ECE295, Data Assimilation and Inverse Problems, Spring 2015

Peter Gerstoft, 534-7768, gerstoft@ucsd.edu

We meet Wednesday from 5 to 6:20pm in HHS 2305A

Text for first 5 classes: Parameter Estimation and Inverse Problems (2nd Edition) here under UCSD license

Grading S

Classes

1 April, Intro; Linear discrete Inverse problems (Aster Ch 1, 2)

8 April, SVD (Aster ch 2 and 3)

15 April, Regularization (ch 4) Numerical Example: **Beamforming**

22 April, Sparse methods (ch 7.2-7.3)

29 April, Sparse methods

6 May, Bayesian methods and Monte Carlo methods (ch 11) Numerical Example: **Ice–flow from GPS**

13 May, Introduction to sequential Bayesian methods, Kalman Filter

20 May, Data assimilation, EnKF

27 May, EnKF, Data assimilation

3 June, Markov Chain Monte Carlo, PF

Homework: You can use any programming language, matlab is the most obvious, Just email the code to me (I dont need anything else).

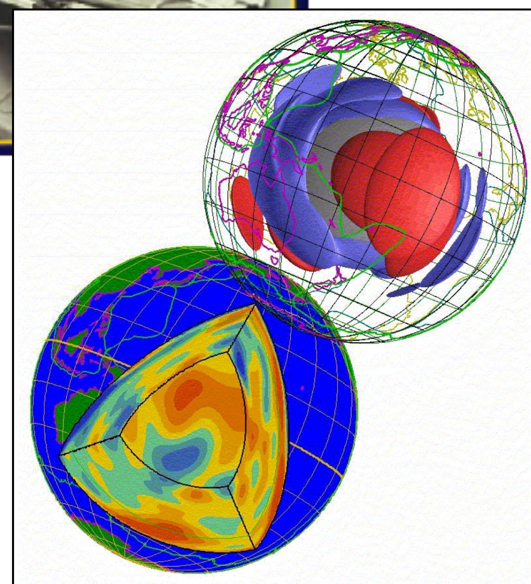
Call the files LastName_ExXX.

Homework is due 8am on Wednesday. That way we can discuss in class.

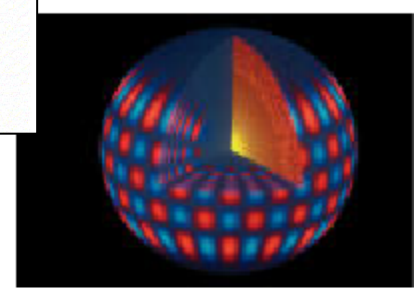
Hw 1: Download the matlab codes for the book (cd_5.3). Run the 3 examples for chapter 2. Come to class with one question about the examples. Due 8 April.

Inverse problems are everywhere

Medical tomography
1970s



Seismic tomography
1980s



Helioseismology
1990s

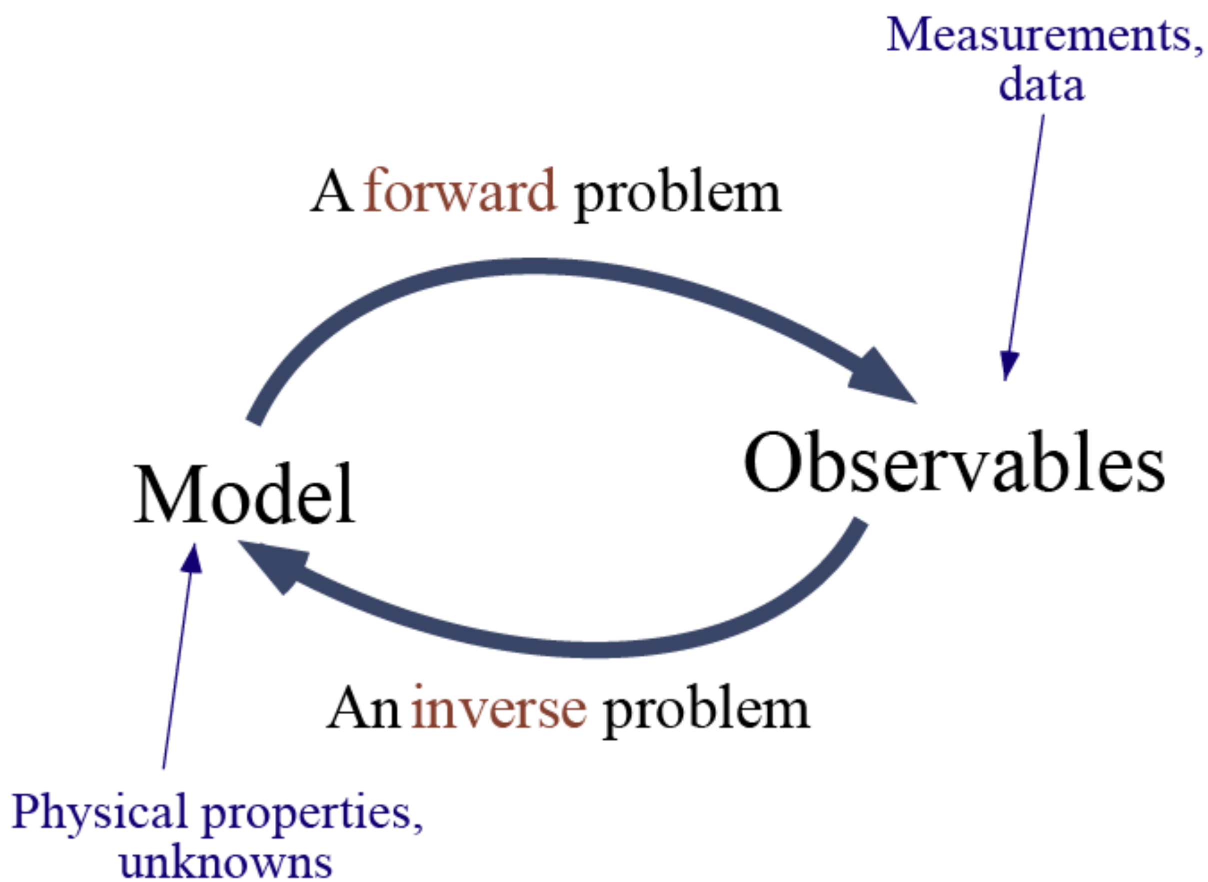
When data only indirectly constrain quantities of interest

Thinking backwards

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

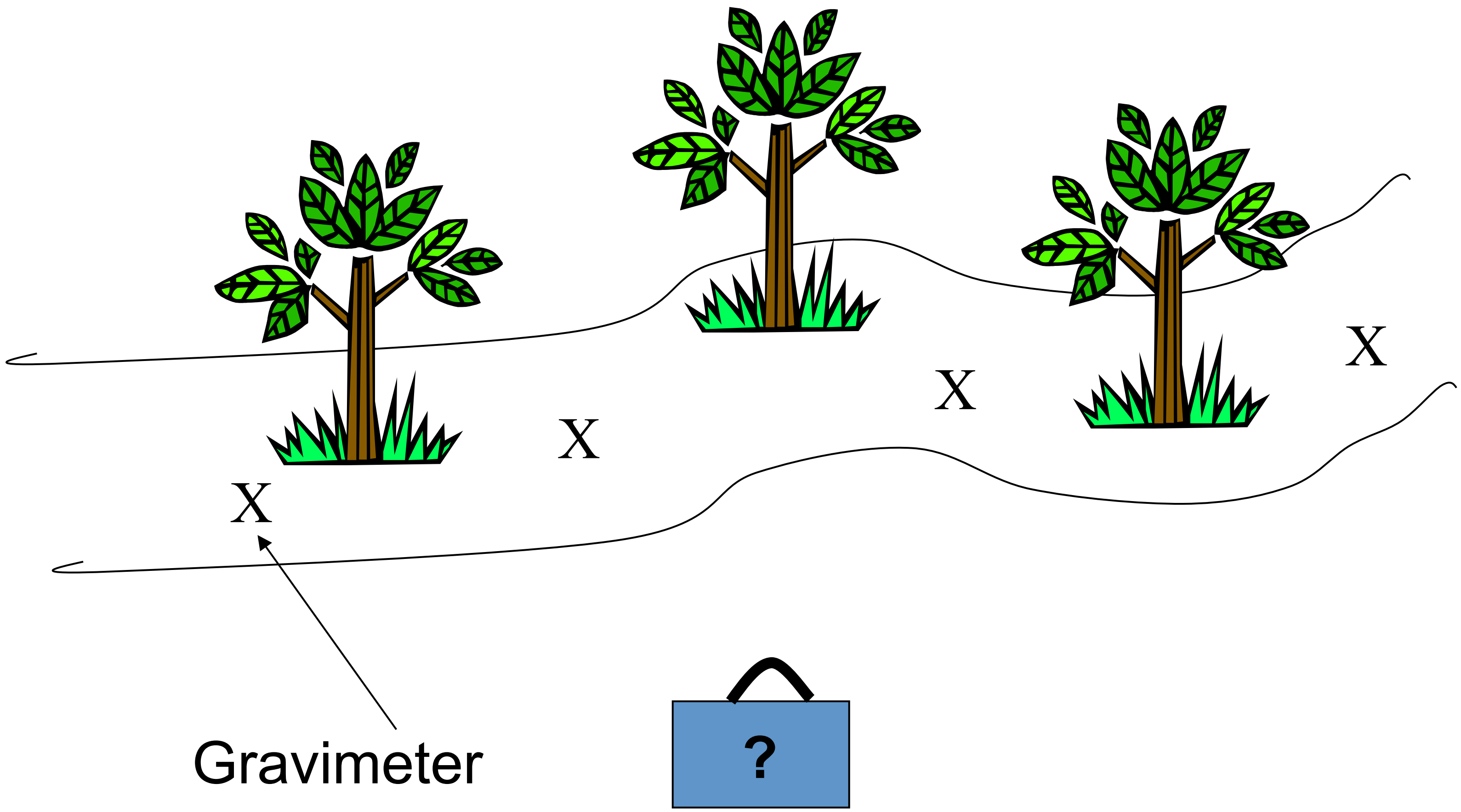
Sherlock Holmes,
A Study in Scarlet,
Sir Arthur Conan Doyle (1887)

Reversing a forward problem



Anatomy of an inverse problem

Hunting for gold at the beach with a gravimeter



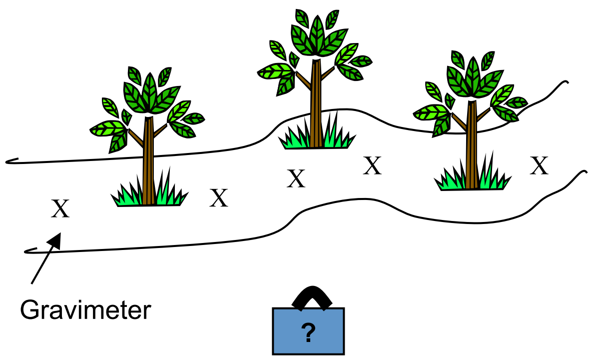
Courtesy Heiner Igel

Forward modelling example: Treasure Hunt

We have observed some values:

10, 23, 35, 45, 56 μ gals

How can we relate the observed gravity values to the subsurface properties?



We know how to do the *forward* problem:

$$\Phi(r) = \int \frac{G\rho(r')}{|r - r'|} dV'$$

This equation relates the (observed) gravitational potential to the subsurface density.

-> given a density model we can predict the gravity field at the surface!

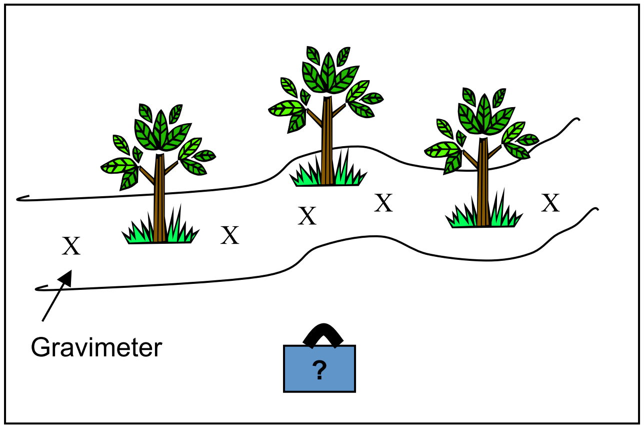
Treasure Hunt: Trial and error

What else do we know?

Density sand: 2.2 g/cm^3
Density gold: 19.3 g/cm^3

Do we know these values *exactly*?

Where is the box with gold?



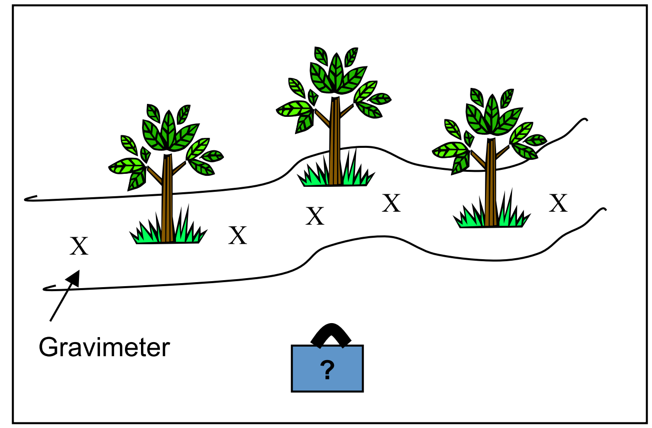
One approach is trial and (t)error forward modelling

Use the *forward* solution to calculate many models for a rectangular box situated somewhere in the ground and compare the *theoretical* (*synthetic*) data to the observations.

Treasure Hunt: model space

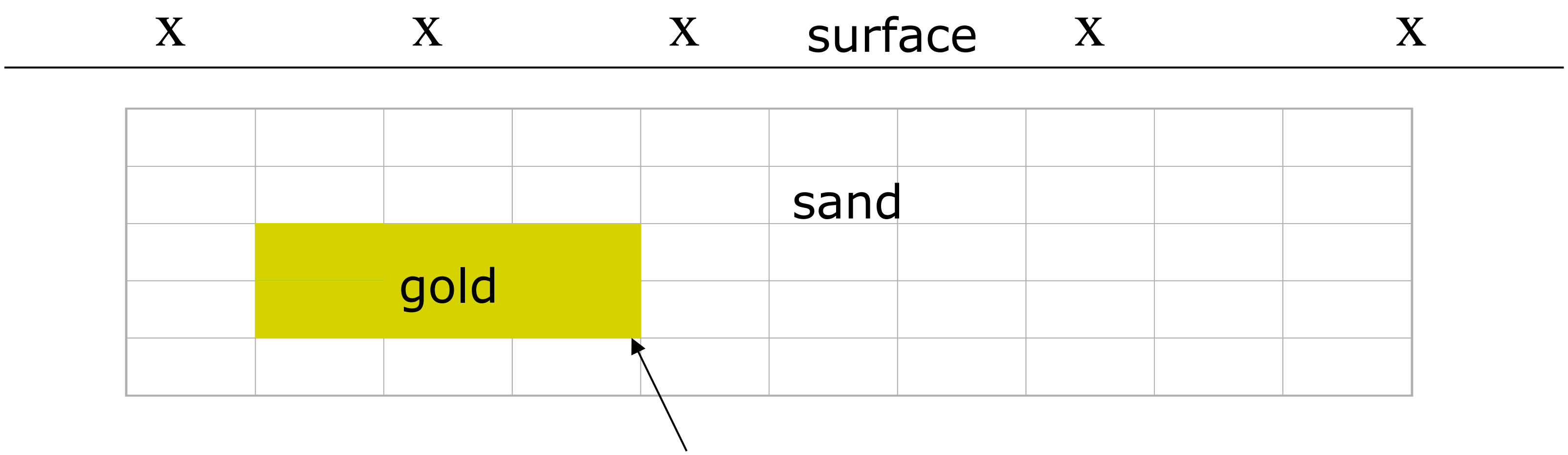
But ...

... we have to define *plausible* models for the beach. We have to somehow describe the model geometrically.



We introduce simplifying approximations

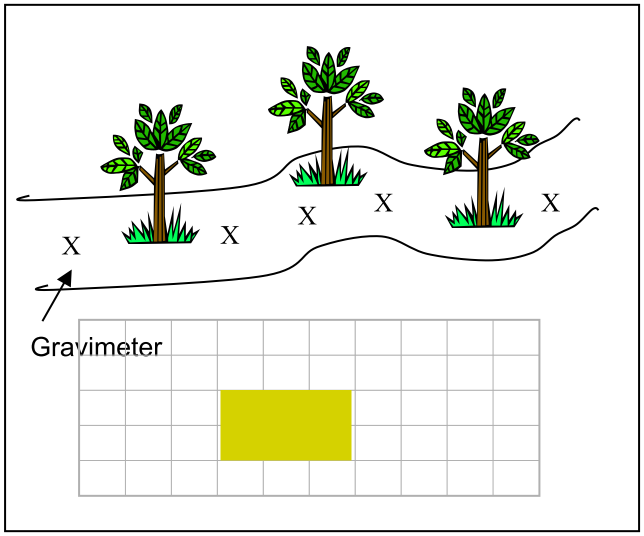
- divide the subsurface into rectangles with variable density
- Let us assume a flat surface



Treasure Hunt: Non-uniqueness

Could we compute all possible models and compare the synthetic data with the observations?

- at every rectangle two possibilities (sand or gold)
- $2^{50} \sim 10^{15}$ possible models



(Age of universe ~10^{17} s)

Too many models!

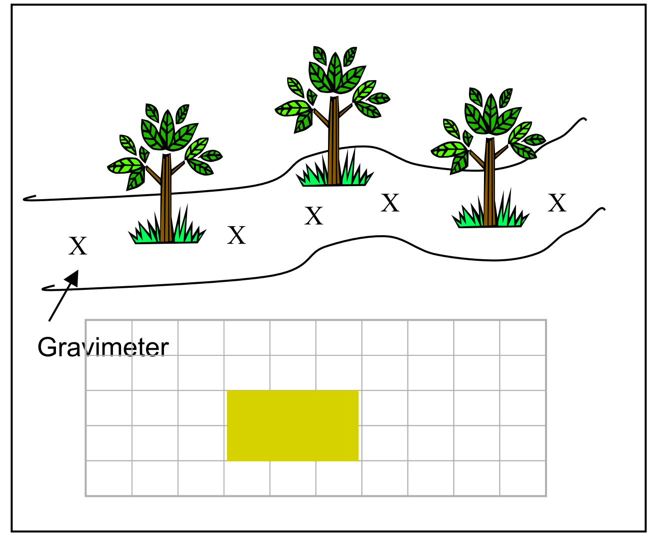
- We have 10^{15} possible models but only 5 observations!
- It is likely that two or more models will fit the data (maybe exactly)

Non-uniqueness is likely

Treasure hunt: a priori information

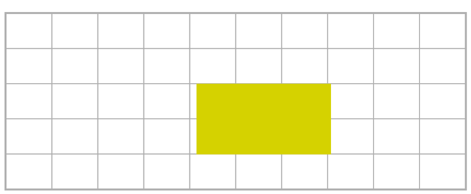
Is there anything we know about the treasure?

How large is the box?
Is it still intact?
Has it possibly disintegrated?
What was the shape of the box?

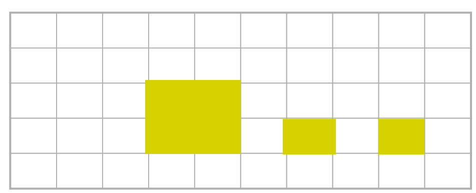


This is called *a priori* (or prior) information.
It will allow us to define plausible, possible, and unlikely models:

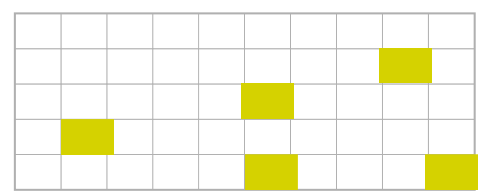
plausible



possible

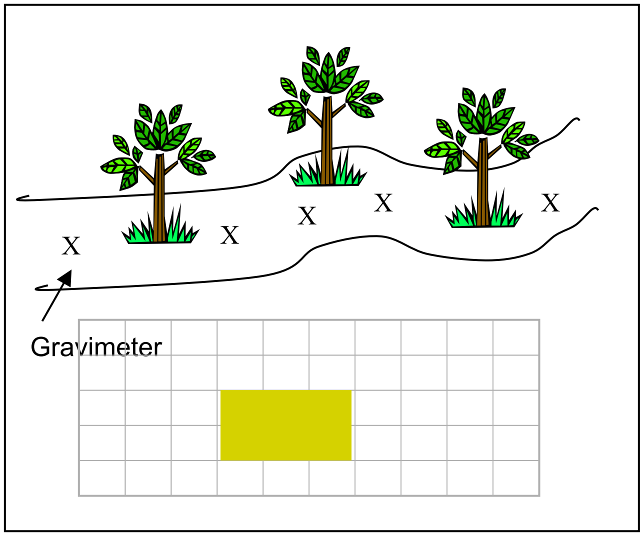


unlikely



Treasure hunt: data uncertainties

Things to consider in formulating the inverse problem



- Do we have errors in the data ?
  - Did the instruments work correctly ?
  - Do we have to correct for anything?

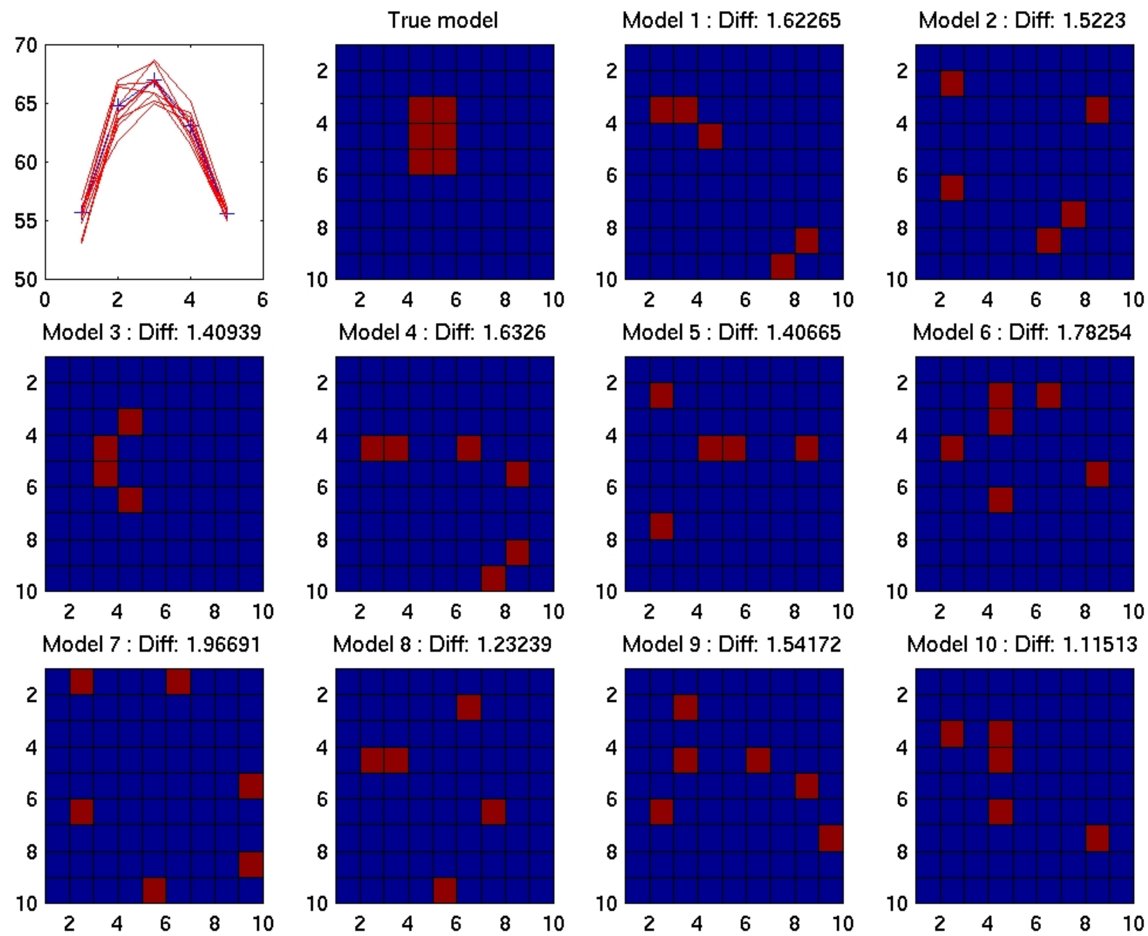
  (e.g. topography, tides, ...)

- Are we using the right theory ?
  - Is a 2-D approximation adequate ?
  - Are there other materials present other than gold and sand ?
  - Are there adjacent masses which could influence observations ?

Answering these questions often requires introducing more simplifying assumptions and guesses.
All inferences are dependent on these assumptions. (GIGO)

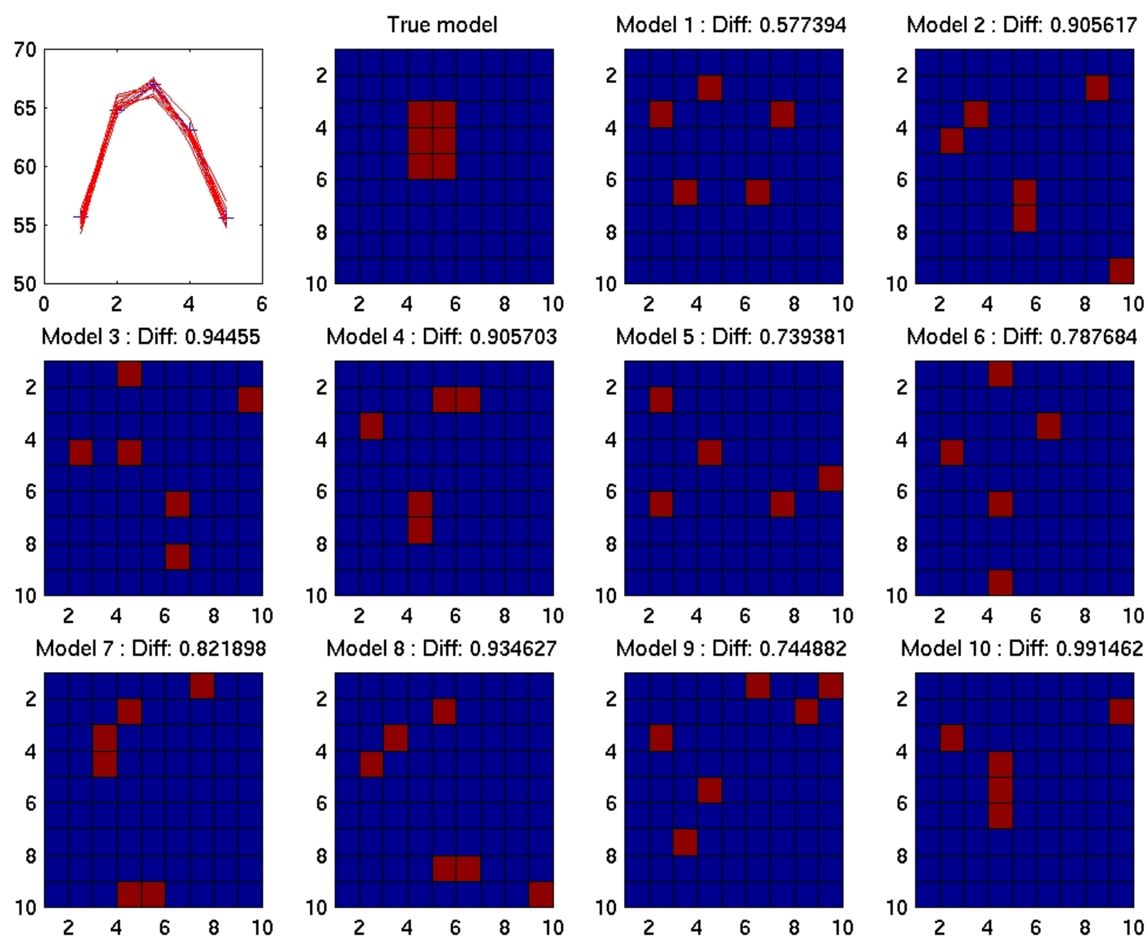
Treasure Hunt: solutions

Models with less than 2% error.



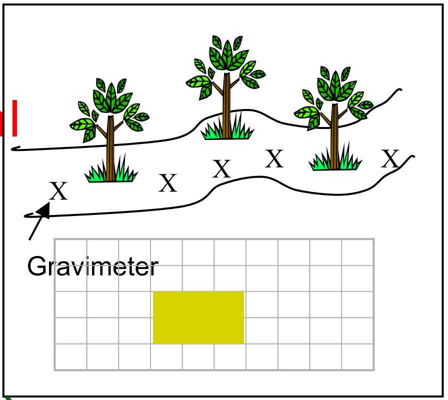
Treasure Hunt: solutions

Models with less than 1% error.



What we have learned from one example

Inverse problems = inference about physical systems from data



- Data usually contain errors (data uncertainties)
- Physical theories require approximations
- Infinitely many models will fit the data (non-uniqueness)
- Our physical theory may be inaccurate (theoretical uncertainties)
- Our forward problem may be highly nonlinear
- We always have a finite amount of data

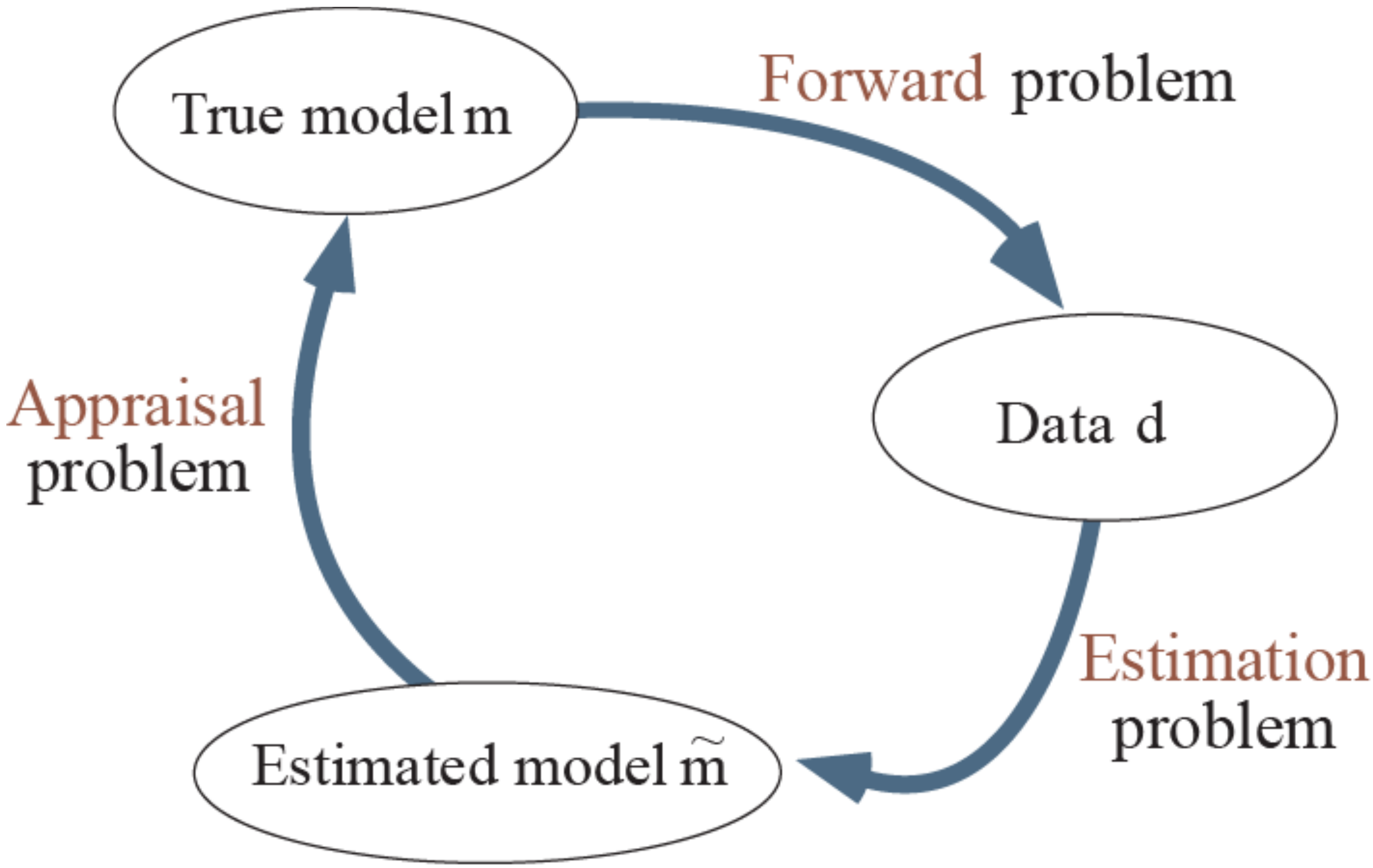
Detailed questions are:

How accurate are our data?

How well can we solve the forward problem?

What independent information do we have on the model space (a priori information) ?

Estimation and Appraisal

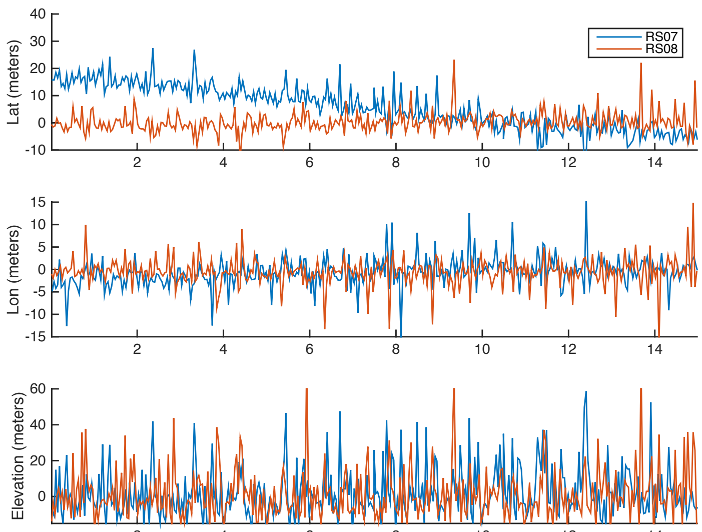


Sequential inverse problems

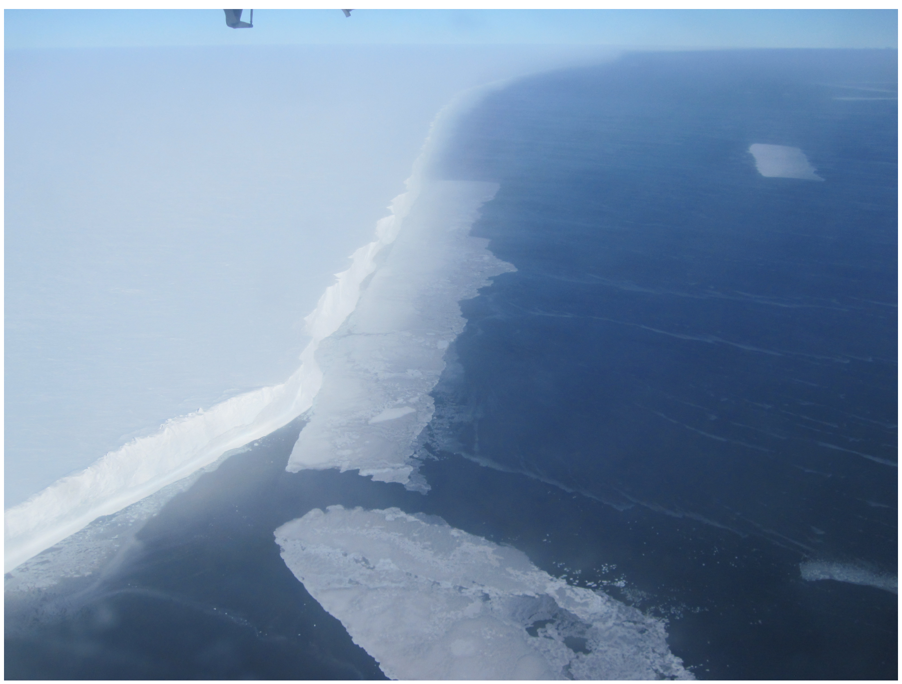
- Ocean circulation
  - Observe temperature, current flow, acoustic traveltimes…
  - Predictict ocean circulation
- Numerical Weather Prediction
- Evolution of an earthquake
- Evolution of a oilfield
- Evolution of a stock market

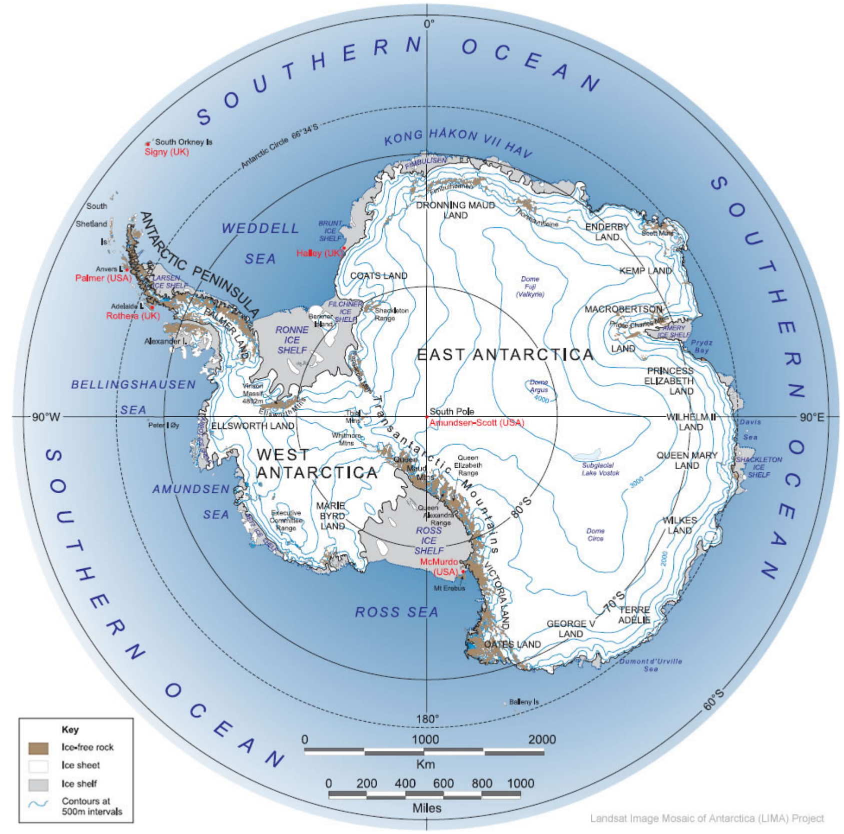
- Structural monitoring
- Tracking an airplane

ICE flow



RS07
RS08
Lat (meters)
Lon (meters)
Elevation (meters)





SOUTHERN OCEAN
SOUTHERN OCEAN
SOUTHERN OCEAN
SOUTHERN OCEAN
WEDDELL SEA
BELLINGSHAUSEN SEA
AMUNDSEN SEA
ROSS SEA
EAST ANTARCTICA
WEST ANTARCTICA
ANTARCTIC PENINSULA
RONNE ICE SHELF
ROSS ICE SHELF
KONG HÅKON VII HAV
DRONNING MAUD LAND
COATS LAND
ENDERBY LAND
KEMP LAND
MACROBERTSON LAND
PRINCESS ELIZABETH LAND
WILHELM II LAND
QUEEN MARY LAND
WILKES LAND
TERRE ADÉLIE
GEORGE V LAND
OATES LAND
VICTORIA LAND
MARIE BYRD LAND
ELLSWORTH LAND
PALMER LAND
Transantarctic Mountains
South Pole
Amundsen-Scott (USA)
McMurdo (USA)
Halley (UK)
Palmer (USA)
Rothera (UK)
Signy (UK)
South Orkney Is
South Shetland Is
Alexander I.
Peter I I.
Mt Erebus
Balleny Is
Dome Fuji (Valkyrie)
Dome Argus
Dome Circe
Subglacial Lake Vostok
Antarctic Circle 66°34'S
0°
90°W
90°E
180°
60°S
70°S
80°S
Key
Ice-free rock
Ice sheet
Ice shelf
Contours at 500m intervals
0
1000
2000
Km
0
200
400
600
800
1000
Miles
Landsat Image Mosaic of Antarctica (LIMA) Project

Data assimilation

- **Data assimilation** is the process by which observations are incorporated into a computer model of a real system. Applications of data assimilation arise in many fields of geosciences, perhaps most importantly in weather forecasting and hydrology.
- Data assimilation proceeds by *analysis cycles*. In each analysis cycle, observations of the current (and possibly past) state of a system are combined with the results from a numerical model (the *forecast*) to produce an *analysis*, which is considered as 'the best' estimate of the current state of the system. This is called the *analysis step*.
- Essentially, the analysis step tries to balance the uncertainty in the data and in the forecast. The model is then advanced in time and its result becomes the forecast in the next analysis cycle.

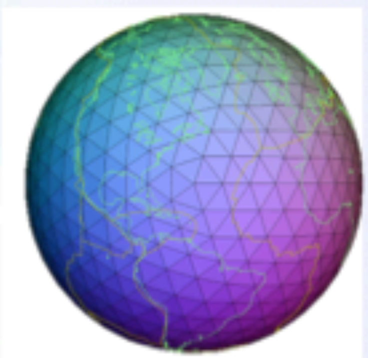
Discretizing a continuous model

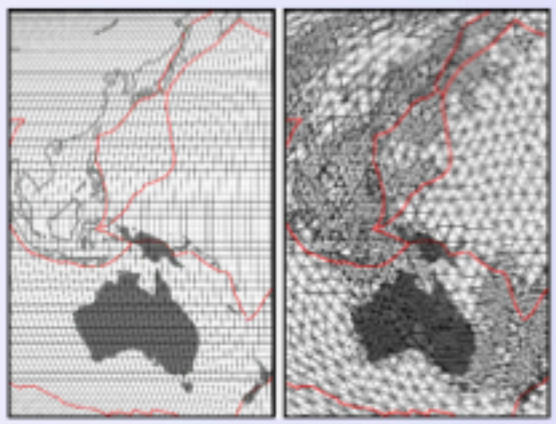
Often continuous functions are discretized to produce a finite set of unknowns. This requires use of *Basis functions*

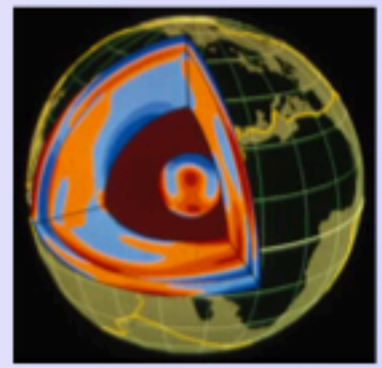
$$m(x) = \sum_{j=1}^{M} m_j \phi_j(x)$$

m_j become the unknowns $(j = 1, \dots, M)$

$\phi_j(x)$ are the *chosen* basis functions





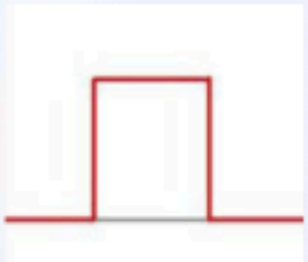


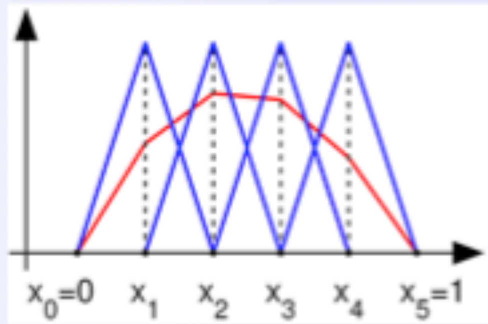
All inferences we can make about the continuous function will be influenced by the choice of basis functions. They must suit the physics of the forward problem. They bound the resolution of any model one gets out.

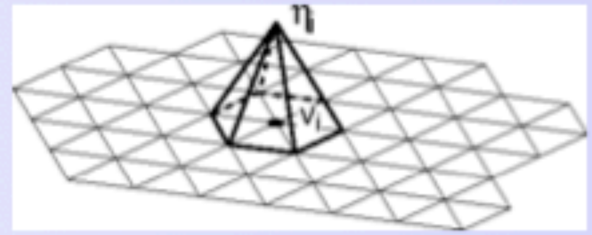
Discretizing a continuous model

Example of *Basis functions*

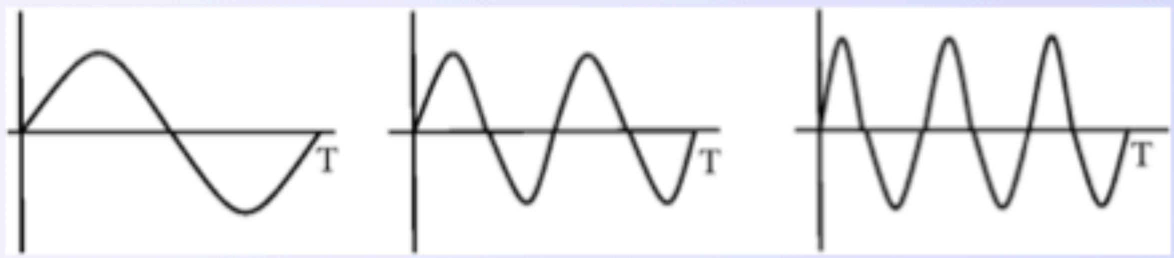
Local support

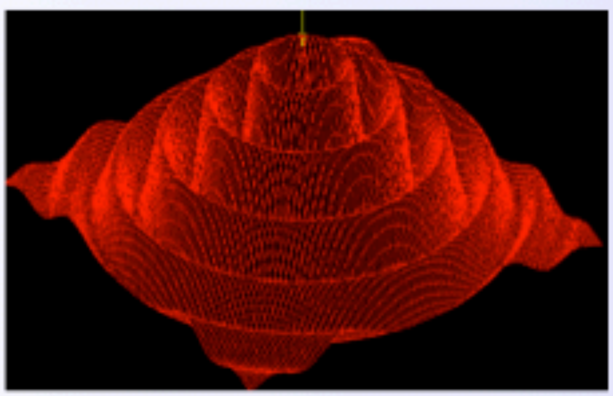


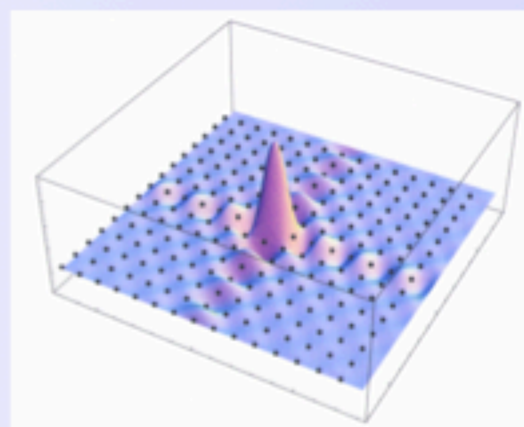




Global support

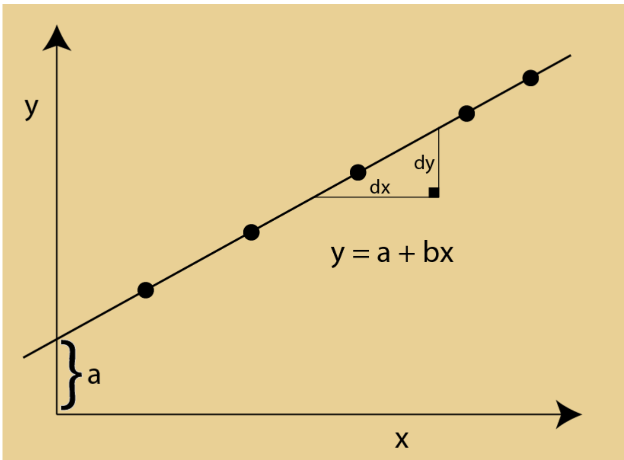






Formulating inverse problems

Regression



$$d = Gm$$

$$d = [y_1, y_2, \ldots, y_N]^T$$

$$m = [a, b]^T$$

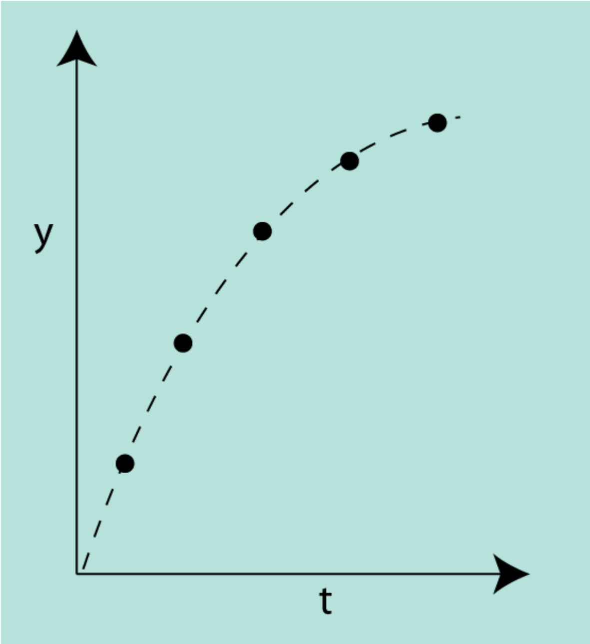
$$G = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$$

Discrete or continuous ?
Linear or nonlinear ? Why ?
What are the data ?
What are the model parameters ?
Unique or non-unique solution ?

y = a + bx

Formulating inverse problems

Ballistic trajectory



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

$$\boldsymbol{d} = [y_1, y_2, \ldots, y_N]^T$$

$$\boldsymbol{m} = [m_1, m_2, m_3]^T$$

$$G = \begin{pmatrix} 1 & t_1 & -1/2t_1^2 \\ 1 & t_2 & -1/2t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_M & -1/2t_M^2 \end{pmatrix}$$

Discrete or continuous ?
Linear or nonlinear ? Why ?
What are the data ?
What are the model parameters ?
Unique or non-unique solution ?

$$y = m_1 + m_2 t - \frac{1}{2} m_3 t^2$$

Recap: Characterizing inverse problems

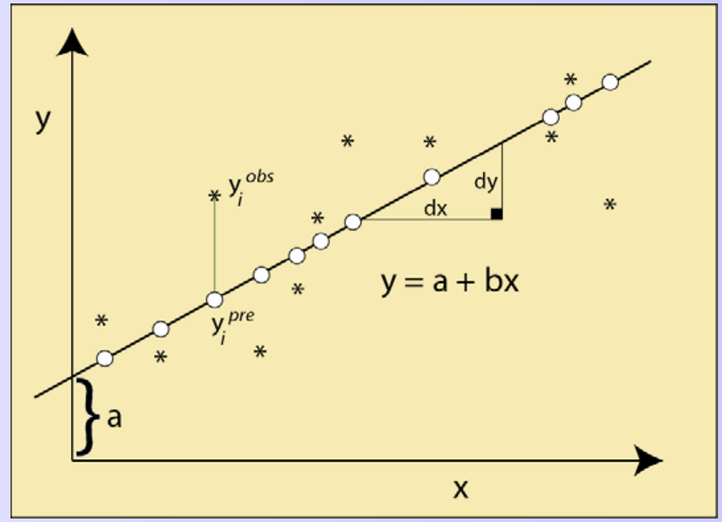
- Inverse problems can be continuous or discrete
- Continuous problems are often discretized by choosing a set of basis functions and projecting the continuous function on them.
- The forward problem is to take a model and predict observables that are compared to actual data. Contains the Physics of the problem. This often involves a mathematical model which is an approximation to the real physics.
- The inverse problem is to take the data and constrain the model in some way.
- We may want to build a model or we may wish to ask a less precise question of the data !

Over-determined: Linear discrete inverse problem

To find the best fit model we can minimize the *prediction error* of the solution

$$r_i(m) = y_i^{obs} - y_i^{pre}(m)$$

$$r = d - Gm$$



But the data contain errors. Let’s assume these are *independent* and *normally* distributed, then we weight each residual inversely by the standard deviation of the corresponding (known) error distribution.

We can obtain a least squares solution by minimizing the *weighted prediction error* of the solution.

$$\phi(m) = \sum_{i=1}^{N} \left(\frac{y_i^{obs} - y_i^{pre}(m)}{\sigma_i} \right)^2 = r^T C_d^{-1} r$$

$$C_d = diag\left[\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2\right]$$

Over-determined: Linear discrete inverse problem

We seek the model vector **m** which minimizes

Compare with maximum likelihood

$$\phi(\boldsymbol{m}) = \frac{1}{2}\boldsymbol{r}^T C_d^{-1}\boldsymbol{r} = \frac{1}{2}(\boldsymbol{d} - G\boldsymbol{m})^T C_d^{-1}(\boldsymbol{d} - G\boldsymbol{m})$$

Note that this is a quadratic function of the model vector.

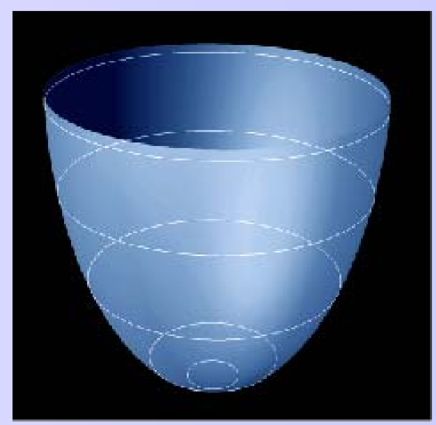
Solution: Differentiate with respect to **m** and solve for the model vector which gives a zero gradient in $\phi(\boldsymbol{m})$

This gives…

$$\nabla\phi(\boldsymbol{m}) = -G^T C_d^{-1}(\boldsymbol{d} - G\boldsymbol{m}) = 0$$

$$\Rightarrow \boldsymbol{m} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1}\boldsymbol{d}$$

This is the least-squares solution.



A solution to the normal equations:

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

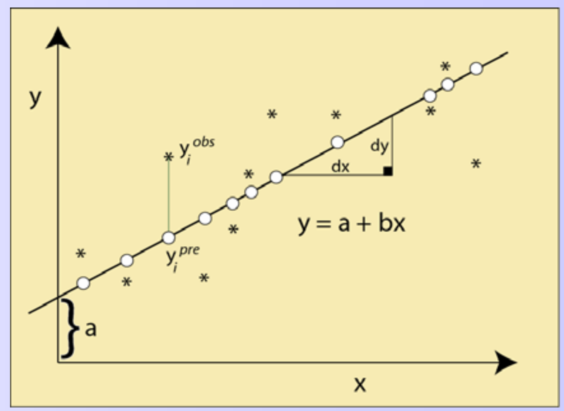
Over-determined: Linear discrete inverse problem

How does the Least-squares solution compare to the standard equations of linear regression ?

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

Given N data y_i with independent normally distributed errors and standard deviations σ_i what are the expressions for the model parameters **m** = $[a,b]^T$?

$$G\boldsymbol{m} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_{N_d} \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_N \end{bmatrix} = \boldsymbol{d}$$



$$\boldsymbol{m} = \begin{bmatrix} N & \sum_{i=1}^N x_i \\ \sum_{i=1}^N x_i & \sum_{i=1}^N x_i^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

$$\boldsymbol{m} = \frac{1}{N\sum_{i=1}^N x_i^2 - \left(\sum_{i=1}^N x_i\right)^2} \begin{bmatrix} \sum_{i=1}^N x_i^2 & -\sum_{i=1}^N x_i \\ -\sum_{i=1}^N x_i & N \end{bmatrix} \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_i y_i \end{bmatrix}$$

Linear discrete inverse problem: Least squares

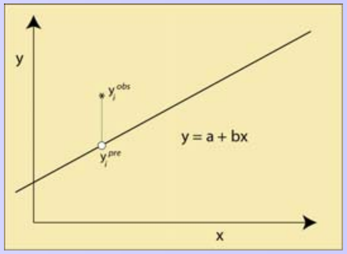
$$\mathbf{m}_{LS} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} \mathbf{d} = G^{-g} \mathbf{d}$$

What happens in the under and even-determined cases ?

$$\mathbf{m} = \frac{1}{N\sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \begin{bmatrix} \sum_{i=1}^{N} x_i^2 & -\sum_{i=1}^{N} x_i \\ -\sum_{i=1}^{N} x_i & N \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

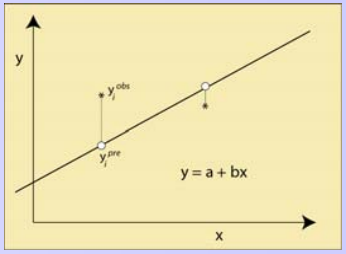
- Under-determined, N=1:

Matrix has a zero determinant
and a zero eigenvalue
an infinite number of solutions exist



- Even-determined, N=2:

$$\mathbf{m} = [m_1, m_2]^T, \quad m_2 = \left[\frac{y_1 - y_2}{x_1 - x_2}\right], \quad m_1 = y_1 - m_2 x_1.$$



$$r = d - Gm = 0$$

Prediction error is zero !

Example: Over-determined, Linear discrete inverse problem

The Ballistics example

$$y_i = m_1 + m_2 t_i - \frac{1}{2} m_3 t_i^2$$

Given data and noise

t	y
1	109:3827
2	187:5385
3	267:5319
4	331:8753
5	386:0535
6	428:4271
7	452:1644
8	498:1461
9	512:3499
10	512:9753

$$C_d^{-1} = \frac{1}{\sigma^2} I$$

$$\sigma = 8m$$

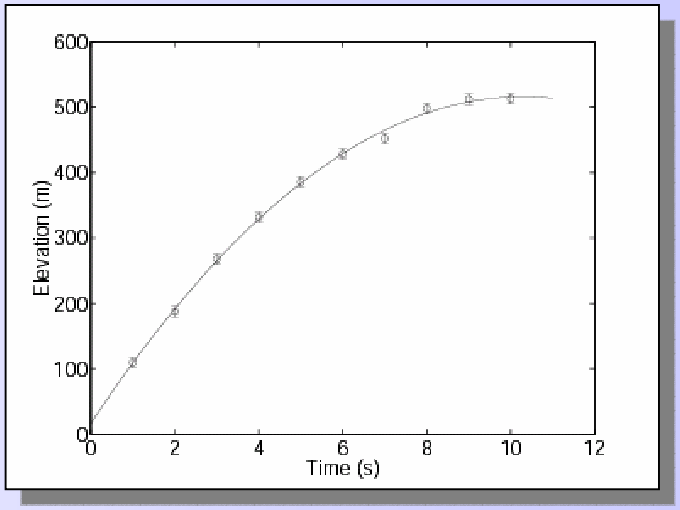
Calculate G

$$G = \begin{pmatrix} 1 & t_1 & -1/2t_1^2 \\ 1 & t_2 & -1/2t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_M & -1/2t_M^2 \end{pmatrix}$$

$$m_{LS} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} \mathbf{d}$$

$$m_{LS} = [16.4m, 97.0m/s, 9.4m/s^2]^T$$

$$m_{true} = [10m, 100m/s, 9.8m/s^2]^T$$



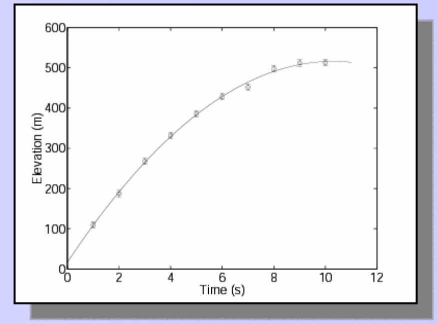
Is the data fit good enough ?

And how to errors in data propagate into the solution ?

The two questions in parameter estimation

We have our fitted model parameters

…but we are far from finished !



We need to:

- Assess the quality of the data fit.

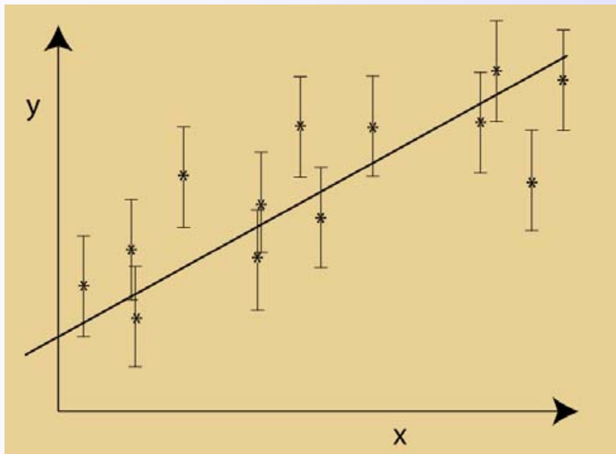
  Goodness of fit: Does the model fit the data to within the statistical uncertainty of the noise ?

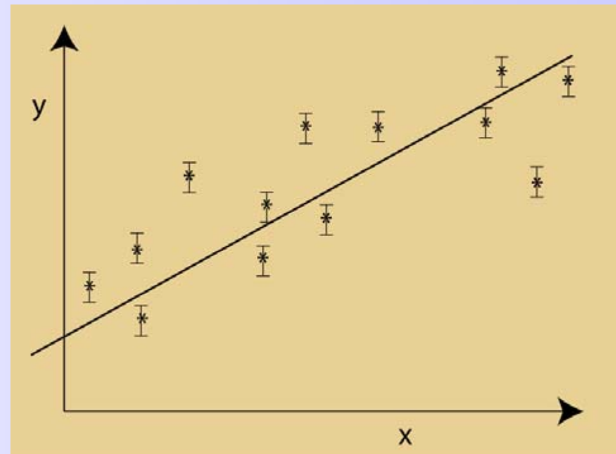
- Estimate how errors in the data propagate into the model

  What are the errors on the model parameters ?

Goodness of fit

Once we have our least squares solution $\mathbf{m}_{LS}$ how do we know whether the fit is good enough given the errors in the data ?





Use the prediction error at the least squares solution !

$$\phi(m_{LS}) = \frac{1}{2}(d - Gm_{LS})^T C_d^{-1}(d - Gm_{LS}) = \sum_{i=1}^{N}\left(\frac{d_i - \sum_{j=1}^{M} G_{i,j} m_j}{\sigma_i}\right)^2$$

If data errors are Gaussian this as a chi-square statistic χ^2_{obs}