Recent advances in full waveform inversion:

(a) data-driven versus model-driven strategies,
(b) truncated Newton approach,
(c) application to real Valhall dataset

A. Asnaashari[1,3], R. Brossier[1], C. Castellanos[2], V. Etienne[2], Y. Gholami[2], G. Hu[2],

L. Métivier[1], S. Operto[2], D. Pageot[2], V. Prieux[2], A. Ribodetti[2] and J. Virieux[1]

[1] ISTerre, CNRS-UJF, France, [2] Geoazur, CNRS-UNSA-IRD-OCA, France, [3] TOTAL E&P, Pau, France

http://seiscope.oca.eu

Seismic imaging of complex structures from multicomponent global offset data by full waveform inversion

FWI DIFFICULTIES

Full Waveform Inversion is a promising technique for macromodel (geodynamic) estimation (and downscaling extraction ?)

but it suffers for difficulties related to its potentially high resolution power.

The main one is the problem of local minima which may trap the optimization procedure and prevent the reconstruction

- Objective function definition
- Data preparation
- Model parametrisation and prior information
- Initial model construction

A hierarchy strategy from low frequency to high frequency promoted by Pratt and co-workers in '90s could overcome this difficulty for single parameter reconstruction.

How low should be the starting frequency?

Data-driven strategy

- Frequency decimation/filtering (Kolb et al, 1986; Bunks et al, 1995; Sirgue and Pratt, 2004)
- Time windowing (Kolb et al, 1986; Shin et al, 2002; Brenders and Pratt, 2007; Brossier et al, 2009a)
- Beam forming (Vigh and Starr, 2008; Brossier and Roux, 2011)
- Data hierarchy (Sears et al, 2008; Brossier et al, 2009b)

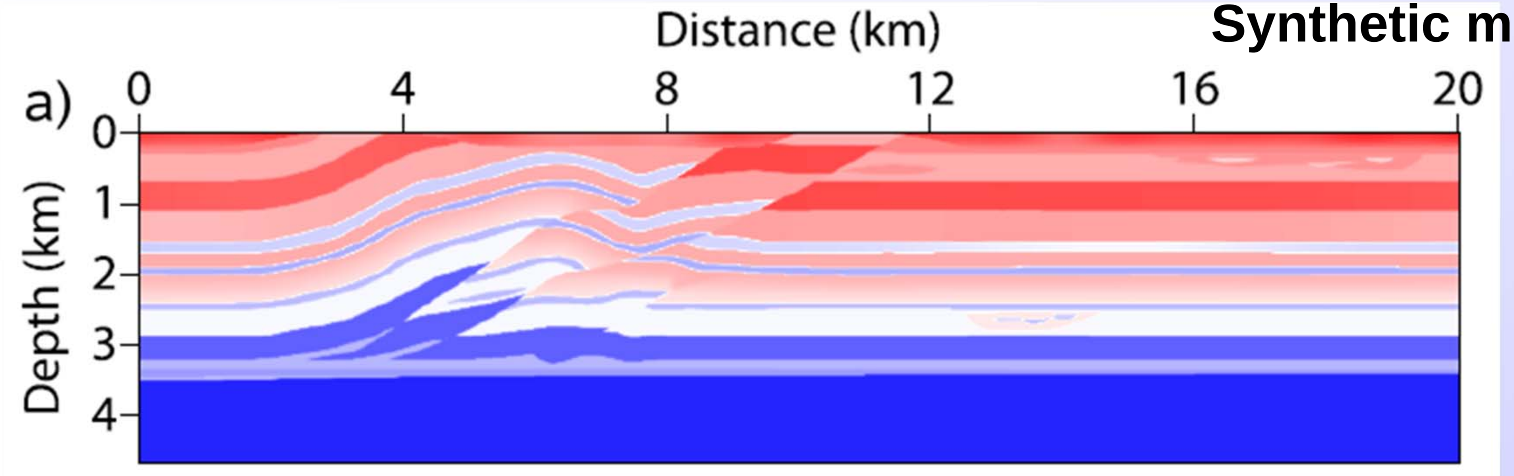
FWI has been considered as a purely data-driven technique

The information content in the data was supposed to be enough for stable imaging

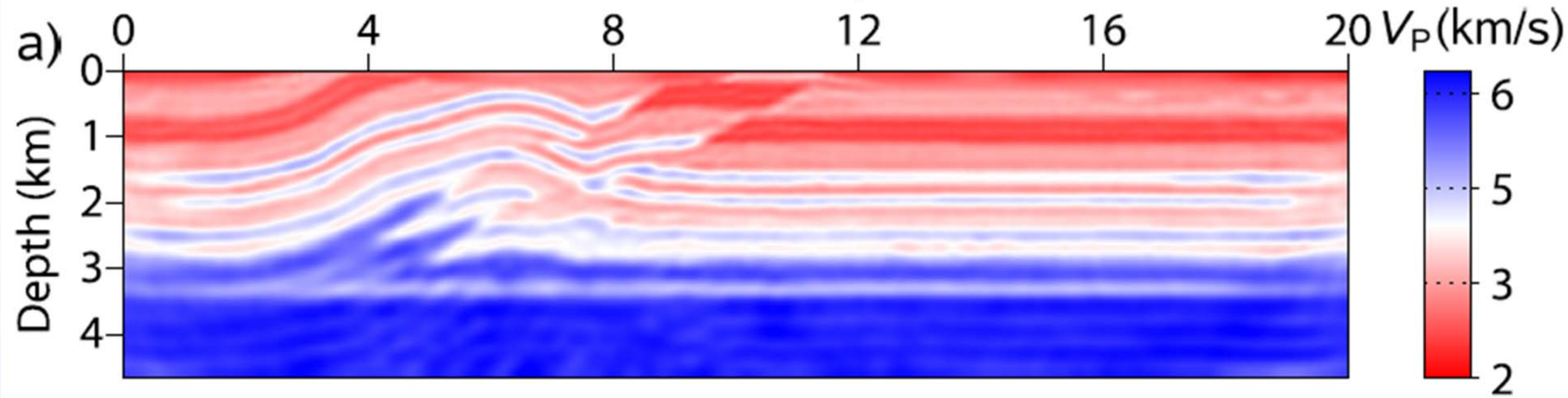
Sequential acoustic FWI

Five sequential frequencies 1.7; 2.5; 3.5, 4.7, 7.2 Hz

Synthetic model



Recovered model

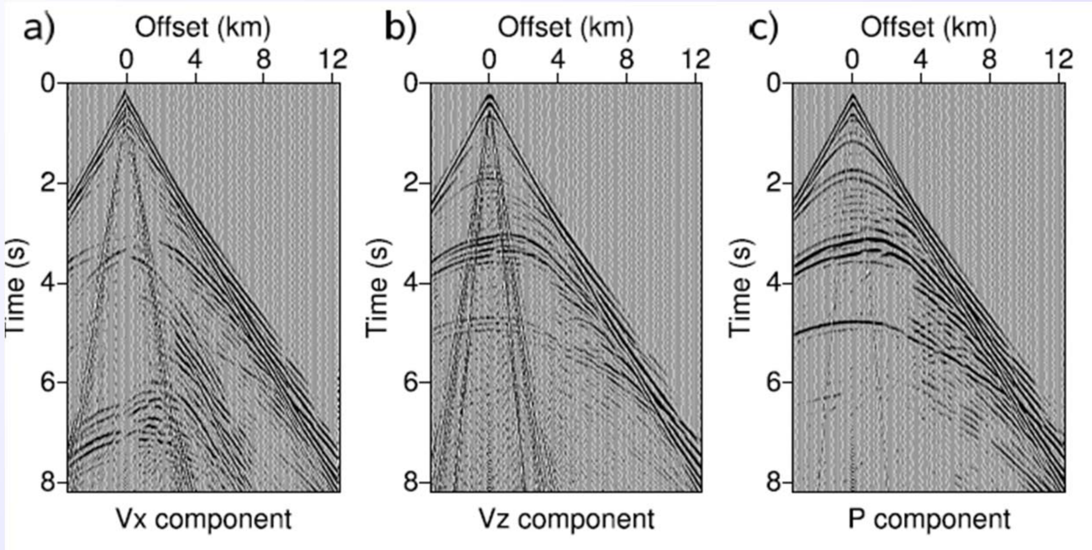


Similar to Pratt et co-workers results

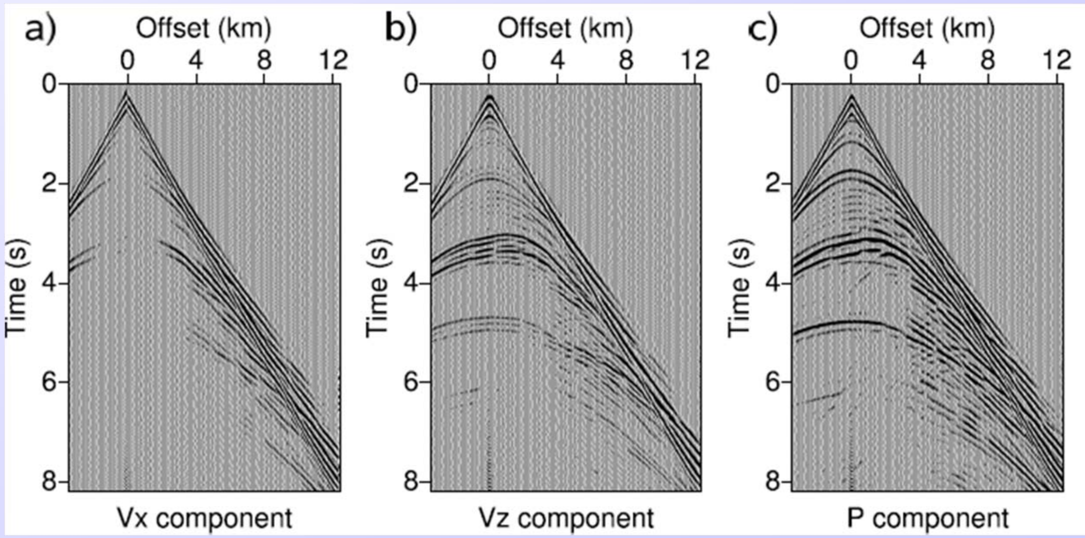
Sourbier et al (2008)

Data: complexity of seismograms

Elastic seismograms



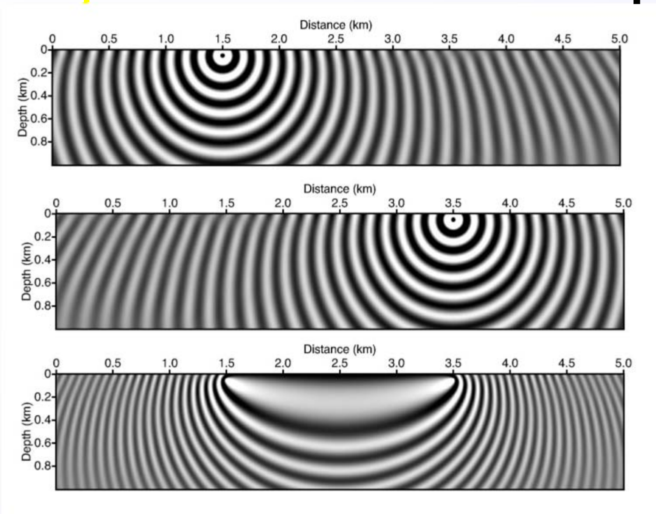
Acoustic seismograms

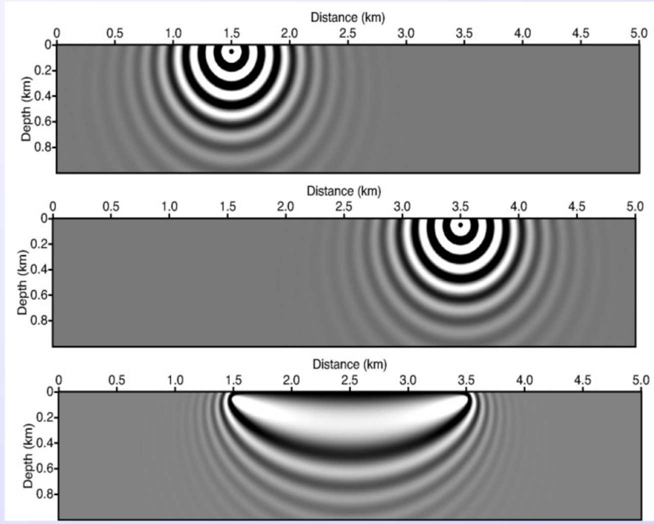


Free surface complexity effects

undamped

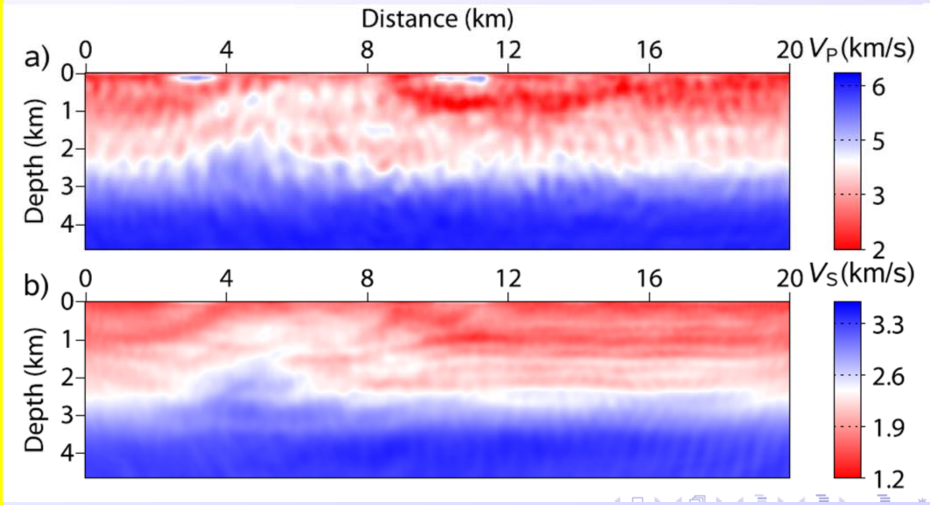
damped

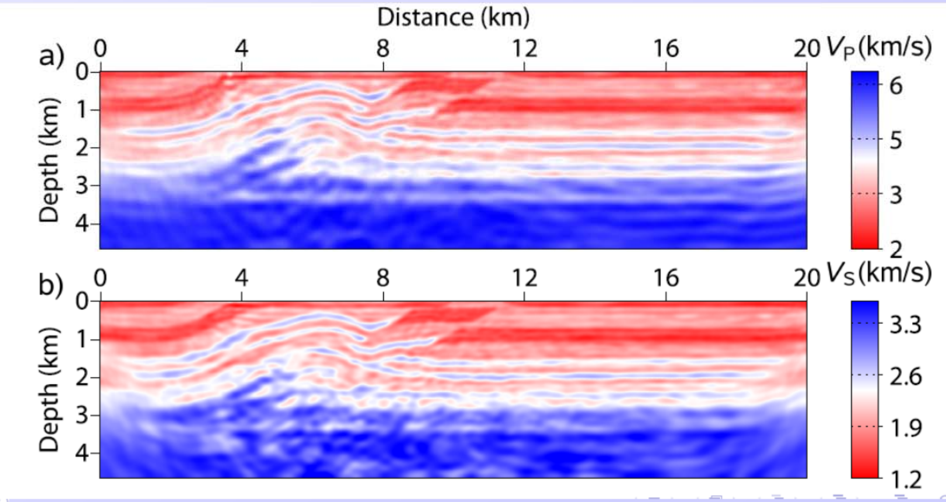




Single loop ω

Double loops ω & γ



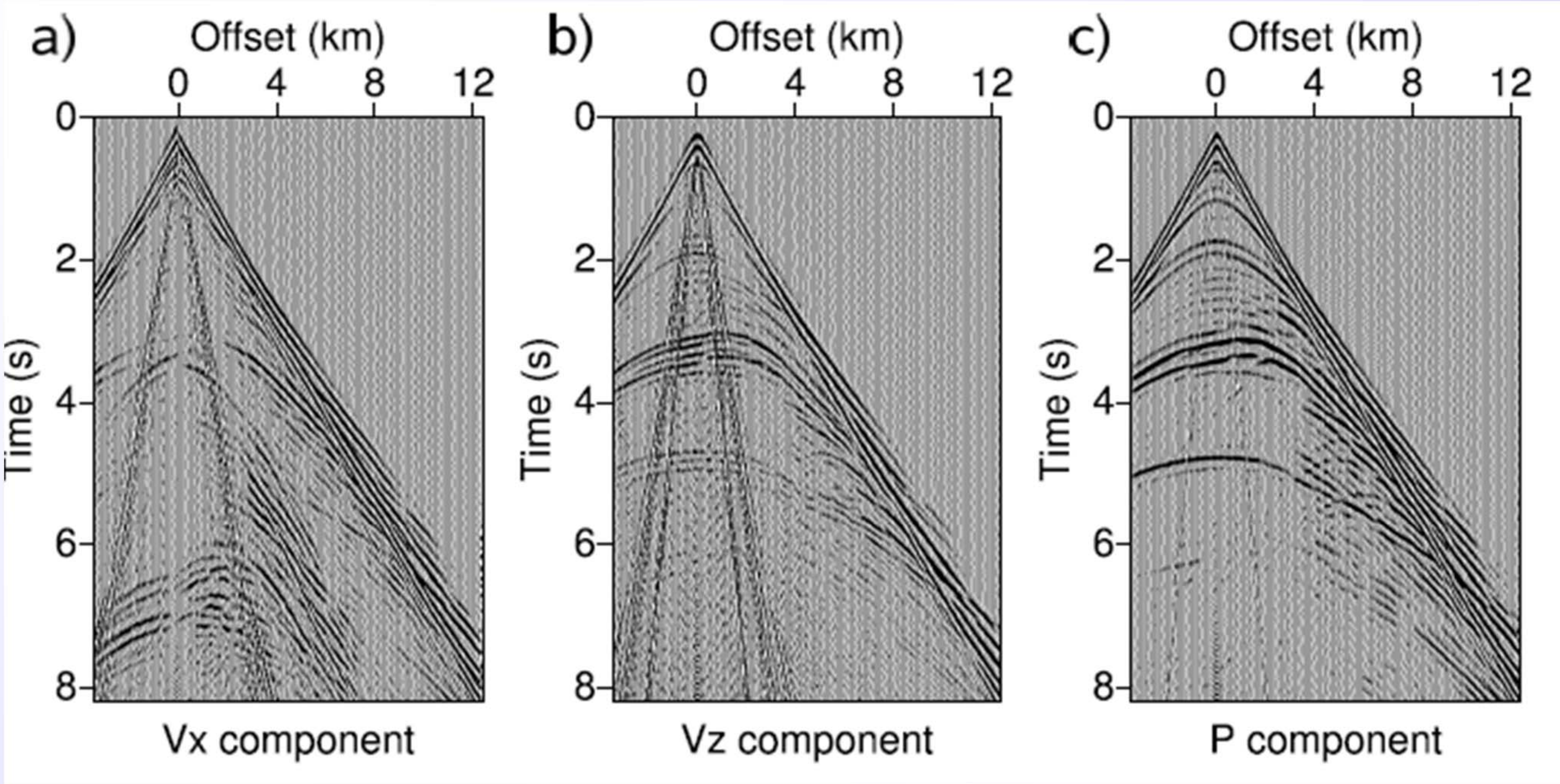


(Brossier et al, 2009a)

Elastic frequency-domain FWI

Valhall type dataset

Selection of components

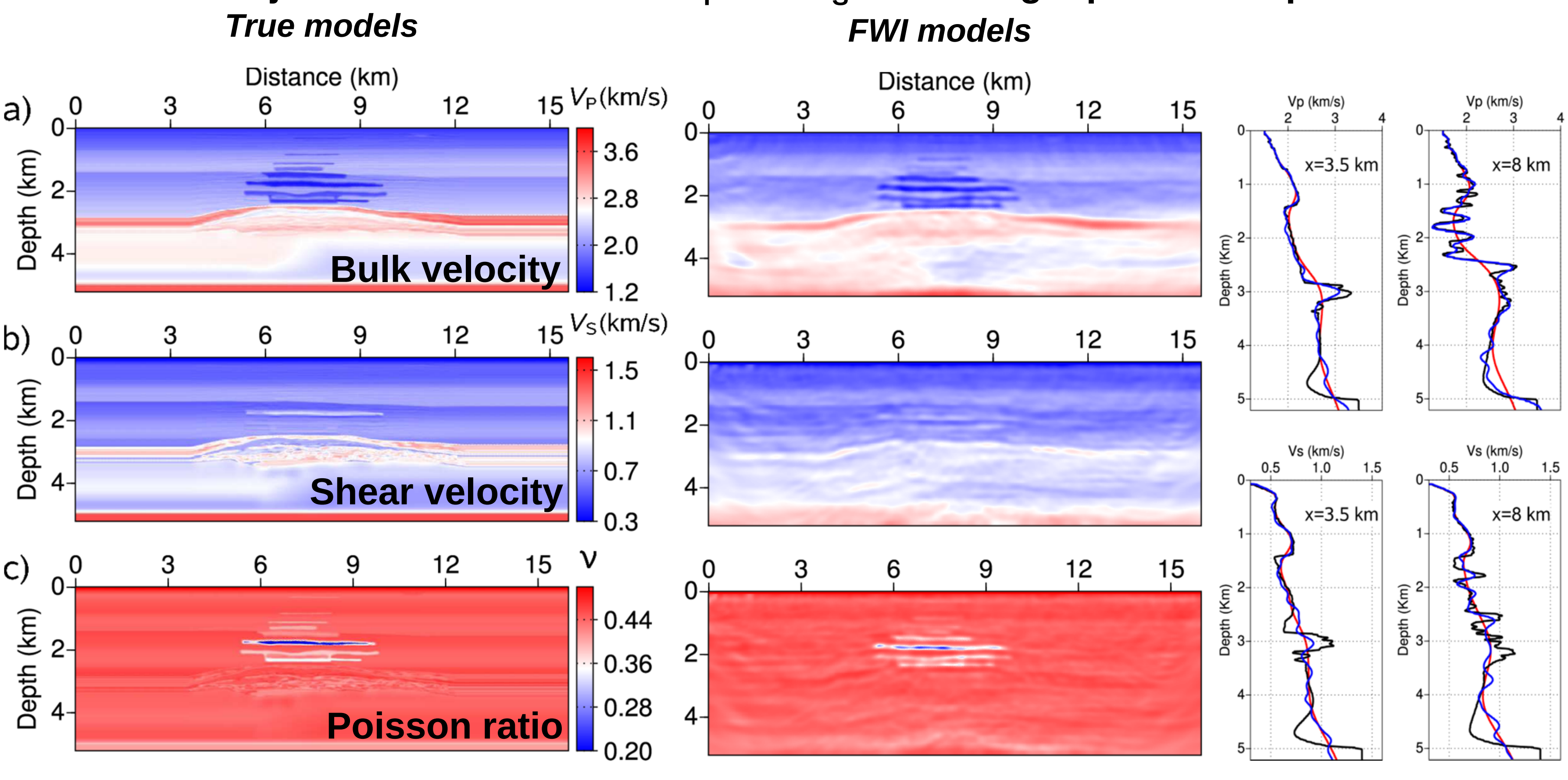


(polarisation … poster of Ho, Quest)

Data selection in FWI (*2D Synthetic Valhall dataset)*

First inversion of Vp from hydrophones and

joint reconstruction of V_P and V_S from the geophone components



(Brossier's work)

Partial conclusion on the data side

Presenting data with gradual complexity to the FWI engine prevents somehow to be trapped into local minima.

Filtering, windowing, beaming, transforming the data for proper convergence towards the global minimum is an efficient technique.

Potential other data manipulation (Laplace and complex Fourier transforms have been promoted: group of Pr. Shin and group of Pr. Pratt)

Optimisation: combining data misfit and model misfit

Inverse theory introduces the hyperparameter $\mathcal{E}$ for the misfit function

$$C(m) = \frac{1}{2}\Delta d^{\dagger} W_d \Delta d + \frac{1}{2}\mathcal{E}(m - m_{prior})^{\dagger}\ W_m(m - m_{prior})$$

Using gradient method and Gauss-Newton approximation, the update

$$m_{k+1} = m_k + \{\Re(J^{\dagger} W_d J) + \mathcal{E} W_m\}^{-1} \Re[J^{\dagger} W_d\, \Delta d_k + \mathcal{E} W_m \Delta m_k]$$

$$m_{k+1} = m_k + \alpha_k \{\mathcal{H}\}^{-1} \Re[J^{\dagger} W_d\, \Delta d_k]$$

Omitting the model gradient, approximating Hessian, smoothing the model perturbation are practical ingredients often used in FWI

MODEL-DRIVEN CONTRIBUTION

Local optimization and regularization: quasi-Newton approach

$$C(m) = \frac{1}{2}\Delta d^{\dagger} W_d \Delta d + \frac{1}{2}\lambda_1 m^t D m + \frac{1}{2}\lambda_2 (m - m_{pi})^{\dagger} W_m (m - m_{pi})$$

$$\nabla C_k(m) = J_k{}^t W_d \Delta d_k + \lambda_1 D m_k + \lambda_2 W_m (m_k - m_{prior})$$

Ridge regression+Tikhonov+ Prior influence = L2 L2
(Tikhonov and Arsenin, 1977)

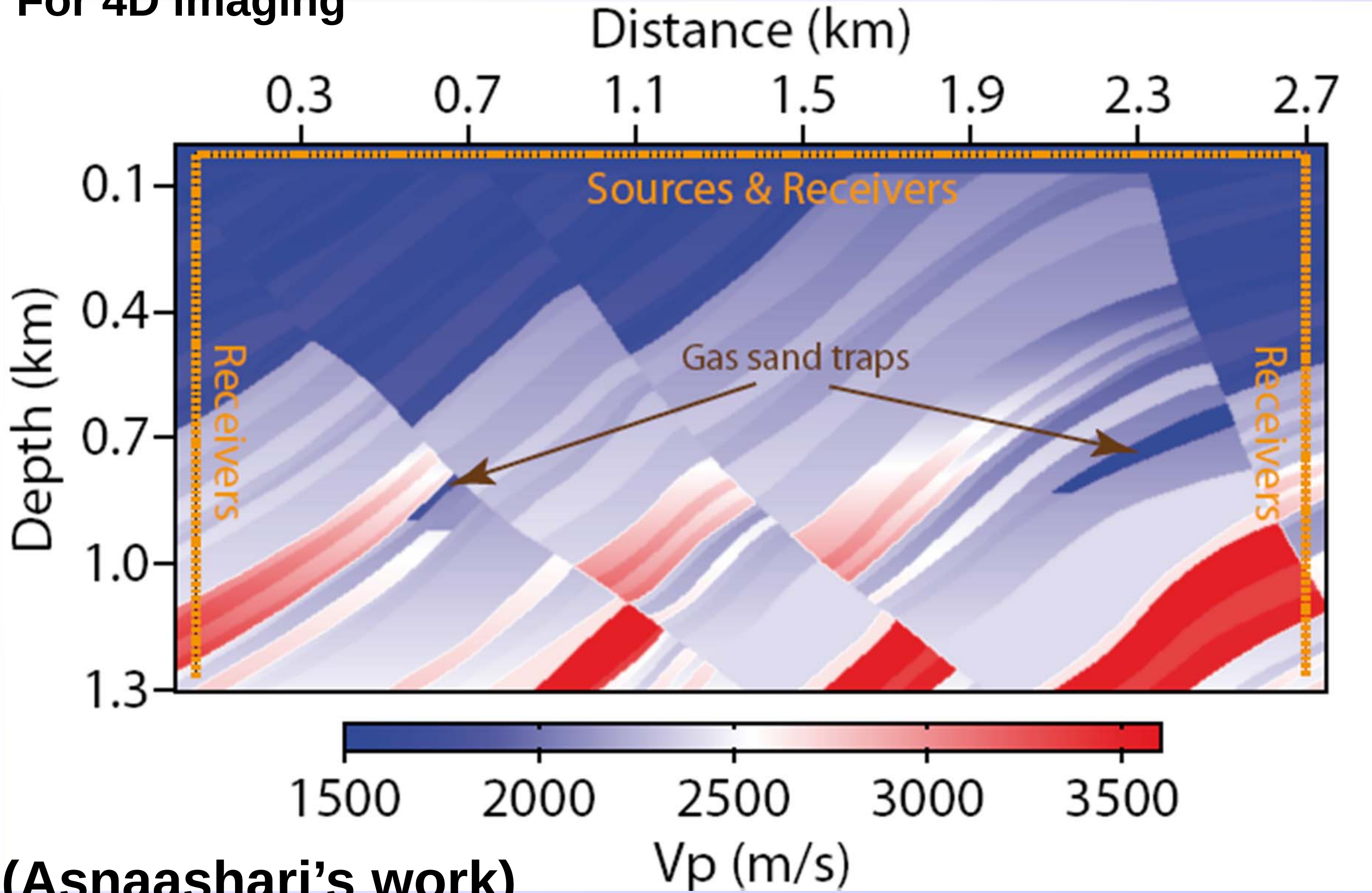
Using ℓBFGS-B for Hessian influence leads to perform only gradient numerical estimations (Byrd et al., 1995)

Two effects of the model gradient: smoothing and prior information

Estimation of hyper-parameters ?
Lasso regression: L2 L1 (preserving the sparsity …) ?

Zoom on the Marmousi II
(Martin et al, 2006)

For 4D imaging



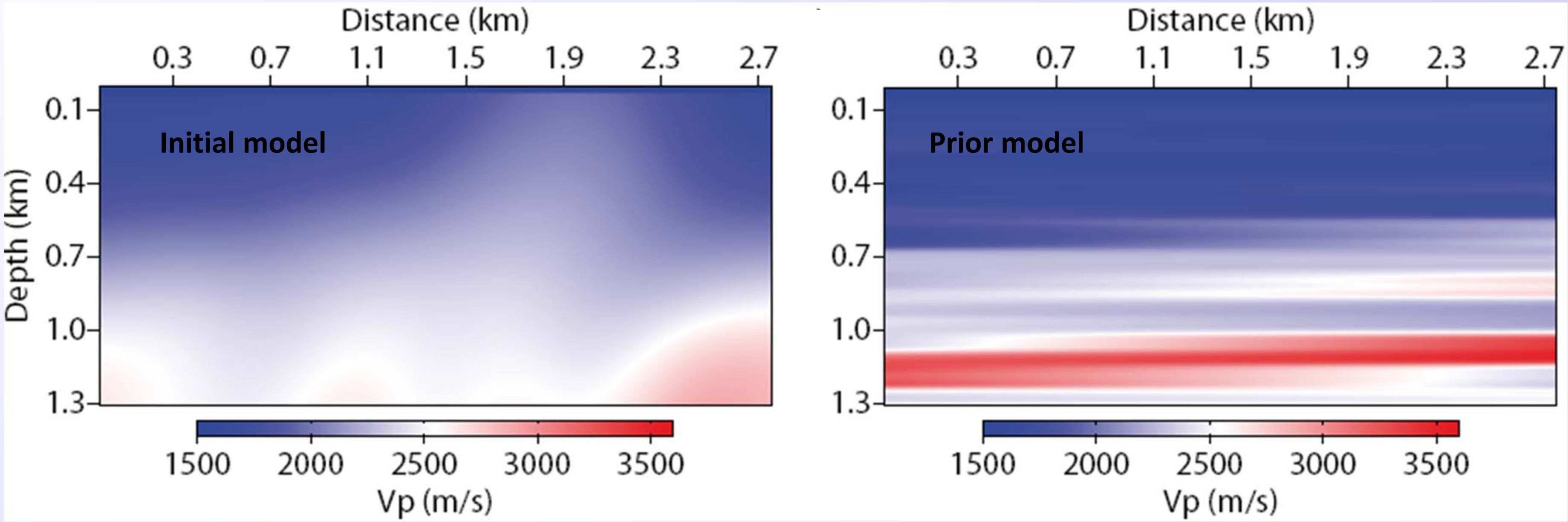
(Asnaashari's work)

Initial and prior models

- ❑ Initial model: highly smoothed true model
- ❑ Prior model: linear interpolation of two velocity profiles in wells

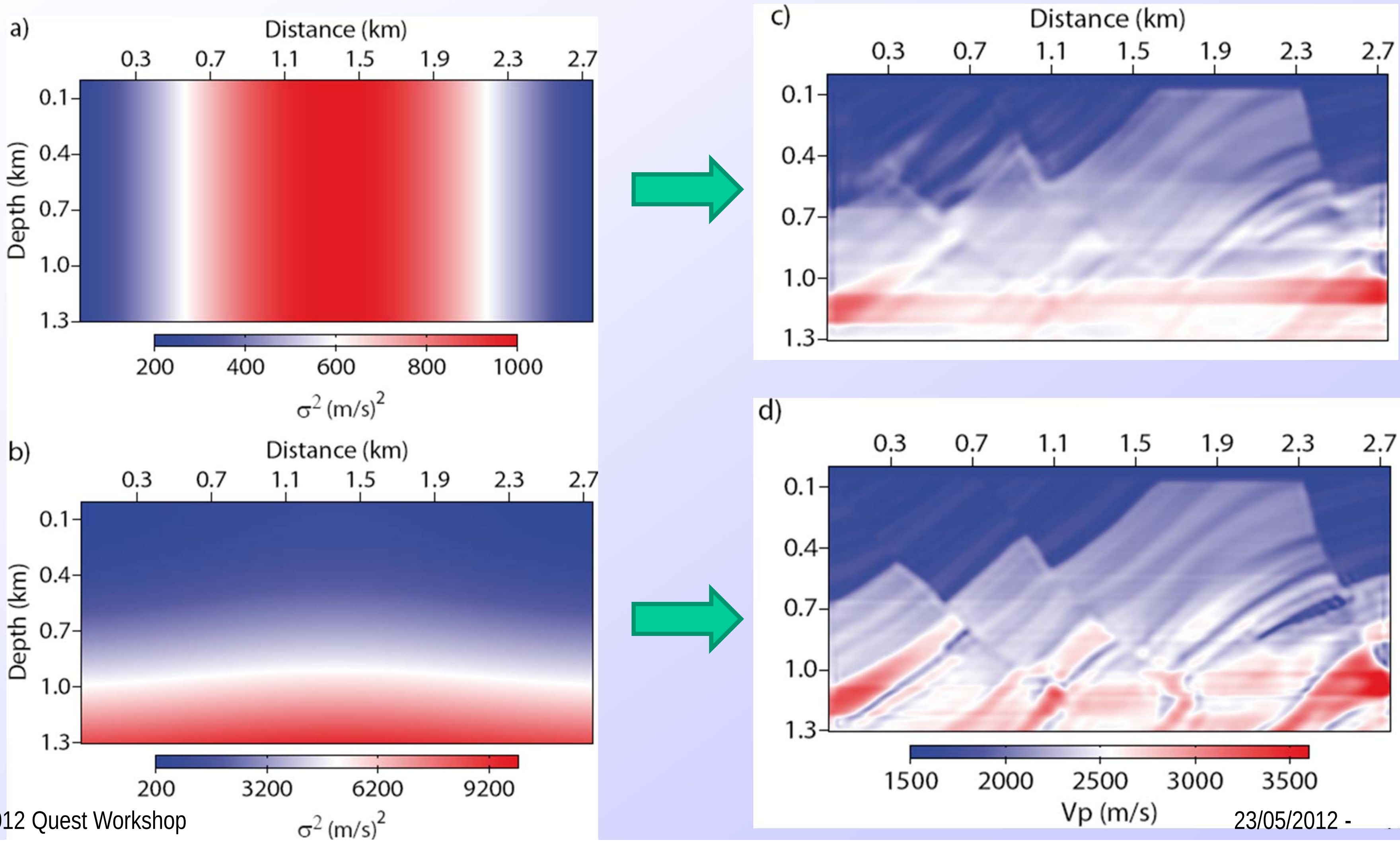
✓ Very small value for λ_1, since we want to investigate only the effect of prior model to constrain the inversion.

✓ $W_d = I[\dim(data)]$ and W_m is chosen as diagonal matrix with ${}^{1}/_{\sigma(m)}$ values, where $\sigma(m)$ is called prior weighting model which is related to both the prior uncertainty and the weighting.



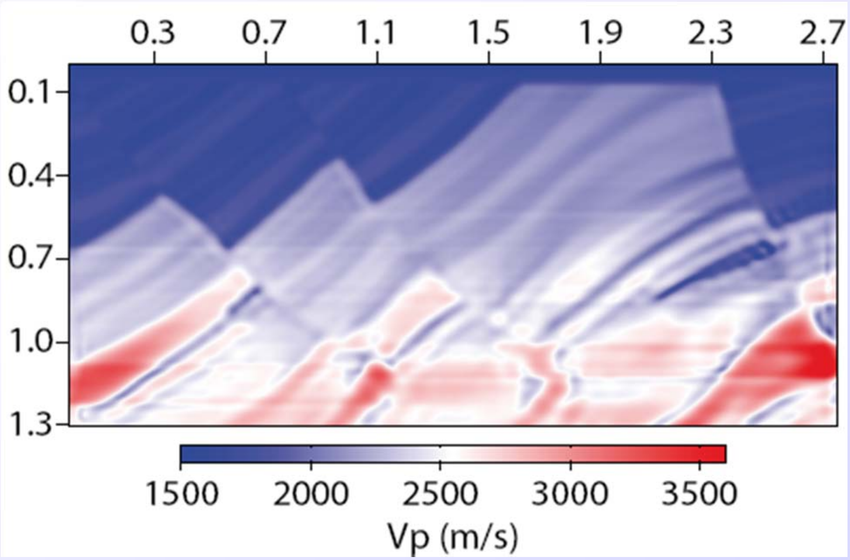
Two types of prior weighting model

Selected hyper-parameters $\lambda_1 = 20\ sec^2$ and $\lambda_2 = 3 \times 10^5$

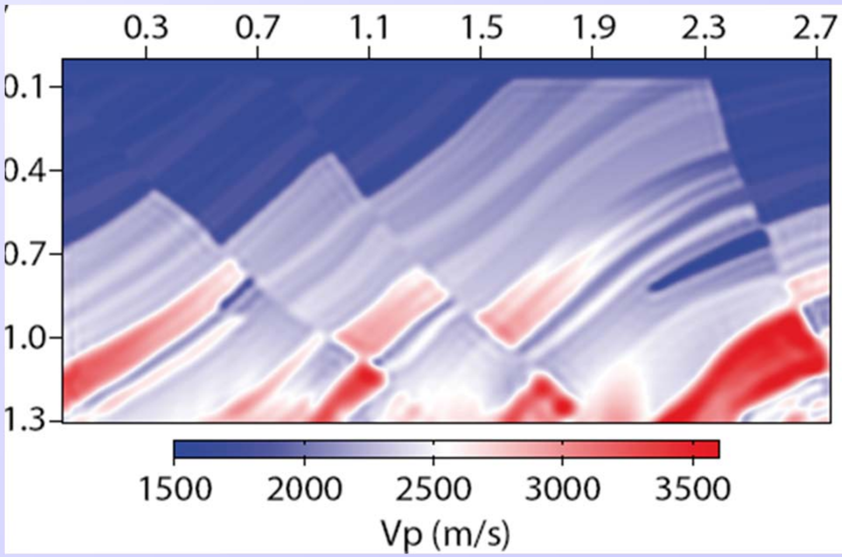


Dynamic prior weighting

- In practice, the prior model can be far from reality and also the final FWI model can keep a significant footprint of the prior model structure due to fixed weight on prior term.
- Dynamic prior weighting in order to decrease gradually λ_2 with iterations, based on derivatives of cost function evolution.



Fixed prior weight



Dynamic prior weight

Partial conclusion on the model side

Elaborating various informations in a prior model is possible in the FWI protocole.

Initial and prior models with lower spatial frequency content showing that

- we could relax our constraint on the low frequency content of the data.
- we could relax our constraint on the accuracy of the initial model.
- we could relax illumination requirement

Adapt our model description (blocky, gridded, compact…)

Optimisation performance

Multi-parameters trade-off and secondary scattering effects

Hessian operator plays a significant role (Pratt et al, 1998)

- proper gradient scaling with respect to parameters

- correct for acquisition biases in the illumination

BUT … sometimes trade-off and secondary scattering effects are important in the optimisation

$$H_{ij} = \Re\left\{J^{\dagger}J_{ij} + \sum_{k=1}^{n} \frac{\partial^2 u_k}{\partial m_i \partial m_j} \Delta d_k{}^{*}\right\} \qquad \mathbf{H(m)=B(m)+C(m)}$$

B => Gauss-Newton method

B+C => Quasi-Newton and full-Newton method

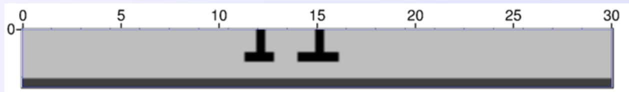
$$\boldsymbol{H}(\boldsymbol{m_k})\Delta\boldsymbol{m_k} = -\boldsymbol{\nabla}C(\boldsymbol{m_k}) \qquad \textbf{(1)}$$

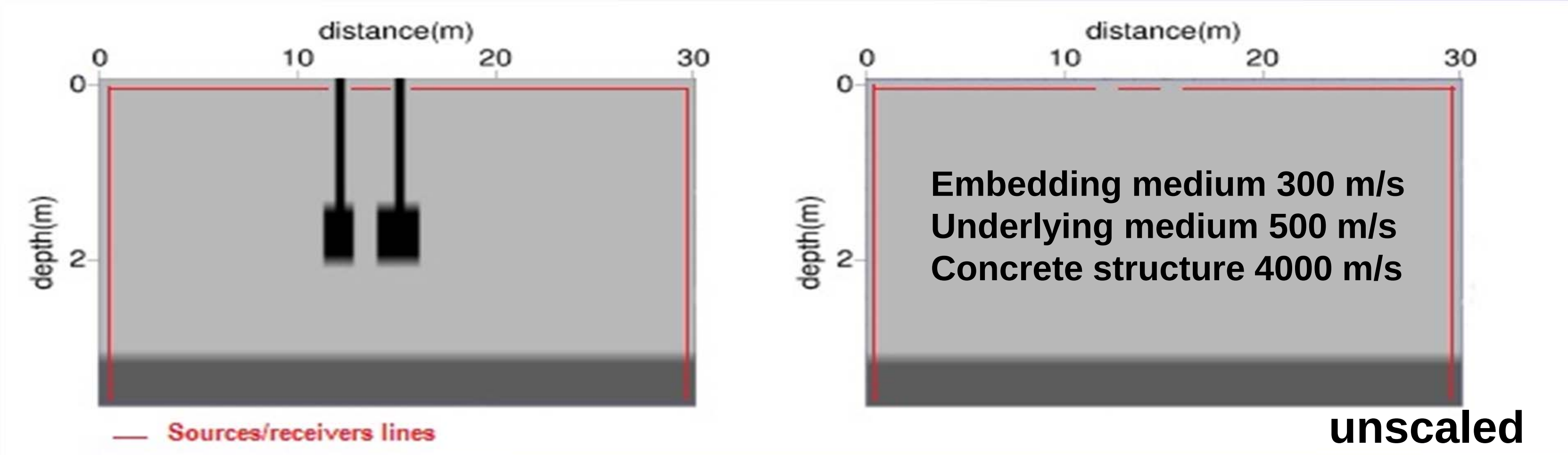
- l-BFGS is solving (1) with a quadratic interpolation through FDs (two modeling + storage)

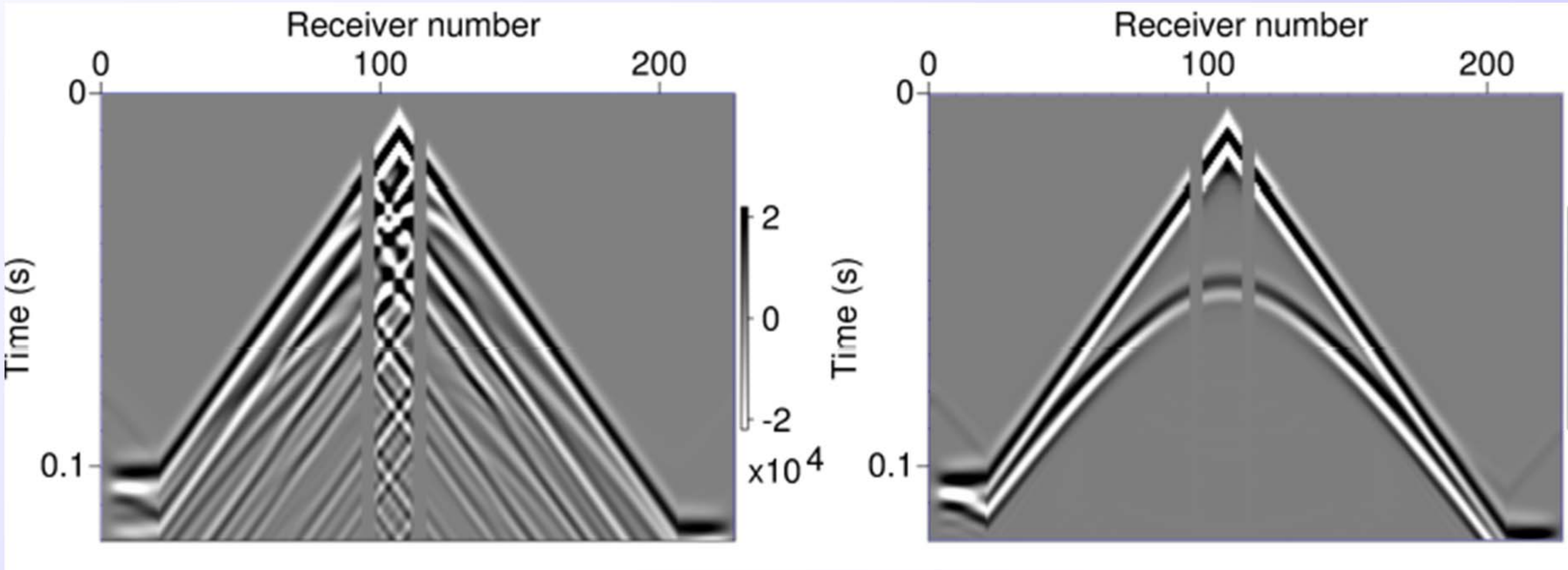
- Truncated Newton is solving (2) while Hessian could not be a quadratic form (Nash, 2000) (four modeling ~ Gauss-Newton overload)

(Métivier's work)

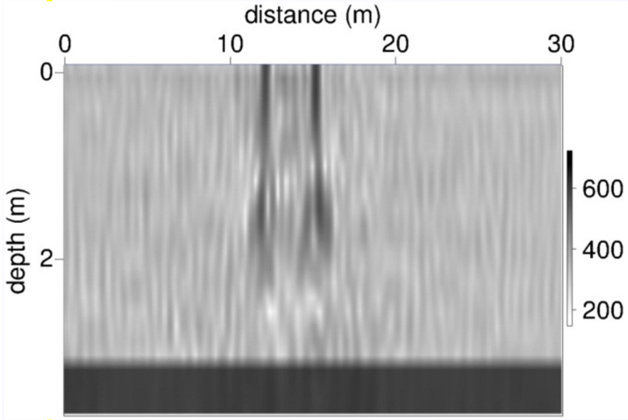
Near-surface bodies with high constrasts



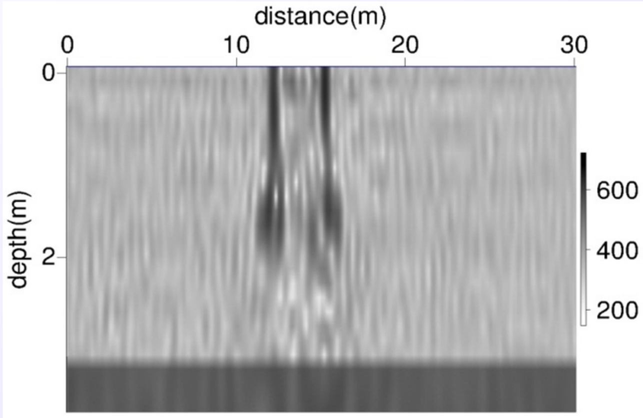




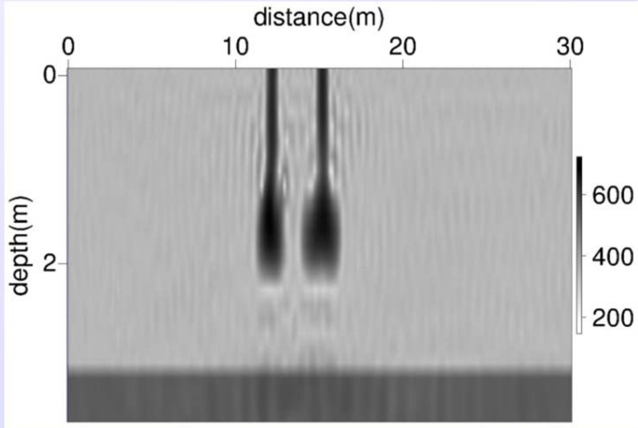
Reconstruction of the embedded concrete structures



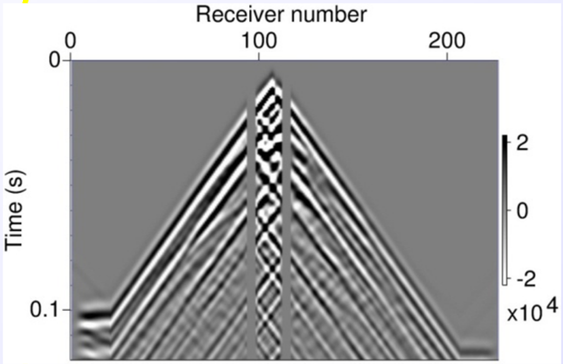
Steepest descent

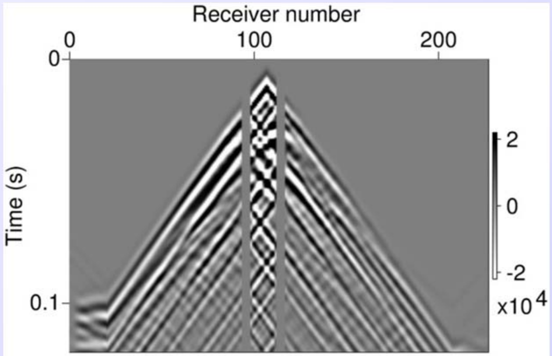


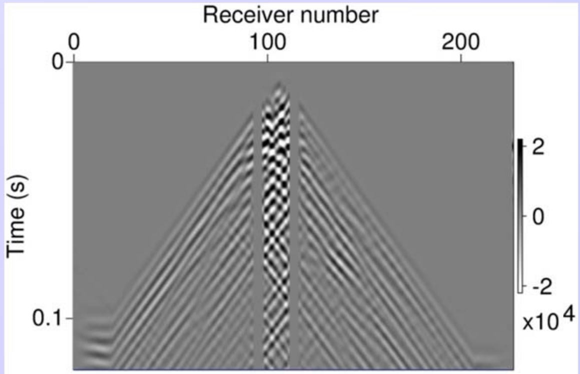
Quasi-Newton (l-BFGS)



Truncated Newton (TN)



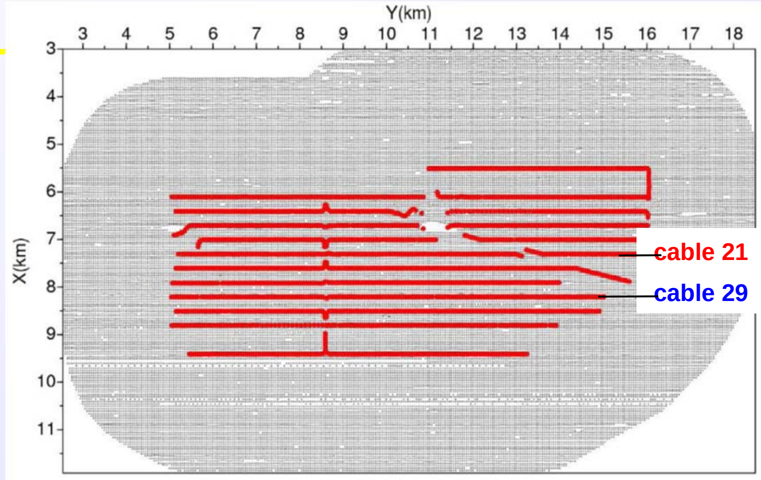




Residues

In most cases, l-BFGS strategy is the most efficient one …

Background on Valhall

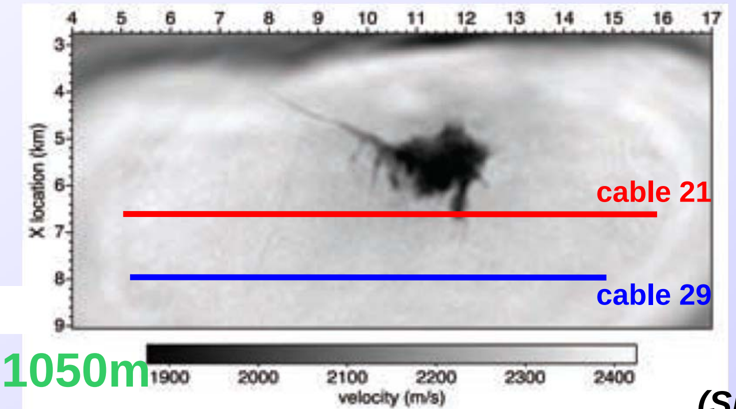


- Exploited since 1982, Life of Field Seismic (LoFS) network since 2003

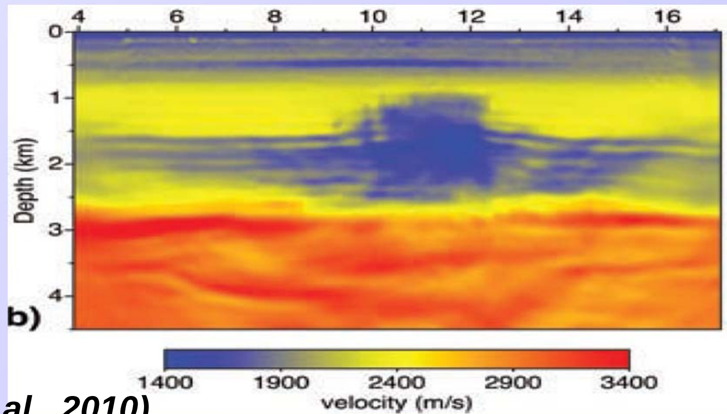
- BP starting models by anisotropic reflection traveltime tomography

- Strong imprint of anisotropy in the seismic Valhall dataset

- 3D isotropic acoustic FWI was presented by Sirgue & al., 2010, using 13 km maximum offset

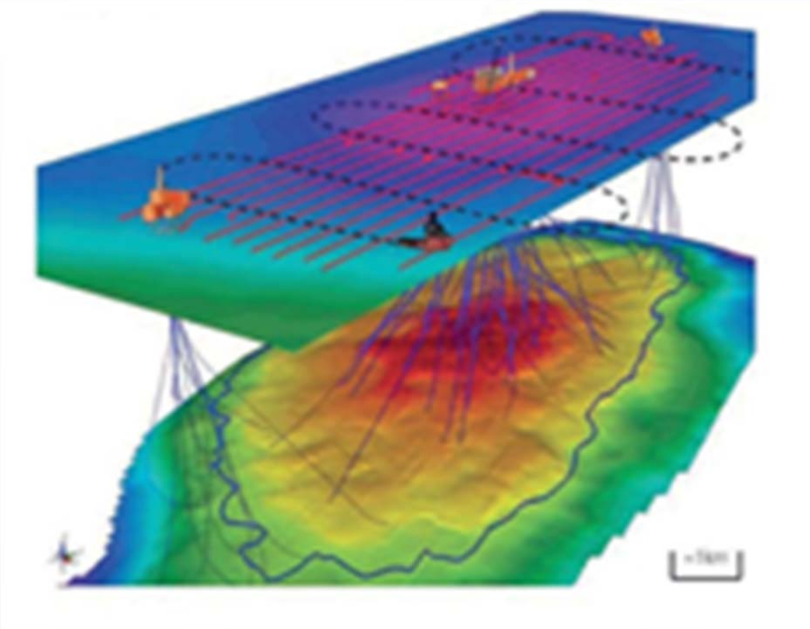


Z=1050m

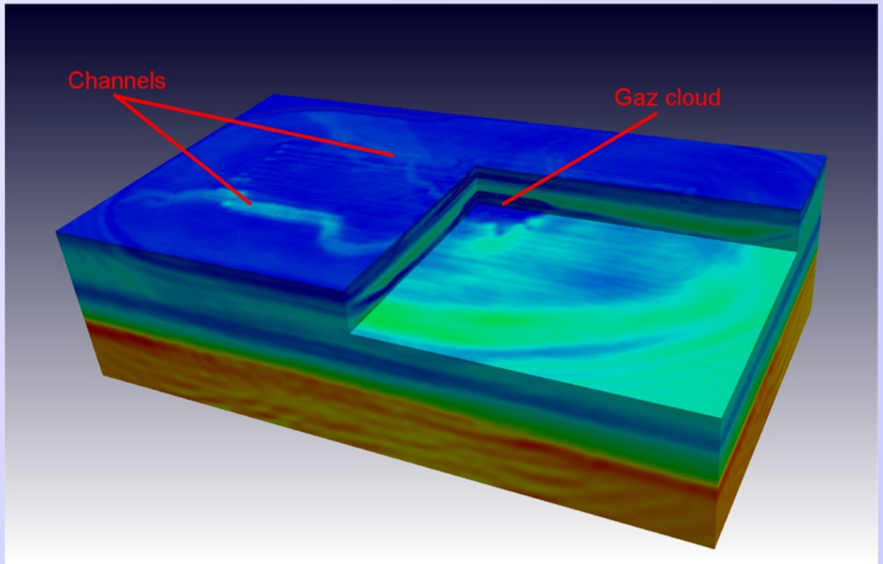


(Sirgue & al., 2010)

3D acoustic FWI



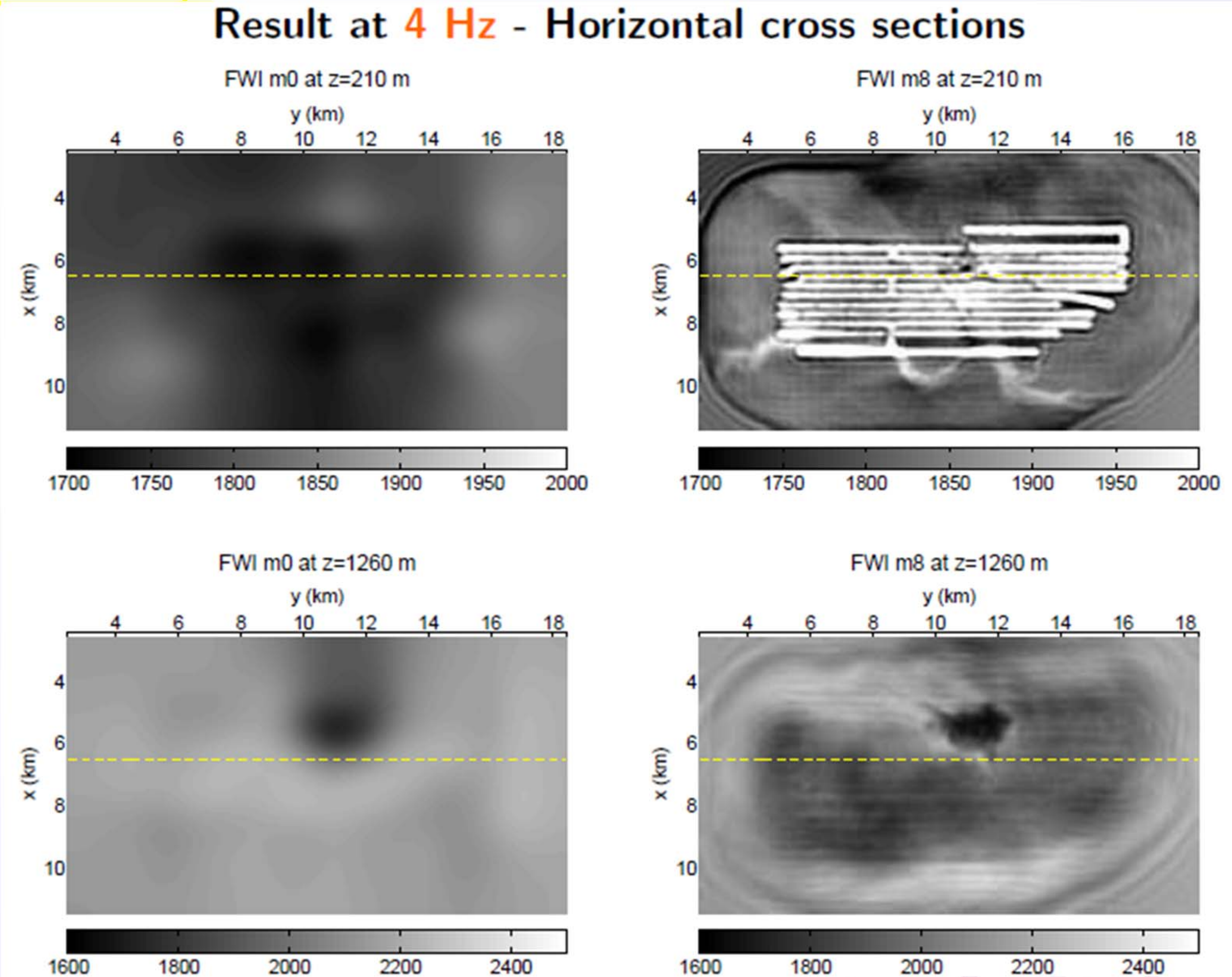
For+Inv	Few cores	Many cores
Time+Freq	20830 s	326 s
Freq+Freq	6209 s	1445 s



3D monoparametric reconstruction (Pratt's strategy)

(Etienne's and Hu's work)

3D acoustic FWI



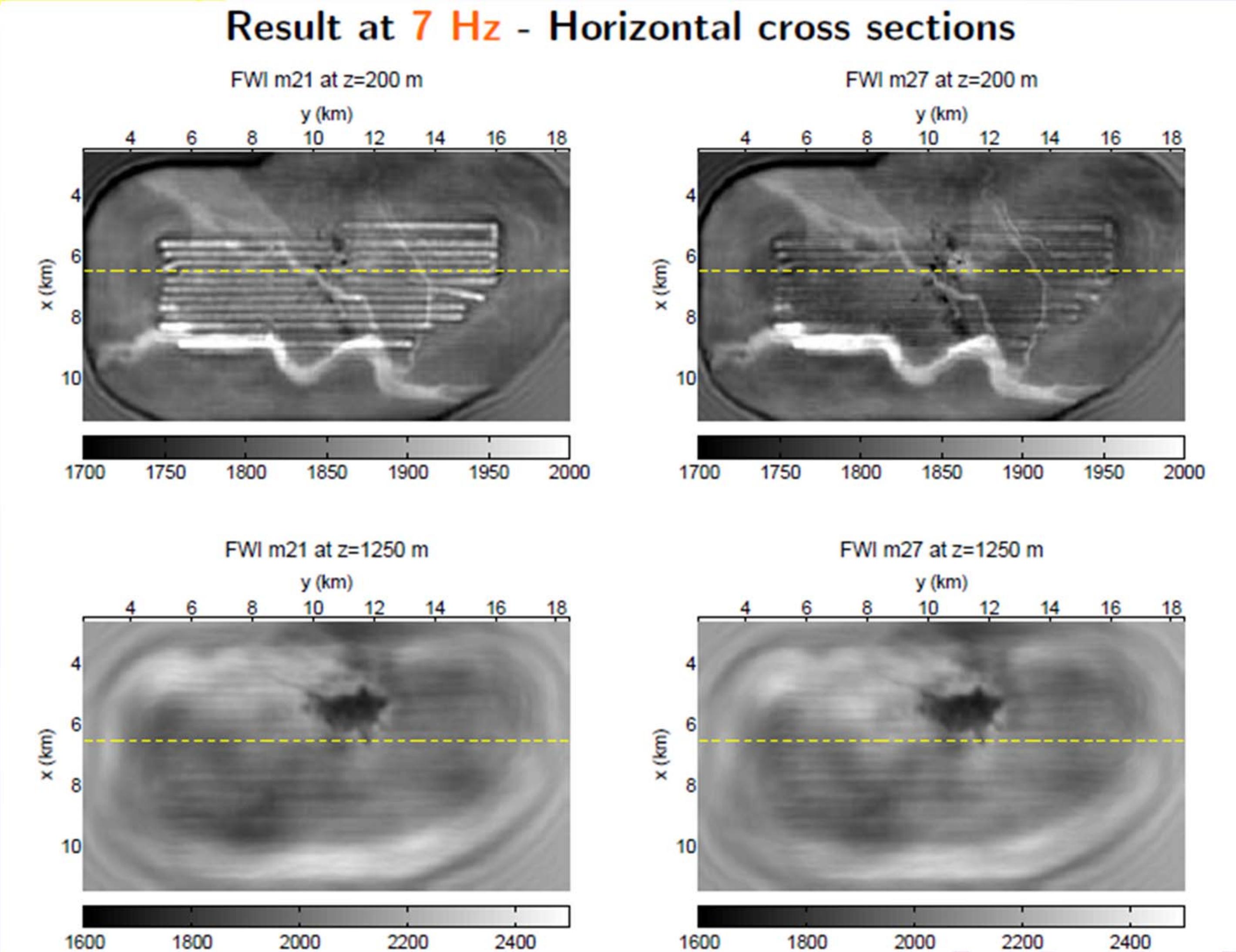
Superficial channels

Gas reservoir

Imprint of the acquisition

Travel-time tomography **Vp FWI**

3D acoustic FWI



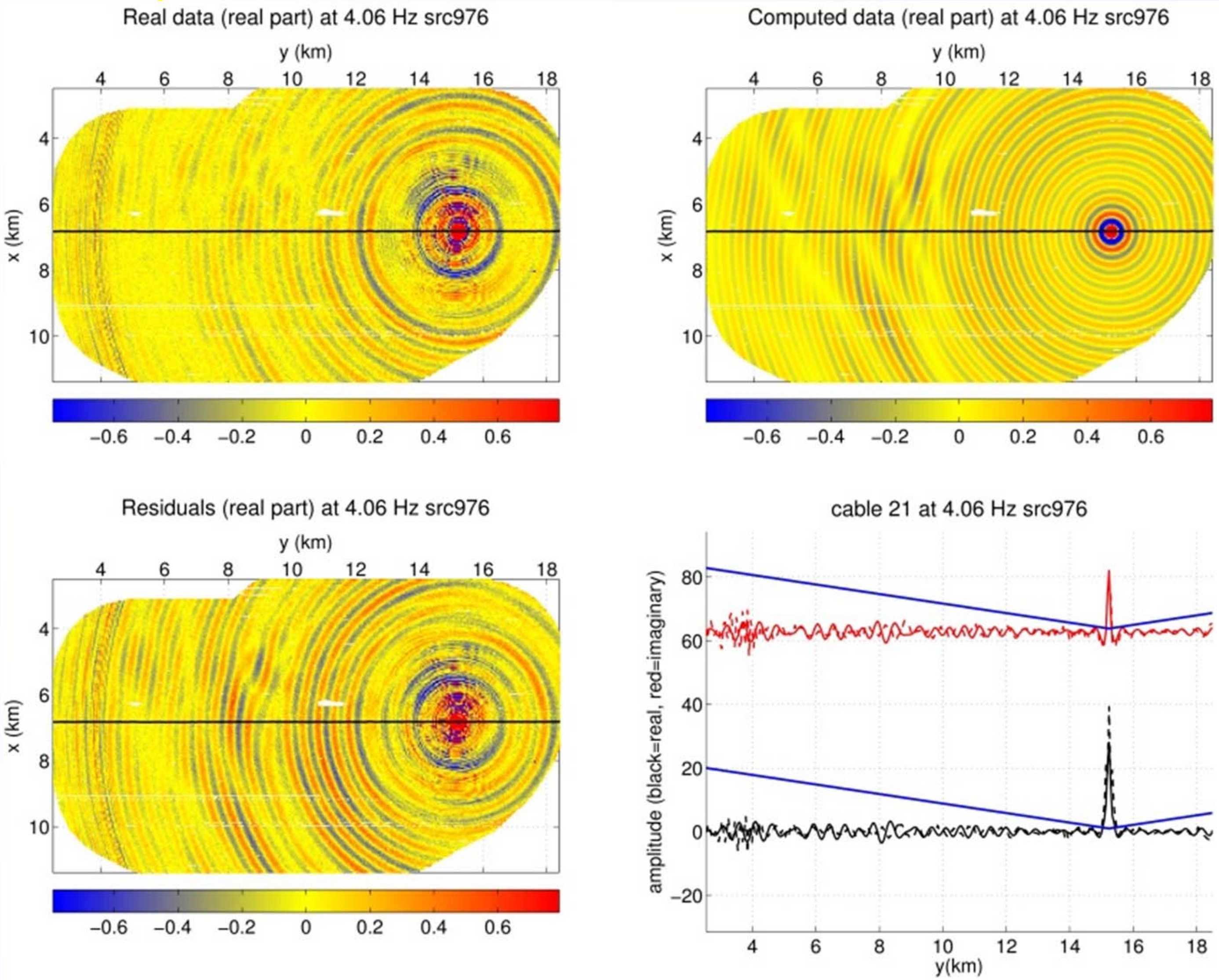
Travel-time tomography **Vp FWI**

Superficial channels

Gas reservoir

Imprint of the acquisition

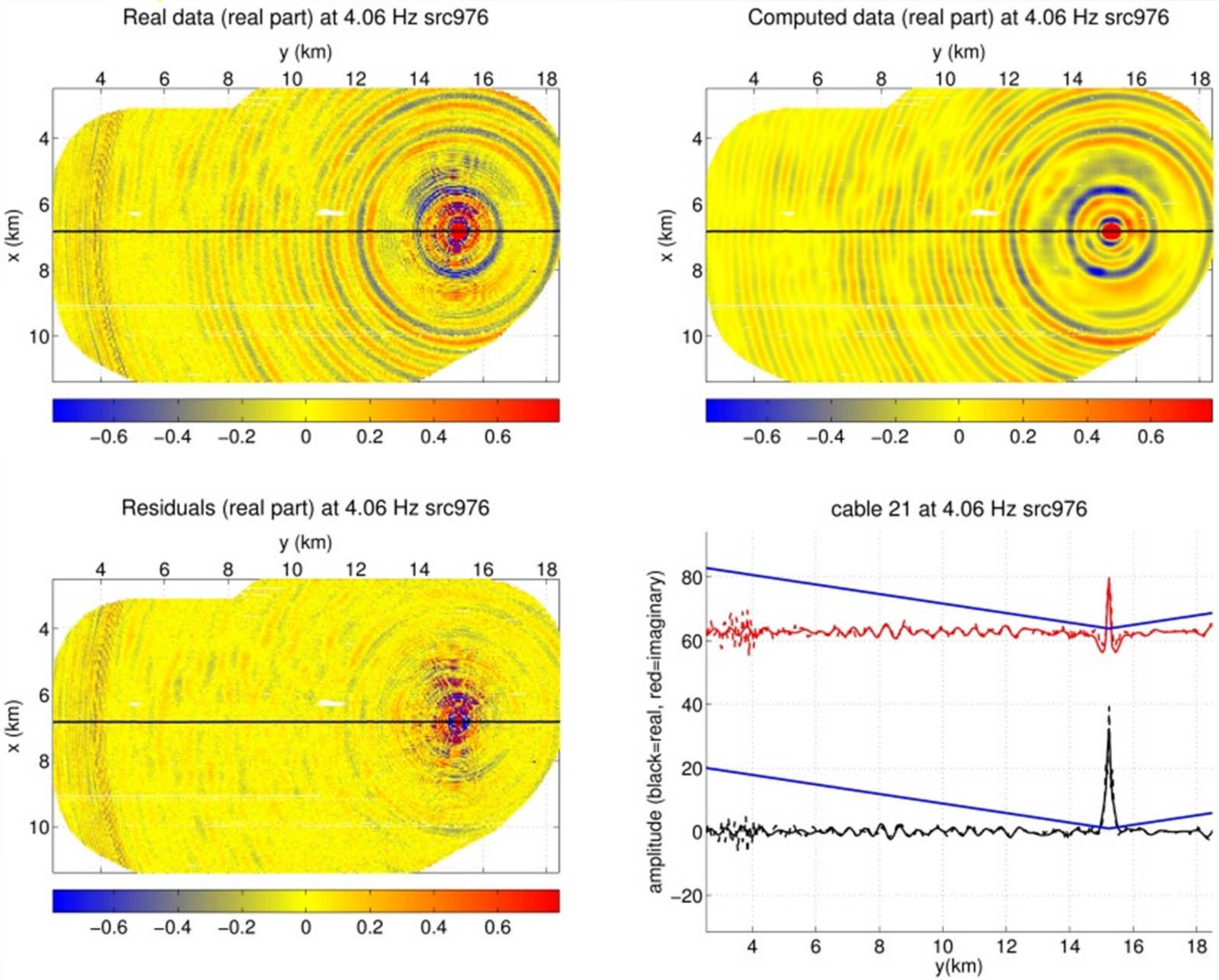
Fitting waveform: residual reduction



QC

- Source repeatability
- Waveform fit

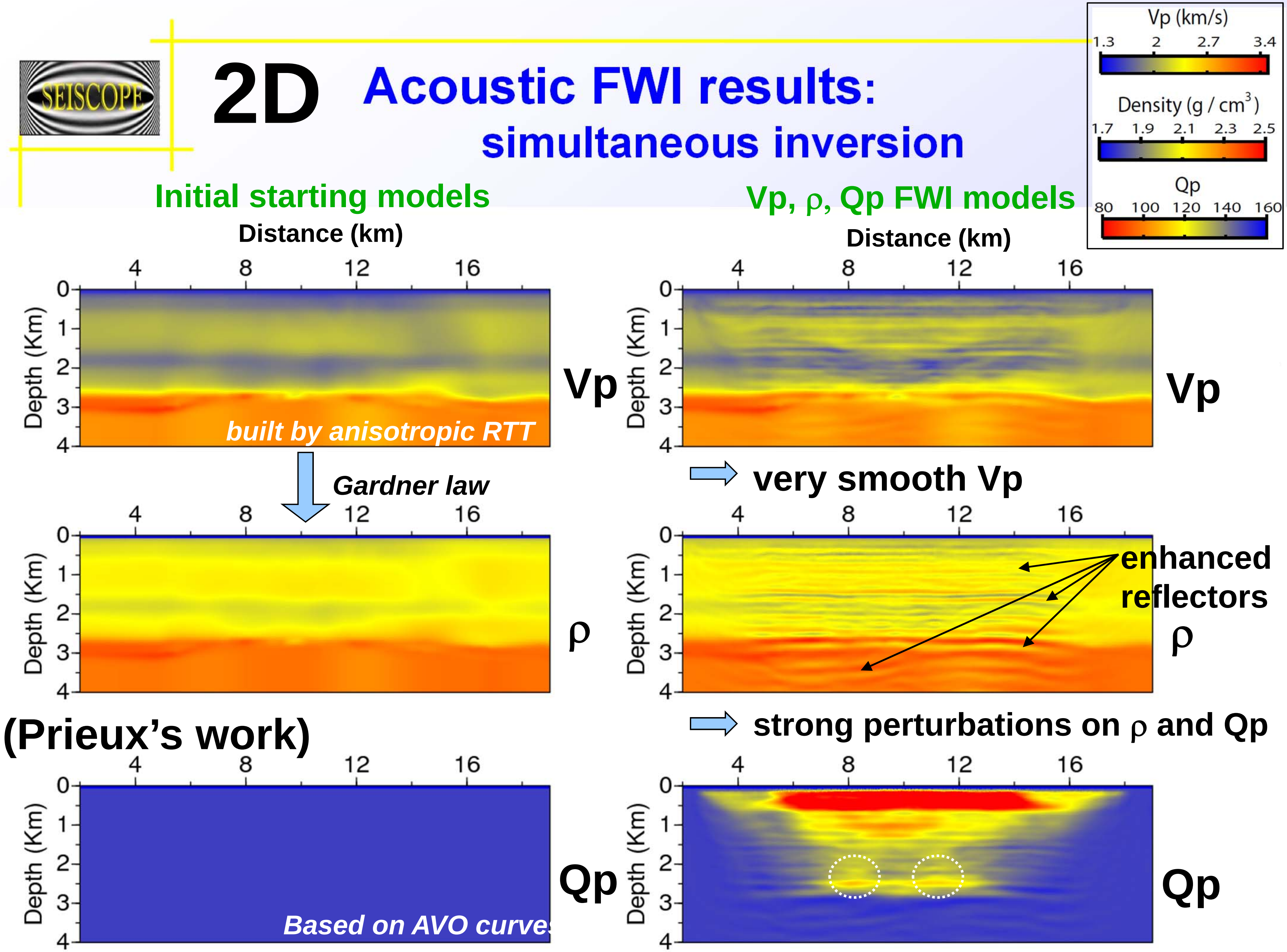
Fitting waveform: residual reduction



QC

- Source repeatability
- Waveform fit

2D Acoustic FWI results: simultaneous inversion



Hierarchical workflow over parameter classes and data components

Test	Step	Inverted data	Inverted parameter	Approximation
	1	hydrophone	V_P	acoustic aniso.
	2		V_P, ρ, Q_P	
G	3	geophone	V_P, V_S	elastic aniso.
H	3	hydrophone		
HG	4	geophone		

Step 3

Tests G and H use same starting models (Vp and smooth Vs)

Two main levels of hierarchy

(Inspired by Tarantola, 1986)

1) Dominant data and parameters

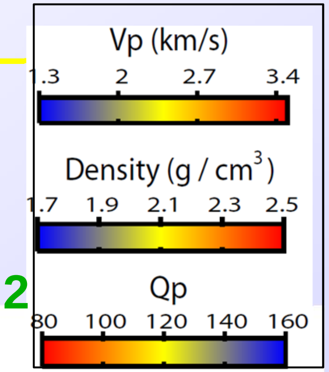
data → **Acoustic before elastic**

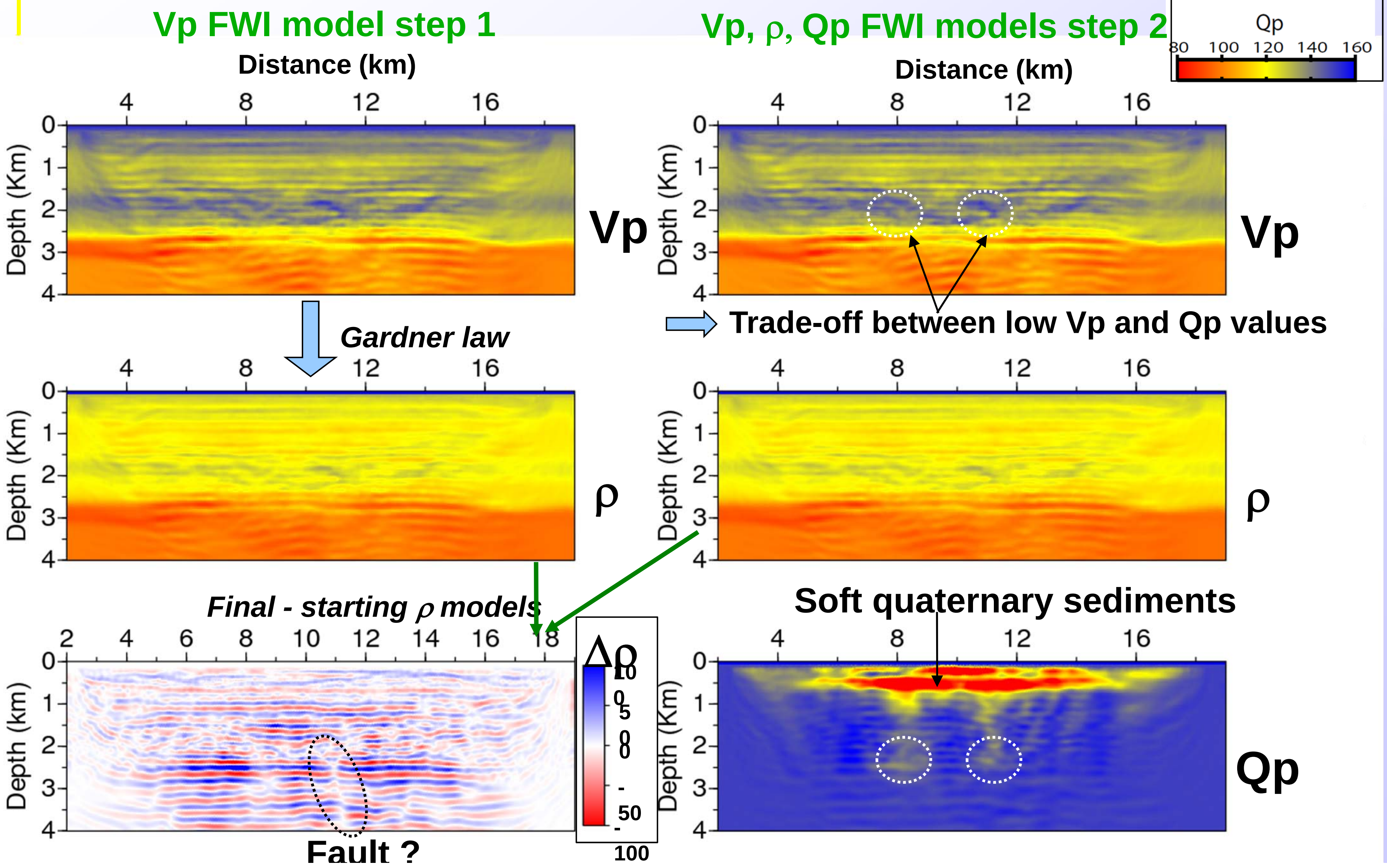
model → **Vp before ρ and Qp**

PP reflections => intermediate wavelengths

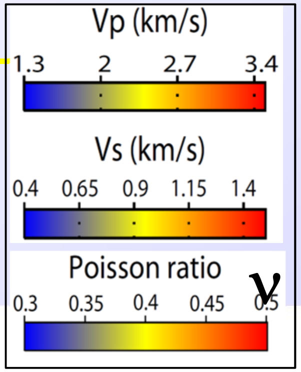
PS reflections => short wavelengths

Acoustic FWI results:
hierarchical inversion

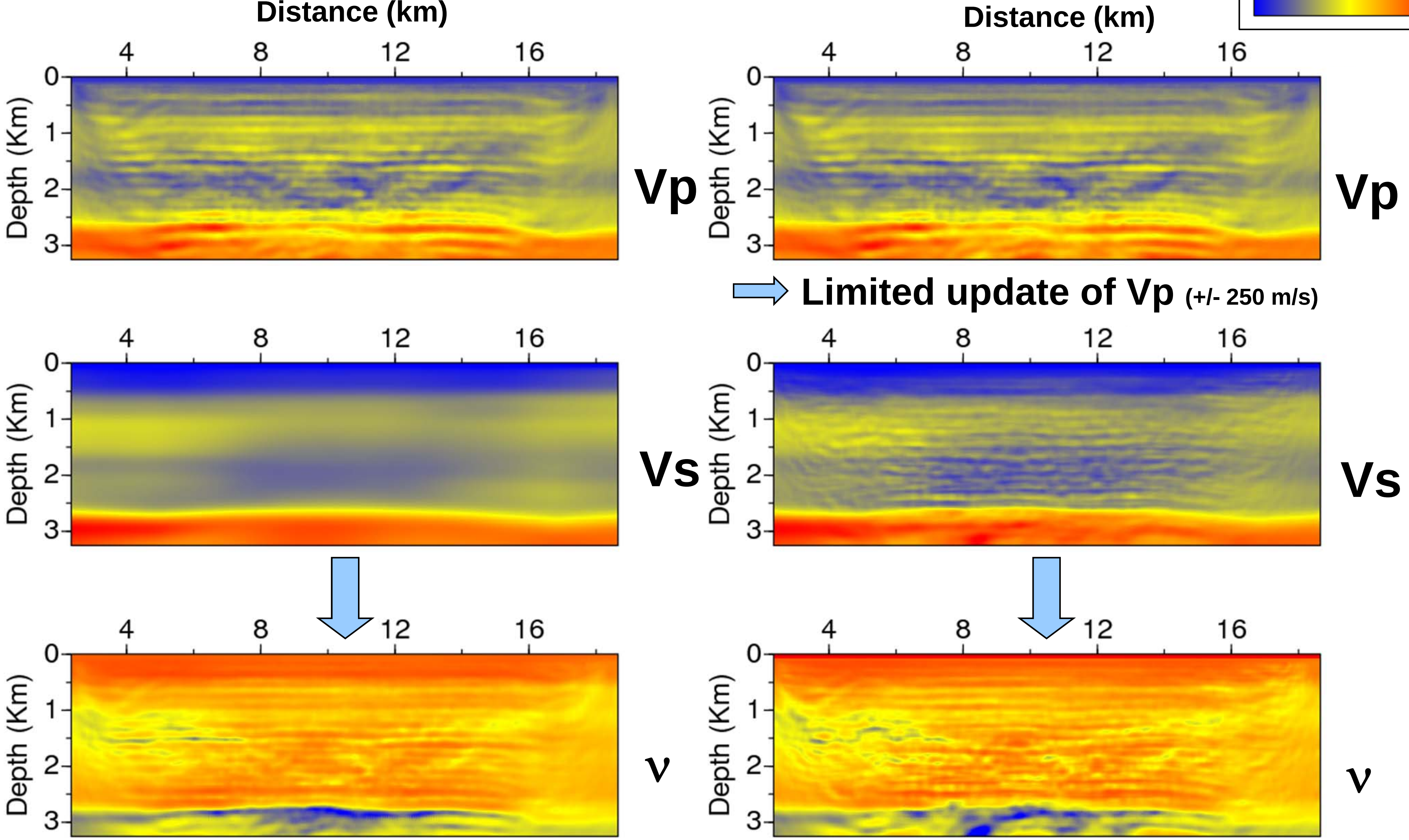




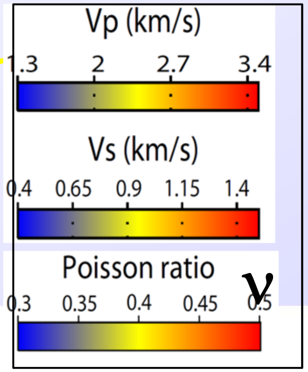
Elastic FWI results: Test H



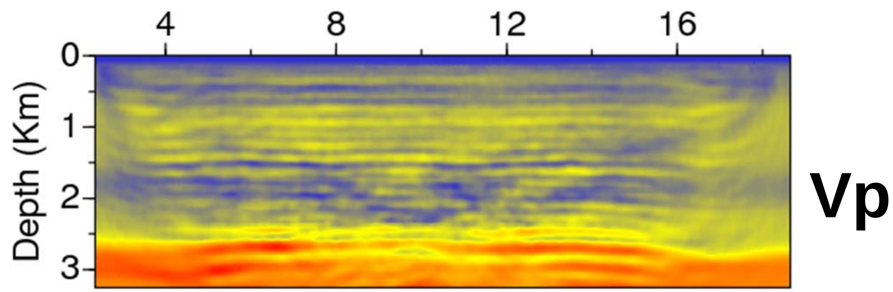
Starting models step 3 | **FWI models Test H**



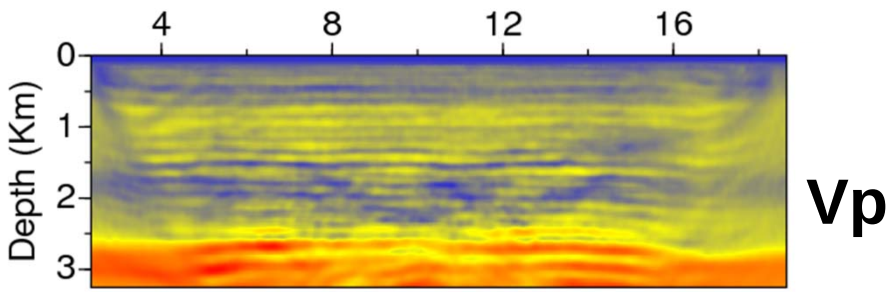
Elastic FWI results: Test G vs Test HG



FWI models Test G

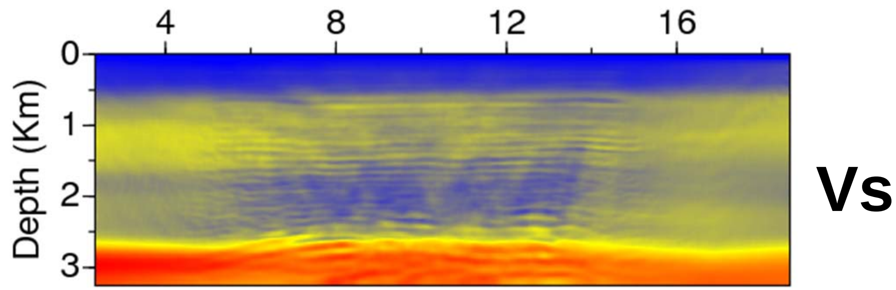


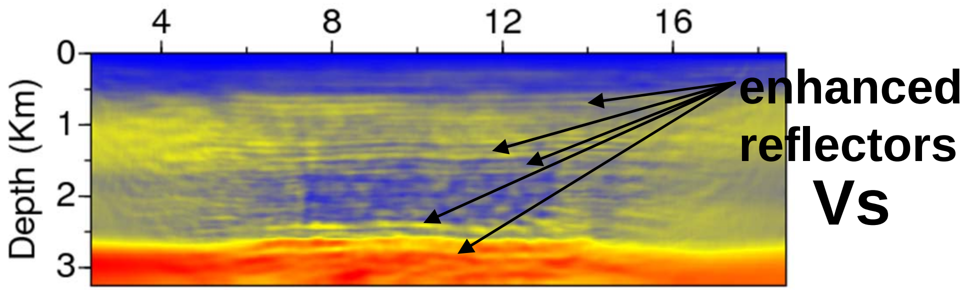
FWI models Test HG



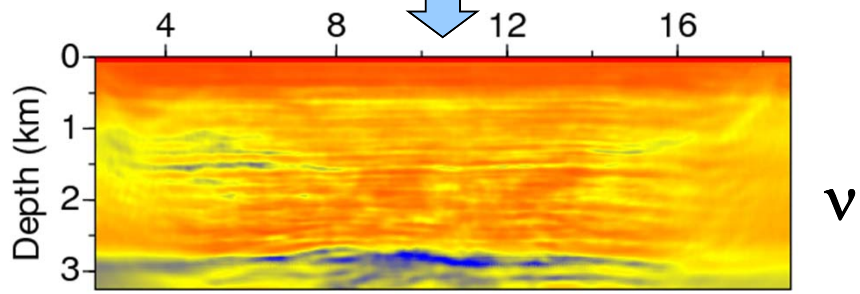
⇨ Reflectors less continuous with Test G

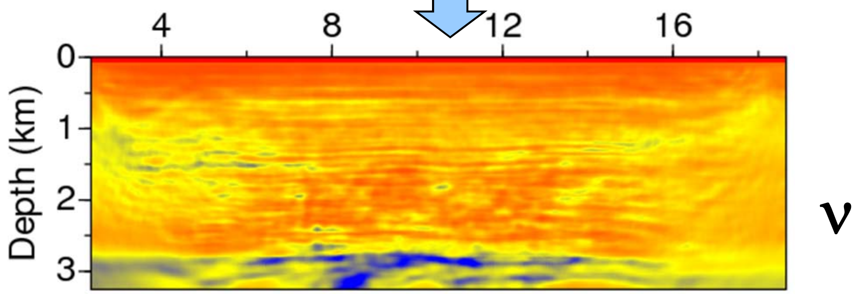
⇨ Vs less smooth than with Test H





⇨ Stronger perturbations with Test HG





Conclusion

- FWI is an high resolution imaging technique
  (still a least-square method)

- Model prior information
  - Acquisition side
    - Low-frequency content request is relaxed
    - Acquisition geometry is less important
  - Model interpretation (expected features)

- Multi-parametric reconstruction
  - Double hierarchy in data space and model space

- Hessian influence
  - Strong contrast
  - Potentialities towards resolution quantification

- Real data application

FWI=λ/2

Thank you very much for your attention

We would like to thank sponsors of the Seiscope consortium

Reshaping the objective function

Making the objective function less sensitive to amplitudes, noise or amplify specific parts of the seismogram (still FWI?)

- **Cross-correlation approach (Fichtner, 2008; Leeuwan & Mulder, 2010)**
- **Deconvolution approach (Luo & Sava, 2011)**
- **Laplace approach (Shin & Cha, 2009)**

Intermediate results based on phase information more than on amplitude information is another way to mitigate the local minimum pitfull.

The evolution of the objective function from one to the other one could be part of the global workflow for the entire interpretation of the full seismogram with various weights.

DISCUSSION (1)
The Low's

A - Cycle skipping

The non-linear optimisation performed by the full waveform inversion could be trapped into local minima.
How could we avoid cycle skipping ?

1. **data content (low frequency)** ?
   sequential frequency procedure from low to high frequencies
2. **increasing the linearity of the inversion (phase versus amplitude manipulation)** ?
   transformed domain : Laplace domain, phase wrapping, misfit shaping through cross-correlation
3. **prior information if there is any** ?
   prior information may fill in the spatial frequency content of the image we want to reconstruct
4. **precise initial model (high frequency)** ? better picking of times could improve travel-time tomography procedures (first-arrival times tomography, stereotomography, finite frequency tomography)

DISCUSSION (2)
The High's

B - Resolution issue

How far could we go in the extraction of the information from seismograms to reach this resolution ?

1. **multi-scattering issues of the high frequency content ?**
   stronger interaction between waves and media balanced by attenuation
2. **illumination issue ?**
   better acquisition geometry for focused energy on the target
3. **macro-model reconstruction versus reflectivity estimation ?**
   velocity reconstruction versus impedance contrast estimation