Optimal Experimental Design

Andrew Curtis

THE UNIVERSITY of EDINBURGH

edinburgh imaging project

Optimal Experimental Design (Survey)

Andrew Curtis

Survey and Experimental Design

- Can not avoid: potentially one of the most common tasks
- Applied in many other fields of work and research
- Many geophysical surveys are designed in large part using tried and tested rules of thumb – *heuristics*
- Heuristics are generally robust, but not optimal: far more sophisticated theory exists and is used in other fields
  - *Statistical Experimental Design*
- We will examine this theory, and both common and state of the art applications

The Leading Edge, 2004
https://blogs.ed.ac.uk/curtis/

Theory of model-based geophysical survey and experimental design

Part A—linear problems

ANDREW CURTIS, Schlumberger Cambridge Research, Cambridge, U.K.

Enormous sums of money are invested by industry and scientific funding agencies every year in seismic, well logging, electromagnetic, earthquake monitoring and microseismic surveys, and in laboratory-based experiments. For each survey or experiment a design process must first take place. An efficient design is usually a compromise—a suitable trade-off between information that is expected to be retrieved about a model of interest and the cost of data acquisition and processing. In some fields of geophysics, advanced methods from design theory are used, not only to optimize the survey design, but also to shift this entire trade-off relationship between information and cost. In others, either crude rules of thumb are used or, indeed, expected model information is not optimized at all.

This is the first part of a two-part tutorial that provides a theoretical framework from the field of statistical experimental design (SED), within which model-based survey and experimental design problems and methods can be understood. Specifically, these two articles describe methods that are pertinent to the detection and inference of physical properties of rocks in the laboratory, or in the earth. The choice of method to use when designing experiments depends greatly on how easily one can measure information. This in turn depends principally on whether the relationship between data that will be measured and model parameters of interest is approximately linear, or significantly nonlinear. Consequently, the first article focuses on the case where this relationship is approximately linear and the next (in next month's issue of *TLE*) deals with theory for nonlinear design.

surface to constrain optimally the shallow subsurface conductivity structure (Maurer and Boerner, GJI, 1998; Maurer et al., 2000); designing the interrogation of human experts to obtain optimal information to condition geophysical surveys (Curtis and Wood, 2004); designing nonlinear AVO surveys (van den Berg et al., 2003); planning crosswell seismic tomography surveys that illuminate the inter-well structure optimally (Curtis, 1999; Curtis et al., 2004); updating shallow resistivity survey designs in real-time as new data, and hence new information are acquired (Stummer et al., 2004); creating seismic acquisition geometries that maximize resolution of the earth model (Gibson and Tzimeas, 2002).

This tutorial considers the case where we would like to perform an experiment to collect data **d** (seismic, electromagnetic, logs, core, etc.) to constrain some model of earth properties or architecture described by a vector **m**. Say we define a set of basis functions $\{\mathbf{B}_j(\mathbf{x}):j=1,\ldots,P\}$ that describe elementary components of earth properties or architecture. Examples of such basis functions used in geophysics are rock properties in each of a set of mutually-exclusive spatial cells, discrete Fourier components over a finite band-width, scatterers of energy at a set of fixed locations, or statistical properties observed over a finite range of length scales. Possible models of the earth can then be expressed as:

$$\mathrm{M} = \sum_{j=1}^{P} m_j \, \mathbf{B}_j(\mathbf{x}) \qquad (1)$$

The problem of estimating earth composition consists of estimating coefficients m_j.

Designing Experiments to Constrain Parameters

Experiments should be designed such that:

- They can be conducted in practise
- ***Expected*** post-experimental model parameter uncertainties are minimised
  ➔ Model information is maximised
- Costs are constrained/minimised

How does experimental design work?

Designs the ***relationship*** between

a set of parameters and some data

$$F(m) = d$$

How does experimental design work?

Designs the ***relationship*** between
a set of parameters and some data

$$F(m) = d$$

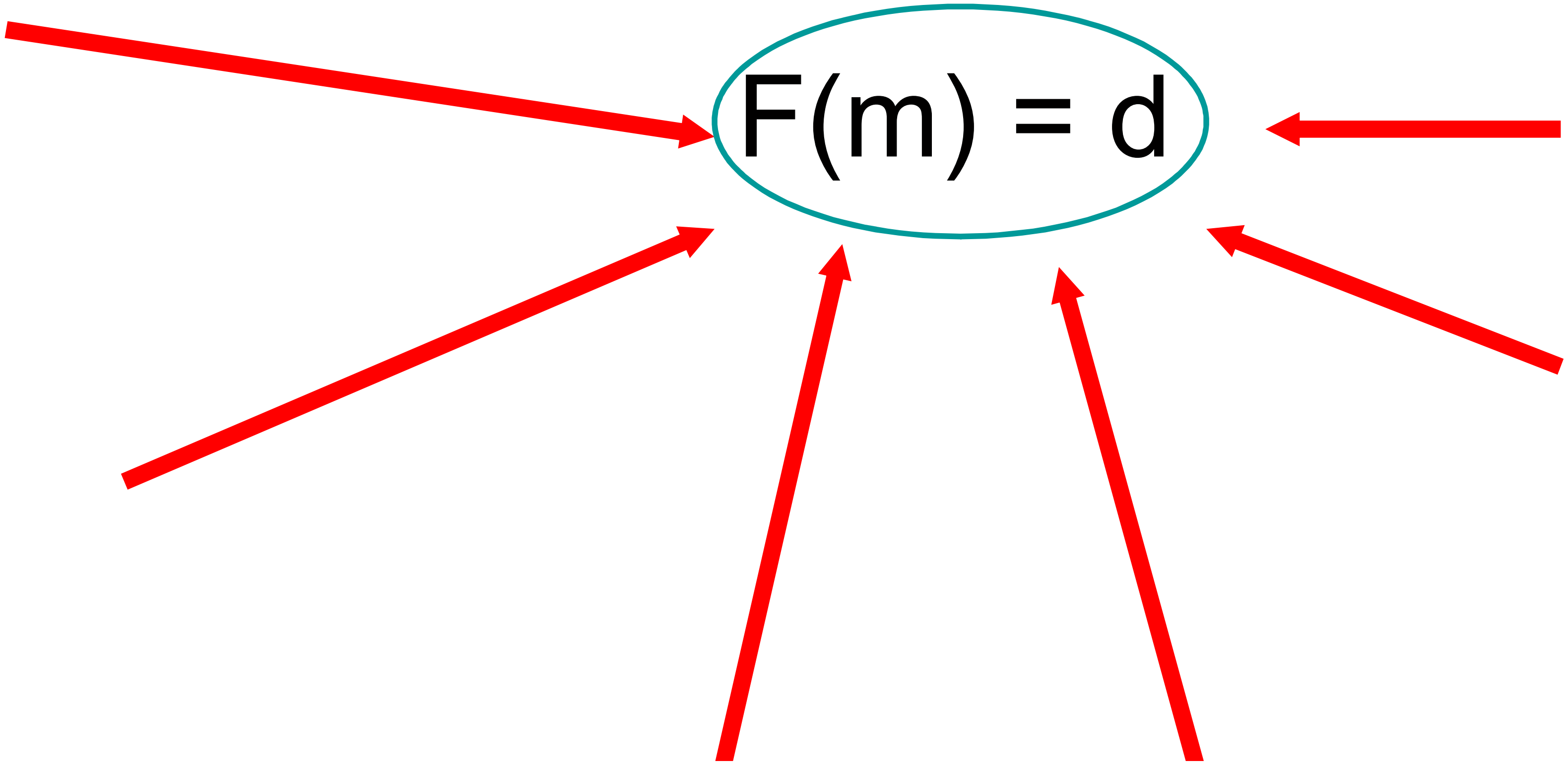
How does experimental design work?

Designs the ***relationship*** between
a set of parameters and some data

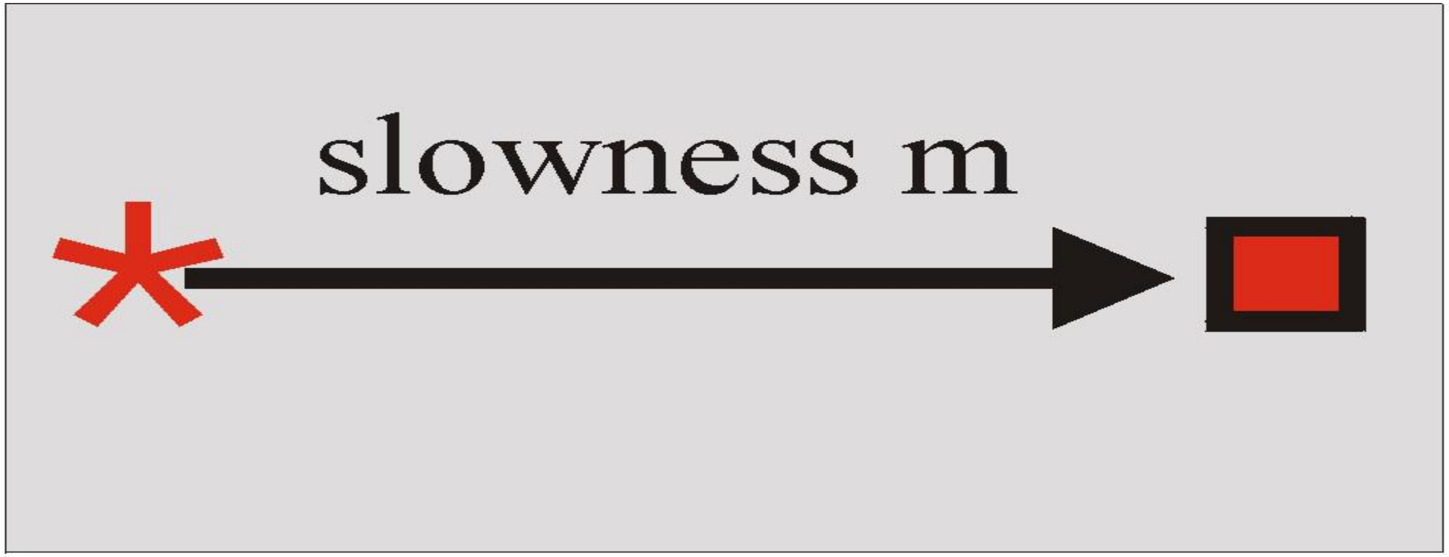
$$F(m) = d$$

How does experimental design work?

Designs the ***problem that we will have to solve***

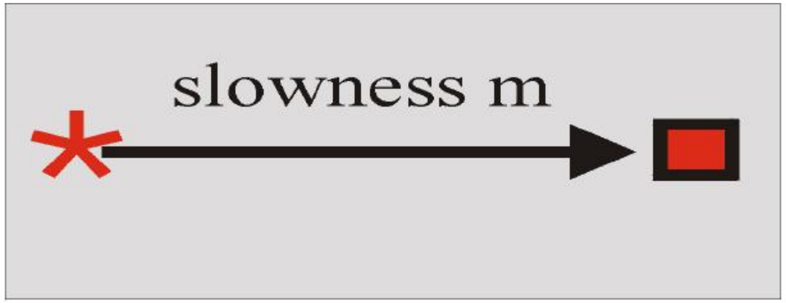


Linear Experimental Design

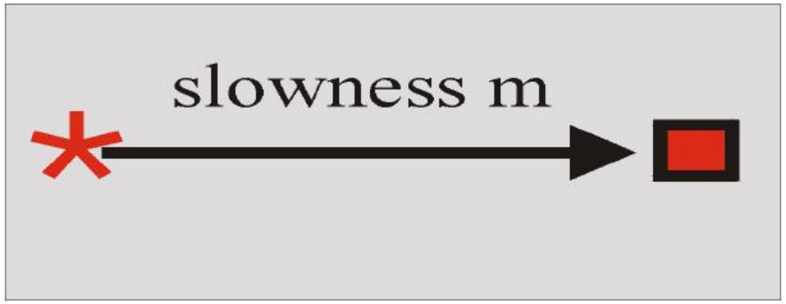


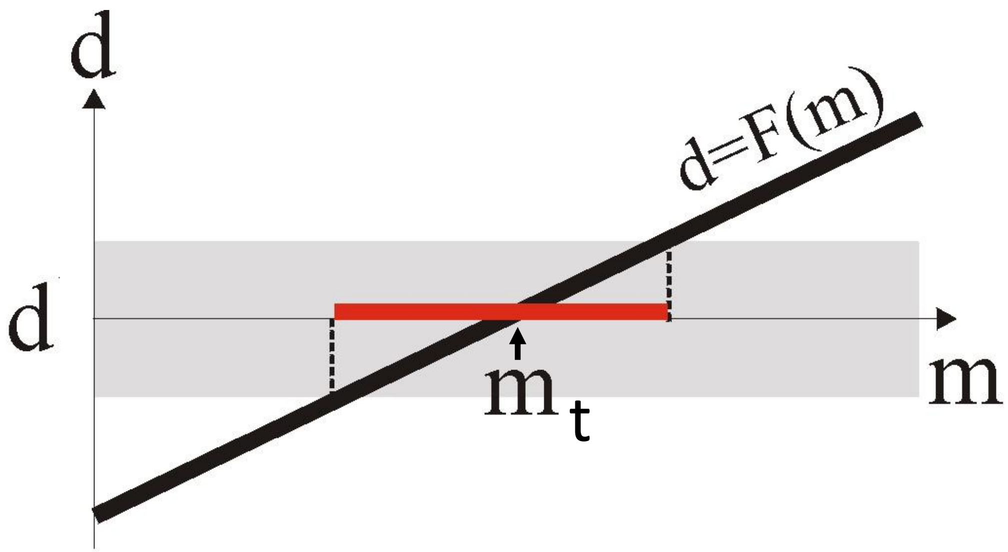
$$m = (\text{wave speed})^{-1}$$

Linear Experimental Design

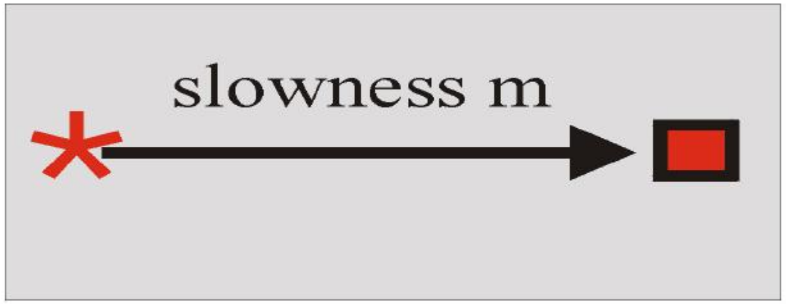


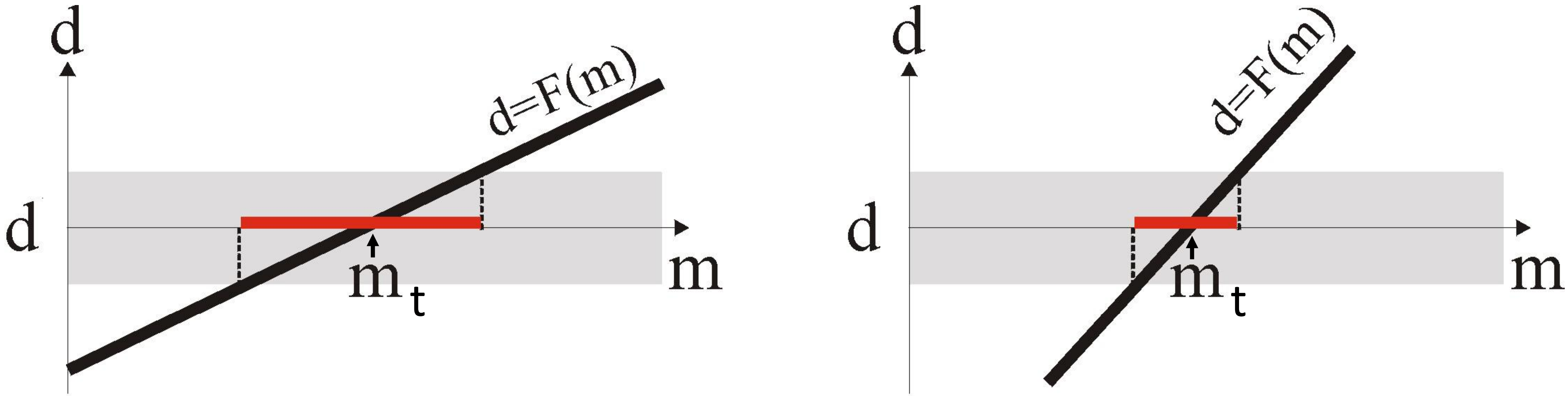
Linear Experimental Design



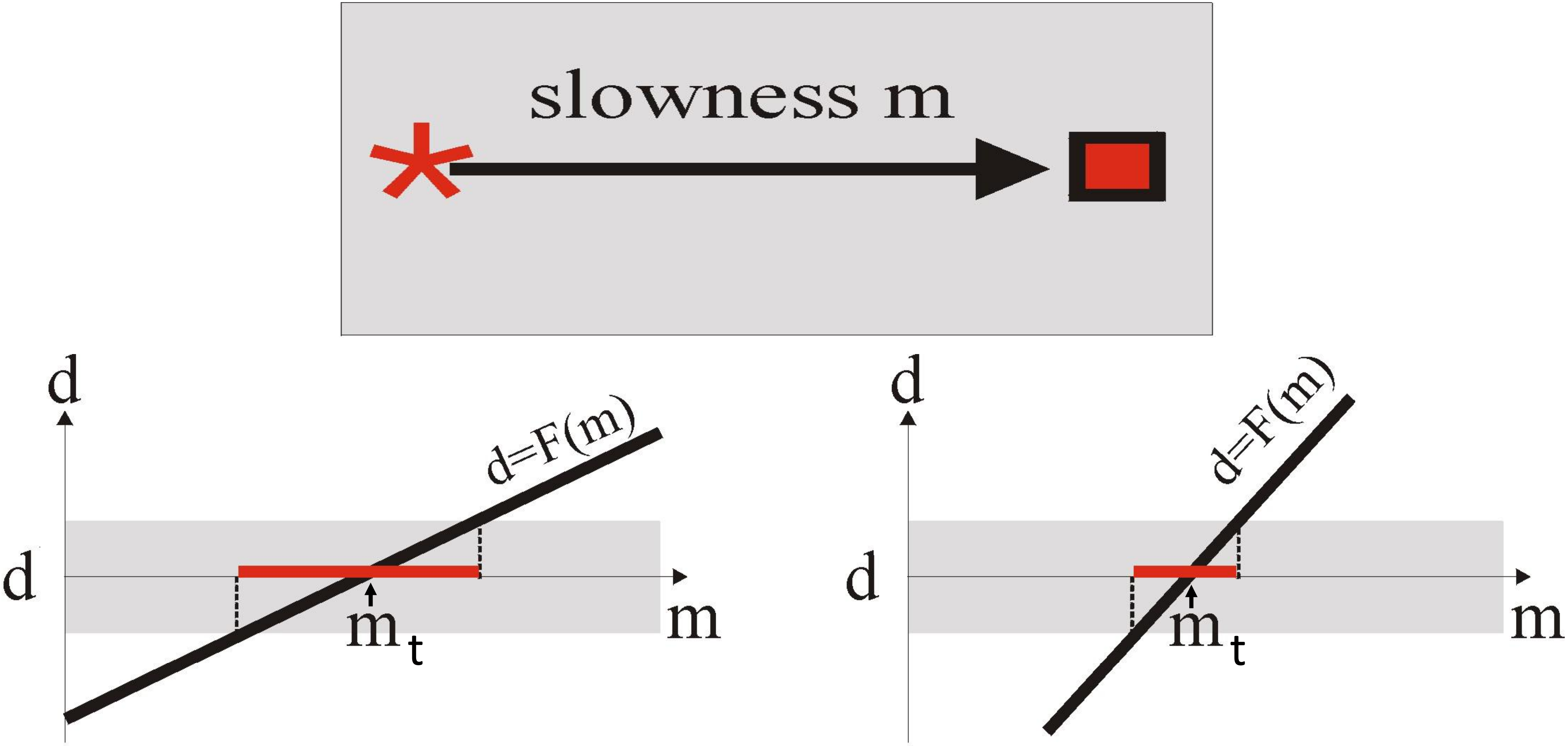


Linear Experimental Design

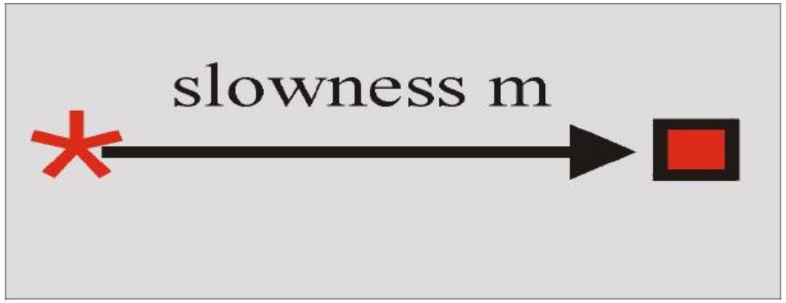


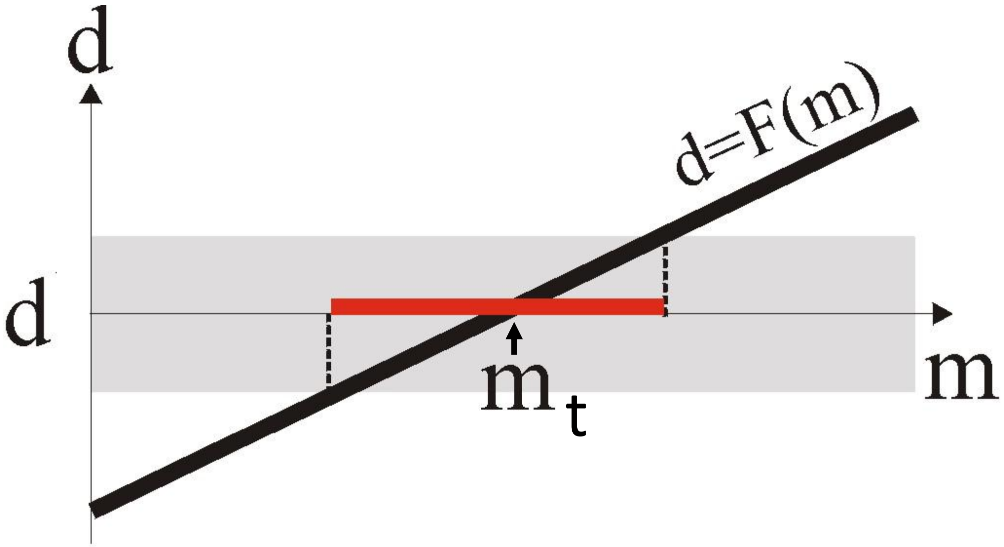


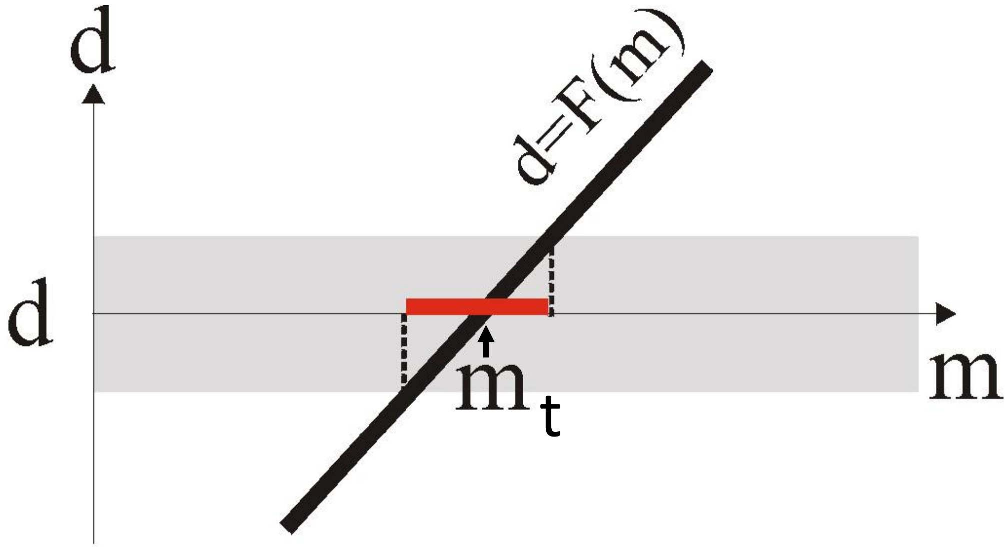
Linear Experimental Design



Linear Experimental Design





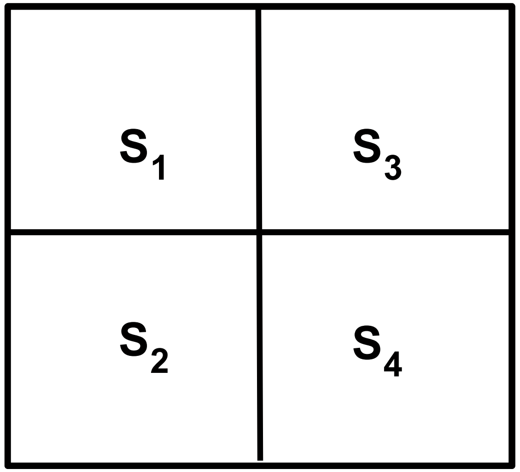


Linear Experimental Design

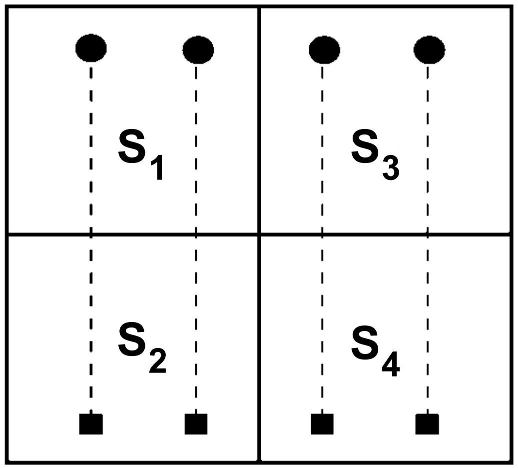
Is this region of the Earth heterogeneous?

Linear Experimental Design

Linear Experimental Design



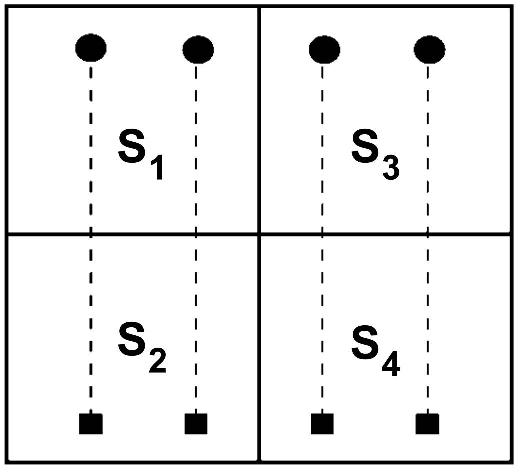
Linear Experimental Design



$$A\,m = d$$

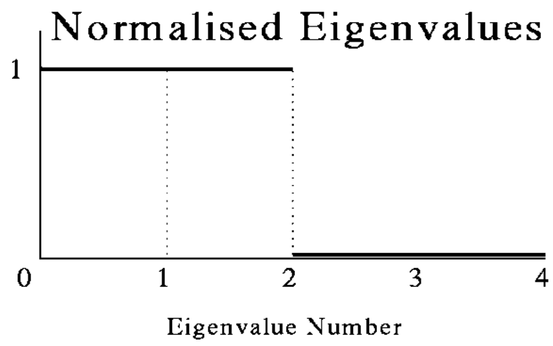
$$m = [A^T A]^{-1} A^T d$$

Linear Experimental Design



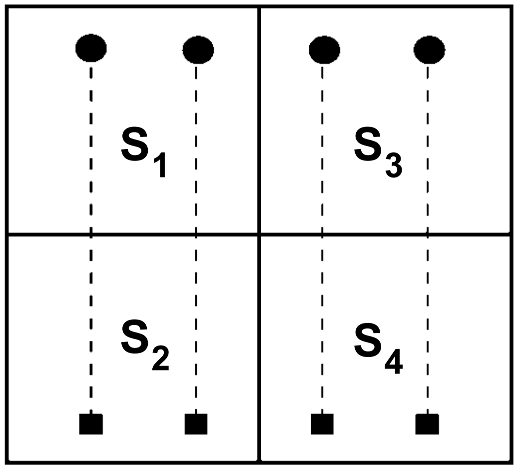
$$A\,m = d$$

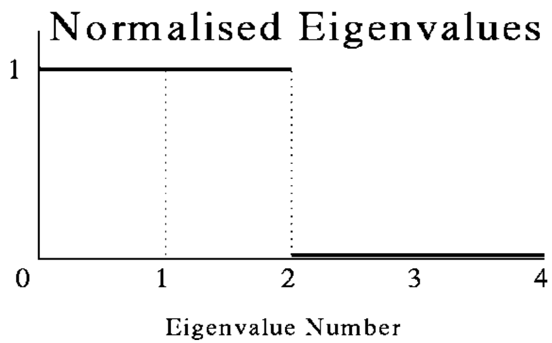
$$m = [A^TA]^{-1}A^Td$$



Linear Experimental Design

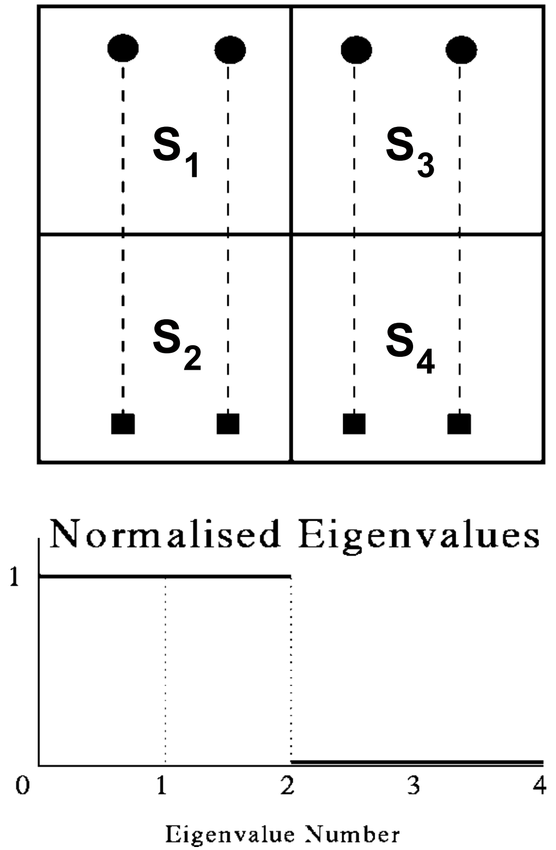
(a)



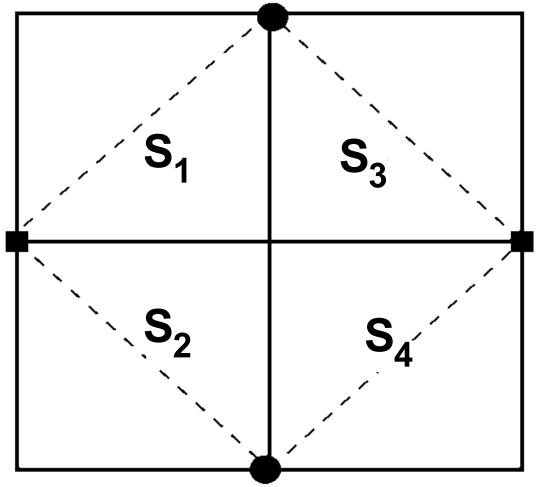


Linear Experimental Design

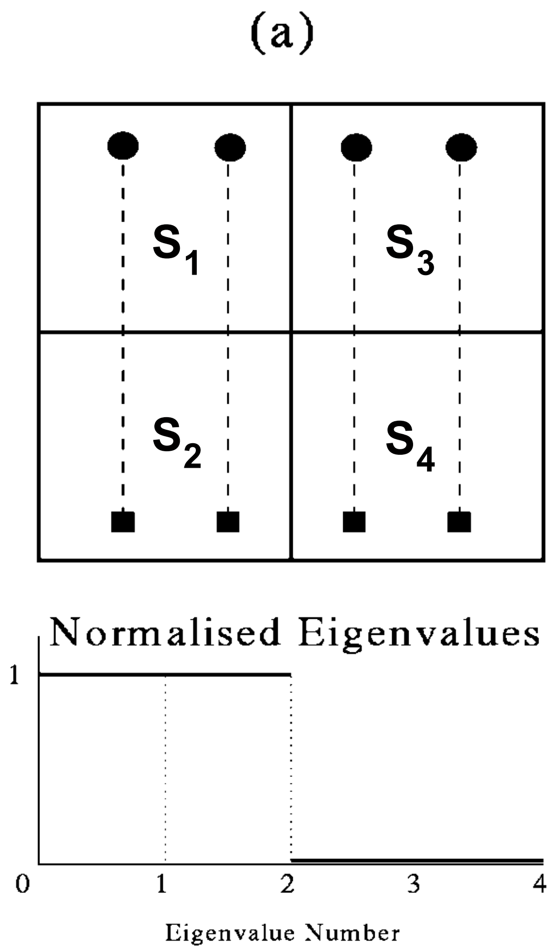
(a)

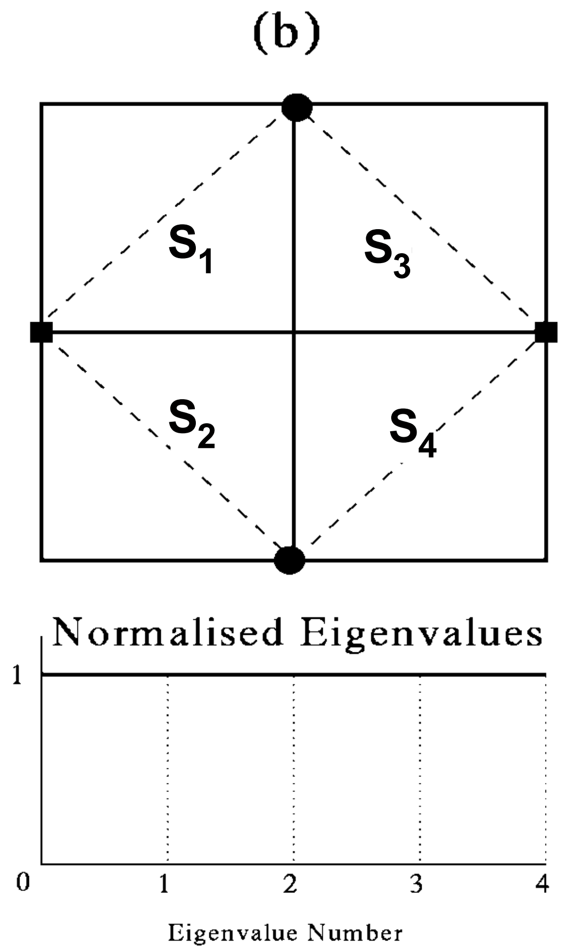


(b)



Linear Experimental Design





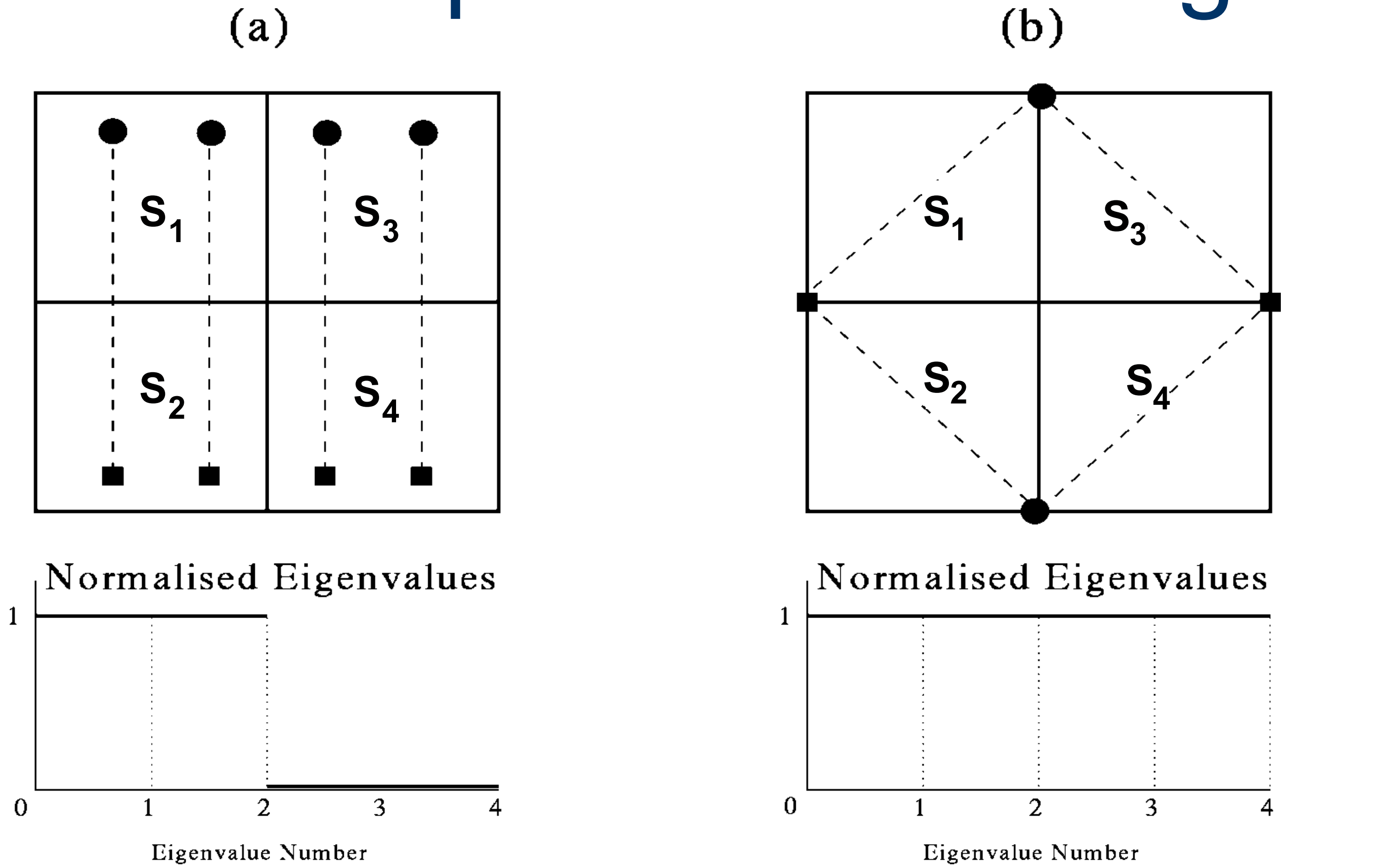
The Survey on the Right…

- Provides twice as many independent pieces of information as the survey on the left, using the same number of data
- Is nominally carried out at half the cost (2 sources + 2 receivers)

General Points

- Eigenvalues specify precisely how many pieces of information can be constrained *in principle (no data uncertainties yet!)*
- For each e-value, corresponding e-vector describes precisely the associated *independent* piece of information
- E-system allows us to create *measures* of design quality…

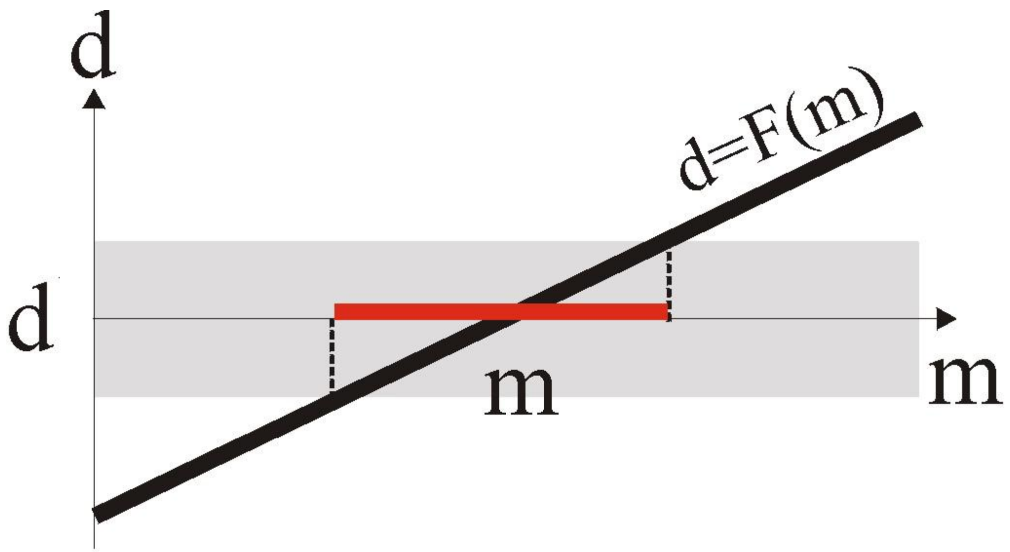
Linear Experimental Design

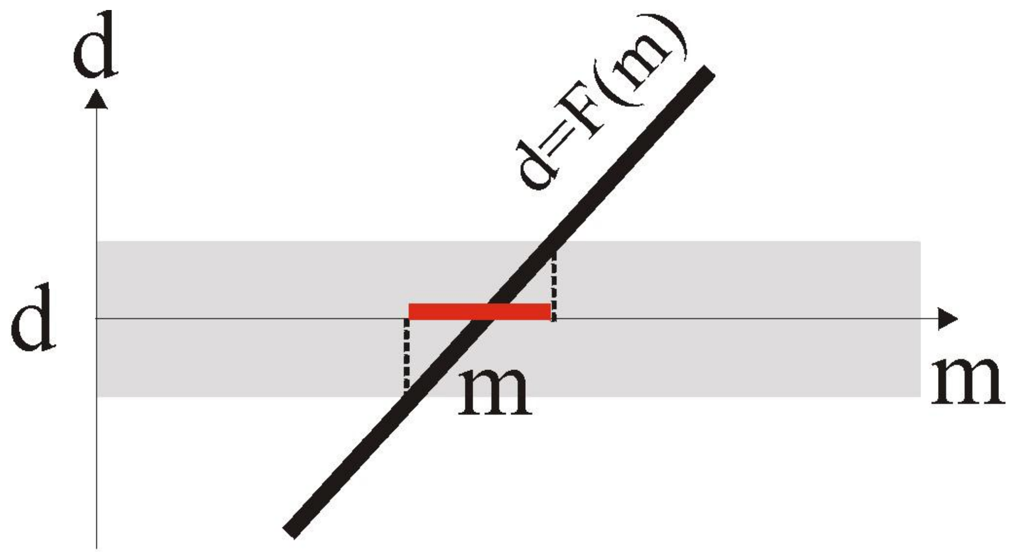


$$\Phi(\mathbf{S}) = \sum_i \frac{\lambda_i}{\lambda_1} \quad or \quad \Phi(\mathbf{S}) = \prod_i \lambda_i$$

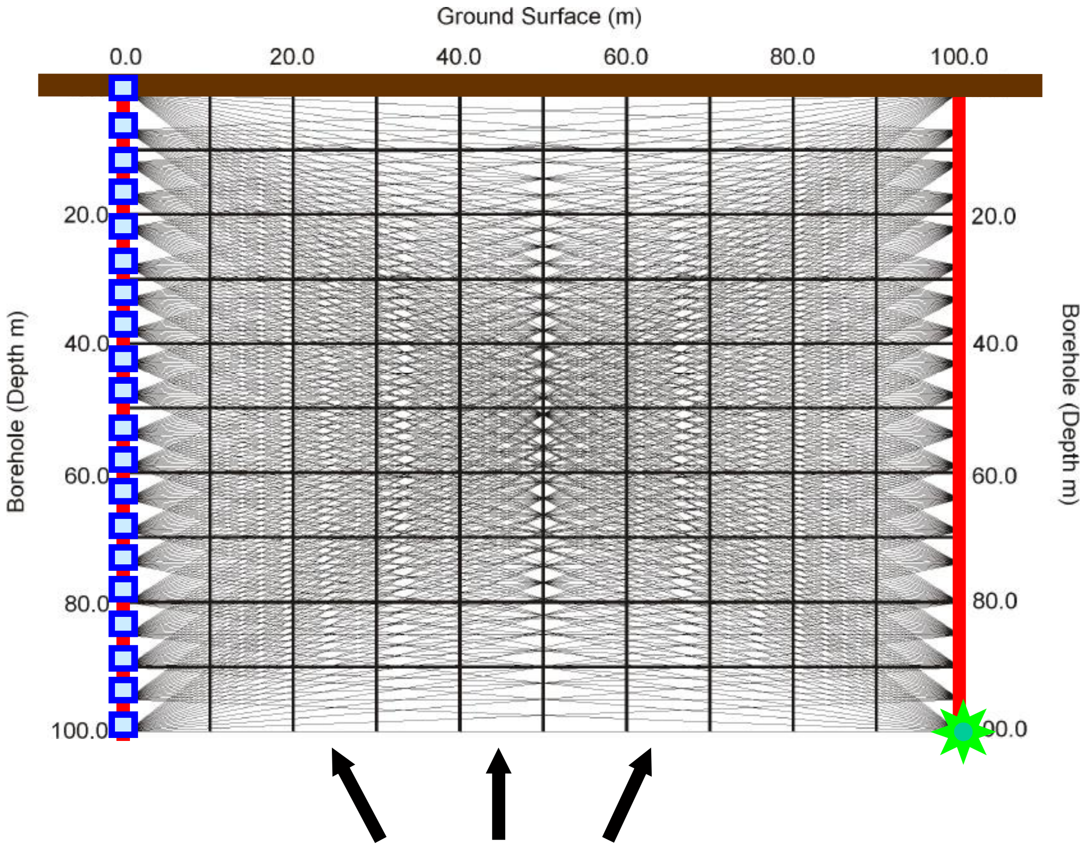
Linear Experimental Design

Eigenvalue = gradient squared



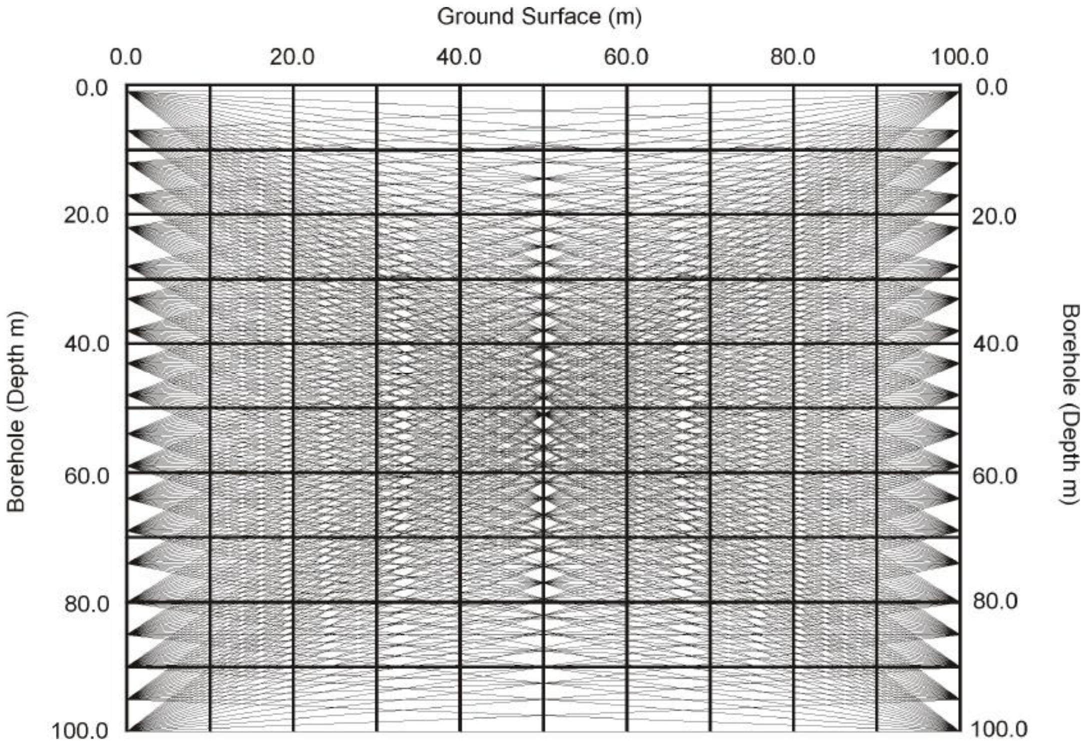


Unfocussed Crosswell Example

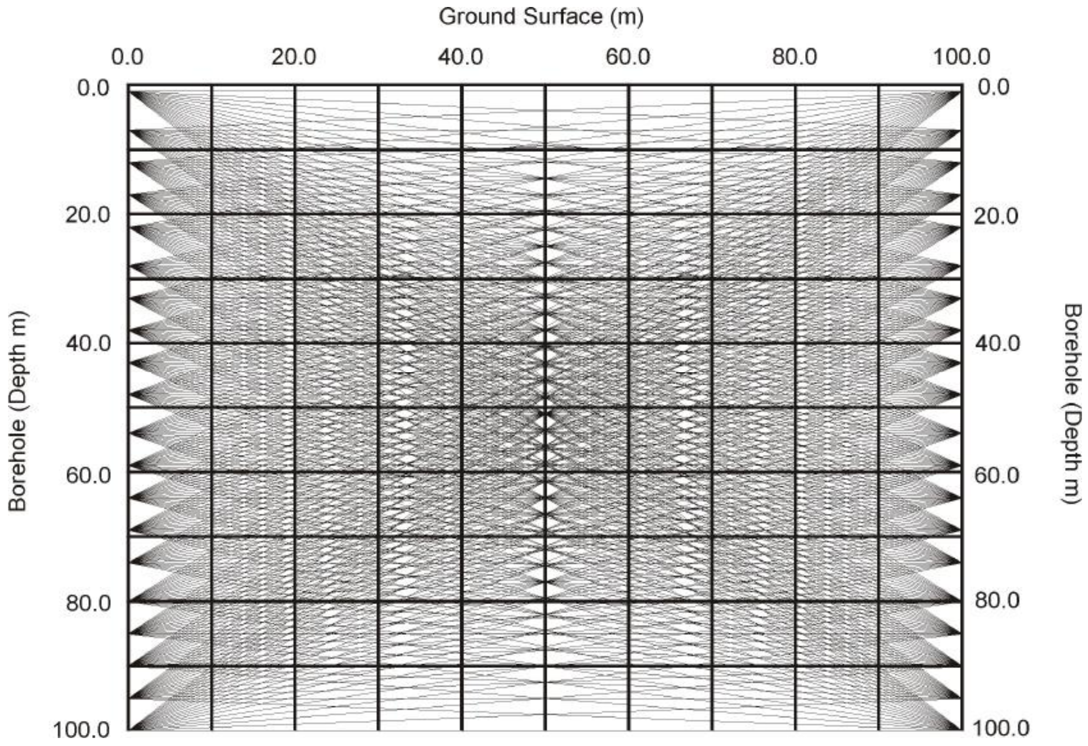


Parameters = cell slownesses

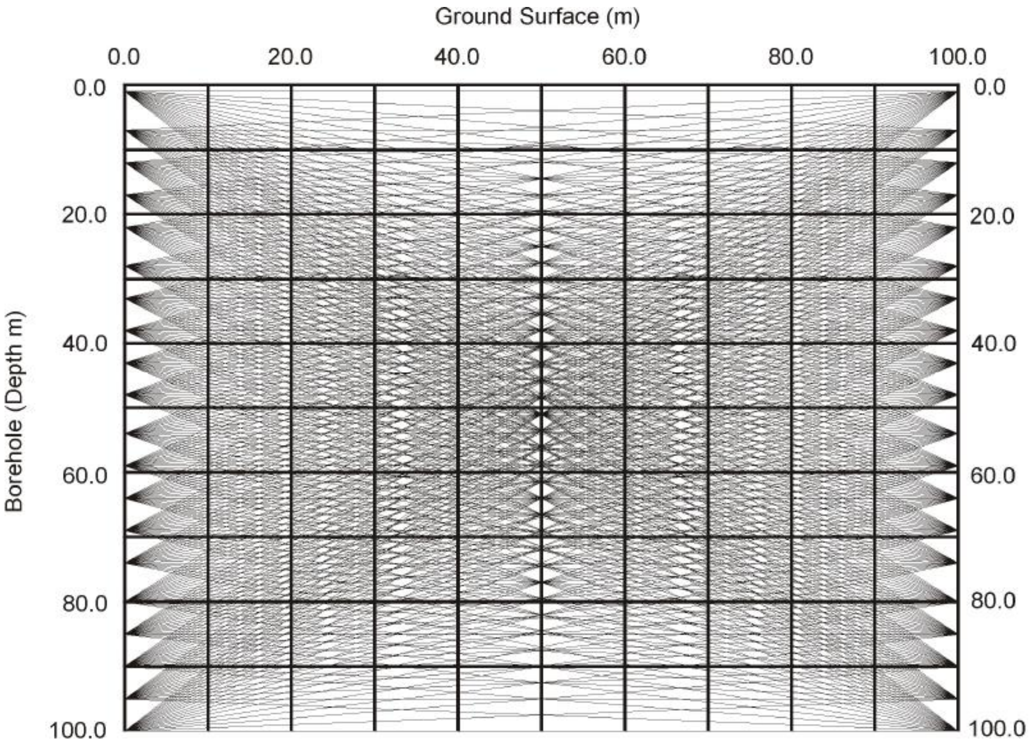
Unfocussed Crosswell Example

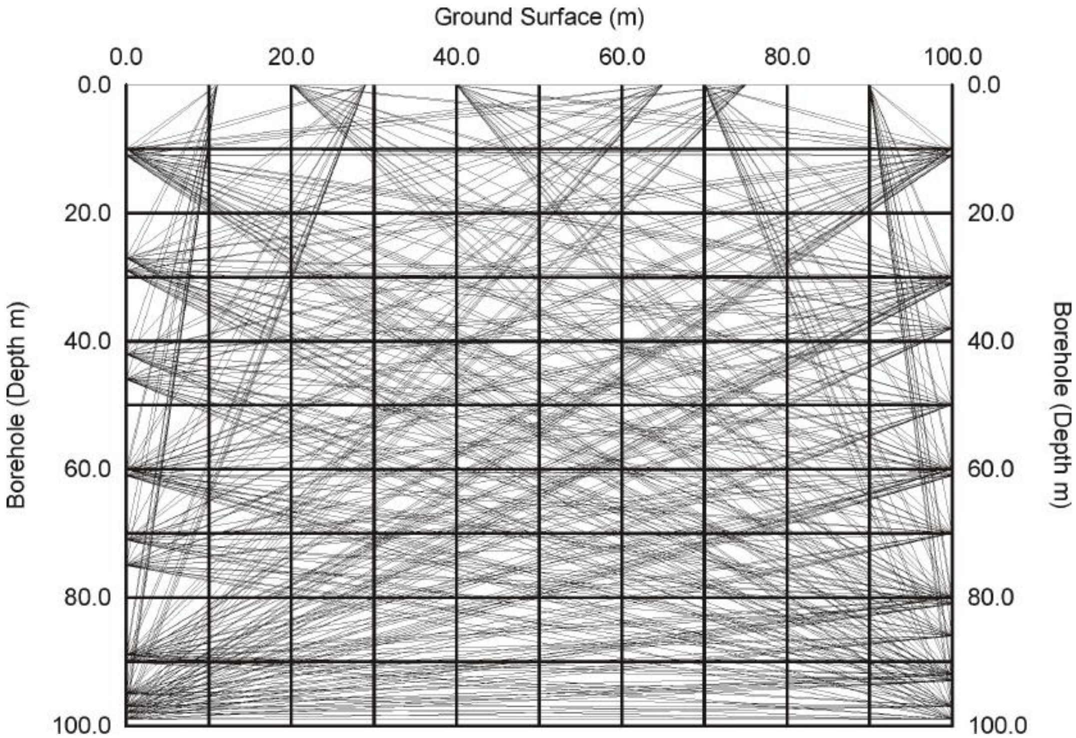


Unfocussed Crosswell Example

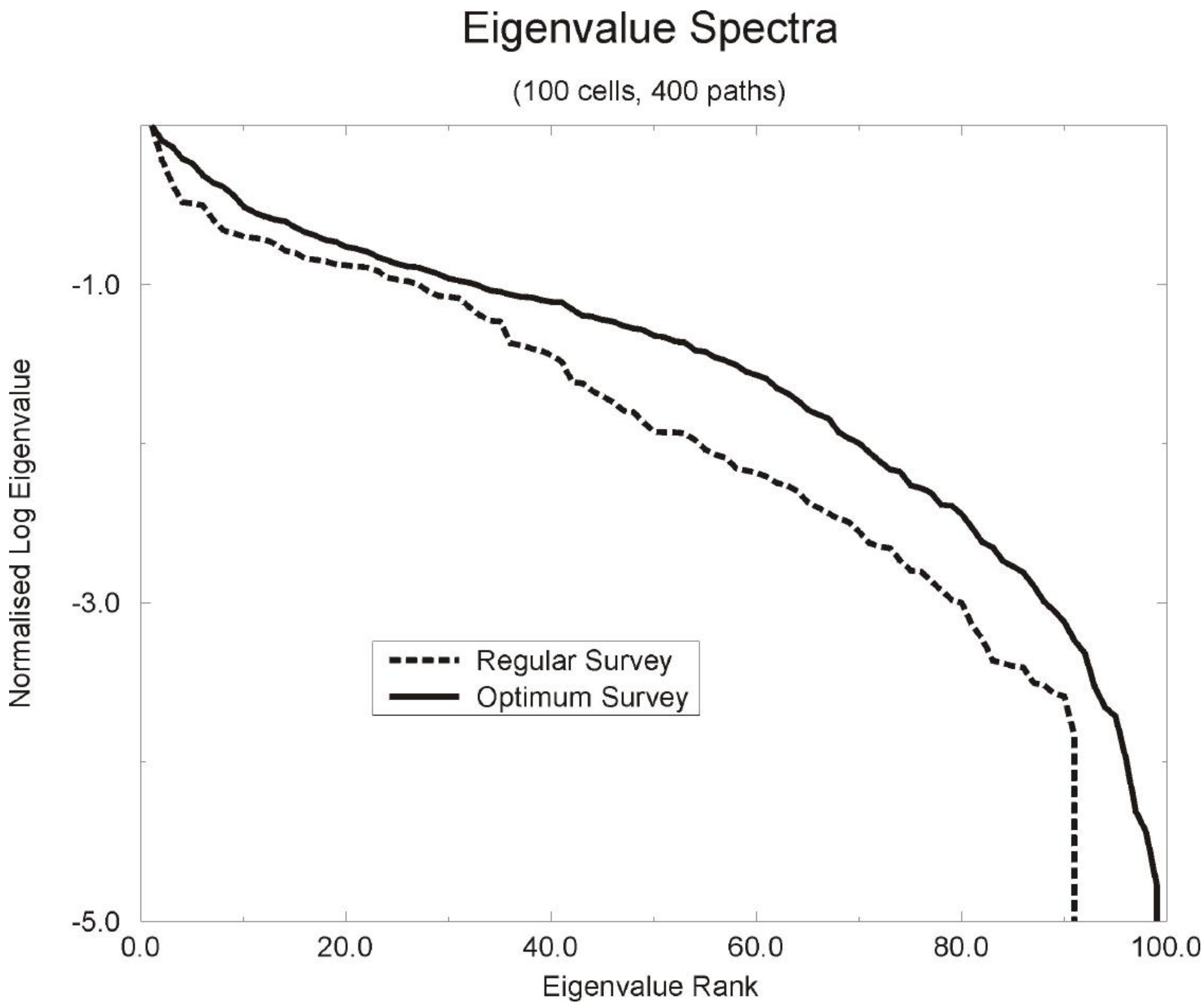


Unfocussed Crosswell Example

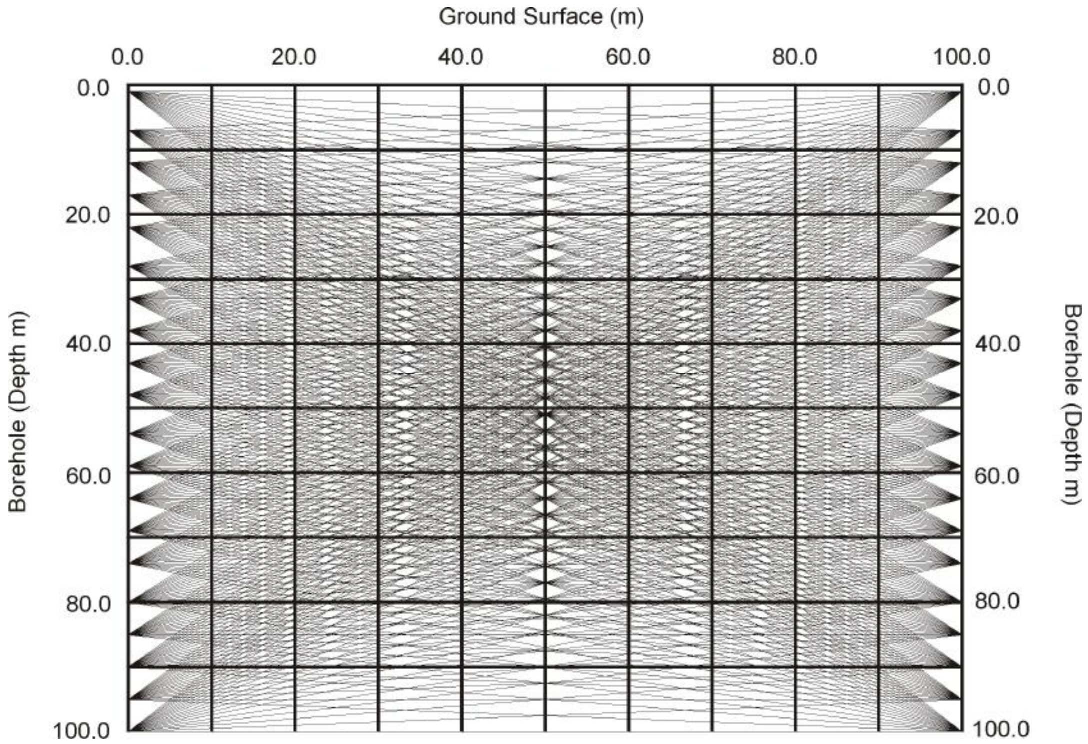


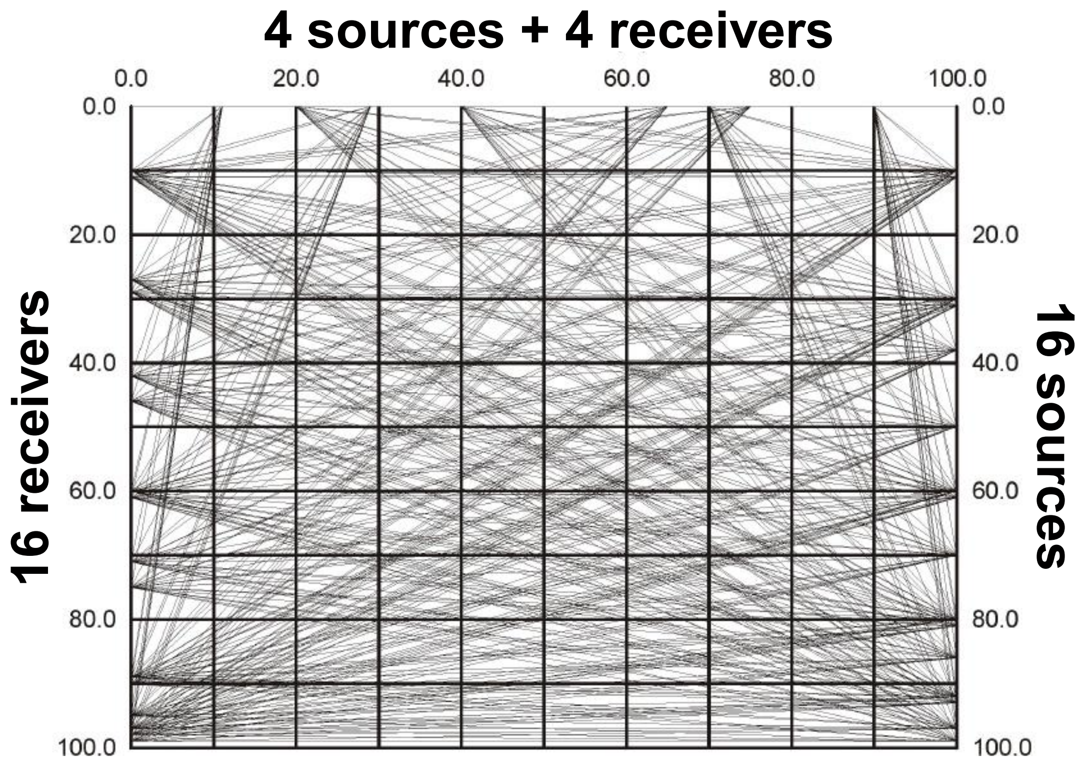


Unfocussed Crosswell Example



Unfocussed Crosswell Example





All long paths: all data average >9 s_i

Greatest path density in centre

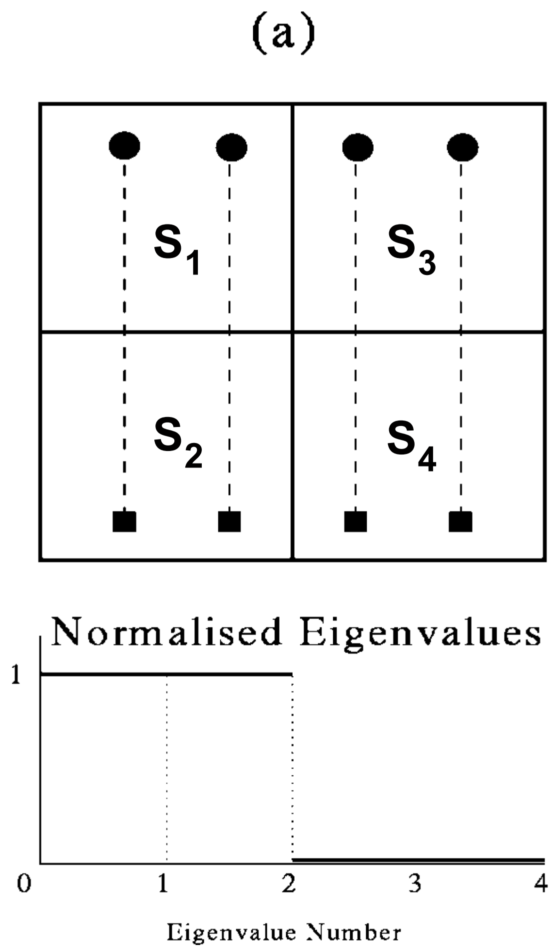
Have not used cheap freedom at surface

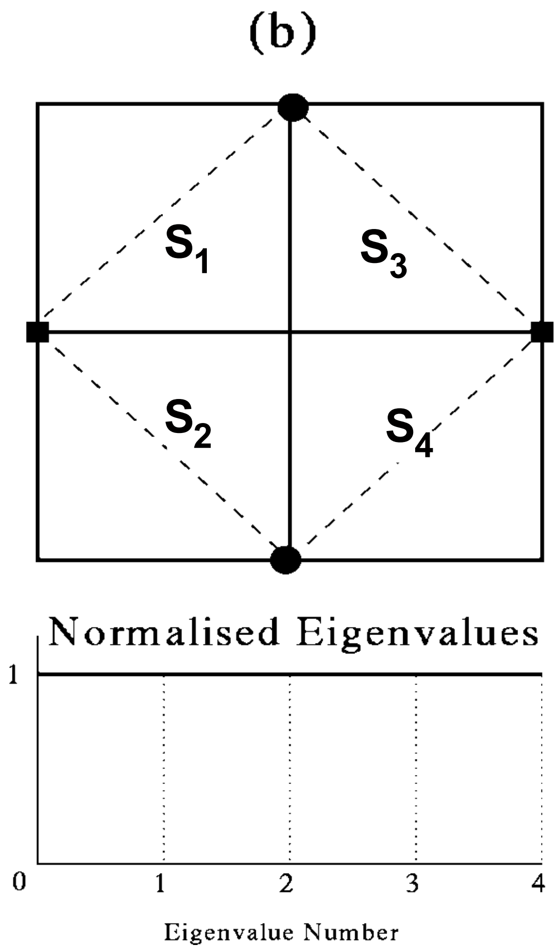
Some short paths

Increasing density with depth
(since only have longer paths at depth)

But **how** dense? **Exactly** where?

Linear Experimental Design

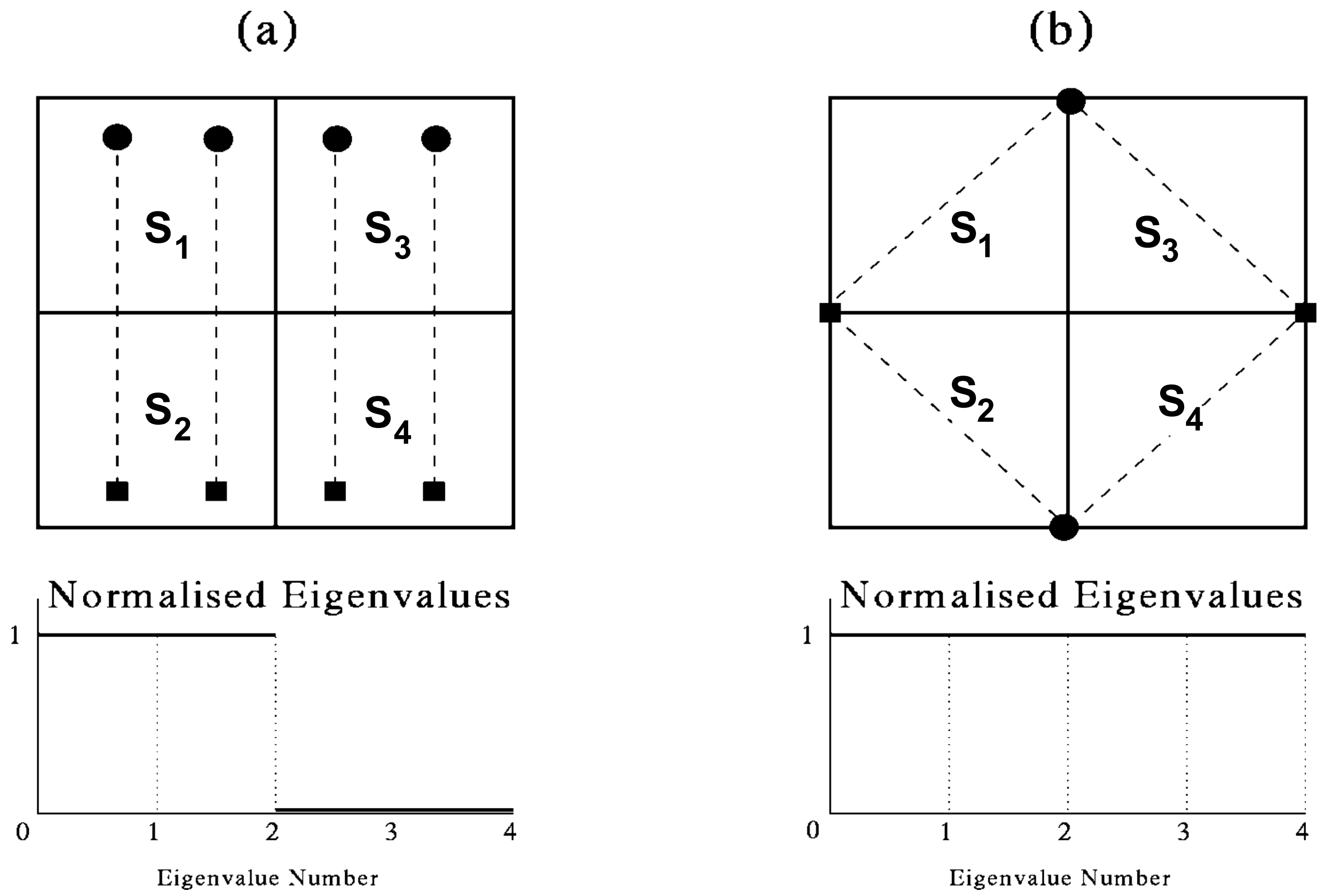




<u>Eigenvalues</u>: *how much* information

<u>Eigenvectors</u>: specifically *what* information

Linear Experimental Design



Eigenvalues: *how much* information

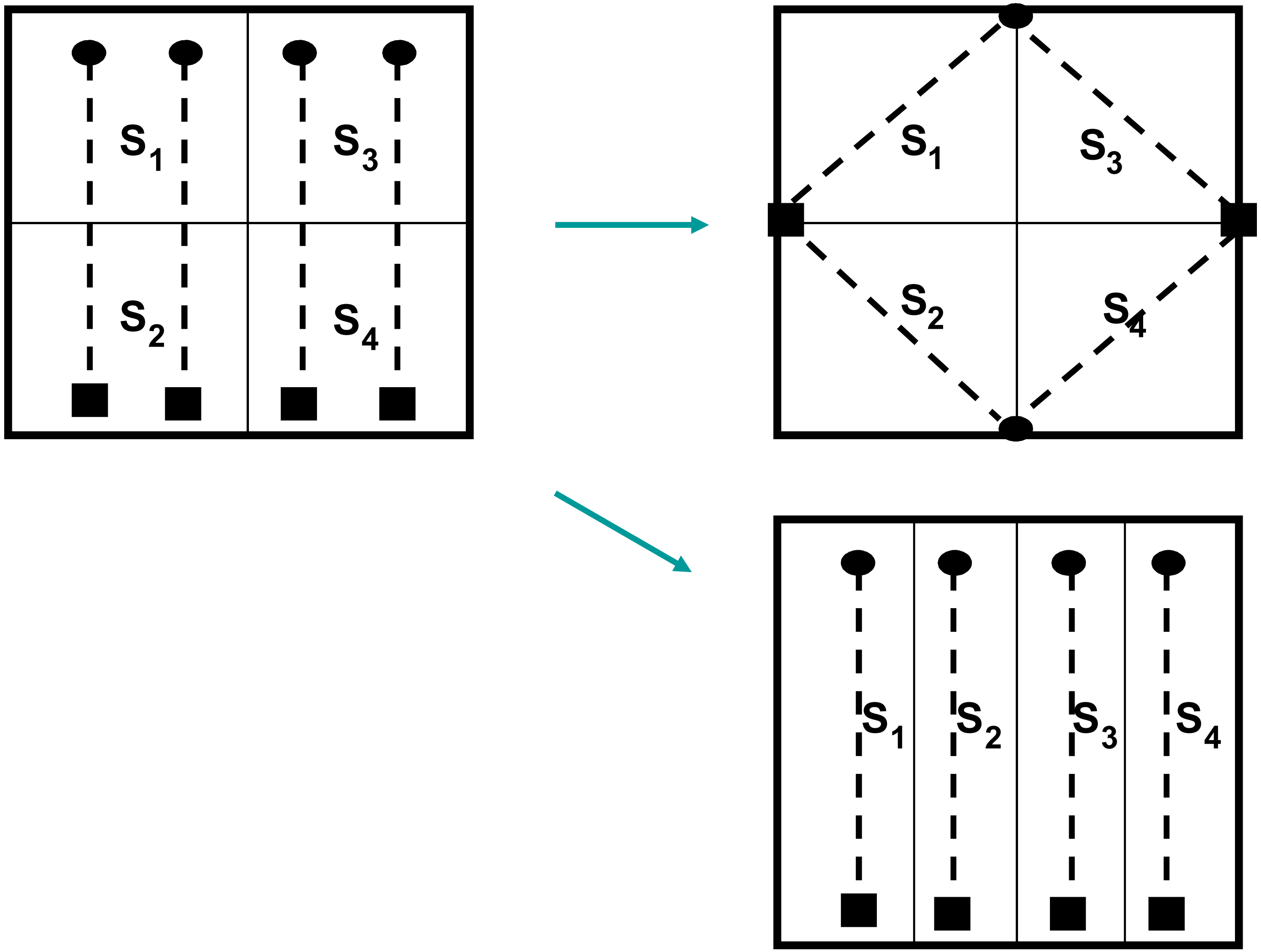
Eigenvectors: specifically *what* information

Two Further Possibilities:

Can design model *parameterisation* using eigenvalues

or, can *focus* information on a model subspace by only maximising eigenvalues of eigenvectors (information) spanning that subspace.

Linear Experimental Design

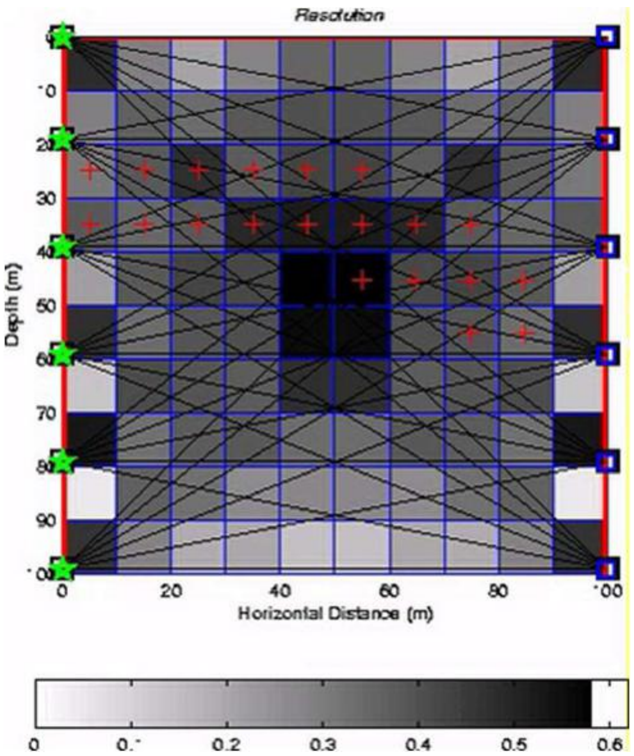


Optimise Data Acquisition

Optimise Model Parameterisation

Is this region of the Earth heterogeneous?

Focussed Crosswell Example



Shading shows diagonal elements of Resolution matrix (max. possible = 1)

Red crosses mark cells spanning model subspace of interest.

H**ow** dense? **Exactly** where?

<u>Not possible to design using intuition alone</u>

➔ Need to solve **Optimisation** Problems

Questions?