Bayesian Design for Minimal Forecast Uncertainties

M. R. Khodja H. A. Djikpesse M. D. Prange

Mathematics & Modeling Department
Schlumberger-Doll Research, Cambridge, MA

Monday, June 8, 2010

Schlumberger

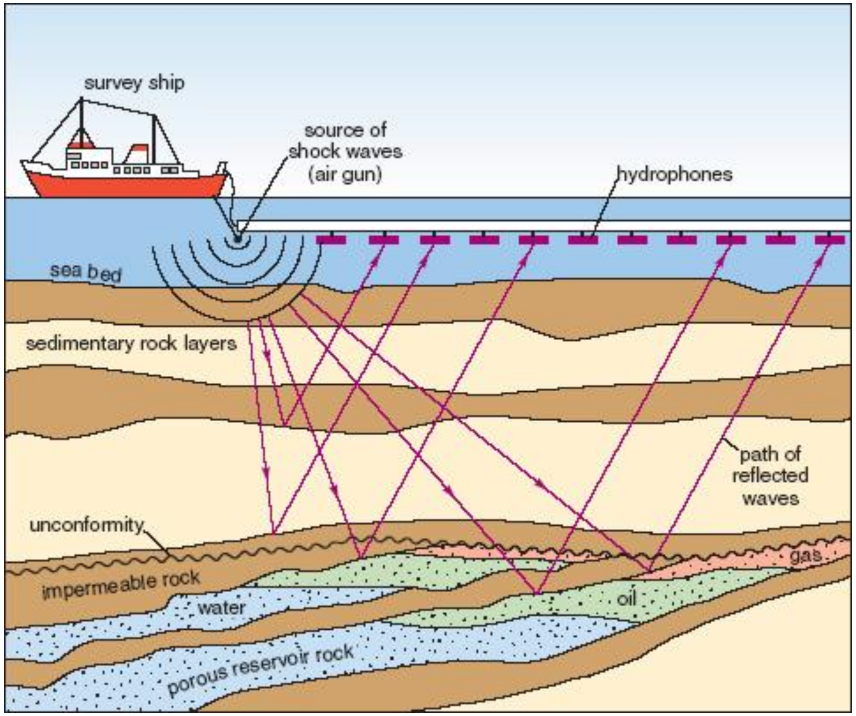
Outline

- Motivation
- Problem Formulation
- Proposed Bayesian Criterion
- Numerical Results
- Summary
- Discussion

Schlumberger

Motivation

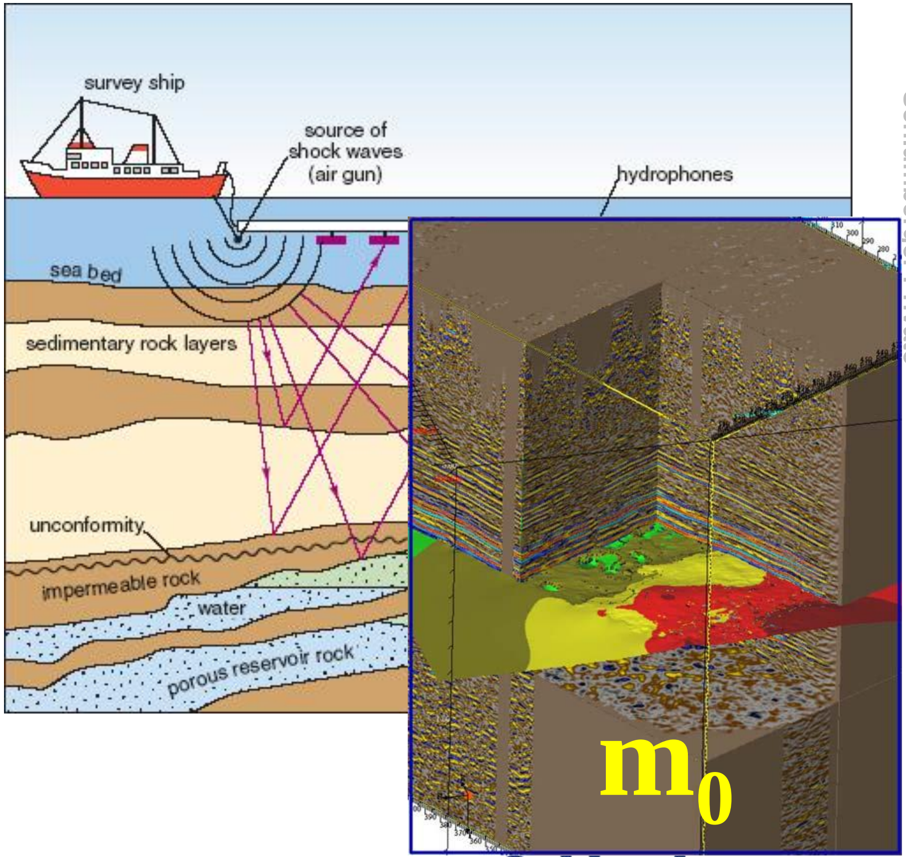
- **A surface seismic survey is performed**



Schlumberger

Motivation

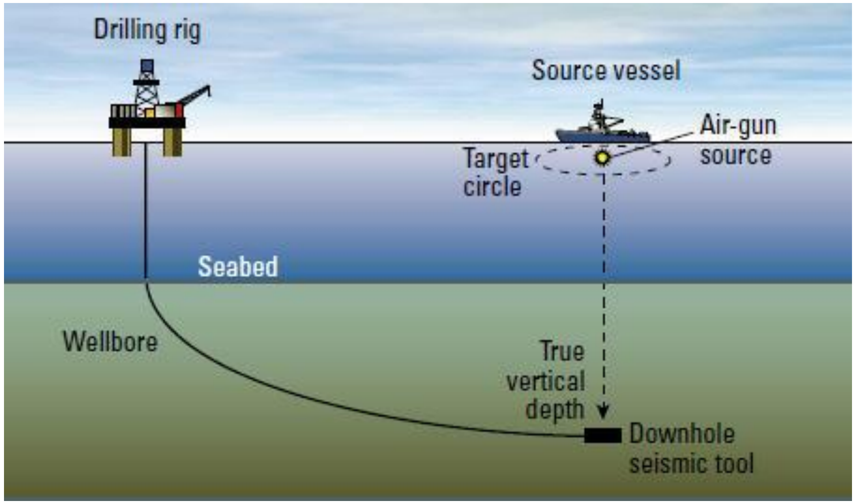
- **A surface seismic survey is performed**
- **The interpreted seismic data yields a primary model, m_0, of the subsurface**
- **The interpretation of the data reveals the presence of a zone of interest**



Schlumberger

Motivation

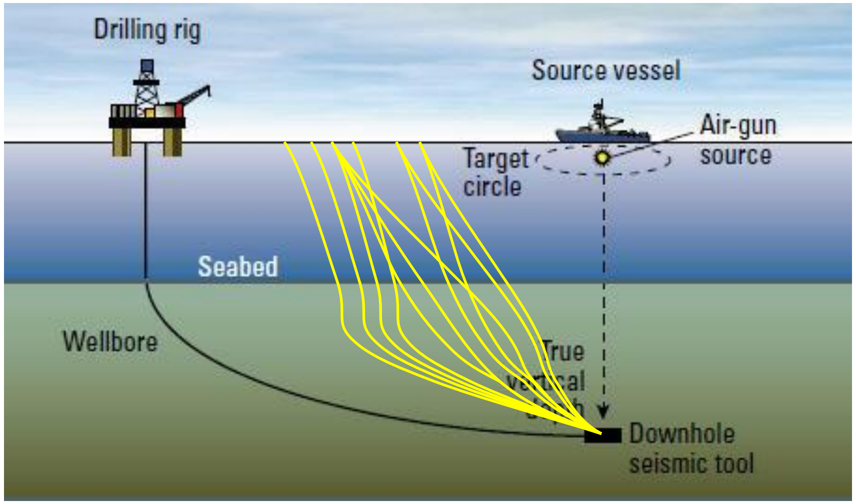
- A more expensive borehole seismic survey is envisaged to gather more information about the zone of interest
- Problem: How to design an optimal survey that maximizes the information gathered



Schlumberger

Current Approach

- Use reciprocity: Shoot rays from the anticipated location of the receiver
- Locate the areas with the highest concentration of rays on the surface
- Put your receivers at these locations



Not optimal !

Problem Formulation

$$g : \mathcal{M} \times \Xi \rightarrow \mathcal{D}$$

$$\left(m;\xi\right) \rightarrow d = g\left(m;\xi\right)$$

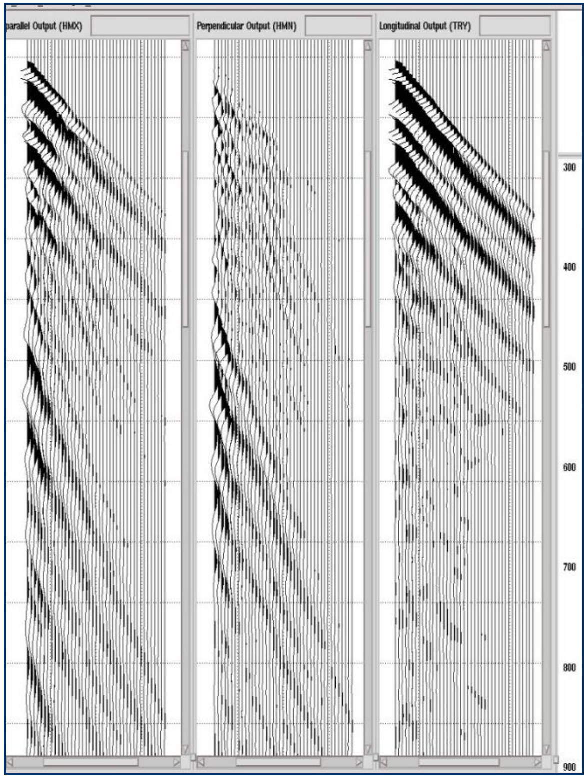
g: nonlinear continuous mapping.

Problem Formulation

$$g : \mathcal{M} \times \Xi \rightarrow \mathcal{D}$$

$$(m;\xi) \rightarrow d = g(m;\xi)$$

g: nonlinear continuous mapping.



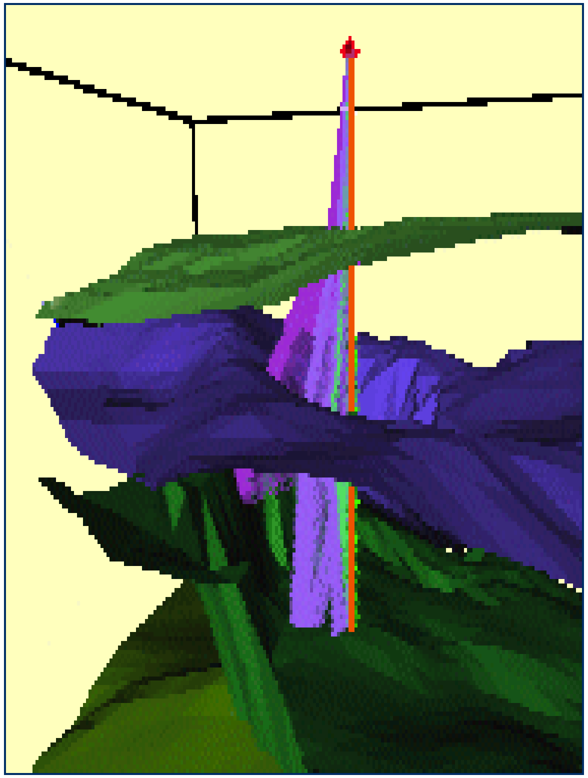
d: data

Problem Formulation

$$g : \mathcal{M} \times \Xi \to \mathcal{D}$$

$$(m;\xi) \to d = g(m;\xi)$$

g: nonlinear continuous mapping.



m: model

Schlumberger

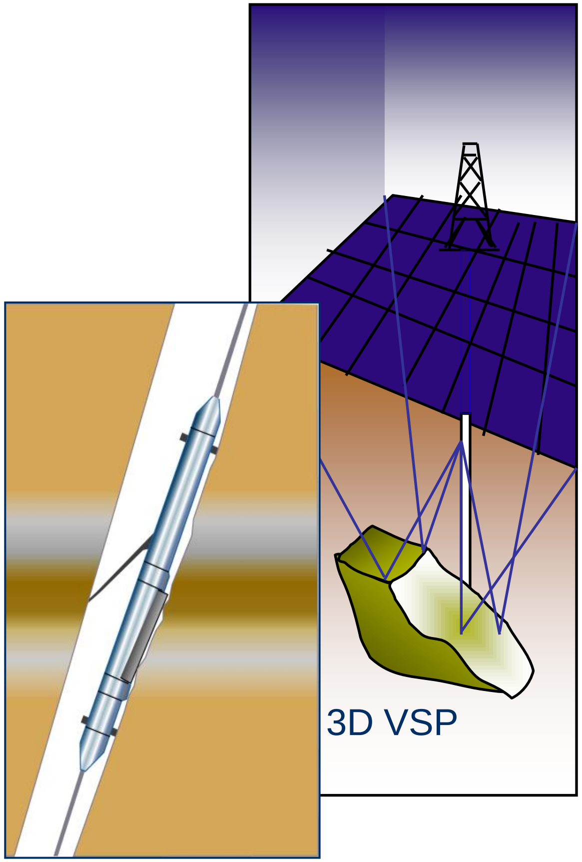
Problem Formulation

$$g : \mathcal{M} \times \Xi \to \mathcal{D}$$

$$(m;\xi) \to d = g(m;\xi)$$

g: nonlinear continuous mapping.

ξ: survey control parameters.



Not inversion !

Forecast optimization for ξ

Schlumberger

Problem Formulation

- **This is a complex nonlinear global optimization problem.**
- **Characteristic of current OEDs:**
  - Use of global, stochastic, search methods: simulated annealing, genetic algorithms, etc.
- **Drawback:**
  - Long computation times: days, weeks.

We are not doing that !

Our Approach

- **Discretization and Linearization**

$$d = g(m, \xi) \rightarrow \mathbf{d} - \mathbf{d}_0 = \mathbf{G}.(\mathbf{m} - \mathbf{m}_0)$$

ξ : survey control parameters

$\mathbf{d_0}$: a reference data vector

$\mathbf{m_0}$: a reference model vector

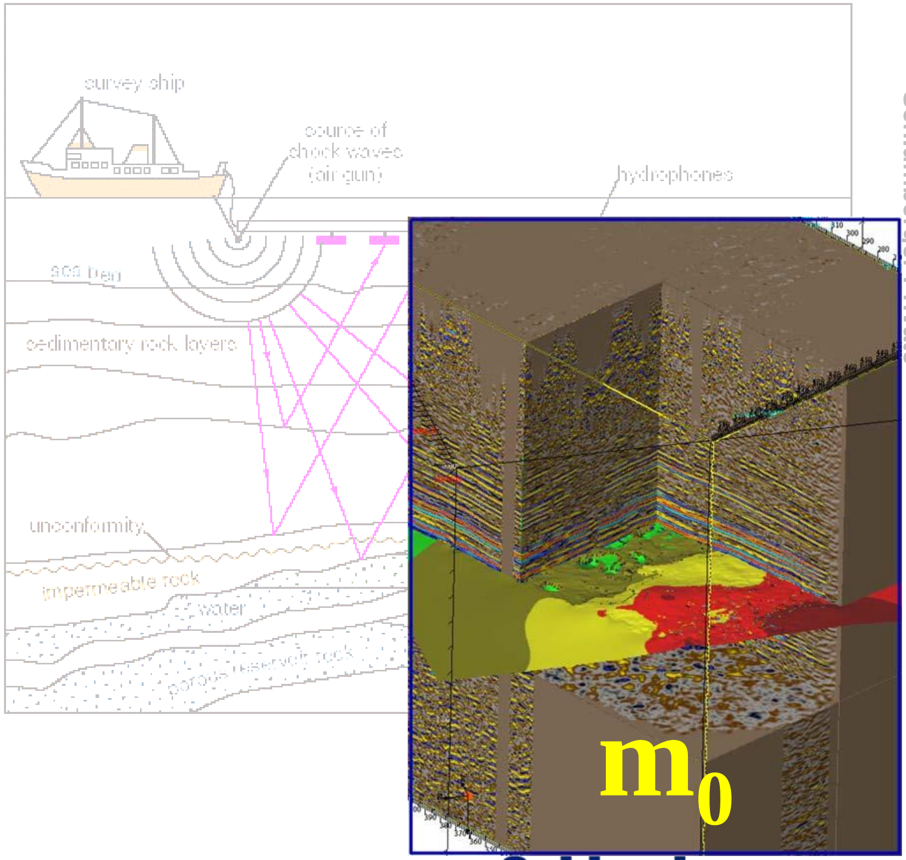
$\mathbf{G}$: *D*x*M* sensitivity matrix ($G_{ij} \equiv \left.\frac{\partial g(\mathbf{m}, \xi_i)}{\partial m_j}\right|_{\mathbf{m}=\mathbf{m}_0}$)

The problem is still nonlinear in ξ !

Schlumberger

Motivation

- A surface seismic survey is performed
- **The interpreted seismic data yields a primary model of the subsurface**
- The interpretation of the data reveals the presence of a zone of interest



Schlumberger

Sensitivity Kernels and Observations

$$\mathbf{G} = \begin{bmatrix} G_{11} & G_{12} & \cdots & G_{1M} \\ G_{21} & G_{22} & \cdots & G_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ G_{D1} & G_{D2} & \cdots & G_{DM} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_1^T \\ \mathbf{g}_2^T \\ \vdots \\ \mathbf{g}_D^T \end{bmatrix}$$

Each row defines a candidate measurement

Non-Bayesian *D*-optimality Criterion

- **Wald (1948):**

$$\arg\max_{\xi \in \Xi} \left| \mathbf{G}^T \mathbf{G} \right|$$

i.e., maximize the determinant of the likelihood covariance matrix.

Sequential Augmentation of an Existing Experiment

Dykstra (1971): What is the best next $\mathbf{g}$?

$$\arg\max_{\xi \in \Xi} \frac{\left| \mathbf{G}_{n+1}^T \left(\mathbf{C}_d \right)_{n+1}^{-1} \mathbf{G}_{n+1} \right|}{\left| \mathbf{G}_n^T \left(\mathbf{C}_d \right)_n^{-1} \mathbf{G}_n \right|}$$

diagonal

$$\mathbf{G}_{n+1} = \begin{bmatrix} \mathbf{G}_n \\ \mathbf{g}_{n+1}^T \end{bmatrix}$$

Augmented experiment

Base experiment (could be empty)

Sensitivity kernel of the candidate data station

Non-Bayesian *D*-optimality Criterion

- **Coles & Morgan (2008):**

$$\arg\max_{\xi \in \Xi} \frac{\left| \mathbf{G}_{n+1}^{T} \left(\mathbf{C}_{d} \right)_{n+1}^{-1} \mathbf{G}_{n+1} \right|}{\left| \mathbf{G}_{n}^{T} \left(\mathbf{C}_{d} \right)_{n}^{-1} \mathbf{G}_{n} \right|}$$

diagonal

  - g_{n+1} is maximally orthogonal to the space of the base experiment.
  - Fast Gram-Schmidt orthogonalization.

What is Missing?

- **The non-Baysian OED criterion**
  - does not account for model uncertainties (and correlated data uncertainties)
  - is incapable of automatically focusing on a particular area of interest

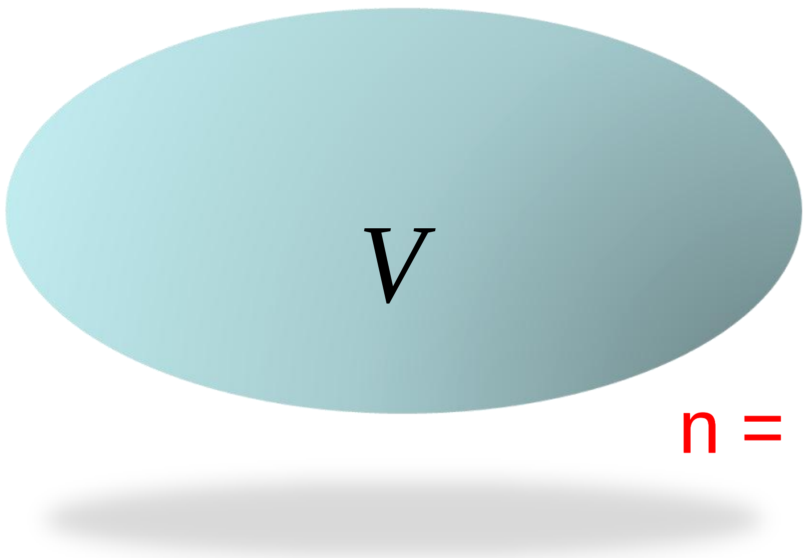
Our Bayesian *D*-optimality Criterion

- **Q: What is the best next $\mathbf{g}$?**

- **A:** $\arg\max_{\xi \in \Xi} \frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+1}|}$

- **Why?**

$$|\mathbf{C}| \sim V$$



n = 2

Confidence ellipsoid

Important Remark

- The sequential approach corresponds to a greedy algorithm
- It is not guaranteed to reach the global optimum
- It is guaranteed, however, to yield a very good approximation of the global optimum

Bayesian Augmentation

- **If the only known moments of the uncertainties are the mean and covariance → optimal choice: Gaussian distributions**

$$\mathbf{C}_{n+1} = \left[\mathbf{G}_{n+1}^{T} (\mathbf{C}_{d}^{-1})_{n+1} \mathbf{G}_{n+1} + \mathbf{C}^{-1} \right]^{-1}$$

Posterior model covariance matrix

Prior data covariance matrix

Prior model covariance matrix

General Result for Non-Diagonal Data Covariance Matrix

$$\left(\mathbf{C}_d\right)_{n+1} = \begin{bmatrix} \left(\mathbf{C}_d\right)_n & \mathbf{c}_{n+1} \\ \mathbf{c}_{n+1}^T & \sigma_{n+1}^2 \end{bmatrix}$$

$$\frac{\left|\mathbf{C}_n\right|}{\left|\mathbf{C}_{n+1}\right|} = \left(1+\mathbf{g}_{n+1}^T\mathbf{A}^{-1}\mathbf{h}\right)\left|\,\mathbf{I}+\mathbf{B}\mathbf{G}_n\mathbf{C}_n\mathbf{G}_n^T\right|\left[1+\mathbf{k}^T\left(\mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\mathbf{h}\mathbf{g}_{n+1}^T\mathbf{A}^{-1}}{1+\mathbf{g}_{n+1}^T\mathbf{A}^{-1}\mathbf{h}}\right)\mathbf{g}_{n+1}\right]$$

$$\mathbf{A} \equiv \mathbf{C}_n^{-1} + \mathbf{G}_n^T\mathbf{B}\mathbf{G}_n$$

$$\phi_{n+1} \equiv 1-\sigma_{n+1}^{-2}\mathbf{c}_{n+1}^T\left(\mathbf{C}_d\right)_n^{-1}\mathbf{c}_{n+1}$$

$$\mathbf{B} \equiv \sigma_{n+1}^{-2}\phi_{n+1}^{-1}\left(\mathbf{C}_d\right)_n^{-1}\mathbf{c}_{n+1}\mathbf{c}_{n+1}^T\left(\mathbf{C}_d\right)_n^{-1}$$

$$\mathbf{h} \equiv \sigma_{n+1}^{-2}\phi_{n+1}^{-1}[\mathbf{g}_{n+1}-\mathbf{G}_n^T\left(\mathbf{C}_d\right)_n^{-1}\mathbf{c}_{n+1}]$$

$$\mathbf{k}^T \equiv -\sigma_{n+1}^{-2}\phi_{n+1}^{-1}\mathbf{c}_{n+1}^T\left(\mathbf{C}_d\right)_n^{-1}\mathbf{G}_n$$

Simplifying Assumption

- **Uncorrelated data:**

$$\mathbf{C}_d = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_D^2 \end{bmatrix}$$

Data covariance matrix

Schlumberger

Bayesian OED

- **For uncorrelated data**

$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+1}|} = 1 + \sigma_{n+1}^{-2} \mathbf{g}_{n+1}^T \mathbf{C}_n \mathbf{g}_{n+1}$$

$$\mathbf{C}_n = \mathbf{C}_{n-1} - \frac{\sigma_n^{-2} \mathbf{C}_{n-1} \mathbf{g}_n \mathbf{g}_n^T \mathbf{C}_{n-1}}{1 + \sigma_n^{-2} \mathbf{g}_n^T \mathbf{C}_{n-1} \mathbf{g}_n}$$

Another Look at the Proposed Objective Function

$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+1}|} - 1 = \sigma_{n+1}^{-2} \mathbf{g}_{n+1}^T \mathbf{C}_n \mathbf{g}_{n+1}$$

$$\equiv \boldsymbol{\gamma}_{n+1}^T \mathbf{C}_n \boldsymbol{\gamma}_{n+1}$$

$$\equiv \left\langle \boldsymbol{\gamma}_{n+1}^T, \boldsymbol{\gamma}_{n+1} \right\rangle_{\mathbf{C}_n}$$

C-norm

σ-scaled sensitivity kernels

$$\boldsymbol{\gamma}_{n+1} \equiv \sigma_{n+1}^{-1} \mathbf{g}_{n+1}$$

How About the non-Bayesian Criterion?

- **The Dykstra-Coles-Morgan criterion maximizes this norm**

$$\frac{\left|\mathbf{G}_{n+1}^{T}\left(\mathbf{C}_{d}\right)_{n+1}^{-1}\mathbf{G}_{n+1}\right|}{\left|\mathbf{G}_{n}^{T}\left(\mathbf{C}_{d}\right)_{n}^{-1}\mathbf{G}_{n}\right|}=1+\sigma_{n+1}^{-2}\mathbf{g}_{n+1}^{T}(\mathbf{G}_{n}^{T}\mathbf{G}_{n})^{-1}\mathbf{g}_{n+1}$$

$$=\left\langle \boldsymbol{\gamma}_{n+1}^{T},\boldsymbol{\gamma}_{n+1}\right\rangle_{\left(\mathbf{G}_{n}^{T}\mathbf{G}_{n}\right)^{-1}}$$

Rank-k Update Versus k Rank-one Updates

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{\Gamma}_1^T \\ \mathbf{\Gamma}_2^T \\ \vdots \\ \mathbf{\Gamma}_k^T \end{bmatrix}, \quad (\mathbf{C}_d)_{n+k} = \begin{bmatrix} (\mathbf{C}_d)_n & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_1^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_k^2 \end{bmatrix}$$

Rank-k update →

$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+k}|} = |\mathbf{S}|^{-1} \left|\mathbf{S} + \mathbf{\Gamma}\mathbf{C}_n\mathbf{\Gamma}^T\right|$$

k rank-one updates →

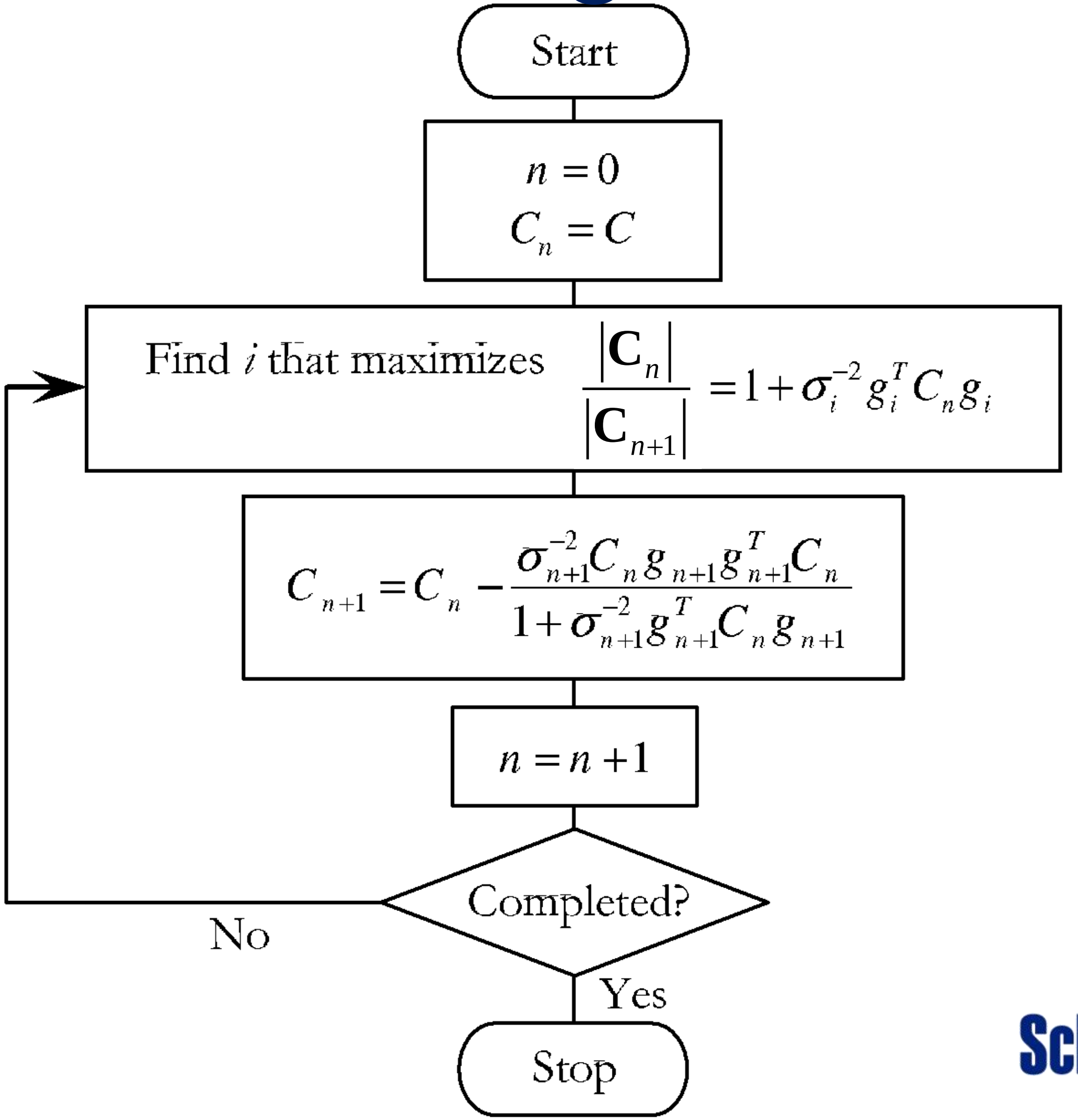
$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+k}|} = \prod_{l=1}^{k}\left(1 + \sigma_{k-l+1}^{-2}\mathbf{\Gamma}_{k-l+1}\mathbf{\Lambda}_{k-l}^{-1}\mathbf{\Gamma}_{k-l+1}^T\right)$$

$$\mathbf{\Lambda}_{k-l} \equiv \mathbf{C}_n^{-1} + \sum_{r=1}^{k-l}\sigma_r^{-2}\mathbf{\Gamma}_r\mathbf{\Gamma}_r^T$$

Both

$$O\left(M^2k\right), k \ll M$$

Basic Algorithm



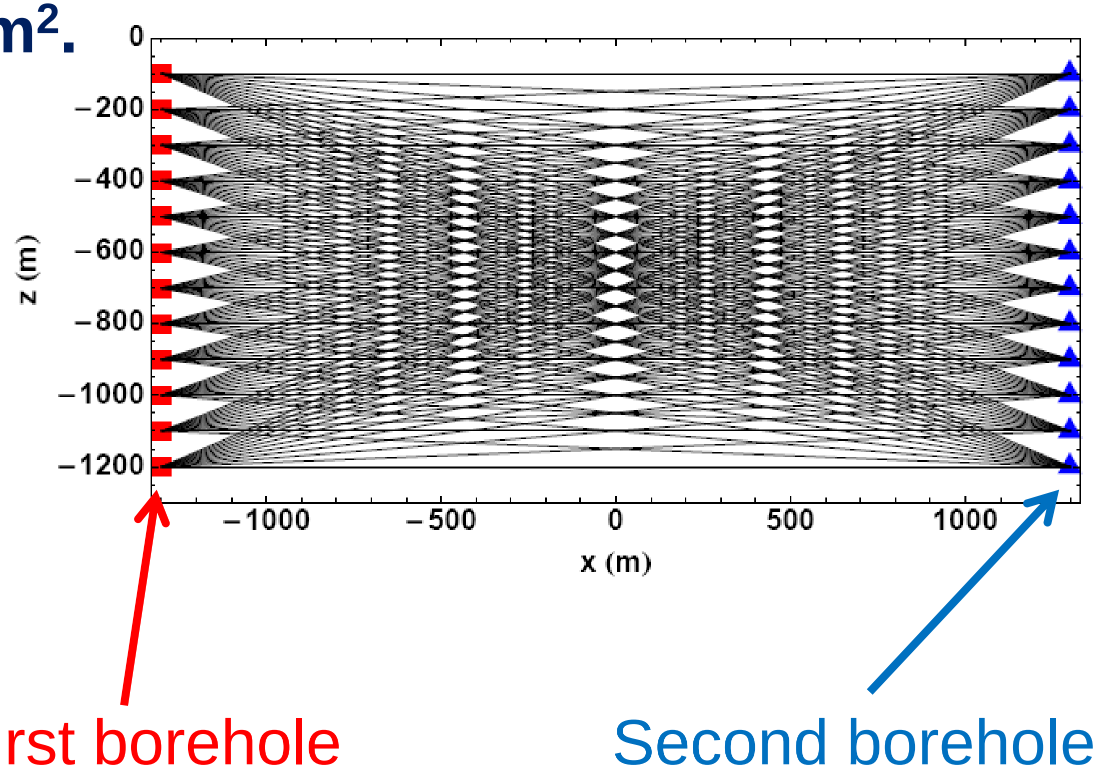
Numerical Results

- **Seismic Tomography**
  1. Proof-of-concept Examples:
     a) Homogeneous velocity model with constant prior uncertainty.
     b) Homogeneous velocity model with a larger uncertainty in a localized area.
     c) Heterogeneous velocity model with a larger uncertainty in a localized area.
  2. Realistic model (Pluto1.5).

Schlumberger

Homogeneous Model

- **Grid: 2600x1300 m^2.**
- **V = 2400 m/s.**
- **12X12 source-receiver pairs.**

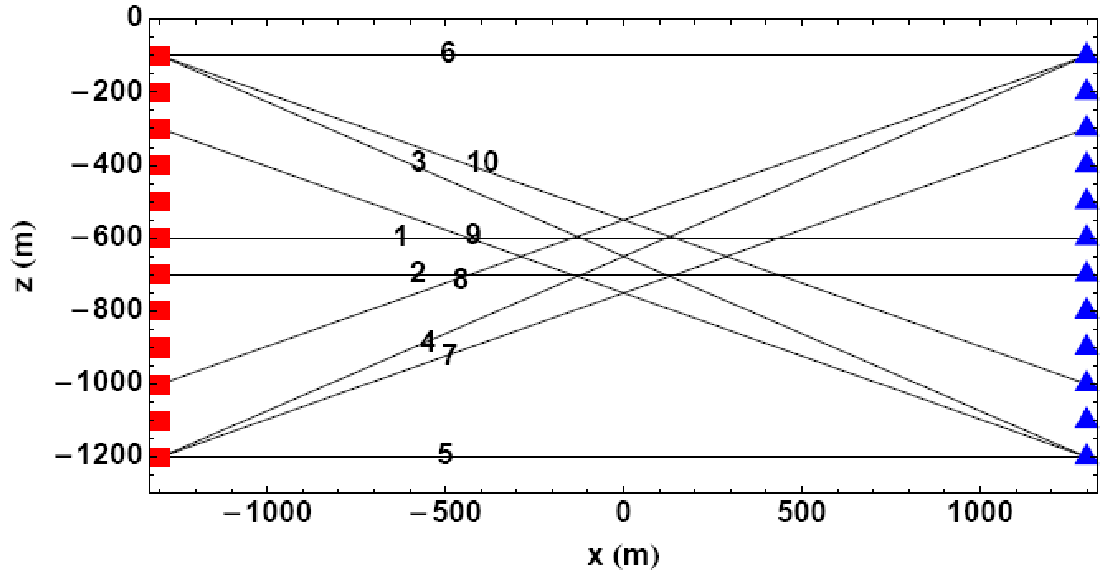


Sources.

Receivers.

Homogeneous Model

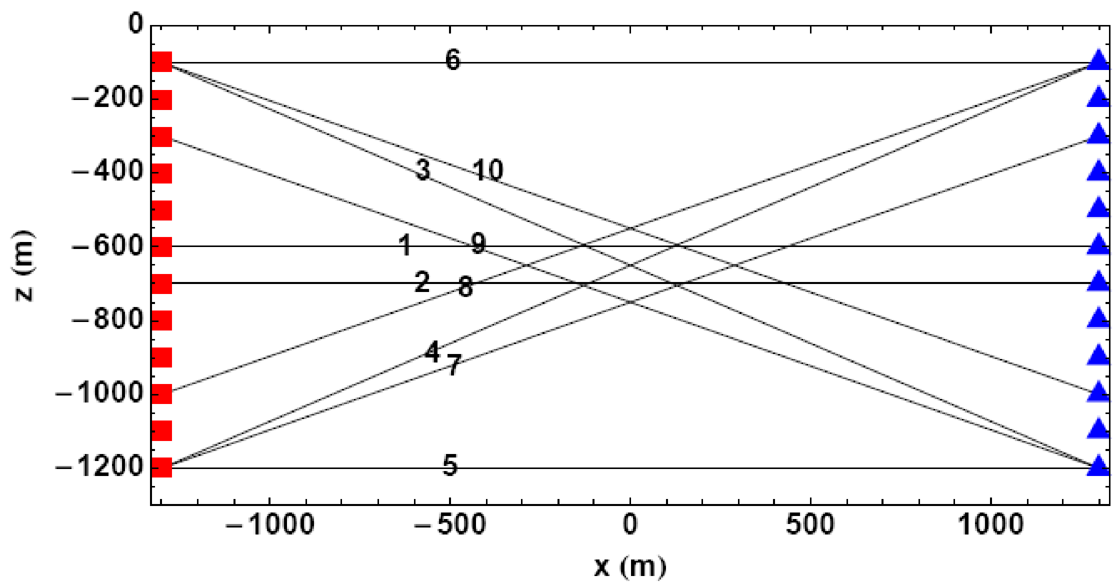
Non-Bayesian



Bayesian

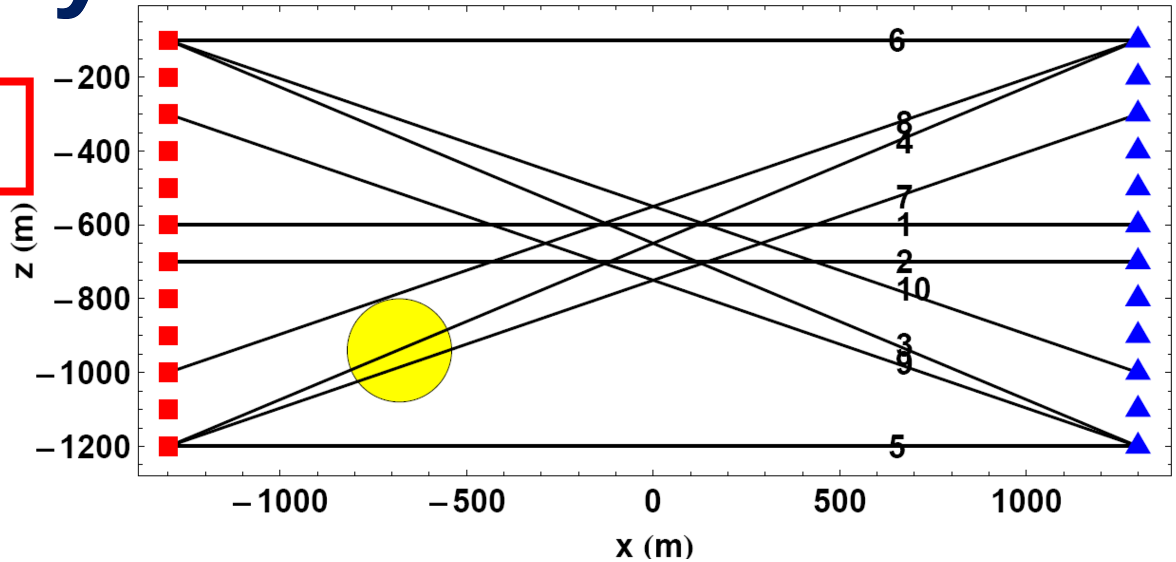
$$\mathbf{C}_d = \left(2\ \mathrm{ms}\right)^2 \mathbf{I}$$

$$\mathbf{C} = \left(10^{10}\ \mathrm{s/m}\right)^2 \mathbf{I}$$



Homogeneous Model with Larger Uncertainty in a Localized Area

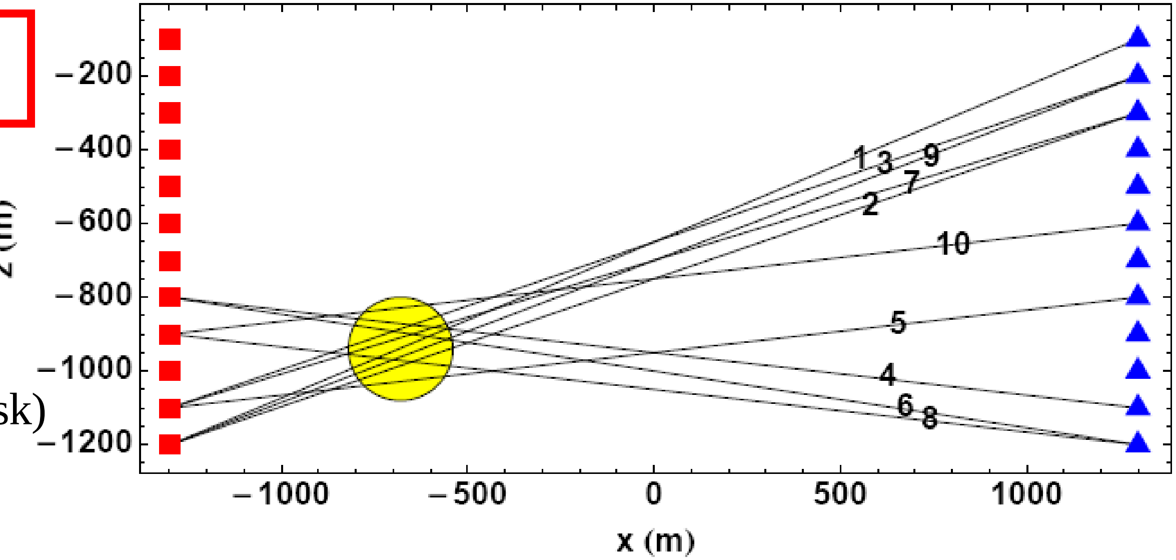
Non-Bayesian



Bayesian

$$\mathbf{C}_d = (2\ \text{ms})^2\,\mathbf{I}$$

$$\mathbf{C} = \begin{cases} (10^{-3}\ \text{s/m})^2\,\mathbf{I}\ \text{(disk)} \\ (10^{-4}\ \text{s/m})^2\,\mathbf{I}\ \text{(outside the disk)} \end{cases}$$



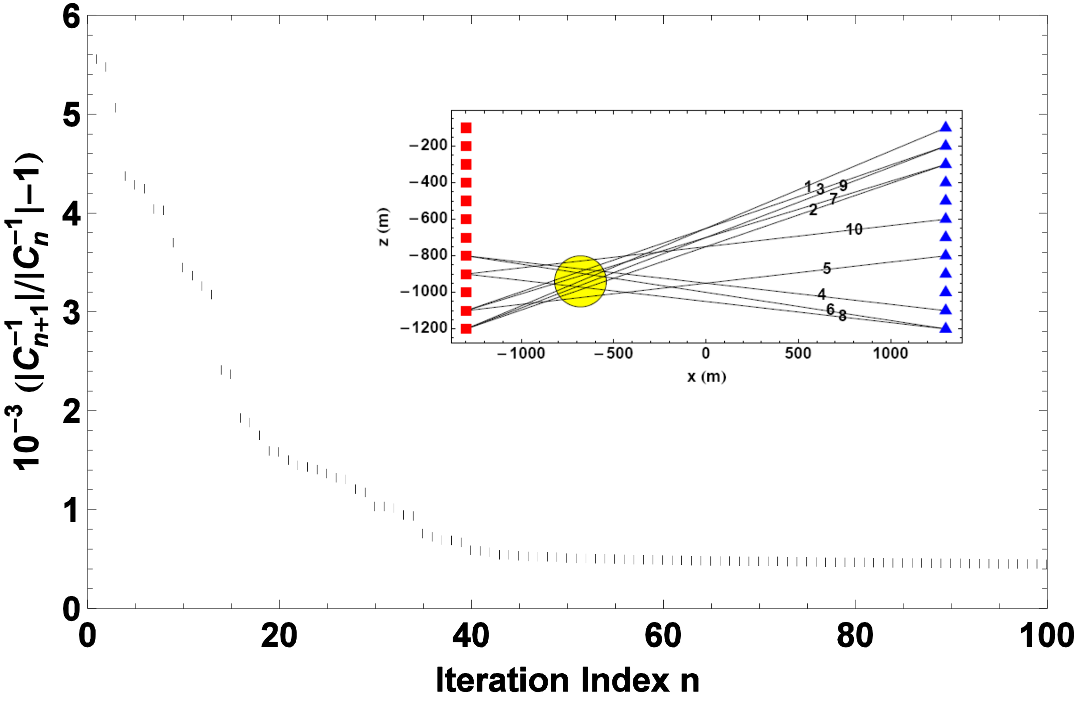
Homogeneous Model with Larger Uncertainty in a Localized Area

Bayesian

Convergence of the norm

$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+1}|} - 1$$

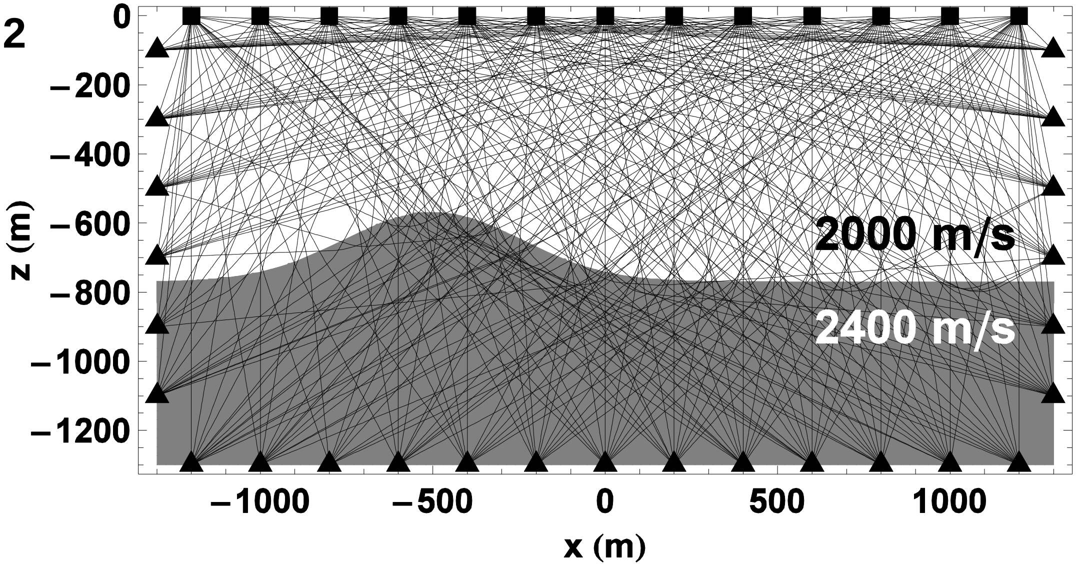
versus the iteration index



Schlumberger

Heterogeneous Model

- **Grid: 2600x1300 m^2**
- **13 sources.**
- **25 receivers.**
- **325 source-receiver pairs.**

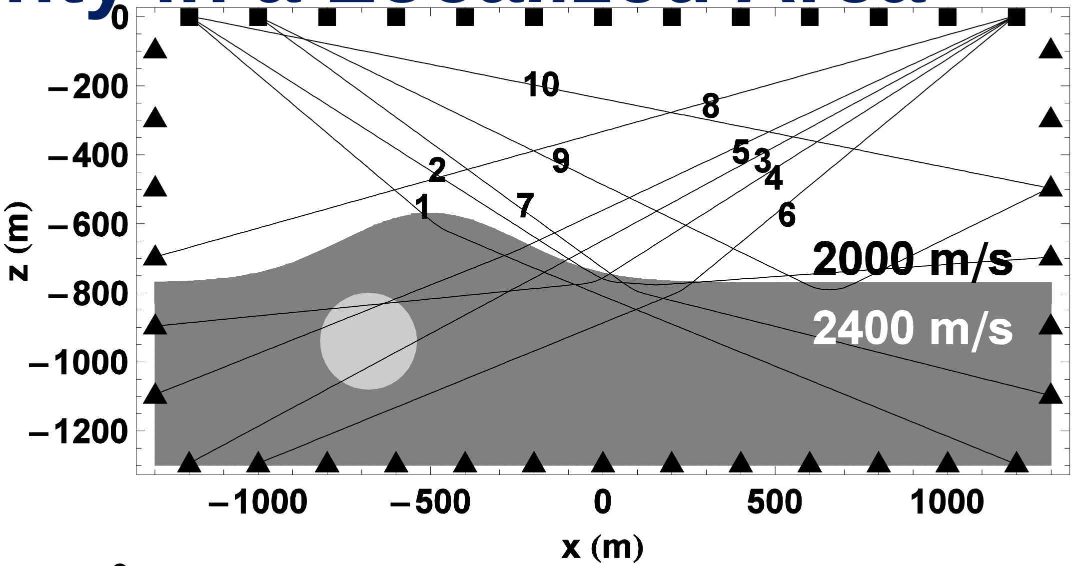


■ **Sources** ▲ **Receivers**

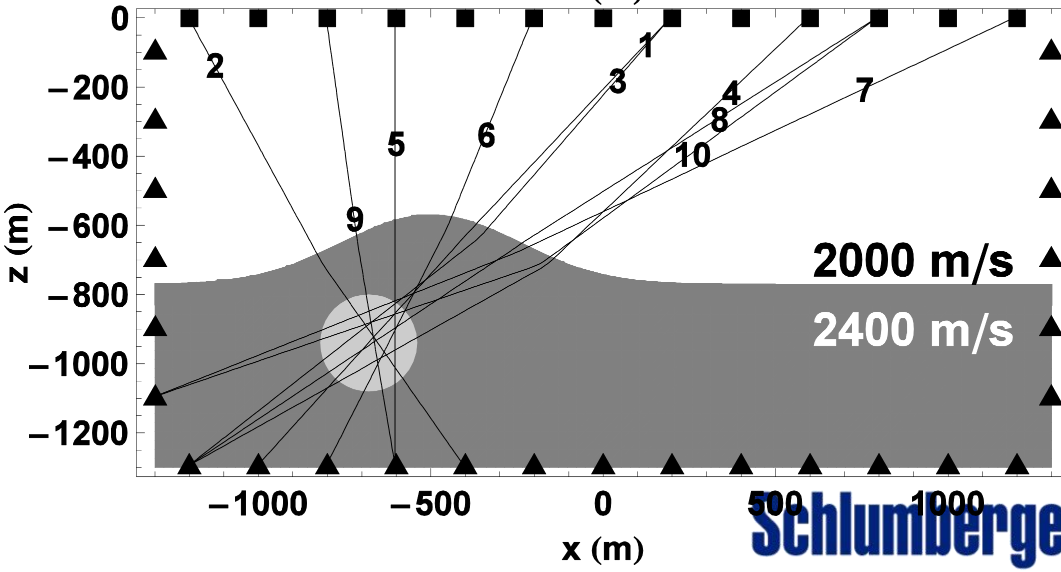
Schlumberger

Heterogeneous Model with a Larger Uncertainty in a Localized Area

Non-Bayesian



Bayesian



$$\mathbf{C}_d = (2\ \text{ms})^2\, \mathbf{I}$$

$$\mathbf{C} = \begin{cases} (10^{-3}\ \text{s/m})^2\, \mathbf{I}\ \text{(disk)} \\ (10^{-4}\ \text{s/m})^2\, \mathbf{I}\ \text{(outside the disk)} \end{cases}$$

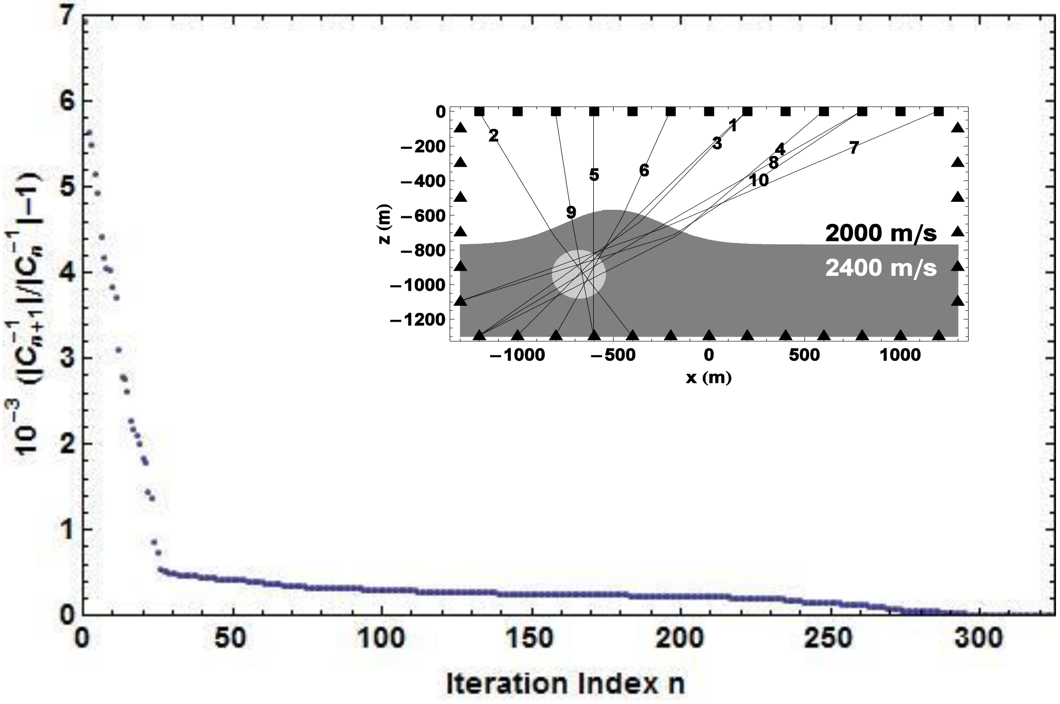
Heterogeneous Model with Larger Uncertainty in a Localized Area

Bayesian

Convergence of the norm

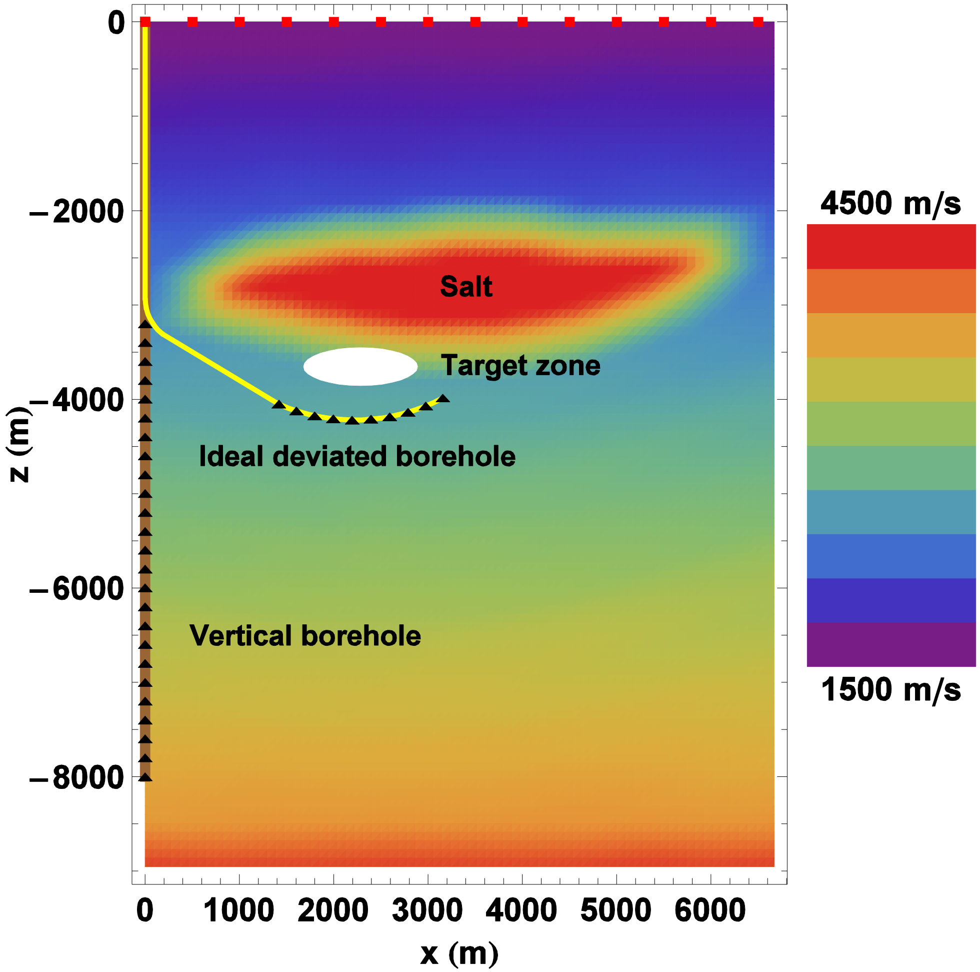
$$\frac{|\mathbf{C}_n|}{|\mathbf{C}_{n+1}|}-1$$

versus the iteration index



Schlumberger

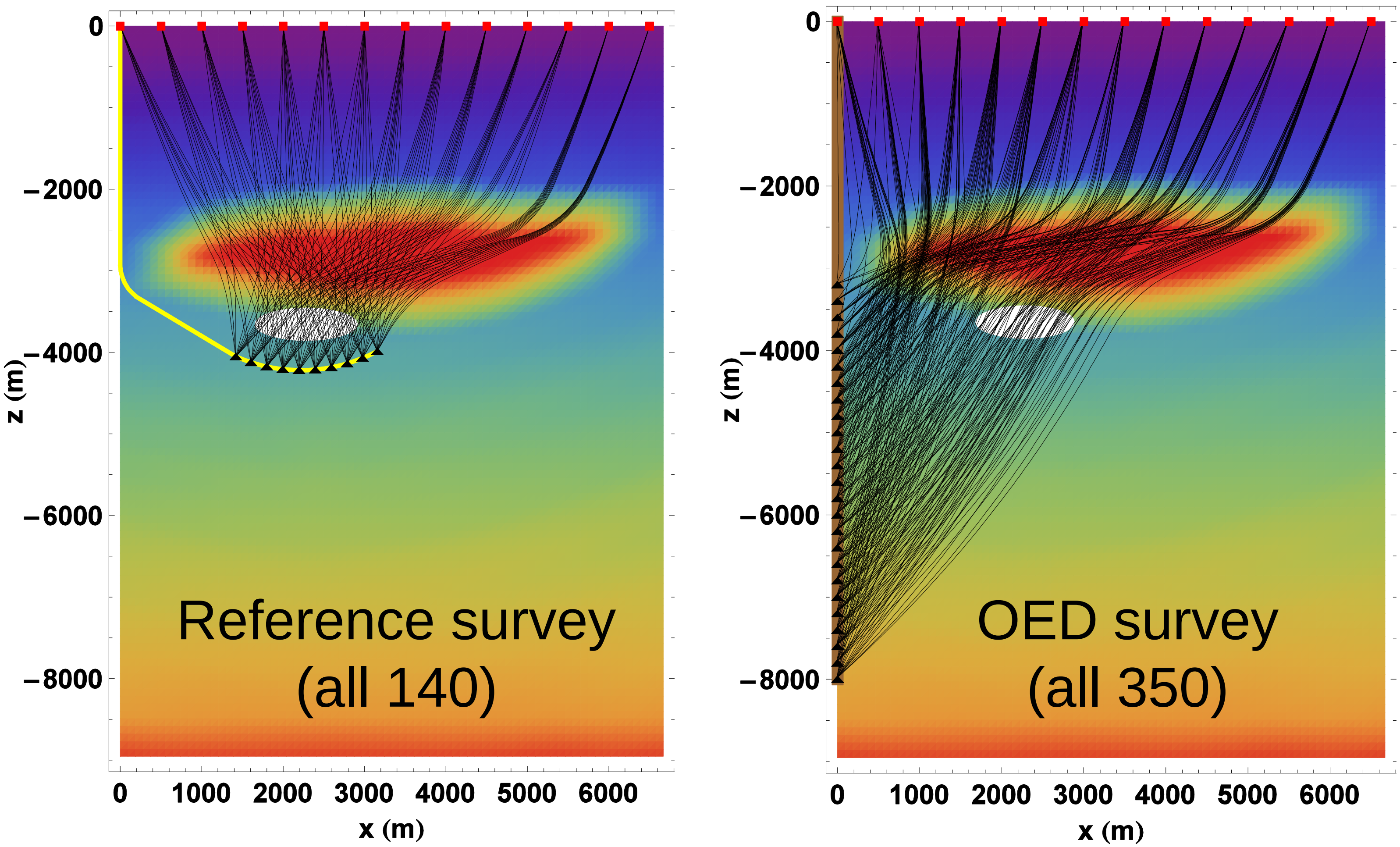
Realistic Model (Pluto 1.5)



$$\mathbf{C}_d = (2\ \text{ms})^2\,\mathbf{I}$$

$$\mathbf{C} = \begin{cases} (10^{-3}\ \text{s/m})^2\,\mathbf{I}\ \text{(disk)} \\ (10^{-4}\ \text{s/m})^2\,\mathbf{I}\ \text{(outside the disk)} \end{cases}$$

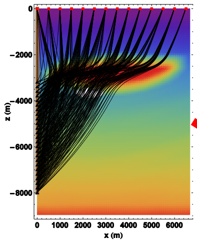
Competing Surveys



Schlumberger

Discriminator

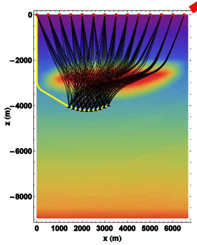
- **We define the quality discriminator**



$$\frac{\left|\mathbf{C}_n^{(OED)}\right|}{\left|\mathbf{C}^{(Ref)}\right|} = \begin{cases} >1 \\ <1 \end{cases}$$

>1: when the reference "ideal" survey outperforms the designed survey

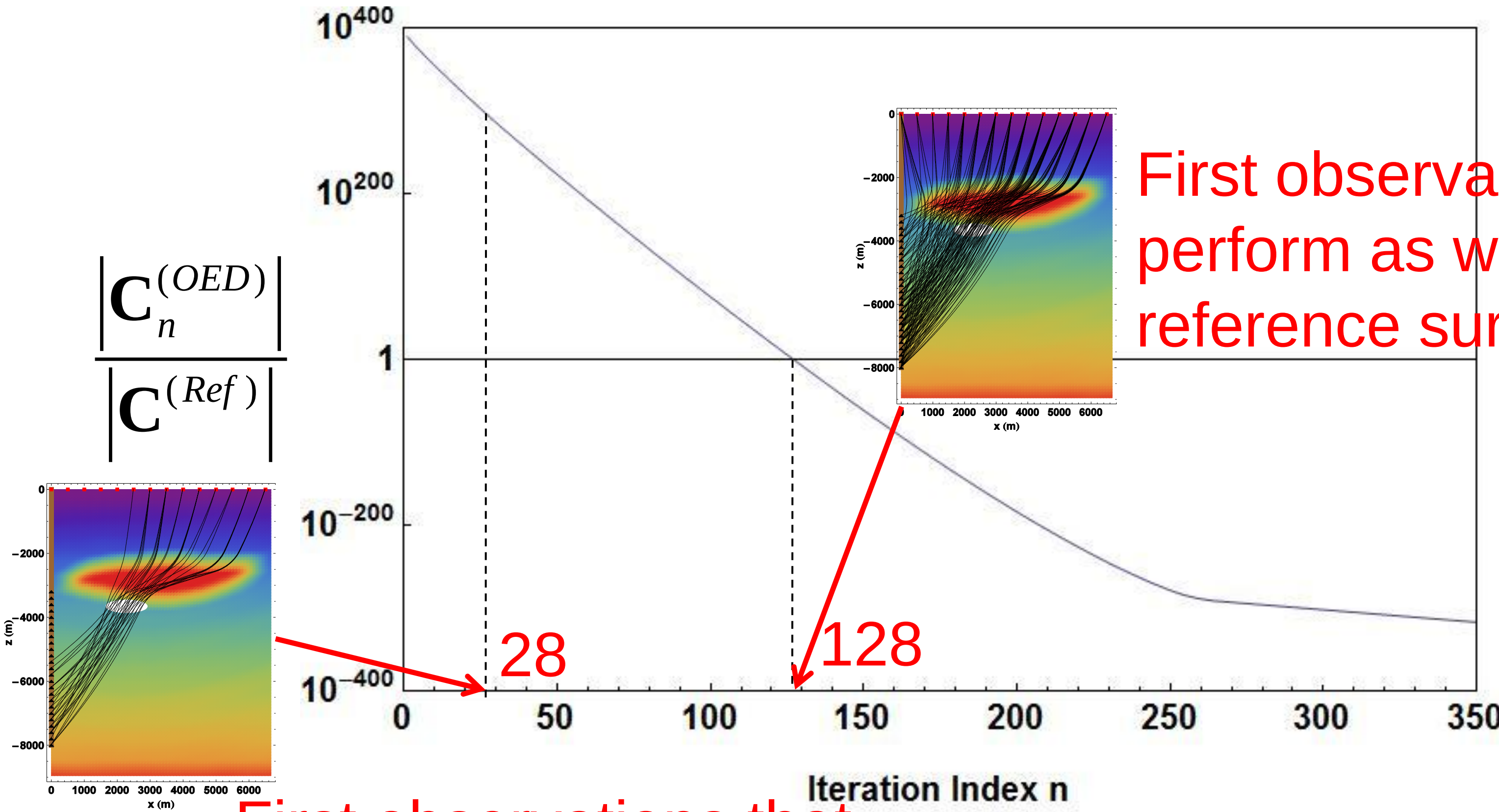
<1: otherwise



(all 140 s-r pairs)

Schlumberger

Discriminator vs. Iteration Index



First observations that perform as well as the reference survey

First observations that focus on the target zone

Summary

- We have described a Bayesian methodology for designing experiments
- This methodology takes advantage of available prior information and aims at reducing posterior uncertainties
- It allows us to focus on a zone of interest
- Although our algorithm is not guaranteed to arrive at the global optimal solution but it is fast

Questions

Schlumberger