History/Overview of Inverse problems
Legacy of A. Tarantola

Jean-Paul Montagner
Institut de Physique du Globe
Université Paris Cité

A.T.

Anisotropic **T**omography -
Anelastic **T**omography -
Adjoint **T**omography

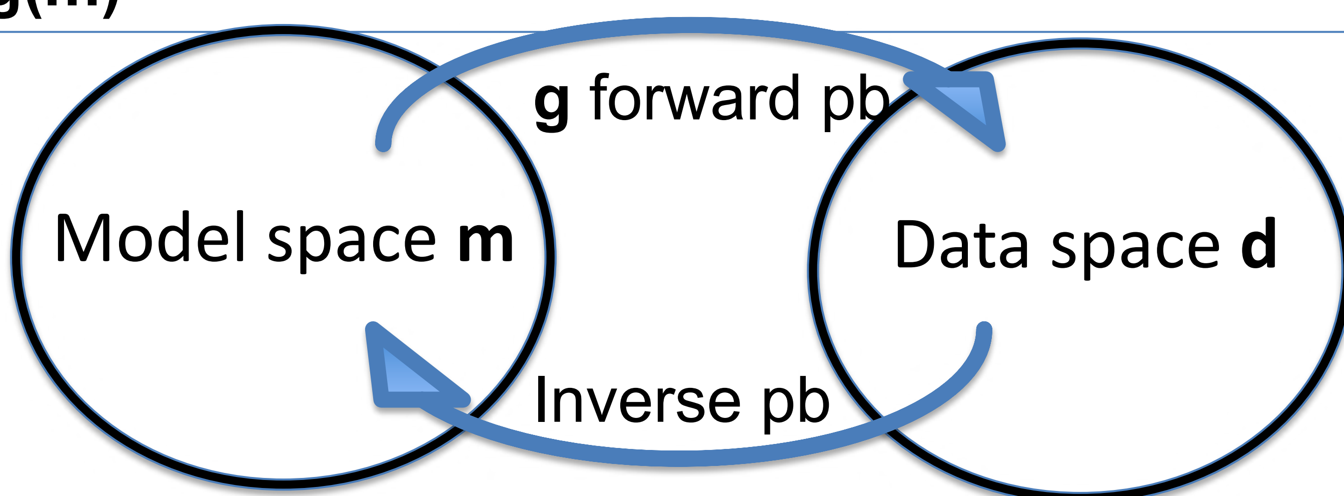
Probabilistic Approach- Bayesian inference

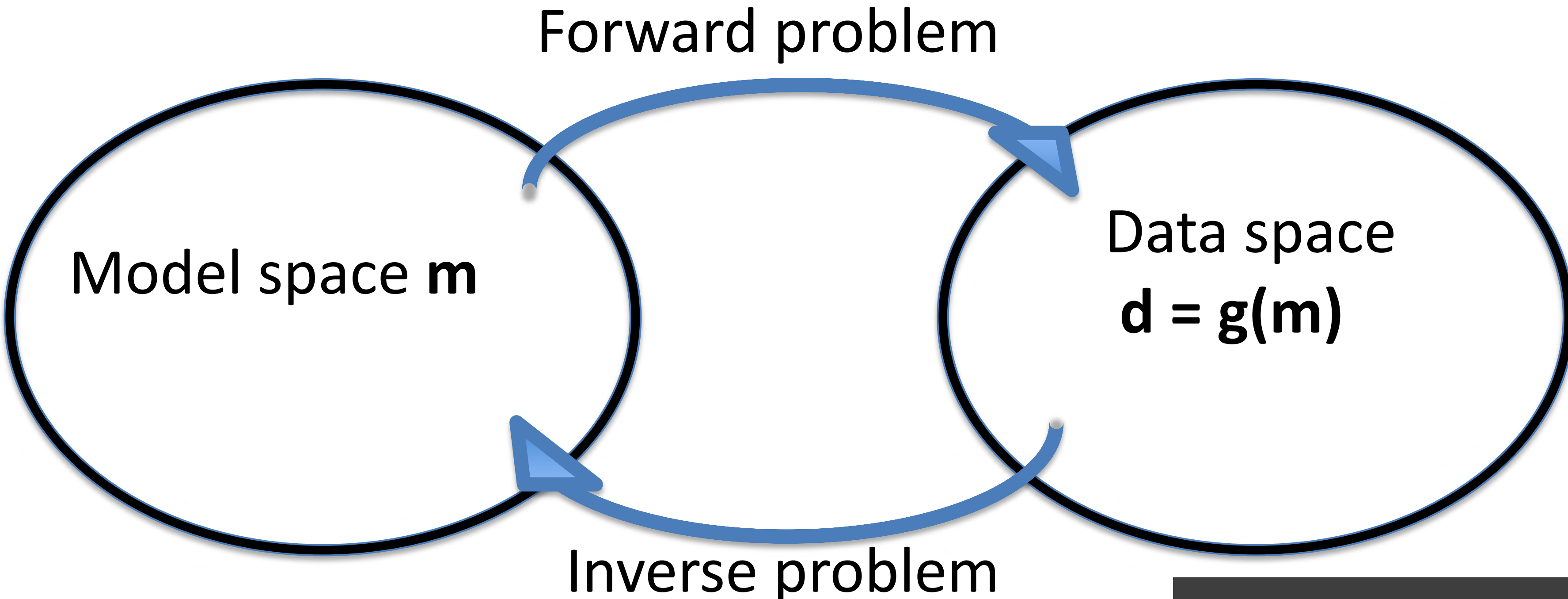
Neural networks, Full waveform inversion

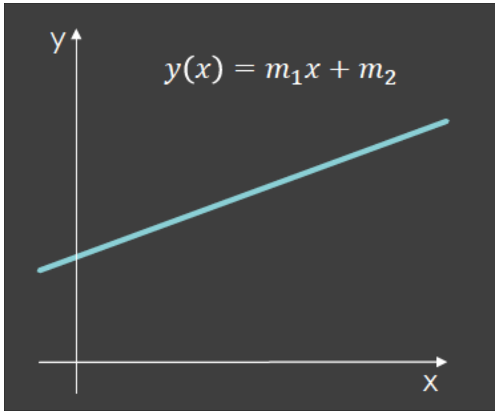
What an Inverse Problem?

- Process of calculating causal model parameters **m** from observational data **d** that produce them.
- Inverse of the forward problem, starting with the causes and then calculating the effects.

- Forward problem associated with a theory but not necessarily (explicit or implicit).
- There is a functional **g** relating **d** and **m**:
  d = g(m)







1- Linear problem: $d = Gm$

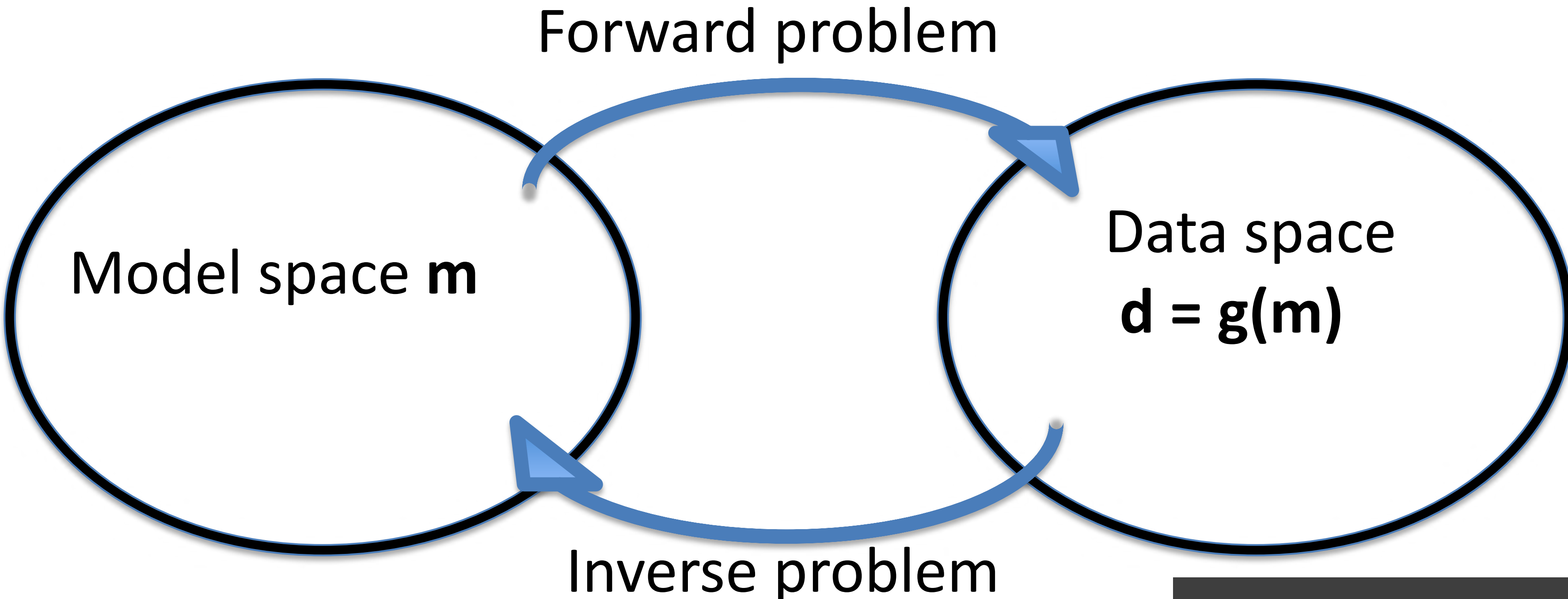
2- Slightly non-linear problem: $d_0 = g(m_0)$

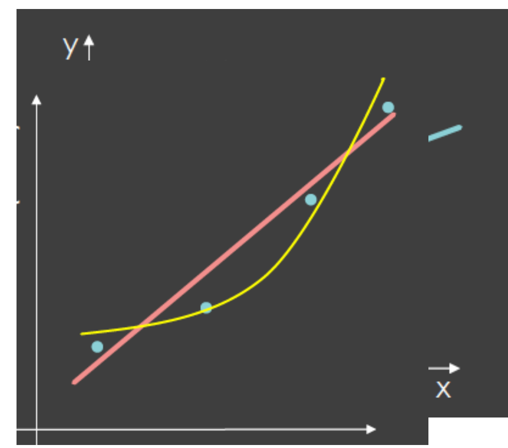
First order perturbation theory:

$d - d_0 = G(m - m_0) + O((m - m_0)^2)$ + Iterations

G sensitivity kernels

3- Non-linear problems: statistical, probabilistic approach (exploration of the whole parameter space when possible)





1- Linear problem: $d = Gm$

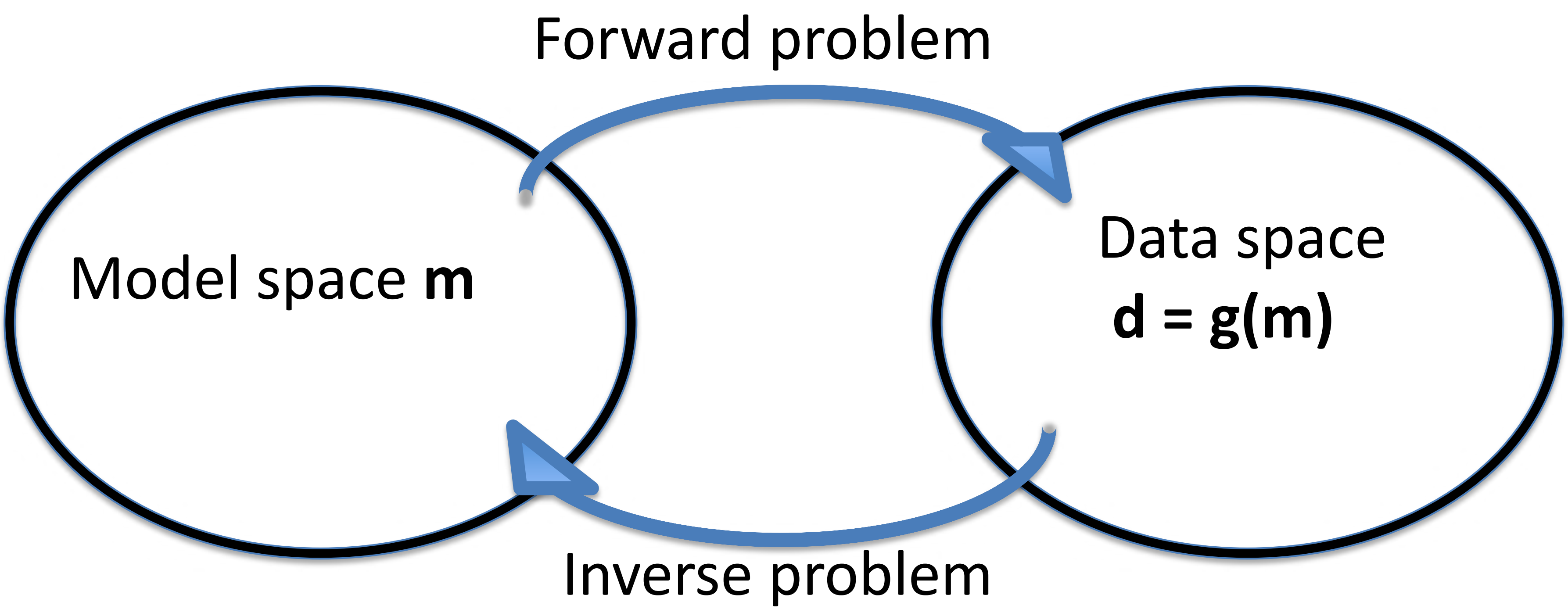
2- Slightly non-linear problem: $d_0 = g(m_0)$

First order perturbation theory:

$$d - d_0 = G(m - m_0) + O((m - m_0)^2) + \text{Iterations}$$

G sensitivity kernels

3- Non-linear problems: statistical, probabilistic approach (exploration of the whole parameter space when possible)



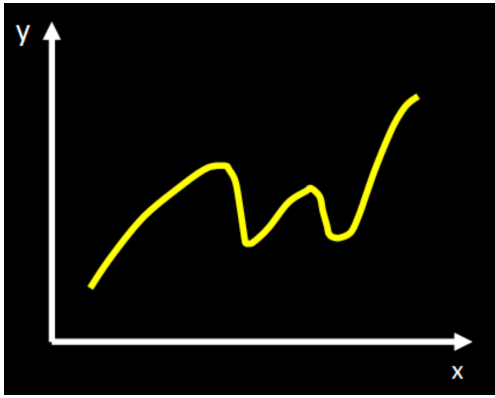
1- Linear problem: $d = Gm$

2- Slightly non-linear problem: $d_0 = g(m_0)$

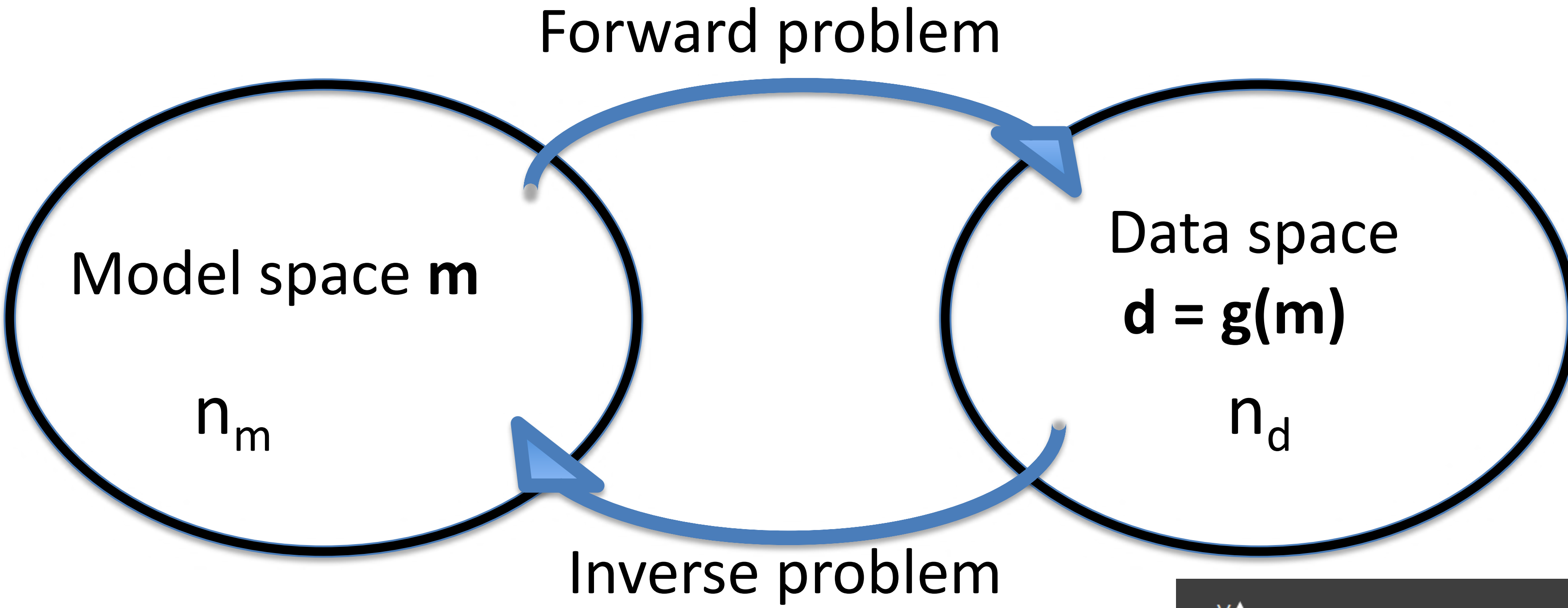
First order perturbation theory:

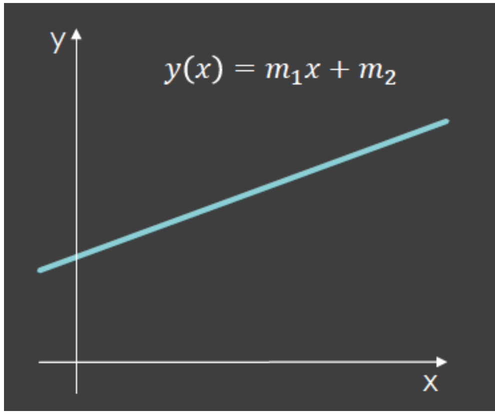
$$d - d_0 = G(m - m_0) + O((m - m_0)^2) + \text{Iterations}$$

G sensitivity kernels



3- Non-linear problems: statistical, probabilistic approach (exploration of the whole parameter space when possible)





1- Linear problem: $d = G\,m$

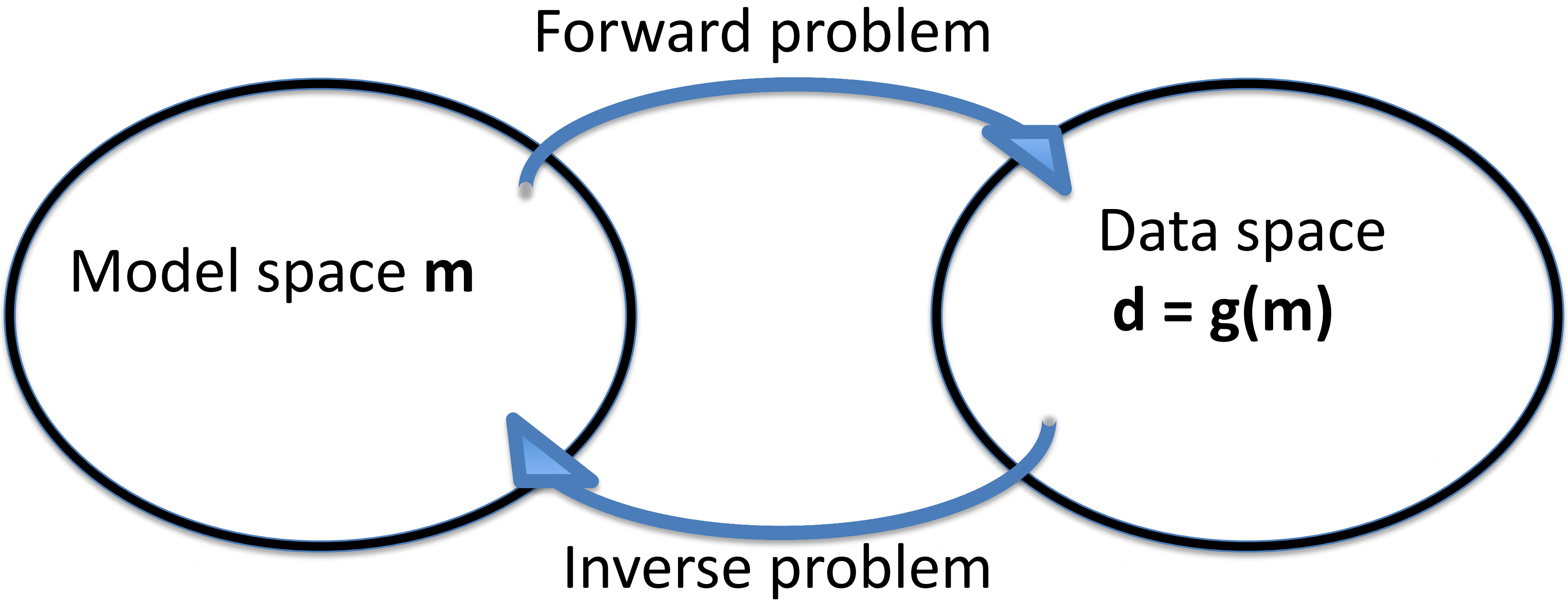
G non-square matrix (n_d x n_m),

n_d number of data, n_m number of model parameters

Least squares Solution:

$$G^T d = G^T G\, m$$

$$m = \underbrace{(G^T G)^{-1}}_{\text{Might be huge}} G^T d$$



2- Slightly non-linear problem: $d_0 = g(m_0)$ $d = g(m)$

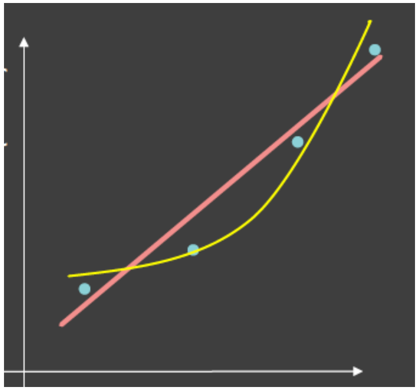
First order perturbation theory:

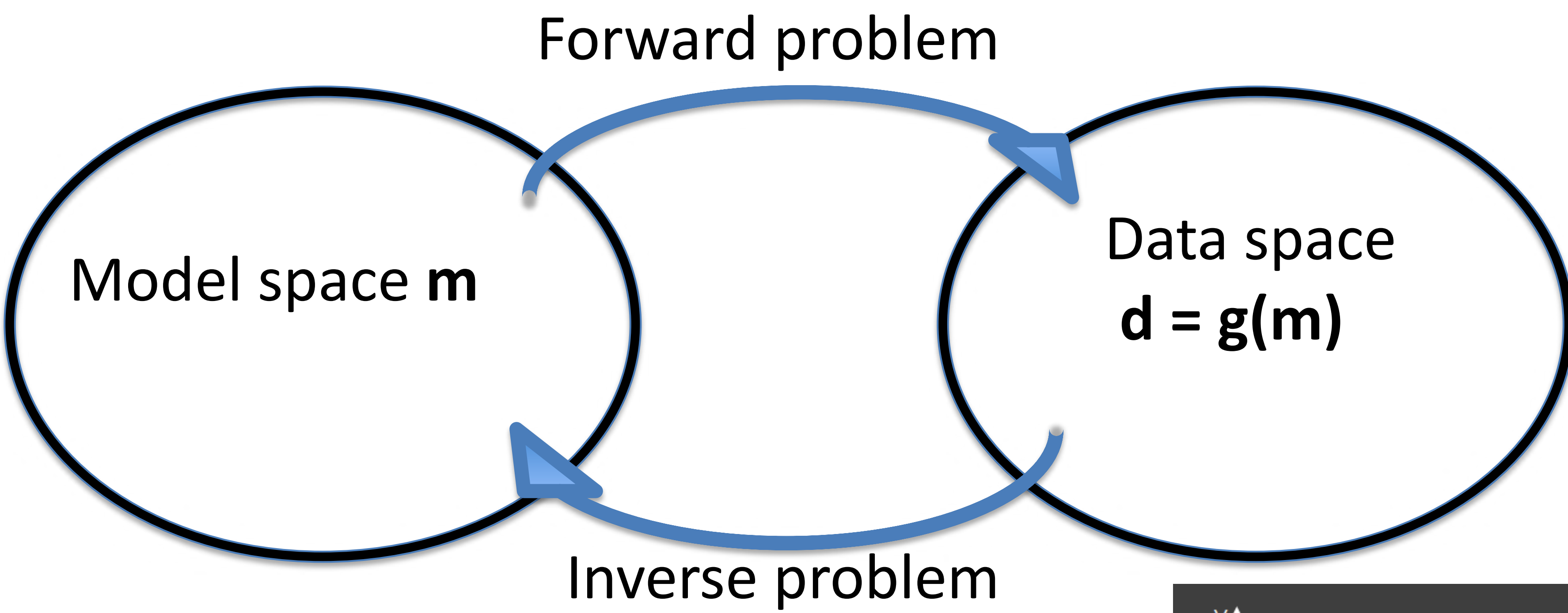
$$d-d_0 = G(m-m_0) + O((m-m_0)^2) + \text{iterations}$$

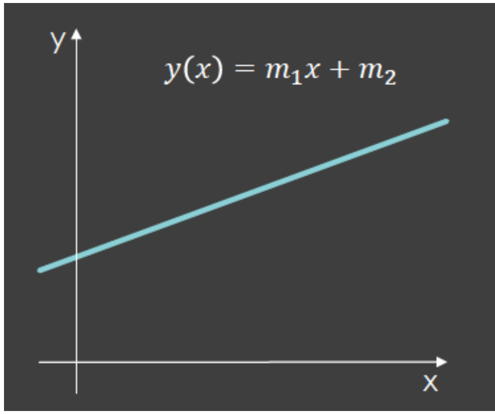
G sensitivity kernels

$$m-m_0 = (G^TG)^{-1}G^T(d-d_0) + \text{iterations}$$

Classical approach:
Backus & Gilbert, 1967, 1968, …
Franklin, 1970, Jackson, ….
Wiggins, 1972, ..







1- Linear problem: $d = G\,m$

G non-square matrix,

n_d number of data, n_m number of model parameters

Least squares Solution: $m = (G^T G)^{-1} G^T d$

2- Slightly non-linear problem: $d_0 = g(m_0)$ $d = g(m)$

$m - m_0 = (G^T G)^{-1} G^T (d - d_0) +$ iterations

Quality of the fit?

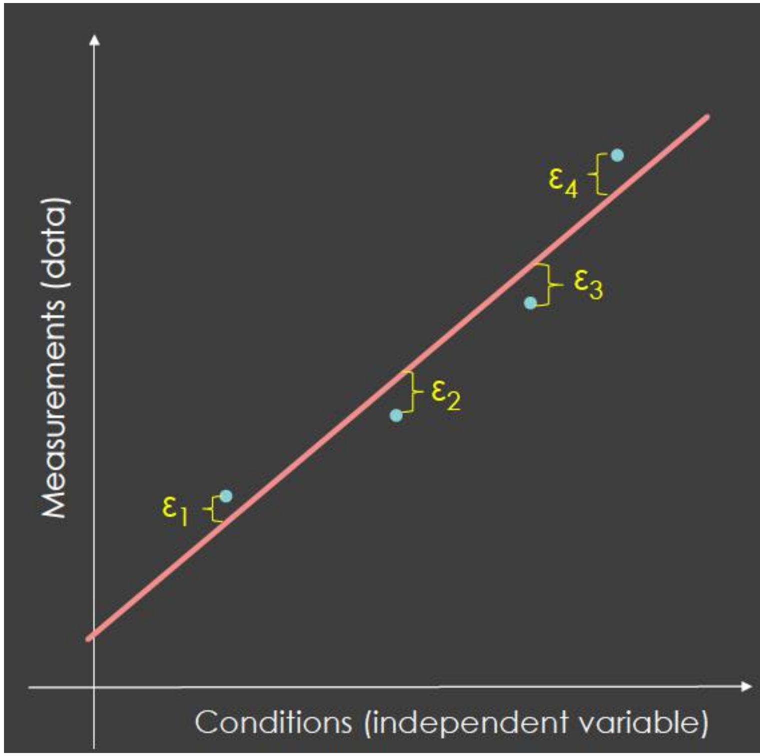
Cost function

Cost Function S (or objective function)

Minimization of all $\varepsilon_j = d_j - g_j(m)$ for getting the best fit

$$S(m) = \sum_{j=1}^{Nd} | d_j - g_j(m) |$$ L1-norm

$$S(m) = \sum_{j=1}^{Nd} (d_j - g_j(m))^2$$ L2-norm

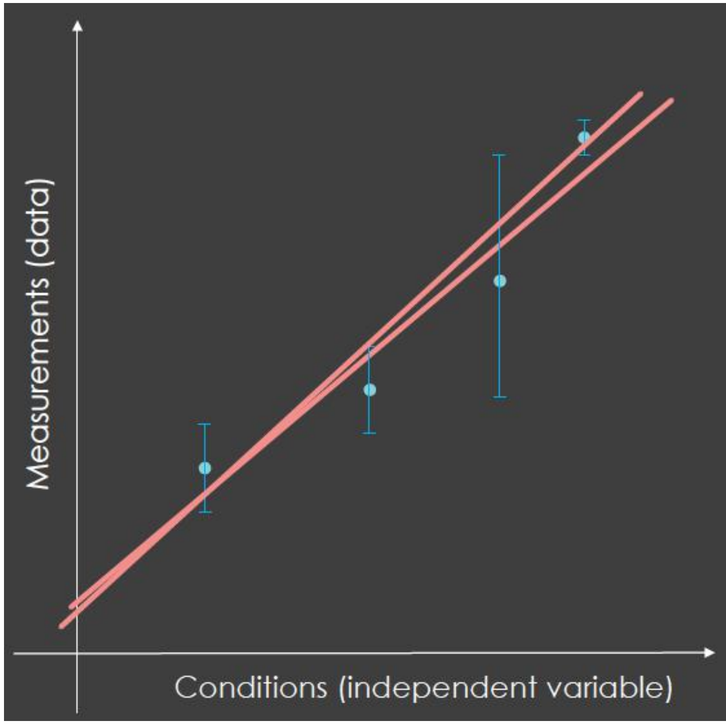


What about data and model uncertainty?

(lecture of Alison Malcolm)

Data uncertainty σ (error bars)

Errors are not necessarily the same for all measurements – quantitative estimate Is not trivial



Cost Function S (or objective function)

$$S(m) = \sum_{j=1}^{Nd} \frac{1}{\sigma} |d_j - g_j(m)| \quad \text{weighted L1-norm}$$

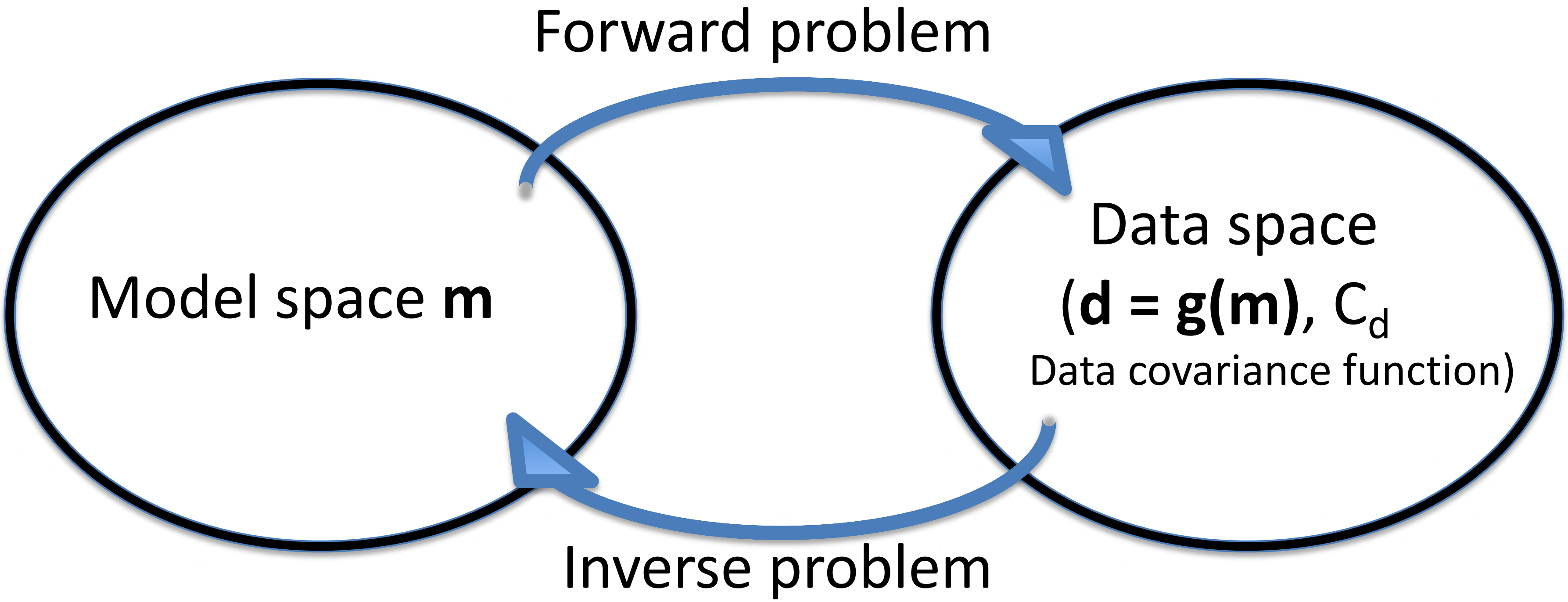
$$S(m) = \sum_{j=1}^{Nd} \frac{1}{\sigma^2} (d_j - g_j(m))^2 \quad \text{weighted L2-norm}$$

Probability distributions – Bayesian Approach

(see Klaus Mosegaard, Andreas Fichtner lectures)

Tarantola & Valette, 1982; 1984, ….2005

Jackson & Matsu'ura, 1985, …

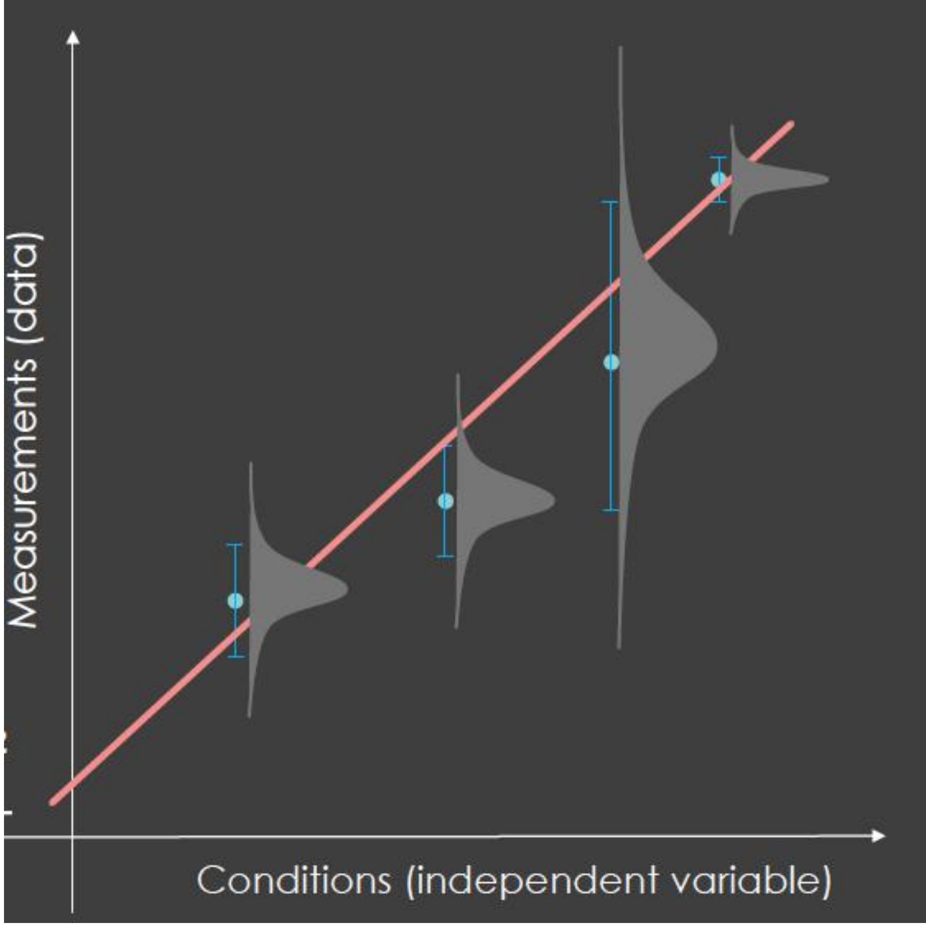


Cost Function S (or objective function)

$$S(m) = \sum_{j=1}^{Nd} \frac{1}{\sigma} | d_j - g_j(m) |$$ weighted L1-norm

$$S(m) = \sum_{j=1}^{Nd} (d - g(m))^T C_d^{-1} (d - g(m))$$ .2-norm

Normally distributed errors:
Gaussian distribution



Cost Function L2 norm

C_d covariance function of data

$S(m) = (d - g(m))^T C_d^{-1} (d - g(m))$
weighted L2-norm

Minimisation of S: $\partial S / \partial m = 0$

For a linear problem:

$m_{est} = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} d$ least square estimate of model parameters

$C_m^{post} = (G^T C_d^{-1} G)^{-1}$ *a posteriori* model covariance function

For a slightly nonlinear problem:

$m_{est} - m_0 = (G^T C_d^{-1} G)^{-1} G^T C_d^{-1} (d - d_0)$ + iterations

Normally distributed errors:
Gaussian distribution

Cost Function L2 norm

$S(m) = (d - g(m))^T C_d^{-1} (d - g(m))$
weighted L2-norm

Equivalence Data Space –Model Space

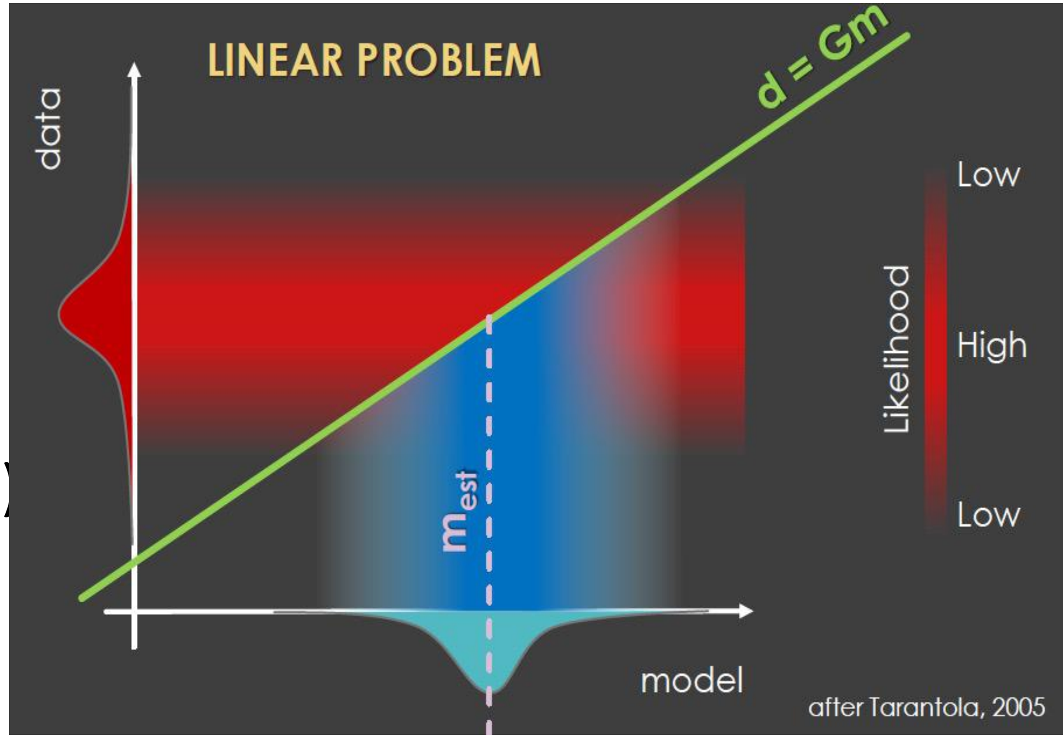
Regularization: C_m covariance function of model parameters
Use of a priori information (Bayesian inference)

$S(m) = (d-g(m))^T C_d^{-1}(d-g(m))$ **$+(m-m_0)^T C_m^{-1}(m-m_0)$**

Normally distributed errors:
Gaussian distribution

Cost Function L2 norm

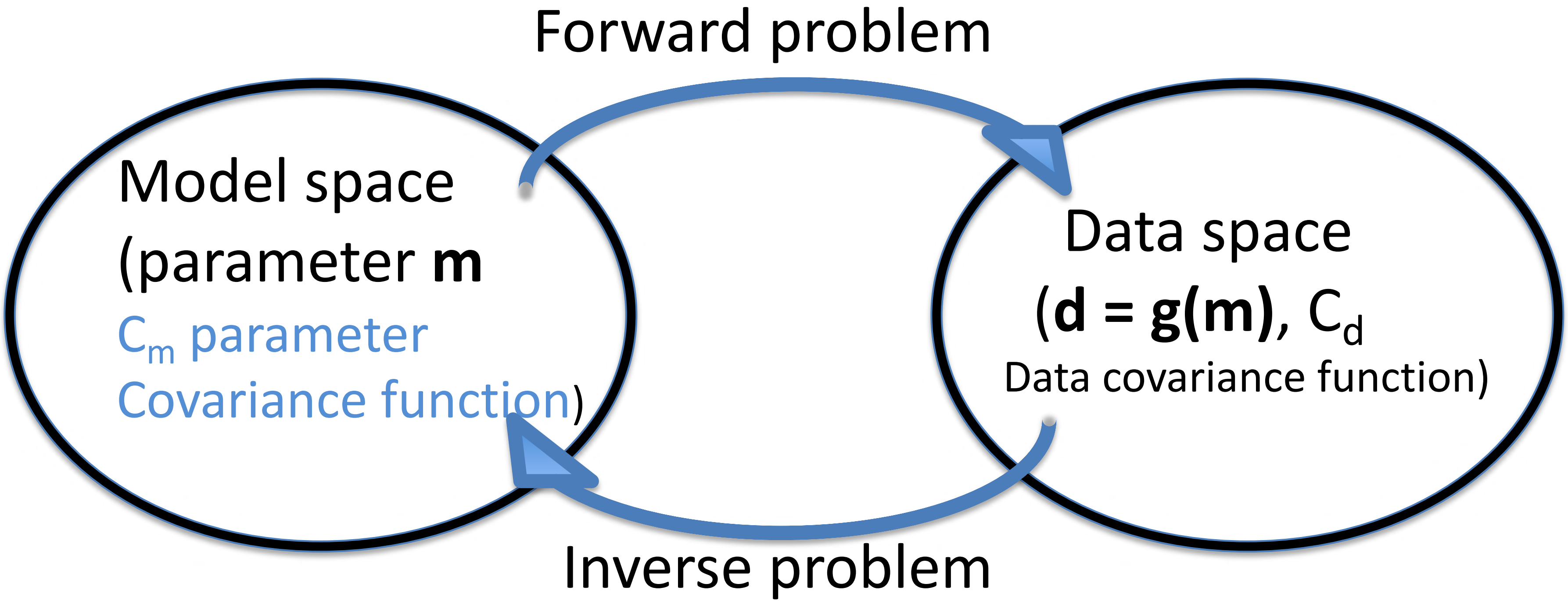
$S(m) = (d - g(m))^T C_d^{-1} (d - g(m))$
weighted L2-norm



Equivalence Data Space –Model Space

Regularization: C_m covariance function of model parameters
Use of a priori information (Bayesian inference)

$$S(m) = (d - g(m))^T C_d^{-1} (d - g(m)) + (m - m_0)^T C_m^{-1} (m - m_0)$$



Cost Function:

$S(m) = (d - g(m))^T C_d^{-1} (d - g(m))$ weighted L2-norm

Regularization:

$S(m) = (d - g(m))^T C_d^{-1} (d - g(m)) + (m - m_0)^T C_m^{-1} (m - m_0)$

Minimisation of S: $\partial S / \partial m = 0$

« Generalized Inverse problem »

Gaussian Statistics (least square criterion)

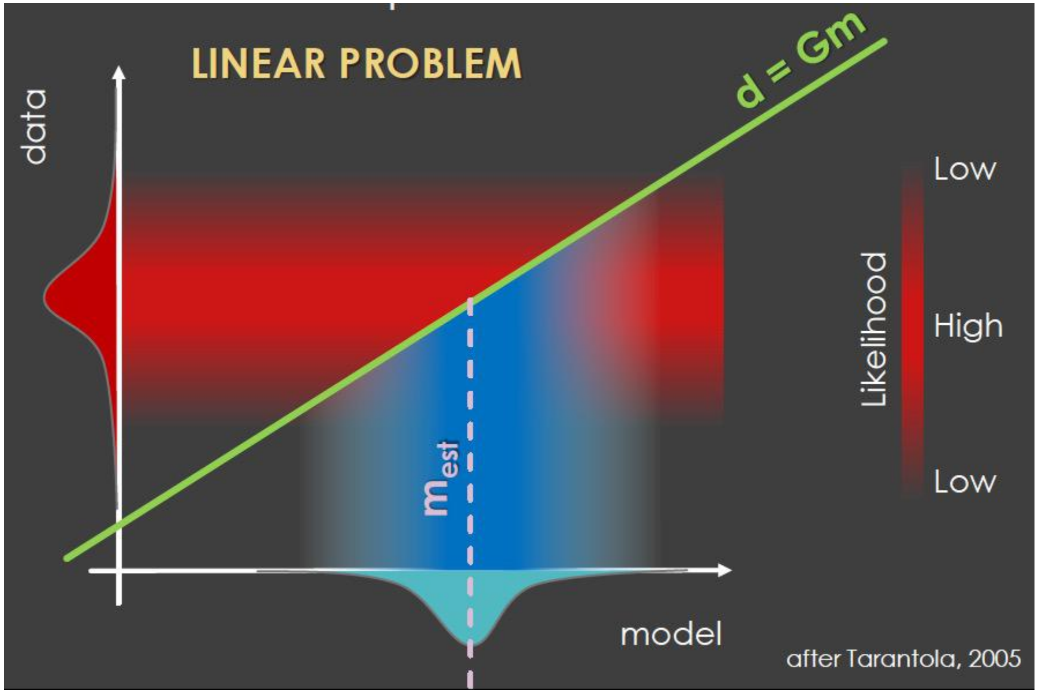
(Tarantola & Valette, Mosegaard & Tarantola Bayesian approach)

•Data space: measurements may have a Gaussian form:

$P_d(d) = k \exp(-1/2\ (d-d_0)^T\ C_d^{-1}\ (d-d_0))$

•Parameters of model space: the *a priori* information may have a Gaussian form:

$P_{apriori}(m) = k' \exp(-1/2\ (m-m_0)^T\ C_m^{-1}\ (m-m_0))$



« Generalized Inverse problem »

Gaussian Statistics (least square criterion)

(Tarantola & Valette, Mosegaard & Tarantola Bayesian approach)

•Data space: measurements may have a Gaussian form:

$P_d(d) = k \exp(-1/2\ (d-d_0)^T C_d^{-1} (d-d_0))$

•Parameters of model space: the *a priori* information may have a Gaussian form:

$P_{apriori}(m) = k' \exp(-1/2\ (m-m_0)^T C_m^{-1} (m-m_0))$

•**Bayes Theorem** under some hypotheses

$P_{post}(m) = 1/\nu\ P_{apriori}(m)\ P_d(d) = K \exp(-S(m))$

With $2\ S(m) = (m-m_0)^T C_m^{-1} (m-m_0)) + (d-d_0)^T C_d^{-1} (d-d_0))$

« Generalized Inverse problem »
Gaussian Statistics (least square criterion)
(Tarantola & Valette: Bayesian approach)

•Data space: measurements may have a Gaussian form:
$P_d(d) = k \exp(-1/2\ (d-d_0)^T C_d^{-1} (d-d_0))$

•Parameters of model space: the *a priori* information may have a Gaussian form:
$P_{apriori}(m) = k' \exp(-1/2\ (m-m_0)^T C_m^{-1} (m-m_0))$

•The *a posteriori* information will have the form:
$P_{post}(m) = P_{apriori}(m)\ P_d(d) = K \exp(-S(m))$

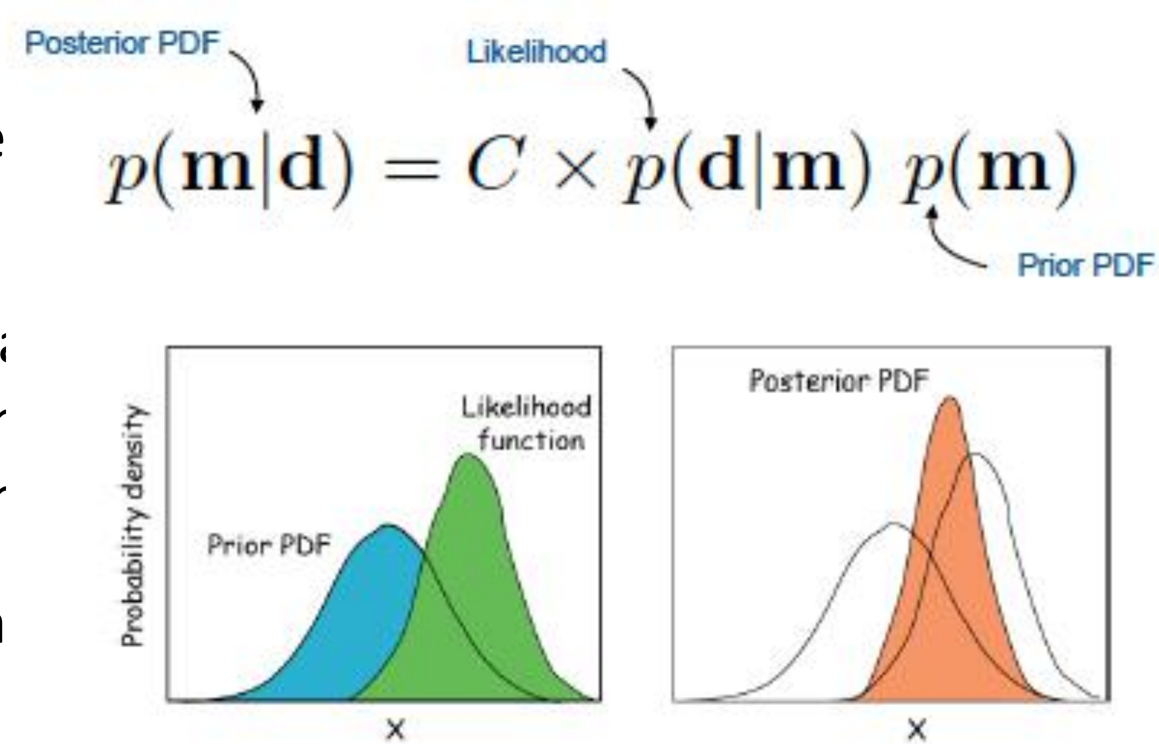
With $2\,S(m) = (m-m_0)^T C_m^{-1} (m-m_0)) + (d-d_0)^T C_d^{-1} (d-d_0))$

=> **Algorithm (dS/dm = 0)**

« Generalized Inverse problem »

Gaussian Statistics (least square criterion)

(Tarantola & Va



Courtesy of M. Sambridge

•Data space: measureme

$P_d(d) = k \exp(-1/2\ (d-d_0)^T$

•Parameters of model sp

may have a Gaussian forr

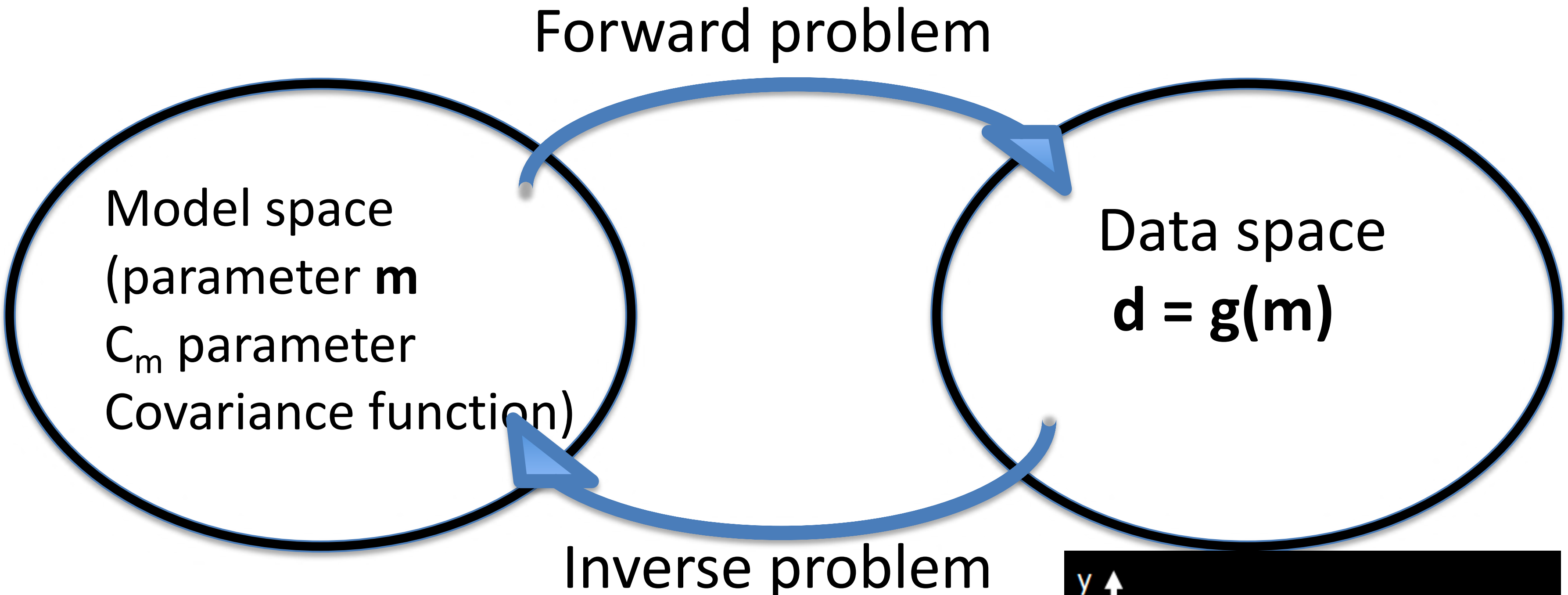
$P_{apriori}(m) = k' \exp(-1/2\ (m$

•The *a posteriori* informa

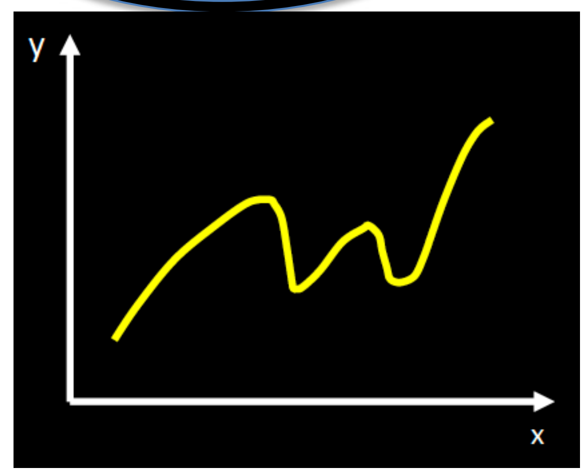
$P_{post}(m) = P_{apriori}(m)\ P_d(d)$

With $2\ S(m) = (m-m_0)^T\ C_m^{-1}\ (m-m_0)) + (d-d_0)^T\ C_d^{-1}\ (d-d_0))$

=> **Algorithm (dS/dm = 0)**



3- Non-linear problems: statistical approach exploration of the whole parameter space when possible



Monte-Carlo approach – Metropolis Algorithm
(Keilis-Borok, Yanovskaya, 1967; Frank Press, JGR, 1968, Mosegaard & Tarantola, 2002, …)
Many different strategies

Neural networks, Simulated annealing, neighborhood algorithm…
Leonard Seydoux, Thomas Bodin,…

First applications of Algorithm of Tarantola & Valette (1982)

$m^{est} = m_0 + C_m\, G^T\, C_d^{-1}\, (I + C_d^{-1}\, G\, C_m G_k^T)^{-1}\, (\, d - g(m^{est}) + G\, (m^{est} - m_0)$

Well suited for underdetermined problems: $N_d < N_m$

$m^{est} = m_0 + (C_m G^T C_d^{-1}\, G + I)^{-1} C_m\, G^T C_d^{-1}\, (\, d - g(m^{est}) + G\, (m^{est} - m_0)$

Well suited for overdetermined problems: $N_d > N_m$

A posteriori covariance function

$C_m^* = (G^t C_d^{-1} G + C_m^{-1})^{-1}$

Resolution

$R = C_m G^t\, (C_d + G C_m G^t)^{-1}\, G = (G^t C_d^{-1} G + C_m^{-1})^{-1} G^t C_d^{-1} G$

First applications of Algorithm of Tarantola & Valette (1982)

$m^{est} = m_0 + C_m G^T C_d^{-1} (I + C_d^{-1} G C_m G_k^T)^{-1} (d - g(m^{est}) + G (m^{est} - m_0))$

Well suited for underdetermined problems: $N_d < N_m$

$m^{est} = m_0 + (C_m G^T C_d^{-1} G + I)^{-1} C_m G^T C_d^{-1} (d - g(m^{est}) + G (m^{est} - m_0))$

Well suited for overdetermined problems: N_d

A posteriori covariance function

$C_m^* = (G^t C_d^{-1} G + C_m^{-1})^{-1}$

Resolution

$R = C_m G^t (C_d + G C_m$ … $C_d^{-1} G$

Complex waveform

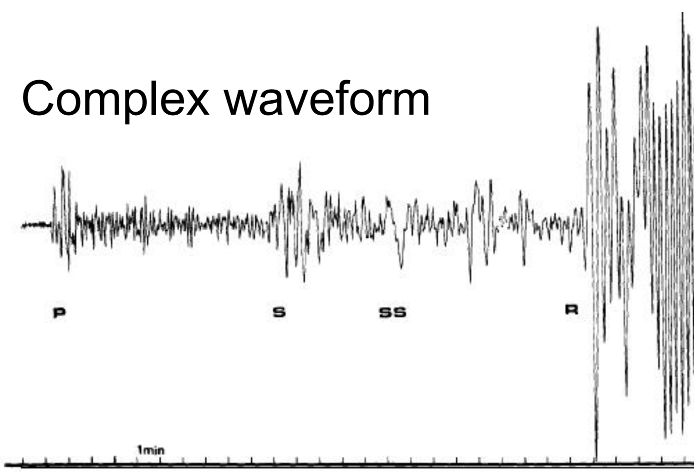


Fig. 3. Record of the Argentina (San Juan) earthquake at Pamatai. A mean line has been drawn to show the lon the Rayleigh phase.

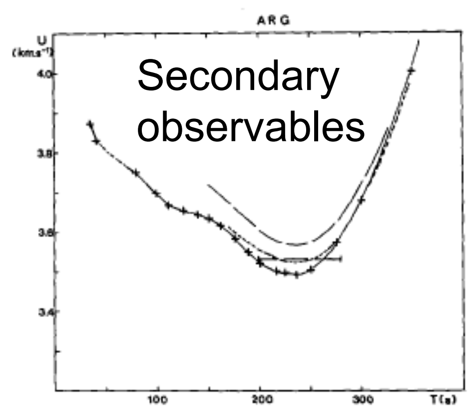


Fig. 5. Rayleigh wave group velocity dispersion curve for the Argentina (San Juan) earthquake recorded at Pamatai, corrected for continental path. - - - - - - Lévêque's (1980) model of a young ocean; — — — model PEMO (Dziewonski et al., 1975); ———— averaged group velocities between 200 and 280 s for 'ridges' (after Mills, 1978).

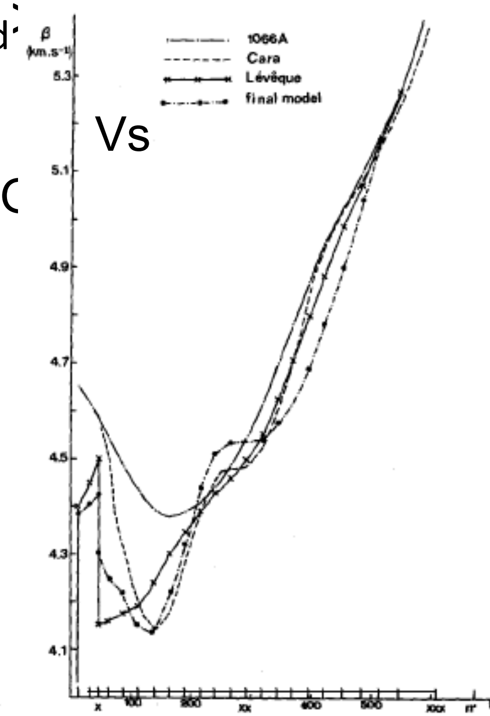


Fig. 15. Model of the shear velocity structure resulting from the inversion of the Argentina dispersion curve compared with model 1066 A, a 90 Ma ocean model of Cara (1979) and the young ocean model of Lévêque (1980).

Montagner and Jobert, 1981

TV inverse theory

- d data space, m parameter space (continuous fns)
- d = g(m)
- (slightly) Non-linear problems
- Errors on data C_d, *A posteriori* errors C_m

- Travel times, phase data too simple =>

Full waveform calculation (SEM) + inversion

=> ***BIG COMPUTERS***

- Isotropic, smooth elastic medium too simple =>

Complete (an)elastic tensor => ***ANISOTROPY, ANELASTICITY***

TV inverse theor

- d data space, m parameter space (contir)
- d = g(m)
- (slightly) Non-linear problems
- Errors on data C_d, *A posteriori* e
- Travel times, phase data too >

Full waveform calc (SEM) + inversion

=> ***BIG COMPUTE***

- Isotropic, smoot medium too simple =>

Complete (an) ensor => ***ANISOTROPY, ANELASTICITY***

THINK BIG AND GENERAL

Connection Machine CM2->5 (1989-> 1992)
First parallel computer in France

Slow vehicle

CM5

Fast and elegant vehicle

Applications: Tomographic Imaging

Practical applications: Ana Ferreira, Stéphanie Durand, …

Advanced inference methods: Andrew Curtis, Thomas Bodin

- ***Forward Problem:*** ***Theory d=g(m)***

  d data space, m model parameter space

- Reference Earth model m_0:

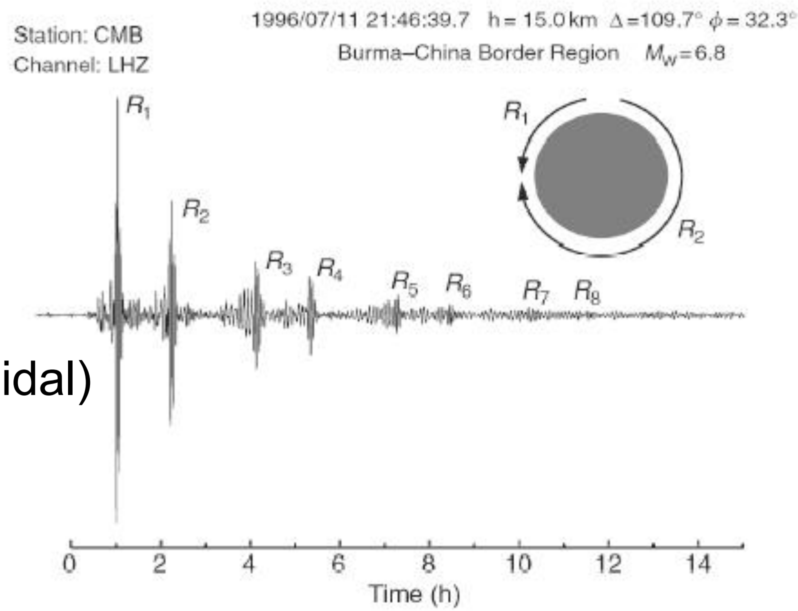
  $d_0 = g(m_0)$
- Kernels $\partial g/\partial m$
- Cd function (or matrix) of covariance of data

- ***Inverse Problem:*** $m-m_0 = \underline{g}^{-1}(d-d_0)$

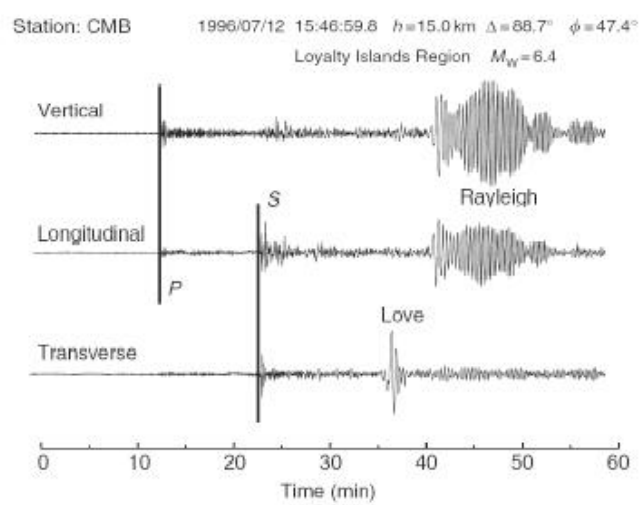
- C_{m0} a priori Covariance function of parameters
- C_{mf} a posteriori Covariance function of parameters
- R Resolution

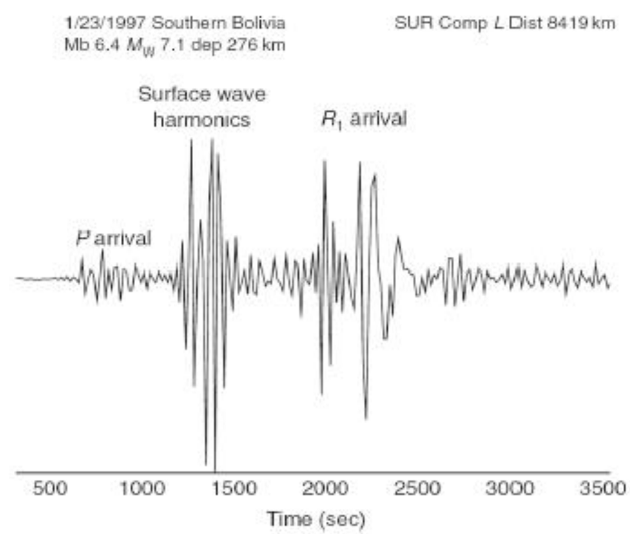
Seismic waveforms- Data Space

Various kinds of **Data**



- Free oscillations (Spheroidal, Toroidal)
- Surface waves (Rayleigh, Love)
- Body waves (P, S, …)

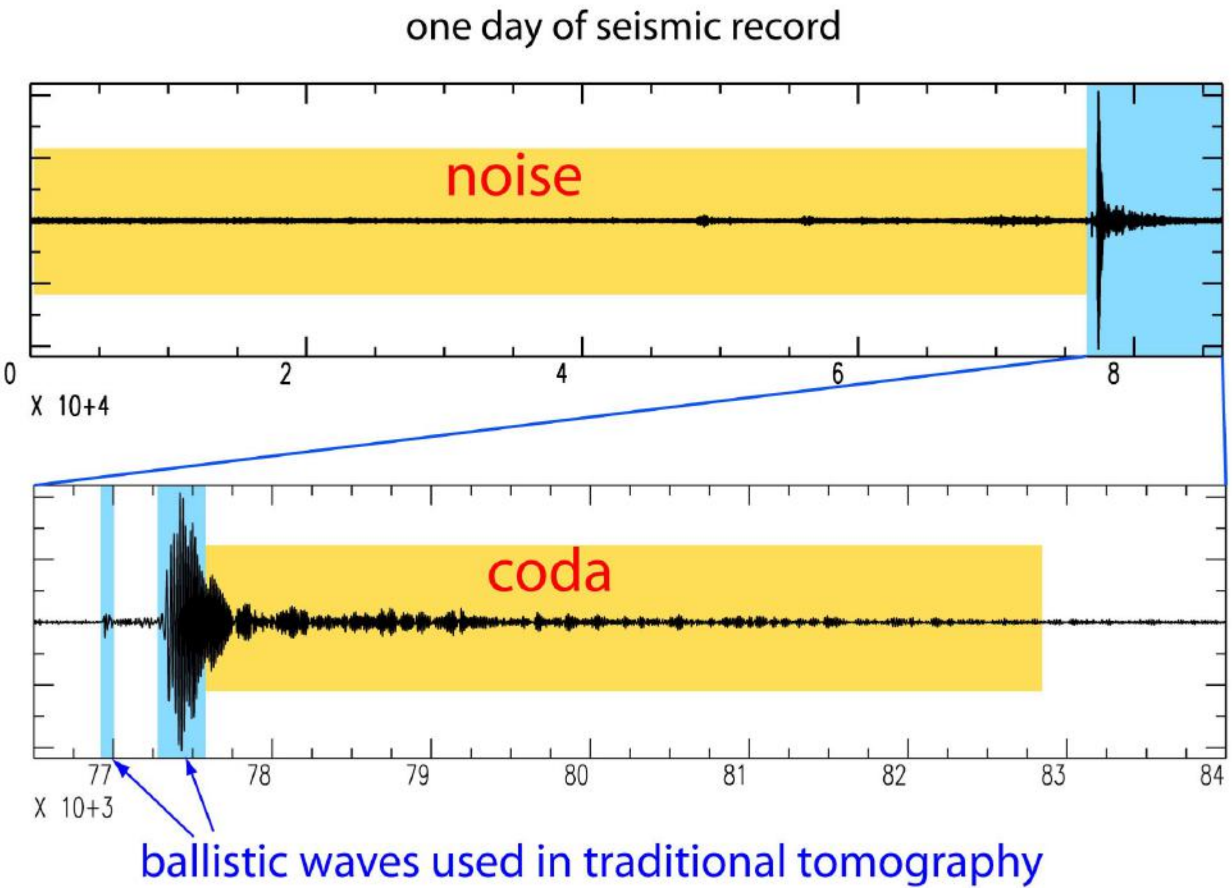




Romanowicz, 2002

Ambient noise

Microseismic noise
T<30s

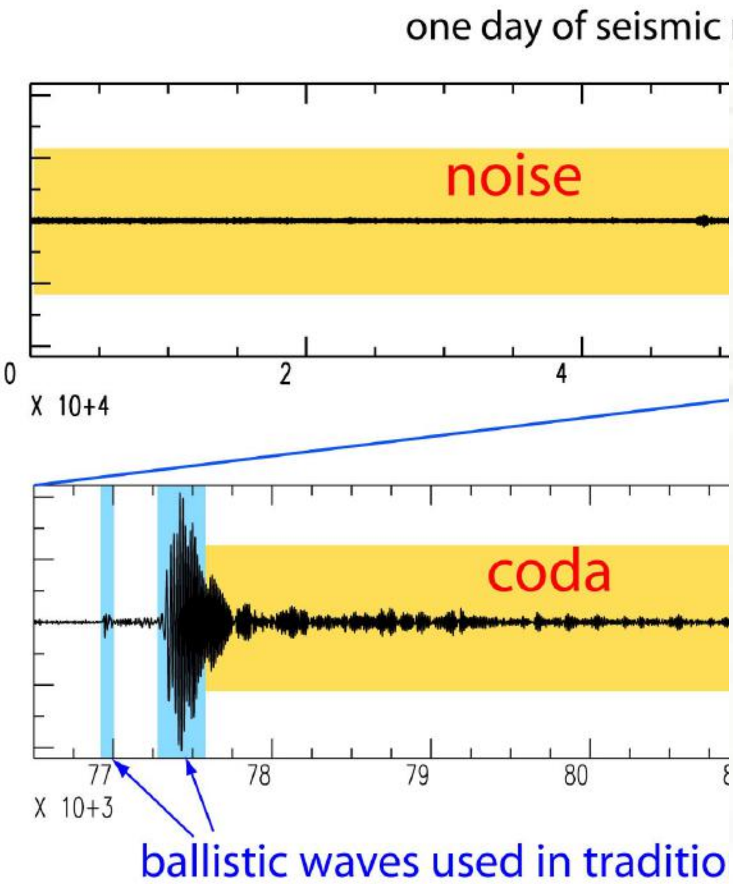


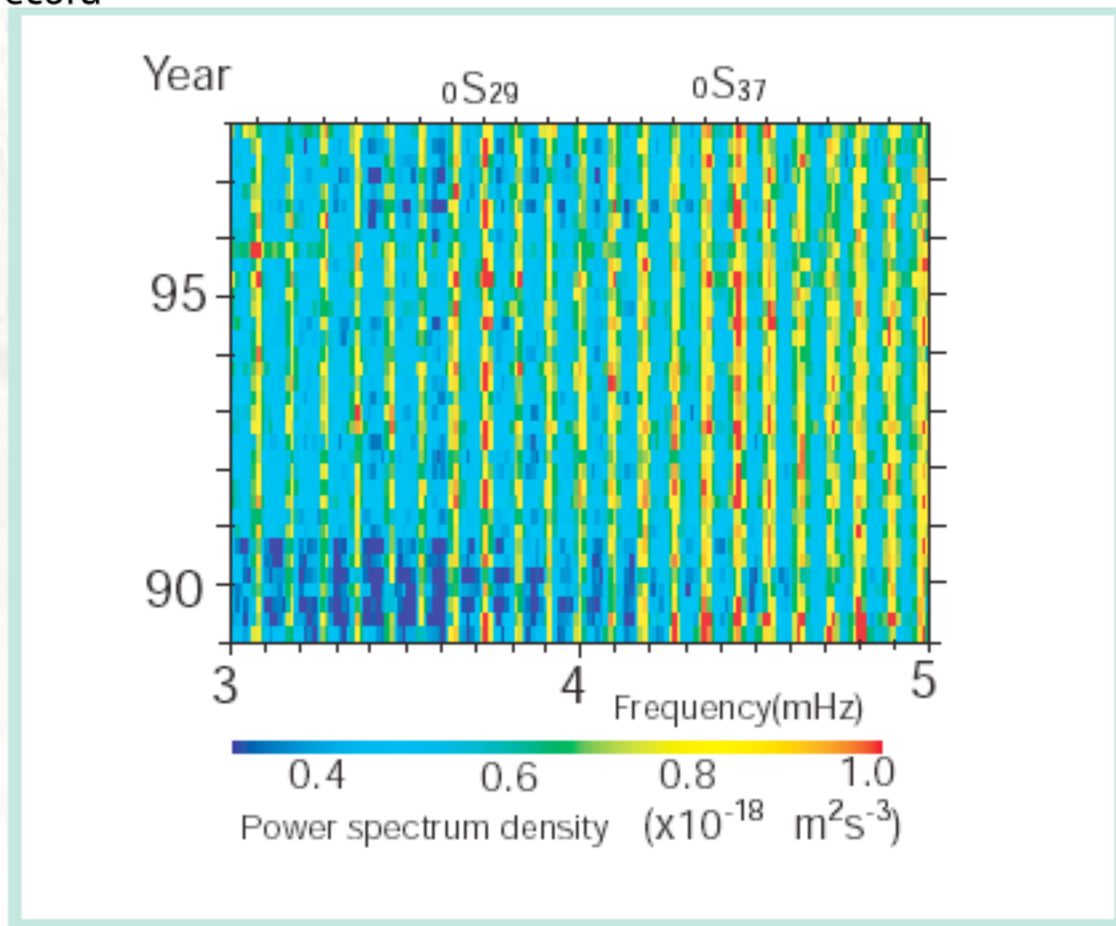
Courtesy of N. Shapiro

Ambient noise

Microseismic noise
T<30s

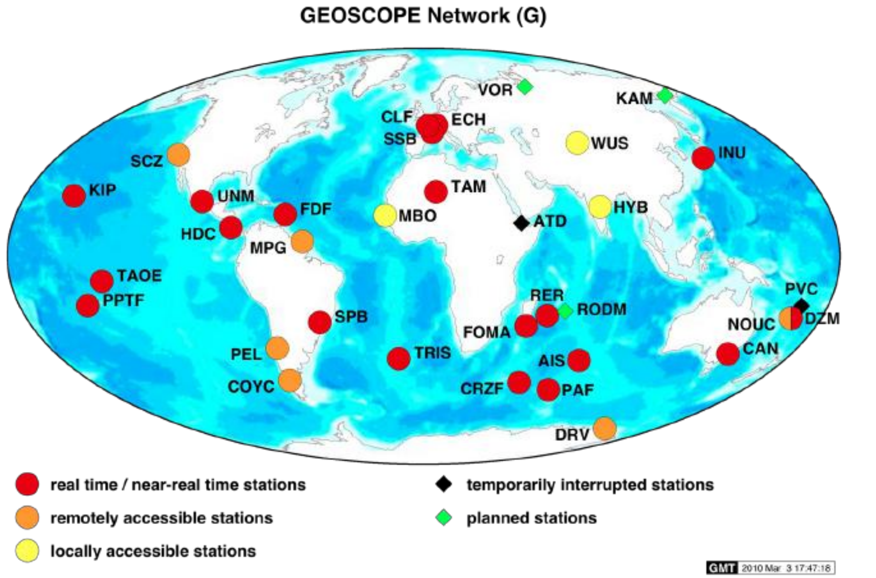
Seismic hum (eigenfrequencies)
T>100s





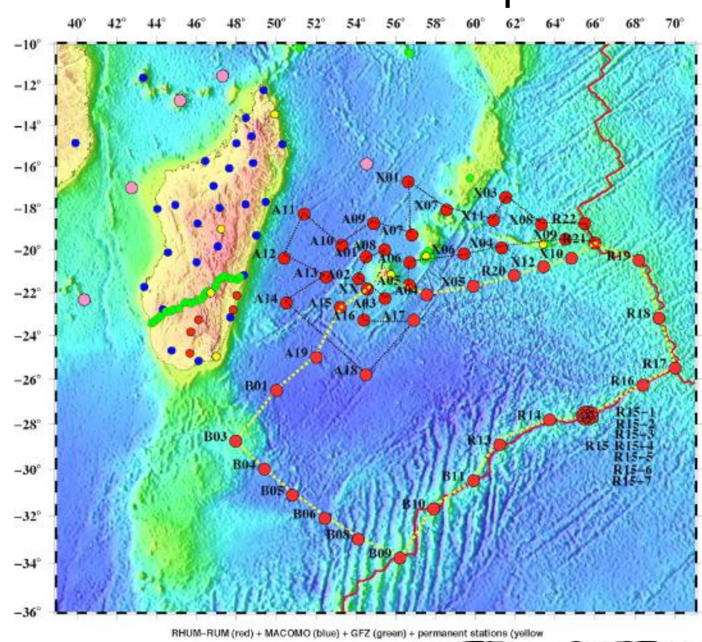
Courtesy of N. Shapiro

Nishida et al., 2000

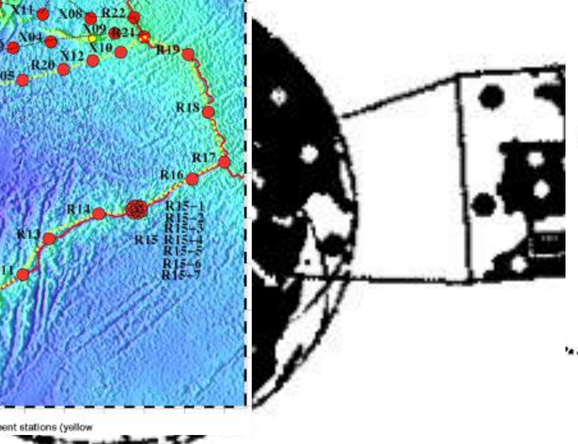


RECEIVERS

OBS -RHUM-RUM Exp.

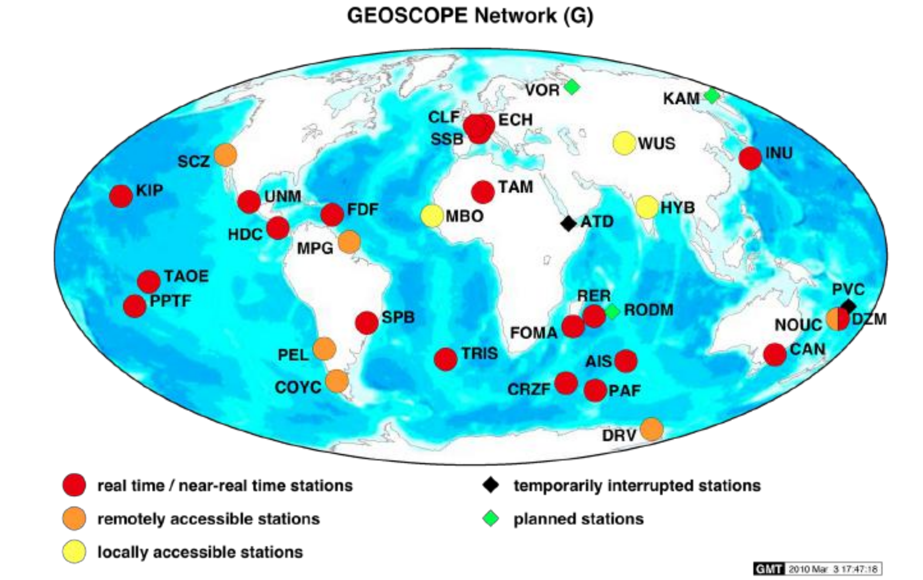


BIG NE



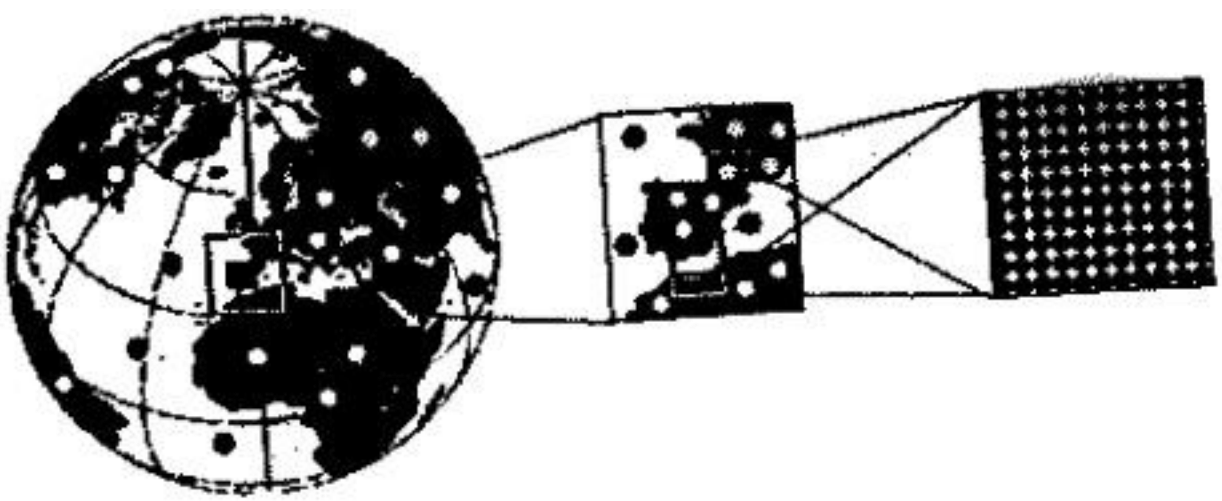
[web.mst.edu]

1995: Concept of Hierarchical Multiscale Network (IPGP) => US-array, OBS

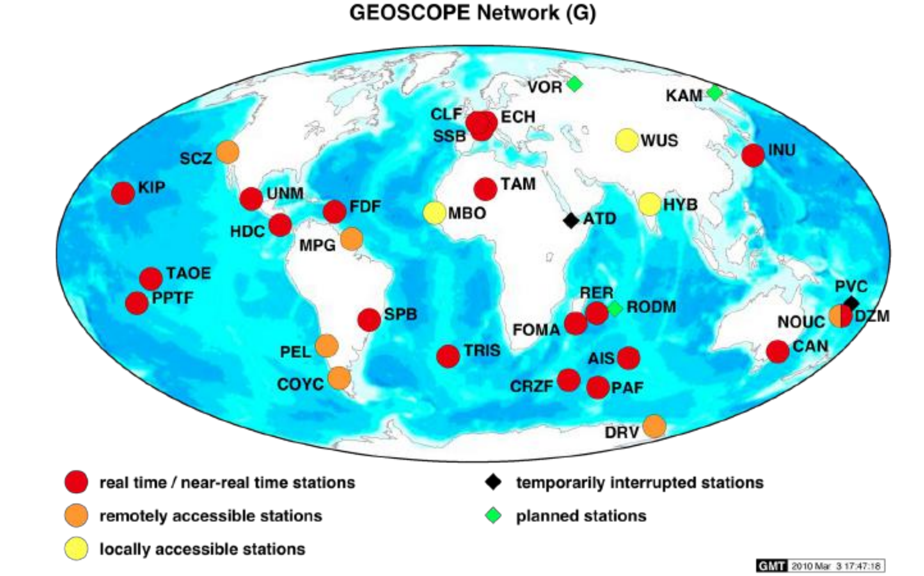


RECEIVERS

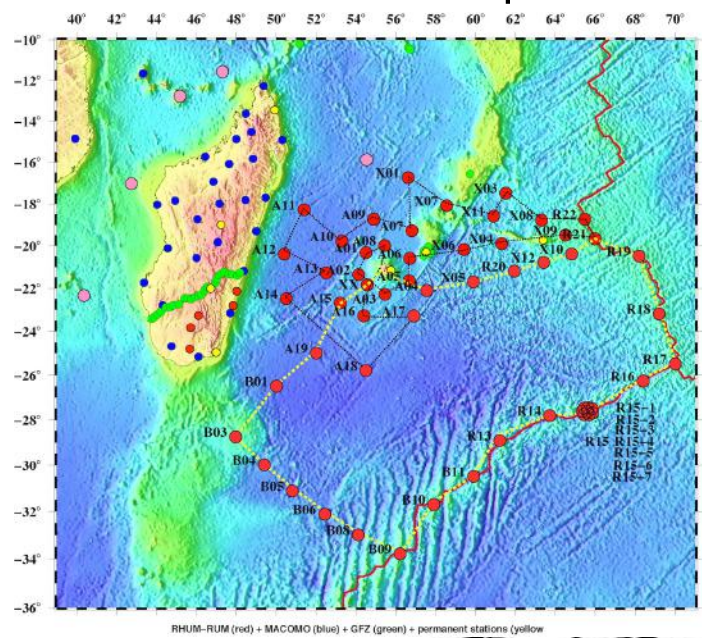
BIG NETWORKS

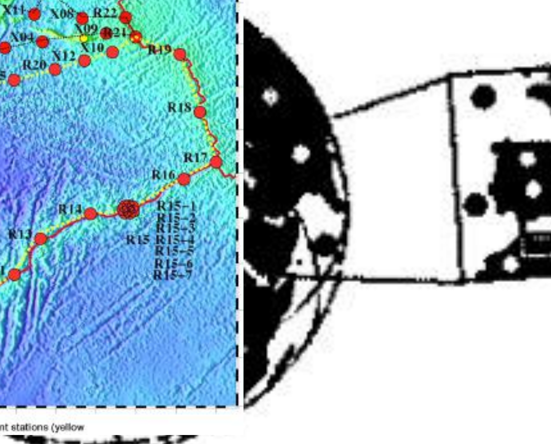


1995: Concept of Hierarchical Multiscale Network (IPGP) => US-array, OBS



OBS -RHUM-RUM Exp.

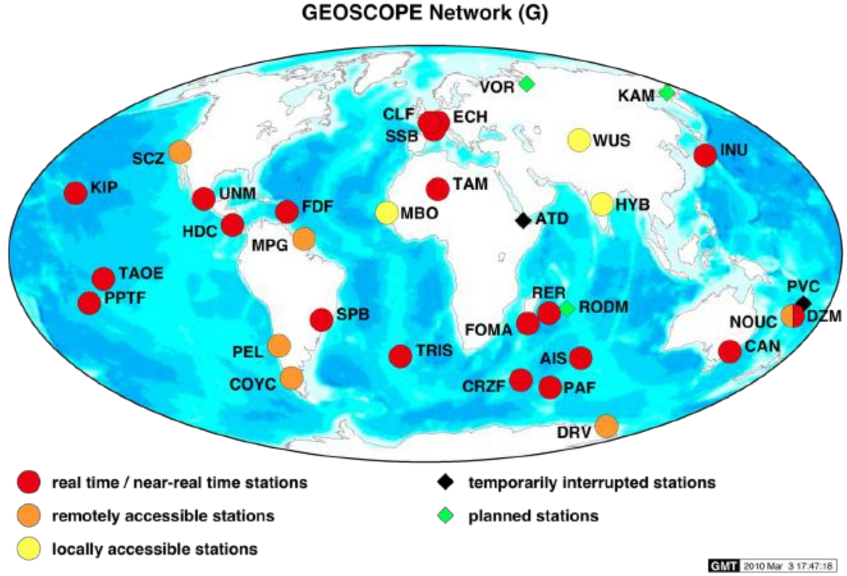




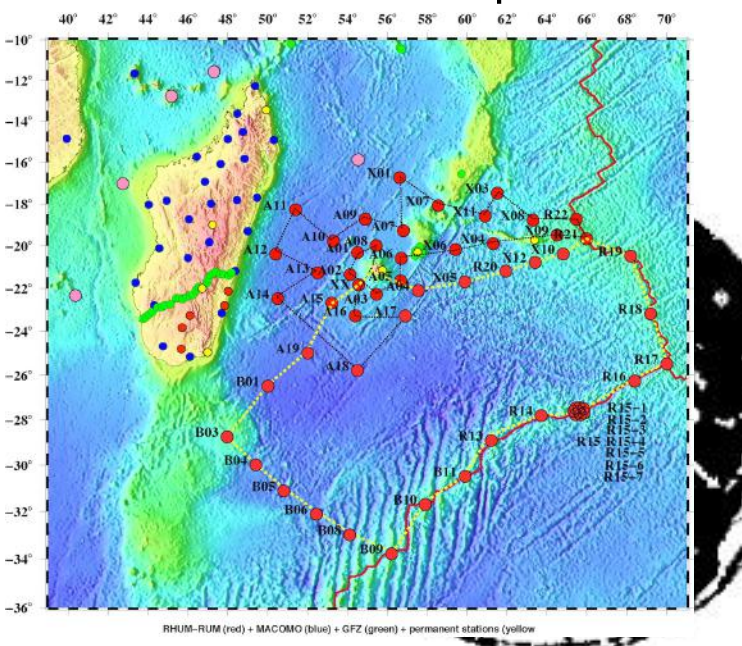
BIG NE

[web.mst.edu]

1995: Concept of Hierarchical Multiscale Network (IPGP) => US-array, OBS, RESIF



OBS -RHUM-RUM Exp.



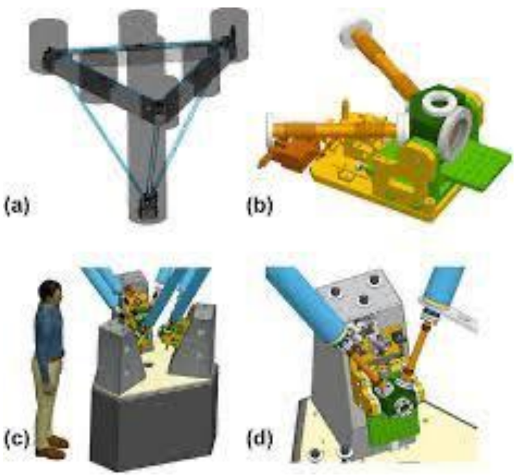
INSIGHT- Exploration of Planet Mars

[web.mst.edu]

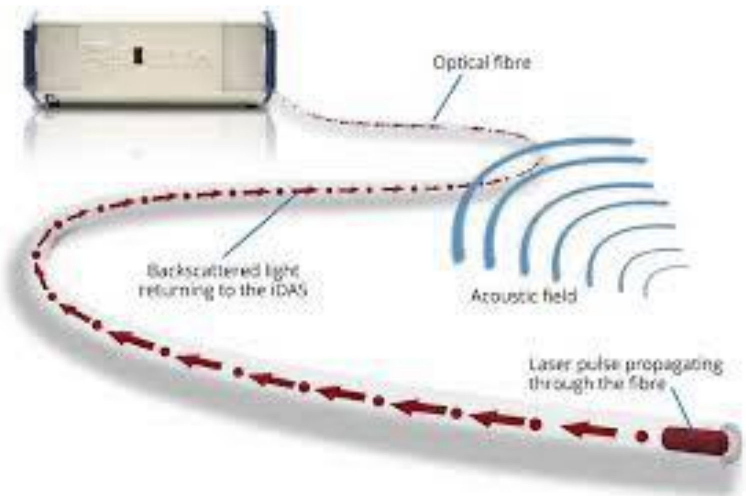
1995: Concept of Hierarchical Multiscale Network (IPGP) => US-array, OBS, RESIF

New Instrumentation

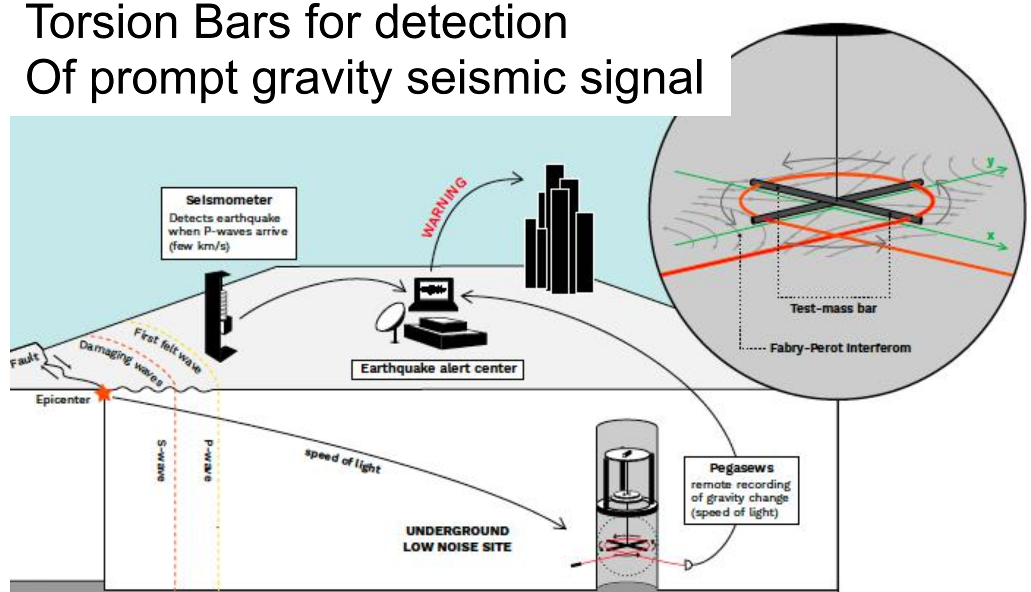
Rotational Seismometers



DAS (Distributed Acoustic sensing)



Torsion Bars for detection Of prompt gravity seismic signal



Tomographic Imaging

- ***Forward Problem:*** ***Theory $d=g(m)$***

  d data space, ***m model parameter space***

- Reference Earth model m_0:

  $d_0 = g(m_0)$

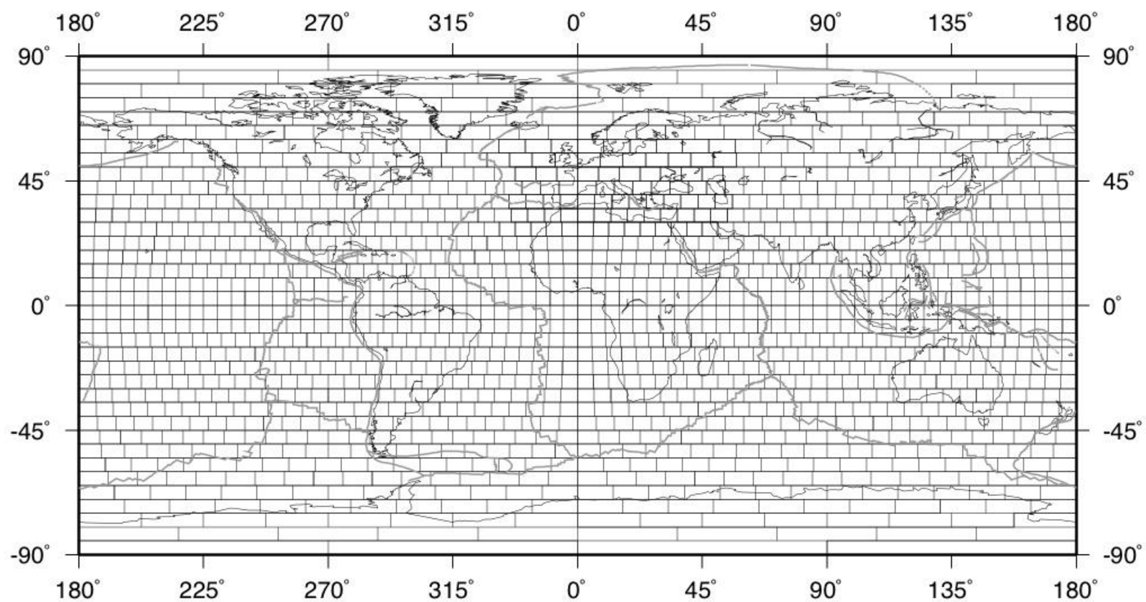
- Kernels $\partial g/\partial m$
- Cd function (or matrix) of covariance of data (error bars, uncertainties)

- ***Inverse Problem:*** ***$m-m_0 = \underline{g}^{-1} (d-d_0)$***

- C_{m0} a priori Covariance function of parameters
- C_{mf} a posteriori Covariance function of parameters
- R Resolution

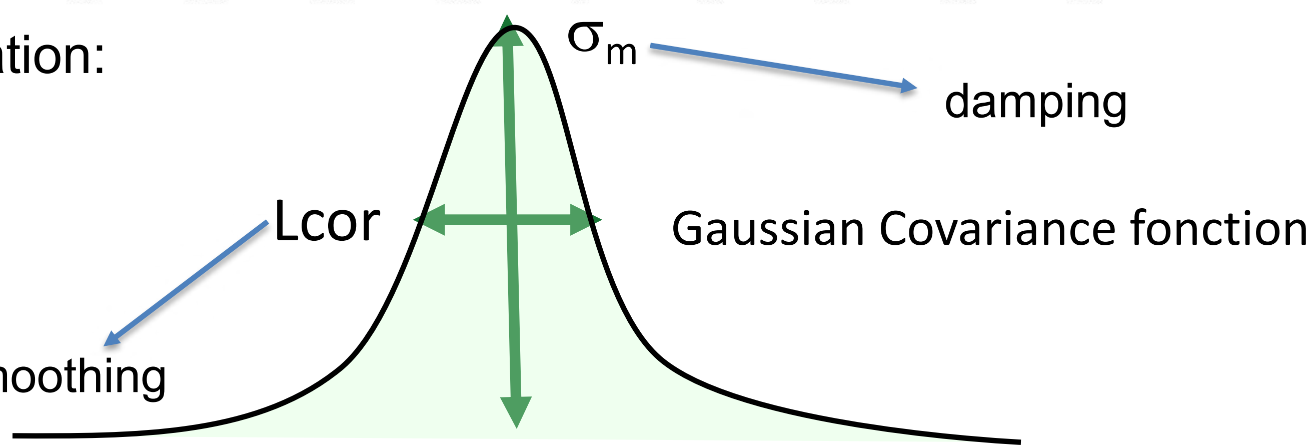
Model Parameter Space
choice of basis functions, covariance function

pixels ("local" basis functions)



Continuous regionalization:

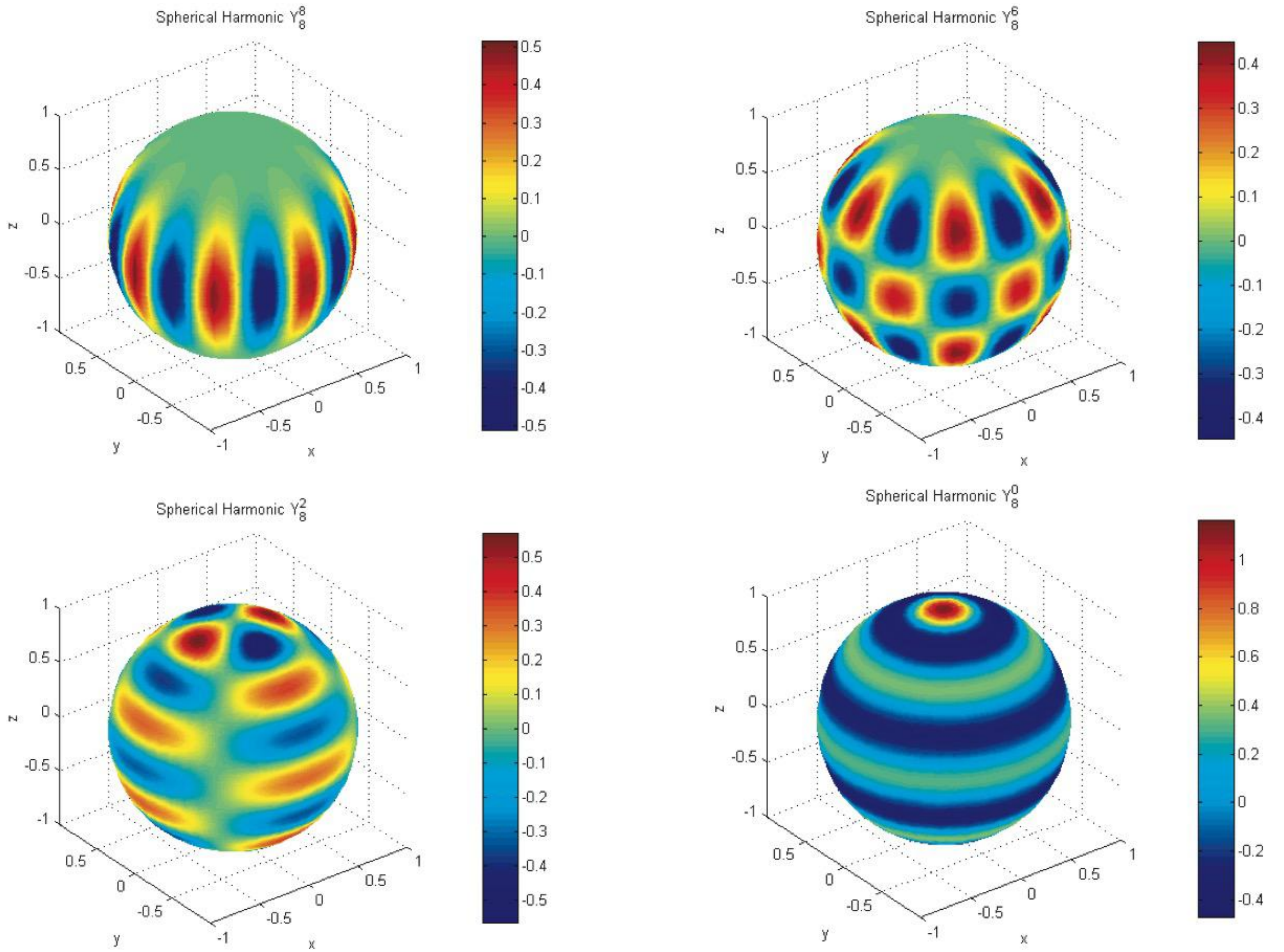
σ_m a priori error
Lcor correlation length



Model Parameter Space

spherical harmonics
("global" basis functions)

$Y_l^m(\theta,\phi)$

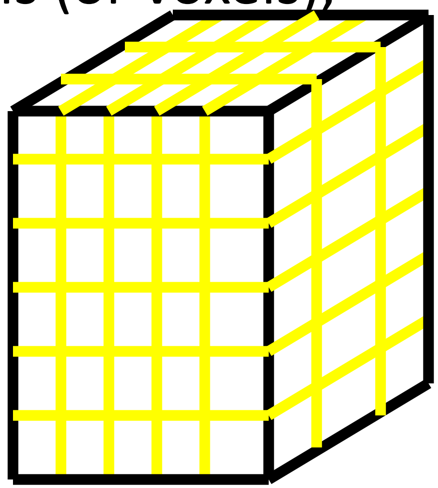


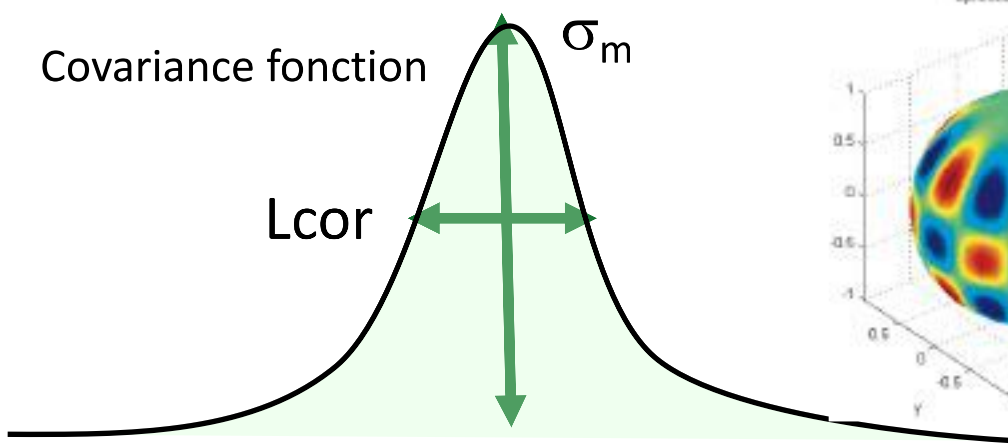
Model Parameter Space:
Principle of Parsimony (Occam's razor)

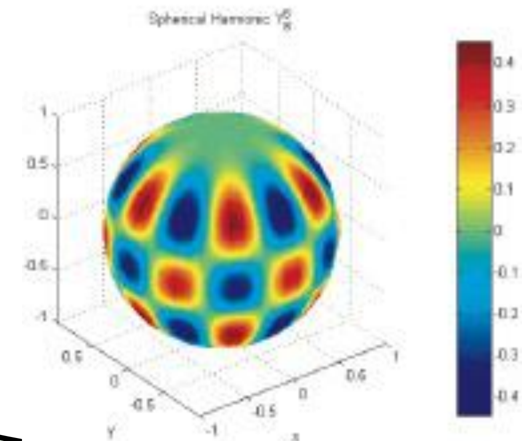
- **Physical parameters**: ρ + 21 (or 13) physical elastic parameters

- Geometrical (geographical) parameterization: $\mathbf{m}(r,\theta,\phi)$

Cells (or voxels), Continuous parameterization, $Y_l^m(\theta,\phi)$







- Spherical harmonic expansion $\Sigma\, a_l^m(r) Y_l^m(\theta,\phi)$

- Lateral, radial resolutions, parameters

Horiz. 1000km, Rad. 50km => 500*60*22 ≈ 660,000 parameters

Different types of seismic data

-Normal Modes of the Earth (eigenfrequencies)
-Surface waves, Rayleigh, Love (phase, group velocities)
-Body waves P, S, ….. (travel times)
-"Ambient Noise" (1-20s microseismic, T>200s hum)

-***Secondary observables (eigenfrequencies, travel times, dispersion velocity) versus full waveform inversion?***

Different types of model parameterisation

-Parametrisation according to the dataset.

-Spatial coverage: Millions, Billions of paths

-***Can we use ray theory, waveform inversion?***

-***Isotropic versus anisotropic, anelastic medium?***

AT inverse problem theory

- d data space, m parameter space (continuous fns) -Bayesian inference
- d = g(m)
- Non-linear problems
- Errors on data C_d, *A posteriori* errors C_m

- Travel times, phase data too simple =>

Full waveform calculation (SEM) +

Inversion by adjoint tomography

Isotropic, smooth elastic, anelastic medium too simple =>

Complete (an)elastic tensor => ***ANISOTROPY, ANELASTICITY***

Waveform Inversion- Adjoint (transpose) Tomography

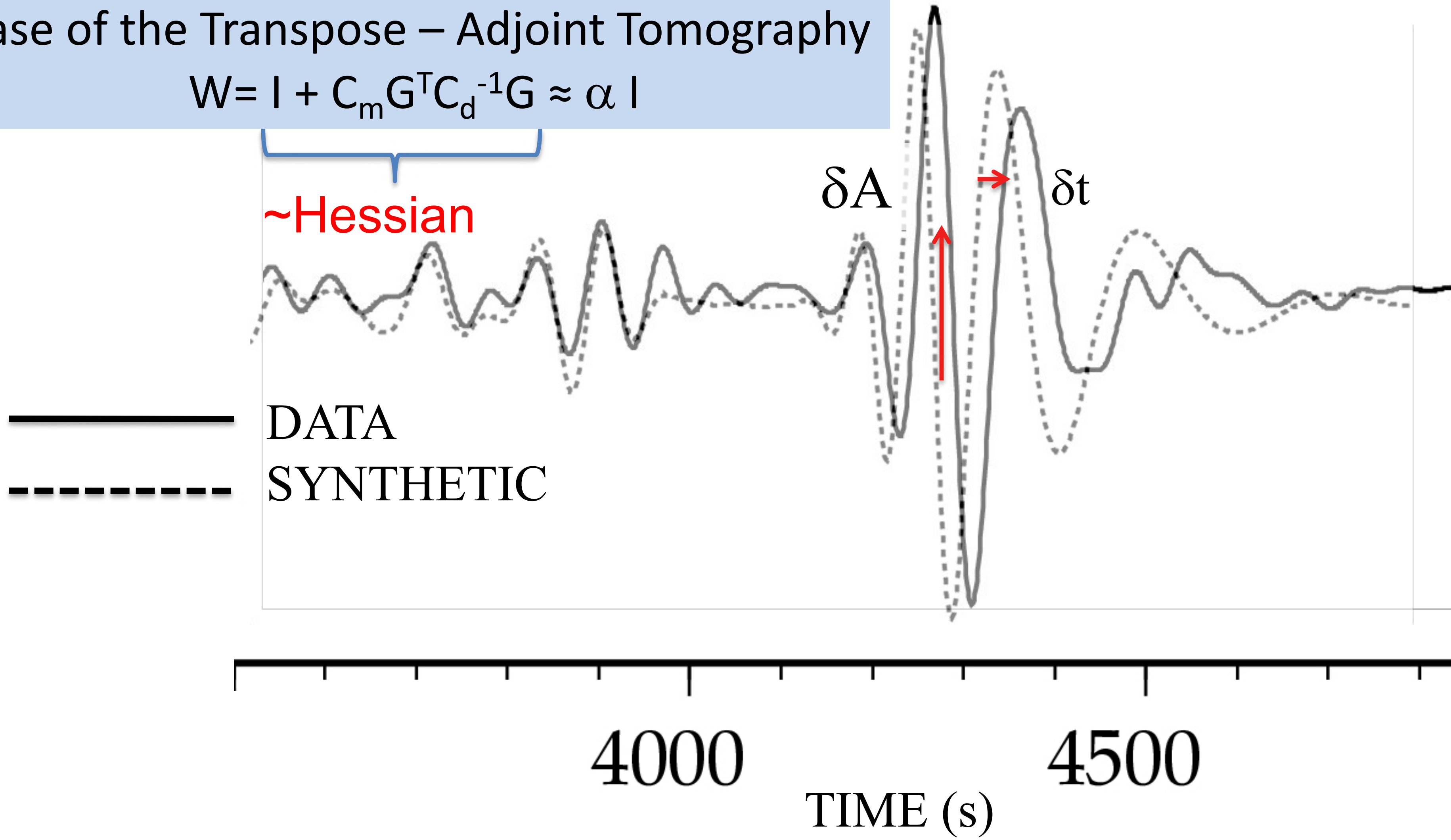
(see also Claerbout, 1985)

$$m^{est} = m_0 + (I + C_m G^T C_d^{-1} G)^{-1} C_m\, G^T C_d^{-1}\, (\, d - g(m^{est}) + G\,(m^{est} - m_0))$$

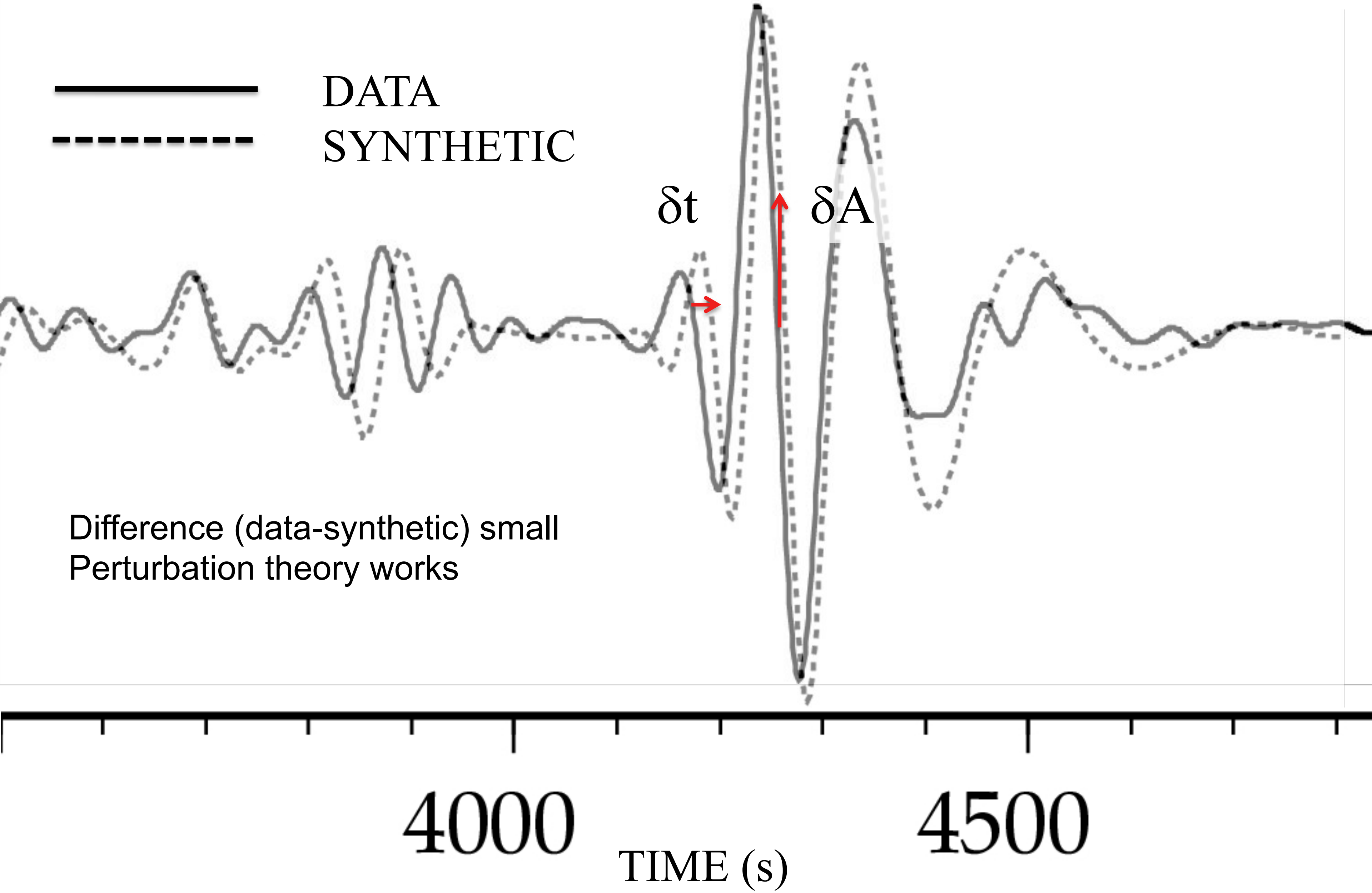
Base of the Transpose – Adjoint Tomography

$$W = I + C_m G^T C_d^{-1} G \approx \alpha\, I$$

~Hessian

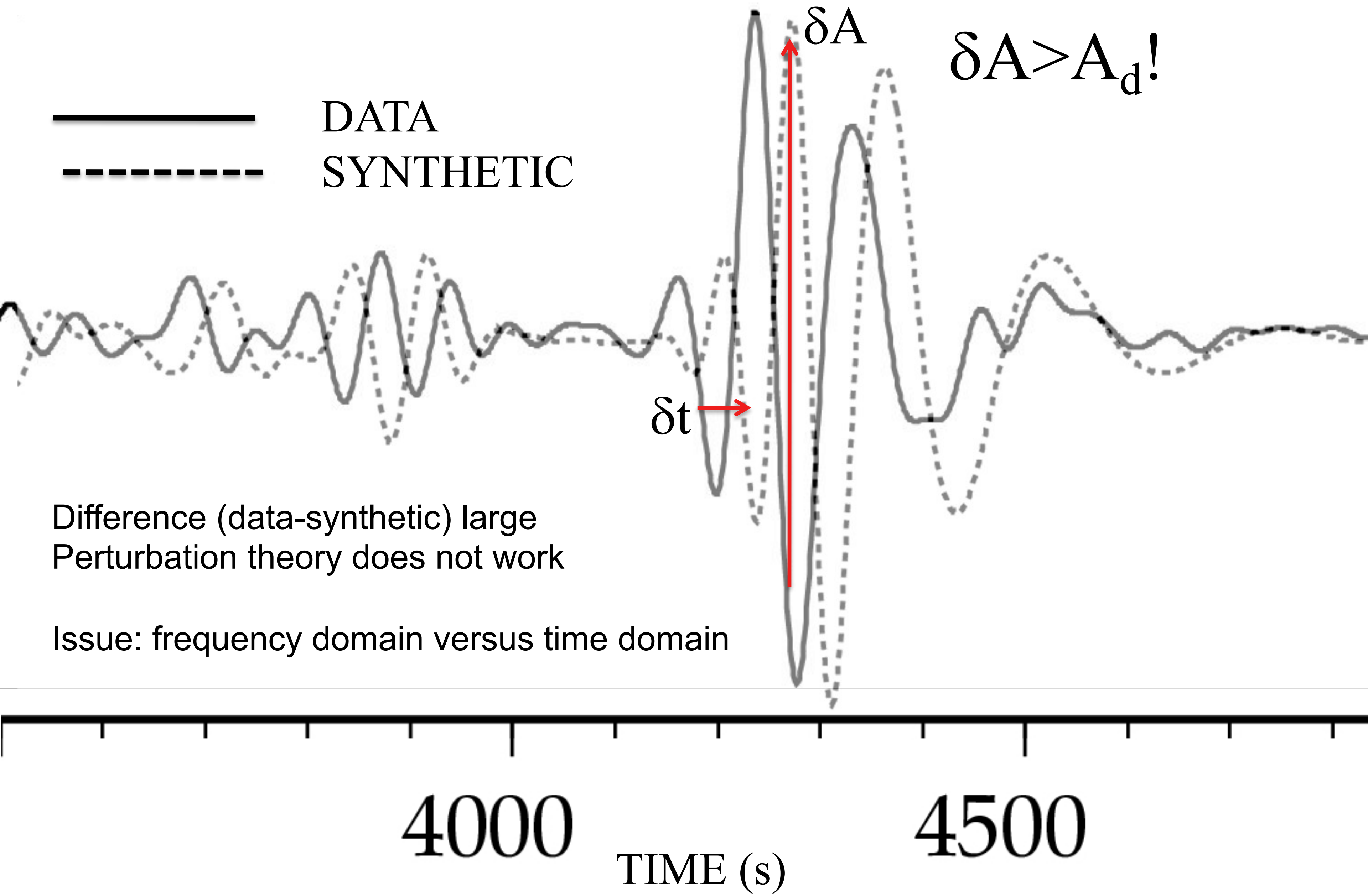


Waveform Inversion- Adjoint (transpose) Tomography



Difference (data-synthetic) small
Perturbation theory works

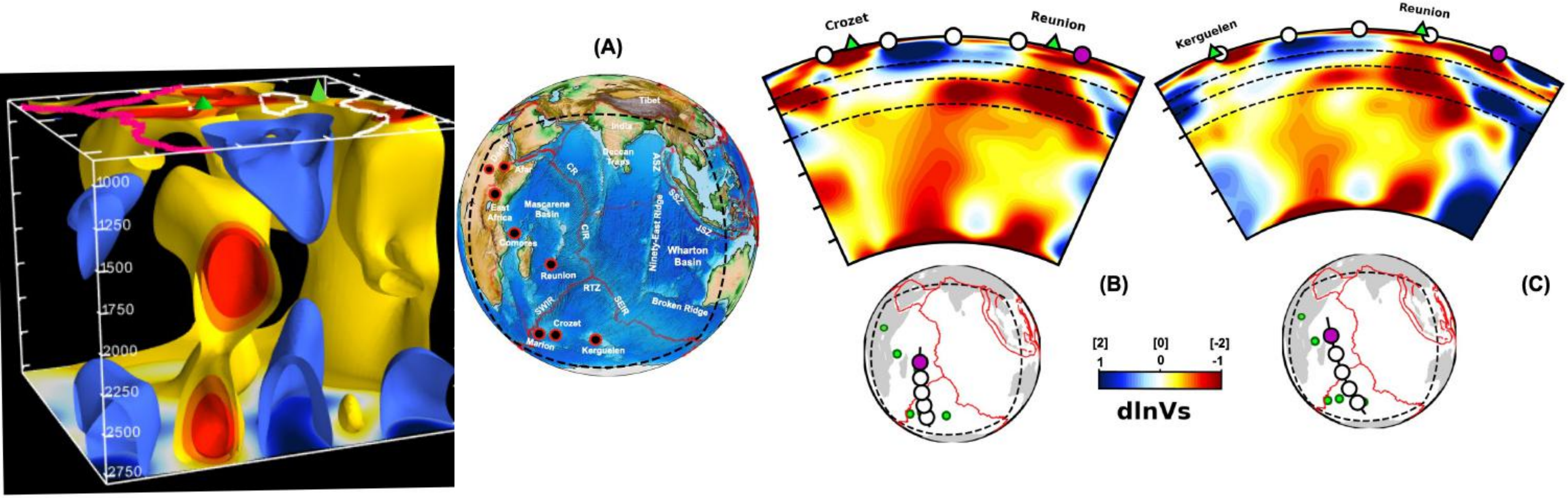
(full) Waveform Inversion



Difference (data-synthetic) large
Perturbation theory does not work

Issue: frequency domain versus time domain

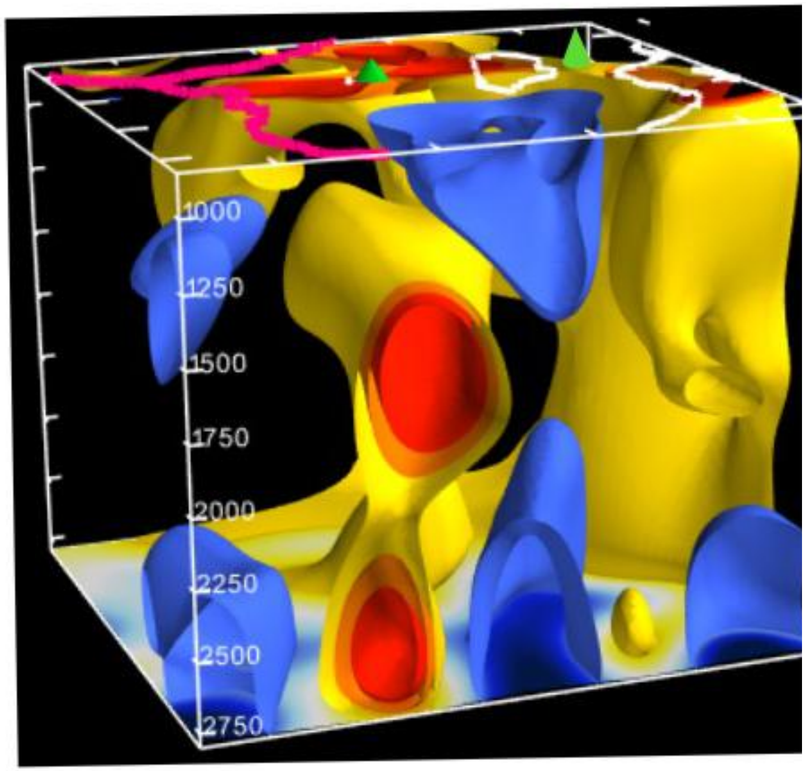
Tomography beneath La Réunion plume by Full Waveform Inversion of broadband seismic data

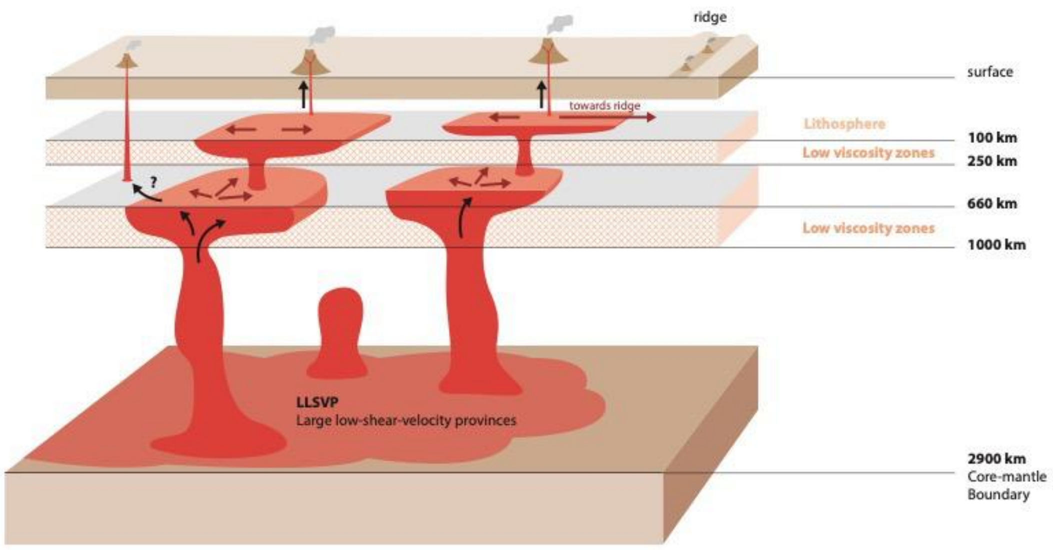


Complex plumbing system with vertical conduits and stagnant ponding zones

Wamba et al., Sci. Adv., 2023

Tomography beneath La Réunion plume by Full Waveform Inversion of broadband seismic data





Complex plumbing system with vertical conduits and stagnant ponding zones

Wamba et al., Sci. Adv., 2023

Time Reversal- Adjoint Tomography

Back propagation

-Source focusing (Green function known)

-Adjoint Tomography => Structure (source known)

Difference Adjoint technique – Time reversal imaging

Inverse problem for source inversion (Kawakatsu and Montagner, 2008)

Forward Problem: Data $\mathbf{d(t)} = \mathbf{G_{*}f}$ ($\mathbf{f}$ force system)

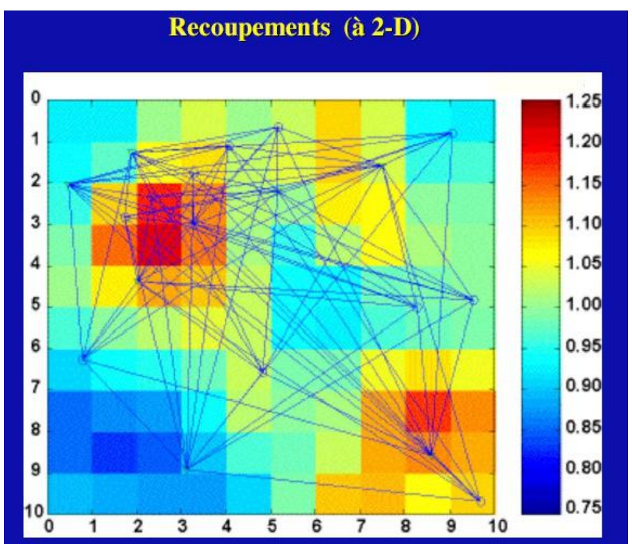
Inverse problem (least square solution) $\mathbf{f} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\,\mathbf{d}$

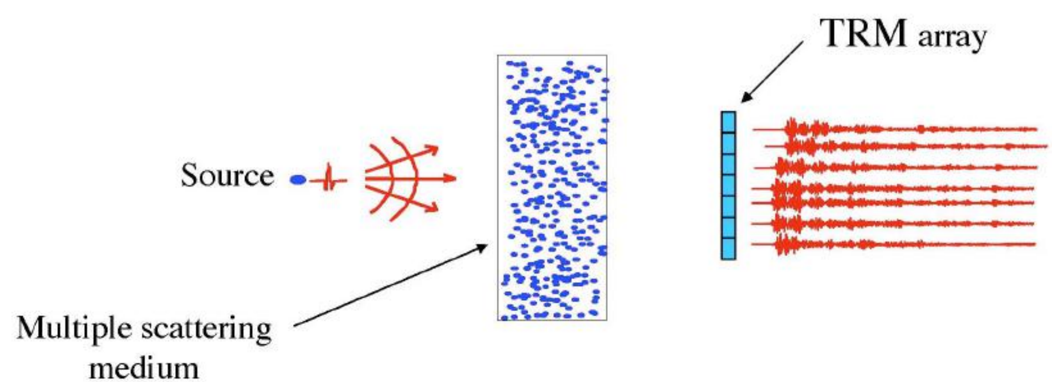
Time reversal Imaging:

$\mathbf{G}^T\mathbf{G} \approx \mathbf{I}$ (Hessian: identity operator)

$\mathbf{f} \approx \mathbf{G}^T\,\mathbf{d(T\text{-}t)}$ Physical Meaning

Need for an excellent Green's function

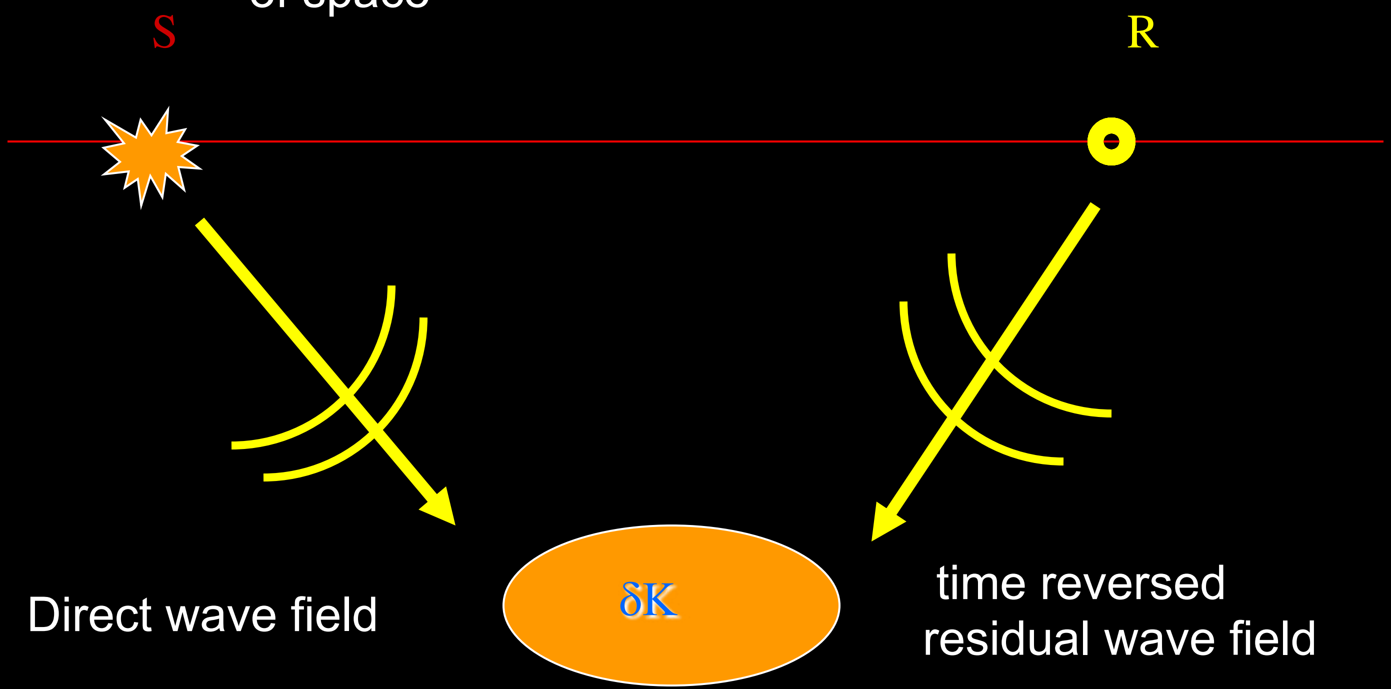




Multiple Sources (scattered waves) explore the whole space

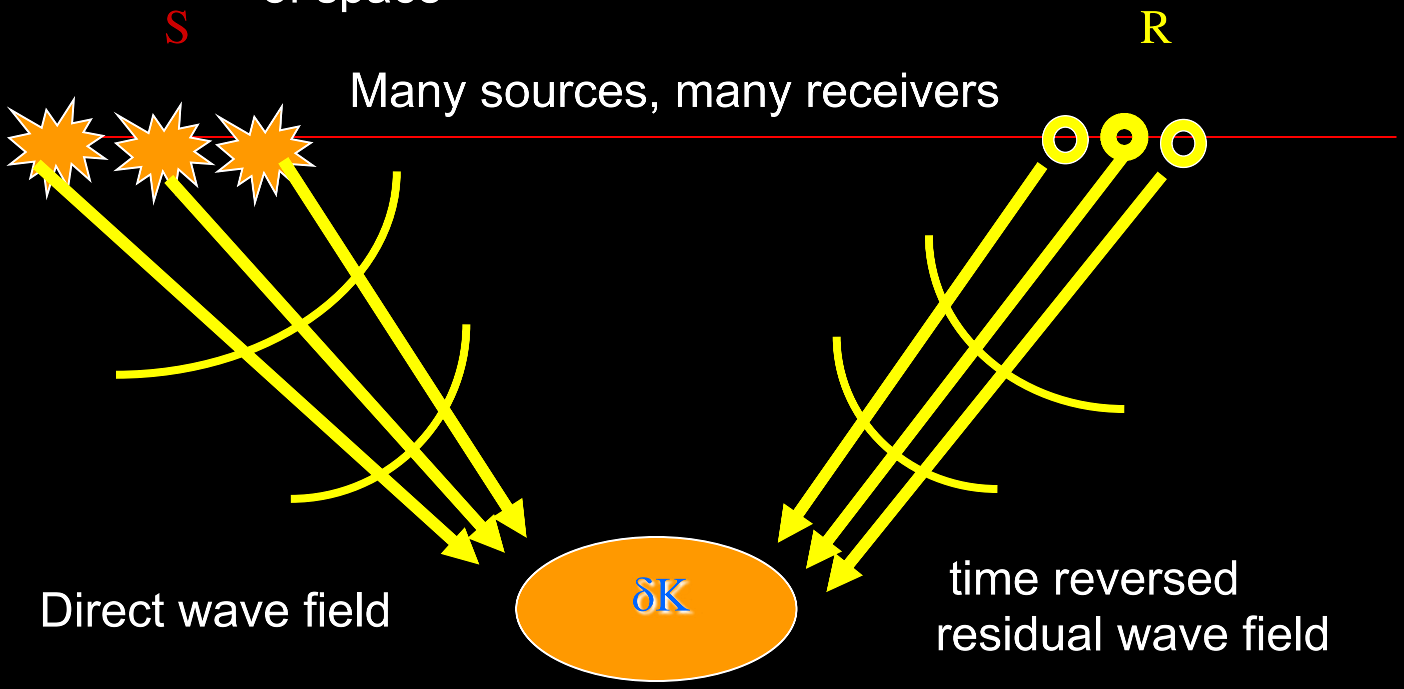
Time Reversal- Adjoint Tomography

Physical Interpretation:
1- Propagation of the residuals backward in time (time reversed residuals)
2- Correlation of the « time-reversed » residual wavefield and the « downgoing » wavefield at each point of space



Time Reversal- Adjoint Tomography

Physical Interpretation:
1- Propagation of the residuals backward in time (time reversed residuals)
2- Correlation of the « time-reversed » residual wavefield and the « downgoing » wavefield at each point of space



CONCLUSIONS

- Be passionate, critical, open-minded, provocative but generous.
- Always address scientific issues in the most general approach
(heterogeneous, anisotropic, anelastic…)
- Be ambitious! Money is not an issue
(big computers, multiscale networks)
- Go beyond the consensus

Albert was a brainstorm.