SAPIENZA
UNIVERSITÀ DI ROMA

Environmental geophysics

Giorgio De Donno

1. Seismic methods

Seismic tomography
Data inversion

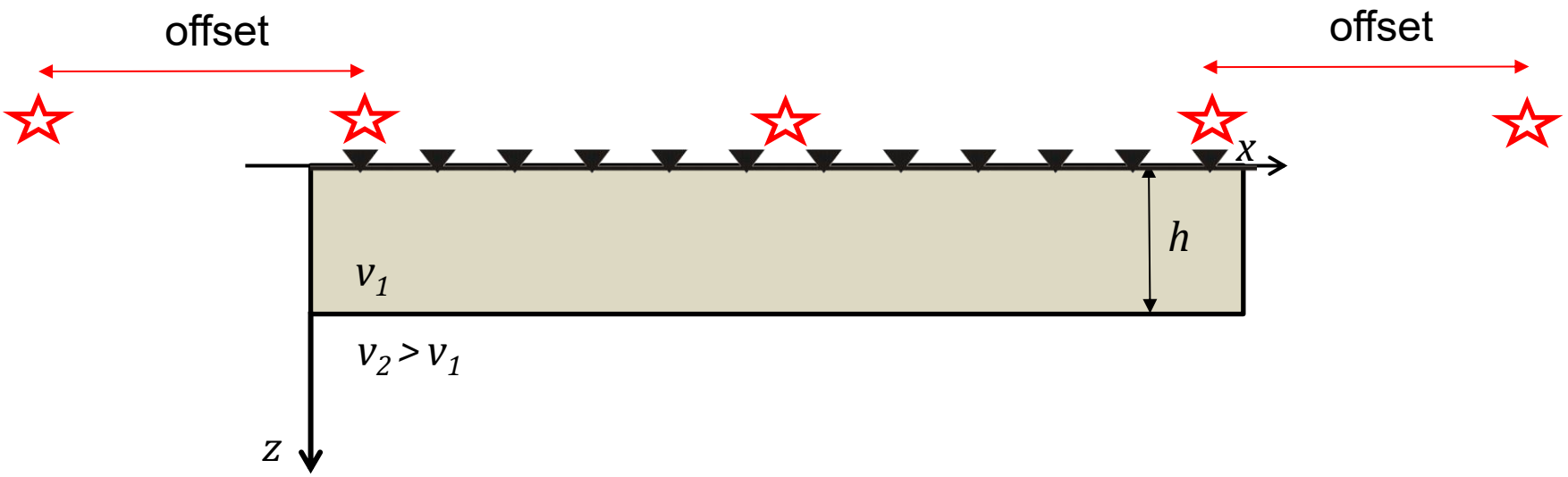
*"**Sapienza" University of Rome** - DICEA Area Geofisica*

phone: +39 06 44 585 078
email: giorgio.dedonno@uniroma1.it

Seismic refraction – Multichannel acquisition

In classical seismic refraction we have 5 shots and 24 or more receivers

hp. 1-D model (plane and parallel layers) -> $v=f(z)$



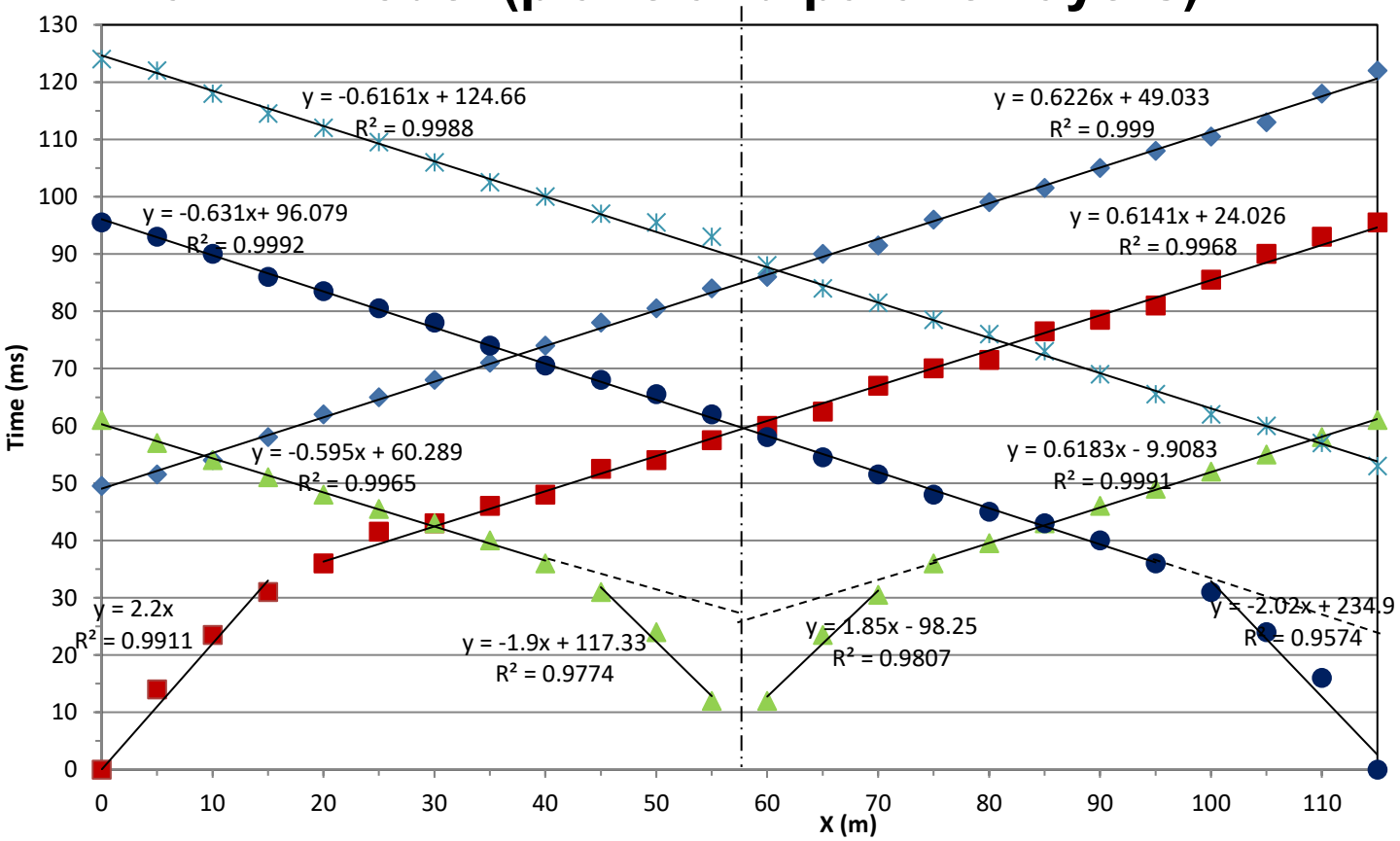
Q. Why do we have 5 shots?

A. To check the hypothesis of 1-D model through the reciprocity of *x-t* curves. Travel-time recorded at G24 due to shot at G1 should be the same of travel-time recorded at G1 due to the shot at G24 and interpolated lines should have the same slope.

Seismic refraction – Multichannel acquisition

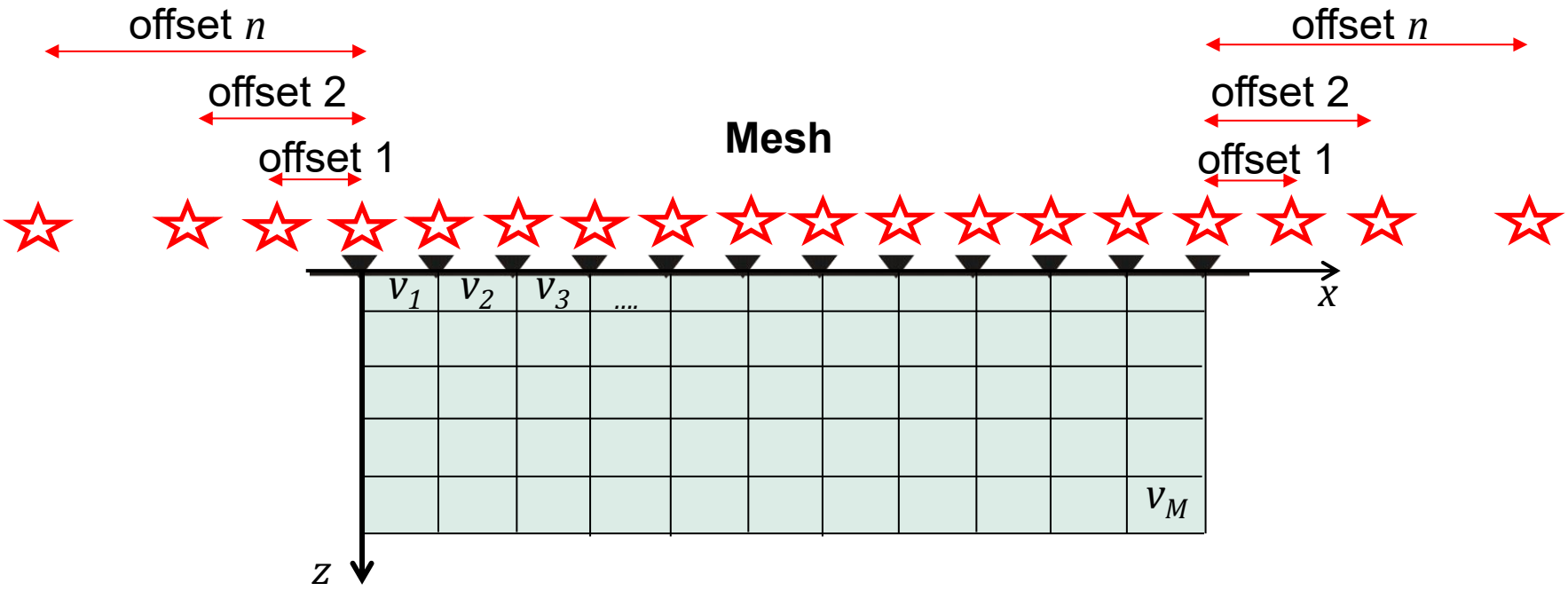
Time-distance plot: compute v_1, v_2 and h and check for the error on the different shots.

However, this is not a high-resolution image of the subsoil but only a 1-D model (plane and parallel layers)



Seismic tomography

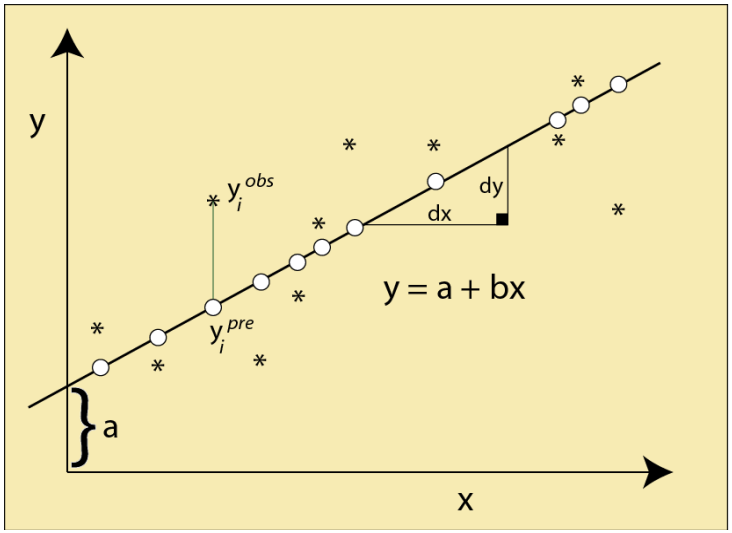
For a tomographic 2-D (or 3-D) reconstruction of the subsoil we need more information: it means more ray-paths and therefore shots should be executed at each receiver position and with many offsets (if feasible). The final image is made by pixels->v=f(x, y, z)



Q. Why I need more information?

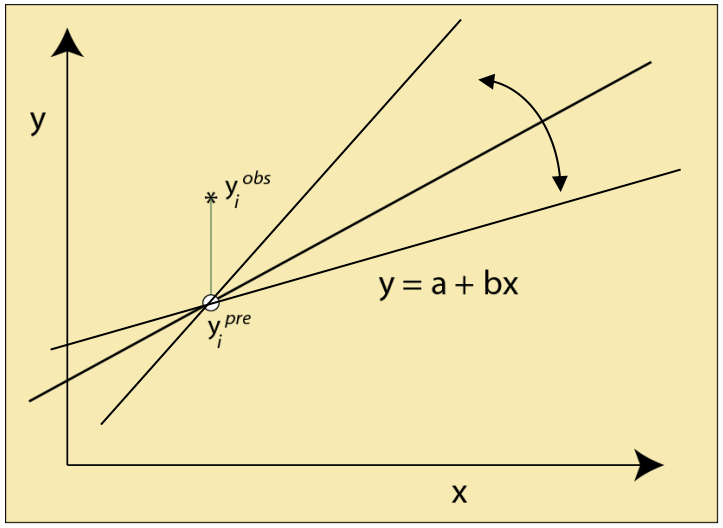
A. Because I have much more unknowns (pixel velocities) to be solved

Can *a* and *b* be resolved having only one point (e.g. in **exercises n.2-3)**?



Over-determined

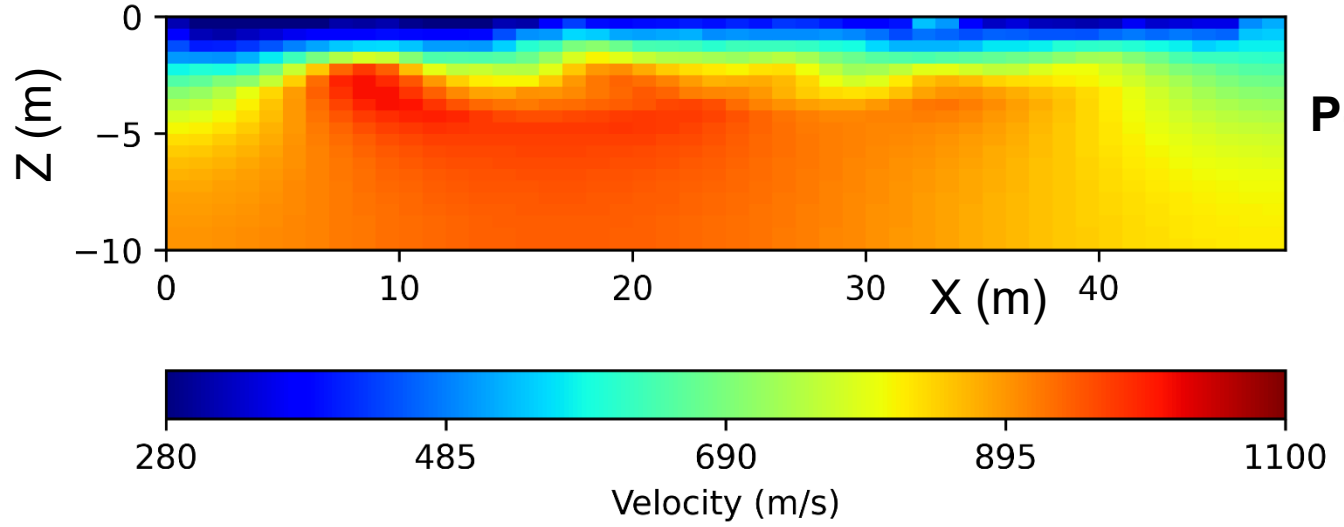
Two unknown parameters and many data points: we can fit these points with a line and estimate the goodness of fit



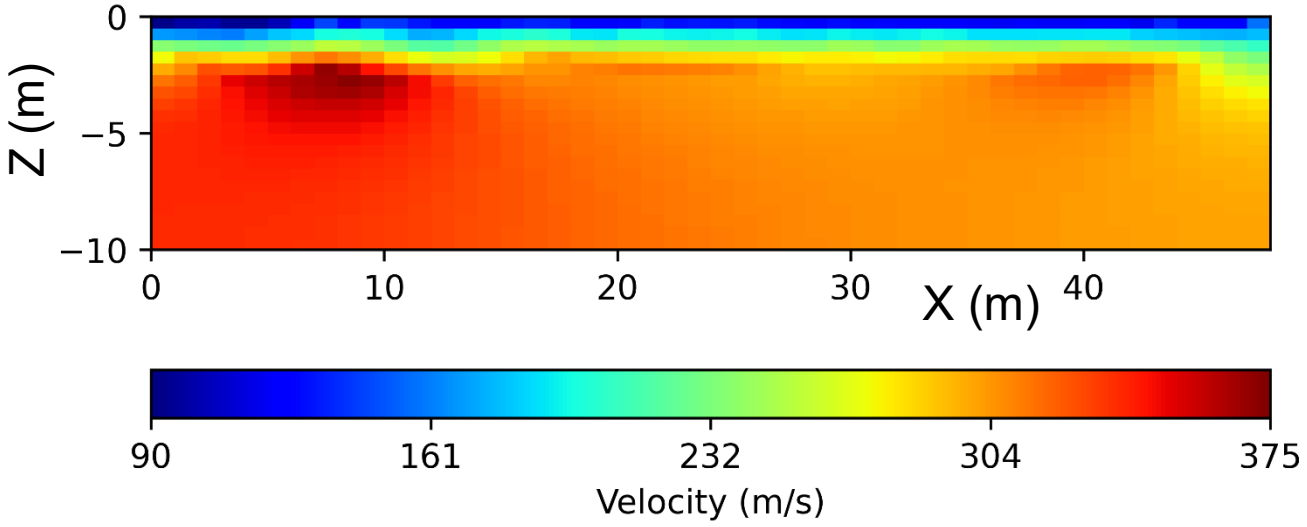
Under-determined

Two unknown parameters and only one data points (infinite lines can fit this point)

Seismic tomography – Example final 2-D output



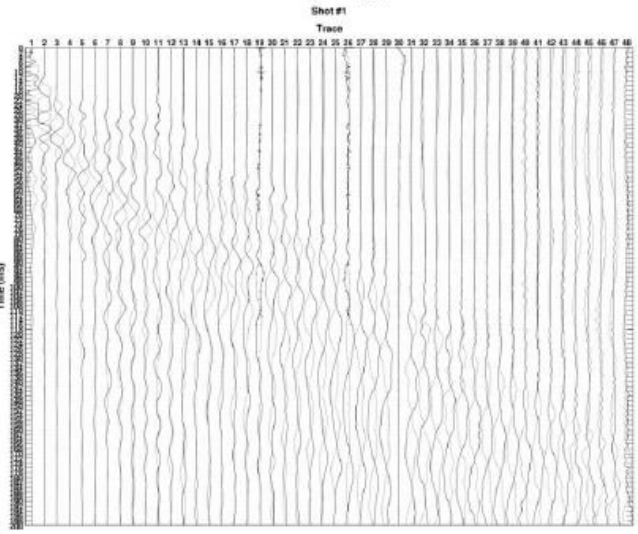
P-wave model



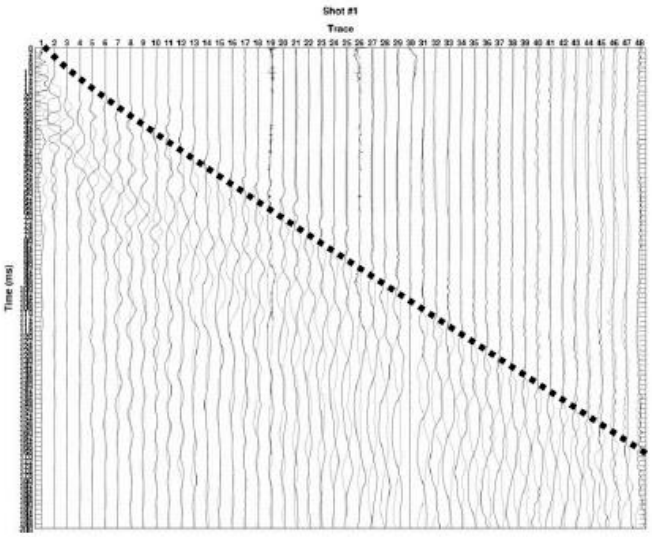
S-wave model

Seismic tomography





Recorded shot gather #1

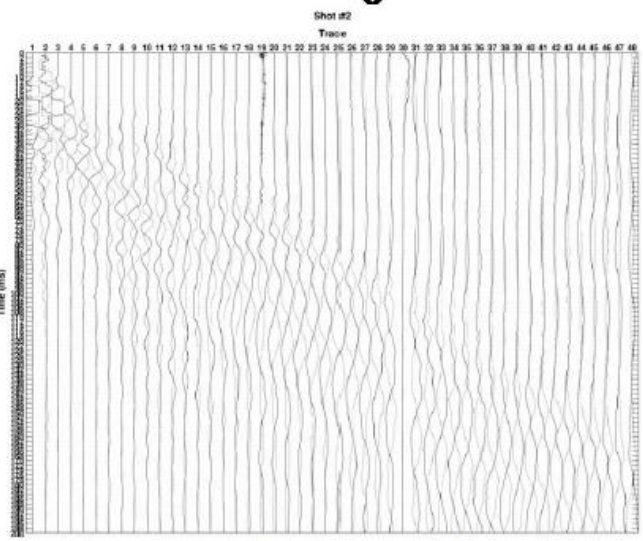


Picking of travel times

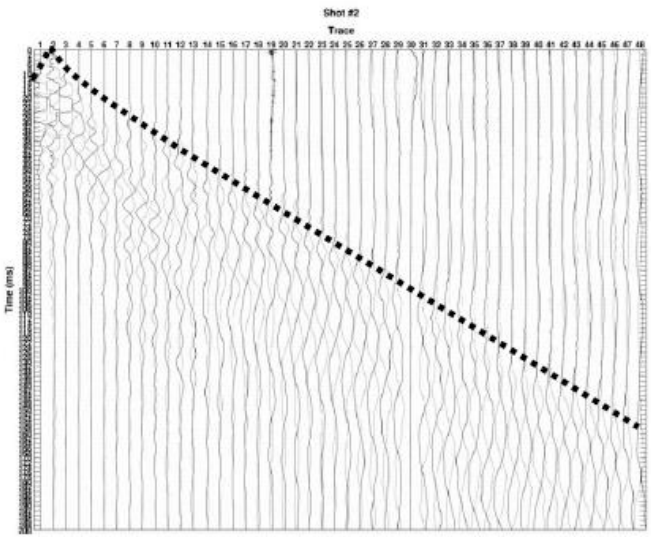
$$\mathbf{t}_{\mathbf{S1}}^{obs} = \begin{Bmatrix} t_{G1}^{S1} \\ t_{G2}^{S1} \\ t_{G3}^{S1} \\ \dots \\ t_{G48}^{S1} \end{Bmatrix}$$

Seismic tomography





Recorded shot gather #2

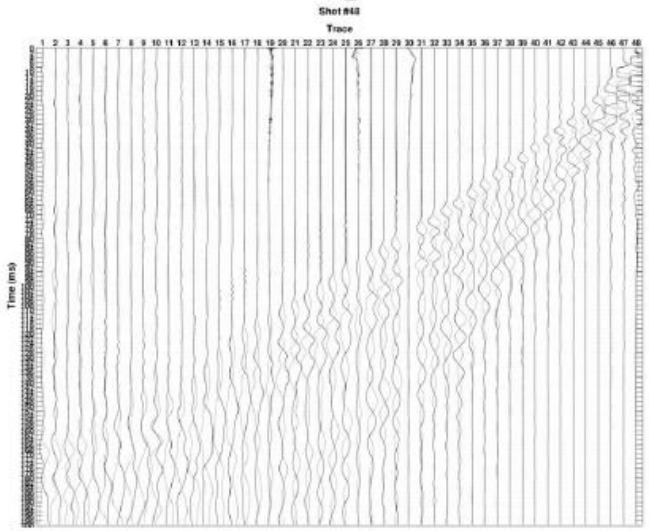


Picking of travel times

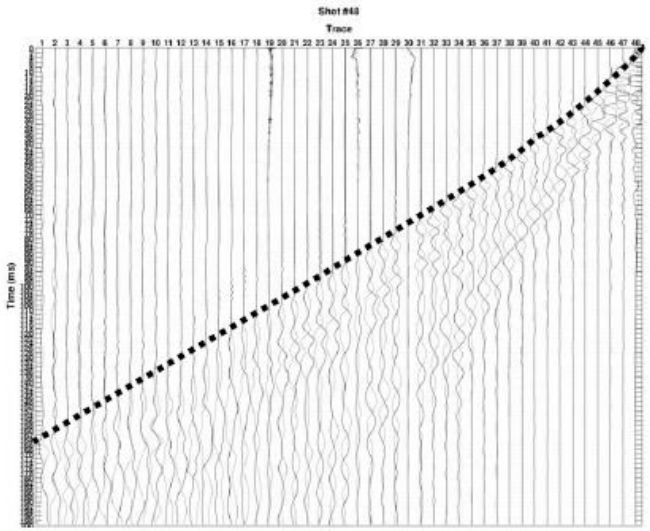
$$\mathbf{t}_{\mathbf{S2}}^{\,obs} = \begin{Bmatrix} t_{G1}^{S2} \\ t_{G2}^{S2} \\ t_{G3}^{S2} \\ \dots \\ t_{G48}^{S2} \end{Bmatrix}$$

Seismic tomography





Recorded shot gather #48



Picking of travel times

$$\mathbf{t}_{S3}^{\,obs} = \begin{Bmatrix} t_{G1}^{S48} \\ t_{G2}^{S48} \\ t_{G3}^{S48} \\ \dots \\ t_{G48}^{S48} \end{Bmatrix}$$

Seismic tomography



$$\mathbf{t}_{TOT}^{obs} = \begin{Bmatrix} \mathbf{t}_{S1}^{obs} \\ \mathbf{t}_{S2}^{obs} \\ \mathbf{t}_{S3}^{obs} \\ \dots \\ \mathbf{t}_{S48}^{obs} \end{Bmatrix}$$

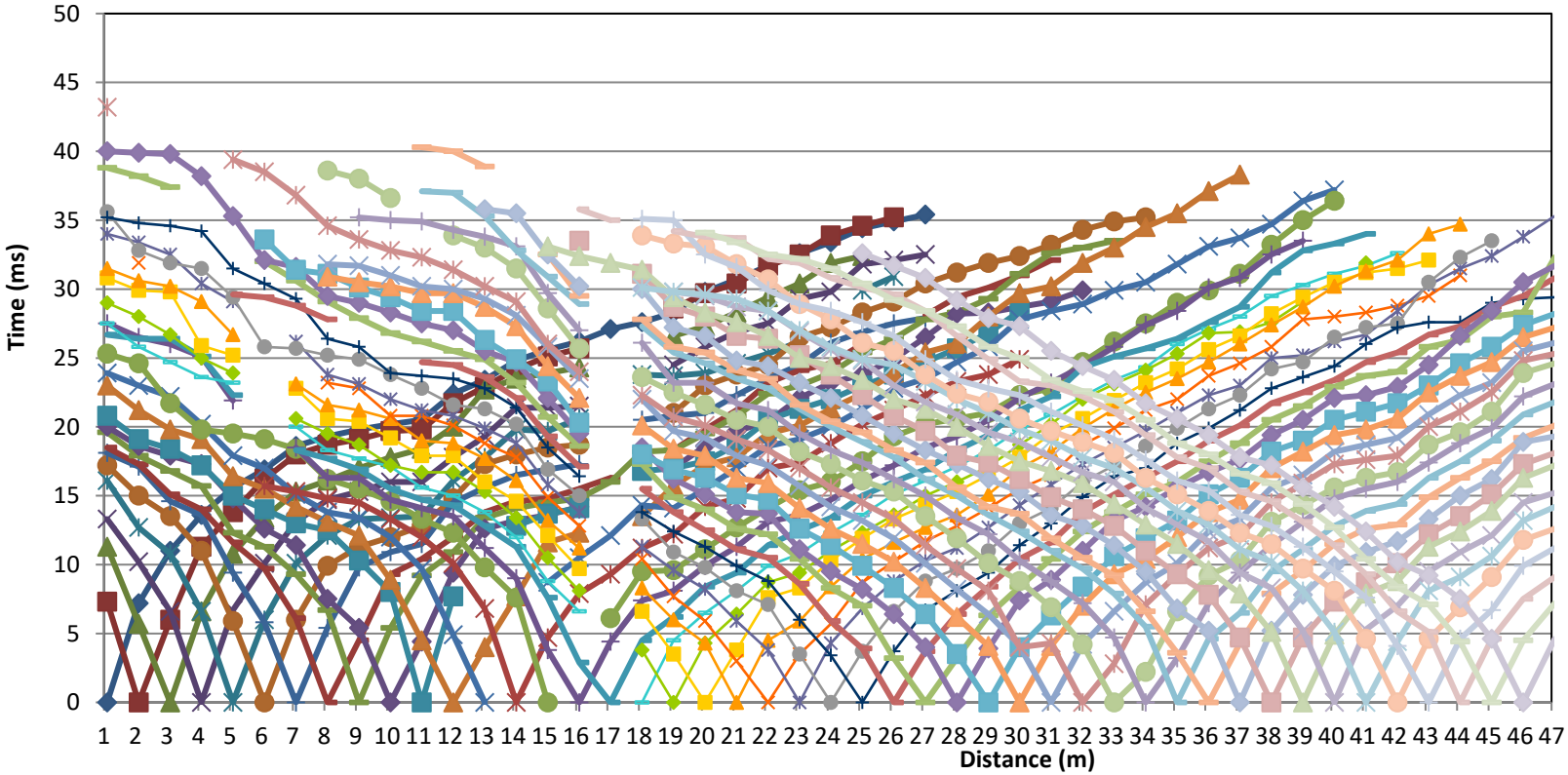
Each term is a vector that contains 48 travel-times

TOTAL NUMBER OF TRAVEL-TIMES AND CONSEQUENTLY OF RAY-PATHS

SIZE($\mathbf{t}_{TOT}$)=48x48=2304

Seismic tomography

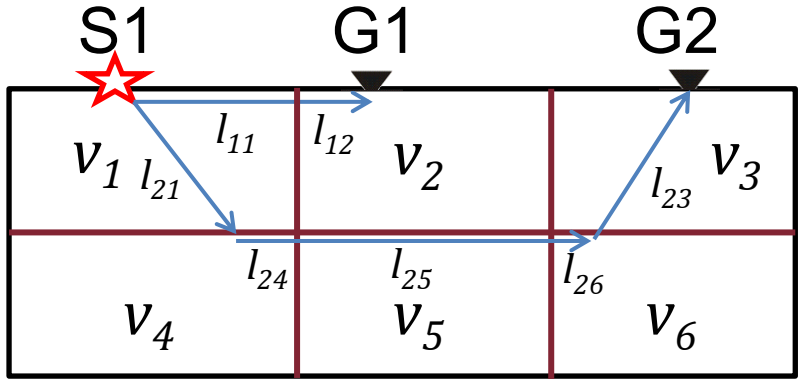
For seismic tomographic we have much more information, and we cannot apply the simple interpretation seen before!
The correct approach is to perform **data inversion!**



Seismic tomography – Data inversion

Ex. 1 – 1 shot, 2 geophones, 6 pixels

First guess of velocity model from prior information $\boldsymbol{v}^{(1)}=(v_1=v_2=v_3<v_4=v_5=v_6)$



We introduce the **slowness** s: $s = \frac{1}{v}$

Therefore, we can **predict** the travel-times:

$$\begin{cases} t_1^{pre} = l_{11}s_1 + l_{12}s_2 \\ t_2^{pre} = l_{21}s_1 + l_{24}s_4 + l_{25}s_5 + l_{26}s_6 + l_{23}s_3 \end{cases} \Rightarrow \begin{Bmatrix} t_1^{pre} \\ t_2^{pre} \end{Bmatrix} = \begin{bmatrix} l_{11} & l_{12} & 0 & 0 & 0 & 0 \\ l_{21} & 0 & l_{23} & l_{24} & l_{25} & l_{26} \end{bmatrix} \begin{Bmatrix} s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \end{Bmatrix}$$

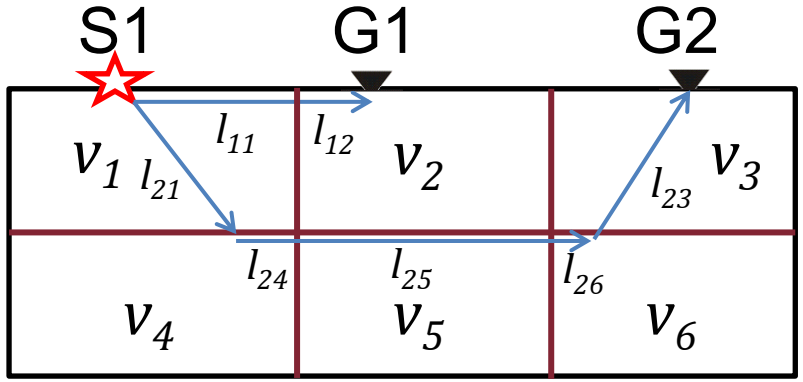
Index of the observation

Index of the pixel velocity

Seismic tomography – Data inversion

Ex. 1 – 1 shot, 2 geophones, 6 pixels

First guess of velocity model from prior information $\boldsymbol{v}^{(1)}=(v_1=v_2=v_3>v_4=v_5=v_6)$



We introduce the **slowness** s: $s = \frac{1}{v}$

Generalizing:

$$t_i^{pre} = \sum_{j=1}^{M} l_{ij} s_j \quad i = 1,2,\dots,N$$

$$\{\mathbf{t^{pre}}\} = [\mathbf{L}]\{\mathbf{s}\}$$

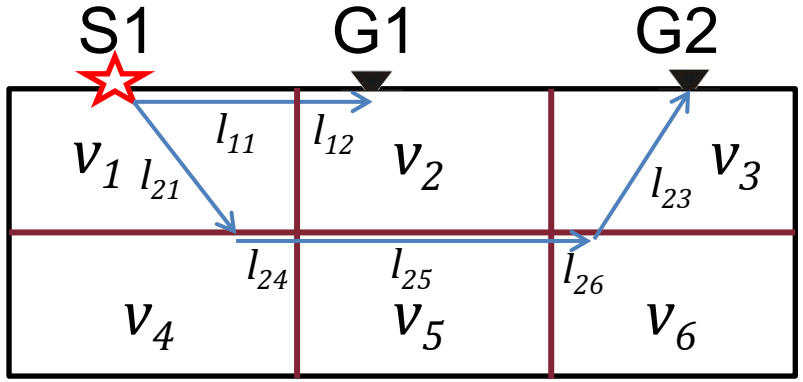
Q. Can I directly find the velocity (slowness) vector by substituting the observed (picked) times to the predicted ones?

A. No! Why?

Seismic tomography – Data inversion

Ex. 1 – 1 shot, 2 geophones, 6 pixels

First guess of velocity model from prior information $\boldsymbol{v}^{(1)}=(v_1=v_2=v_3>v_4=v_5=v_6)$



We introduce the **slowness** s: $s = \frac{1}{v}$

Substituting the observations:

$$\{\mathbf{t}^{obs}\} = [\mathbf{L}]\{\mathbf{s}\} \quad \not\Rightarrow \quad \{\mathbf{s}\} = [\mathbf{L}]^{-1}\{\mathbf{t}^{obs}\}$$

A1. Because **[L] is generally not a square matrix** and cannot be inverted.

A2. I can manage to make **[L]** a square matrix i.e. forcing the number of pixels to be equal to the number of observations (difficult but feasible). But the observed **travel-times are affected by errors** (both **systematic and due to measurements** and picking). Hence **the system is inconsistent**: no model can solve it exactly. The best approach is therefore to perform an **iterative inversion process.**

Data inversion

INVERSION IS AN ITERATIVE PROCEDURE

First guess on the velocity model (prior information) $v^{(1)}$

Seismic data acquisition

Ray-tracing [L] - Fermat's principle

Observed Travel-times t_{OBS} ⟺ $t_{PRE}^{(k)}$ **Predicted Travel-times**

Is the error small enough?

NO → **Try a new model $v^{(k+1)}$**

YES → **$v^{(k)}$ is the final velocity model**

Data inversion – Misfit functions

t^{OBS} - $t^{(k)}_{PRE}$ = **DATA MISFIT**

We do not minimize the simple sum of data misfit; otherwise, I can have negative values that cancel out the positive ones. It is misleading…

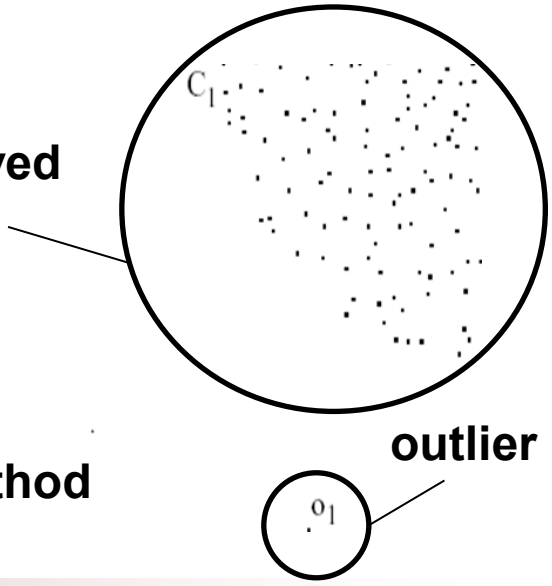
Ex.1 Dataset $\begin{cases} t^{OBS}(1)=5 \text{ ms} \\ t^{OBS}(2)=10 \text{ ms} \\ t^{OBS}(3)=15 \text{ ms} \end{cases}$ hp. At the first iteration $t^{PRE}(1,2,3)=10 \text{ ms}$

Sum of data misfit = 0

$$E(\mathbf{m}) = \sum_{i=1}^{N} \left(t_i{}^{obs} - t_i{}^{pre}(\mathbf{m})\right)^2 = \min$$

Least-squares method

In presence of inconsistent observations (**outliers**), the minimization of the absolute misfit is more suitable. This method is called **ROBUST INVERSION**



$$E(\mathbf{m}) = \sum_{i=1}^{N} \left|t_i{}^{obs} - t_i{}^{pre}(\mathbf{m})\right| = \min$$

Robust method

Data inversion – LS vs robust inversion

Ex.2 Dataset

$t_i^{OBS}(1)$=8 ms
$t_i^{OBS}(2)$=10 ms
$t_i^{OBS}(3)$=10 ms
$t_i^{OBS}(4)$=12 ms

Hp. At the first iteration
$t_i^{PRE}(1,2,\ldots,4)$=10 ms

$$E(\mathbf{m}) = \sum [4,0,0,4] = 8 \quad \textbf{LS}$$

$$E(\mathbf{m}) = \sum [2,0,0,2] = 4 \quad \textbf{Robust}$$

Same weights in the error function to high and low values

Ex.3 Dataset

$t_i^{OBS}(1)$=1 ms
$t_i^{OBS}(2)$=10 ms
$t_i^{OBS}(3)$=10 ms
$t_i^{OBS}(4)$=30 ms

Hp. At the first iteration
$t_i^{PRE}(1,2,\ldots,4)$=10 ms

$$E(\mathbf{m}) = \sum [81,0,0,400] = 481 \quad \textbf{LS}$$

$$E(\mathbf{m}) = \sum [9,0,0,20] = 29 \quad \textbf{Robust}$$

Low values weight less than high values and LS method is highly affected by **outliers**

Data inversion – Model search

Q. How the new trial models should be selected?
A. I can search for a new model:

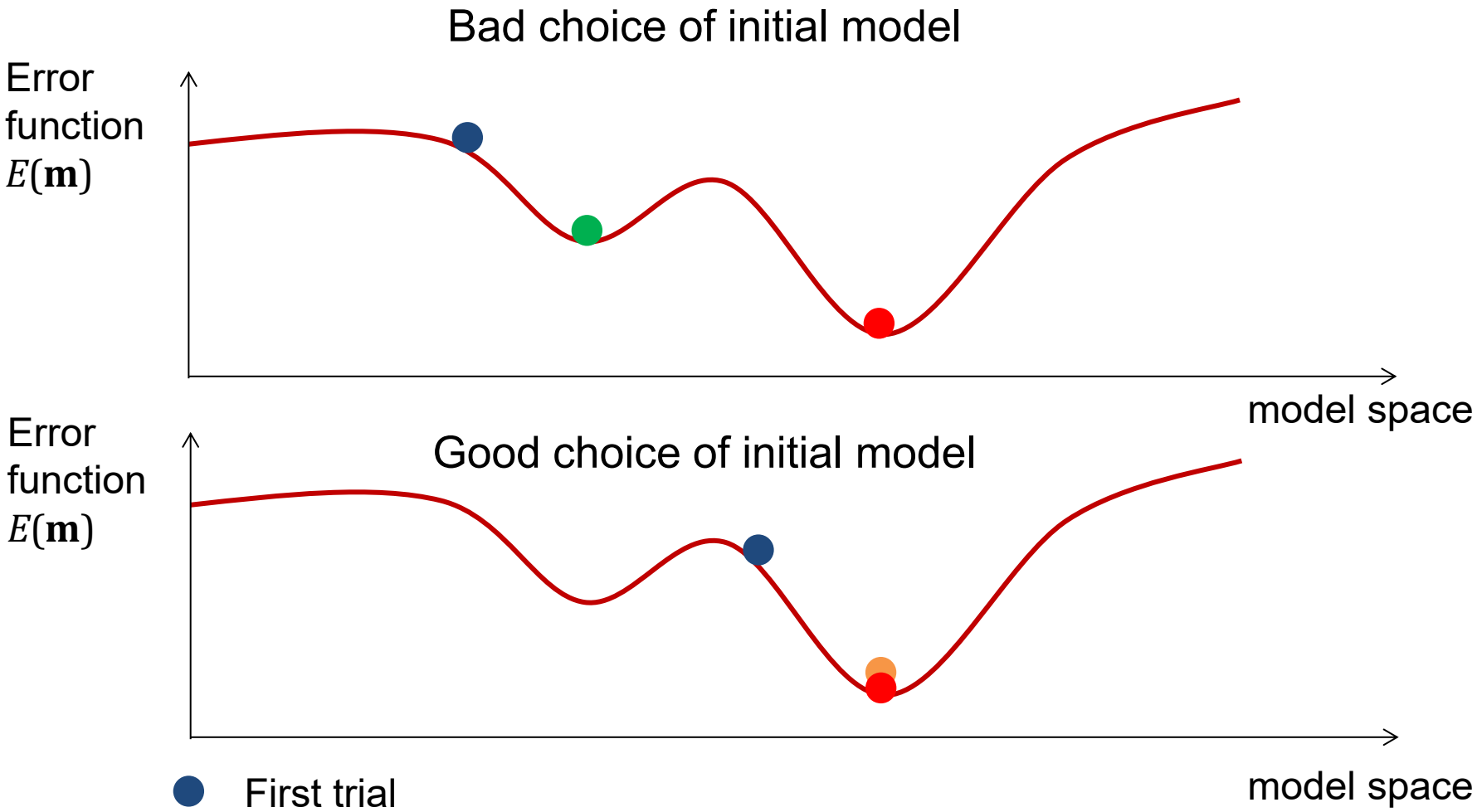
GLOBALLY: investigating the whole model space	**LOCALLY**: investigating only around the initial guess



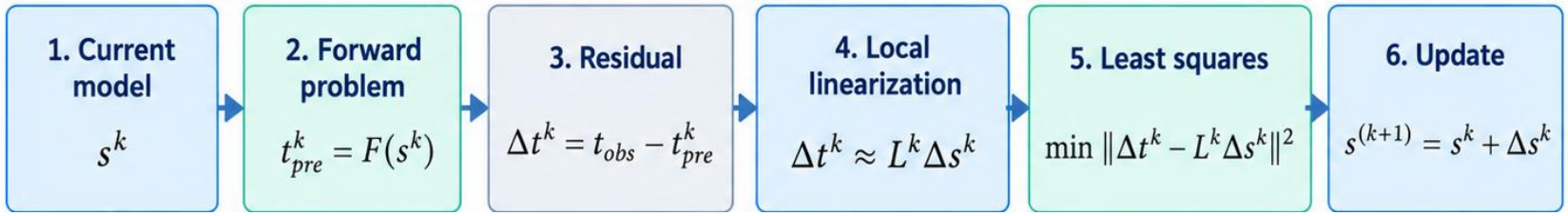
Global methods are slow because they require a number of trials in the model space (often > 10.000): if the forward solver is slow this process can take more than 1 day...

However, using local methods the solution can remain entrapped in a relative minimum: therefore **may be not the best one (it depends on the initial guess)**

Data inversion – Model search



Data inversion – Local (Gauss-Newton) LS iteration



Gauss-Newton LS method $\min\left\|\Delta\mathbf{t}^k - \mathbf{L}^k\Delta\mathbf{s}^k\right\|^2$

$$E\left(\mathbf{s}^{k+1}\right) = \left(\Delta\mathbf{t}^k - \mathbf{L}^k\Delta\mathbf{s}^k\right)^{\mathrm{T}}\left(\Delta\mathbf{t}^k - \mathbf{L}^k\Delta\mathbf{s}^k\right) \rightarrow \min$$

Expanding and omitting the *k* subscript:

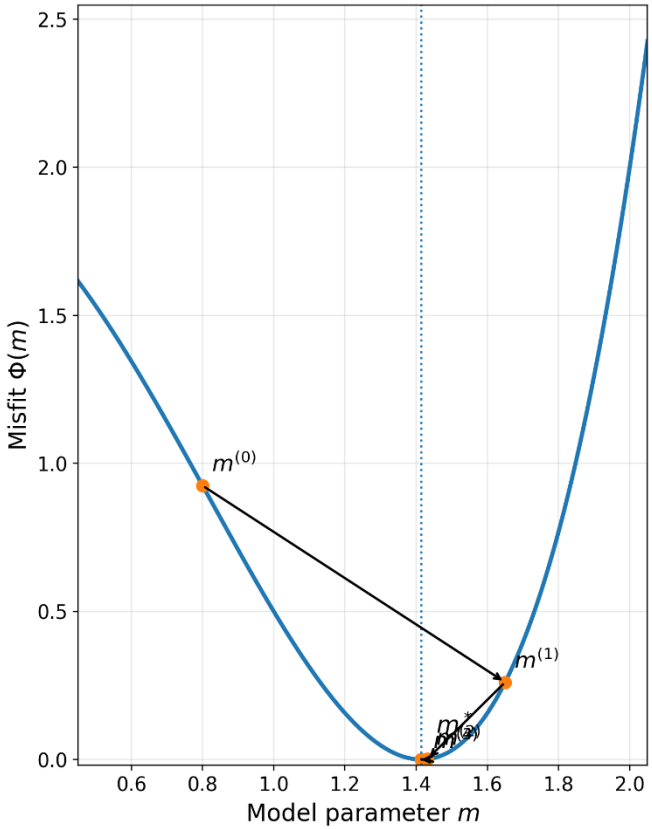
$$E\left(\mathbf{s}^{k+1}\right) = \Delta\mathbf{t}^{\mathrm{T}}\Delta\mathbf{t} - 2\Delta\mathbf{s}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\Delta\mathbf{t} + \Delta\mathbf{s}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\Delta\mathbf{s} \rightarrow \min$$

$$\nabla_{\mathbf{s}^{k+1}}\left(\Delta\mathbf{t}^{\mathrm{T}}\Delta\mathbf{t}\right) - 2\nabla_{\mathbf{s}^{k+1}}\left(\Delta\mathbf{t}^{\mathrm{T}}\mathbf{L}\Delta\mathbf{s}\right) + \nabla_{\mathbf{s}^{k+1}}\left(\Delta\mathbf{s}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{L}\Delta\mathbf{s}\right) = 0$$

$$-2\mathbf{L}^{\mathrm{T}}\Delta\mathbf{t} + 2\mathbf{L}^{\mathrm{T}}\mathbf{L}\Delta\mathbf{s} = 0 \quad \Rightarrow \quad \boxed{\mathbf{L}^{\mathrm{T}}\mathbf{L}\Delta\mathbf{s} = \mathbf{L}^{\mathrm{T}}\Delta\mathbf{t}}$$

Gauss-Newton step solution

$$\boxed{\Delta\mathbf{s}^k = \left(\mathbf{L}^{k\mathrm{T}}\mathbf{L}^k\right)^{-1}\mathbf{L}^{k\mathrm{T}}\Delta\mathbf{t}^k}$$



Data inversion - Goodness of fit

Q. When will the iterative process end?

A. If error at the current iteration is below a certain threshold.

Error is expressed at each iteration *k* as Root Mean Square Error (RMSE) for LS inversion:

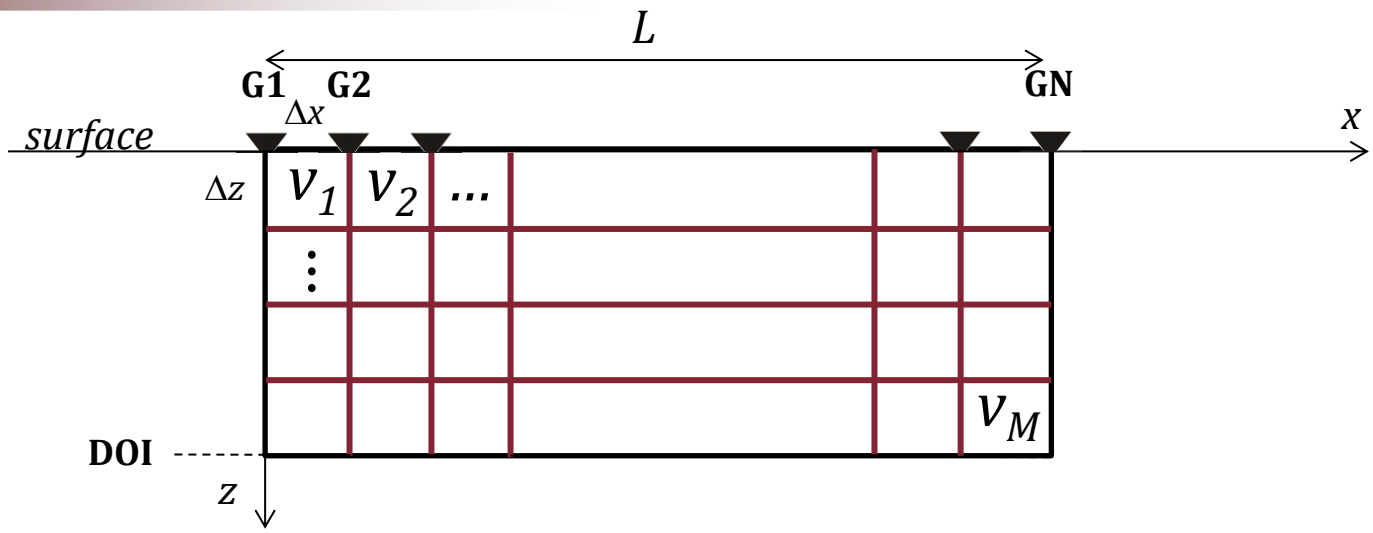
$$RMSE_k(\%) = 100\sqrt{\frac{\sum_{i=1}^{N}\left(\frac{t_i{}^{obs} - t_i{}^{pre}(\mathbf{m}_k)}{t_i{}^{obs}}\right)^2}{N}}$$

and Absolute Error (AE) for robust inversion:

$$AE_k(\%) = 100\,\frac{\sum_{i=1}^{N}\left|\frac{t_i{}^{obs} - t_i{}^{pre}(\mathbf{m}_k)}{t_i{}^{obs}}\right|}{N}$$

When the difference between the RMSE value of the iteration k and k-1 is below a certain acceptable value (i.e. 1%) the iterative procedure ends.

Data inversion – Mesh choice (Resolution and DOI)



Rule of thumb (both resolution and DOI depend on frequency and geophone spacing)

- **Resolution**:
  $hp. \Delta x = \Delta z = \Delta d$

$$\Delta d_{pixel} \geq \sqrt{\frac{v \Delta x}{2f}}$$

Δx: geophone spacing
v: seismic (P- or S-wave) expected velocity
f: seismic wave frequency

Using low-frequency sources and/or large geophone spacings we only get low-resolution models.

- **Depth Of Investigation**: $DOI \sim \frac{1}{4} \div \frac{1}{5} L \sim 20 - 25\% L$ L: line length

Using low-energy or high-frequency sources and/or short lines we only get information on the shallow layers.

Data inversion – Model roughness

Inversion can theoretically lead to a very **"rough" model** that is only the result of a mathematical operation (least squares) with sharp variations and jumps. It can be very different from the geological "**smoother" background.**
Therefore, we **do not use "pure" least-squares** but "**regularised" least-squares** in order to regularize the final model.

The model regularization is obtained by adding a second term to the least squares functional, in order to avoid huge jumps between adjacent nodes:

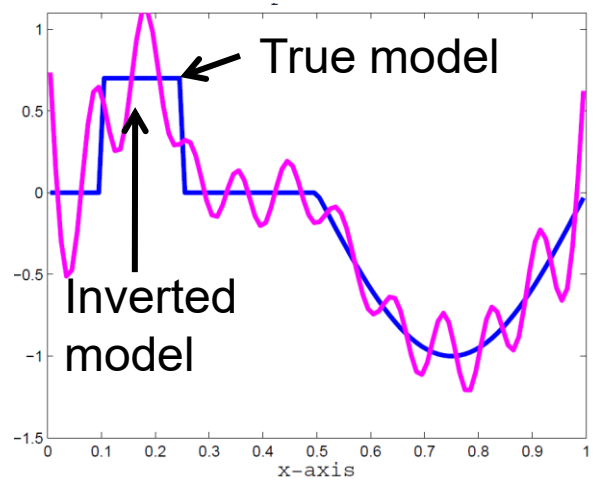
$$E(\mathbf{m}) = \sum_{i=1}^{N} \left(t_i{}^{obs} - t_i{}^{pre}(\mathbf{m})\right)^2 + \alpha^2 \sum_{j=1}^{M} m_j \sum_{l=1}^{M} R_{jl} m_l = \min$$

Damping factor:
a positive number

Regularisation matrix

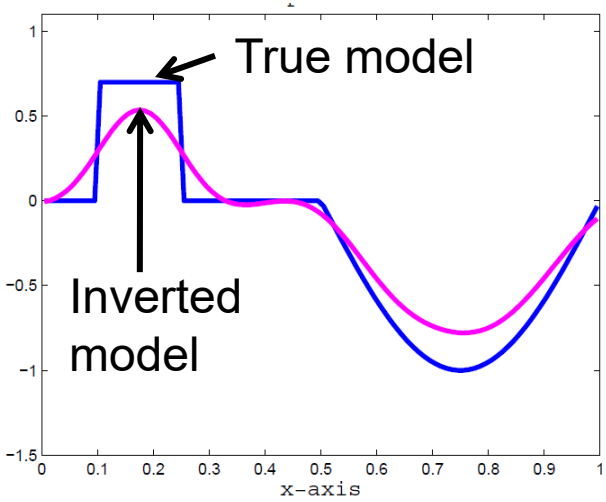
Data inversion – Damping factor

α = 0 or too small



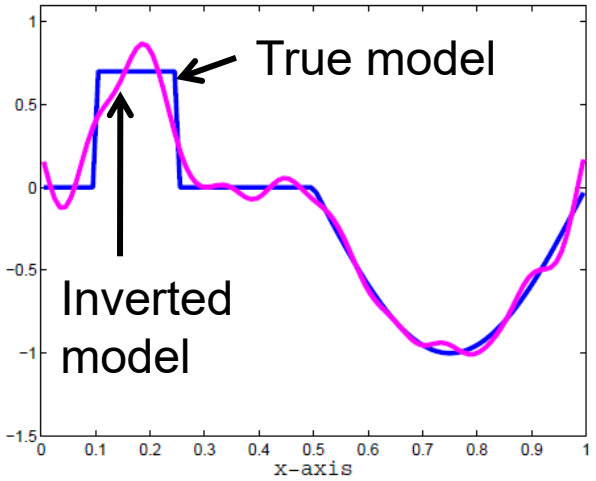
Solution is too rough

α too large



Solution is too smooth

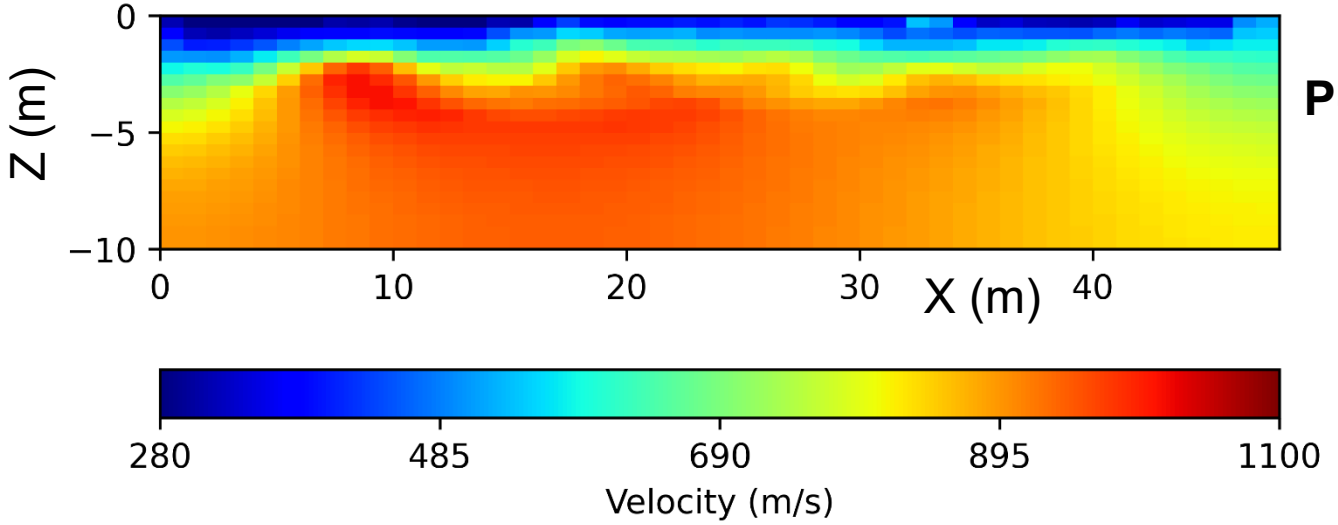
good α value



Compromise between roughness and smoothing

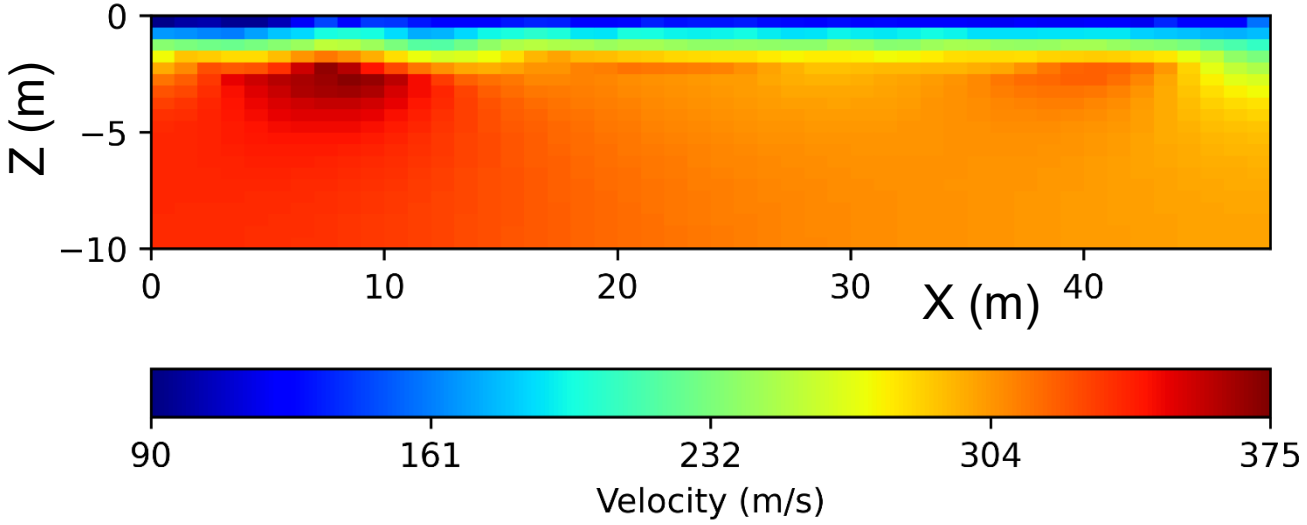
Generally, values of α between 0.1 and 10 are suitable for seismic data inversion

Seismic tomography – Inverted models



P-wave model

RMSE=5.2%

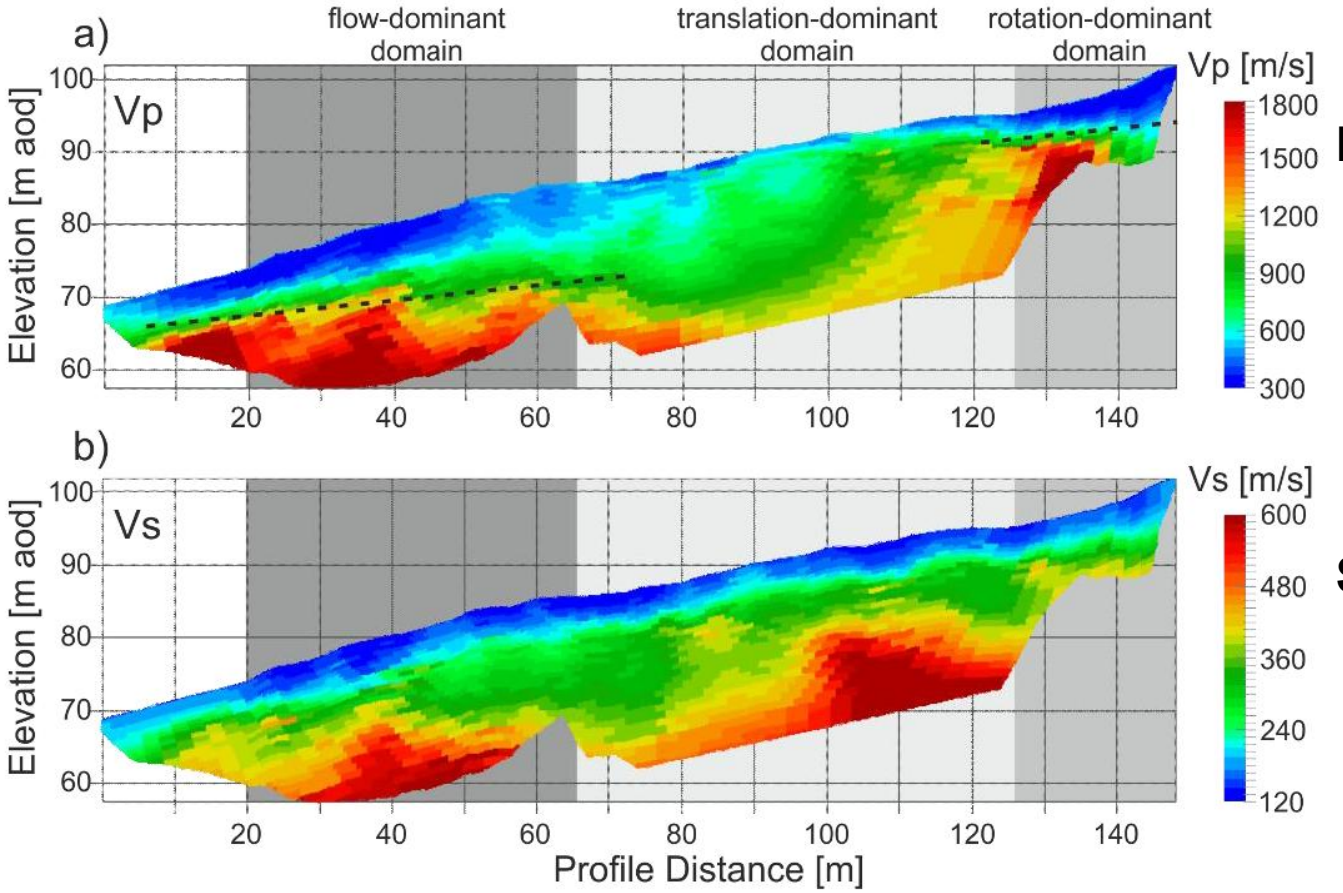


S-wave model

RMSE=3.8%

Seismic tomography – Inverted models

Uhlemann et al. (2016)



P-wave model
RMSE=1.7 ms

S-wave model
RMSE=3.4 ms