Tomography/Inversion
by Monte Carlo sampling

Wojciech Dębski

debski@tono.jnc.go.jp
debski@igf.edu.pl

June 16, 2003

Plan of the seminar

- Bayesian Inversion
  - Inverse Problems
  - Bayesian technique
  - (Not too) serious example
- Monte Carlo Technique
  - Global Optimization
  - Monte Carlo Sampling
    (Markov Chain Monte Carlo)
- Examples
  - Velocity tomography
  - Estimation seismic sources

Inverse Problems - Selected Topics

- Parameter estimation
- Error analysis
- Resolution analysis
- Experiment planning
- Model discrimination

Inverse problem - Indirect Measurements

$$\mathbf{d}^{obs} \implies \mathbf{m}$$

Solution

$$||\mathbf{d}^{obs} - \mathbf{d}^{th}(\mathbf{m})|| + \lambda||\mathbf{m}^{ml} - \mathbf{m}^{apr}|| = \min$$

Errors

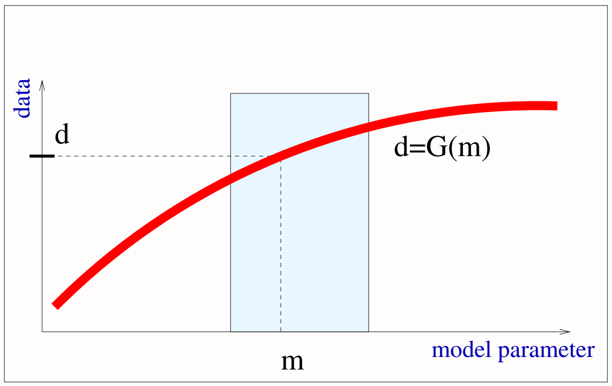
$$\mathbf{m}^{true} = \mathbf{m}^{ml} + \epsilon_{\mathbf{m}}$$

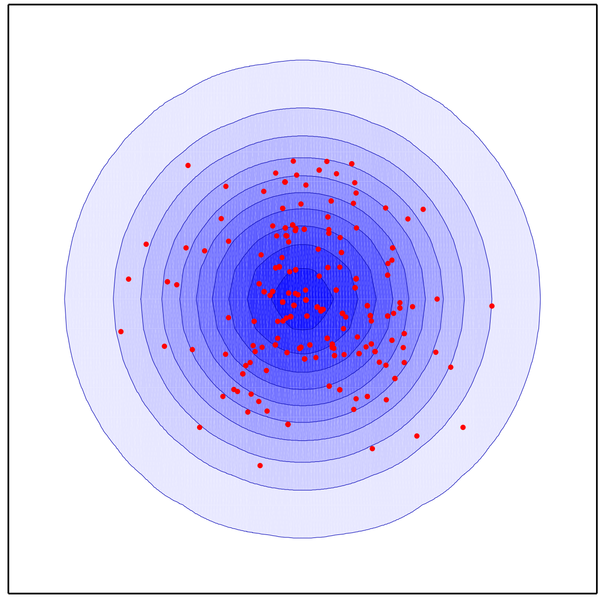
$$\boxed{\epsilon_{\mathbf{m}} = ???}$$

Inversion Algorithms

Method	Advantages	Limitations
Algebraic (LSQR)	- Simplicity	- Only linear problems
$\mathbf{m}^{ml} = (\mathbf{G}^T\mathbf{G} + \gamma\mathbf{I})^{-1}\mathbf{G}^T \cdot \mathbf{d}^{obs}$	- Large scale problems	- Lack of robustness
Optimization	- Simplicity	- Difficult error estimation
$\lVert\mathbf{G}(\mathbf{m}) - \mathbf{d}^{obs}\rVert + \lambda\lVert\mathbf{m}) - \mathbf{m}^a\rVert = \mathrm{min}$	- Fully nonlinear	
Bayesian	- Fully nonlinear	- More complex theory
$\sigma(\mathbf{m}) = f(\mathbf{m})L(\mathbf{m}, \mathbf{d}^{obs})$	- Full error handling	- Requires efficient sampler

Inversion algorithms

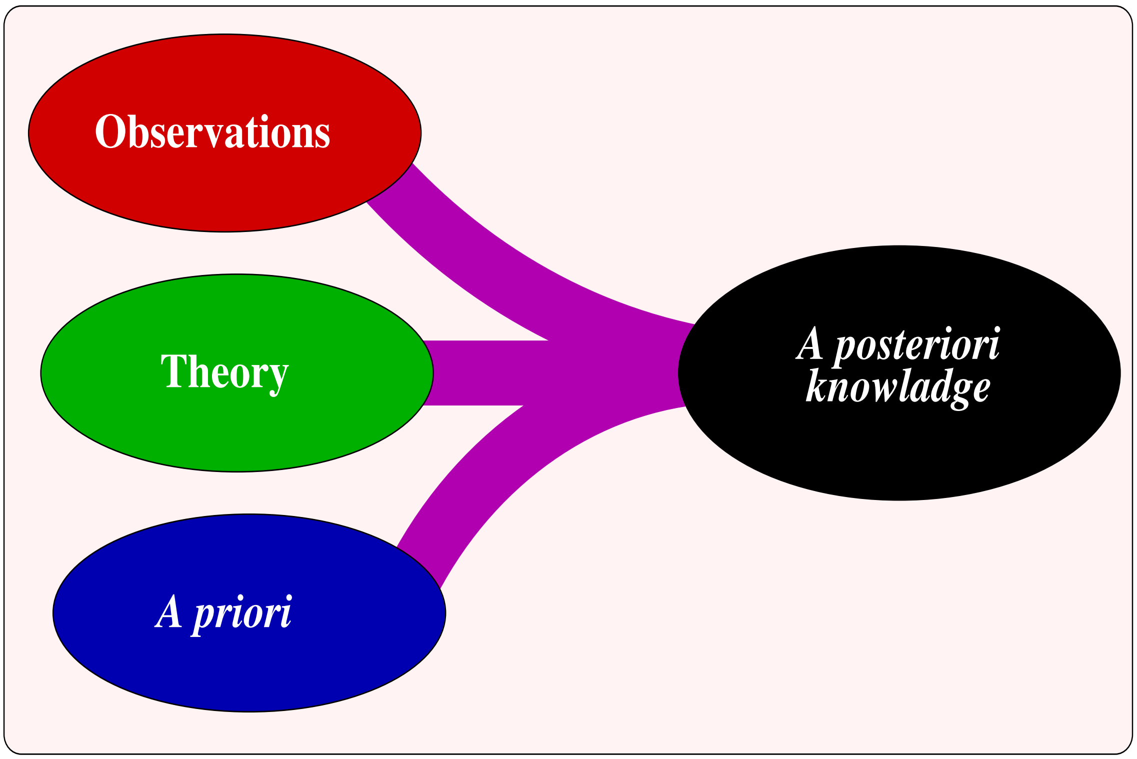




Back projection

Model space search

Bayesian Inversion - Basic Ideas



$$p_{post}(\mathbf{m}|\mathbf{d}) = \frac{p_{th}(\mathbf{d}|\mathbf{m})p_{apr}(\mathbf{m})}{p_{obs}(\mathbf{d})}$$

Bayesian Inversion

A posteriori pdf

A posteriori pdf

$$\sigma(\mathbf{m}) = f(\mathbf{m}) L(\mathbf{m}, \mathbf{d})$$

- f - A priori pdf (prob. dens. funct.)
- L - Likelihood function

Errors independent of $\mathbf{m}$ and $\mathbf{d}$

$$\mathbf{d}^{th} = \mathbf{G}(\mathbf{m})$$

$$\sigma(\mathbf{m}) = f(\mathbf{m}) \exp\left(-||\mathbf{d}^{obs} - \mathbf{G}(\mathbf{m})||\right)$$

$|| \cdot ||$ - a norm in D

Probabilistic approach

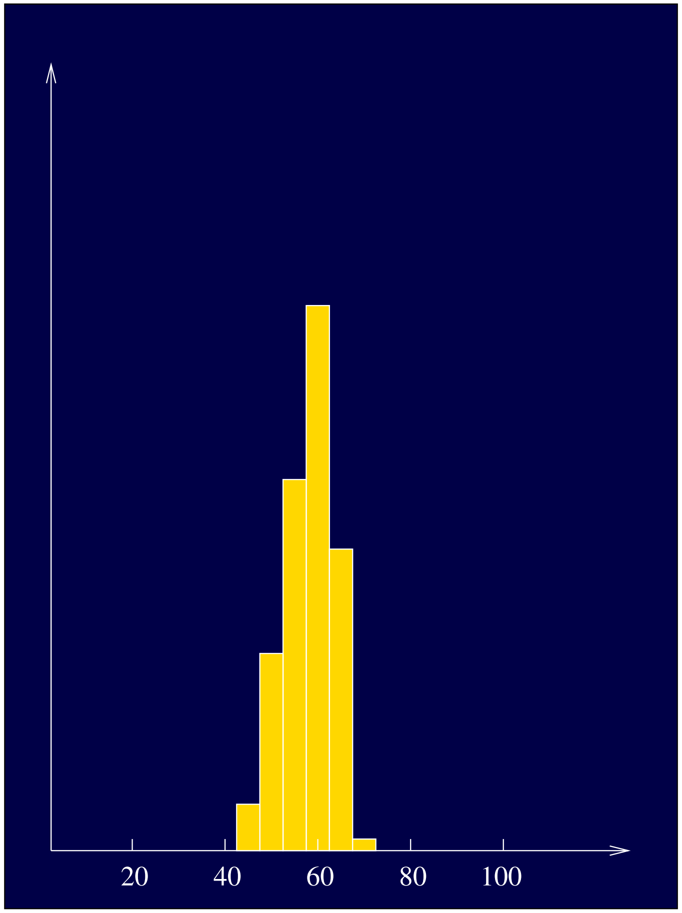
A posteriori pdf $\sigma(m, d)$:

- always exists
- is unique
- describes all information
- is the solution of an inverse problem

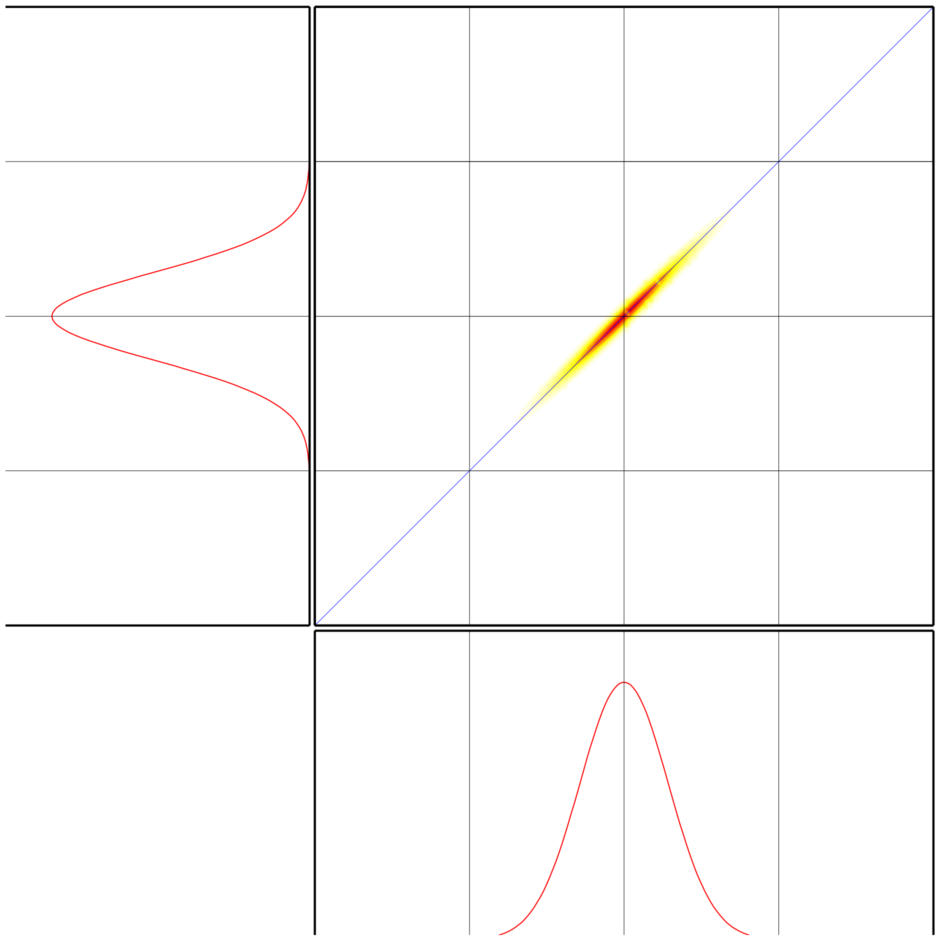
When and why we need to use this approach ???

- nonlinear forward problems
- error analysis

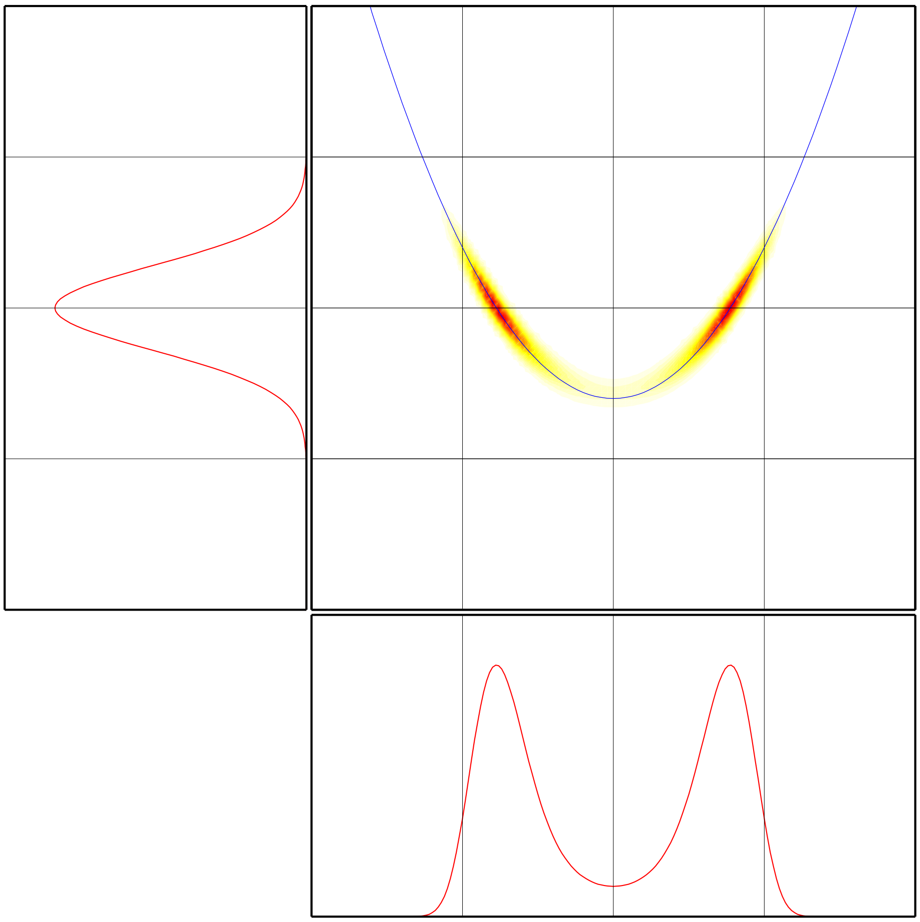
Probabilistic approach is very intuitive!



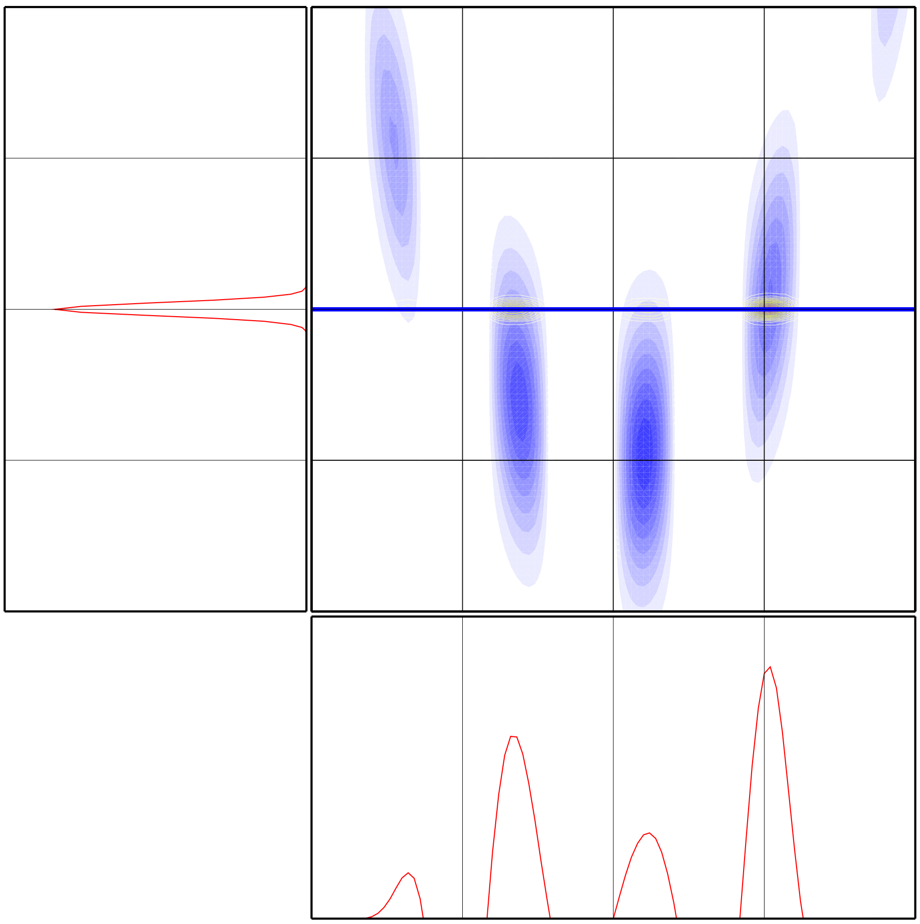
Exact theory: linear case



Exact theory: non-linear case



“Exact” data: uncertain theory



A posteriori pdf

$$\sigma_m(m) = f_M(m) \cdot L(m, d^{obs})$$

$$L(m, d^{obs}) = e^{-||d^{obs} - G(\mathbf{m})||}$$

Examination of $\sigma(\mathbf{m})$

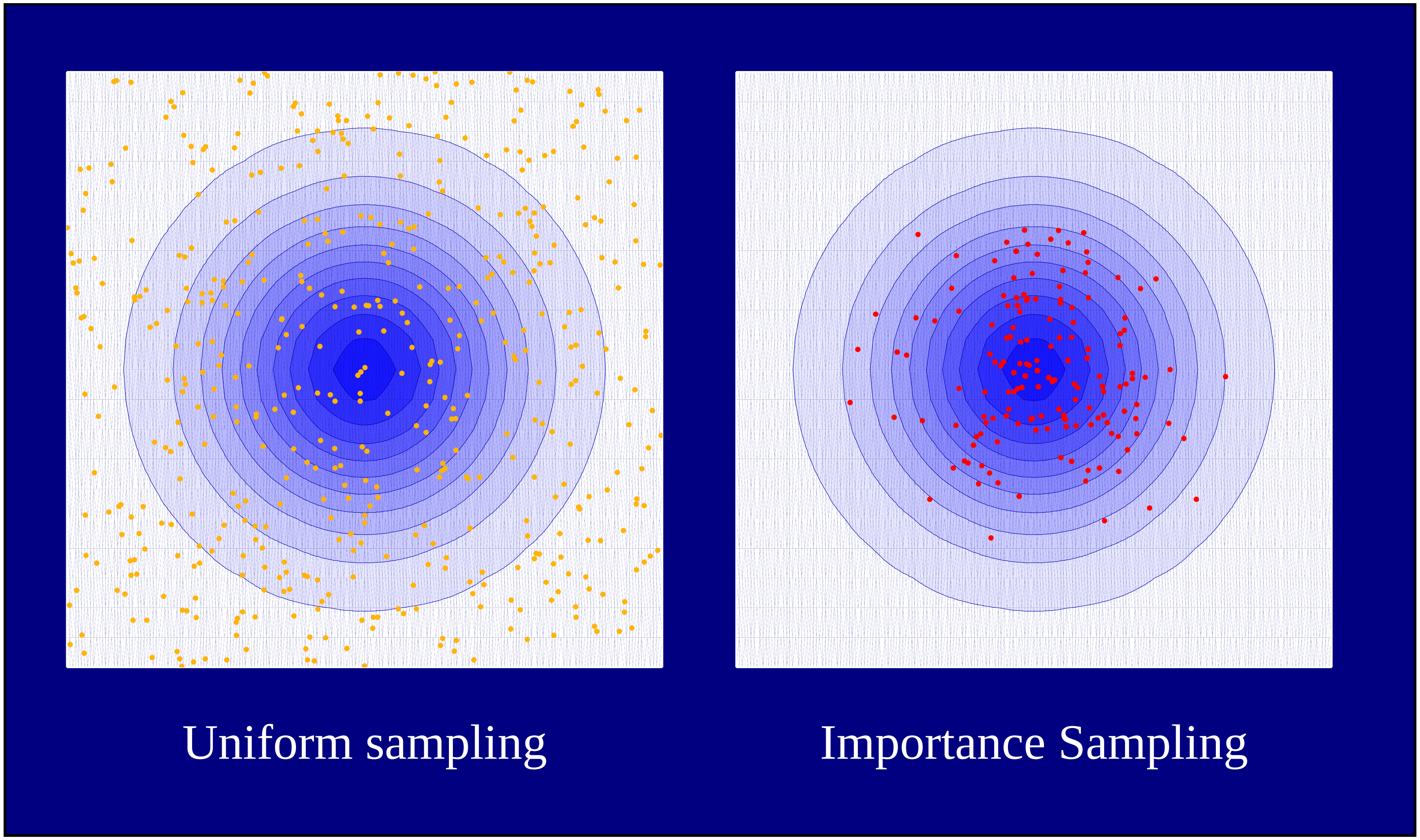
- Searching *Maximum Likelihood* solution
  - Gradient methods
  - Monte Carlo - Global Optimization
- Sampling over regular grid
- Random (Monte Carlo) Sampling

Monte Carlo

Monte Carlo

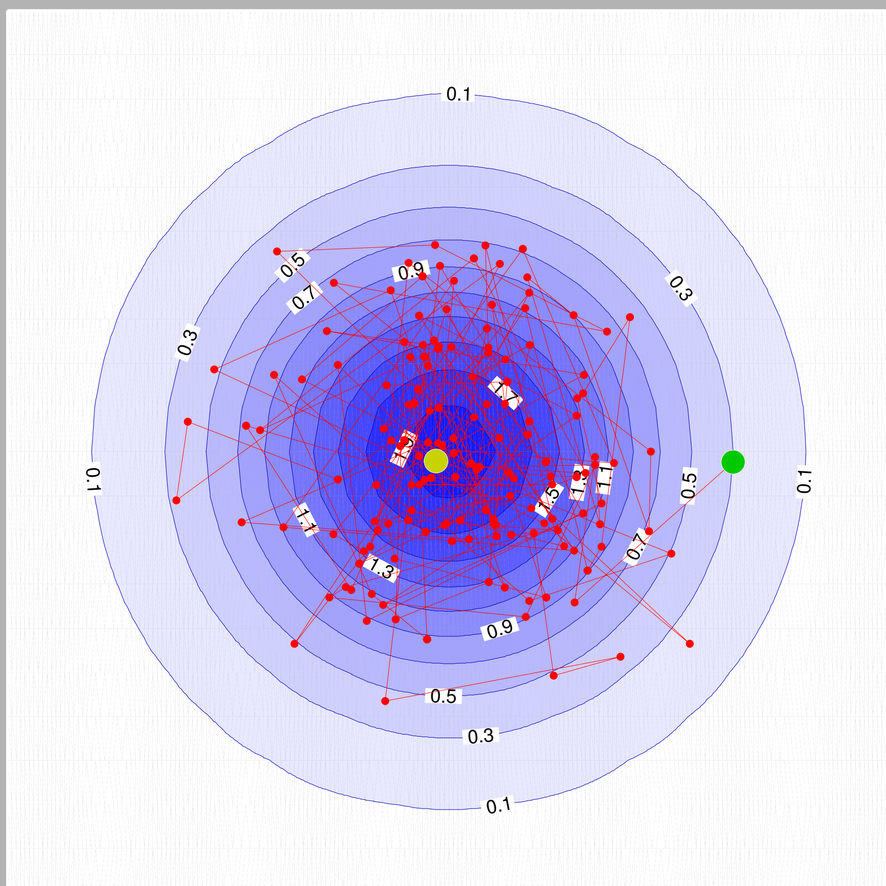
Monte Carlo

Monte Carlo sampling



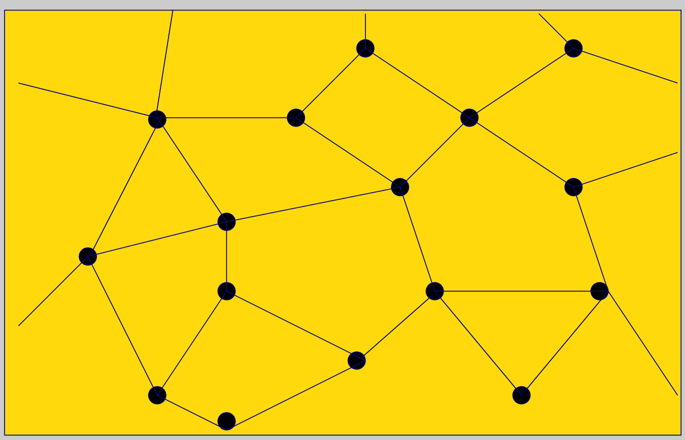
Monte Carlo as Random Walk

Random Walk in the model space



Random Walks (Markov Chain)

Random walk is a stochastic process of sampling (visiting) of the space. Being at given point $\mathbf{m}^i$ it can move only 'one' step either to $\mathbf{m}^j$ or stay in $\mathbf{m}^i$ with some probabilities.



Movements (transition probability) of "walker" depends only on the current position and the nearest nodes it can reach and not on the history of previous movements ("short memory" process).

Random Walks

Random walk is fully defined by transition probability matrix (operator) $P^{i \to j}$ being the probability of moving from the node i to the node j
When the number of steps go to infinity then probability that "walker" is in point $\mathbf{m}$ converge to some probability $p(\mathbf{m})$. Thus, "walker" is sampling this probability.

Metropolis Algorithm

$$P^{i \to j} = \min \left\{ 1, \ \frac{p(\mathbf{m}^j)}{p(\mathbf{m}^i)} \right\}$$

"Chain" of points:

$$m_1, m_2, \cdots m_i, m_j \cdots$$

Travel time tomograpphy

Problems

- Data: $\{\Delta t_1, \ldots \Delta t_N\} \qquad N = 10^2 - -10^6$
- Parameters $\{v_1, v_2, \ldots v_M\} \qquad M = 10^2 - 10^6$
- Forward problem strongly nonlinear: possible large theoretical errors
- Most often problem has non-unique solution: important *a priori* term

Travel time tomograpphy

Classical approach

Parameterization:

$$v(x,y,z) \longrightarrow \{v_i\}_{i=1\cdots M}$$

$$\Delta t_i = \sum_j \frac{\Delta l_j^i}{v_j} = \sum_j s_j \Delta l_j^i$$

$$t_{sr} = \int_{sr} \frac{dl}{v} \quad \longrightarrow \quad \mathbf{d} = \mathbf{G} \cdot \mathbf{m}$$

Regularization (non-unique solution)

$$\det(\mathbf{G}^T\mathbf{G}) = 0$$

Numerical optimization

$$S(\mathbf{m}) = ||\mathbf{d}^{obs} - \mathbf{G} \cdot \mathbf{m}|| + \lambda ||\mathbf{m}^{apr} - \mathbf{m}|| = min$$

Travel time tomograpphy

Probabilistic (MC) approach

Regularization:

$$f(\mathbf{m}) = e^{-||\mathbf{m}-\mathbf{m}^{apr}||/Cm}$$

$$\sigma_M(\mathbf{m}) = f(\mathbf{m}) \exp^{-||\mathbf{d}^{obs}-\mathbf{G}\cdot\mathbf{m}||/Cd}$$

Monte Carlo sampling

$$\{\mathbf{m}_1, \mathbf{m}_2 \cdots\cdots\cdots \mathbf{m}_N\}$$

$$N_{[\mathbf{m}-\Delta,\mathbf{m}+\Delta]} \sim \sigma(\mathbf{m})$$

Solution

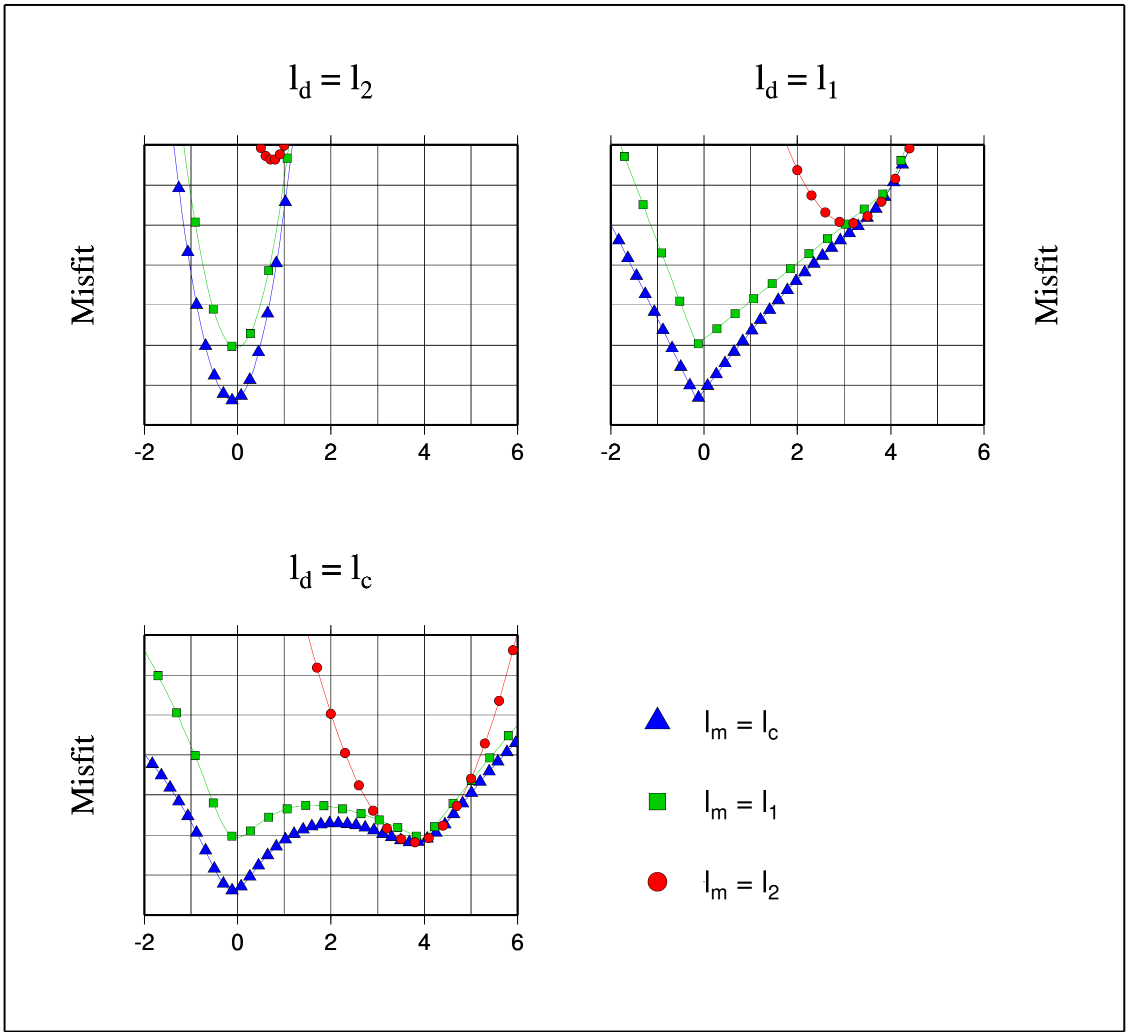
- $\mathbf{m}^{ml} = \mathbf{m}_i : \sigma_M(\mathbf{m}_i) = \max$
- $\mathbf{m}^{avr} = \sum_{samples} \mathbf{m}_s$
- $\epsilon_{\mathbf{m}} = \sum_{samples} (\mathbf{m}^{avr} - \mathbf{m}_s)^2$

Norms & Nonlinearity

l_1	$\|\|\mathbf{d}\|\| = \sum_i \left\| \frac{d^i}{C^i} \right\|$	Absolute value norm
l_2	$\|\|\mathbf{d}\|\| = \frac{1}{2} \sum_{ij} d^i (\mathbf{C})^{ij} d^j$	Gaussian norm
l_c	$\|\|\mathbf{d}\|\| = \sum_i \log \left(1 + \left(\frac{d^i}{C^i} \right)^2 \right)$	Cauchy norm

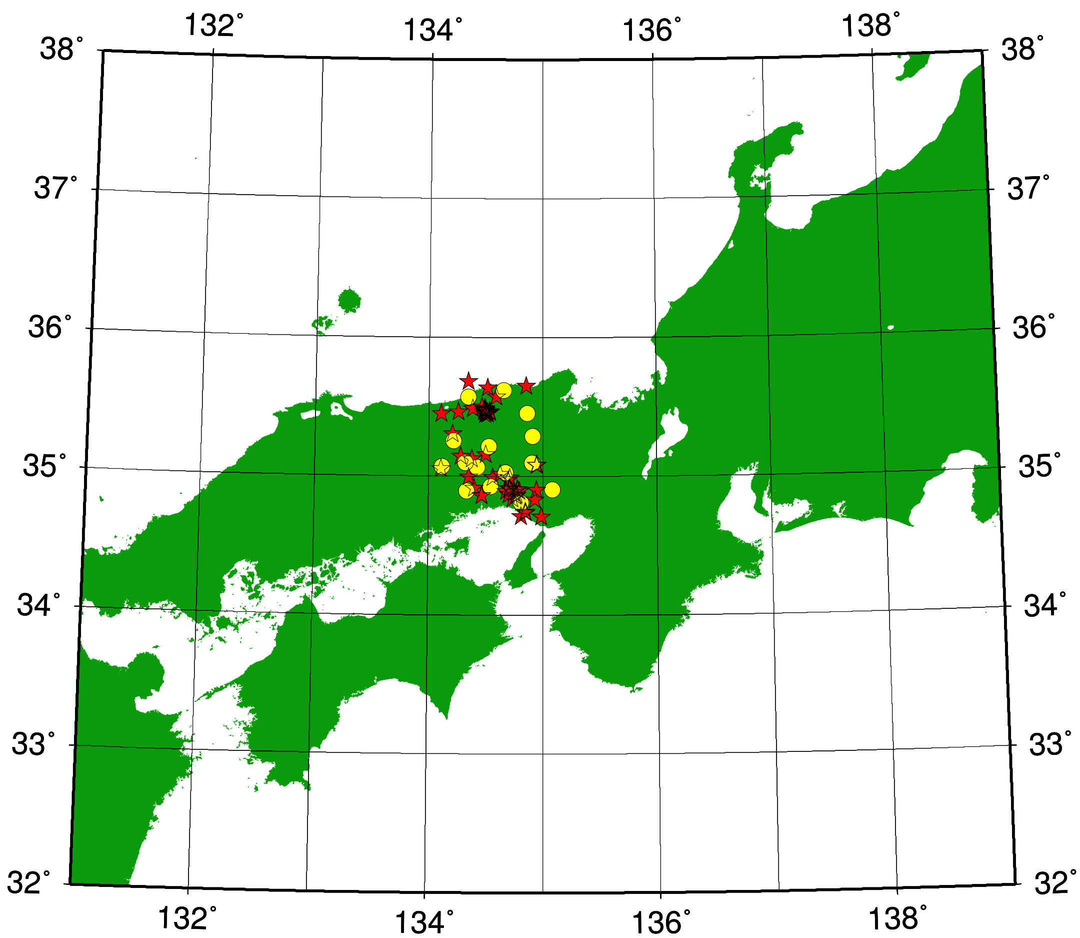
Norms & Nonlinearity

Misfit functions



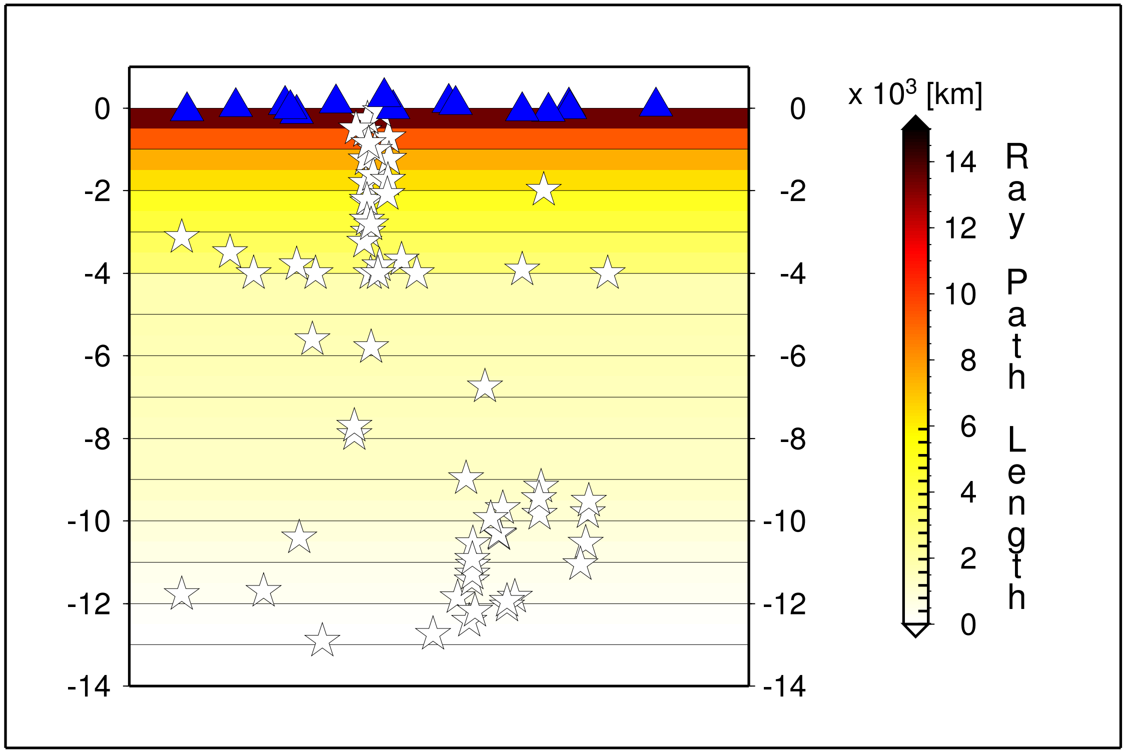
Example-1
Eastern Chugoku

Western Honshu

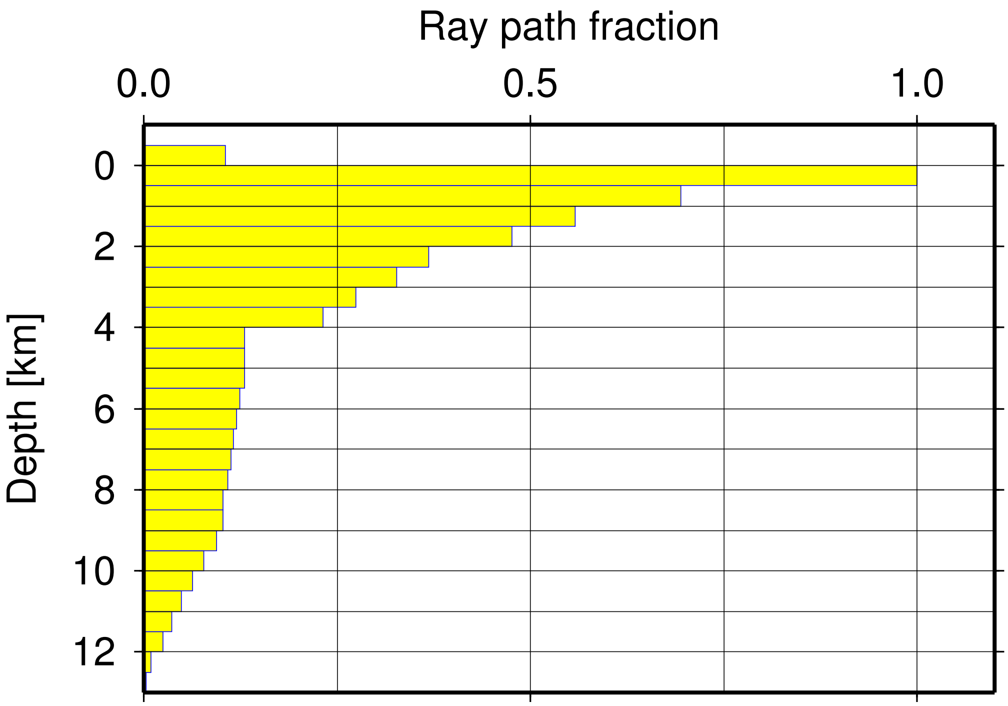


Eastern Chugoku

Ray path Coverage

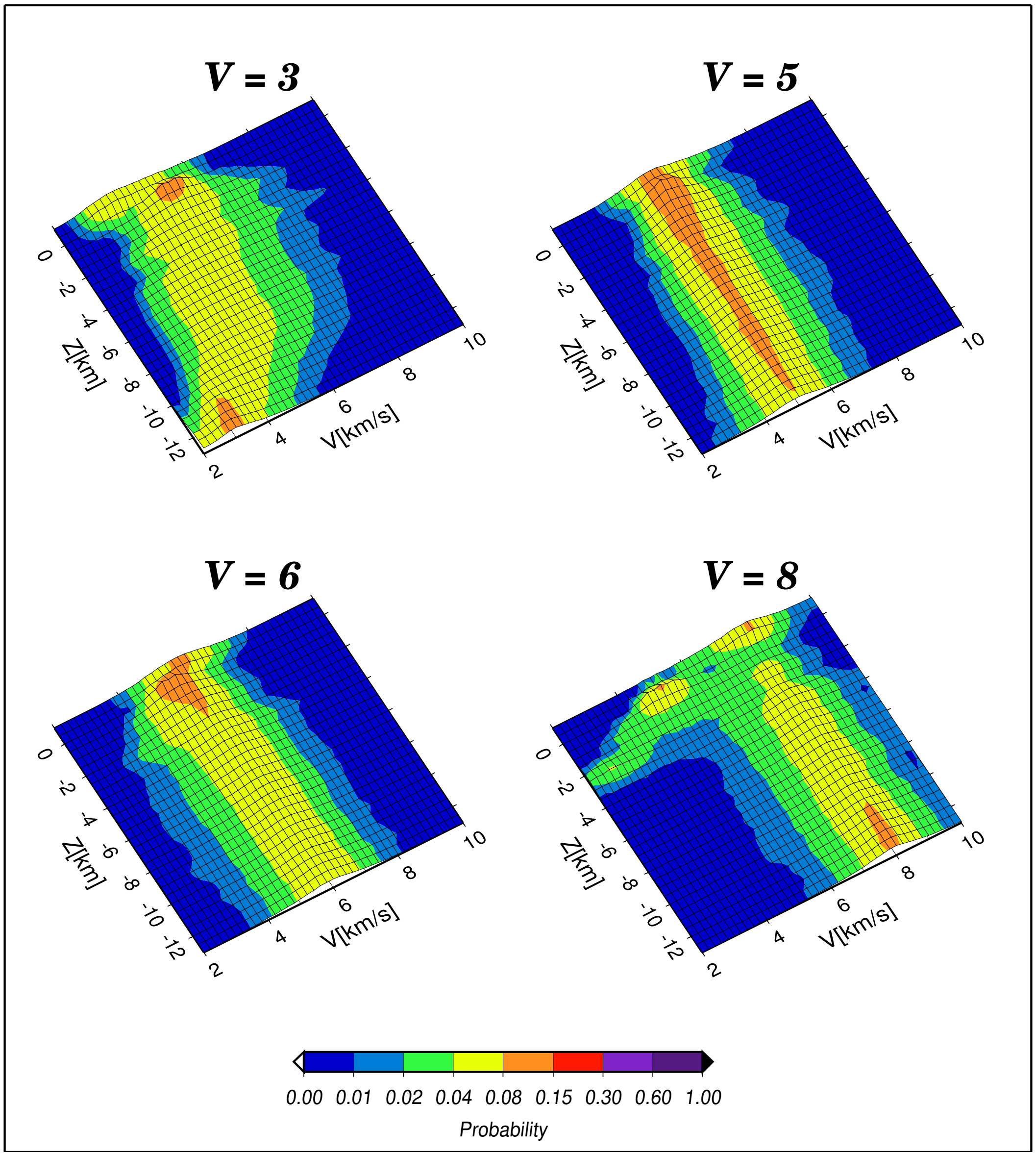


Eastern Chugoku



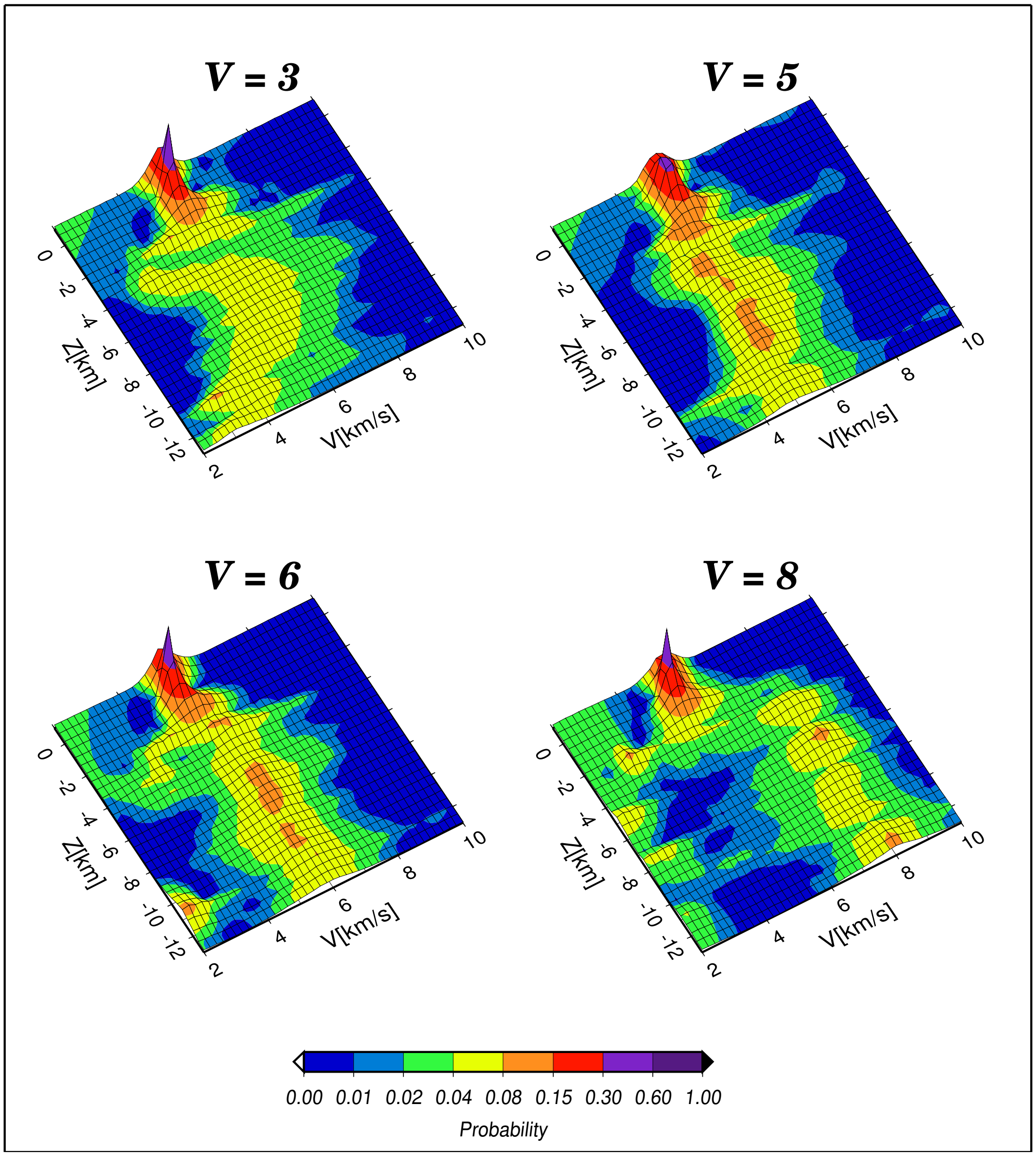
Cd=10 Cm=1

Case A



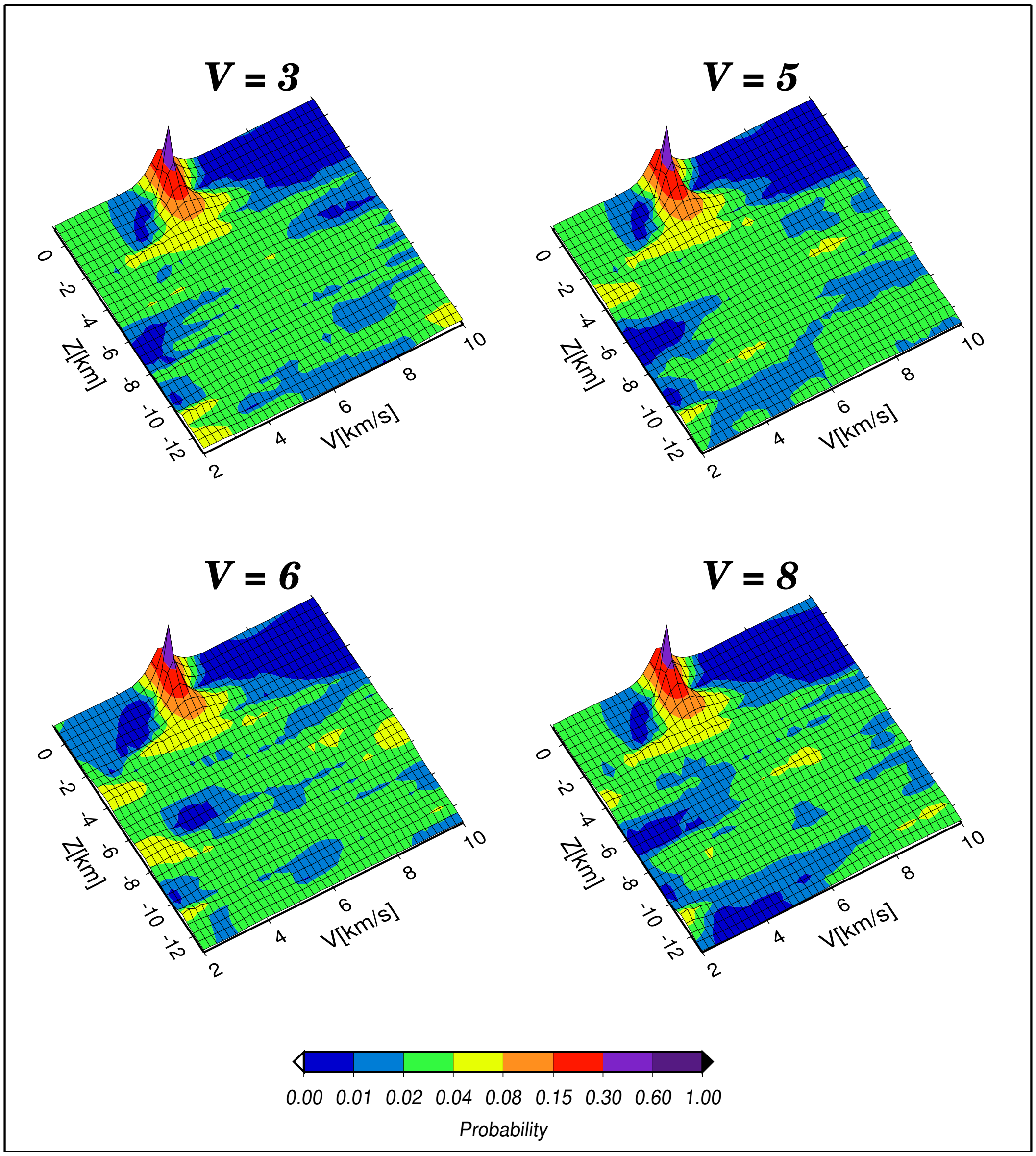
Cd=1 Cm=1

Case B



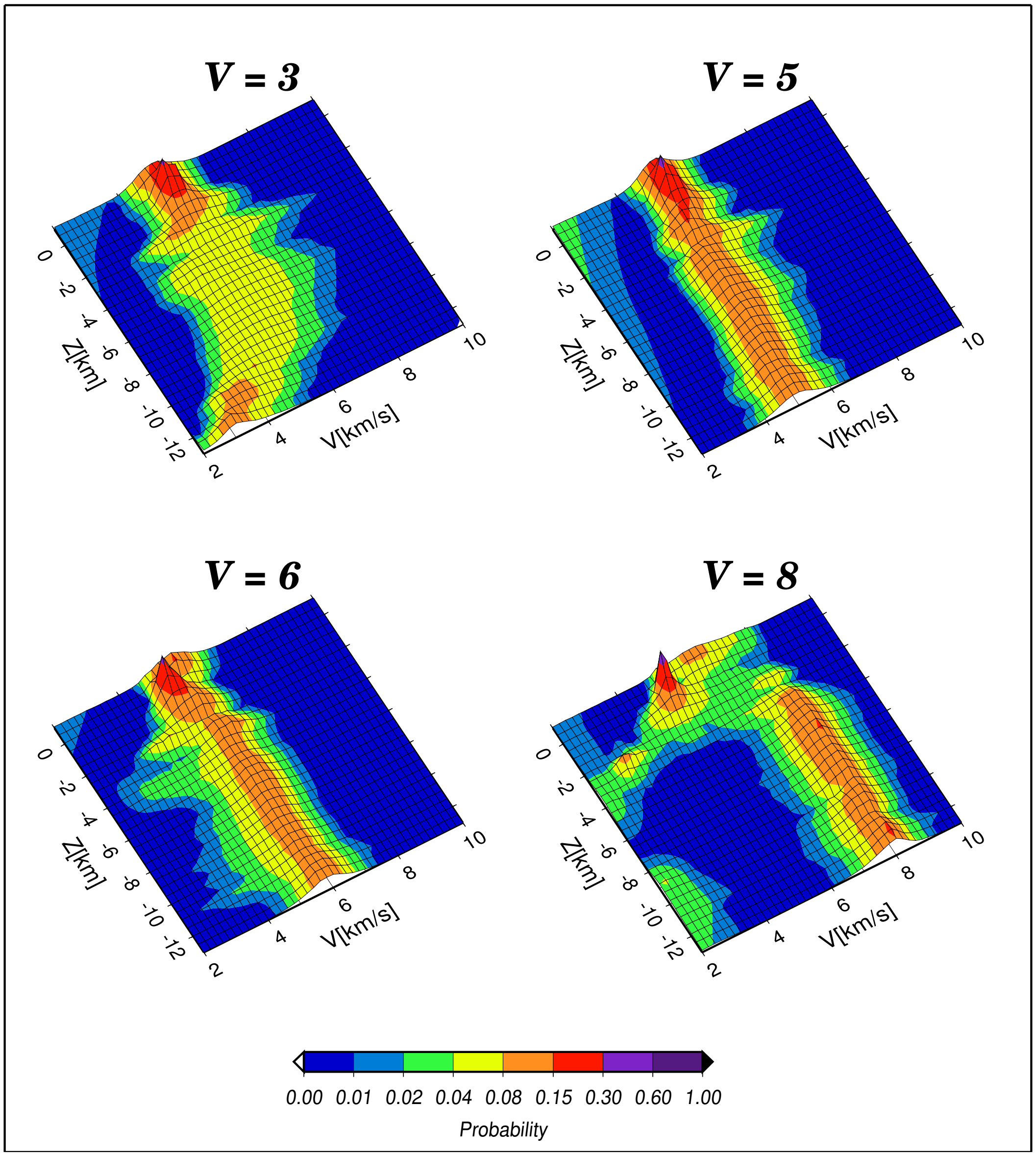
Cd=2 Cm=0.5

Case C



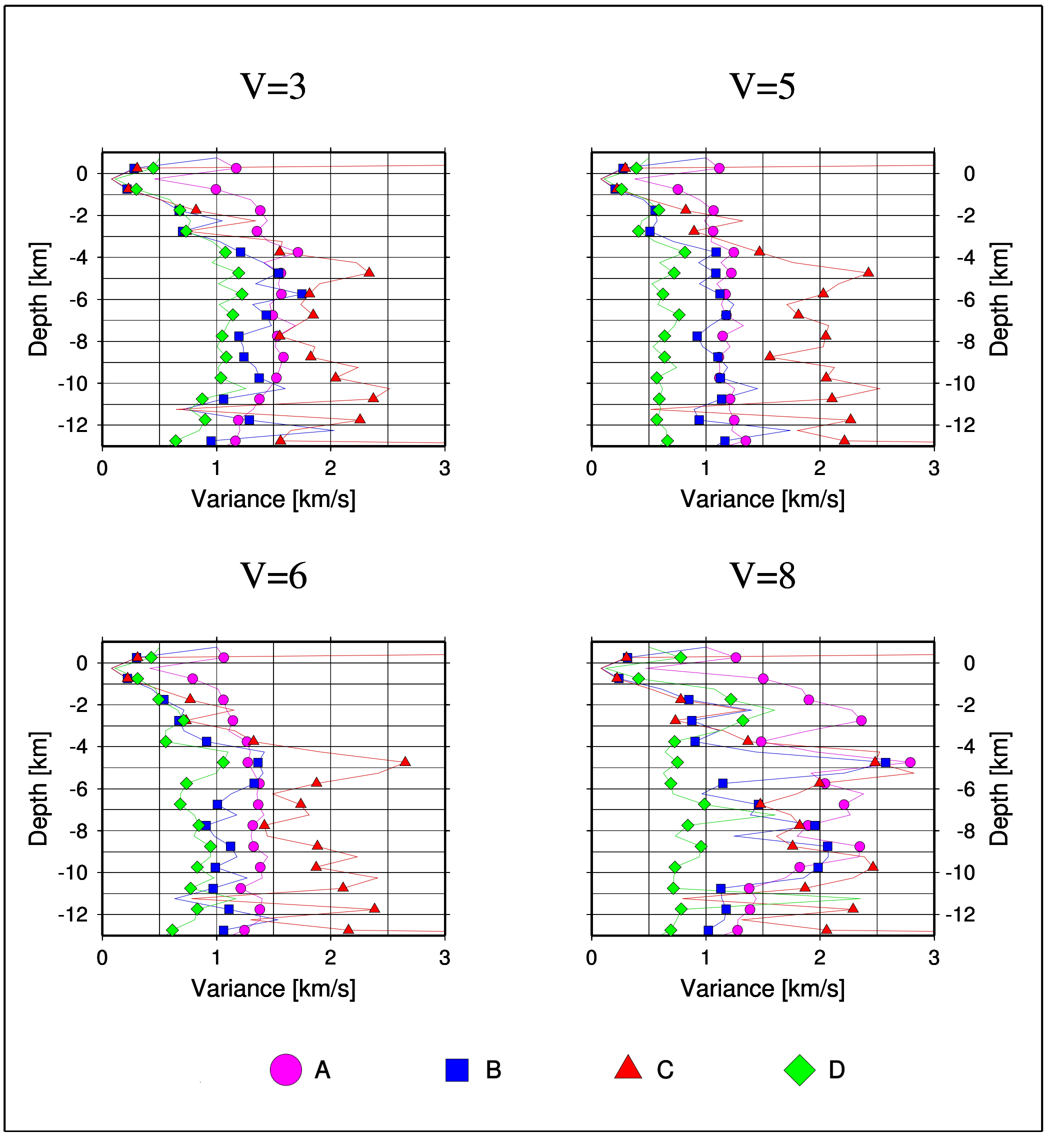
Cd=1 Cm=10

Case D



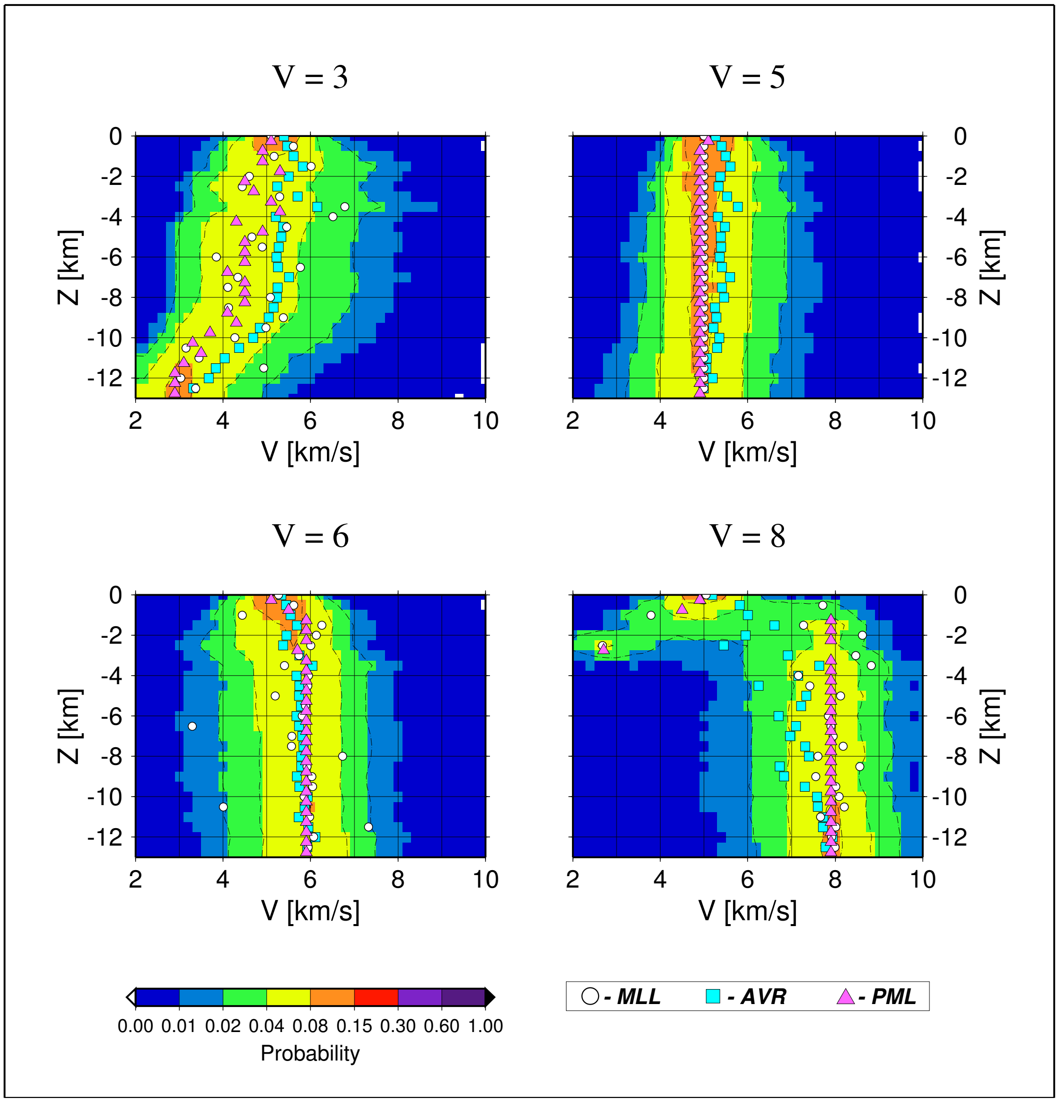
Velocity Errors

Inversion errors



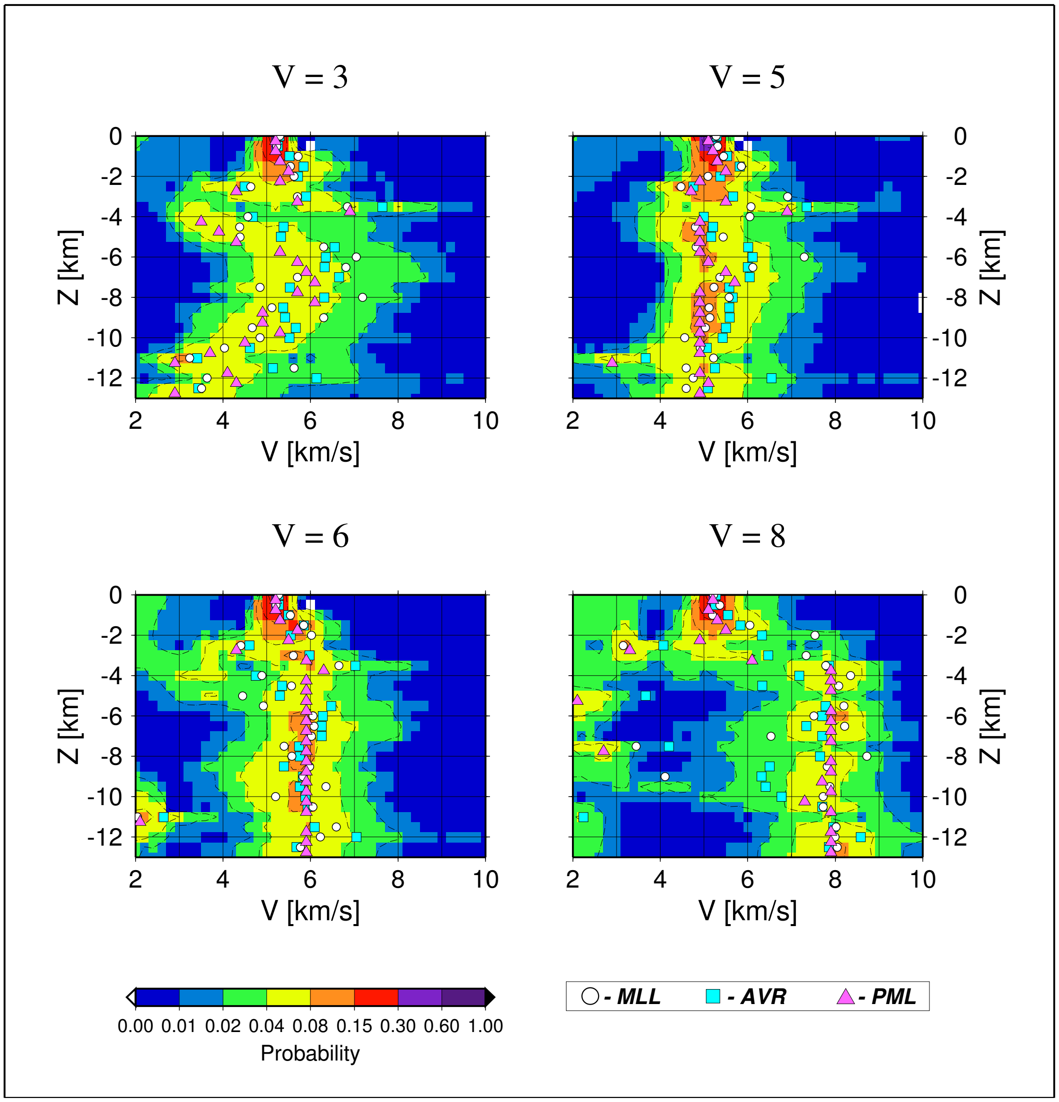
Cd=10 Cm=1

Case A



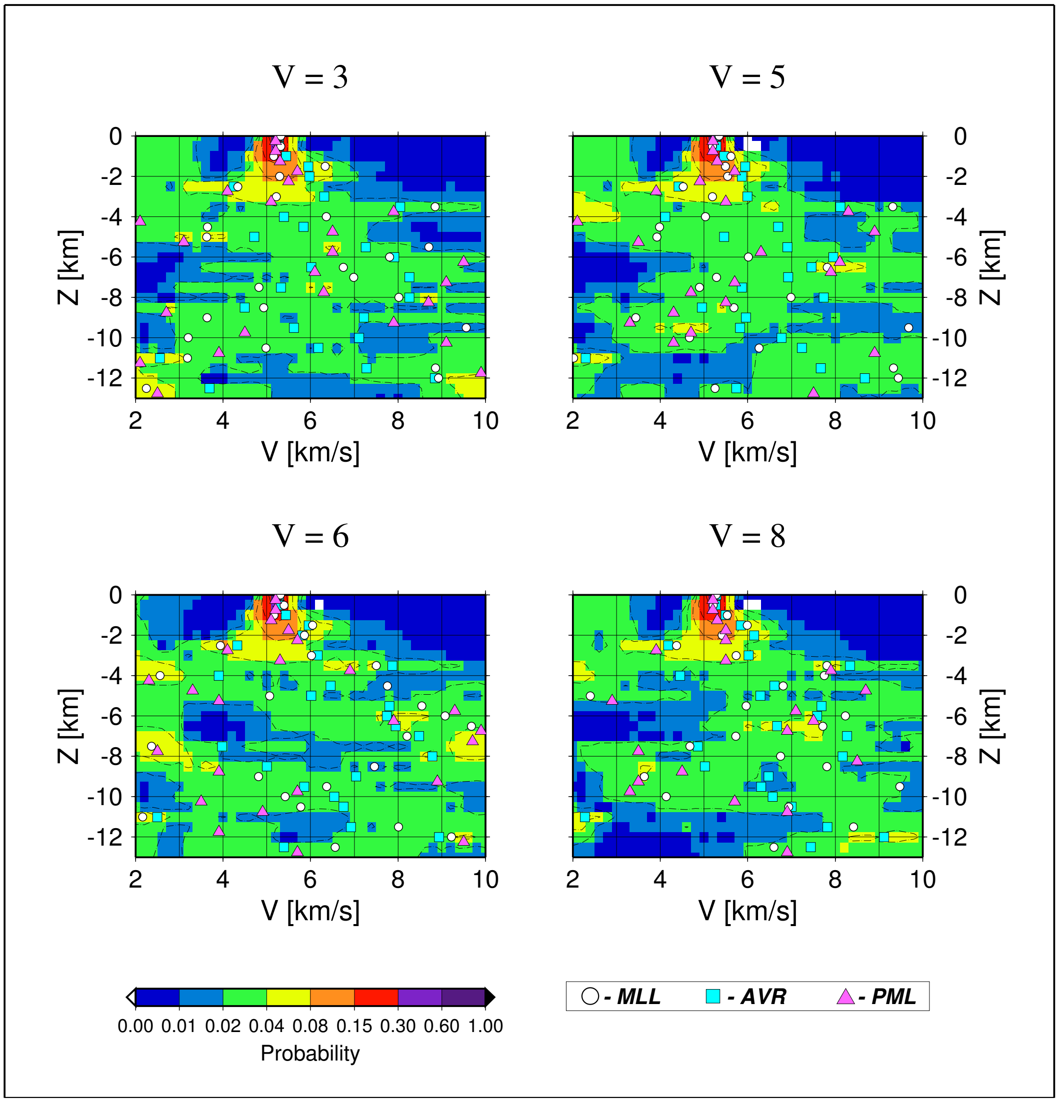
Cd=1 Cm=1

Case B



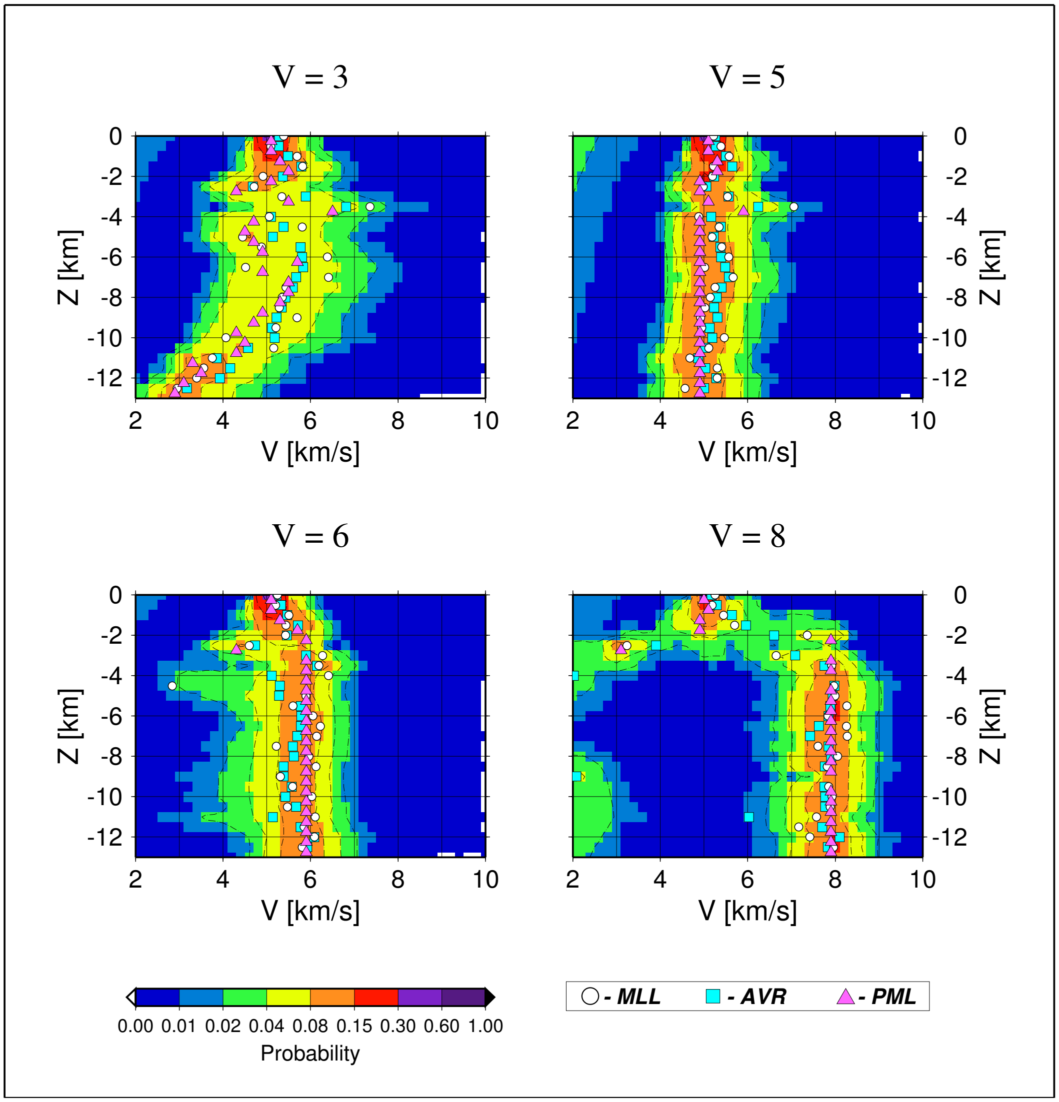
Cd=2 Cm=0.5

Case C

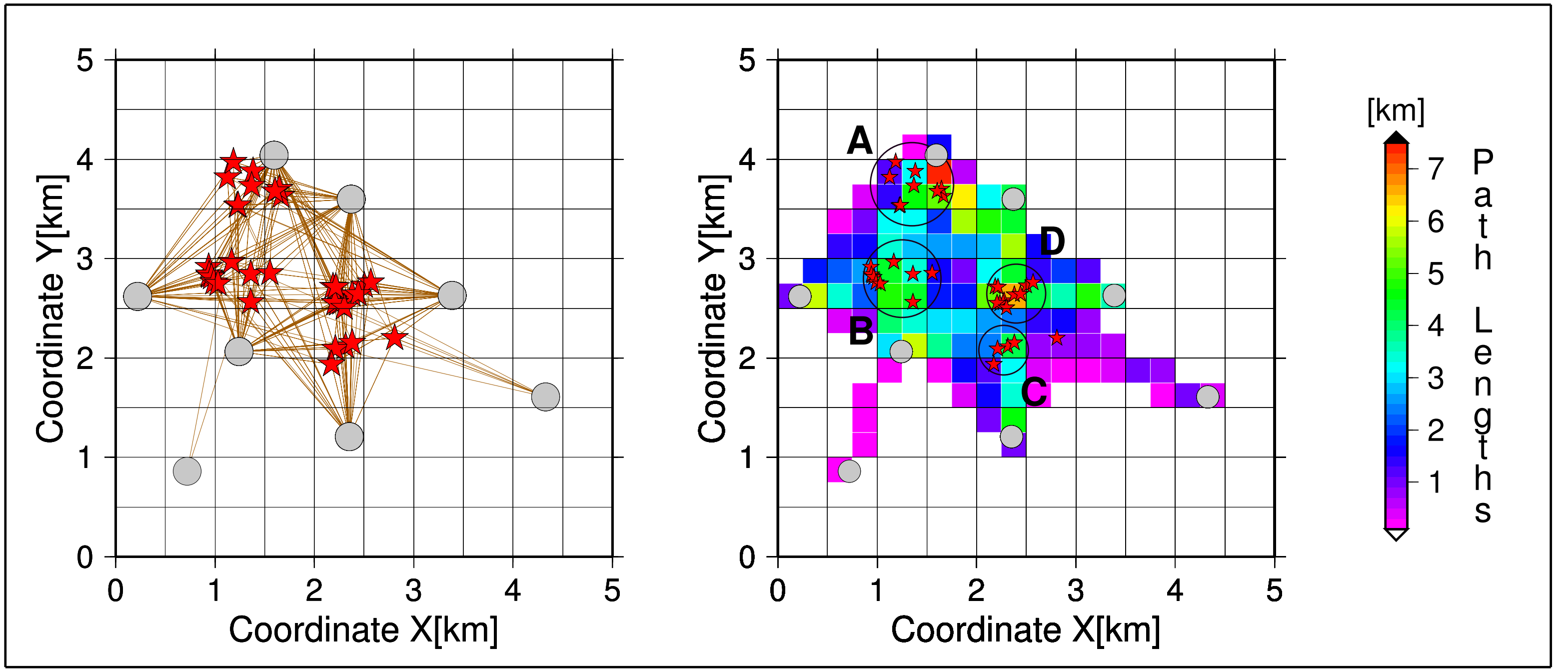


Cd=1 Cm=10

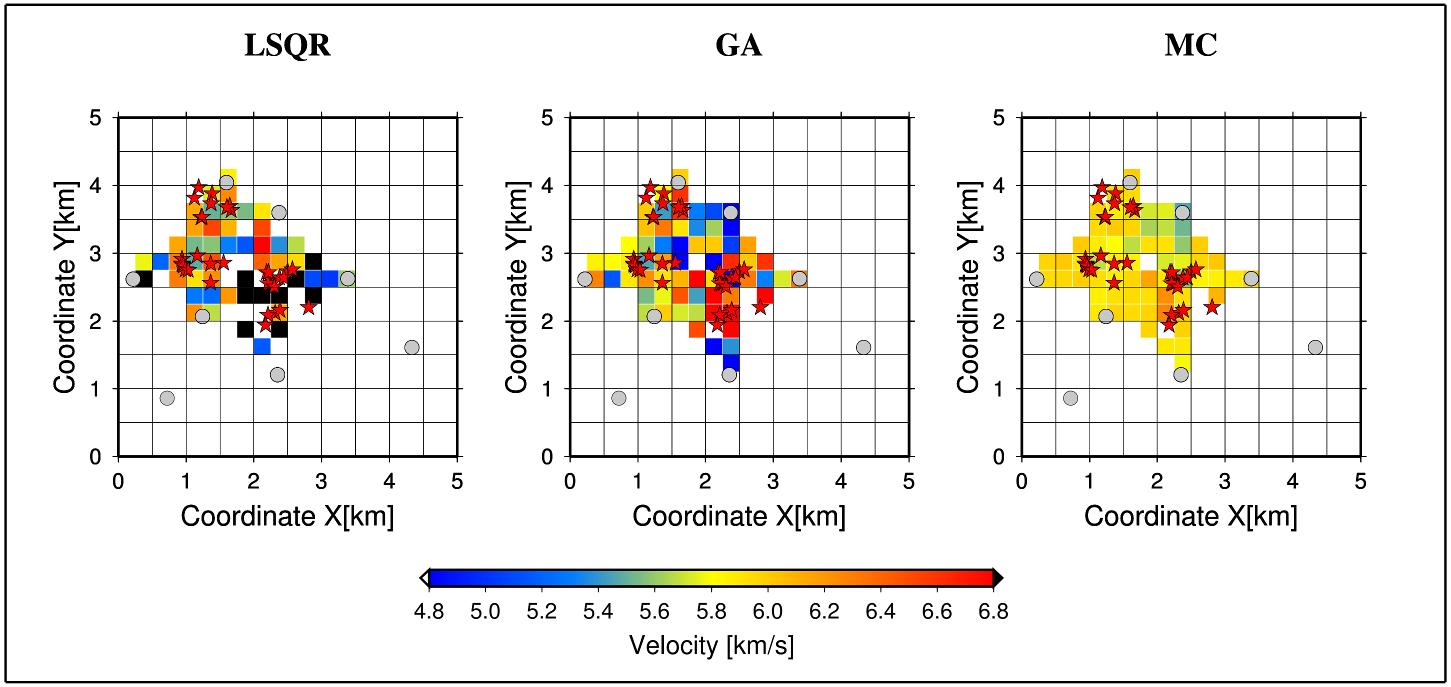
Case D



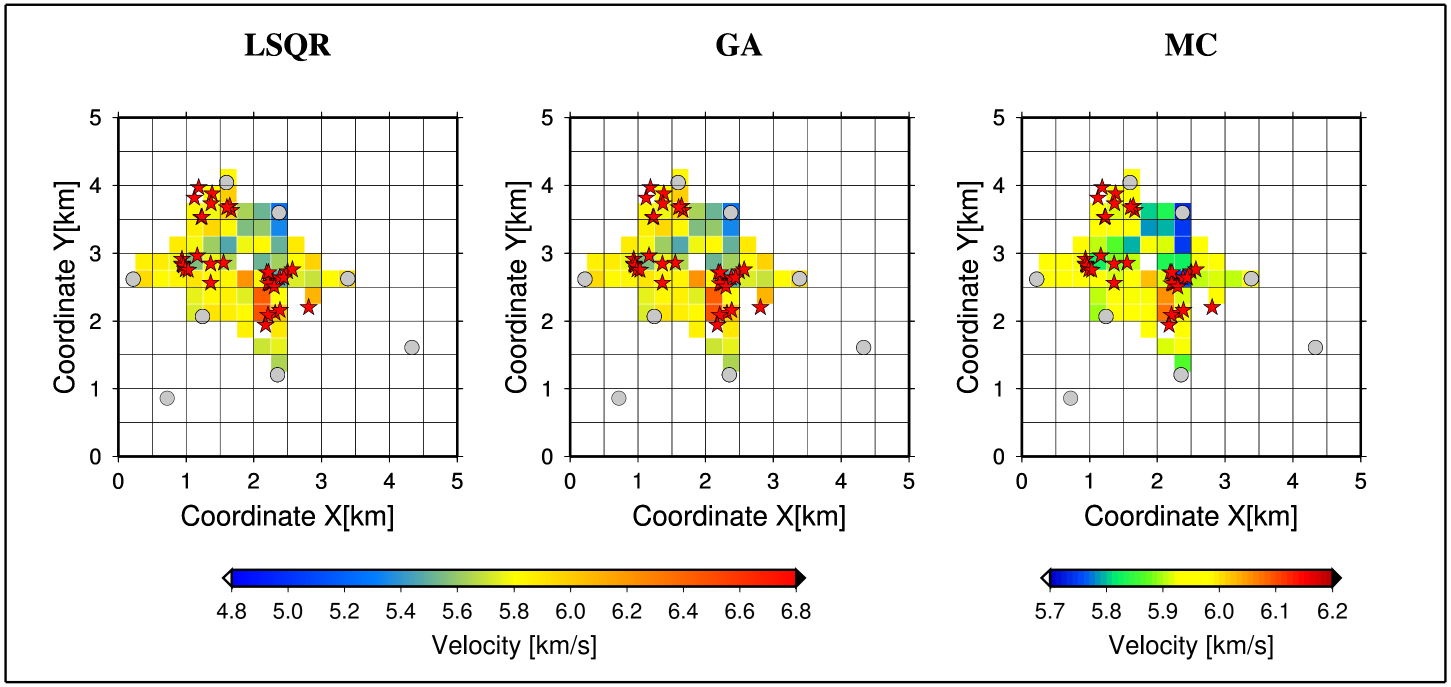
Example-2 Rudna (Poland) copper mine



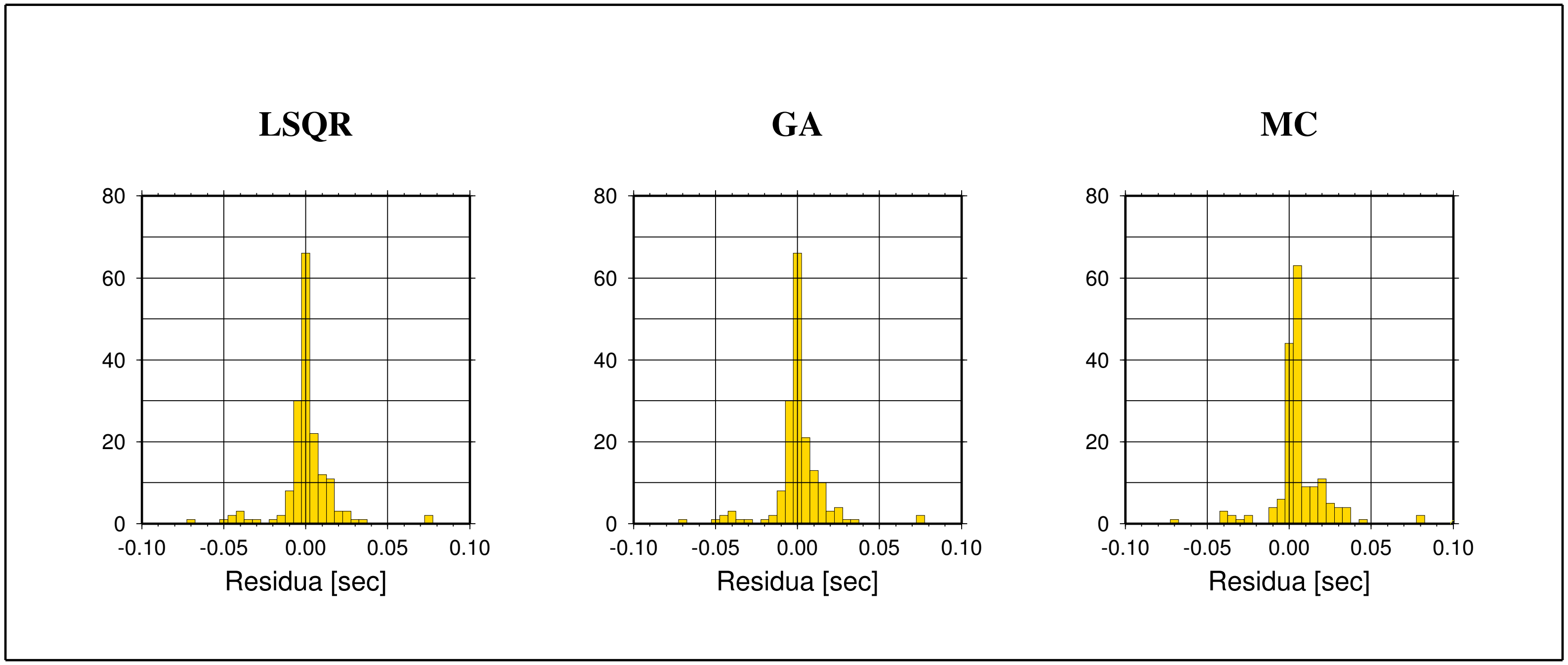
Rudna copper mine



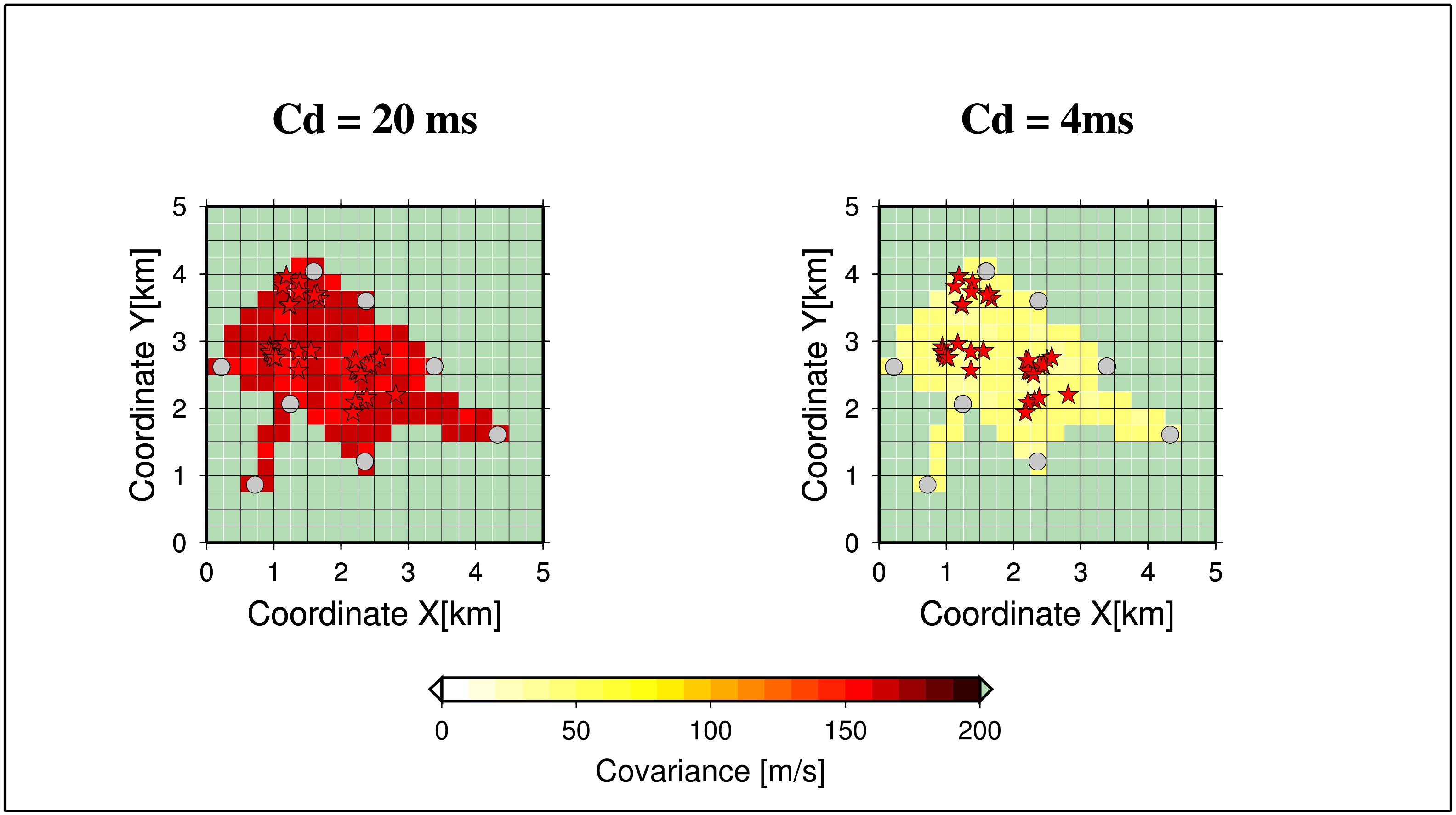
Rudna copper mine



Rudna copper mine

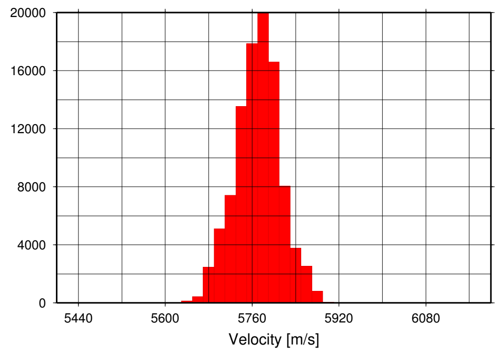


Rudna copper mine

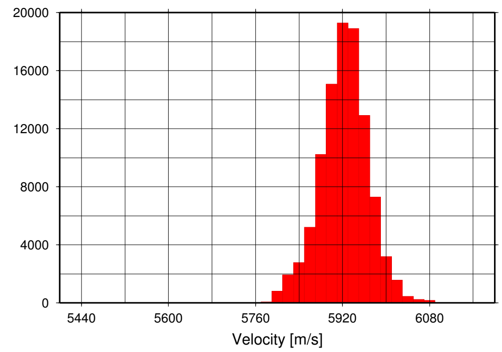


Rudna copper mine

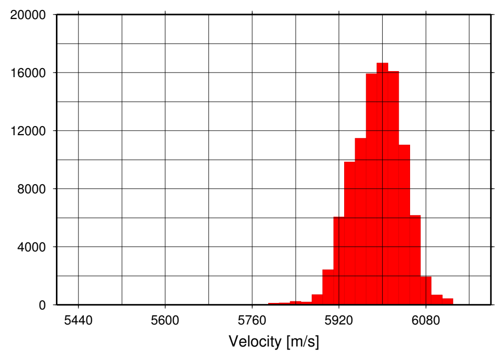
Cd=0.004 Cm=400 Cs=267



Cd=0.004 Cm=400 Cs=226



Cd=0.004 Cm=400 Cs=169



Cd=0.004 Cm=400 Cs=201