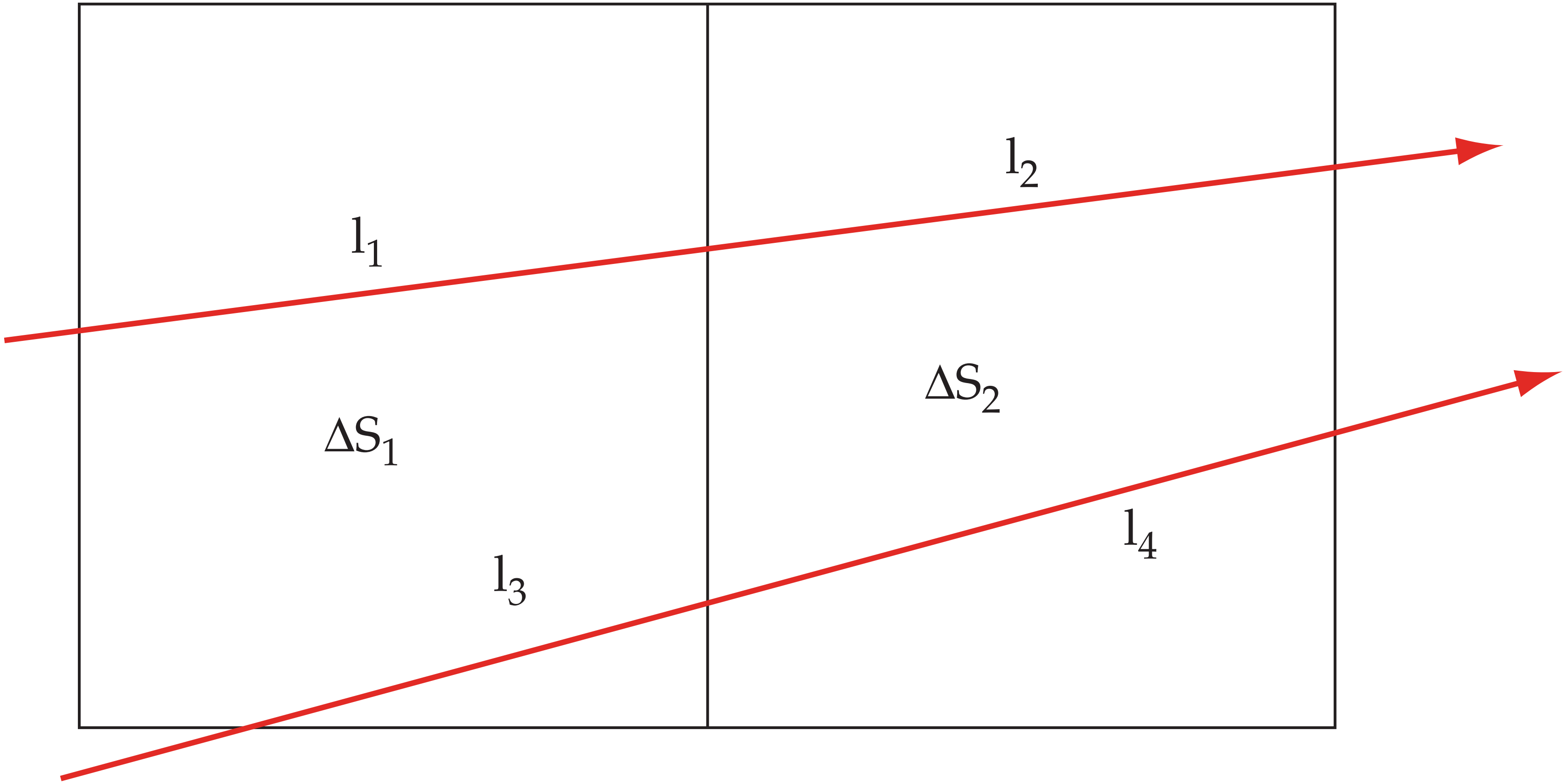
Seismic imaging: a simple experiment



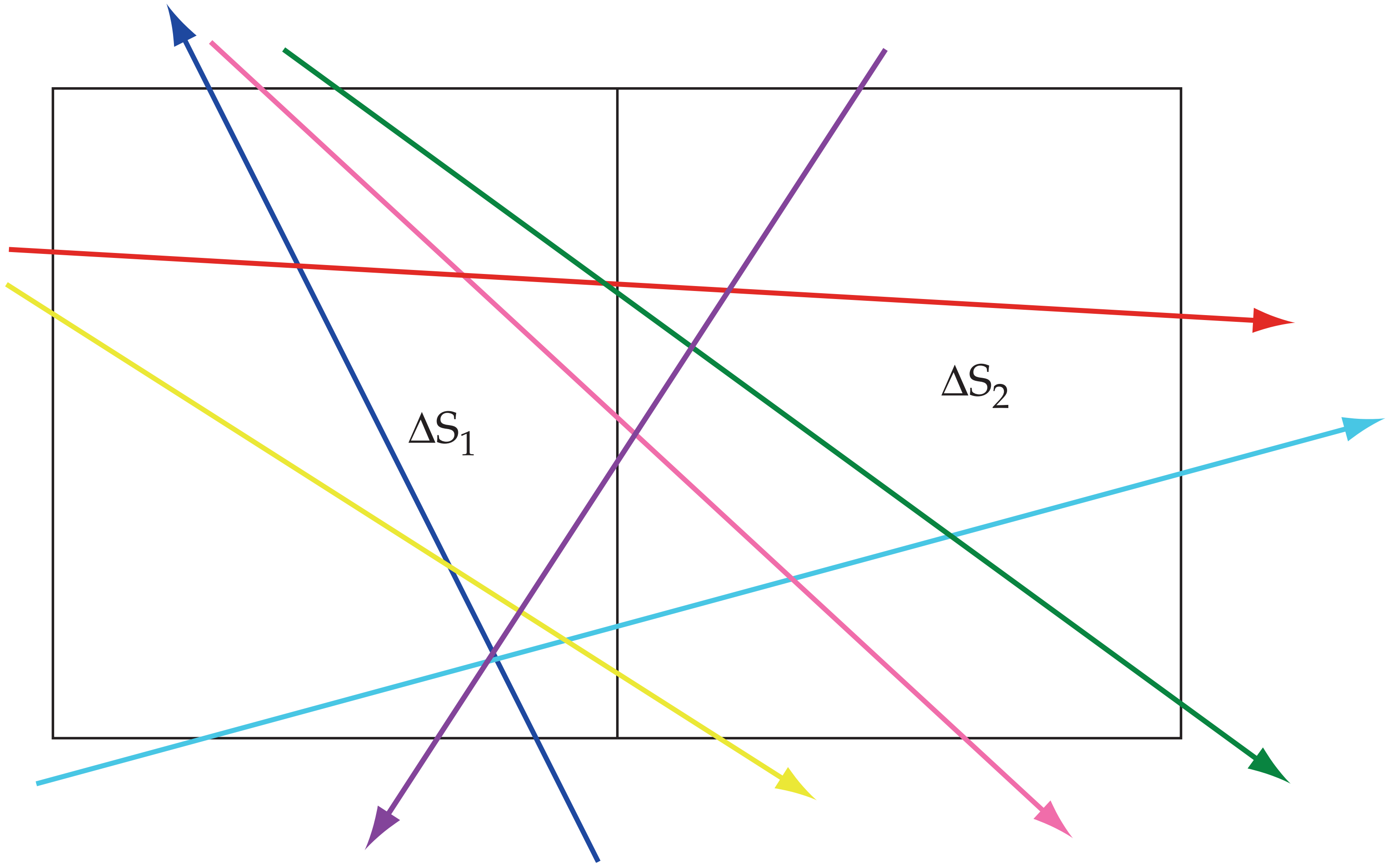
Seismic imaging: a simple experiment



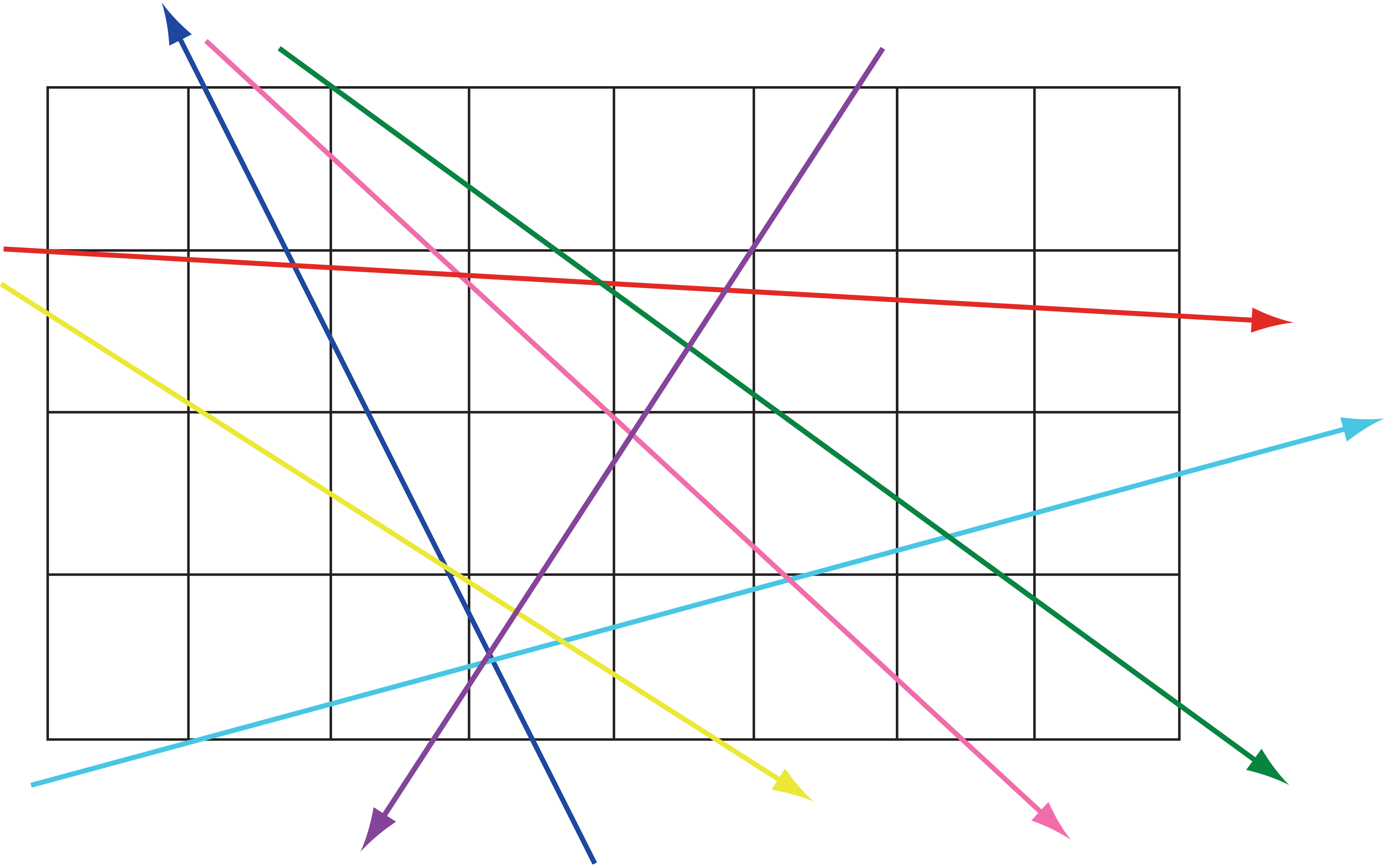
Seismic imaging: a simple experiment



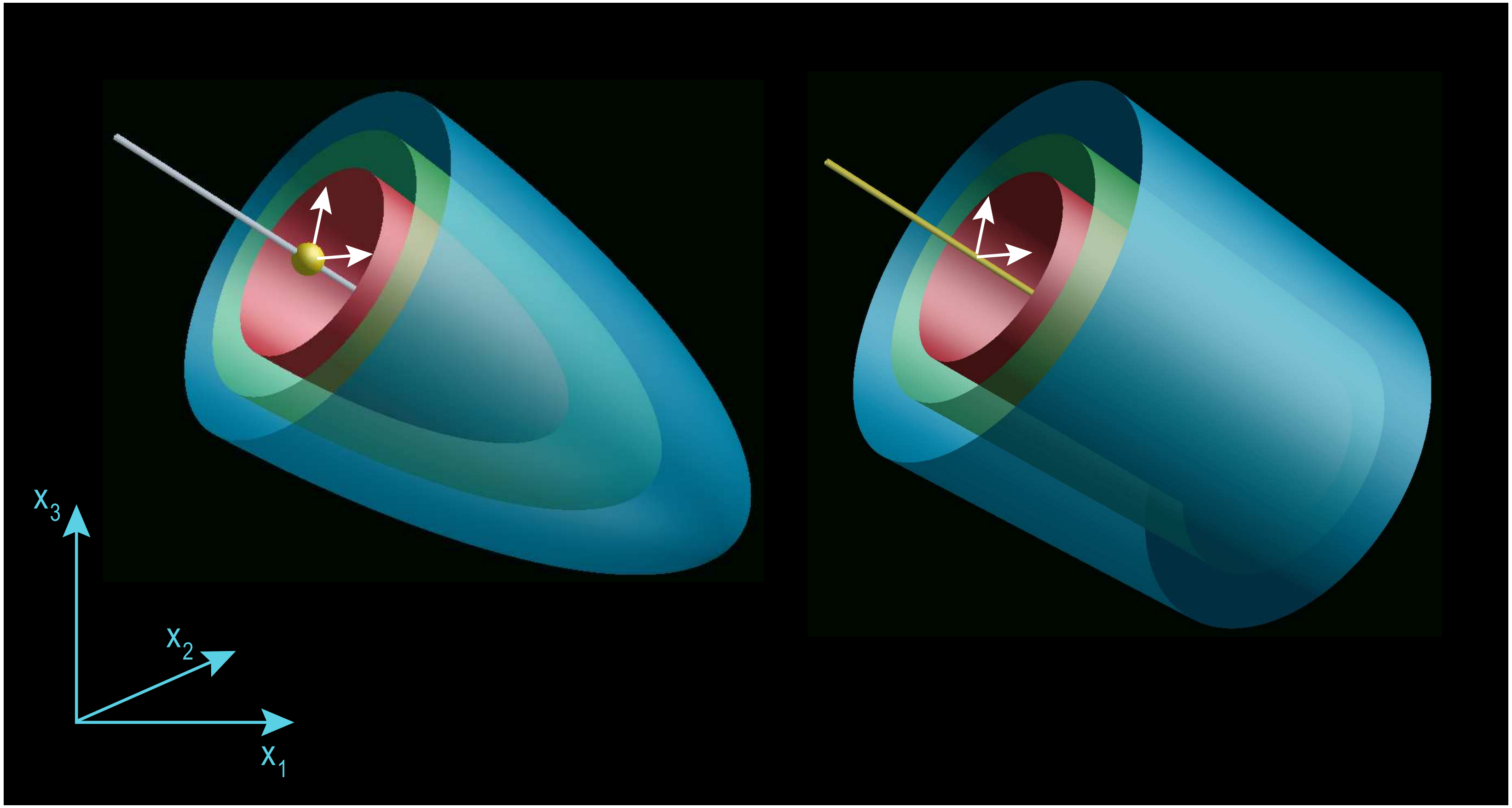
Seismic imaging: a simple experiment



Seismic imaging: a simple experiment



Linear problems and non-uniqueness



Should we just optimize data misfit ?

$$\phi(\boldsymbol{d},\boldsymbol{m}) = (\boldsymbol{d} - G\boldsymbol{m})^T C_D^{-1}(\boldsymbol{d} - G\boldsymbol{m}).$$

Regularization in inverse problems

When the problem is under or mixed-determined we can minimize a combination of data fit and model control.

$$\Psi(\boldsymbol{m}) = \phi(\boldsymbol{d}, \boldsymbol{m}) + \lambda^2 \psi(\boldsymbol{m})$$

λ is a trade-off parameter that must be chosen. It adds stability but decreases resolution. If the regularization is chosen $\psi(\boldsymbol{m}) = (\boldsymbol{m} - \boldsymbol{m}_o)^T C_M^{-1} (\boldsymbol{m} - \boldsymbol{m}_o)$, we get

$$\boldsymbol{m}_{n+1} = \boldsymbol{m}_n + (G^T C_D^{-1} G + \lambda^2 C_M^{-1})^{-1} (G^T C_D^{-1} \delta \boldsymbol{d} - \lambda^2 C_M^{-1} (\boldsymbol{m}_n - \boldsymbol{m}_o))$$

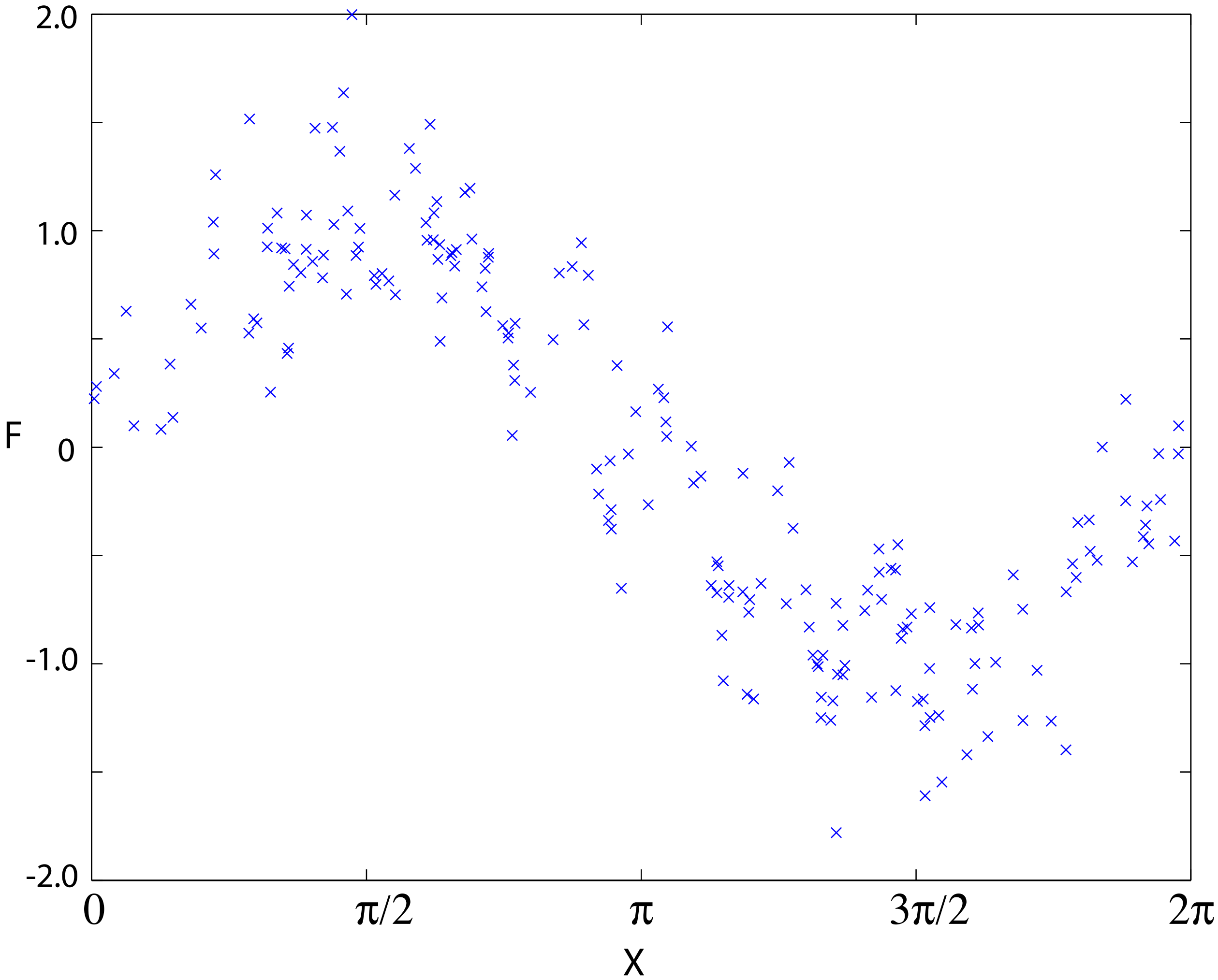
This gives a minimum variance solution. The poorly constrained parts of the model are damped towards the reference model.
An alternative is a Laplacian operator

$$\psi(\boldsymbol{m}) = |||L\boldsymbol{m}||^2 = \boldsymbol{m}^T L^T L \boldsymbol{m}$$

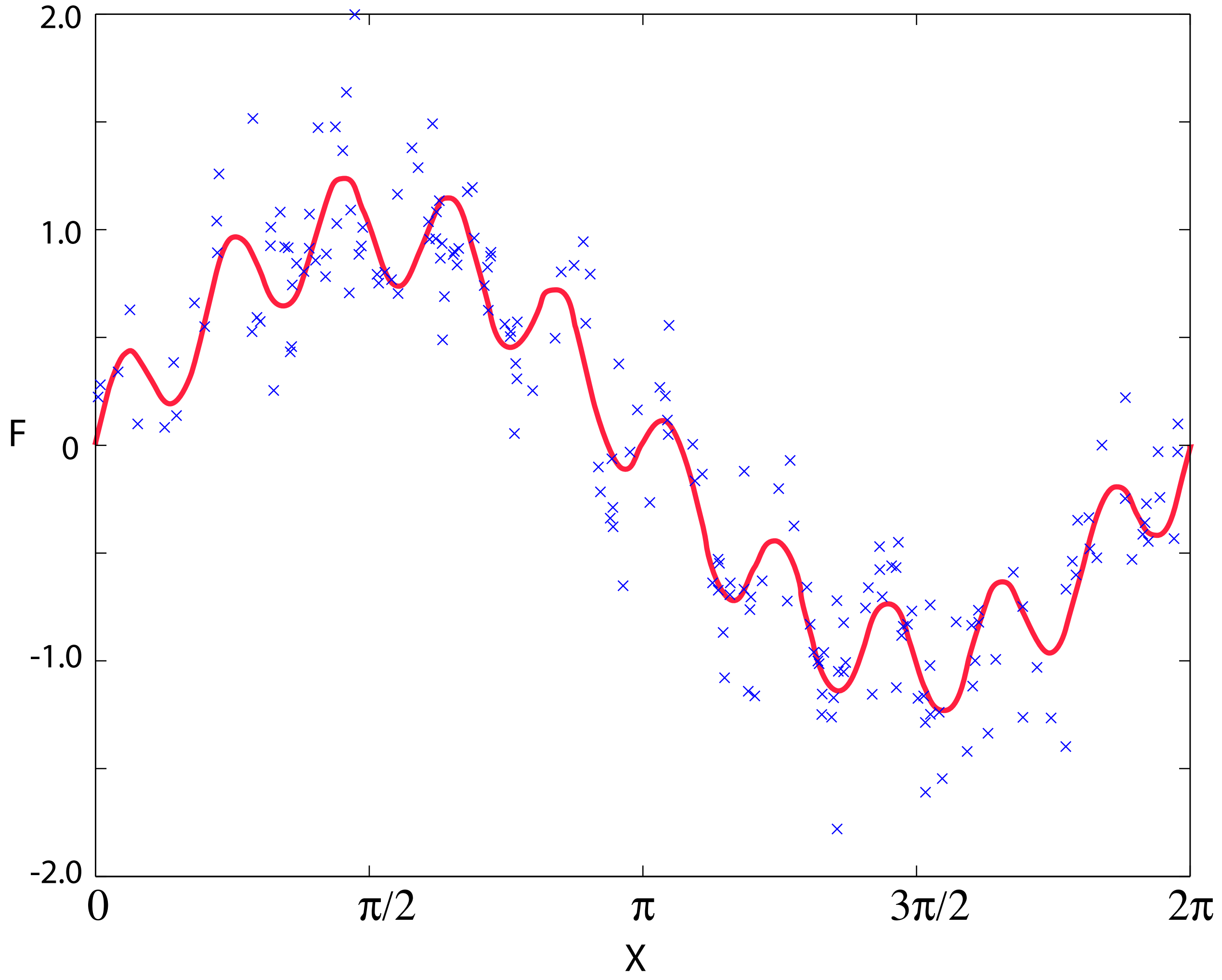
L is a finite difference approximation to ∇. This minimizes model roughness ($\frac{\partial^2 m}{\partial x^2}$) or flatness $\frac{\partial m}{\partial x}$.

$$\min_{\boldsymbol{m}} \{\psi(\boldsymbol{m})\} \quad \text{s.t.} \quad \phi(\boldsymbol{d}, \boldsymbol{m}) < \phi^*$$

Example: smoothing data



Example: smoothing data



Constructing smooth models - theory

Typically we would want to fit the data and regularize or smooth the model at the same time.

$$\psi(\boldsymbol{d}, \boldsymbol{m}) = \sum_{i=1}^{N} (d_i - s(\boldsymbol{x}_i, \boldsymbol{m}))^2 + \mu J(s)$$

Where,

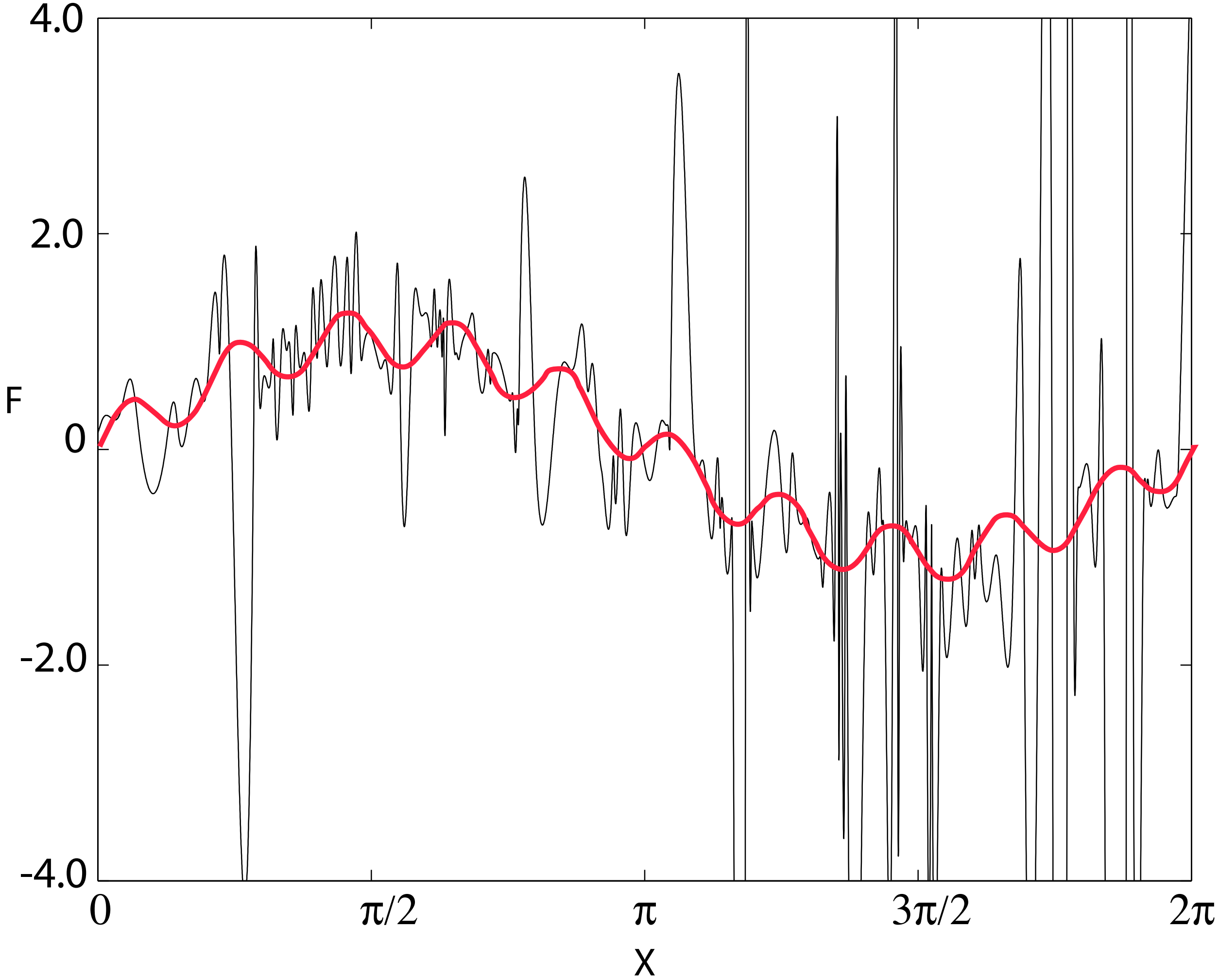
$$J(s) = \int \left[\left({\frac{\partial^2 s}{\partial x^2}}^{2} \right) + 2 \left({\frac{\partial^2 s}{\partial x \partial y}}^{2} \right) + \left({\frac{\partial^2 s}{\partial y^2}}^{2} \right) \right] d\boldsymbol{x}$$

Can we find a smooth model that fits the data exactly ?

$$s(\boldsymbol{x}, \boldsymbol{m}) = p(\boldsymbol{x}) + \sum_{i=1}^{N} \lambda_i \phi(\boldsymbol{x} - \boldsymbol{x}_i)$$

Yes ! use *Thin Plate Splines* for $\phi(\boldsymbol{x})$ (Duchon, 1976)

Smooth models - practice

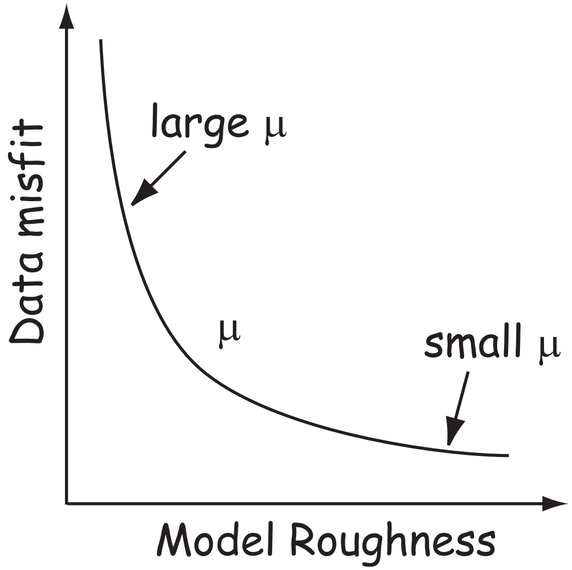


Relaxing the fit to data

We do not want to fit noisy data exactly !

$$\psi(\boldsymbol{d}, \boldsymbol{m}) = \sum_{i=1}^{N} (d_i - s(\boldsymbol{x}_i, \boldsymbol{m}))^2 + \mu J(s)$$

In order to relax the requirement to fit the data we must find a value of the trade-off parameter μ.



Choosing trade-off parameter

One way of finding a balance between data fit and model smoothness is Generalized Cross Validation - which essentially means use the data to find a value for μ.

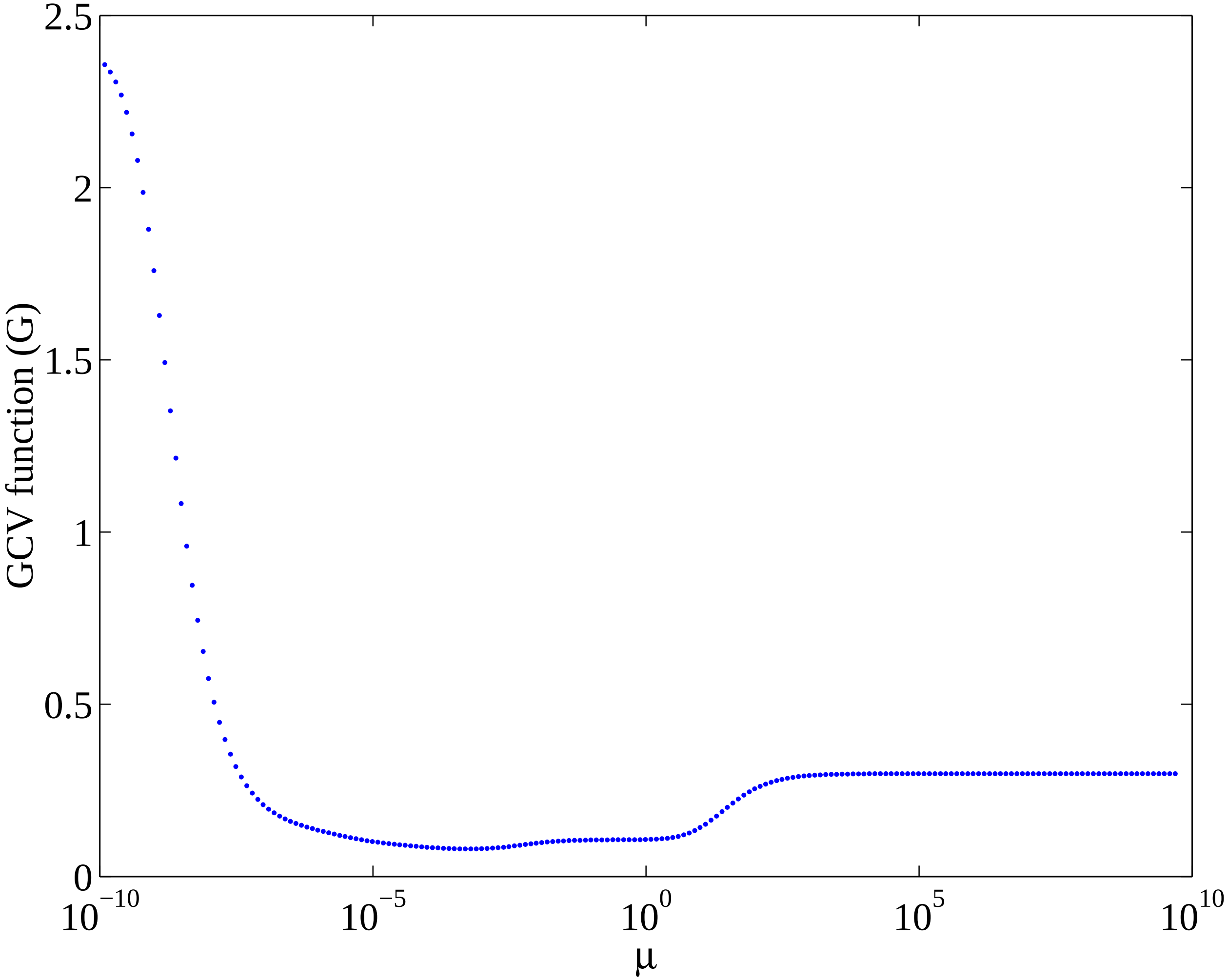
$$G(\mu) = \sum_{i=1}^{N} (d_i - s_i(\boldsymbol{x}_i, \boldsymbol{m}))^2$$

Where $s_i(\boldsymbol{x}, \boldsymbol{m})$ is the TPS interpolant produced when the ith datum is removed. Find μ that minimizes $G(\mu)$. Note

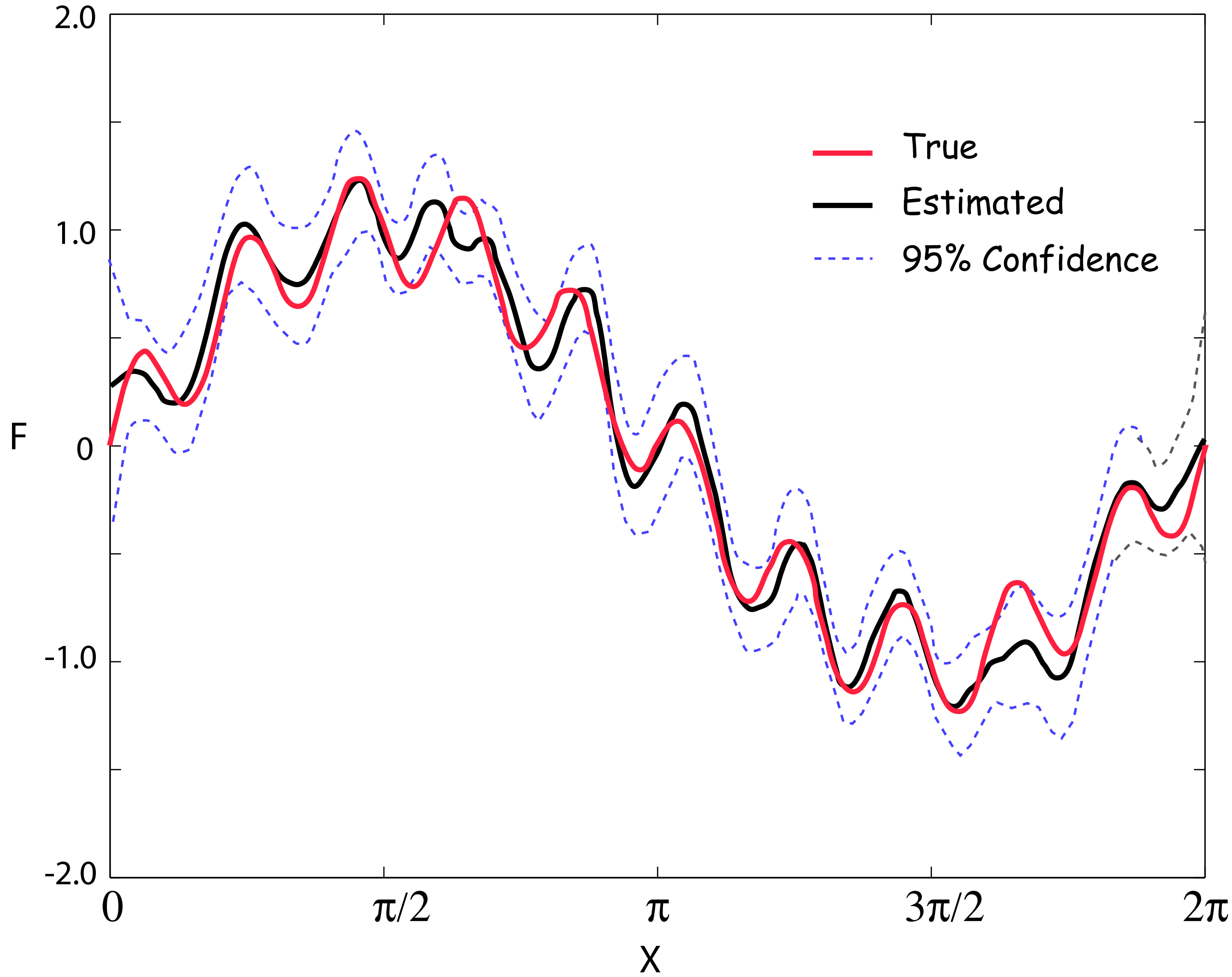
$$\mu \to \infty \Rightarrow G(\mu) \uparrow$$
$$\mu \to 0 \Rightarrow G(\mu) \uparrow$$

$G(\mu)$ is a bootstrap measure of interpolation error.

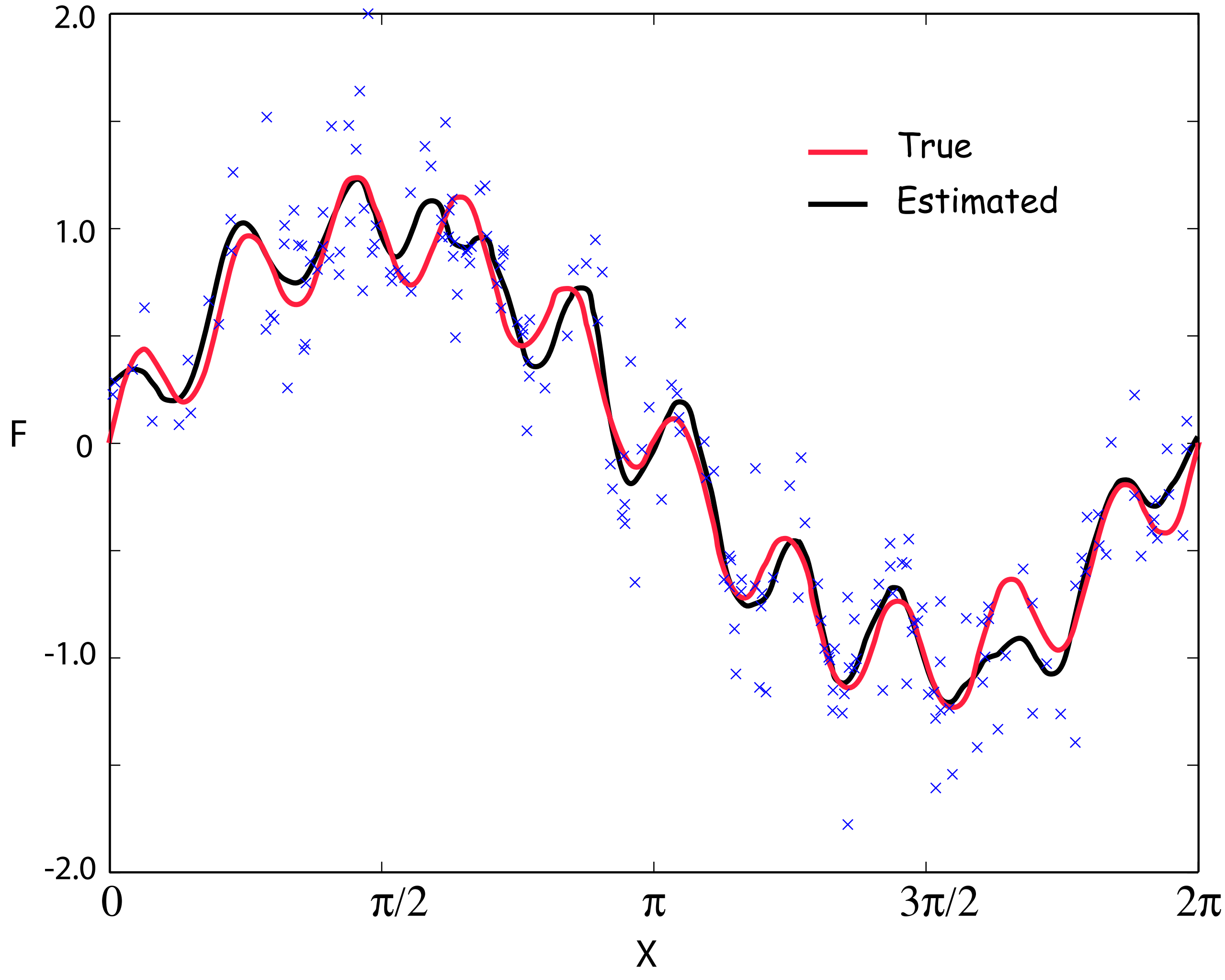
Minimizing GCV to find μ



Generalized cross validation



Generalized cross validation



Features of discrete inverse problems

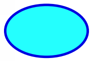
- Linearization is an approximation
- Parametrization is a choice
- Nonuniqueness can occur
  - over determined
  - even determined
  - under determined
- More data reduces input noise but independent data matters most.
- Trade-off between model variance and resolution (spread)

More worked examples available: Over and under-determined linear systems, error propagation, SVD, resolution and covariance matrices.

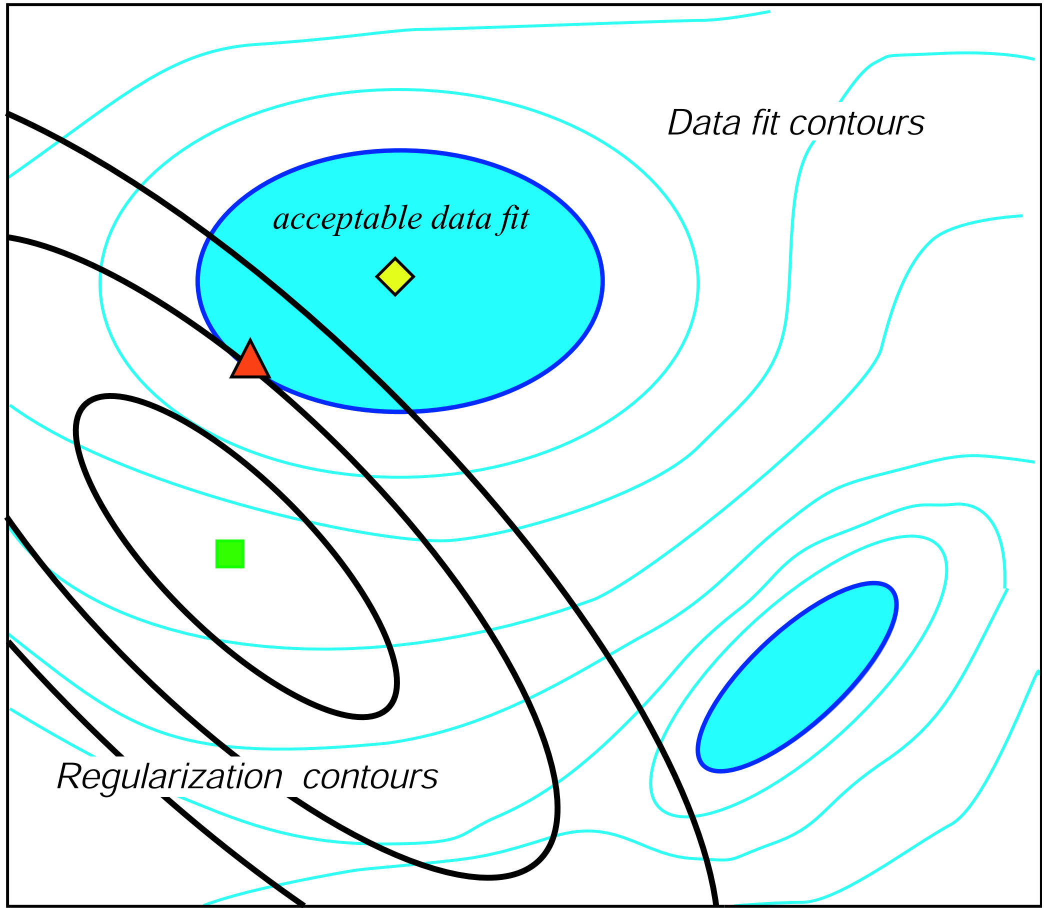
What is a solution to an inverse problem ?

- Optimal data fit solution (c.f. MAP)

- Extremal solution

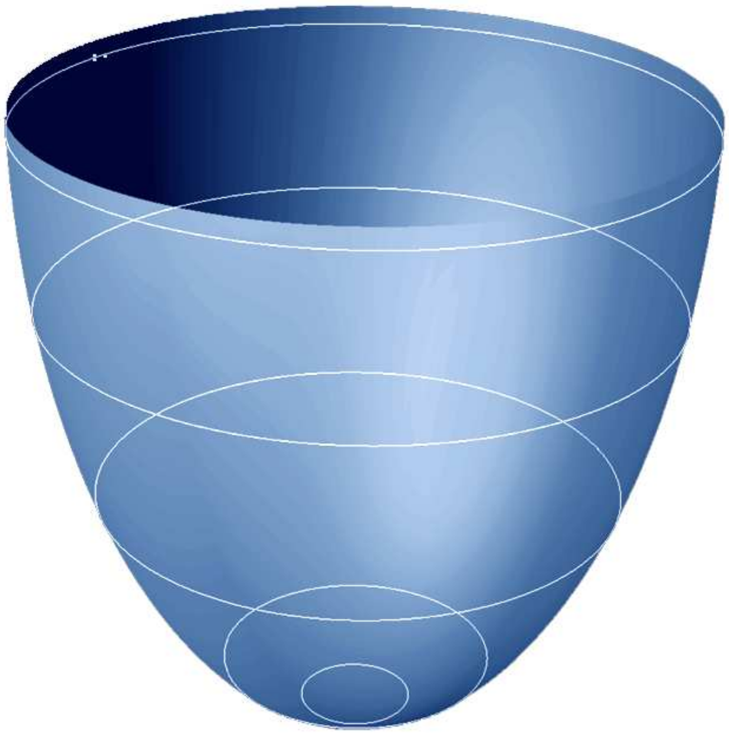


- Data acceptable solutions



Linear problems

- Single minima,
- Gradient methods work,
- Quadratic convergence,
- Many unknowns,



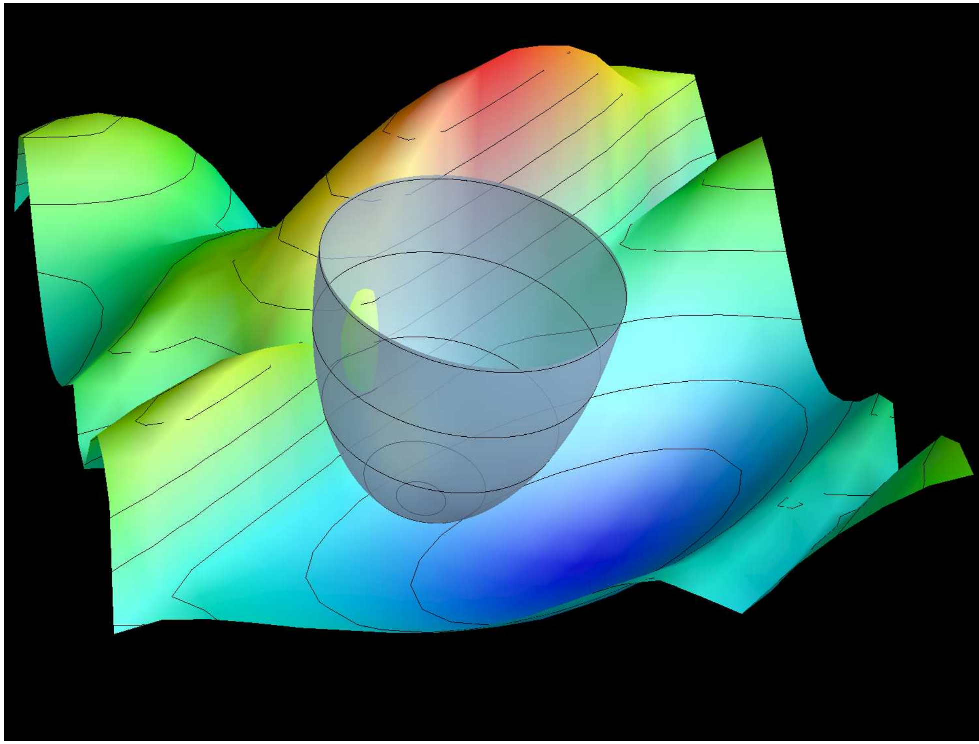
$$\mathbf{d} = G\mathbf{m}$$

$$G_{i,j} = \frac{\partial d_i}{\partial m_j}$$

$$\phi(\boldsymbol{d}, \boldsymbol{m}) = (\boldsymbol{d} - G\boldsymbol{m})^T C_D^{-1} (\boldsymbol{d} - G\boldsymbol{m}) + \lambda^2 (\boldsymbol{m} - \boldsymbol{m}_o)^T C_M^{-1} (\boldsymbol{m} - \boldsymbol{m}_o)$$

Weakly nonlinear problems

- Single minimum (?)
- Gradient methods work,
- Many unknowns,



$$\delta \boldsymbol{d} = \mathbf{G} \delta \boldsymbol{m}$$

$$G_{i,j} = \frac{\partial d_i}{\partial m_j}$$

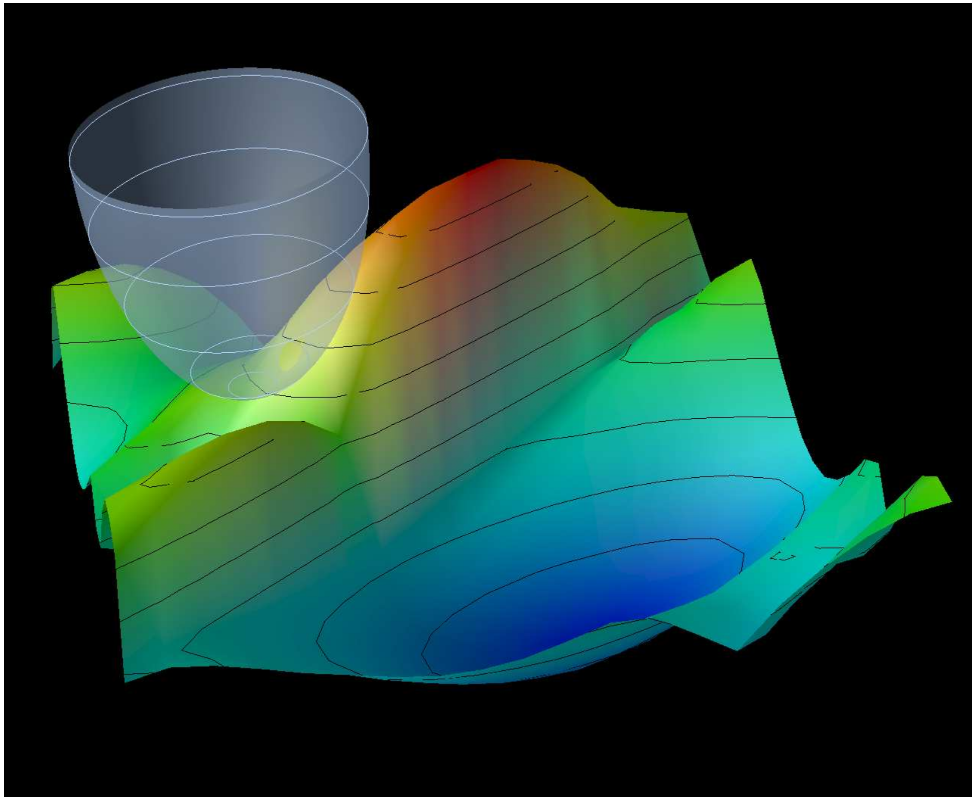
$$\phi(\boldsymbol{d}, \boldsymbol{m}) = (\delta \boldsymbol{d} - G \delta \boldsymbol{m})^T C_D^{-1} (\delta \boldsymbol{d} - G \delta \boldsymbol{m}) + \lambda^2 (\boldsymbol{m} - \boldsymbol{m}_o)^T C_M^{-1} (\boldsymbol{m} - \boldsymbol{m}_o)$$

Weakly nonlinear problems II

But gradient methods can still fail . . .

$$\delta \boldsymbol{d} = \mathbf{G} \delta \boldsymbol{m}$$

$$G_{i,j} = \frac{\partial d_i}{\partial m_j}$$

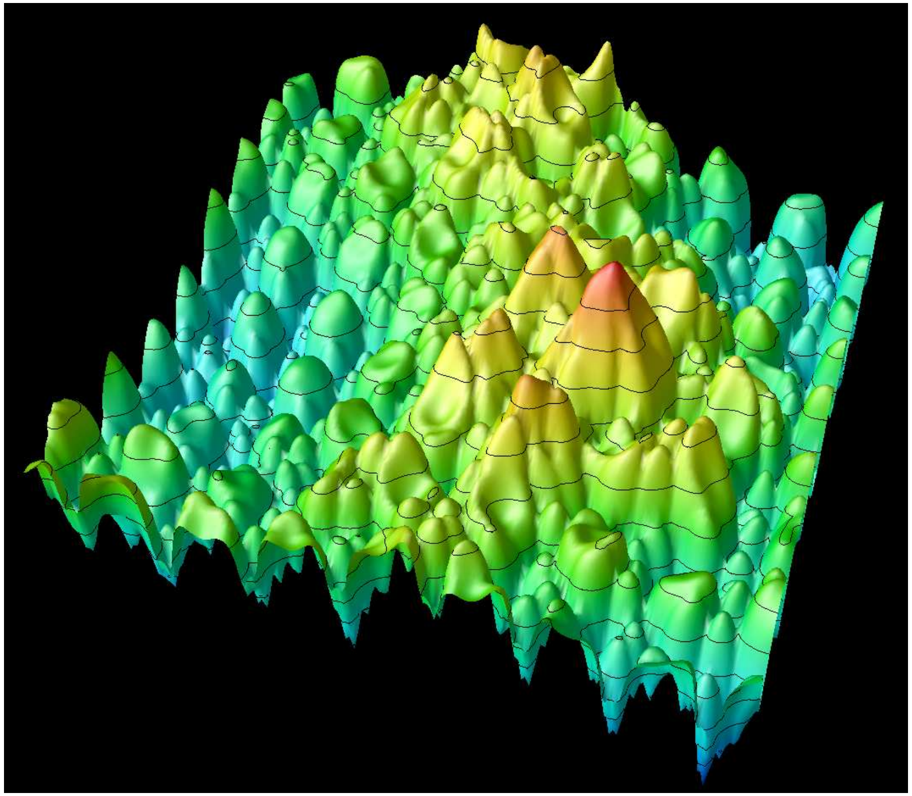


. . . if you start in the wrong place.

Strongly nonlinear problems

- Multi-modal misfits
- Linearization fails
- Direct search techniques might work
- $10^0 - 10^2$ unknowns

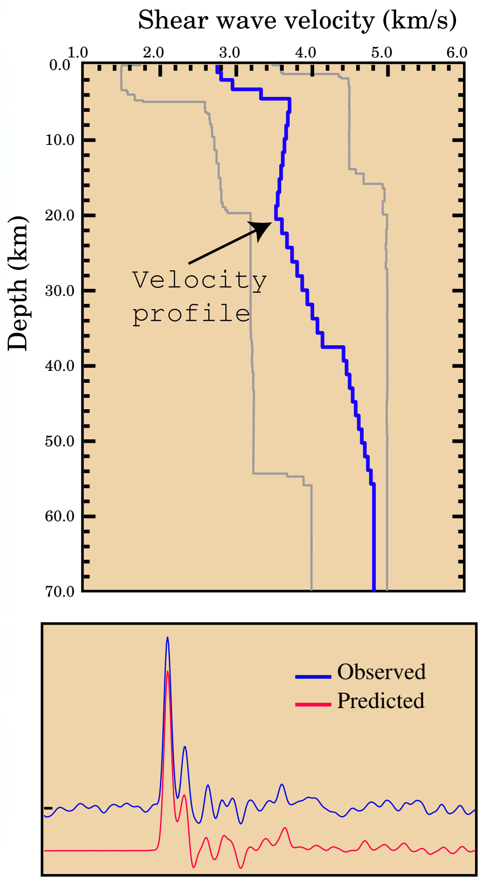
$$\mathbf{d} = g(\mathbf{m})$$

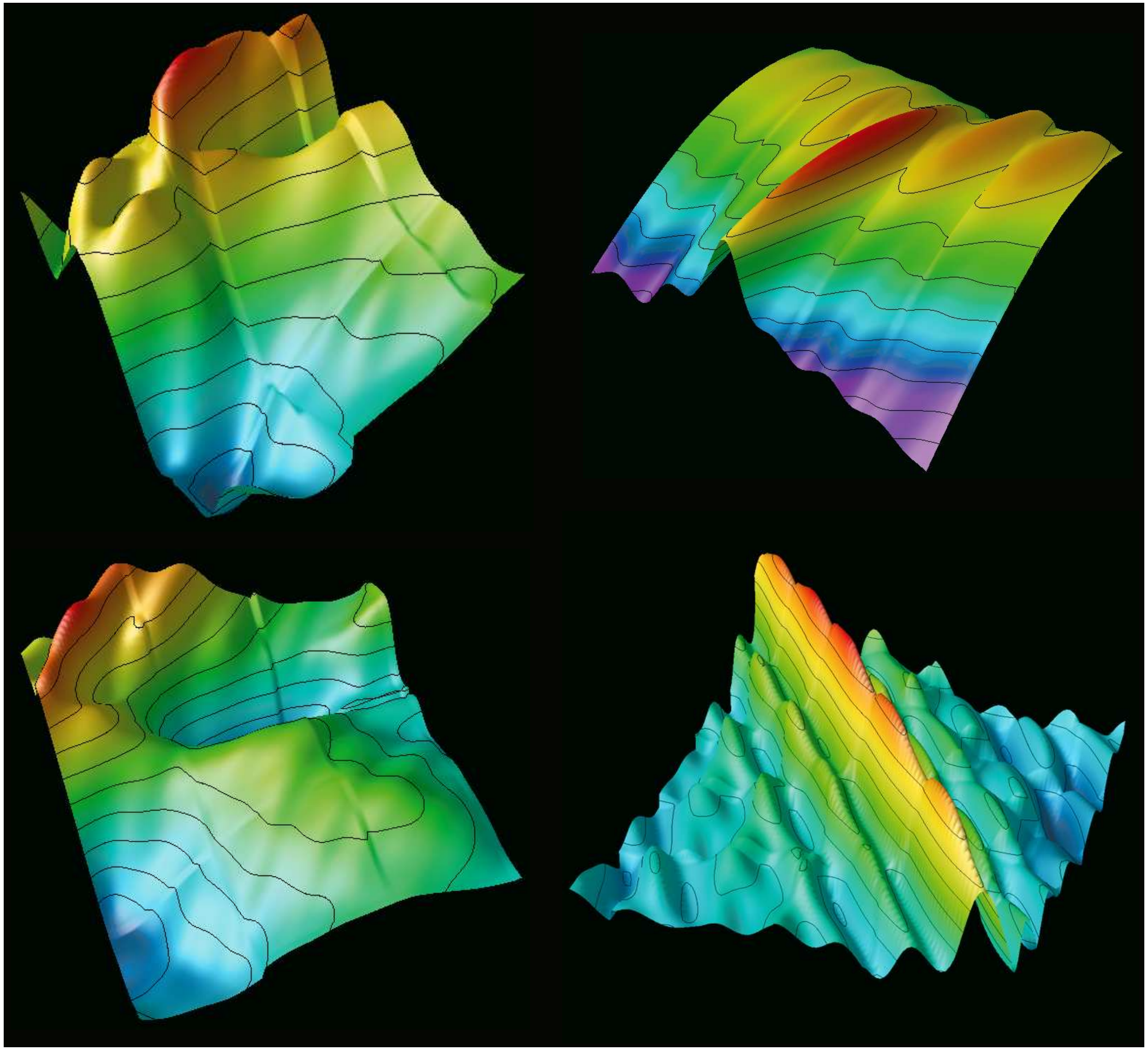


Derivatives, $\frac{\partial d_i}{\partial m_j}$, of little use !

$$\phi(\boldsymbol{d}, \boldsymbol{m}) = (\boldsymbol{d} - g(\boldsymbol{m}))^T C_D^{-1} (\boldsymbol{d} - g(\boldsymbol{m})) + \lambda^2 (\boldsymbol{m} - \boldsymbol{m}_o)^T C_M^{-1} (\boldsymbol{m} - \boldsymbol{m}_o)$$

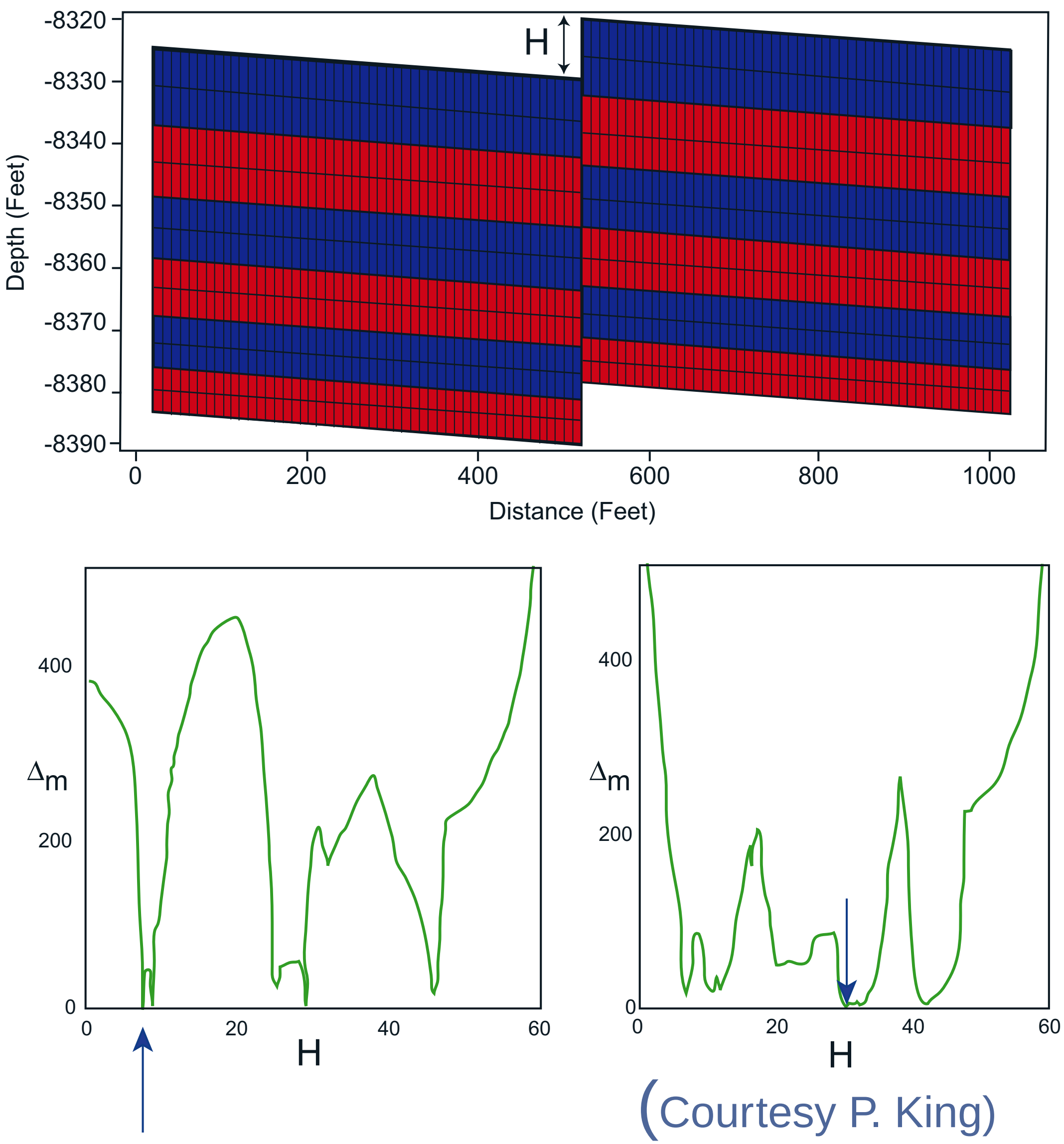
Data misfit surfaces: Receiver functions





Data misfit in History matching

A synthetic example with one unknown.



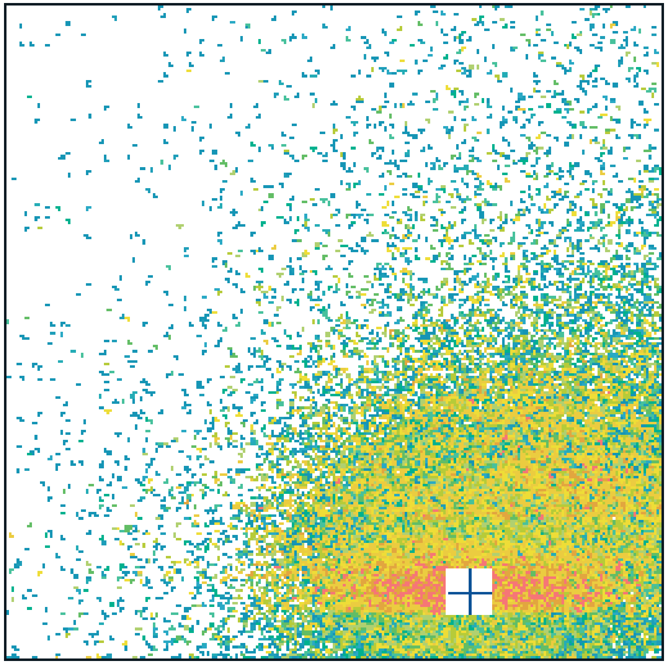
(Courtesy P. King)

Parameter space search techniques

Derivative free or direct search techniques can be useful for weakly and strongly nonlinear problems.

Classes of direct search algorithm:

- Uniform random search
- Simulated annealing (thermodynamics)
- Evolutionary algorithms (biology)
- Neighbourhood sampling (geometry)

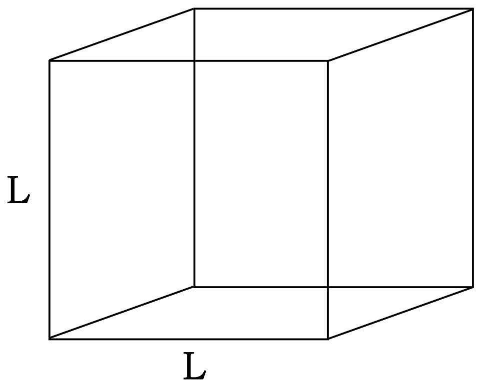


Uniform sampling

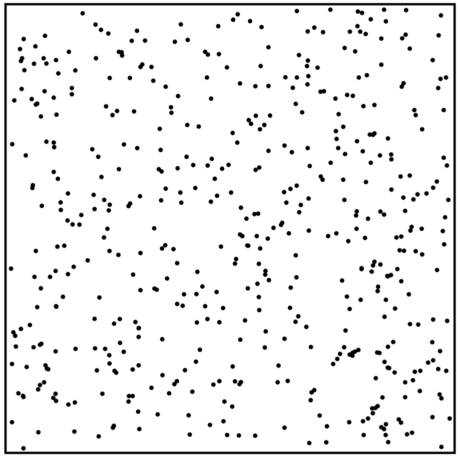
Uniform random sampling means uniform in volume !

Volume of the hypercube in d dimensions,

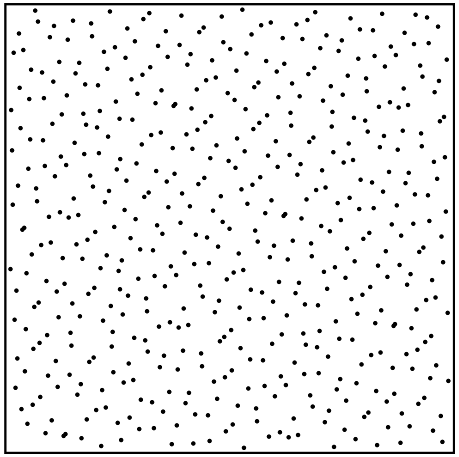
$$V = L^d$$



Curse of dimensionality



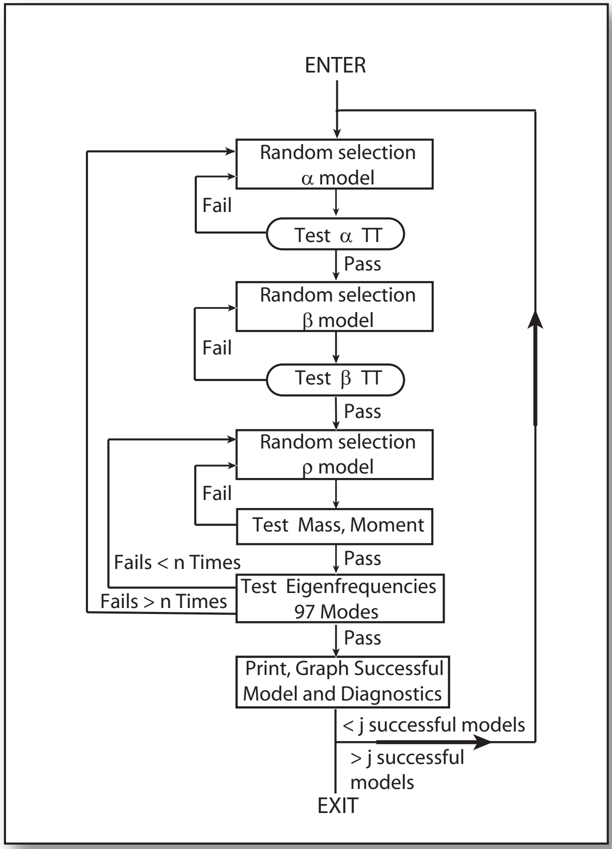
Pseudo - random

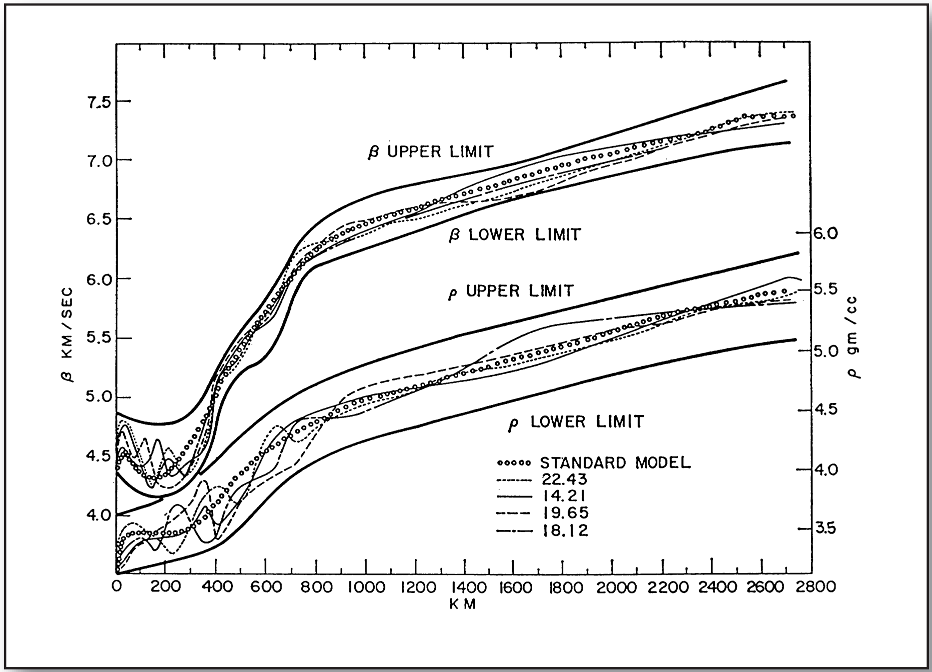


Quasi - random

Uniform Monte Carlo Inversion

A whole earth Monte Carlo inversion by Press (1968)





Keilis-Borok & Yanovskaya (1967) first introduced Monte Carlo inversion into geophysics.

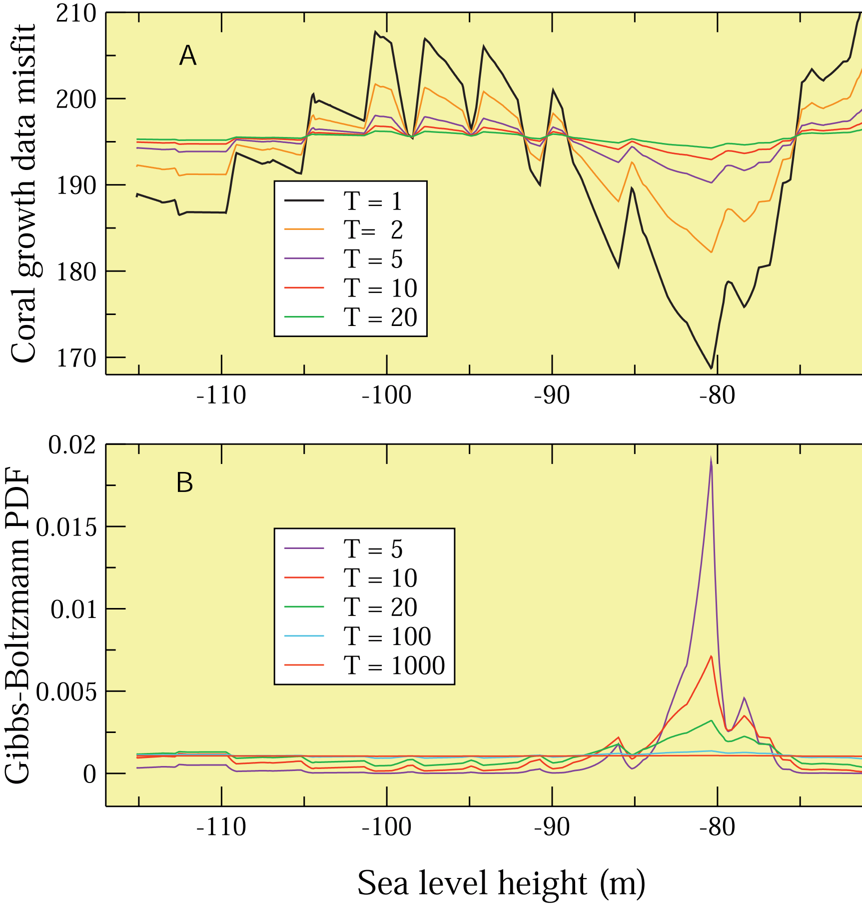
Simulated annealing

A Global optimization technique.

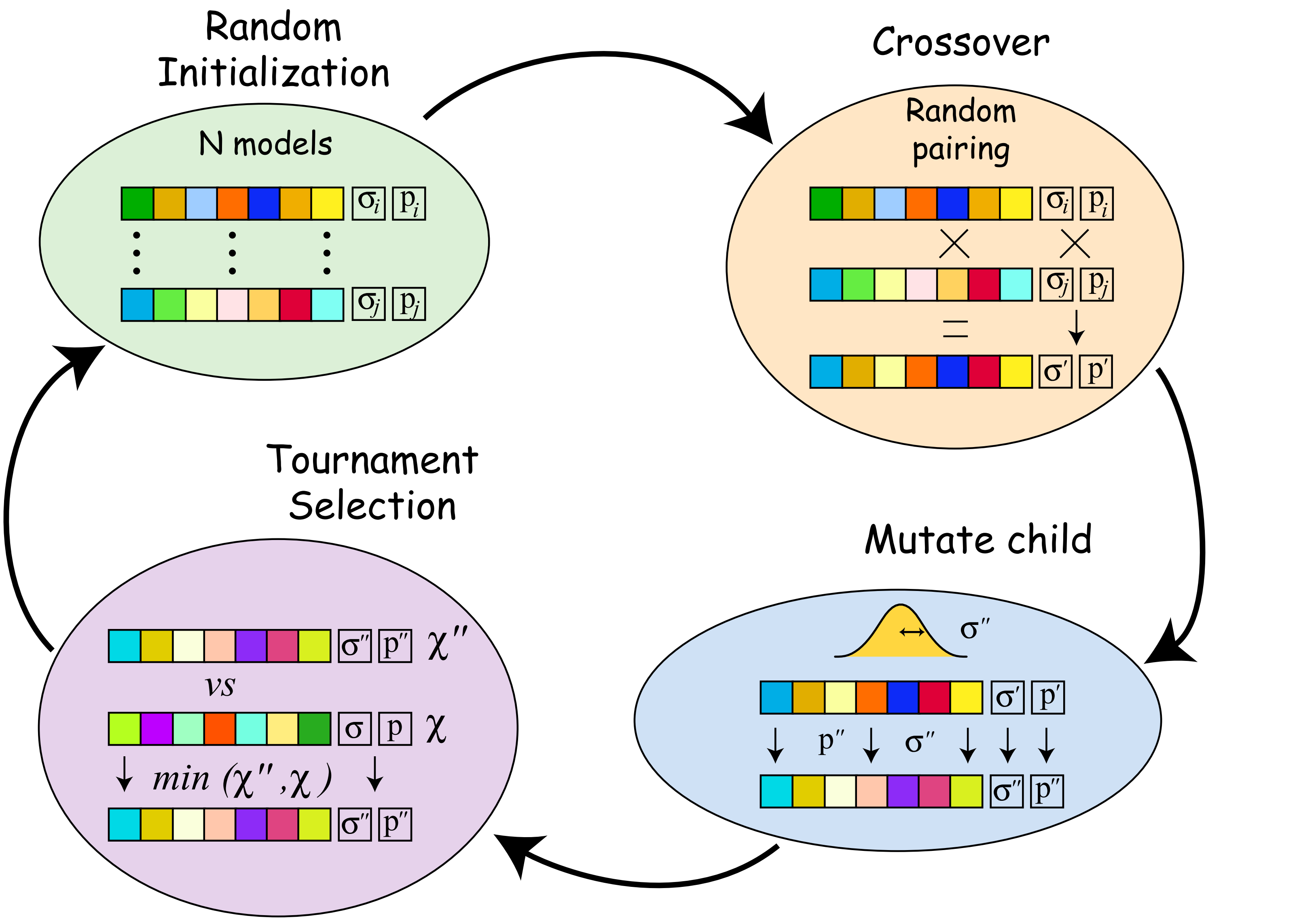
Sampling from a Gibbs-Boltzmann distribution,

$$\boxed{\sigma(\boldsymbol{m}) = e^{\frac{-\phi(\mathbf{m})}{T}}}$$

- *Temperature schedule, T decreases with time,*
- *Metropolis algorithm used to generate samples with an equilibrium distribution of $\sigma(\boldsymbol{m})$.*

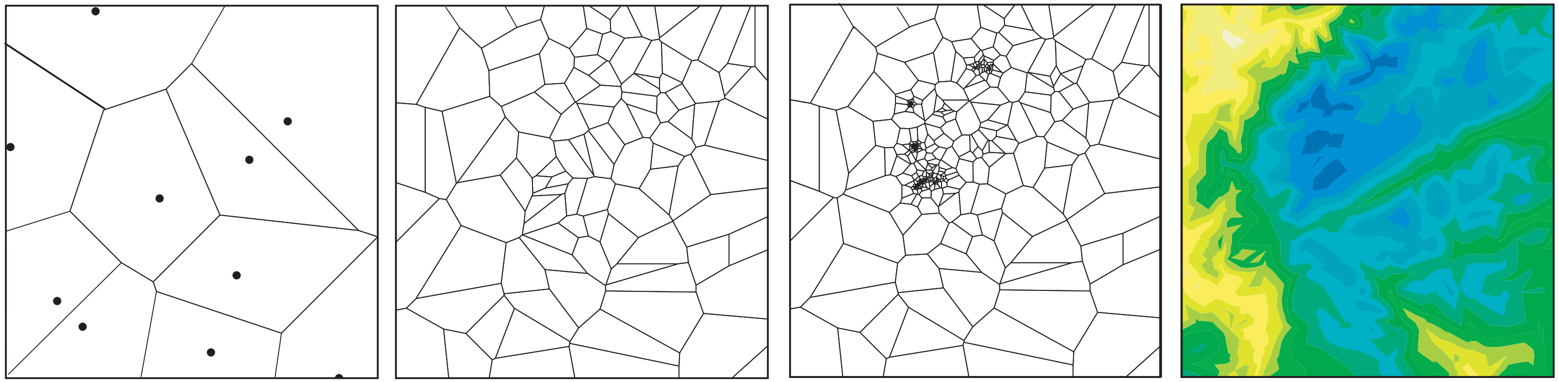


Evolutionary and genetic algorithms

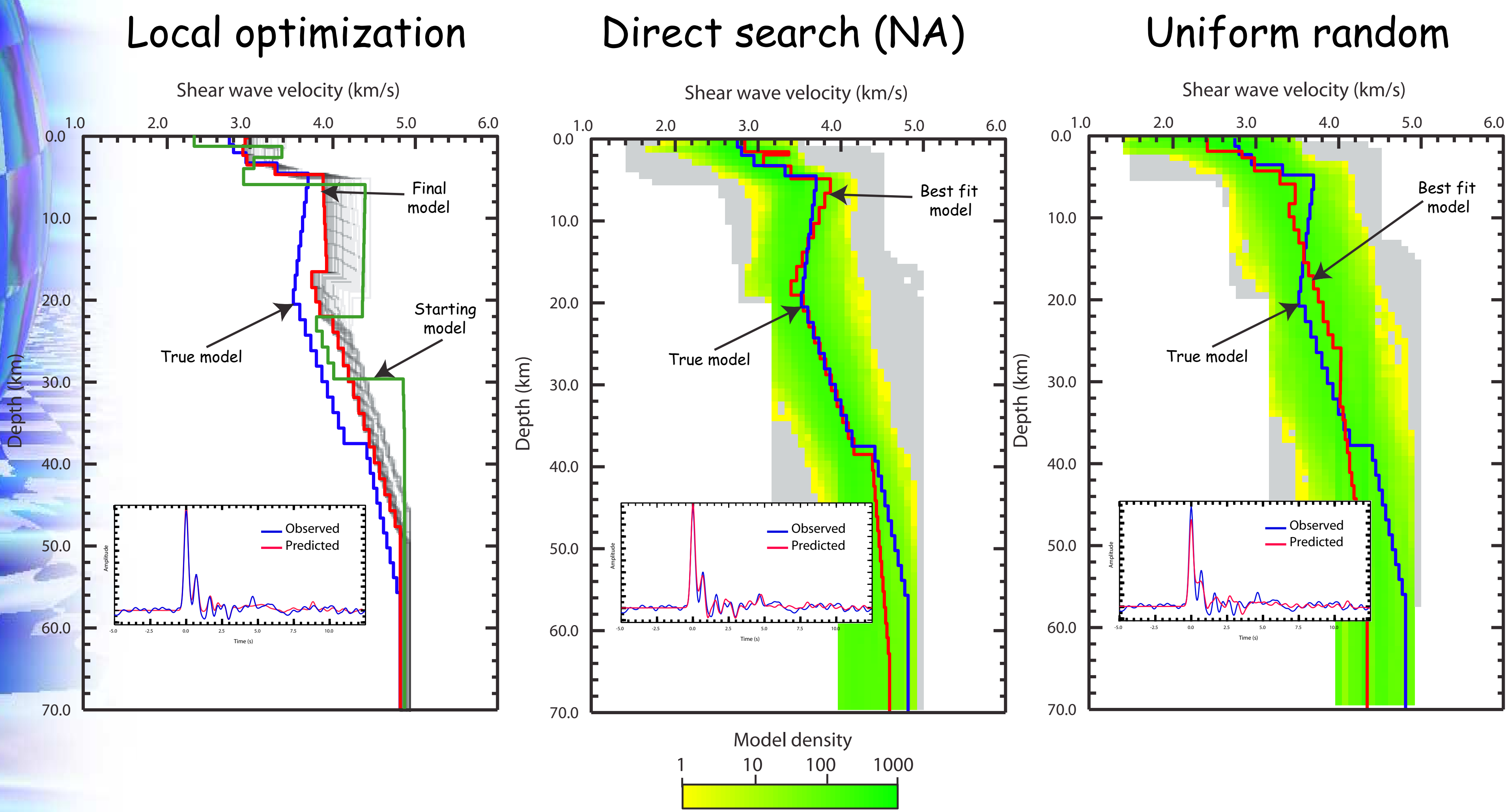


Adaptive neighbourhood sampling

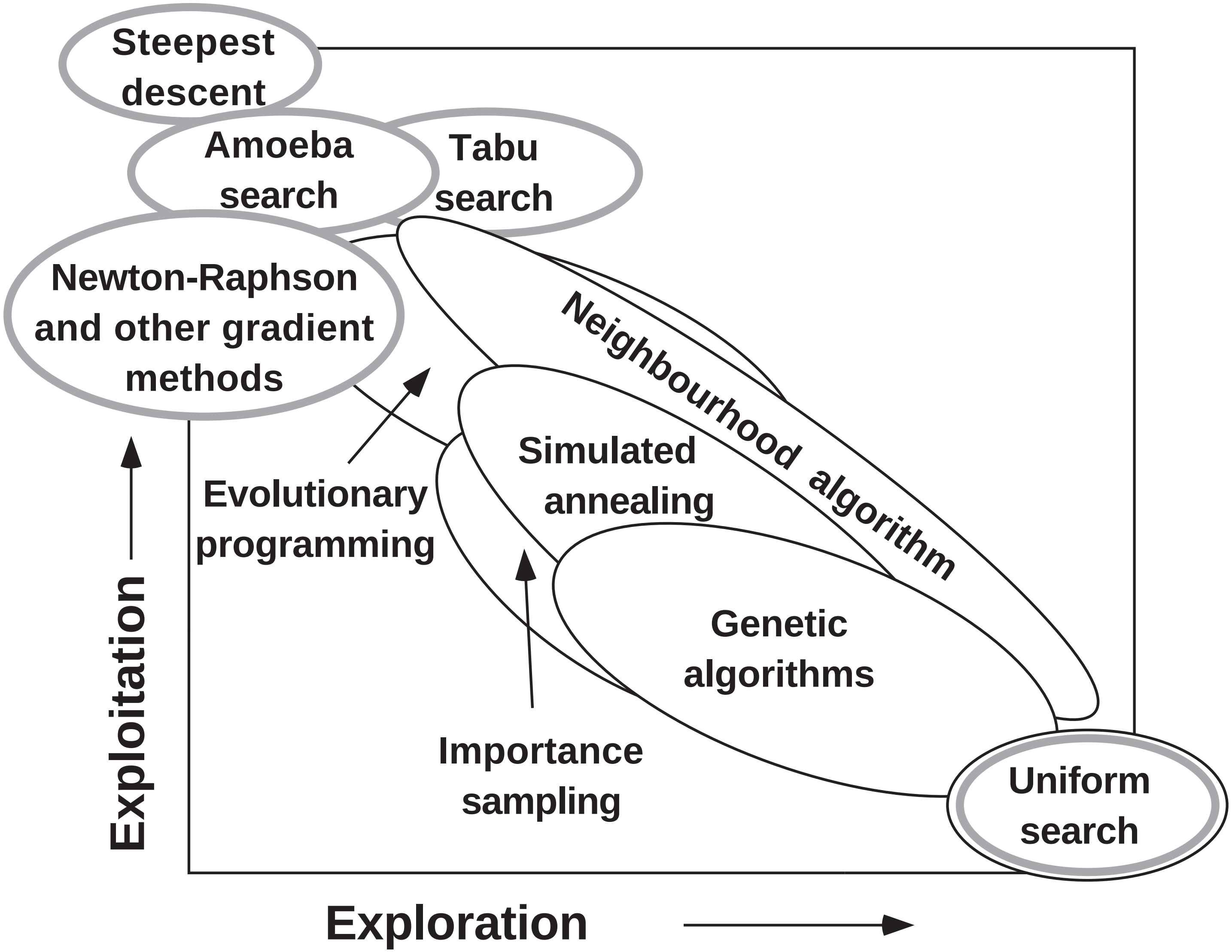
Partitioning the model space adaptively re-sampling using the Neighbourhood algorithm



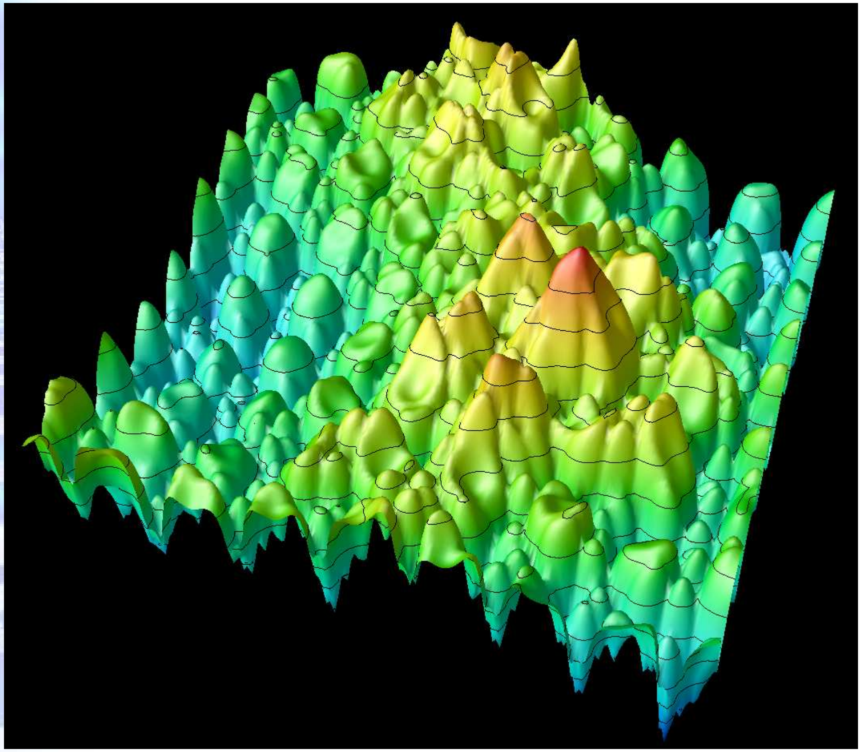
A comparison of approaches

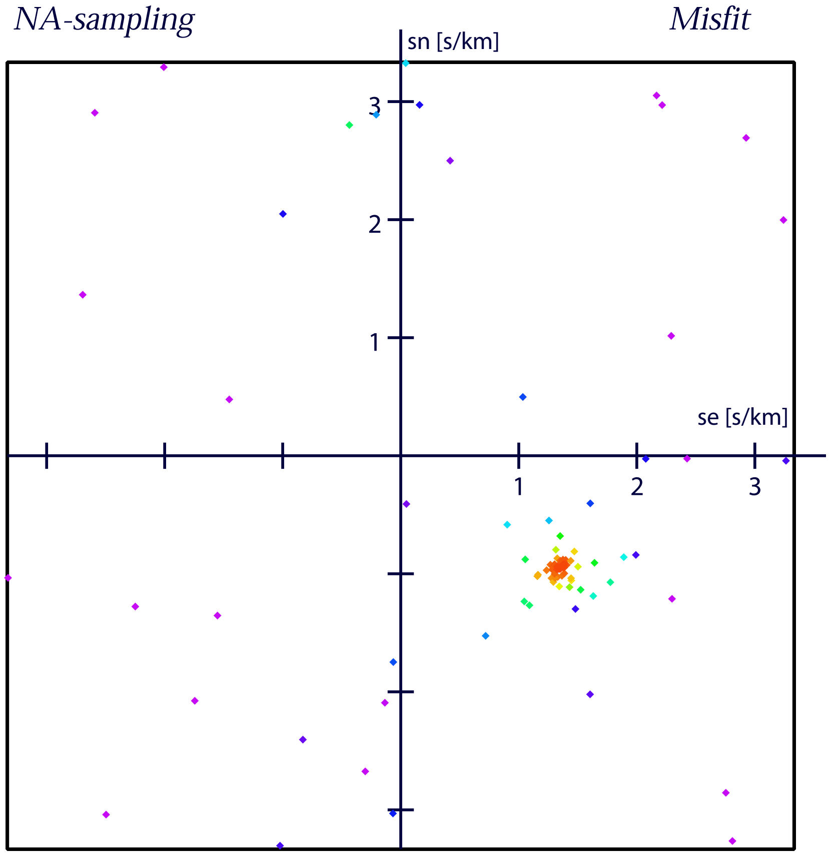


Exploitation vs Exploration



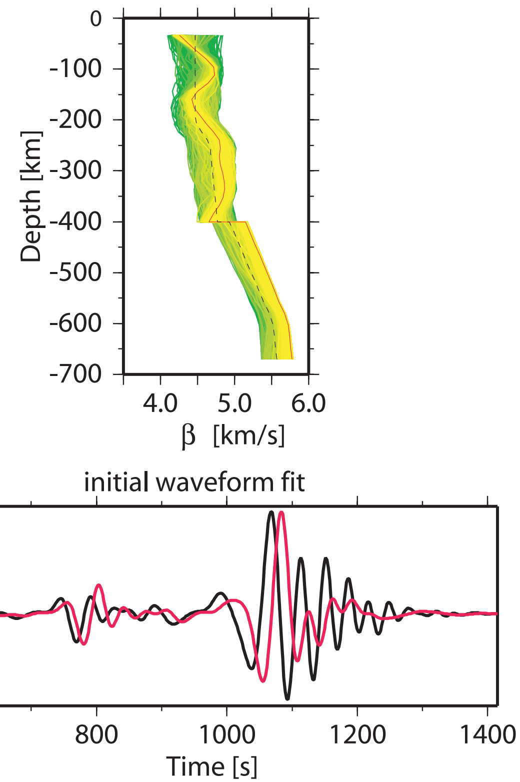
Examples: Beam power maximization

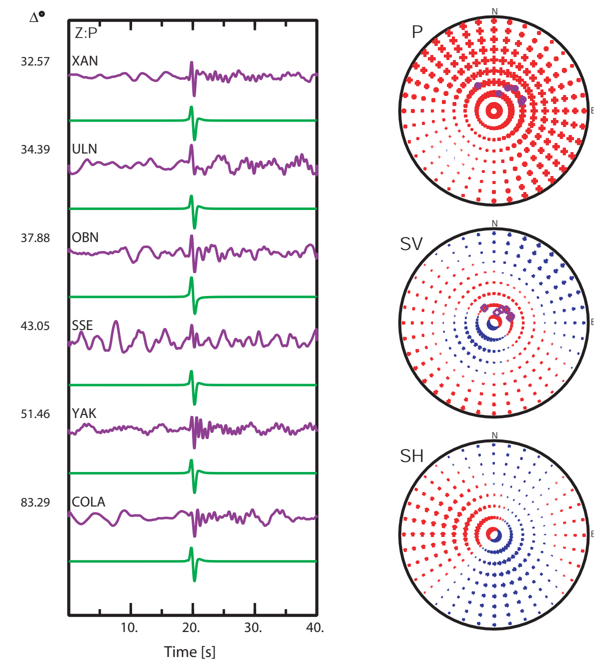




(From Kennett et. al. 2003)

Waveform fitting





Seismic waveforms

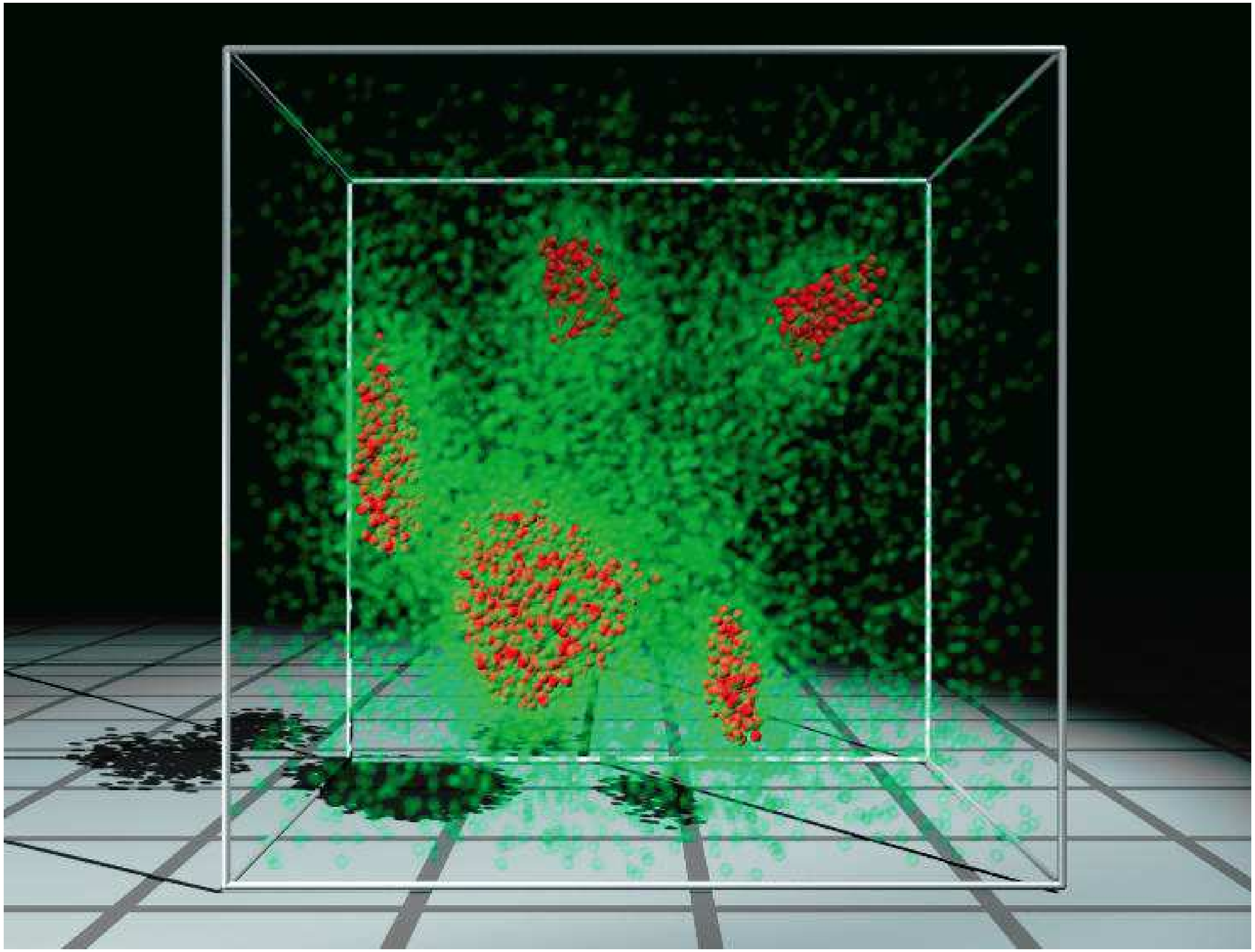
Receiver functions &
Surface waves

Seismic sources

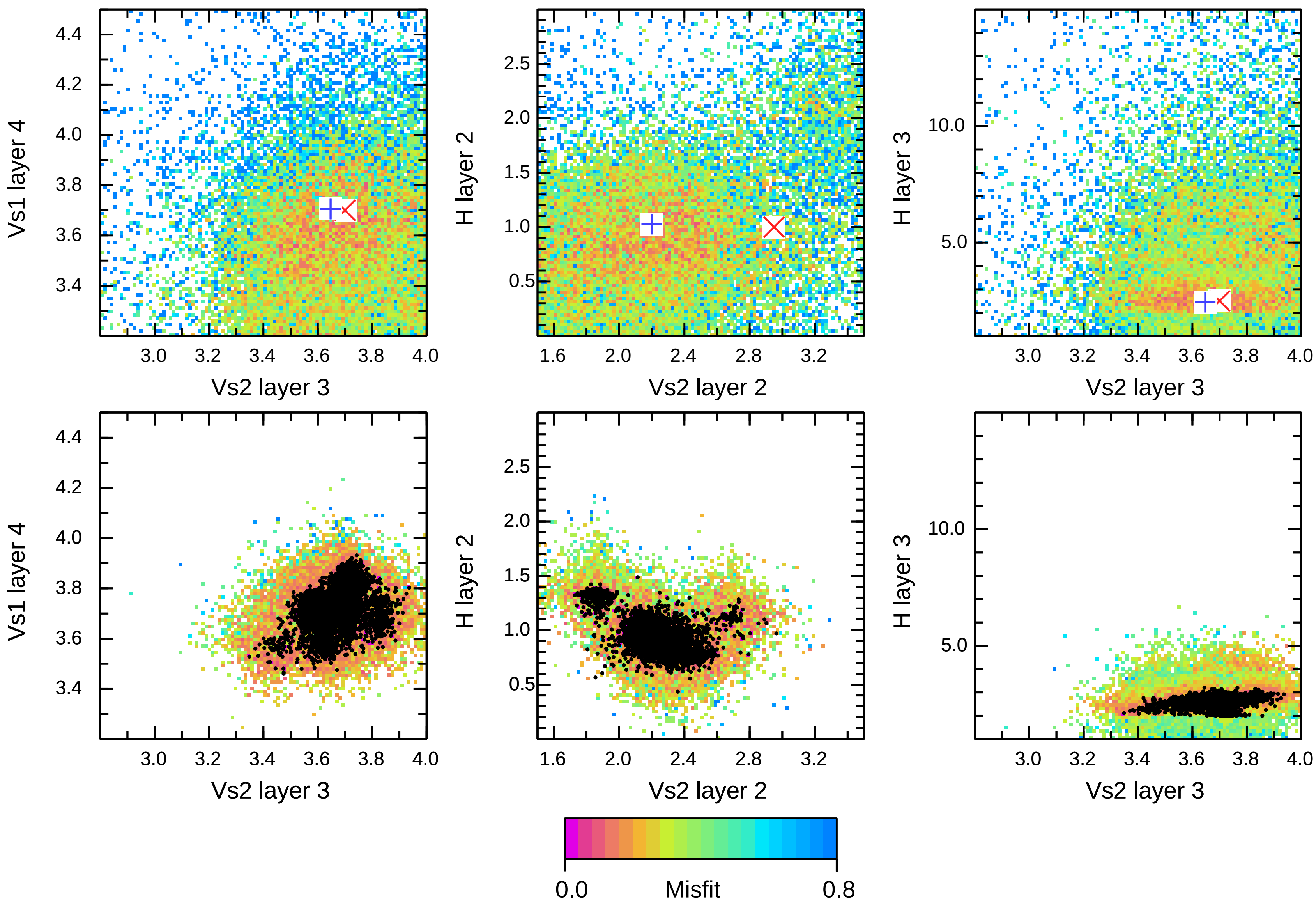
Coupled source moment
tensor & depth location

(From Yoshizawa & Kennett (2002); Marson-Pidgeon et al.(2000))

Mapping out multiple acceptable regions



Mapping out acceptable regions



Probabilistic approach to inverse problems

All information in the form of probability density functions.
Bayes rule

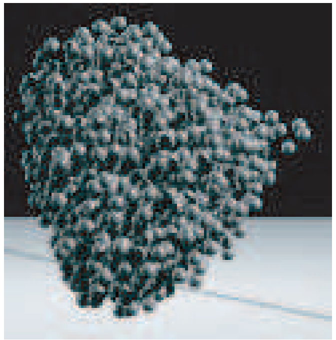
$$p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m}),$$

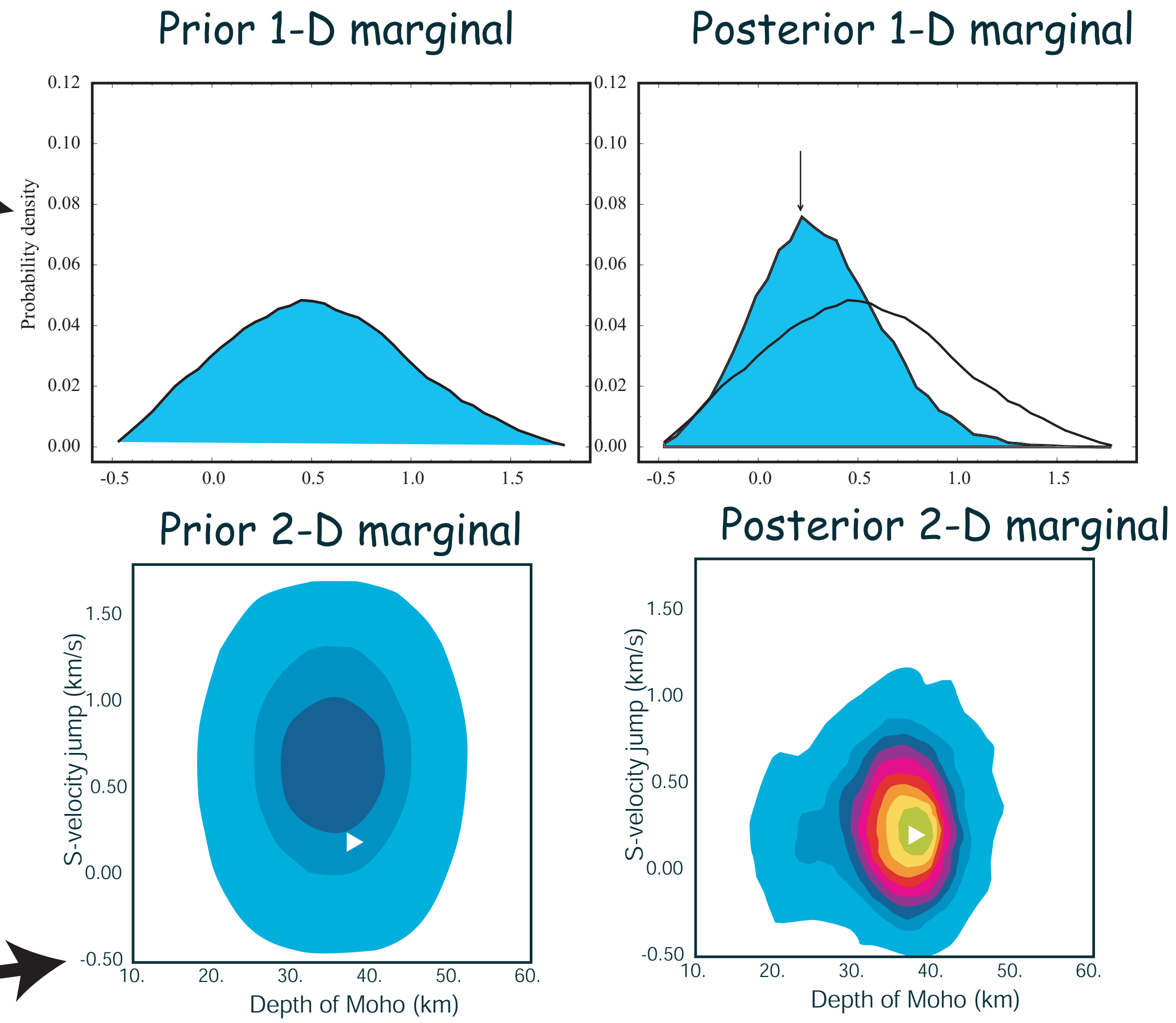
Posterior = Likelihood x prior

$$p(\boldsymbol{m}|\boldsymbol{d}) = \exp\left\{-\frac{1}{2}(\boldsymbol{d}-g(\boldsymbol{m}))^T C_D^{-1}(\boldsymbol{d}-g(\boldsymbol{m})) -\frac{1}{2}\,(\boldsymbol{m}-\boldsymbol{m}_o)^T C_M^{-1}(\boldsymbol{m}-\boldsymbol{m}_o)\right\}$$

Statistical sampling methods are needed to draw samples from the posterior.
Markov chain Monte Carlo (MCMC) is the workhorse technique.

Bayesian sampling

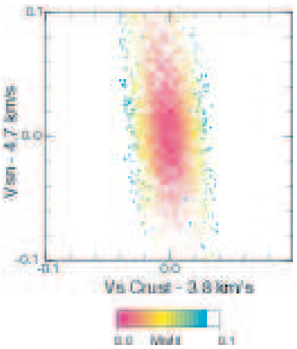




Bayesian inference

Bayesian inference can be applied to:

- The model inference problem
  Estimating the unknowns

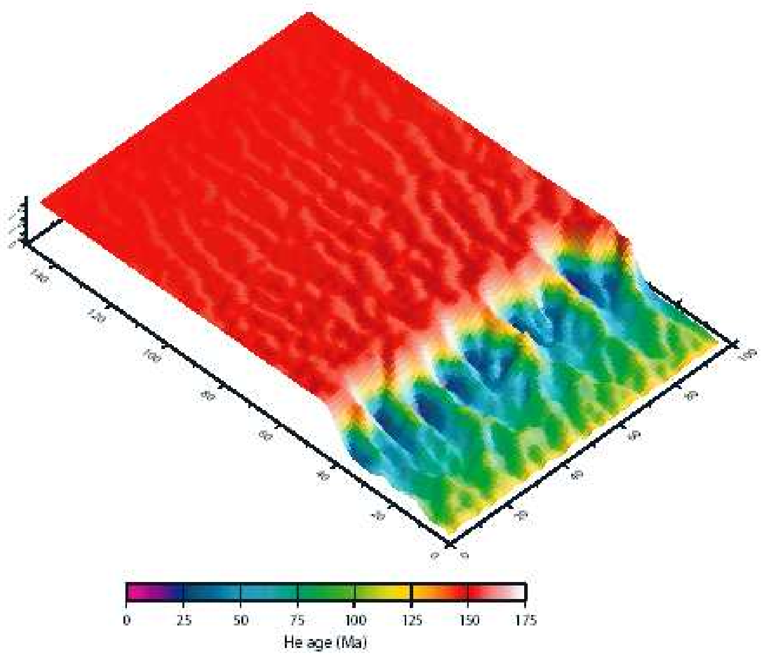


- The model comparision problem
  Hypothesis testing
  When the number of unknowns is one of your unknowns !

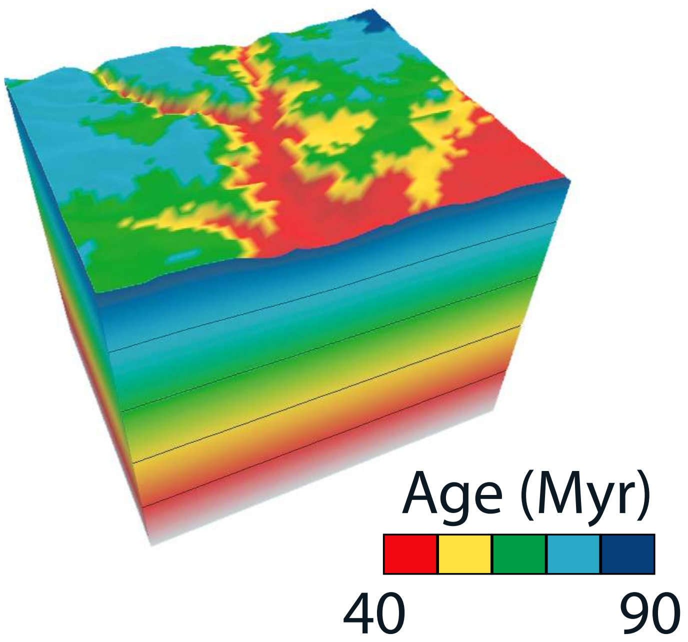
Example in additional material.

Intensive forward models

Current research trends are aimed at computationally intensive forward problems.



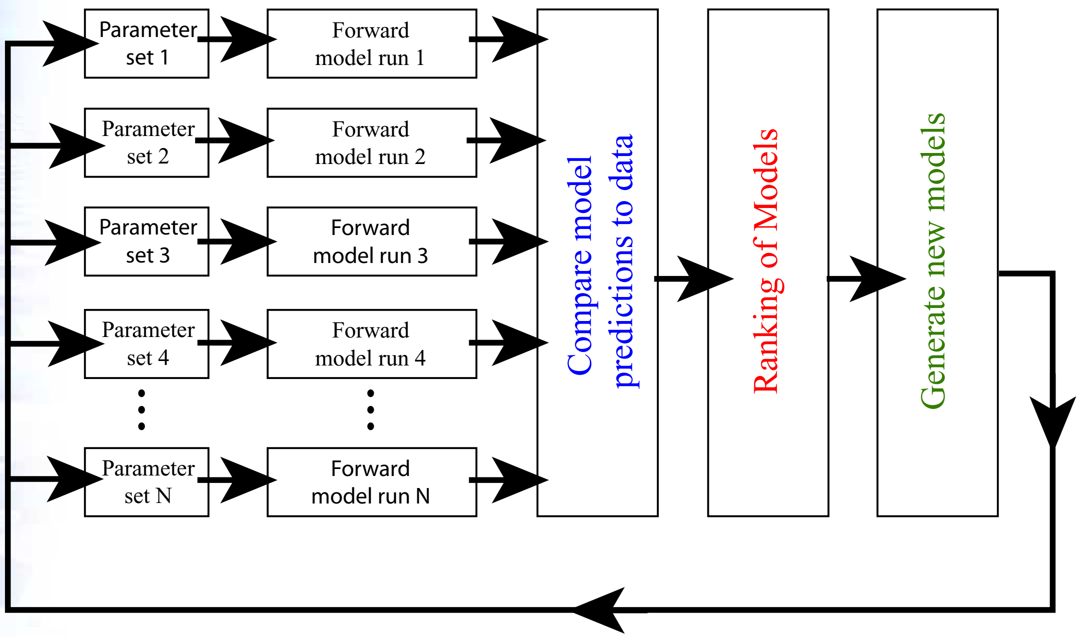
Using morphological data to constrain landscape processes



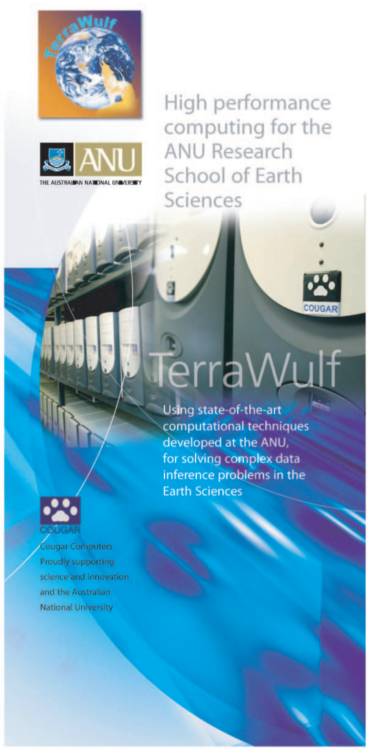
Using Thermo-chronological data to constrain deformation processes

Parallelism

An ensemble based approach is ideally suited to exploit parallel computing architectures



Inversion software

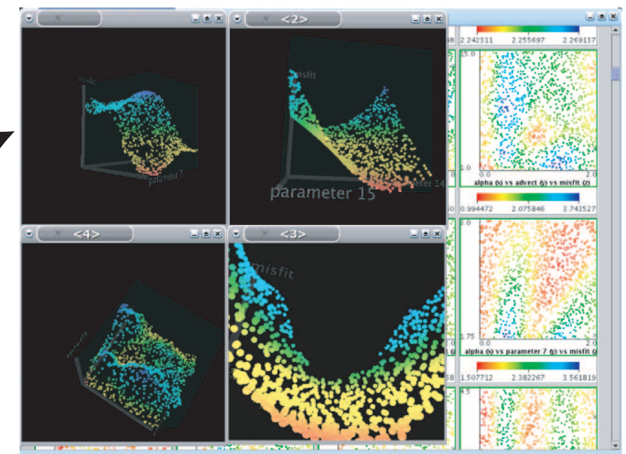


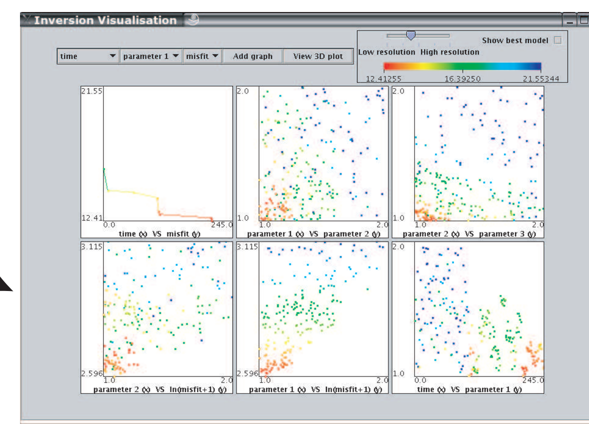
Java inversion toolkit

Insert forward code

Choose inversion algorithm

Sensitivity visualization





Real-time monitoring