Inverse problems

- Introduction

- Probabilistic approach

Wojciech Dębski

Instytut Geofizyki PAN

debski@igf.edu.pl

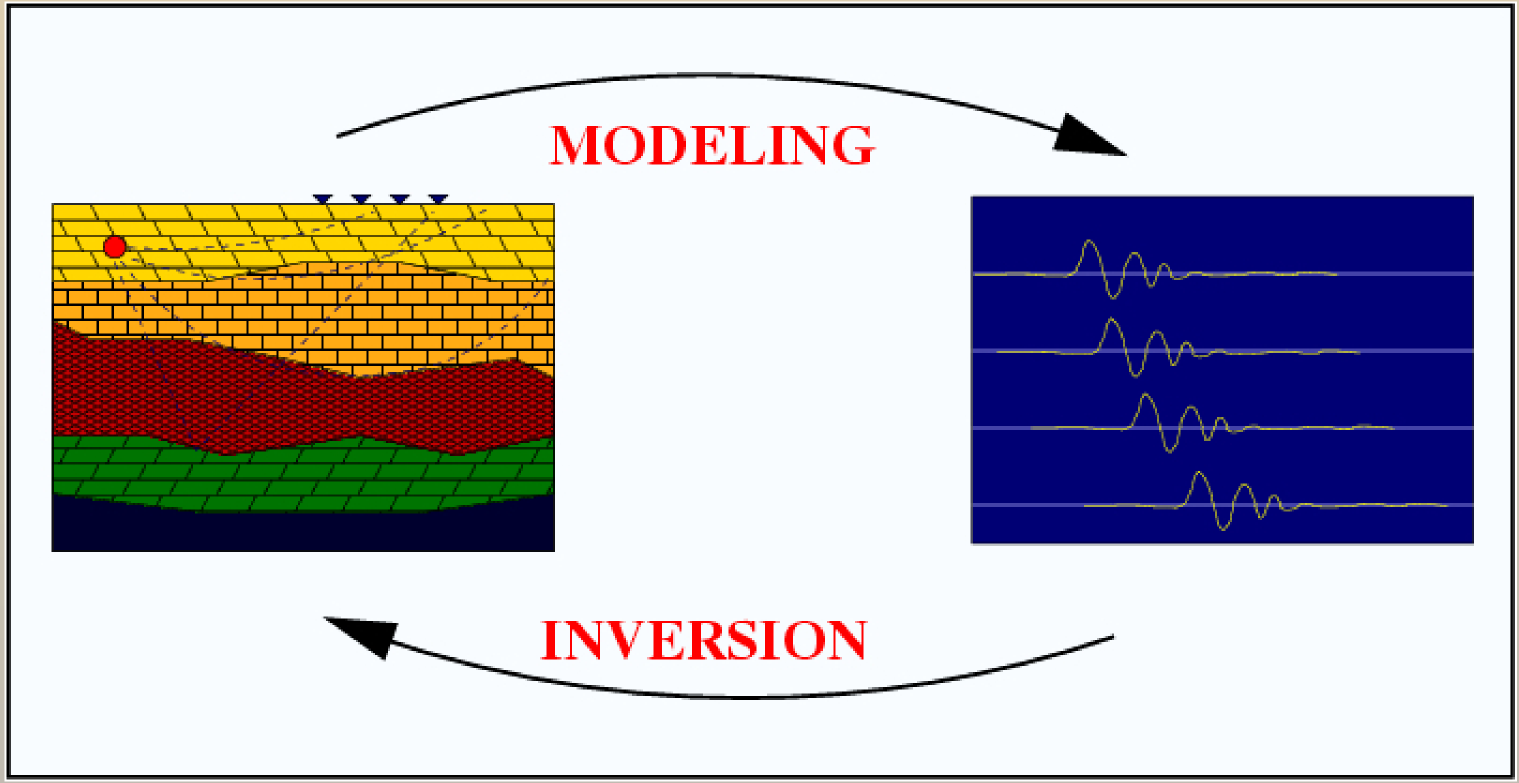
Wydział Fizyki UW , 13.10.2004

Plan of the talk

- I. INTRODUCTION
  - ⋆ forward/inverse problems
  - ⋆ inversion as a parameter estimation task
  - ⋆ direct/indirect measurements
- II. PROBABILISTIC APPROACH
  - ⋆ inversion as an inference process
    - ∗ description of information - probability
    - ∗ experimental, theoretical, and *apriori* information
  - ⋆ Building *A posteriori* pdf
    - ∗ Bayesian rule - *a posteriori* probability
    - ∗ Tarantola's (probabilistic) approach
  - ⋆ *A posteriori* pdf - examples
- II. GEOPHYSICAL EXAMPLES

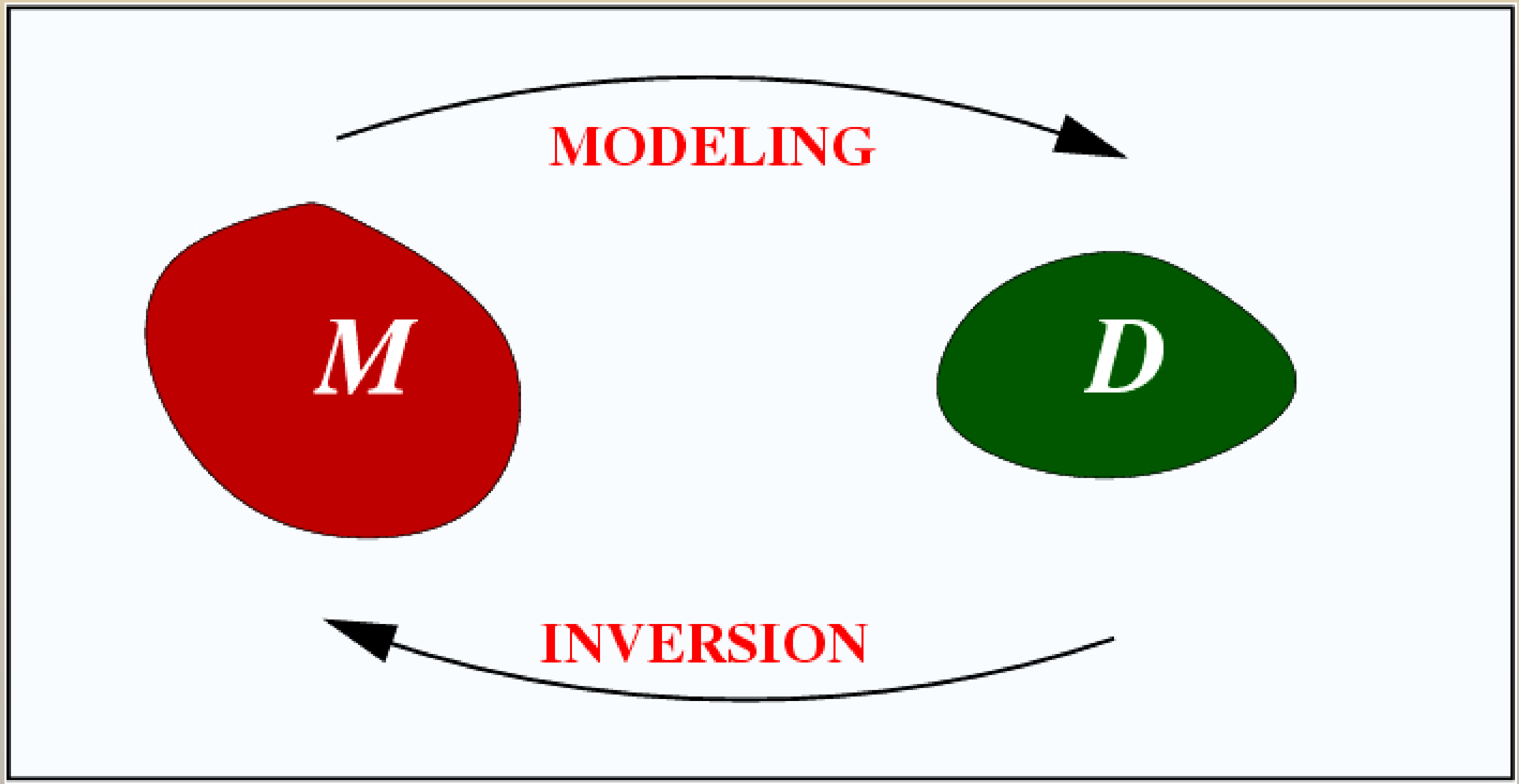
Forward and inverse problems

forward modelling



Forward and inverse problems

inversion



Forward and inverse problems

parameter estimation

Forward problem:

model parameters => observed data

Inverse problem:

observed data => model parameters

Model and Data Parameters

Physical System:

$$p_1, p_2, \cdots p_K$$

Model parameters:

$$\mathbf{m} = (m_1, m_2, \cdots m_M)$$

Predicted (Observed) data:

$$\mathbf{d} = (d_1, d_2, \cdots d_N)$$

Forward modelling:

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

Linear Inverse Problem

Algebraic approach

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

Naive inversion:

$$\mathbf{m}^{est} = \mathbf{G}^{-1} \cdot \mathbf{d}$$

Algebraic method:

$$\mathbf{G}^T \cdot \mathbf{d} = \mathbf{G}^T \mathbf{G} \cdot \mathbf{m}$$

$$\mathbf{G}^T \mathbf{G} \implies \mathbf{G}^T \mathbf{G} + \lambda \mathbf{I}$$

$$\mathbf{m}^{est} = (\mathbf{G}^T \mathbf{G} + \lambda \mathbf{I})^{-1} \cdot \mathbf{d}$$

Inverse Problem

Optimization approach

Searching the model space for the model which "best fits" the data

Least squares method:

$$(\mathbf{d}^{obs} - f(\mathbf{m}))^T(\mathbf{d}^{obs} - f(\mathbf{m})) = \min$$

More generally:

$$||(\mathbf{d}^{obs} - f(\mathbf{m}))||_D + ||\mathbf{m} - \mathbf{m}^{apr}||_M = min$$

Direct / Indirect

measurements

Counts:

- events
- No. of particles
- quantified parameters

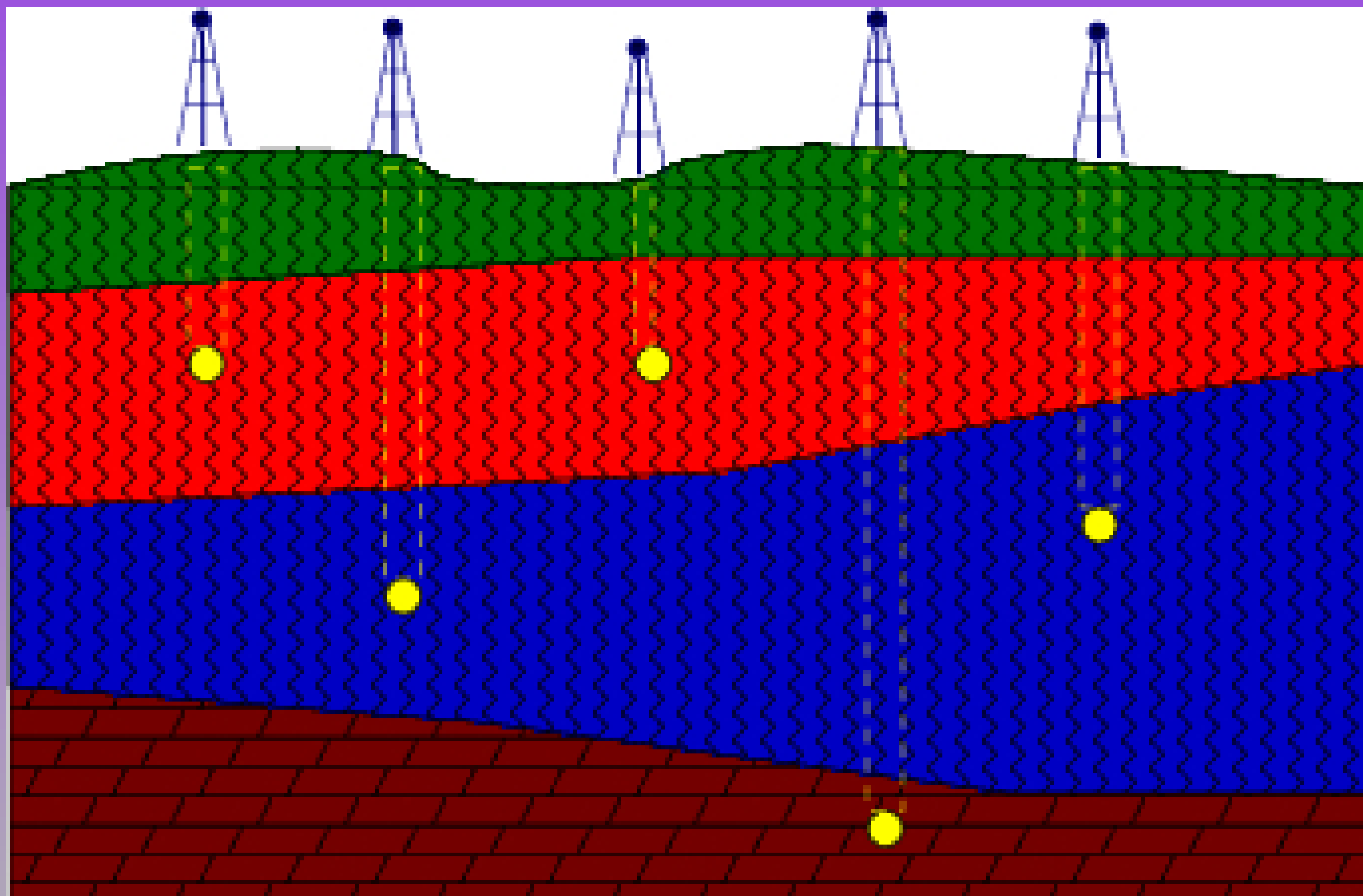
Counts in a scale unit:

- mass
- luminosity
- temperature

Quantities unmeasureable directly:

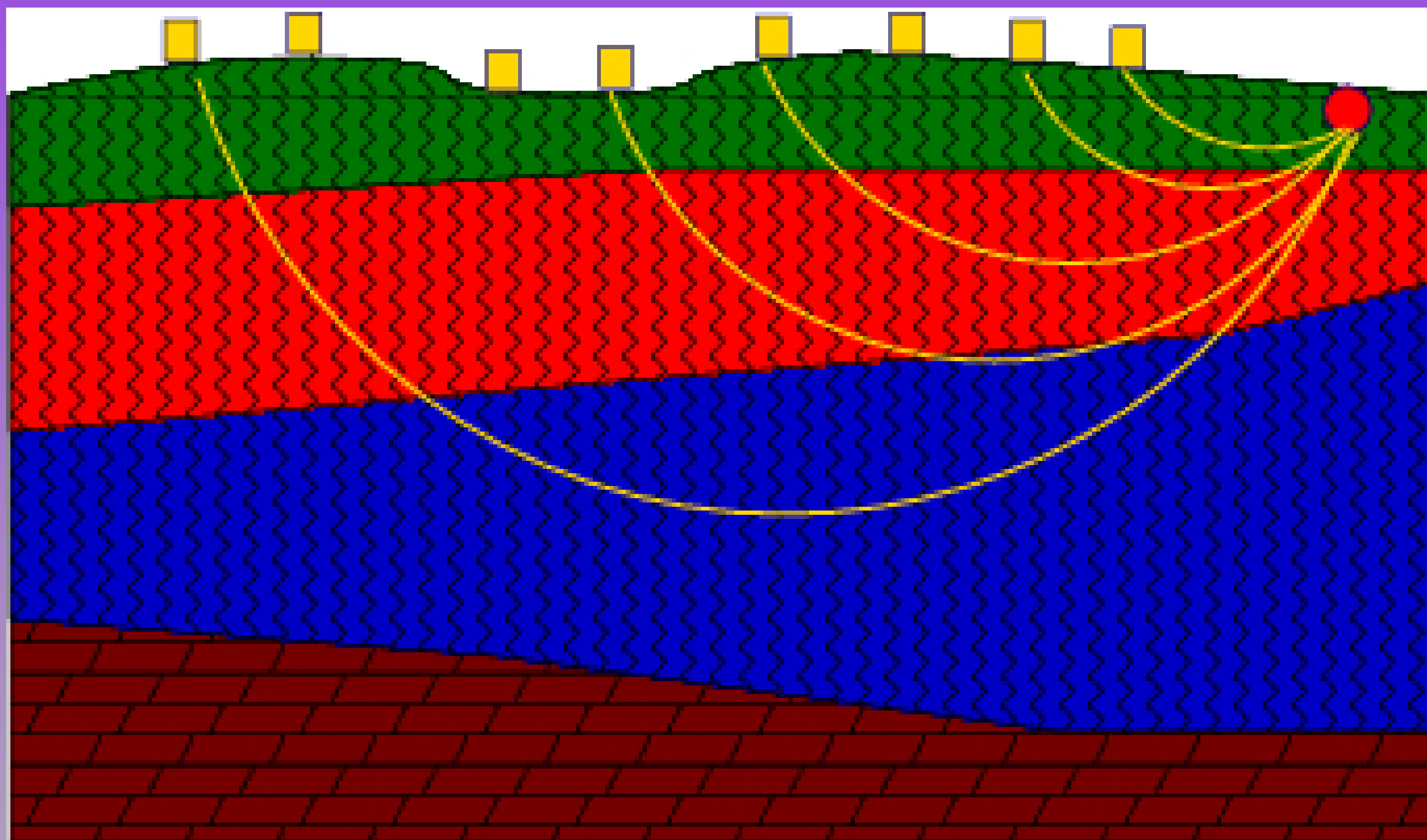
- mas of a earth, star, galaxy
- temperature distribution in the Earth, the sun, etc.
- mass of elementary particles
- ...

Direct & Indirect
measurements

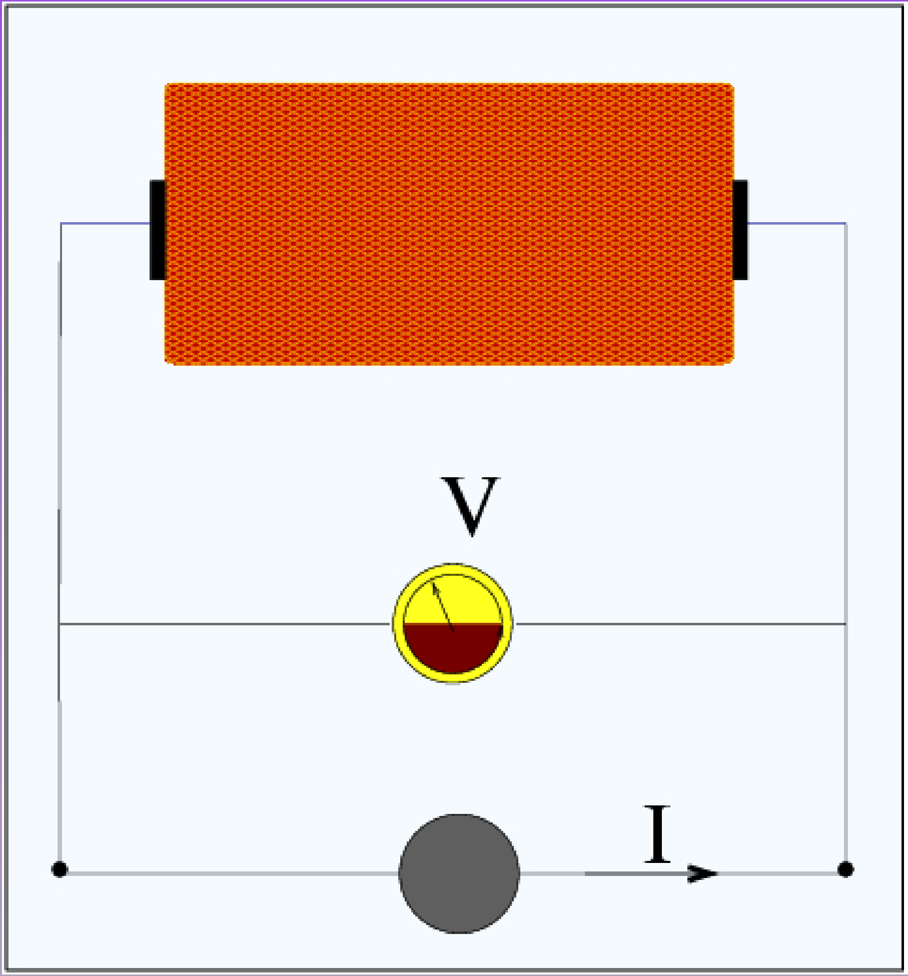


Direct & Indirect

measurements



Indirect measurement - Example



Data: V

Theory (Ohm law + ...)

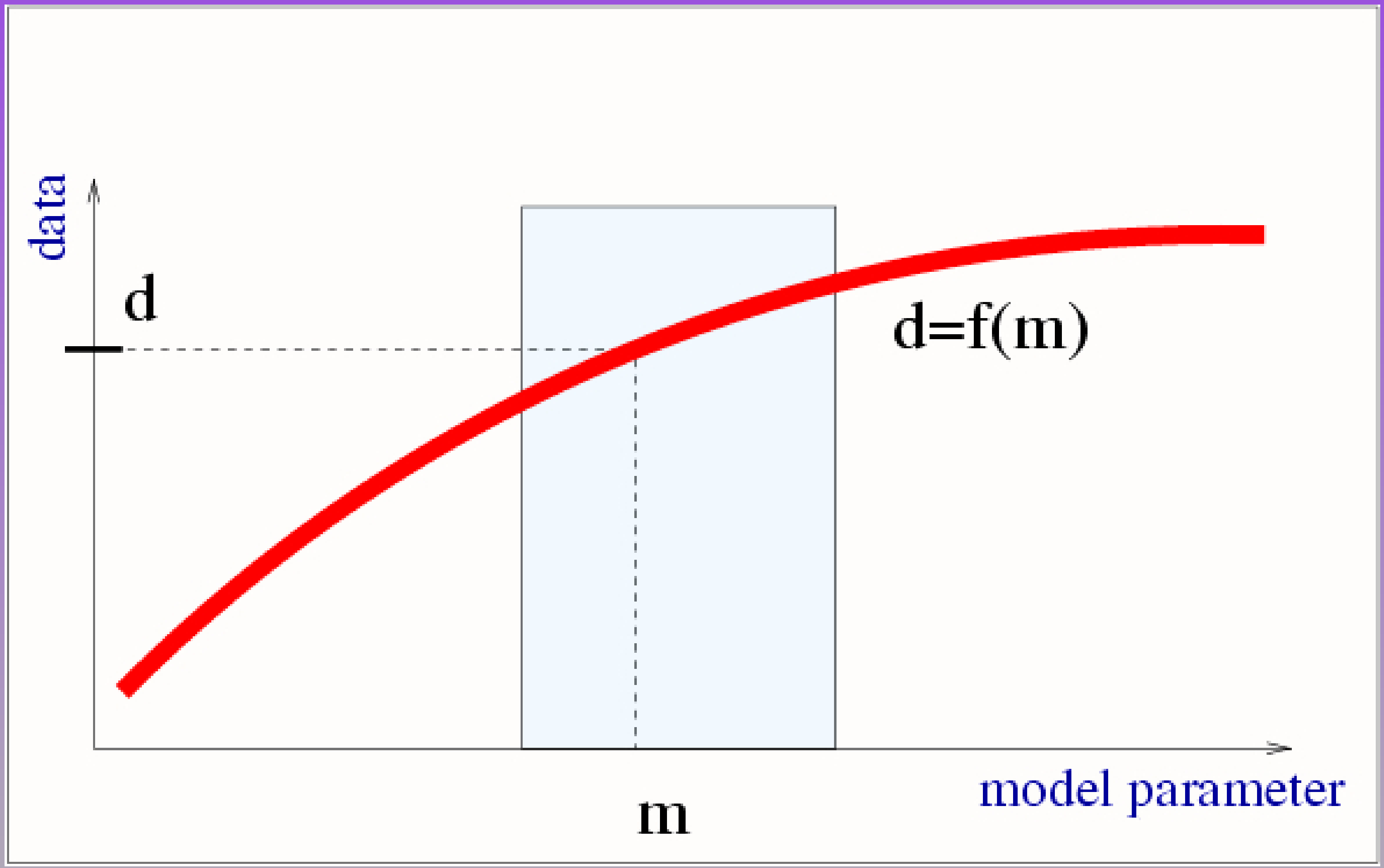
$$V = I\,\rho + n(\rho, \dot{\rho})$$

A priori

$$1 < \rho < 10$$

Inversion

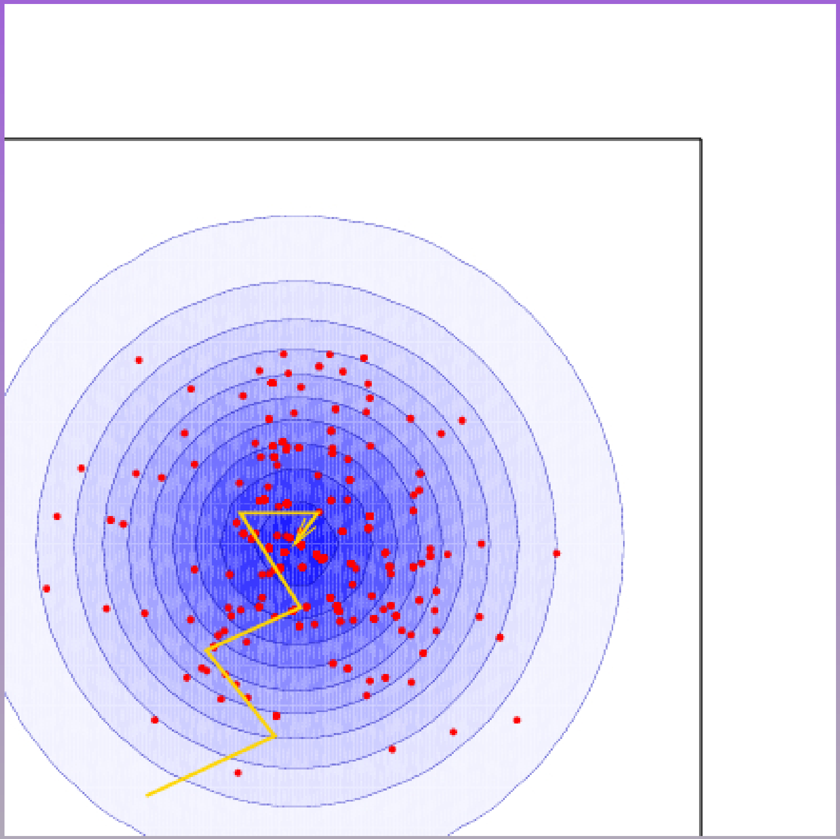
Back projection



Inversion

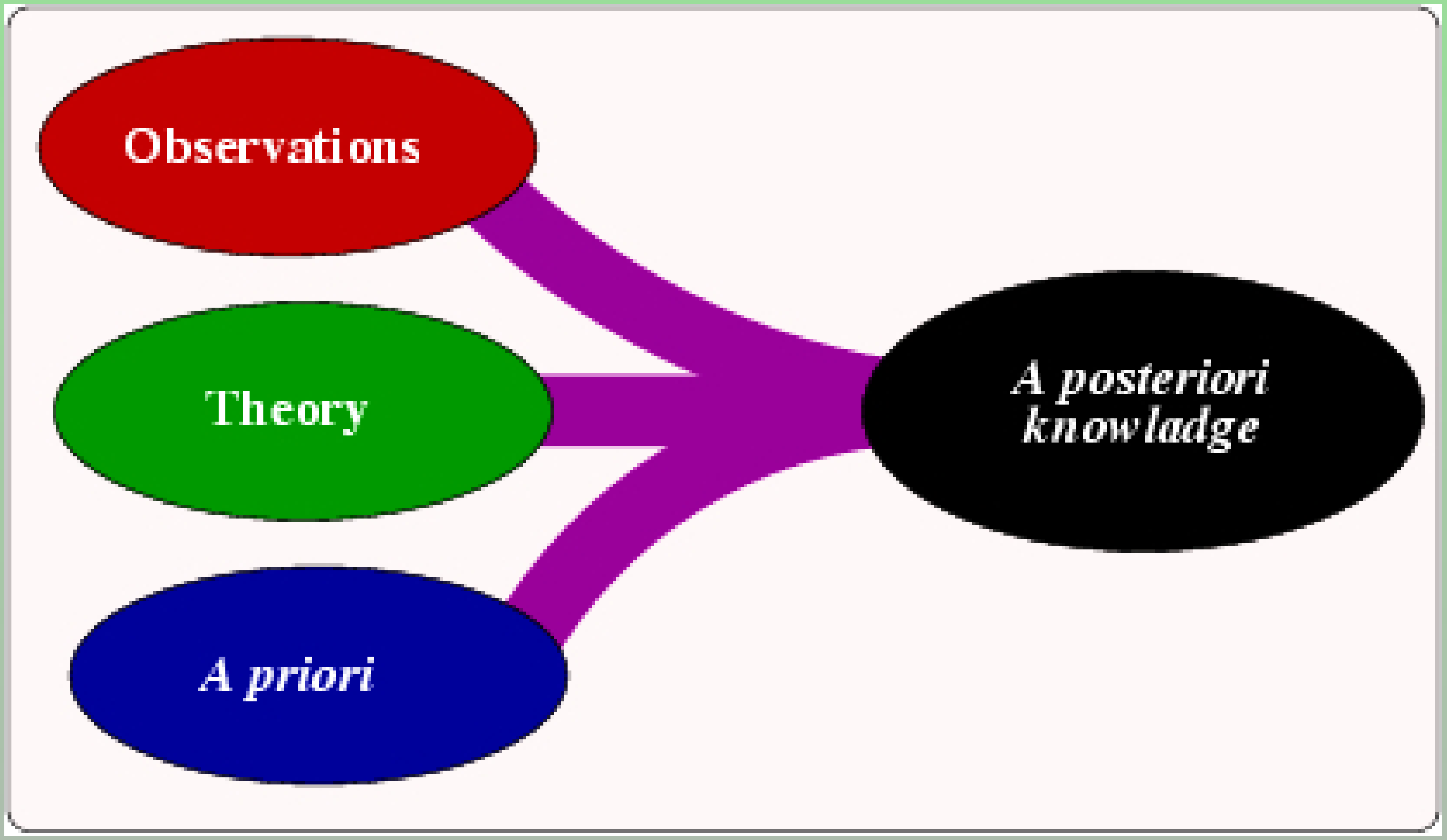
optimization

$$S(\rho) = ||V^{obs} - V(\rho)|| = \min$$



Inversion as Joining Information

(Inference)

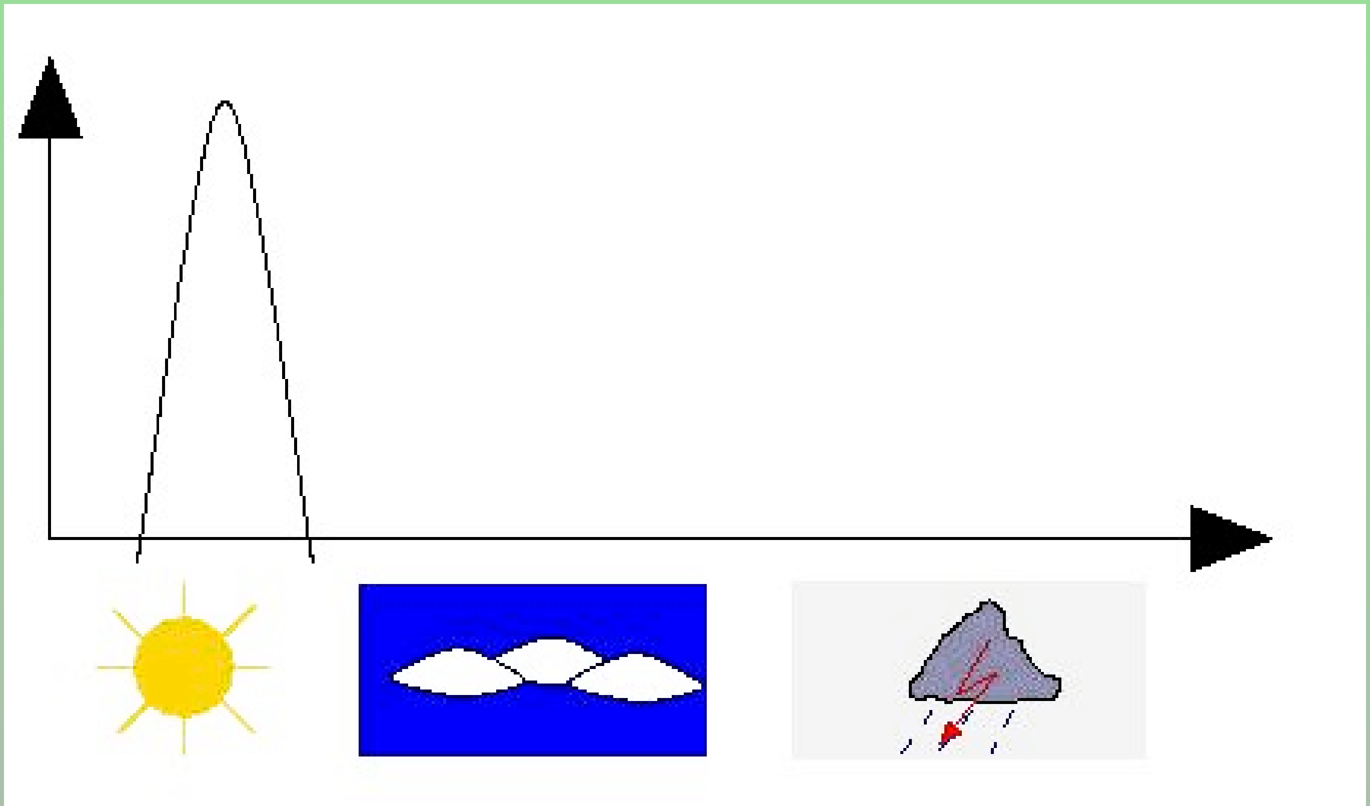


Description of information (a)

(data)

Description of information (a)

(probability)

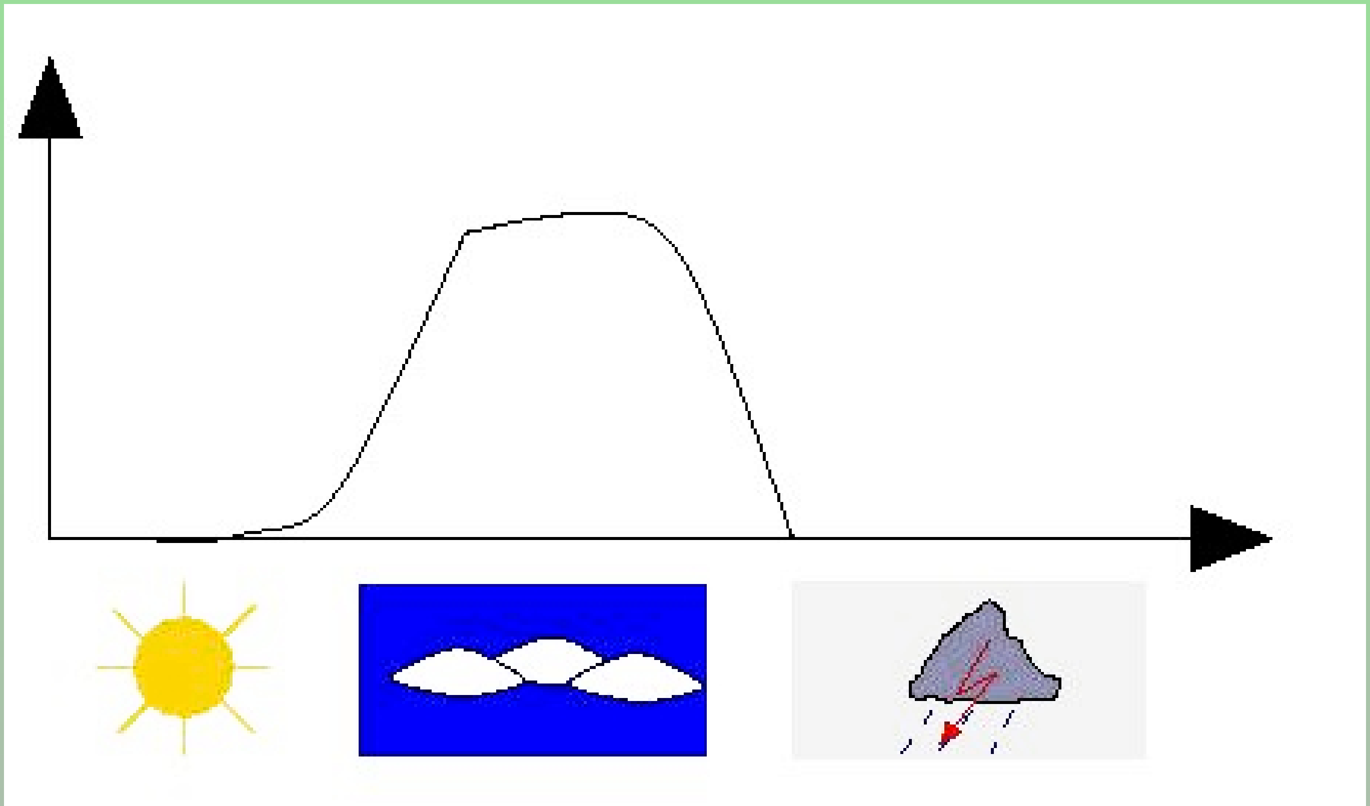


Description of information (b)

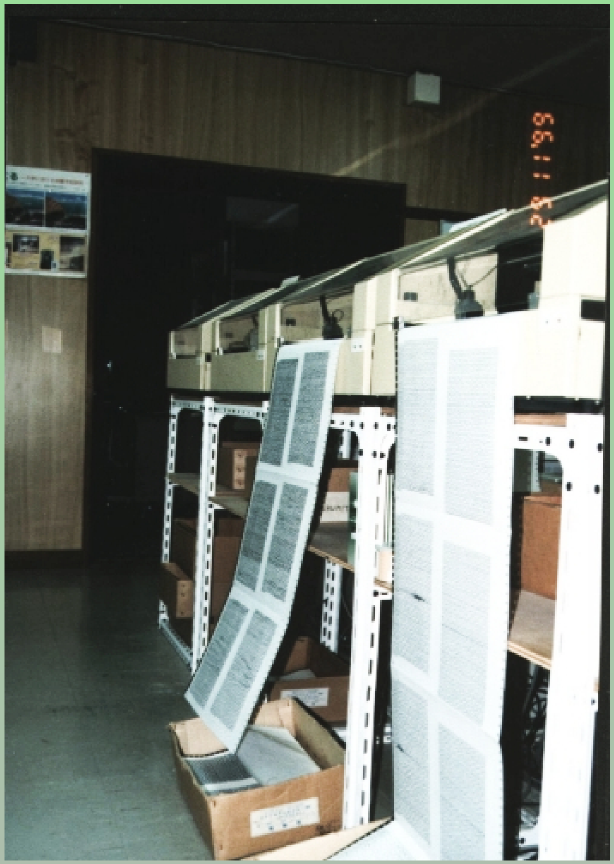
(data)

Description of information (b)

(probability)

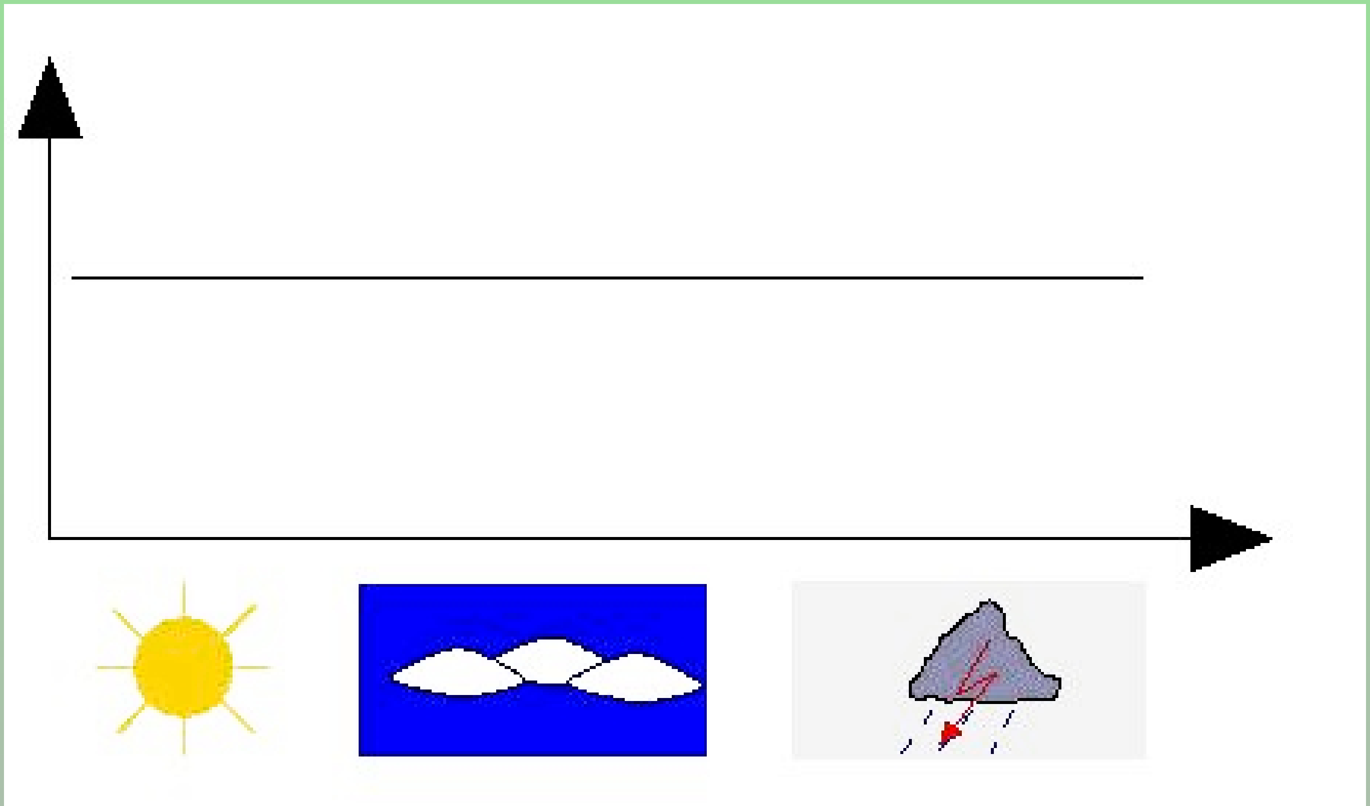


Description of information (b)
(data)

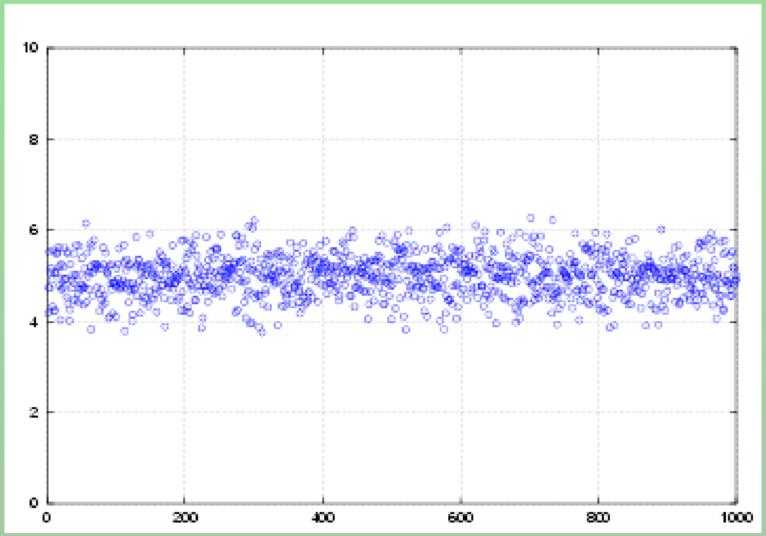


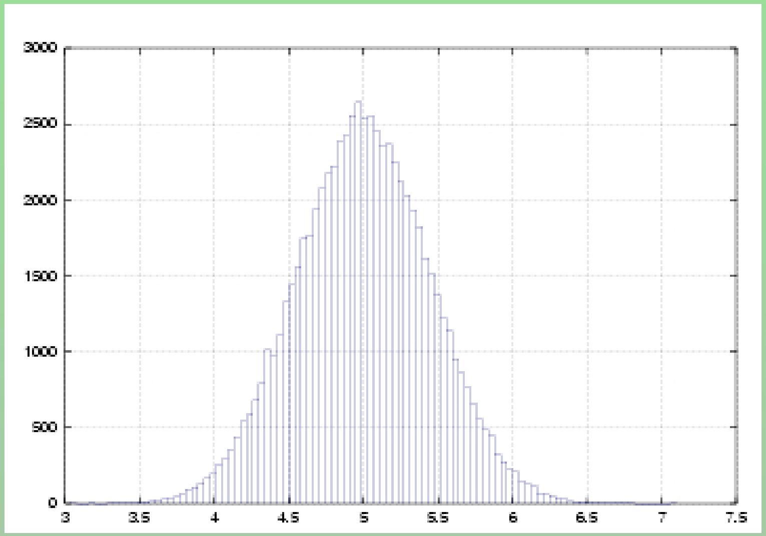
Description of information (b)

(probability)

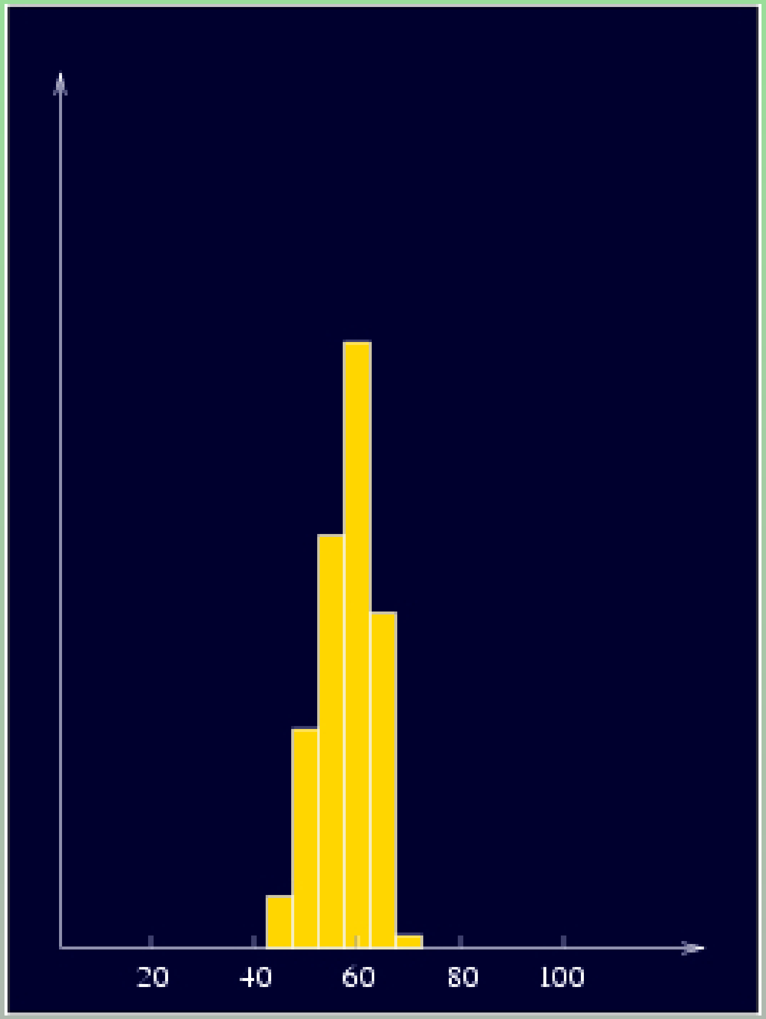


Frequentists interpretation of probability

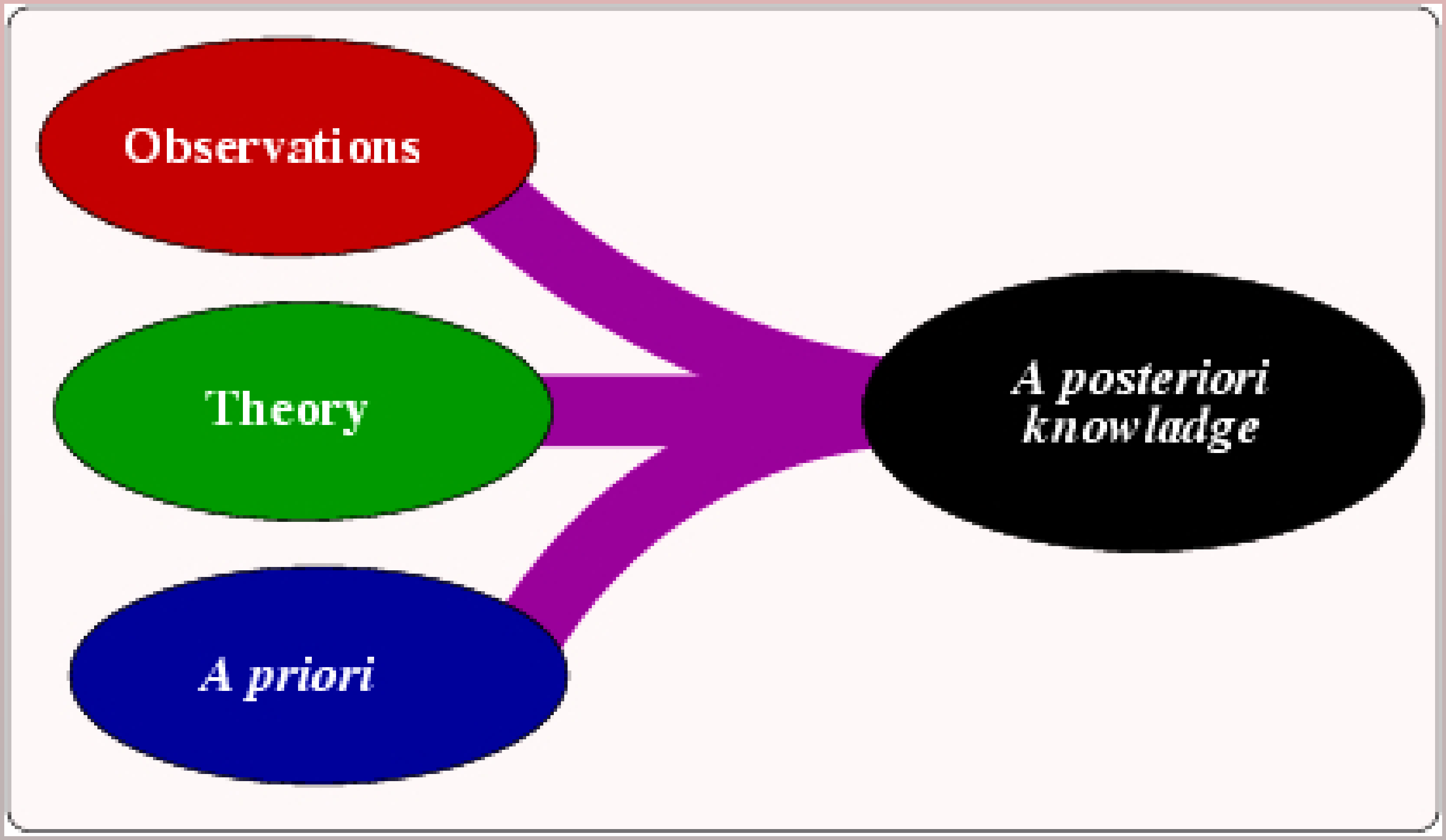




Bayesian interpretation of probability



Joining Information
(Inference)



Experimental information

Experimental uncertainties:

$$d^{obs} = d^{true} + \epsilon$$

ϵ: random $\Longrightarrow p(\epsilon)$

$$P([\epsilon - \delta/2, \, \epsilon + \delta/2])$$

$$\mathbf{P(d = d^{true}) = p_{\epsilon}(d - d^{obs})}$$

But how to find $p(\epsilon)$???

Experimental information

Experimental uncertainties statistic:

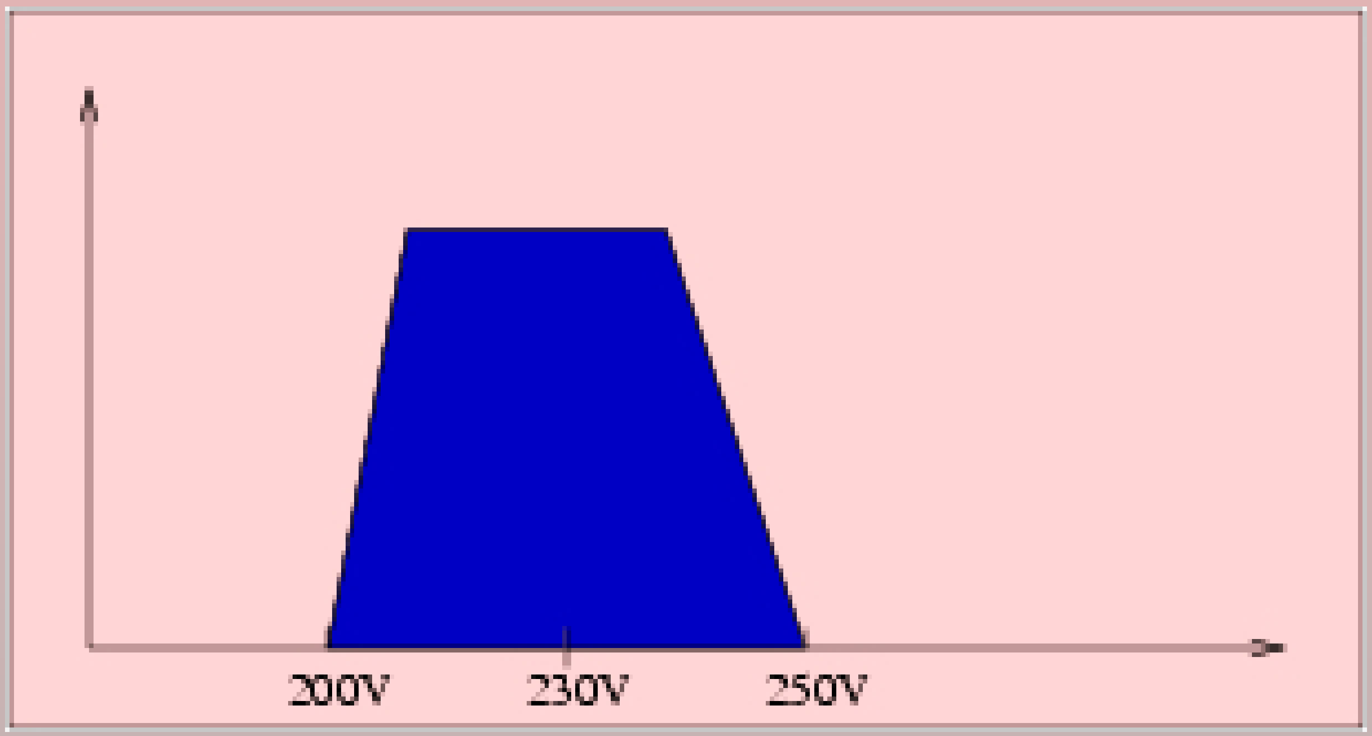
- postulate:

$$p(\epsilon) = k\, e^{-\frac{\epsilon^2}{2\sigma^2}}$$

- estimated by repeating measurement

$$P([x - \delta/2,\, x + \delta/2]) = \frac{N_i}{N}$$

A priori information



Theoretical information

Theoretical information: correlation

$$d = G(m) \Longrightarrow p(d = G(m)|m)$$

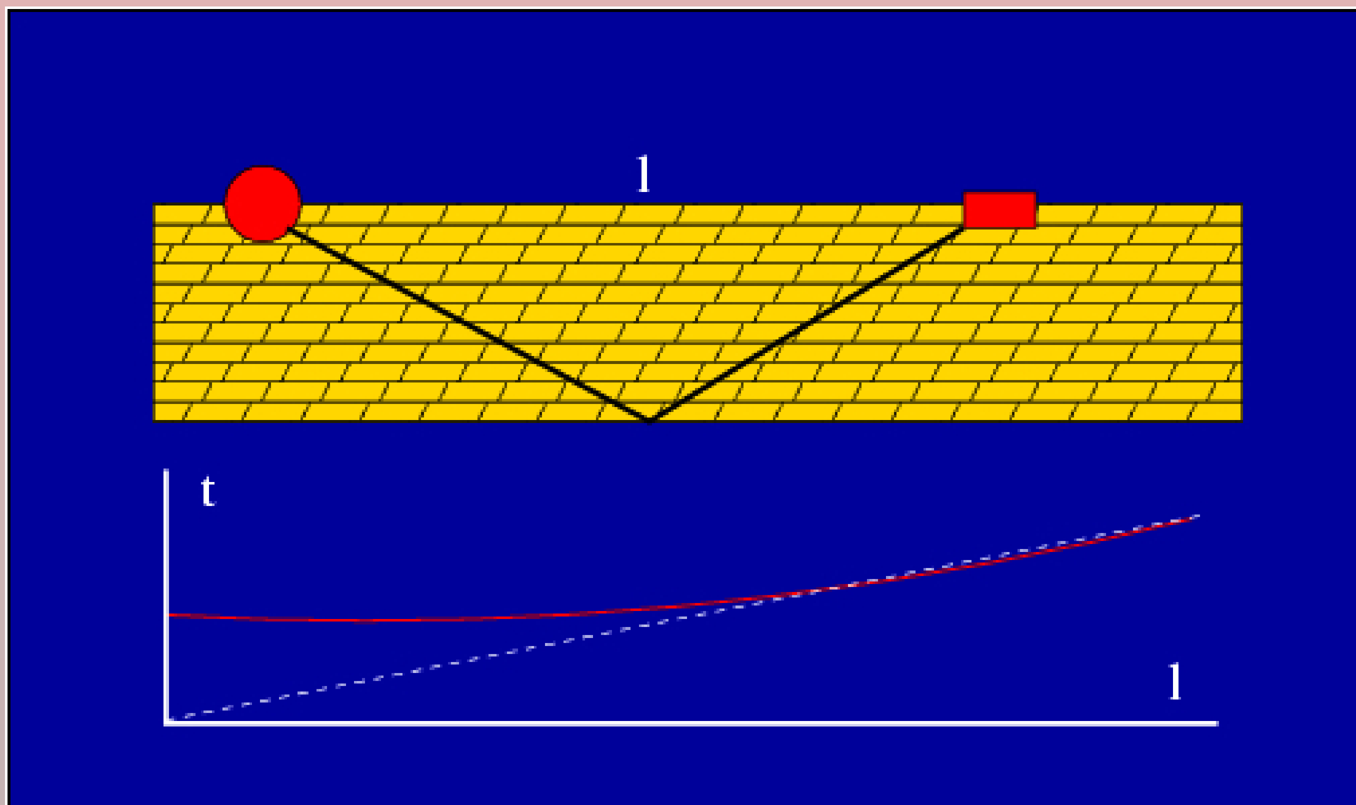
G may be known approximatelly:

$$G(m) = G_o \cdot m + G_1 \times m^2 + \dots$$

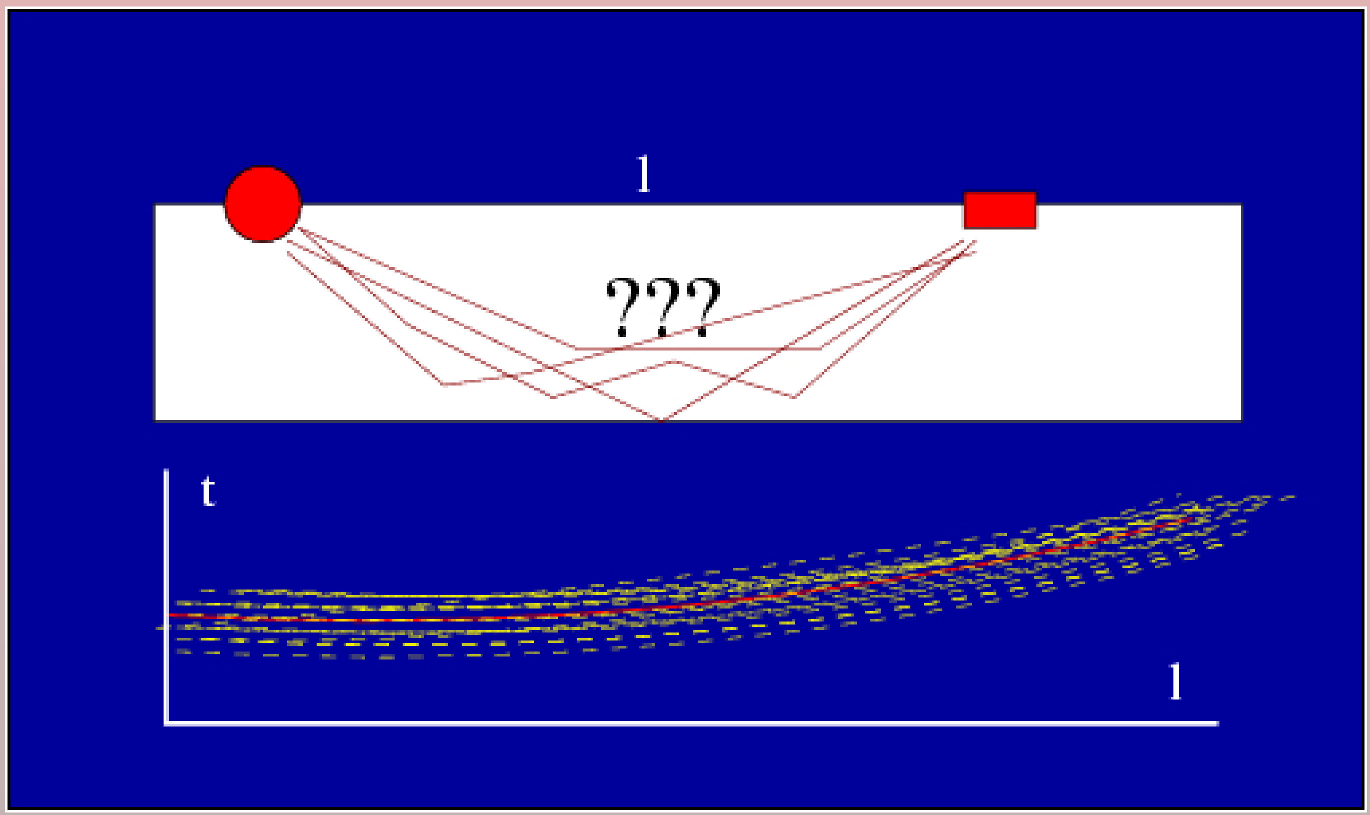
Exact theory:

$$p(d|m) = \delta(d - G(m))$$

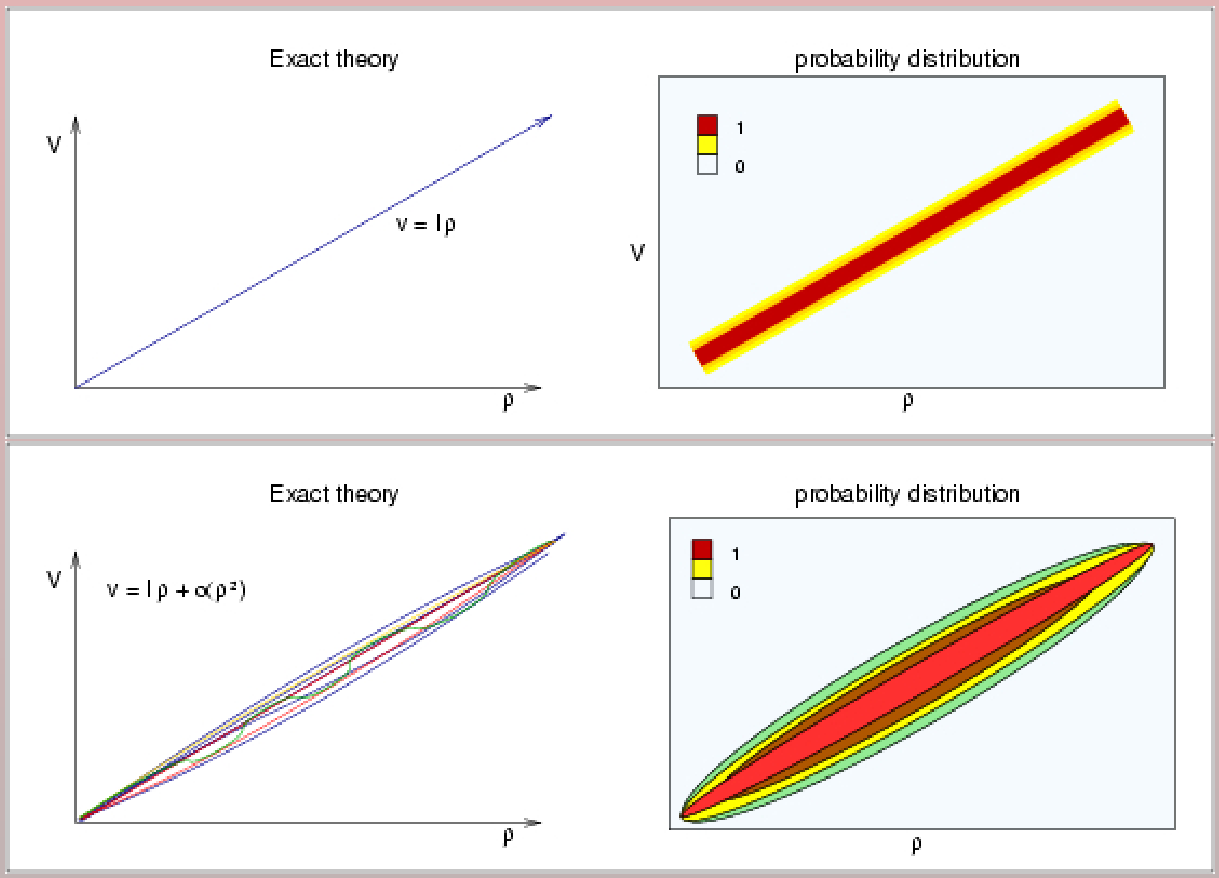
Theoretical information



Theoretical information



Theoretical information



Multi-dimensional pdf

$$p(x, y)$$

- marginal pdf $p_x(x) = \int_Y p(x, y) dy$
- marginal pdf $p_y(y) = \int_X p(x, y) dx$
- conditional pdf: $p_{x|y}(x|y) = p(x, y)/p_y(y)$
- conditional pdf: $p_{y|x}(y|x) = p(x, y)/p_x(x)$

Multi-dimensional pdf

$$p(x,y) = p_{x|y}(x|y)p_y(y)$$

$$p(x,y) = p_{y|x}(y|x)p_x(x)$$

$$p_{x|y}(x|y)p_y(y) = p_{y|x}(y|x)p_x(x)$$

Multi-dimensional pdf

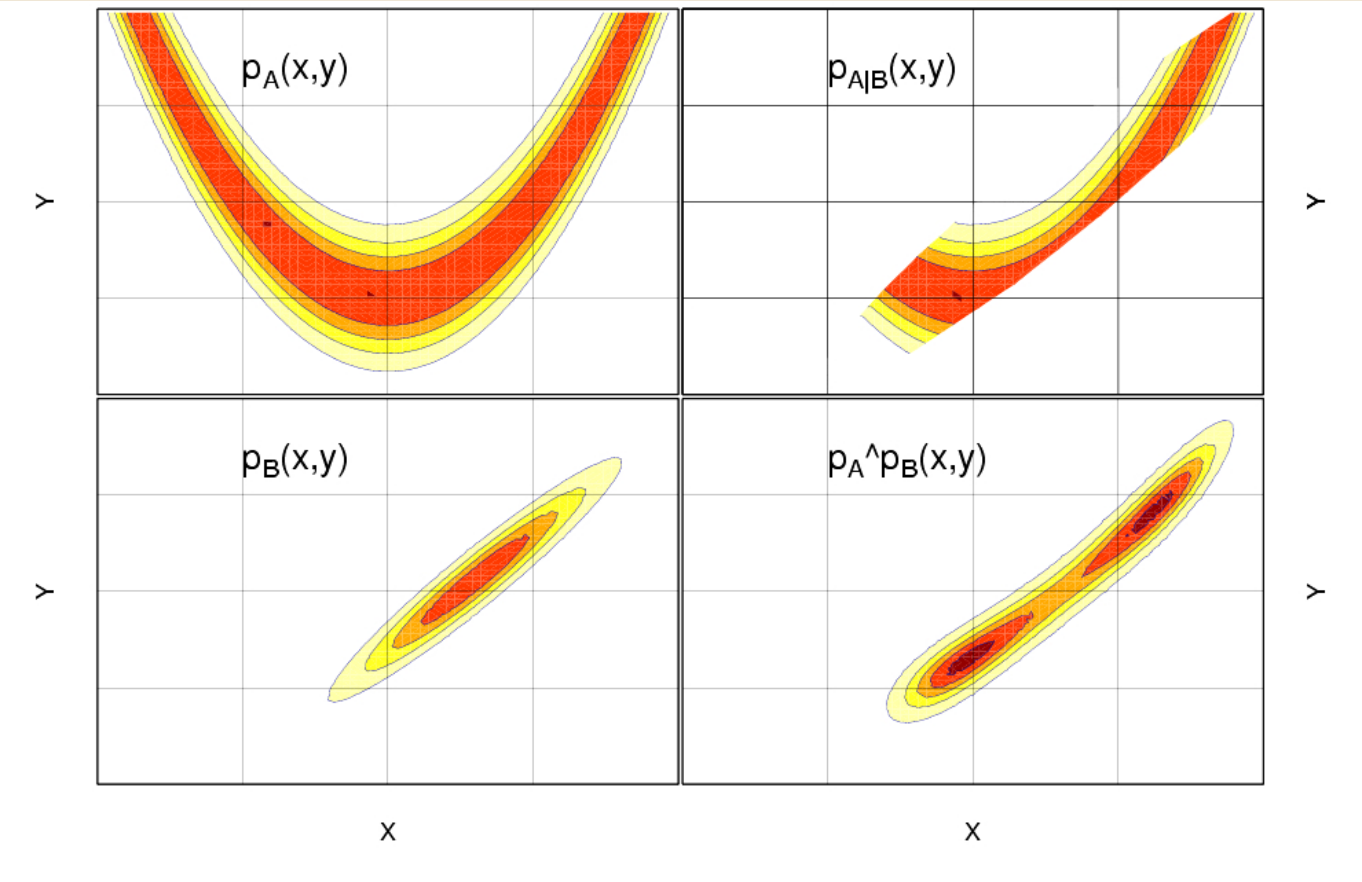
$$p_{\mathbf{m}}(\mathbf{m}|\mathbf{d}) = \frac{p_{d|m}(\mathbf{d}|\mathbf{m})p_m(\mathbf{m})}{p_d(\mathbf{d})}$$

Joining information

1. observation: $p(d)$
2. theory: $q(m,d)$
3. *a priori* $f(m,d)$

$$\mathbf{p} \wedge \mathbf{q}(\mathbf{x}) = \frac{\mathbf{p}(\mathbf{x})\,\mathbf{q}(\mathbf{x})}{\mu(\mathbf{x})}$$

Conditional vs. joint pdf



A posteriori pdf

A posteriori pdf:

$$\sigma(m,d) \approx p(d) \cdot q(m,d) \cdot f_M(m) f_D(m)$$

describes all information on d, m.

$$\sigma_m(m) = \int_D \sigma(m,d) dD$$

$$\sigma_d(d) = \int_M \sigma(m,d) dM$$

A posteriori pdf

$$\sigma_m(m) = f_M(m) \cdot L(m, d^{obs})$$

A posteriori pdf

A posteriori pdf $\sigma(m, d)$:

- always exists
- is unique
- describes all information
- is the solution of an inverse problem

Exact theory

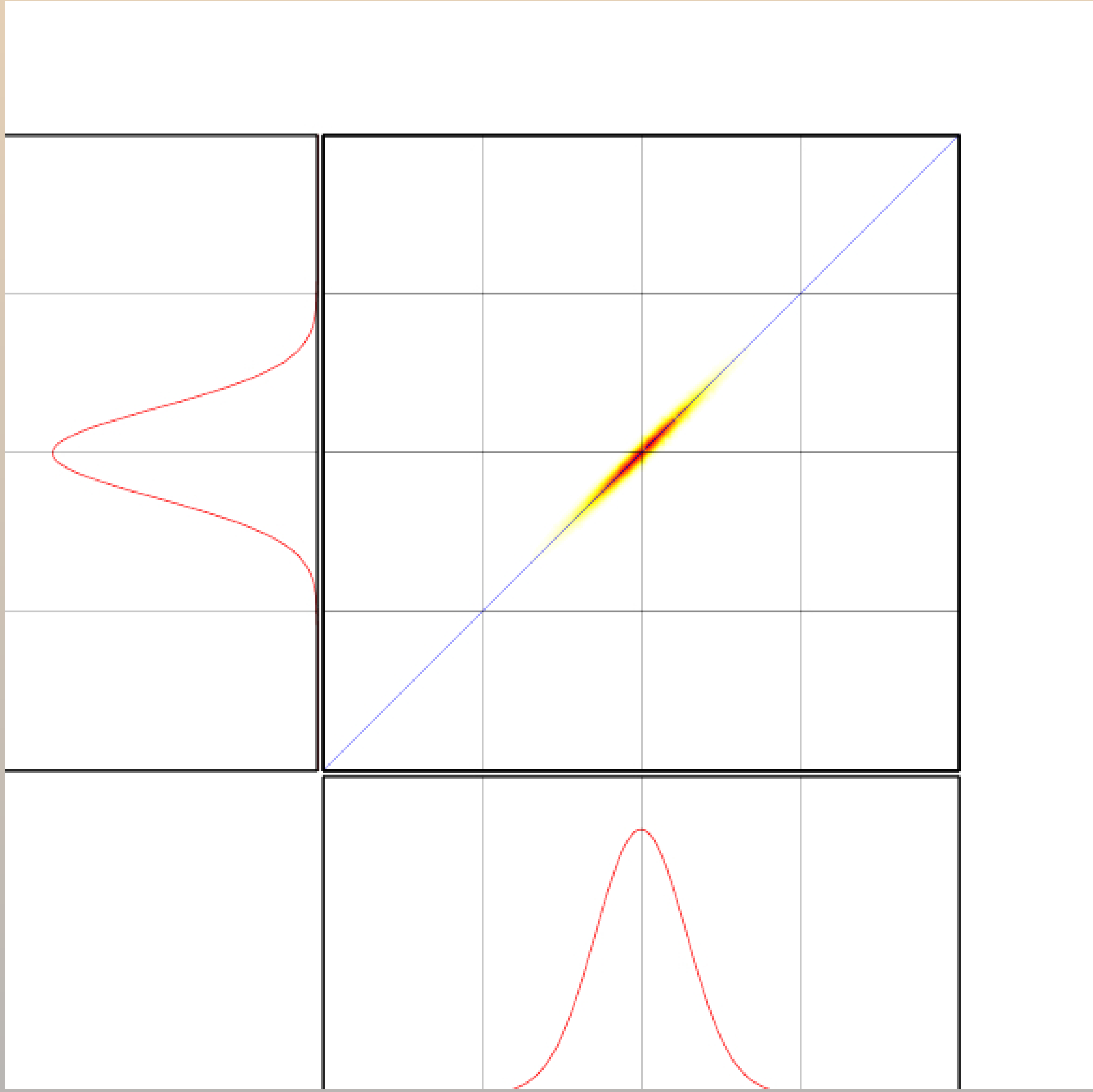
$$\sigma(m) = f_M(m) \int_D p(d) q(m,d) \mathrm{d}d$$

$$q(d,m) = \delta(d - G(m))$$

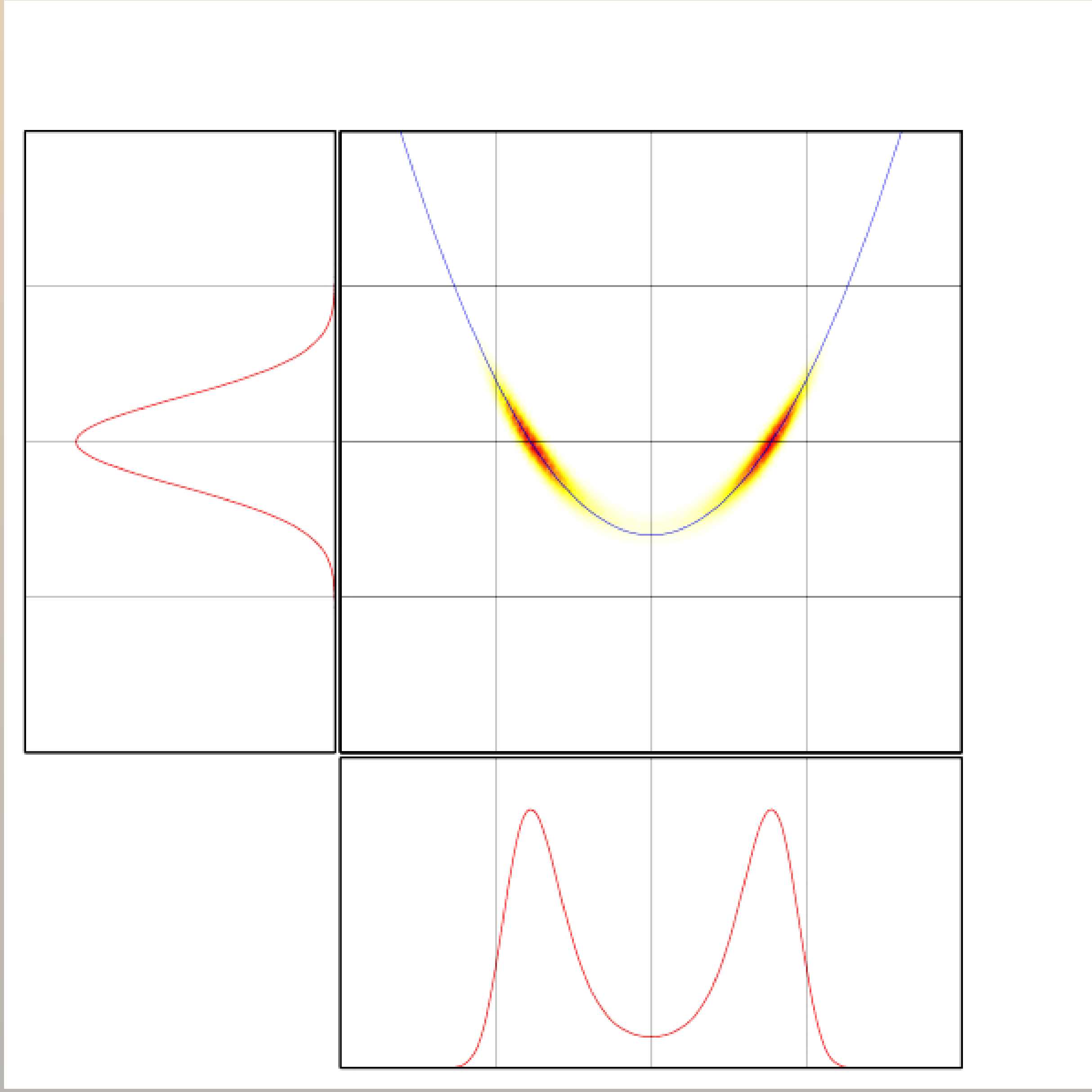
$$\sigma(m) = f_M(m)\, p(d - G(m))$$

$$\sigma(m) = f_M(m) \, \exp\left(-(d - G(m))^T \mathbf{C}^{-1} (d - G(m))\right)$$

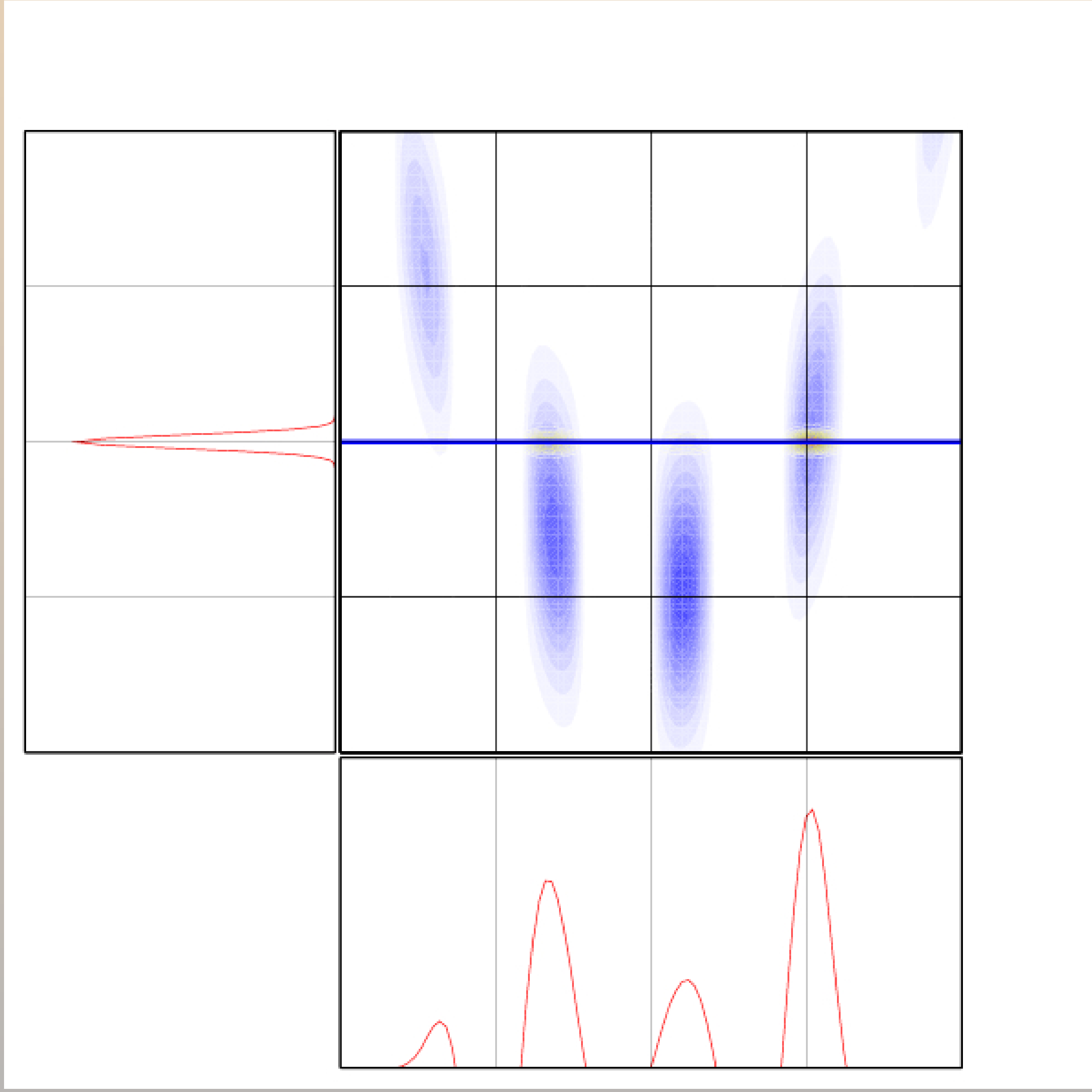
Exact theory: linear case



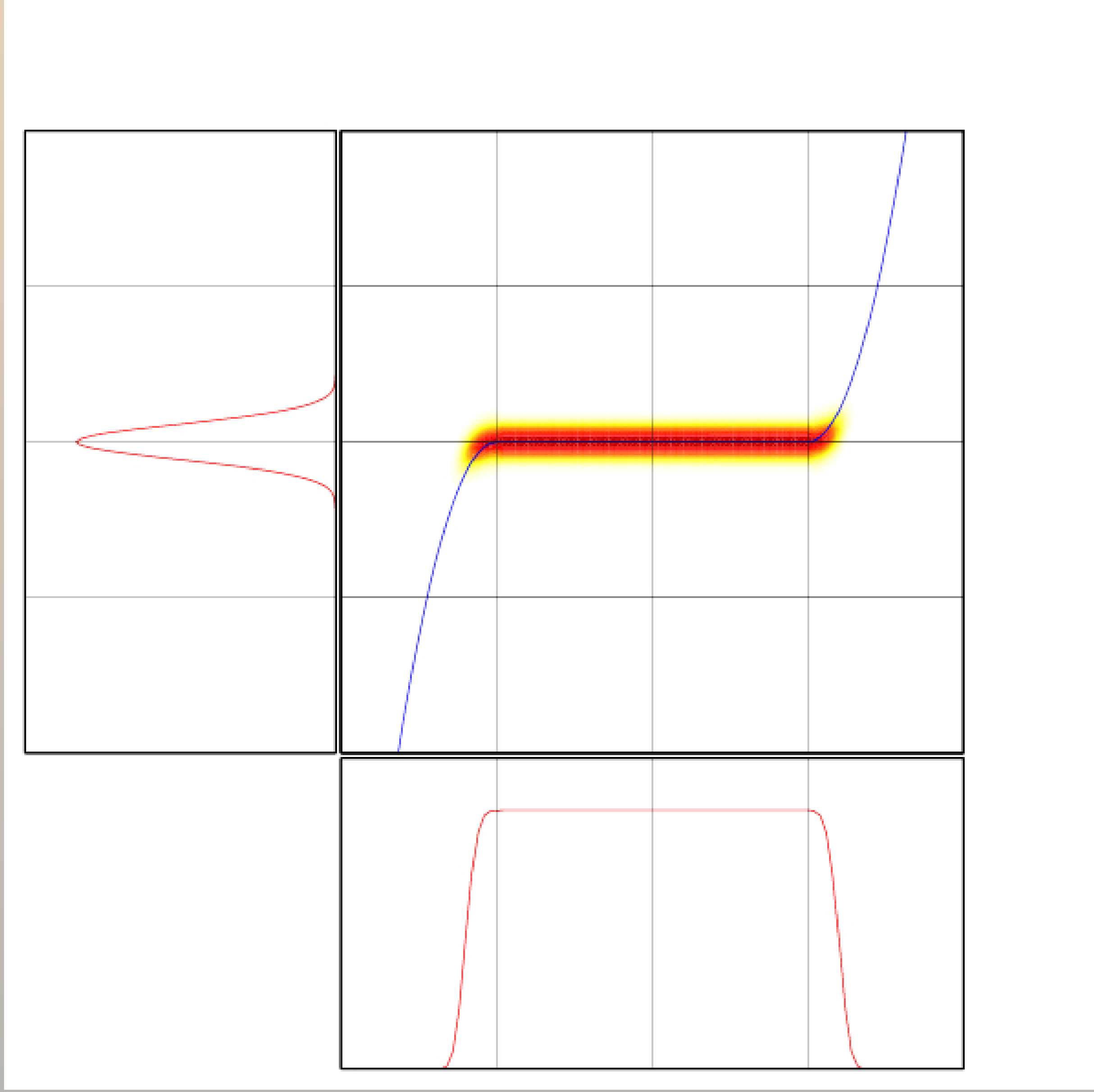
Exact theory: non-linear case



“Exact” data: uncertain theory



Non-resolved parameters



Summary

- Parameter estimation
- Error analysis
- Resolution analysis
- Experimental planning
- Model discrimination
- Others (non-parametric) inference

Inversion Algorithms- Summary

Method	Advantages	Limitations
Algebraic (LSQR)	- Simplicity	- Only linear problems
$\mathbf{m}^{ml} = (\mathbf{G}^T\mathbf{G} + \gamma\mathbf{I})^{-1}\mathbf{G}^T \cdot \mathbf{d}^{obs}$	- Large scale problems	- Lack of robustness
Optimization	- Simplicity	- Difficult error estimation
$\|\mathbf{G}(\mathbf{m}) - \mathbf{d}^{obs}\| + \lambda\|\mathbf{m}) - \mathbf{m}^a\| = \min$	- Fully nonlinear	
Bayesian	- Fully nonlinear	- More complex theory
$\sigma(\mathbf{m}) = f(\mathbf{m})L(\mathbf{m}, \mathbf{d}^{obs})$	- Full error handling	- Requires efficient sampler