Full Waveform Inversion: challenges

By A. Asnaashari[1], F. Bretaudeau[1], R. Brossier[1], C. Castellanos[3], M. Dietrich[1], S. Garambois[1], F. Lavoué[1], I. Masoni[1], L. Métivier[1,2], S. Operto[3], A. Ribodetti[3] and J. Virieux[1]

[1] ISTerre, UJF-CNRS, [2] LJK, UJF-CNRS, [3] GEOAZUR, CNRS-UNSA-IRD-OCA, France

SEISCOPE I http://seiscope.oca.eu (2005-2011)

SEISCOPE II http://seiscope2.osug.fr (2013-2018)

AGU SESSION S008

- **Advances in seismic imaging: Towards integrated GeoModels on all scales**
- **Conveners: Andreas Fichtner, Paula Koelemeijer, Carène Larmat, Monica Maceira**

Seismic imaging is rapidly evolving due to the advent of high-density networks, new modeling techniques, and unprecedented HPC capacity. Our view of the Earth is transforming – from crust to core, on local to global scales. We invite presentations on theoretical and data driven developments in seismic analysis that contribute to the construction of the next generation GeoModels. Emphasis will be on new techniques that harness large emerging data sets and modern computational methods. Topics include the joint inversion of complementary geophysical data sets, forward and inverse modeling of full waveforms, and probabilistic approaches. We welcome studies of the near surface and deep Earth, from reservoir to global scale.

6 August: deadline

(https://fallmeeting.agu.org/2013/scientific-program/session-search/sessions/s008-advances-in-seismic-imaging-towards-integrated-geomodels-on-all-scales)

The broad range of topics will be reflected by our invited speakers: Chao Wang (ION GX Technology), Sebastian Rost (University of Leeds), Antonio Villasenor (ICTJA Barcelona) and Victor Tsai (Caltech).

(Jeroen Ritsema & Jean Virieux help in promoting this session)

Two statements

À la manière d'Albert Tarantola

- FWI is a data-driven imaging technique as we collect billions of real numbers.

- FWI is very democratic as each update is an average over sources and receivers

Adjoint formulation - Chavent (1974), Lailly (1983), Tarantola (1984), Pratt (1996)…

Normal mode - Woodhouse & Dziewonski (1984)…

Asymptotic formulation: (mainly) surface waves - Romanowicz/Snieder (1988a, 1988b)…
Asymptotic formulation: (mainly) body waves - Jin et al (1992), Lambaré et al (1992)…

The difficult time of the FWI !

Non-linearity or the cycle skipping or the phase ambiguity or secondary minima !

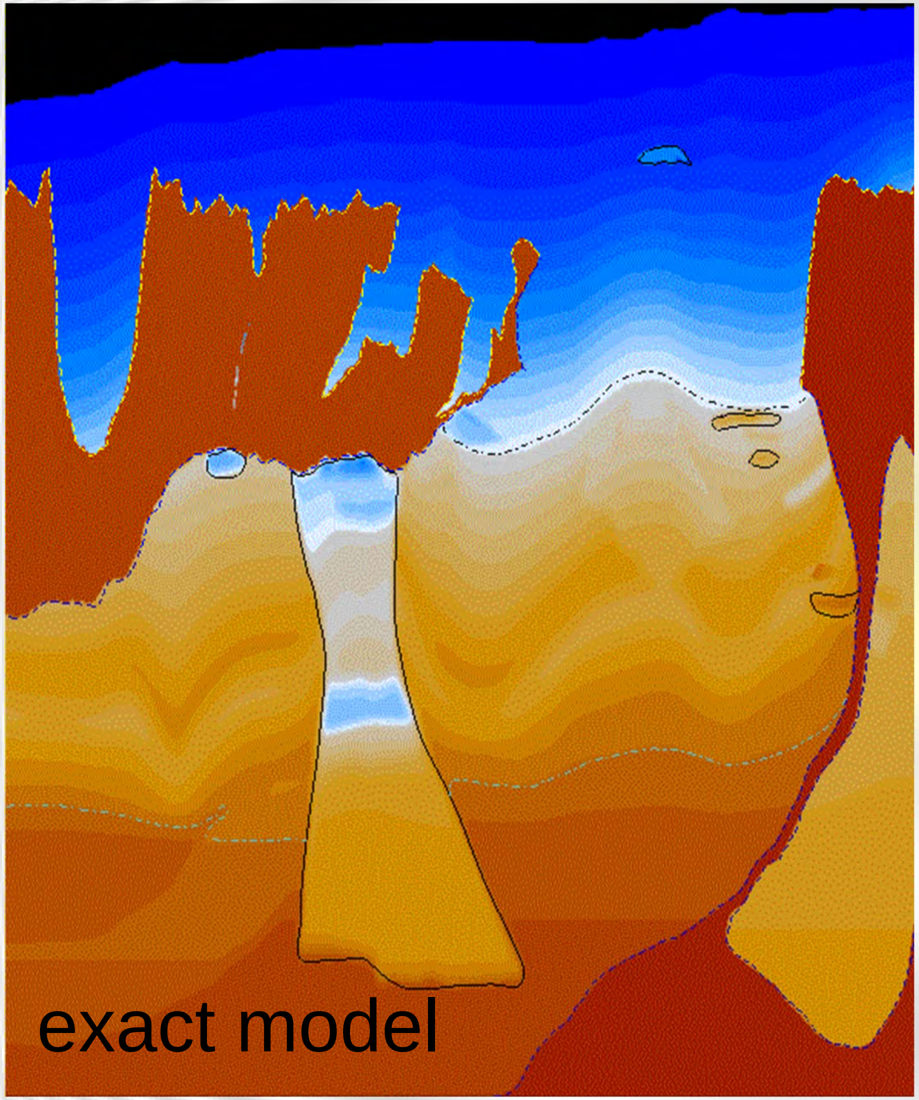
Reflection seismic data are not enough for the simultaneous reconstruction of short and long wavelength content of the medium using linearized techniques

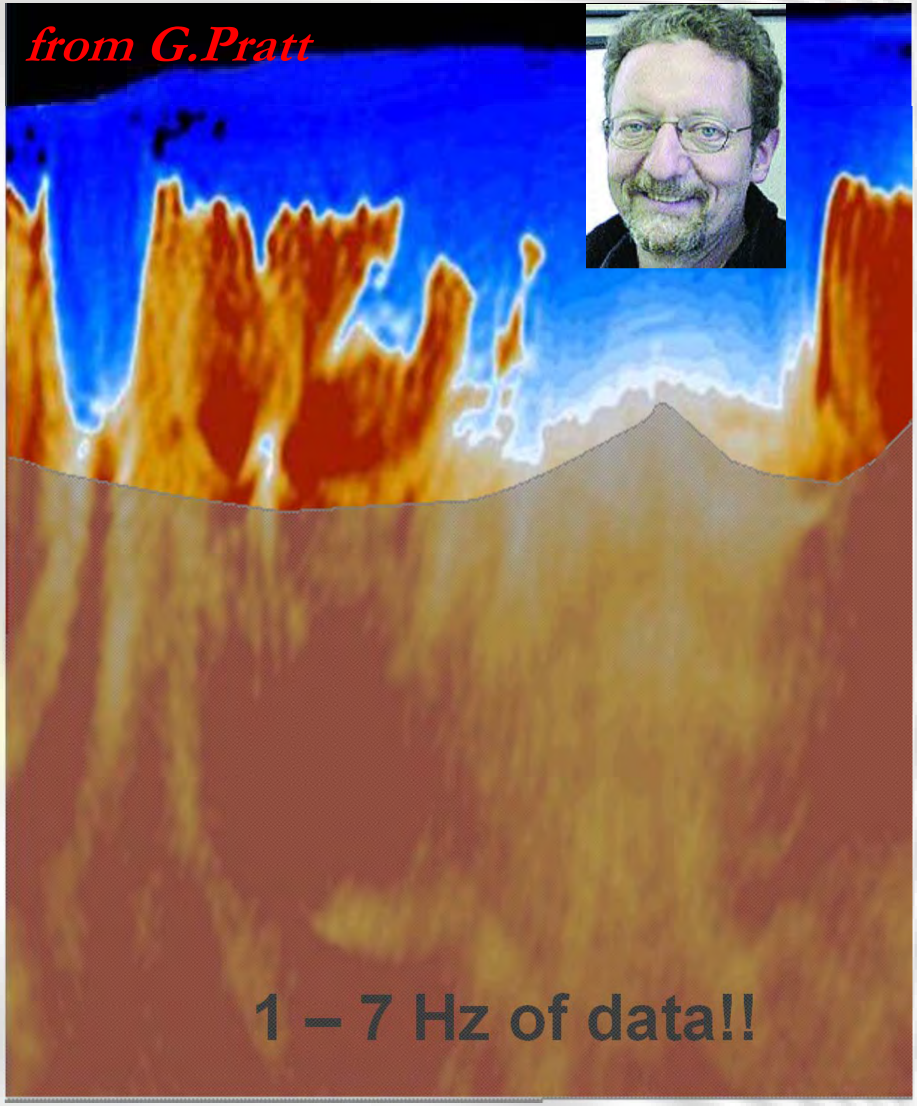
Message of Albert with his work at the late 80's (Jannane et al, 1989)

But data acquisition have changed …. during the 90's !

Both reflection and refraction data are recorded simultaneously using long streams or global offsets (12-15 km)

BP 2004 model using Full Waveform Inversion



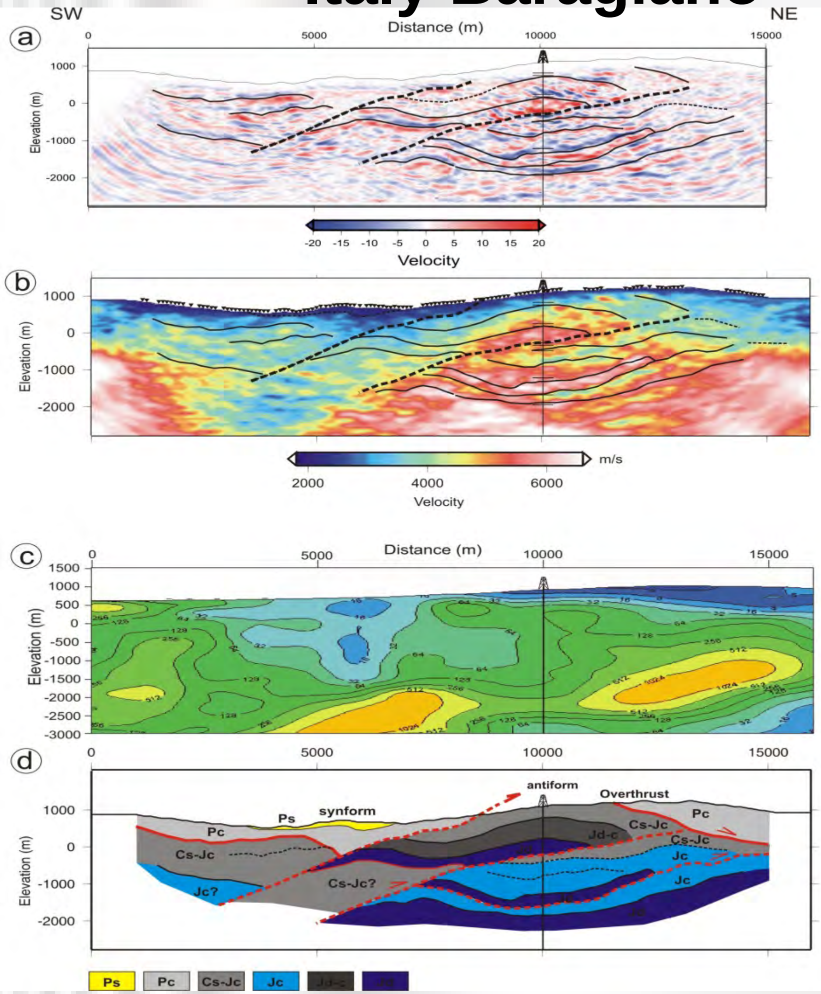


Low frequency information

2D synthetic example

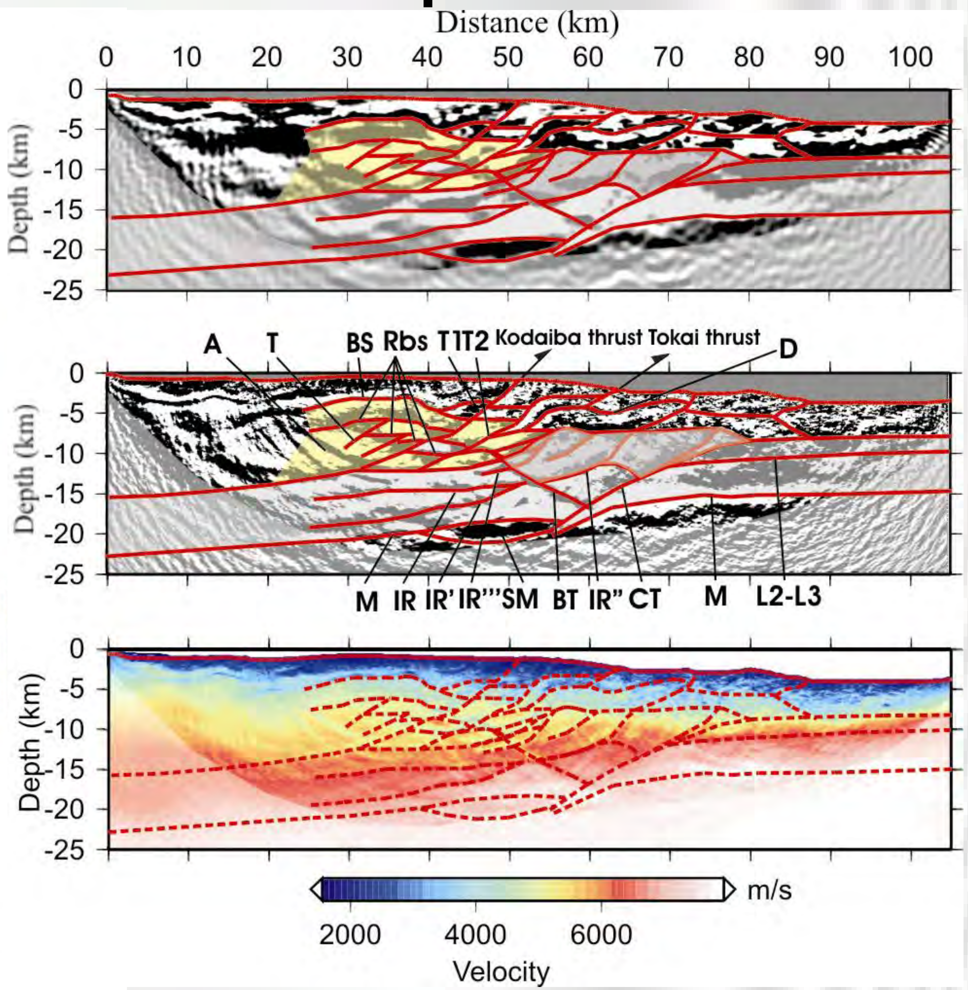
HOPE WAS BACK AGAIN !!!

Italy Baragiano



Crustal target (Operto et al, 2004)

Japan Nankai-Tokai



Lithospheric target (Operto et al, 2006)

Exploration Seismology

FWI is there as a tool!

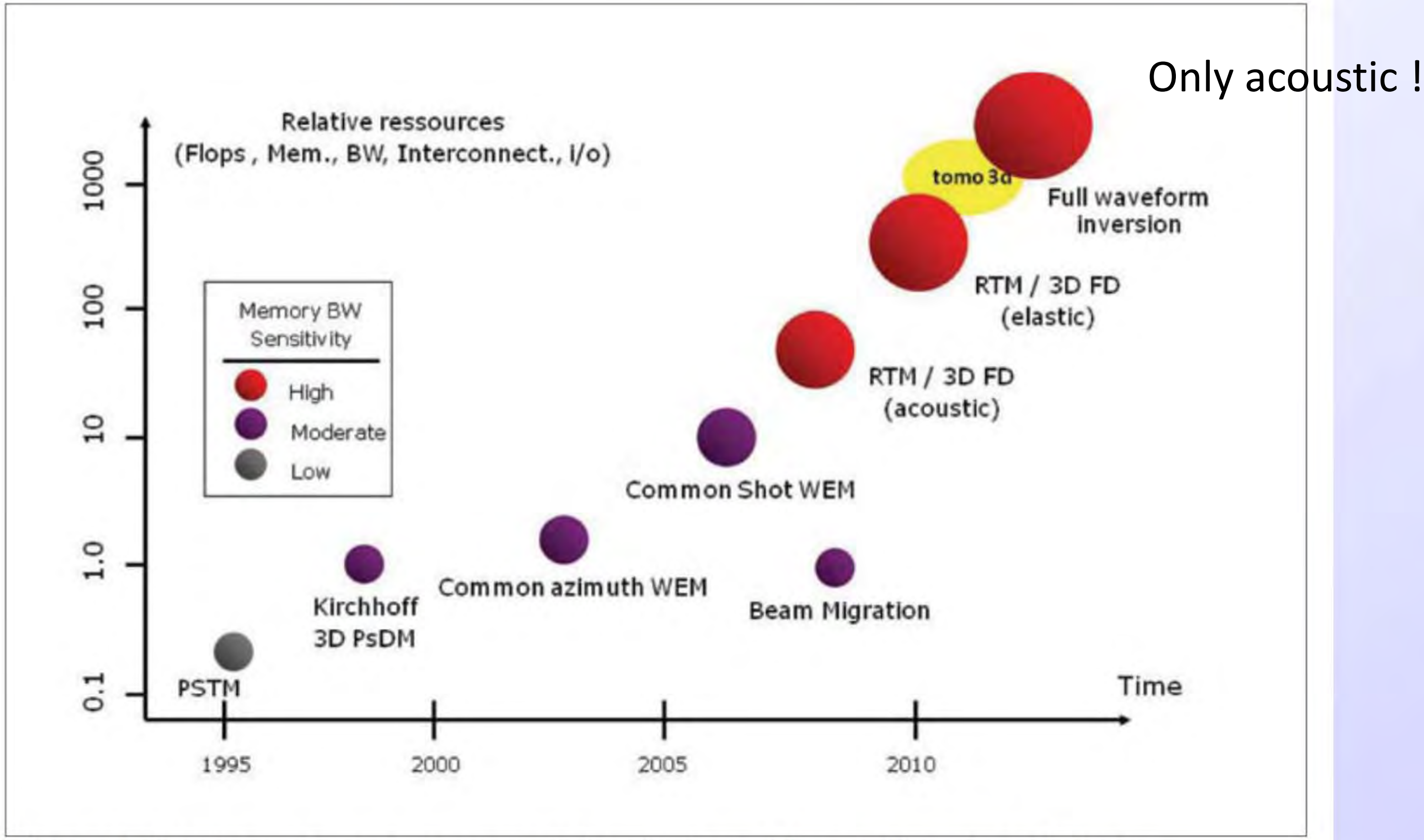


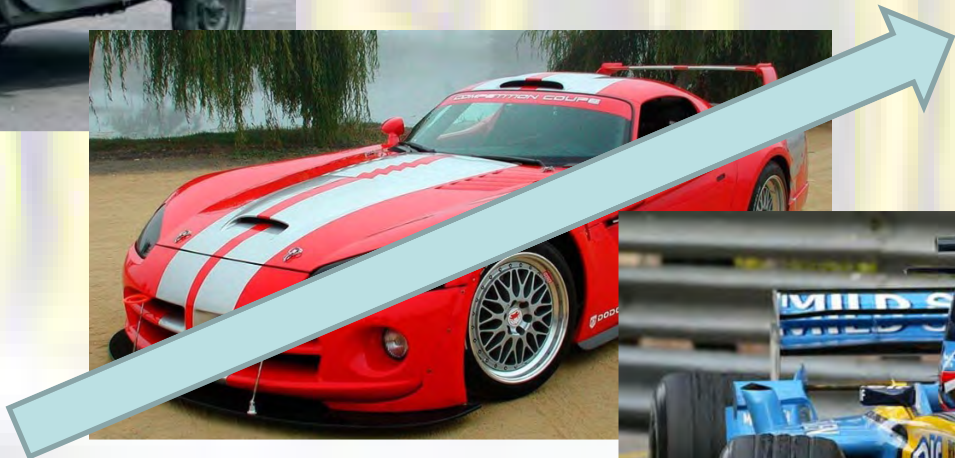
Figure 1. *Relative computing resource needs for seismic modeling and imaging techniques.*

(Camp & Thierry, The Leading Edge, 2010)

CLUSTERS DE CALCUL

Période Pascal Amand (1996)

Clusters de calcul de plus en plus compliqués !



Période Caroline Ramel (2001)

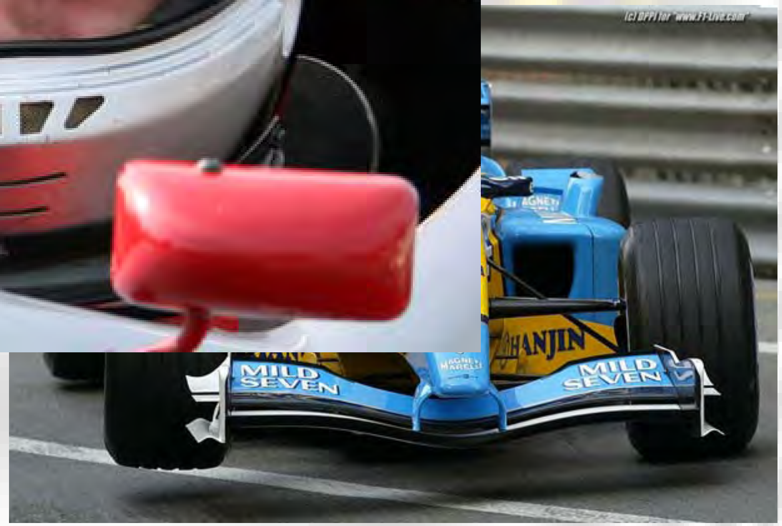
Période Alain Miniussi (2005)

CLUSTERS DE CALCUL

Période Pascal Amand (1996)

Clusters de calcul de plus en plus compliqués !

Un besoin de professionnels !



Période Alain Miniussi (2005)

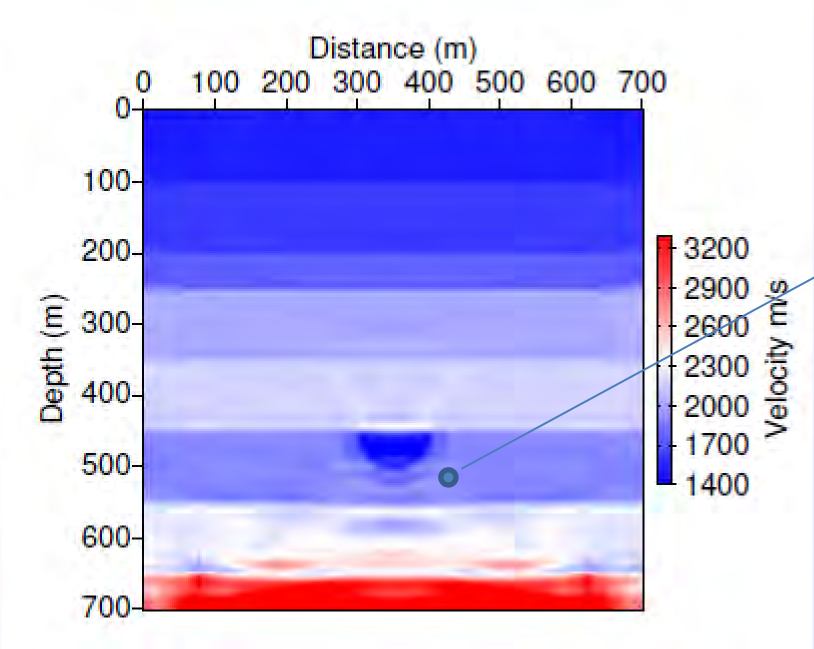
Introduction

- Most FWI applications deal with a single parameter reconstruction
- Density should have a strong impact on amplitude of reflection data
- Anisotropic parameters are involved on the modeling but not in the inversion (only few cases)

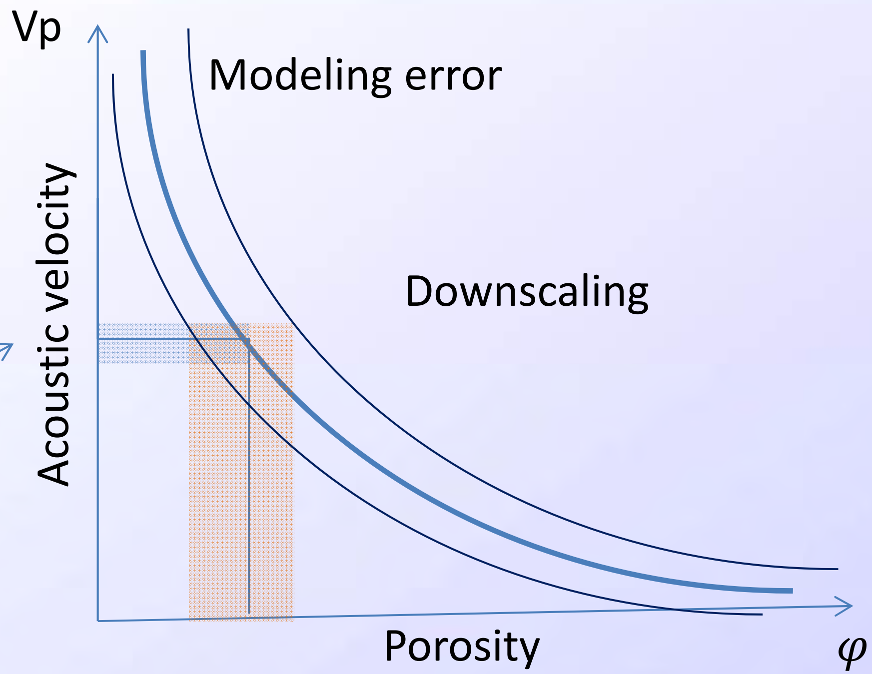
What do we need
for multi-parameter FWI?

Motivation

Macro-scale or FWI-imaged parameter



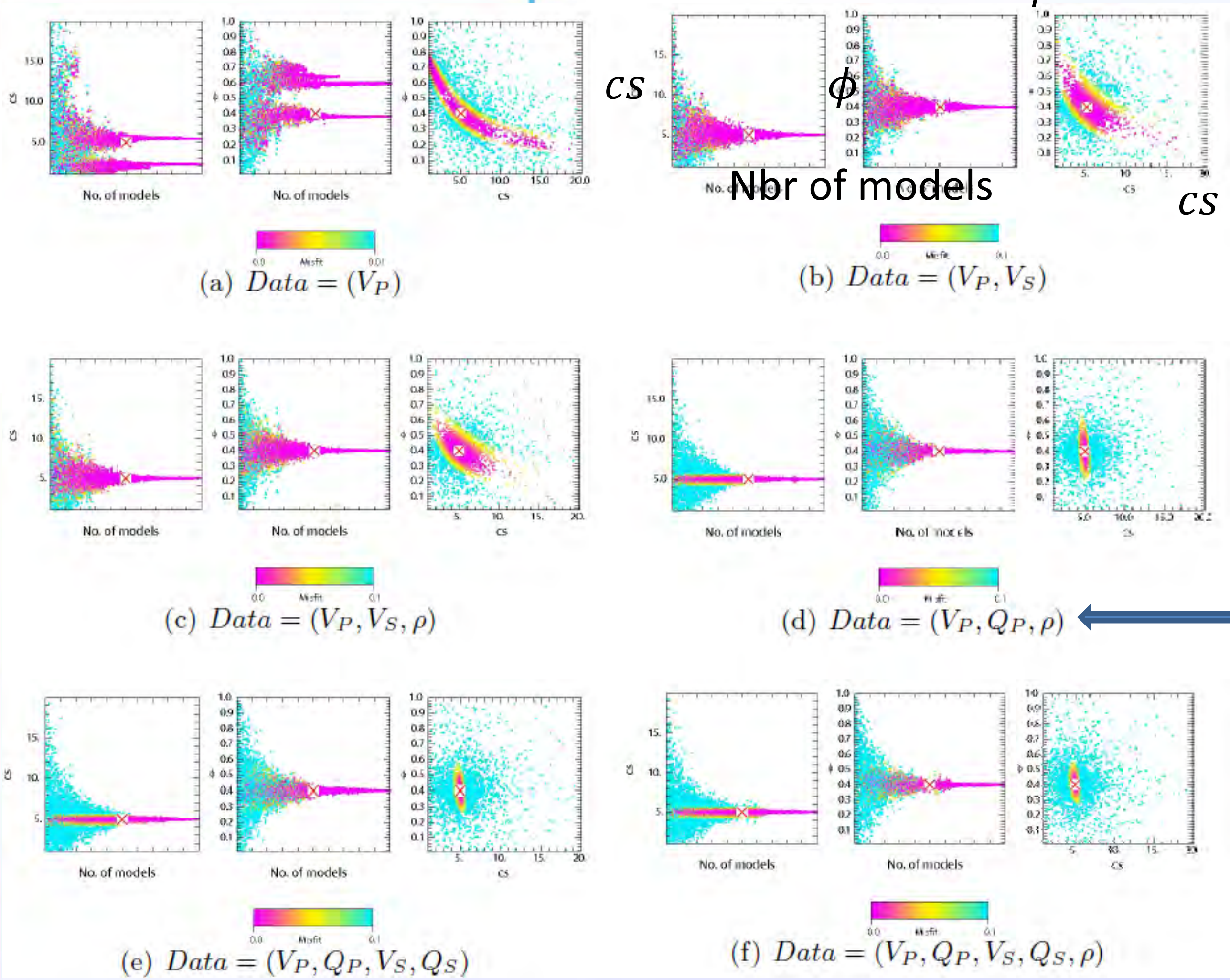
FWI reconstruction



Micro-scale parameter
or
« Local » parameter …

SEISCOPE

Downscaling



Dupuy (2011)

Reconstruction of

porosity ϕ and consolidation parameter cs

using Gassmann law

from FWI parameters

Even with only acoustic information

Importance of multiparametric reconstruction

Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling and regularization
- Hessian approximations
- Model impacts
- Conclusion

FWI: an optimisation problem in the data space

Misfit function definition: Ridge regression L2L2 (Tarantola, 1987)

$$C(m) = \frac{1}{2}\Delta d^{\dagger}\Delta d + \frac{1}{2}\,\varepsilon (m - m_{prior})^{\dagger}\,(m - m_{prior})$$

Updating the model at iteration k+1 from iteration k

$$m_{k+1} = m_k + \{\mathbb{H}\}^{-1}\Re\left[J^{\dagger}\,\Delta d_k + \varepsilon\Delta m_k\right]$$

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

Hessian approximation

$$\mathcal{H}^{-1} \sim \left(\Re\left(J^{\dagger} J\right) + \varepsilon I\right)^{-1}$$

$$\mathcal{H}^{-1} \sim \left(diag\left(\Re\left(J^{\dagger} J\right)\right) + \varepsilon I\right)^{-1} \ldots$$

(other approximations as pseudo-Hessian (Shin et al, 2001) or the l-BFGS approach (Nocedal & Wright, 1986))

The FWI mainly a data-driven technique

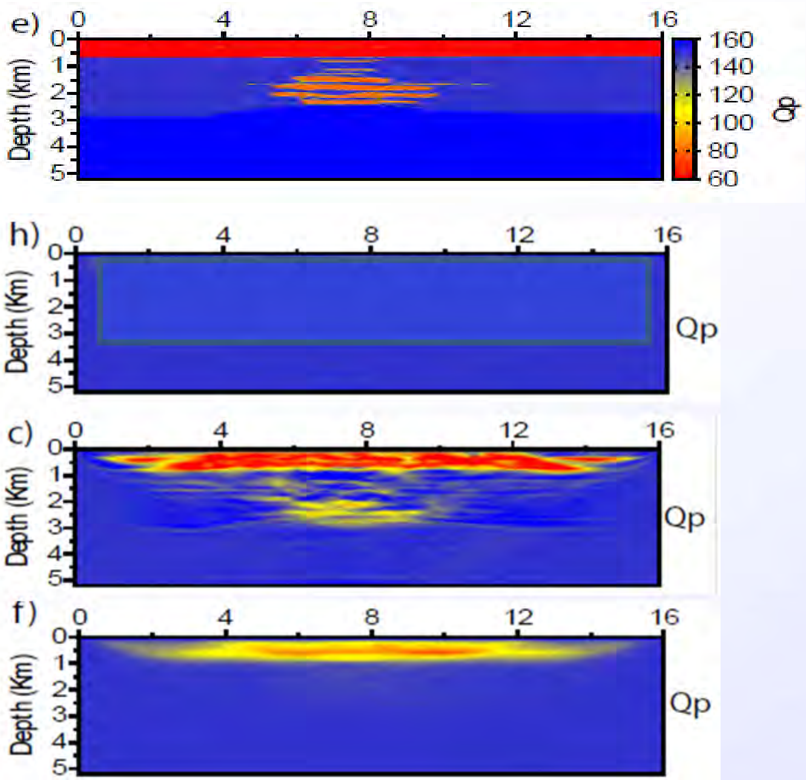
Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling and regularization
- Hessian approximations
- Model impacts
- Conclusion

FWI reconstruction of three parameters

Strategy A : simultaneous inversion

- Acoustic velocity Vp
- Density ρ
- Attenuation factor Qp



True model in Qp

Initial model in Qp (constant value=150)

Small regularization

Strong regularization

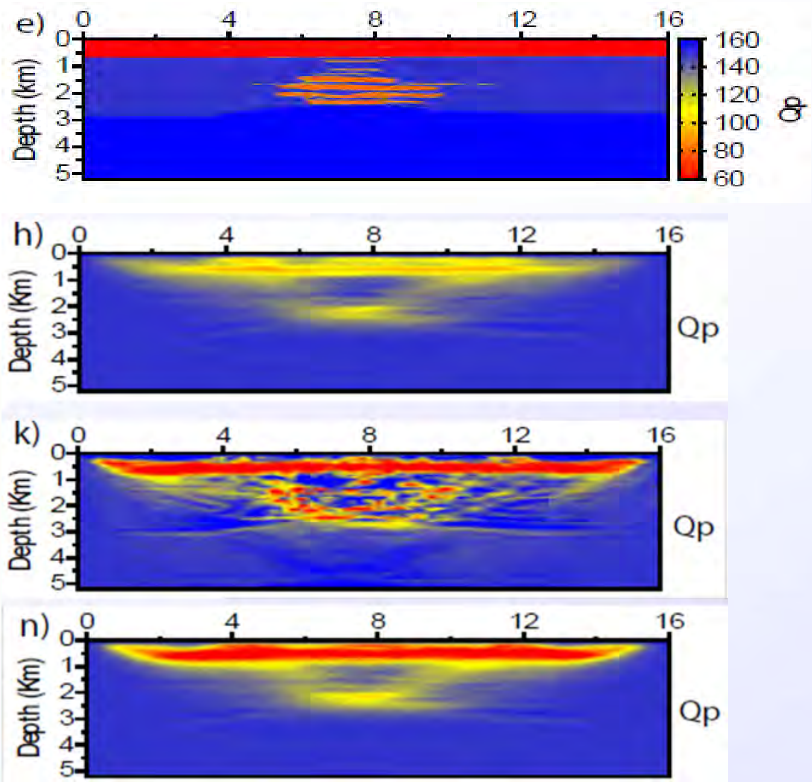
l-BFGS
quasi-Newton approach

Different FWI reconstructions of the parameter Qp for synthetic Valhall dataset
(Prieux, 2013)

FWI reconstruction of three parameters

Strategy B : two-steps inversion

1 step (Vp, Qp)
2 step (Vp, ρ, Qp)



True model in Qp

Starting model of Qp after the first step

Small regularization

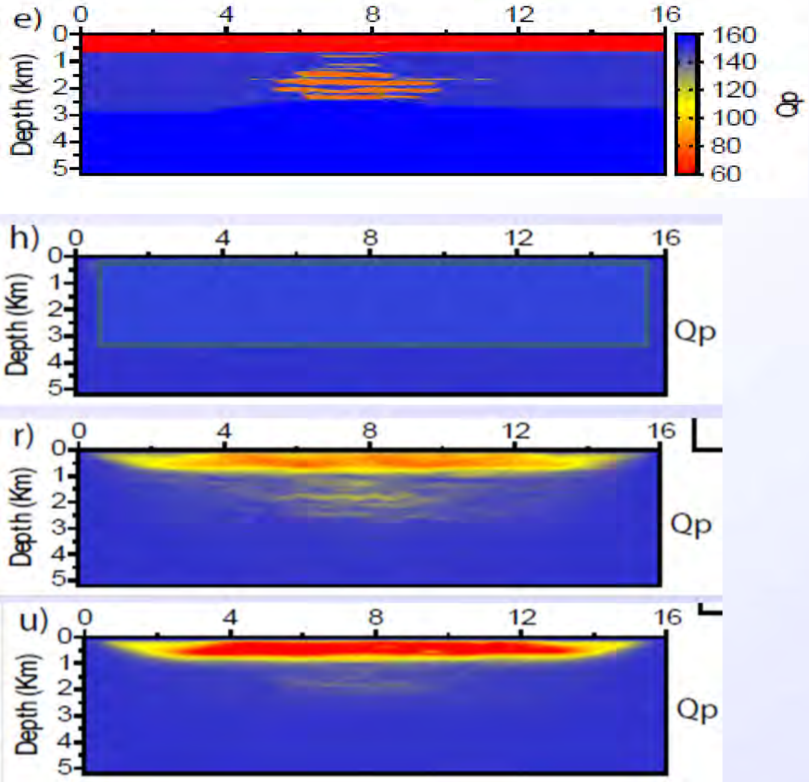
Strong regularization

Different FWI reconstructions of the parameter Qp for synthetic Valhall dataset (Prieux, 2013)

FWI reconstruction of three parameters

Strategy C : two-steps inversion

1 step (Vp)
2 step (Vp, ρ, Qp)



True model in Qp

Starting model of Qp (150 constant value)

Small regularization

Strong regularization

Different FWI reconstructions of the parameter Qp for synthetic Valhall dataset (Prieux, 2013)

Sensitivity of data

Seismic data may have various imprints of parameters and we must consider these variable imprints for better reconstruction

Deciphering the optimisation formulation

- Hunting for the Hessian inverse operator
- Reducing the model sampling (prior info.)

Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling and regularization
- Hessian approximations
- Model impacts
- Conclusion

Gradient structure

$Bu = s$ Forward problem with a symmetric matrix B where the RHS is zero except at sources: the source term is connected to the definition of B

$B^* a = R^t \Delta d$ Adjoint problem where the RHS is zero except at receivers with a restriction operator R

Incident field — Adjoint field

$$\gamma_l = -\sum_{j'}^{S} \sum_{j}^{\omega} \Re \left\{ {u_{j,j'}}^t \frac{\partial B_j}{\partial m_l} {a_{j,j'}}^* \right\}$$

Cross section

- Propagation combines diffraction/reflection and transmission features

- Radiation acts as a local operator on material properties

The so-called FWI democracy

Gradient structure

$Bu = s$ Forward problem with a symmetric matrix B where the RHS is zero except at sources: the source term is connected to the definition of B

$B^* a = R^t \Delta d$ Adjoint problem where the RHS is zero except at receivers with a restriction operator R

Incident field — Adjoint field

$$\gamma_l = -\sum_{j'}^{S} \sum_{j}^{\omega} \Re \left\{ {s^t}_{j'} \, {B_j}^{-1} \frac{\partial B_j}{\partial m_l} {B_j}^{-1} R^t \, {\Delta d^*}_{j'} \right\}$$

Cross section

- Propagation combines diffraction/reflection and transmission features

- Radiation acts as a local operator on material properties

The so-called FWI democracy

Cross sections
(radiation, scattering …)

Isotropic acoustic FWI

At least three possible classes

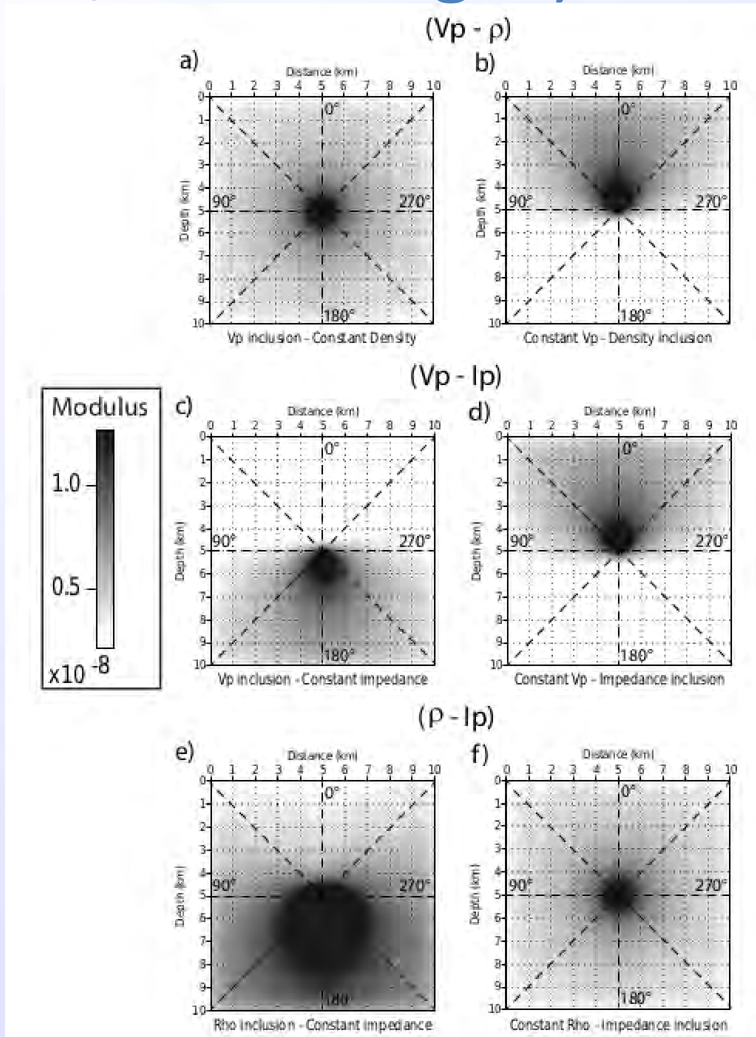
From P.D.E., setting (V_p, ρ)

From reflection, setting (V_p, I_p)

Alternative setting (ρ, I_p)

Narrow offsets: same radiations from velocity and density …

In fact, more complex trade-off in the FWI …



Normal equations: scaling

Updating the model requires an accurate estimation of the Hessian

$$\Delta m = \left[\Re\{J^t J^*\} + \Re\left\{ \frac{\partial J^t}{\partial m^t} (\Delta d^* \ldots \Delta d^*) \right\} + \varepsilon I \right]^{-1} \Re\{J^t \Delta d^*\}$$

Dimensionality analysis

Only the action of the Hessian on the gradient gives the physical units of the model perturbation

$$[M] = \left(\frac{[M][M]}{[D][D]} \right) \left(\frac{[D]}{[M]} [D] \right)$$

Hessian structure (often an approximation with some smooth regularization)

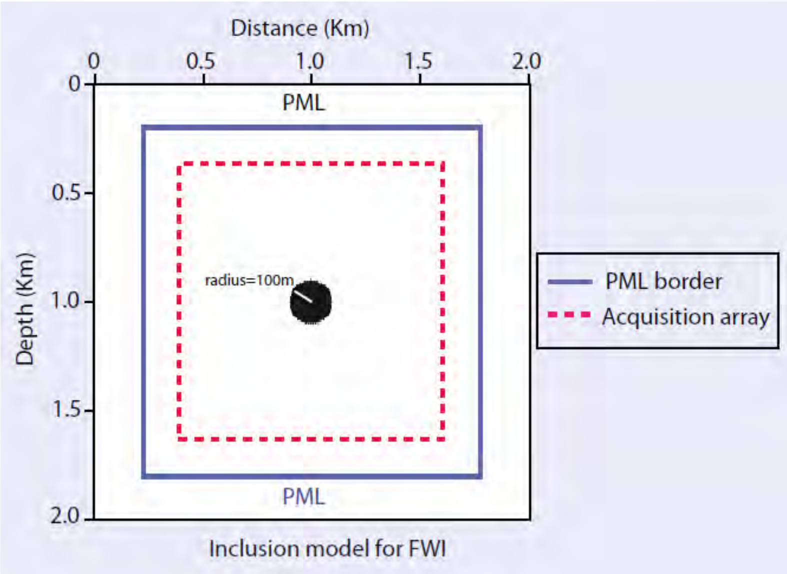
Model parameter description such that the Hessian has a more or less diagonal shape in order to cancel out trade-offs, if possible.

SVD decomposition (Plessix and Cao, 2011)
Resolution analysis (Fichtner & Trampert, 2011)
Asymptotic analysis (Jin et al, 1992, Lambaré et al, 2003)
Multi-scale investigation (Op't Root et al, 2012)

Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling (and regularization)
- Hessian approximations
- Model impacts
- Conclusion

A very simple example



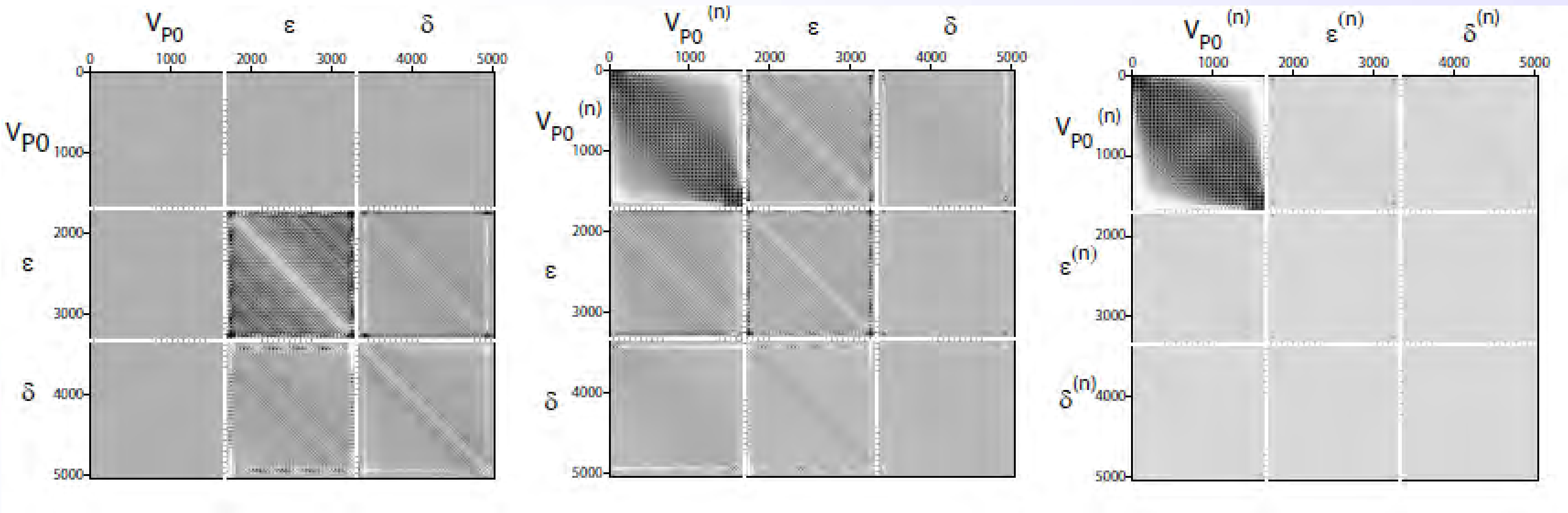
Three parameters to be inverted in this particular case of anisotropic paramters

- Vp Acoustic velocity
- ε Thomsen parameter
- δ Thomsen parameter

Gholami et al (2013)

Scaling influence

Gholami et al (2013)

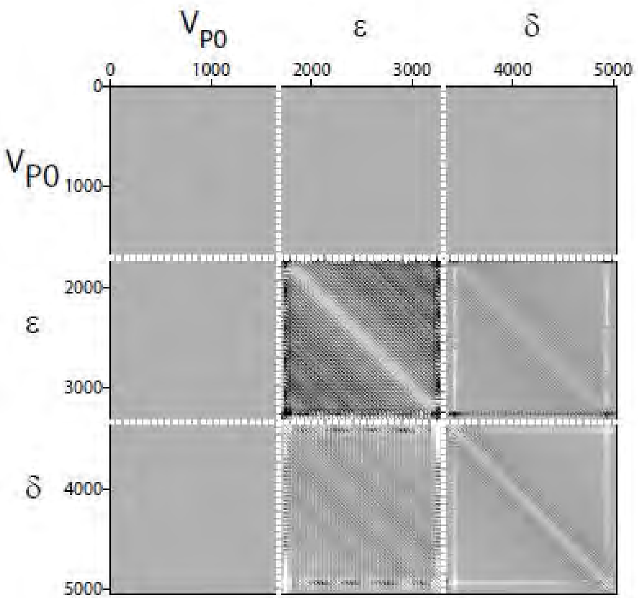


No scaling

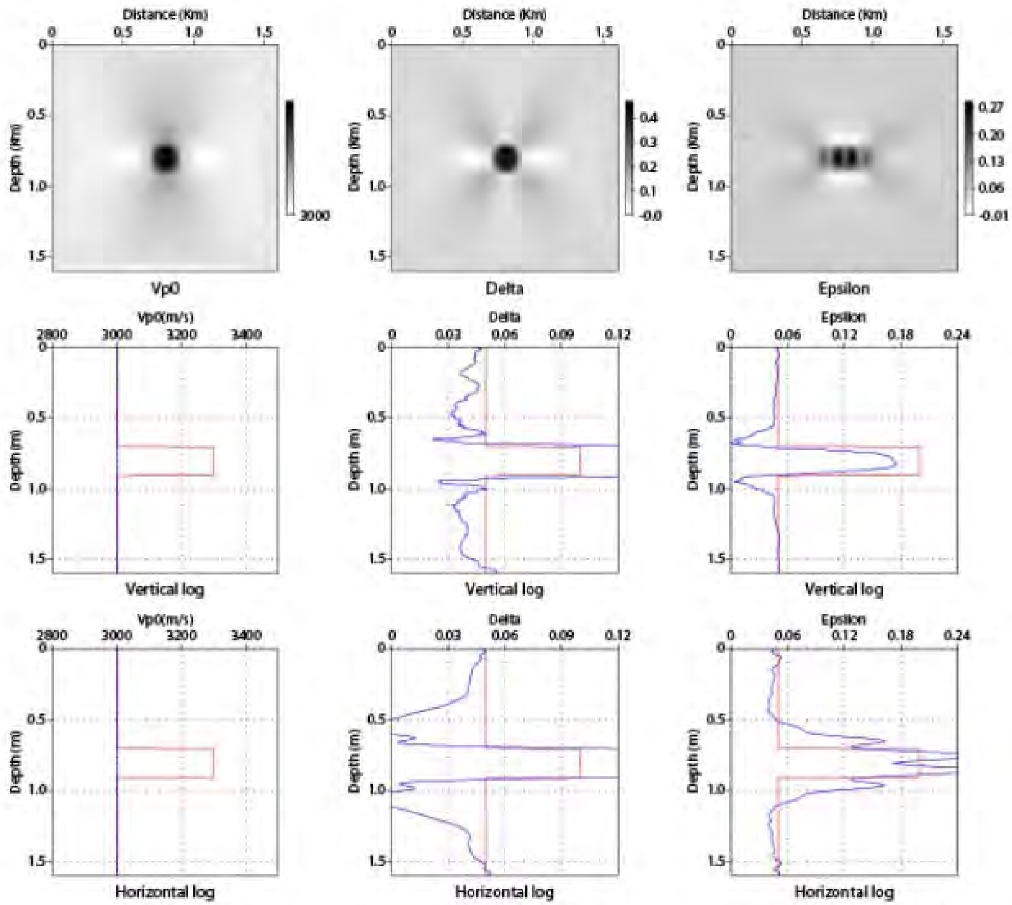
Vp is scaled by its background

All classes are scaled by their background

No scaling

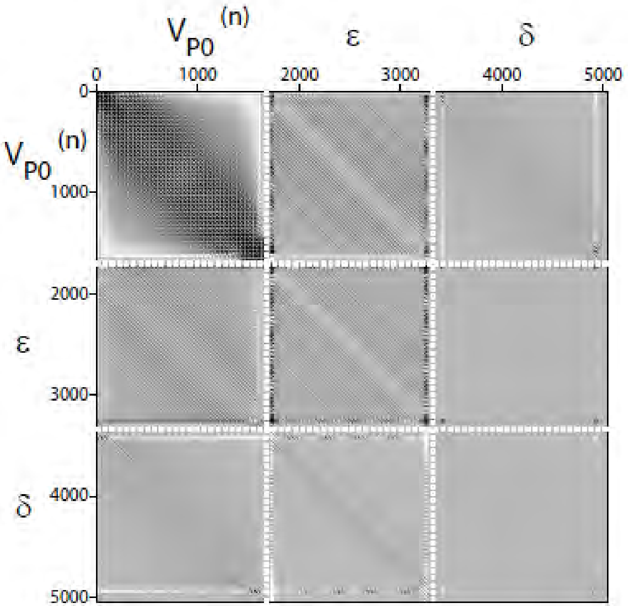


V_{P0}
ε
δ
0
1000
2000
3000
4000
5000

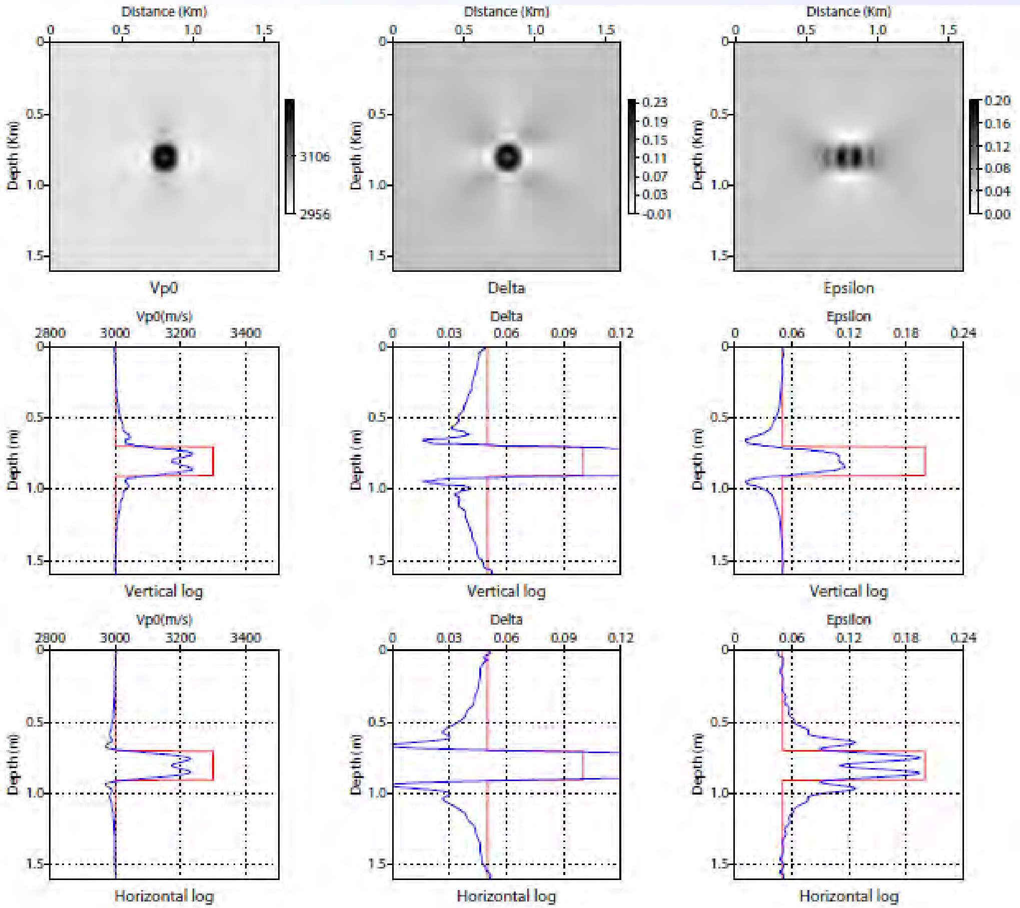


Distance (Km)
Depth (Km)
Vp0
Delta
Epsilon
Vertical log
Horizontal log
Vp0(m/s)
Depth (m)

Vp scaling

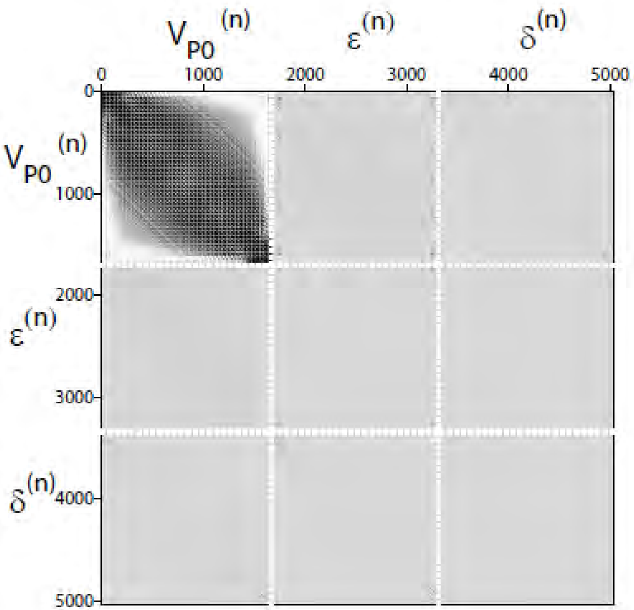


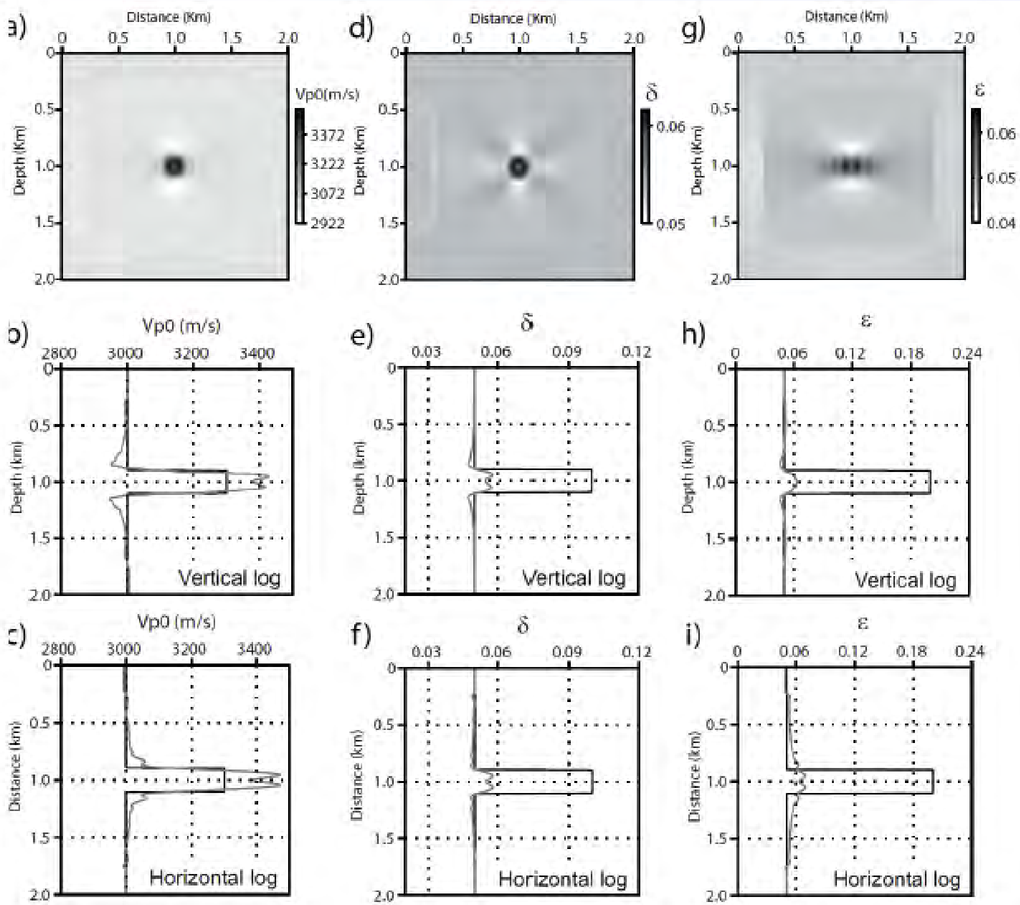
V_P0 (n)
ε
δ
0
1000
2000
3000
4000
5000



Distance (Km)
Depth (Km)
Vp0
Delta
Epsilon
3106
2956
Vp0(m/s)
Depth (m)
Vertical log
Horizontal log

Class scaling





Scaling is required although Hessian operator should take it into account

Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling and regularization
- Hessian approximations
- Model impacts
- Conclusion

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 point sampling (illumination)

BUT … trade-off and secondary scattering effects are key issues when considering multi-parameters

$$\mathcal{H}_{ij} = \Re\left\{J^{\dagger}J_{ij} + \sum_{k=1}^{n} \frac{\partial^2 u_k}{\partial m_i \partial m_j} \Delta d_k{}^{*}\right\}$$

H(m)=B(m)+C(m)

B(m) => Gauss-Newton method

B(m)+C(m) => Quasi-Newton and full-Newton method

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

Brossier et al (2008)

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

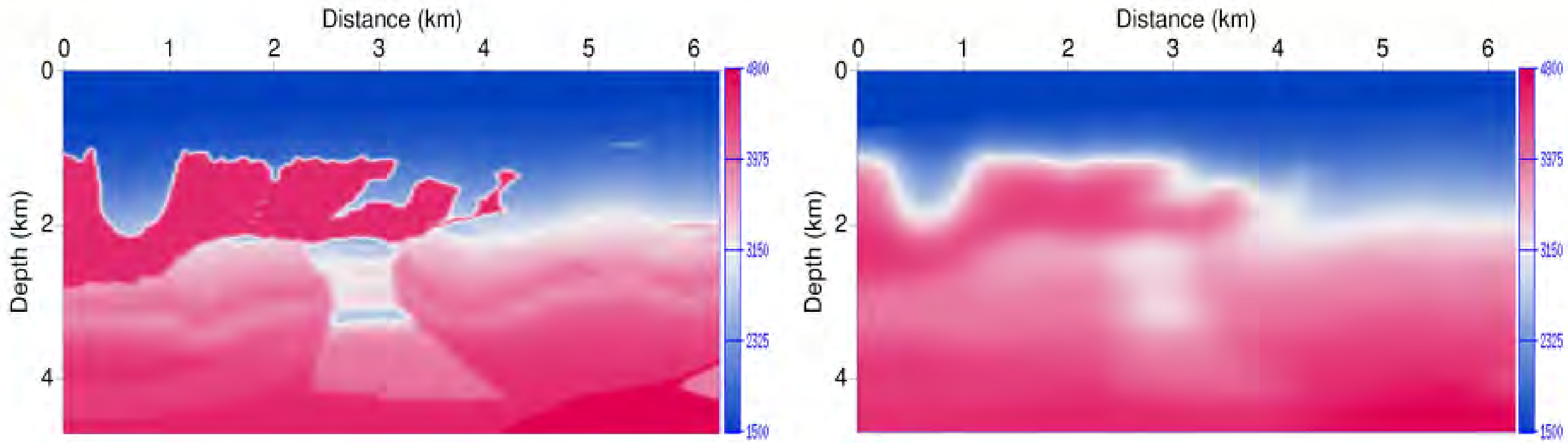
- Truncated Newton is solving (1) while Hessian could not lead to a quadratic form (Nash, 2000) (four modeling ~ Gauss-Newton overload)

Métivier et al (2013)

BP 2004 model

Configuration

- 62 sources each 100 m, 248 receivers each 25 m, free surface condition
- 6 overlapping groups of 4 frequencies from
  - 1 Hz to 19 Hz (experiment 1)
  - 2 Hz to 20 Hz (experiment 2)
  - 2.5 Hz to 20.5 Hz (experiment 3)



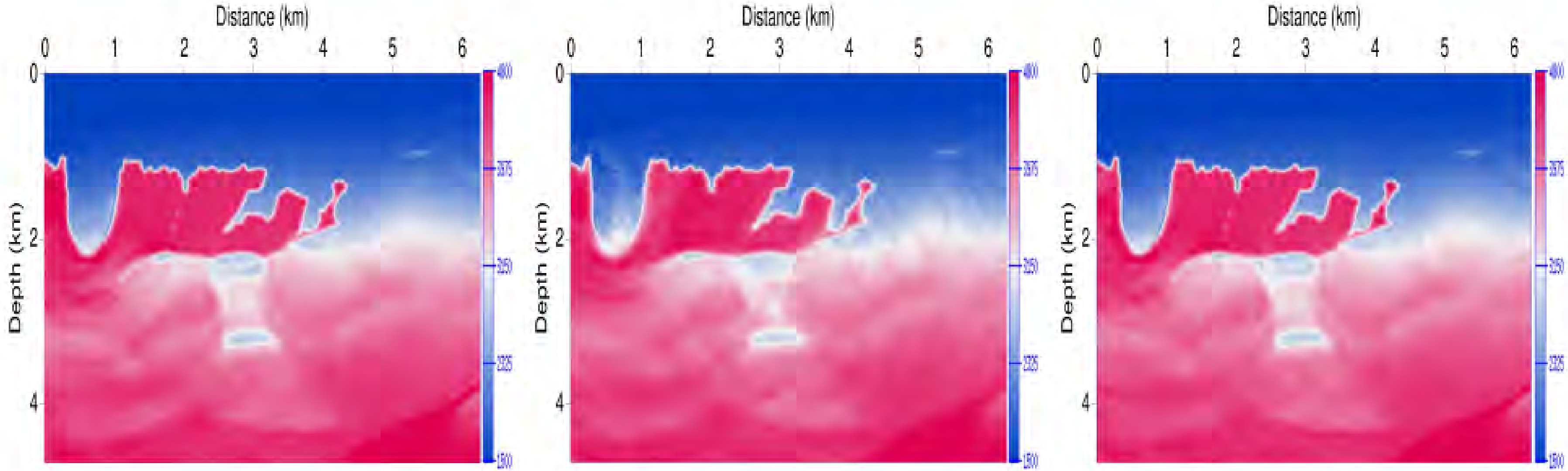
Exact model (left), initial smooth model (right)

Low frequency content

Experiment 1

- Inversion of 6 overlapping groups of 4 frequencies from 1 Hz to 19 Hz

Three methods work well !



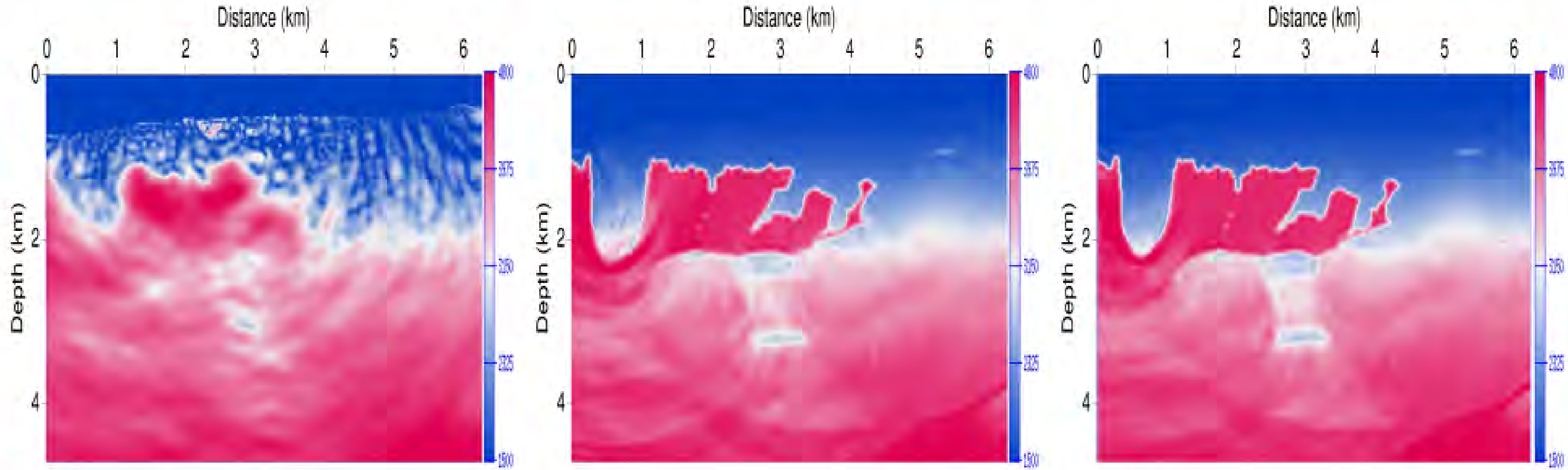
Estimated P-wave velocity models: preconditioned *l*-BFGS (left), preconditioned nonlinear conjugate gradient (middle), preconditioned truncated Newton (right)

Intermediate frequency content

Experiment 2

- Inversion of 6 overlapping groups of 4 frequencies from 2 Hz to 20 Hz

L-BFGS method has difficulties because of high contrasts !



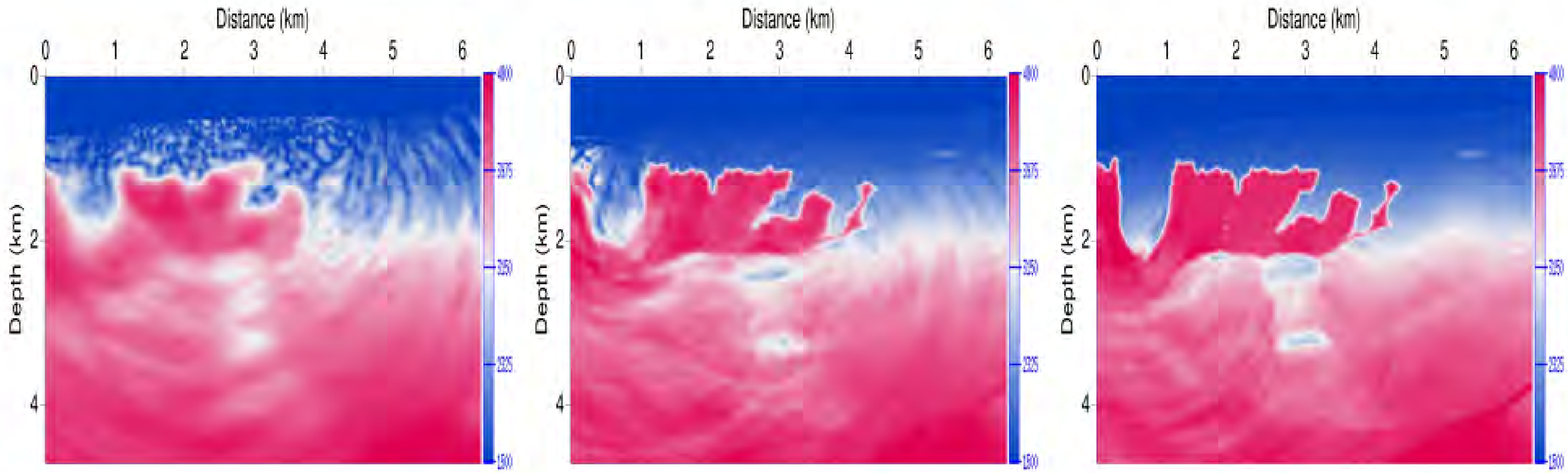
Estimated P-wave velocity models: preconditioned *l*-BFGS (left), preconditioned nonlinear conjugate gradient (middle), preconditioned truncated Newton (right)

Higher frequency content

Experiment 3

- Inversion of 6 overlapping groups of 4 frequencies from 2.5 Hz to 20.5 Hz

Truncated Newton method gives the best result !



Estimated P-wave velocity models: preconditioned *l*-BFGS (left), preconditioned nonlinear conjugate gradient (middle), preconditioned truncated Newton (right)

When doing multi-parametric reconstruction, no compromise on the Hessian inverse operator !

We expect a mitigation of trade-off effects between parameters while considering truncated Newton methods: relax constrains on scaling (on-going work ☺) (Lavoué/Bretaudeau … in our group)

Outline

- Introduction
- FWI ingredients
- Multi-parameter reconstruction Qp
- Gradient and Hessian features
- Scaling and regularization
- Hessian approximations
- Model impacts
- Conclusion

MODEL CONTRIBUTION

A regularization approach

The misfit function: a dimensionless quantity

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

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

Ridge regression+Tikhonov+ Prior influence = L2 L2

(Tikhonov and Arsenin, 1977)

Using l-BFGS-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 (easy to introduce in existing FWI workflow)

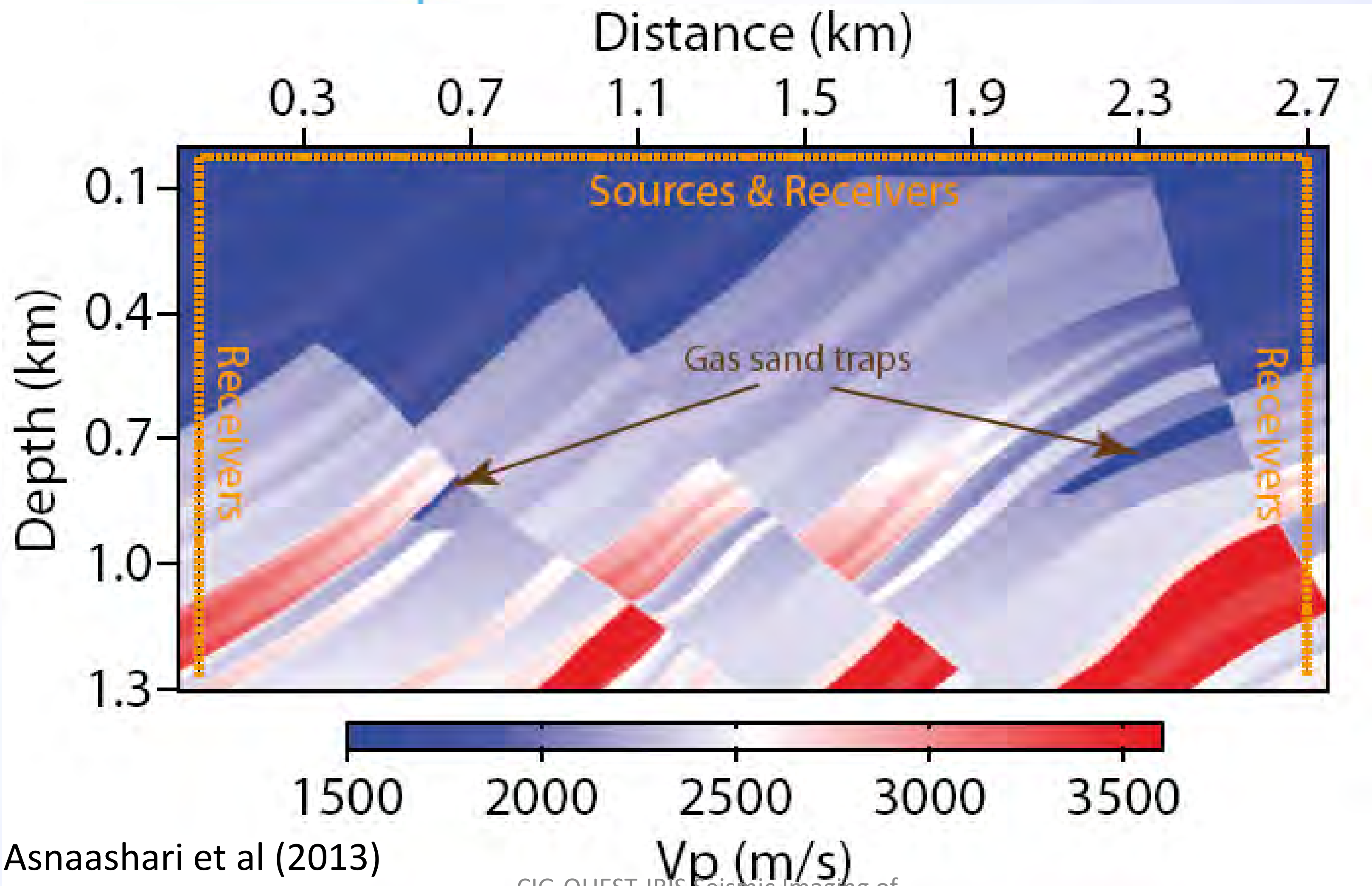
Estimation of hyper-parameters ?
Lasso regression: L2 L1 (preserving the sparsity in model space …) ?

New ingredients

- Hyper-parameters
  - Smoothing λ_1 (it has a dimension if misfit has)
  - Prior importance λ_2
- Weighting matrices or covariance matrices
  - Wd is diagonal because seismic data are not correlated (simple hypothesis): used for normalization of the data
  - Wm is more difficult to design and an easy strategy is through the two terms

Zoom on the Marmousi II

(Martin et al, 2006)

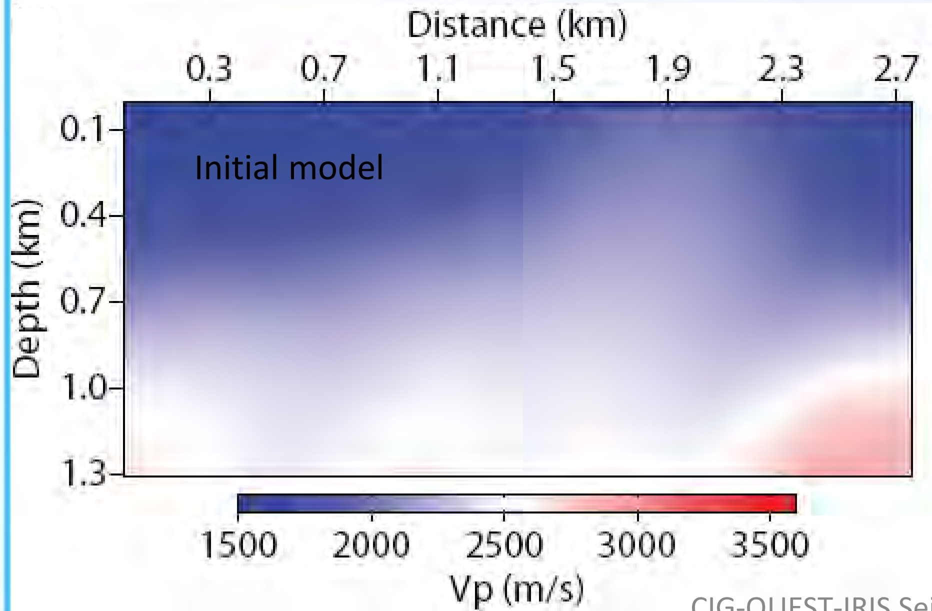


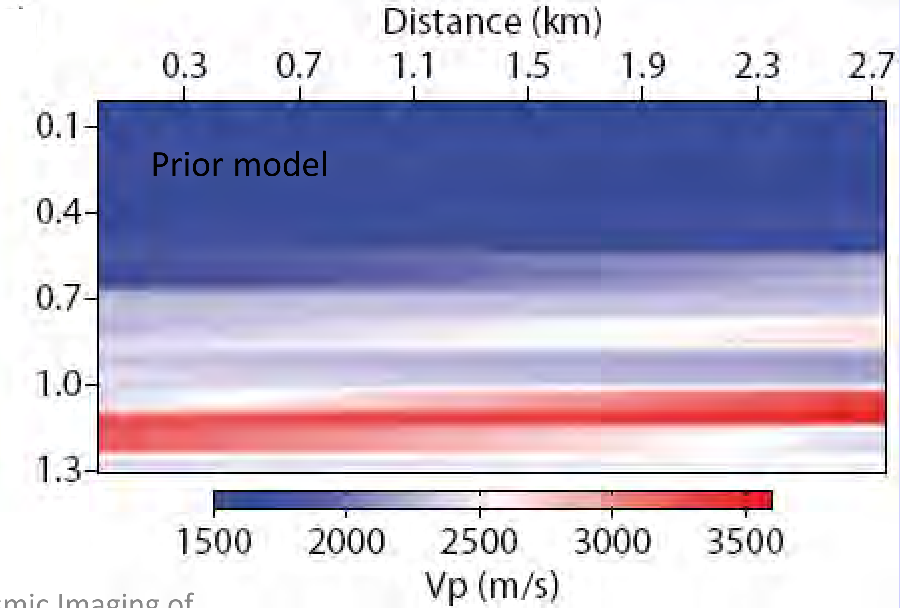
Asnaashari et al (2013)

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



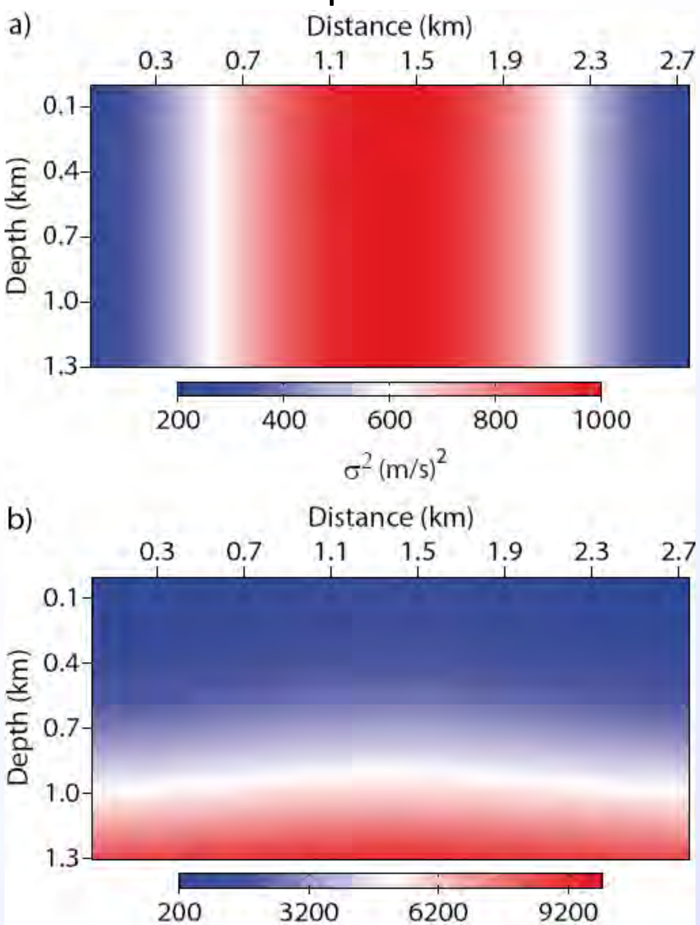


Wm definition

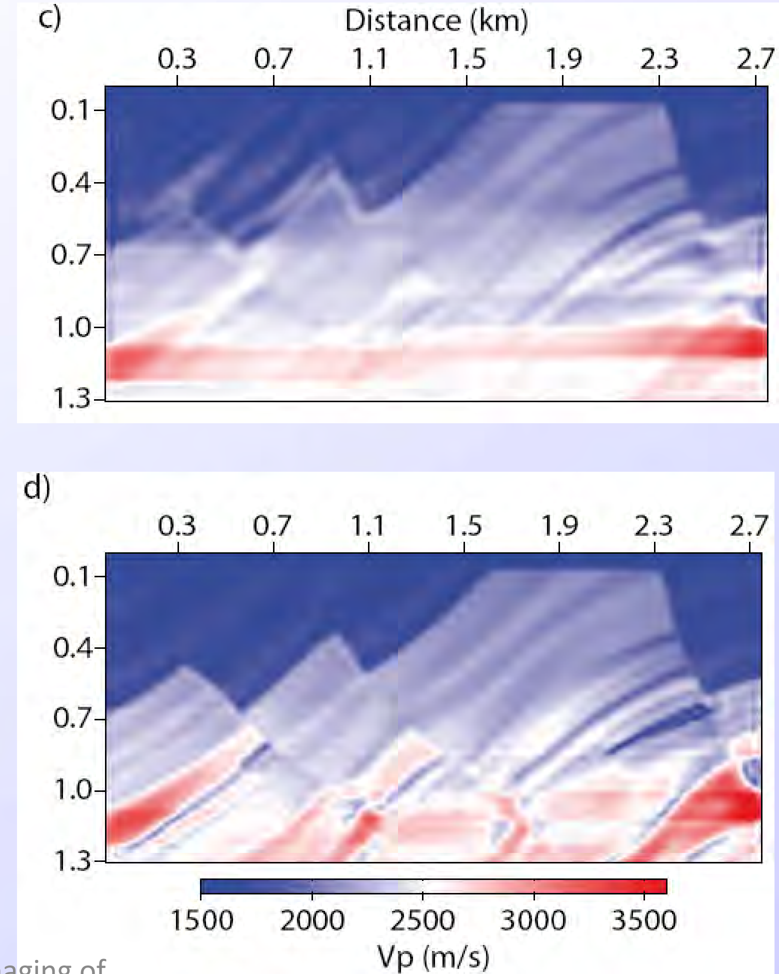
Selected hyper-parameters

$\lambda_1 = 20\ sec^2$ and $\lambda_2 = 3 \times 10^5$

Two wells prior information

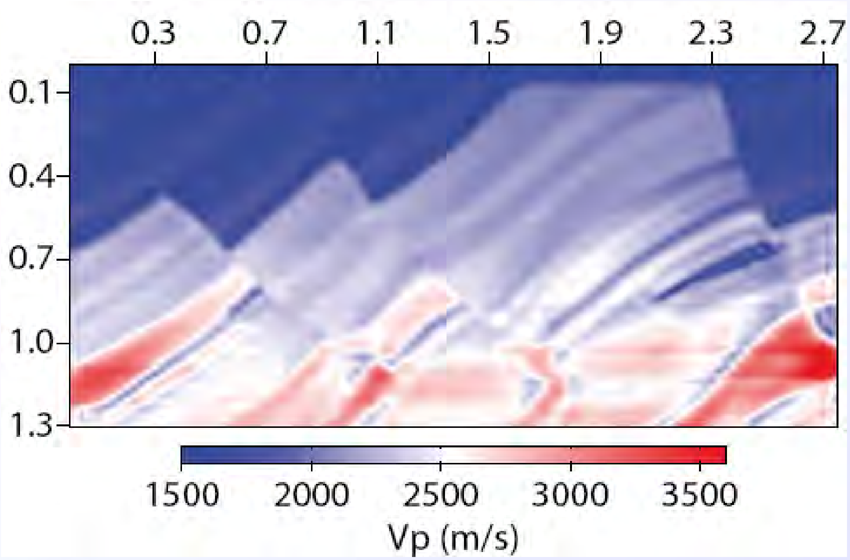


combined with a depth variation

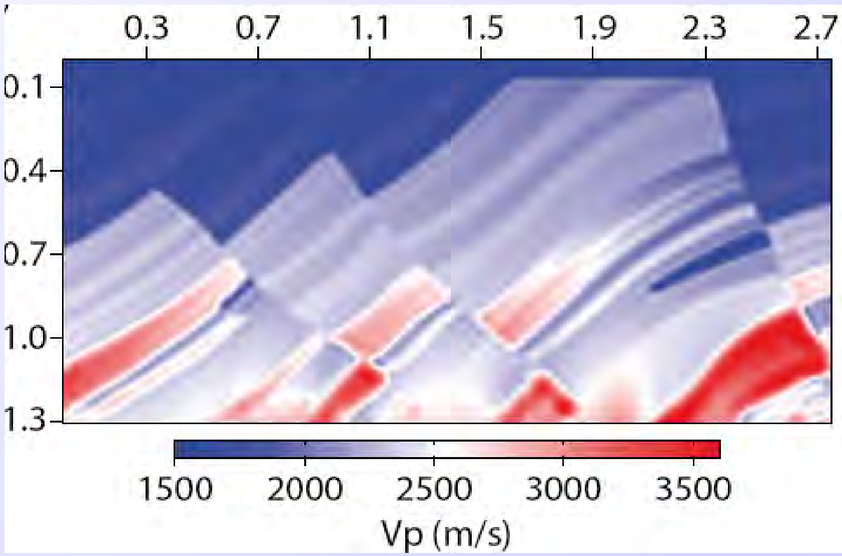


Only data should speak

- 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.

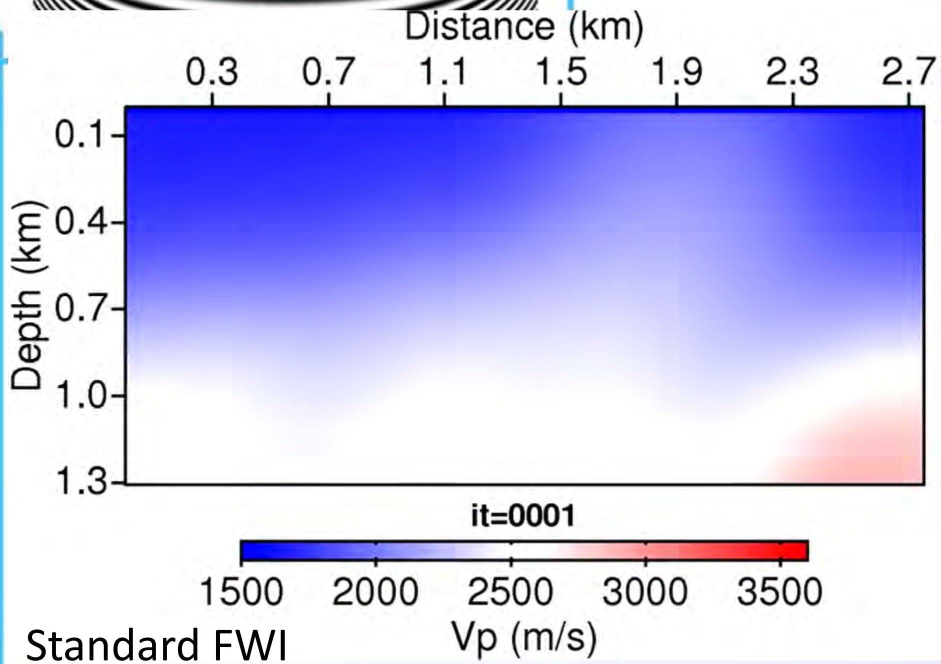


Fixed prior weight



Dynamic prior weight

Model-driven reconstruction



Standard FWI

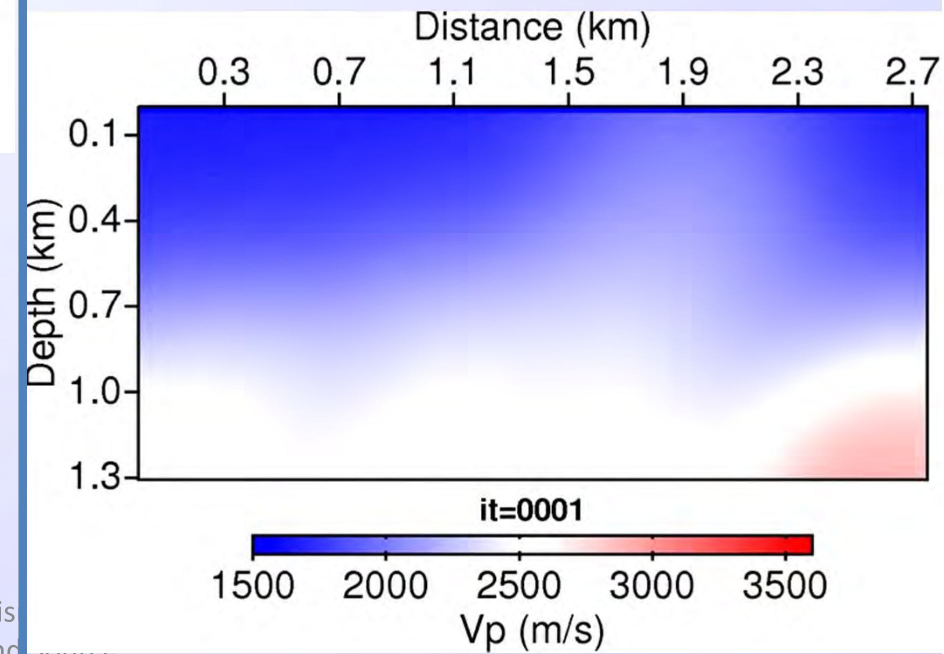
$$C(m) = \frac{1}{2}\Delta d^{\dagger} W_d \Delta d + \frac{1}{2}\lambda_1\, m^t D m$$

Data misfit only with simple regularization (Tikhonov term)

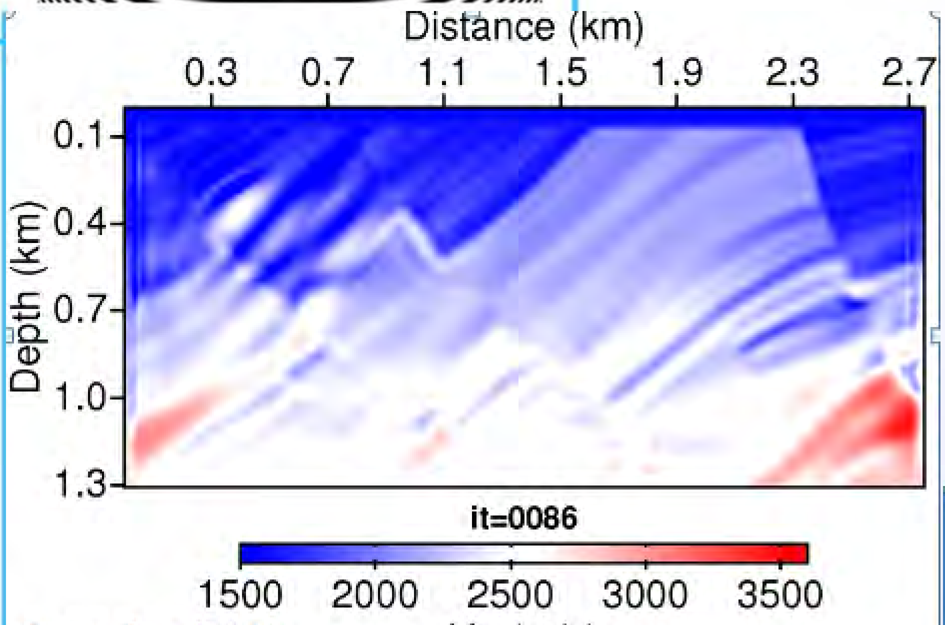
Automatic reduction of hyper parameter

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

Prior model/dynamic weight



Model-driven reconstruction



Standard FWI

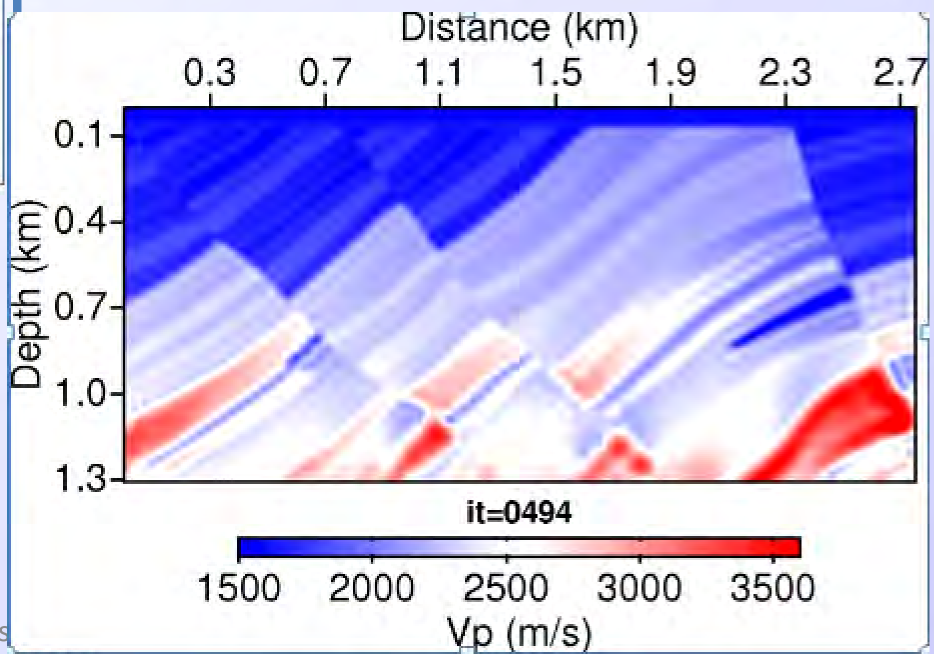
$$C(m) = \frac{1}{2}\Delta d^{\dagger} W_d \Delta d + \frac{1}{2}\lambda_1\, m^t D m$$

Data misfit only with simple regularization (Tikhonov term)

Automatic reduction of hyper parameter

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

Prior model/dynamic weight



Prior information should improve our reconstruction of parameters with different imprints in the seismic data

- Density
- Anisotropy parameters
- Attenuation parameters
- Elastic parameters

Conclusion

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

- Multi-parameter reconstruction is feasable as long as

- Accurate estimation of the Hessian inverse operator
- Prior model information for balancing the different imprints of parameters

Once HR multi-parametric imaging is validated on synthetic and real data, downscaling extraction of « local » parameters (porosity, saturation, consolidation parameter etc) could be performed if an experimental or theoretical downscaling law is considered

SEISCOPE

FWI ↗ $\lambda/2$

Thank you very much for your attention

Many thanks to sponsors

bp
CGG Passion for Geoscience
Chevron Human Energy
ExxonMobil
JGI
ارامكو السعودية Saudi Aramco
Shell
中国石化 SINOPEC
Statoil
TOTAL
WesternGeco

of the Seiscope consortium

One-day lecture

pdf file for these lecture notes for those interested by it

Lecture Notes on Full Waveform Inversion

Romain Brossier and Jean Virieux
Romain.Brossier@ujf-grenoble.fr Jean.Virieux@ujf-grenoble.fr
Université Joseph Fourier, Grenoble, France

SEISCOPE CONSORTIUM
http://seiscope2.osug.fr
see DOCUMENTS/LECTURES

Seiscope contribution

Gholami, Y., Brossier, R., Operto, S., Prieux, V., Ribodetti, A., and Virieux, Jean, 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? - part 2: application to Valhall, *Geophysics*, **2**, R107–R124.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A., and Virieux, Jean, 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? - part 1: sensitivity and trade-off analysis, *Geophysics*, **2**, R81–R105.

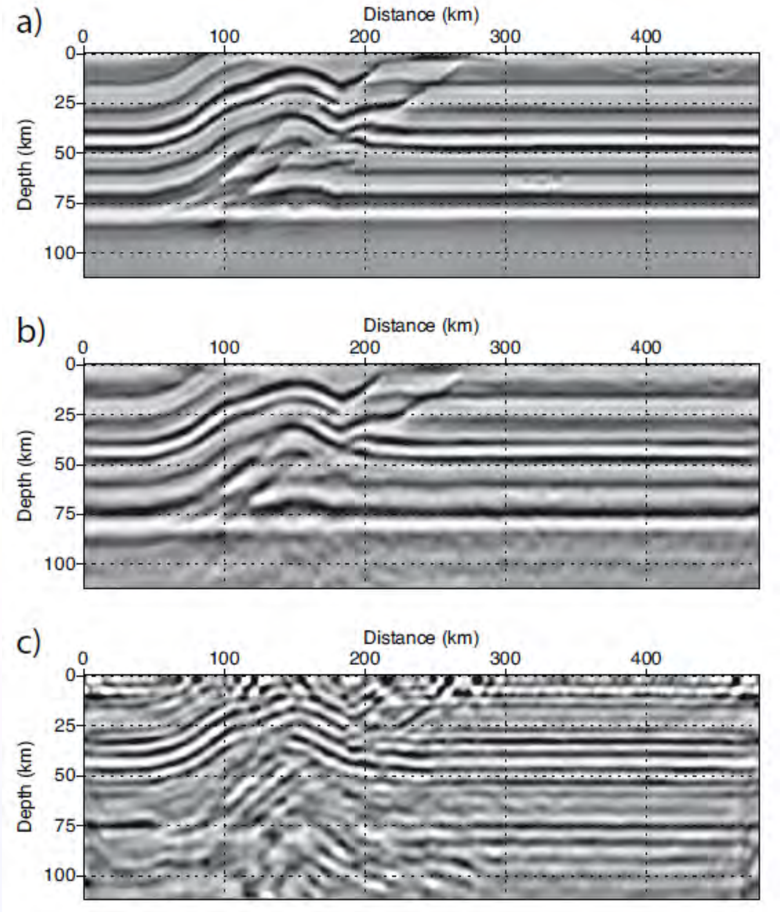
Malinowski, M., Operto, S., and Ribodetti, A., 2011. High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion, *Geophysical Journal International*, **186**(3), 1179–1204.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O.I., and Kommedal, J.H., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophysical Journal International*, **187**, 1495–1515.

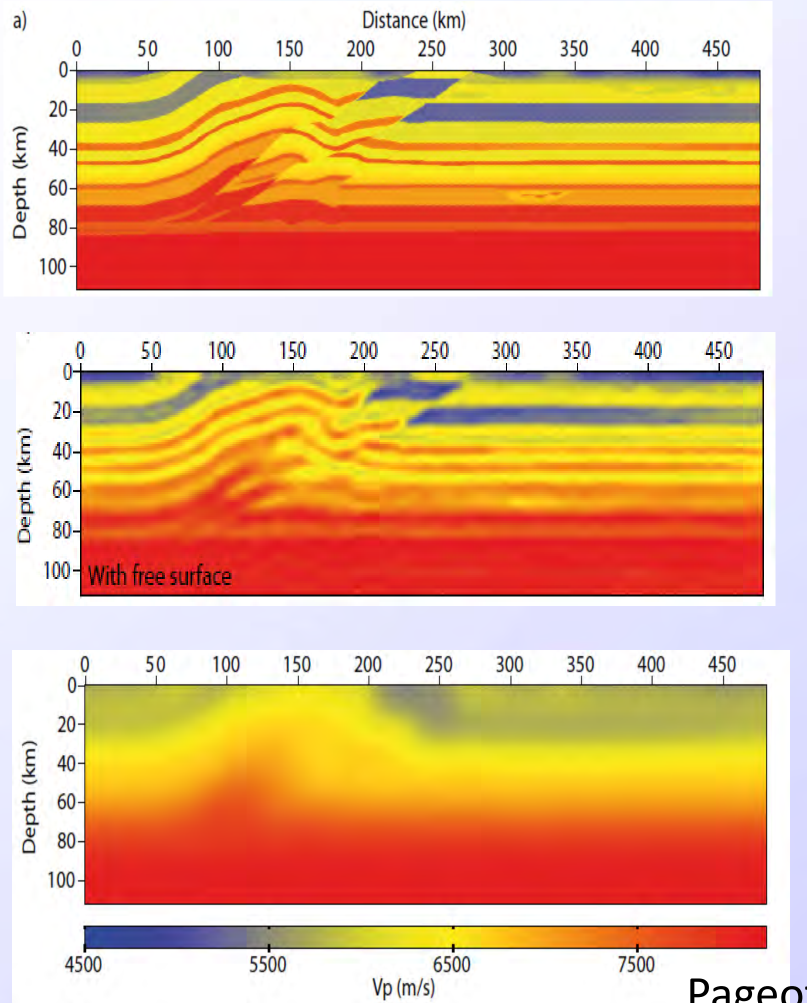
Prieux, V., Brossier, R., Operto, S., and Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent OBC data from valhall. part 1: imaging compressional wavespeed, density and attenuation, *Geophysical Journal International*, **in press**.

Prieux, V., Brossier, R., Operto, S., and Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent OBC data from valhall. part 2: imaging compressional and shear-wave velocities, *Geophysical Journal International*, **in press**.

Macromodel building for improved migration



RTM



Exact

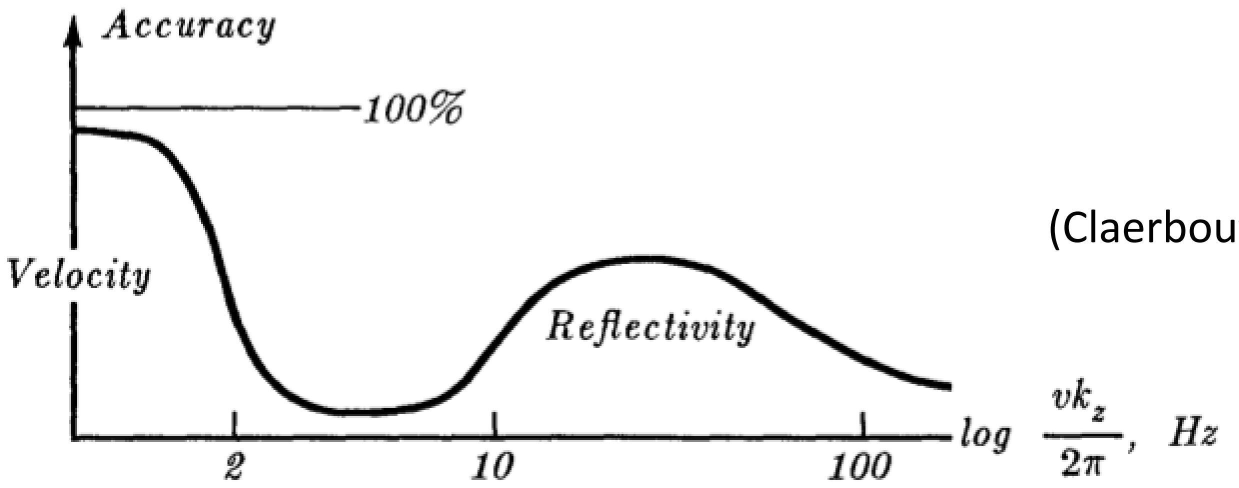
FWI

Initial

Pageot et al (2013)

Separation between two scales

Recorded seismic traces bring different information from medium properties



(Claerbout, 1984)

FIG. 1.4-3. Reliability of information obtained from surface seismic measurements.

Macro-model Velocity from low frequency content
Impedance/Reflectivity from high frequency content

Various signatures of parameters on data

Separation between two scales

Recorded seismic traces bring different information from medium properties

How to define the initial model for FWI?

(We do not address this problem in this presentation)

Macro-model Velocity from low frequency content
Impedance/Reflectivity from high frequency content

Various signatures of parameters on data

Parameterization

Reminder

(V_p, I_p) (V_p, ρ)

$$\left.\frac{\partial q_i}{\partial V_p}\right]_{I_p} \# \left.\frac{\partial q_i}{\partial V_p}\right]_{\rho}$$

The partial derivative with respect to one parameter depends on the other selected parameters.

$$q_i = \begin{pmatrix} v_x \\ v_y \\ v_z \\ p \\ r \end{pmatrix}$$