Inversion
by
Monte Carlo sampling

Wojciech Dębski*

debski@tono.jnc.go.jp

debski@igf.edu.pl

July 18, 2003

*TGC-JNC, Japan - on leave of absence from IGF PAS, Poland

Plan of the seminar

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

Inverse problems == Indirect Measurement

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

Solution

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

Errors

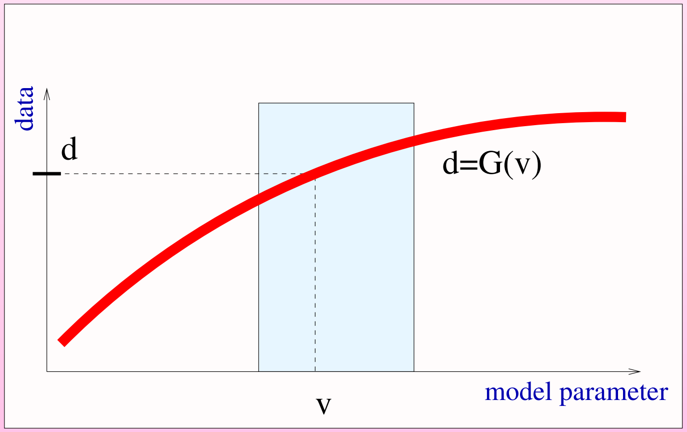
$$\mathbf{m}^{true} = \mathbf{m}^{ml} + \epsilon_{\mathbf{m}}$$

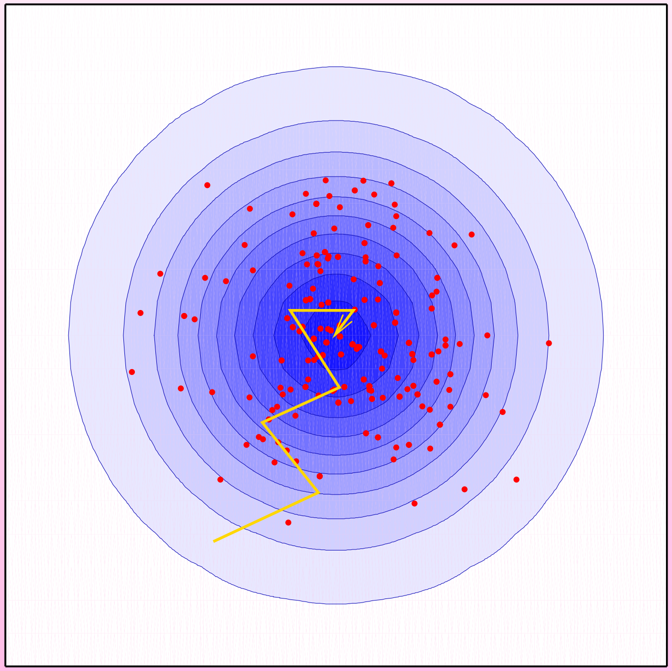
$$\boxed{\epsilon_{\mathbf{m}} = ???}$$

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 problems	- Lack of robustness
Optimization	- Simplicity	- No error estimation
$S(\mathbf{m}) = \Vert\mathbf{G}(\mathbf{m}) - \mathbf{d}^{obs}\Vert + \gamma\Vert\mathbf{m} - \mathbf{m}^a\Vert$	- Fully nonlinear	
Bayesian	- Fully nonlinear	- Complex theory
$\sigma(\mathbf{m}) = f(\mathbf{m})L(\mathbf{m}, \mathbf{d}^{obs})$	- Full error handling	- Efficient sampler

Inversion algorithms - Classical Techniques





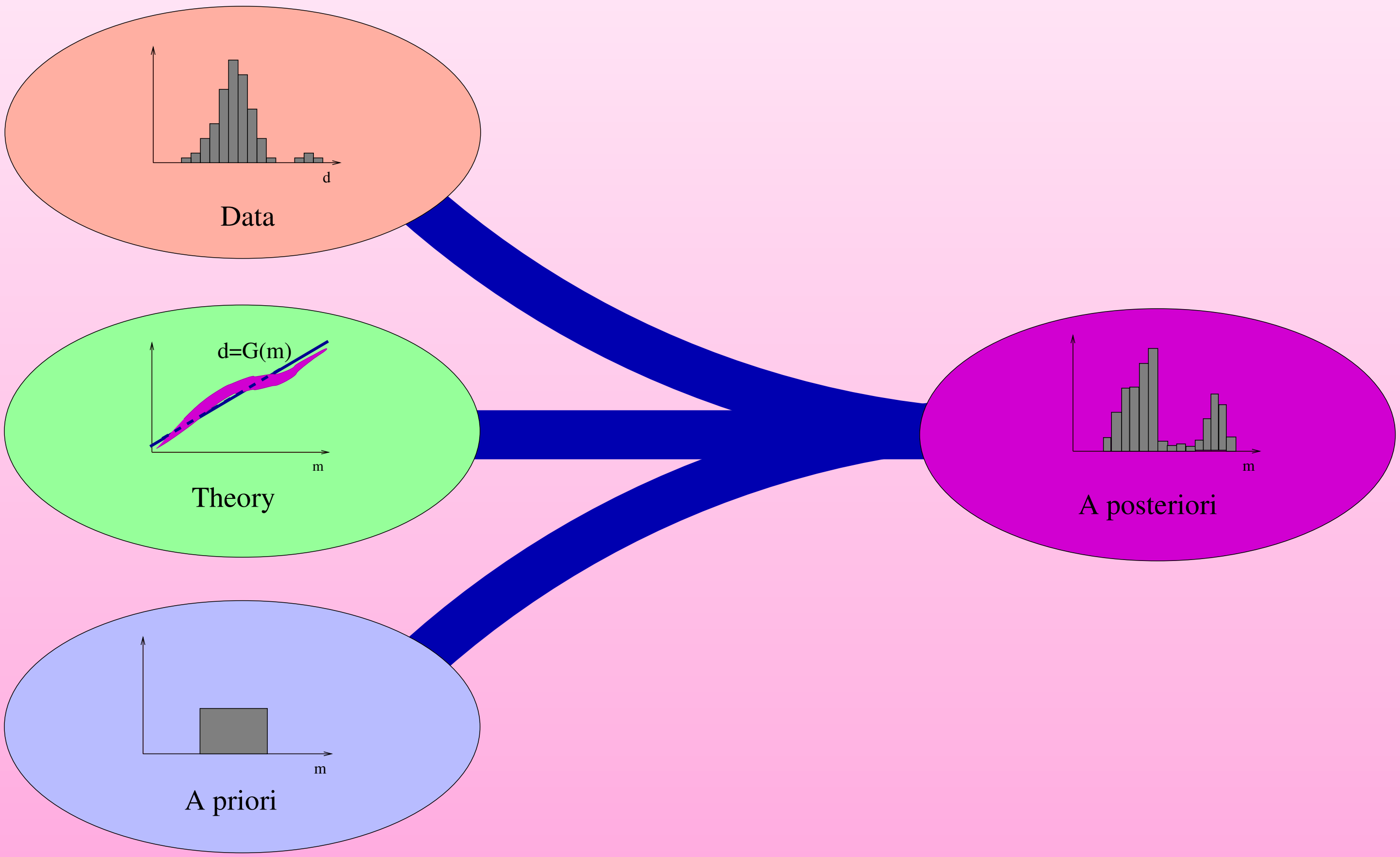
Algebraic

(Back projection)

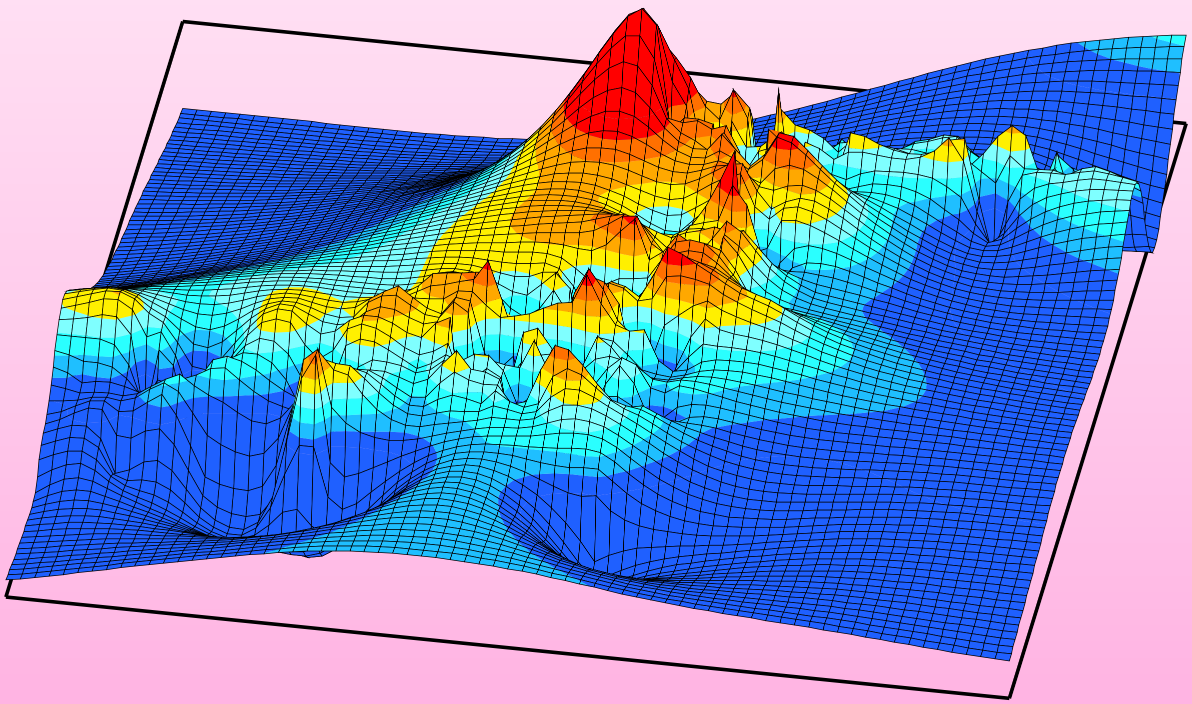
Optimization

(Model space search)

Bayesian Inversion - Basic Ideas



Bayesian inversion -Inspection *A posteriori* pdf



A posteriori pdf

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

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

Probabilistic approach - features

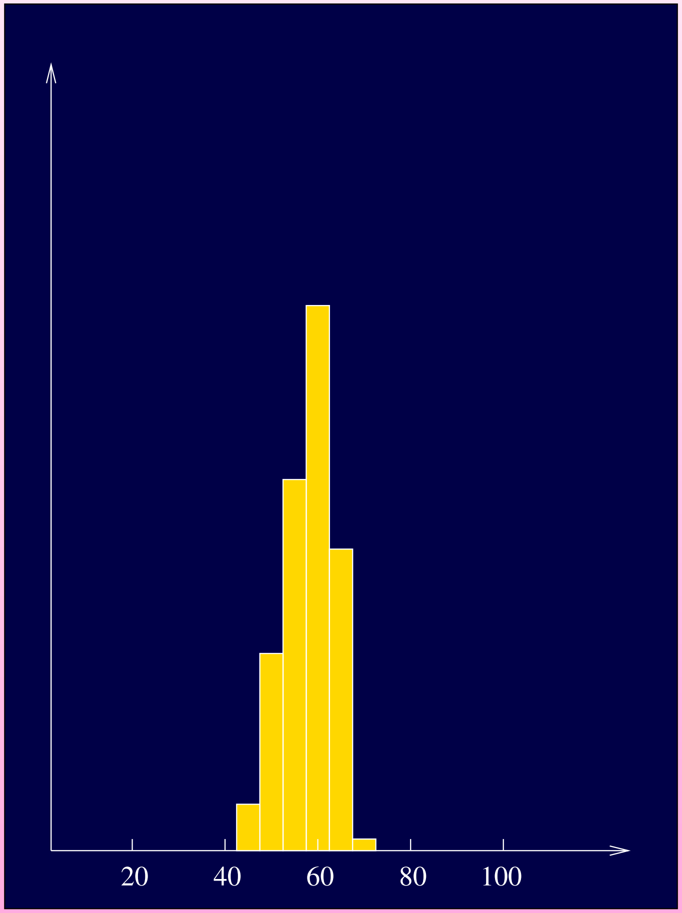
A posteriori pdf $\sigma(m, d)$:

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

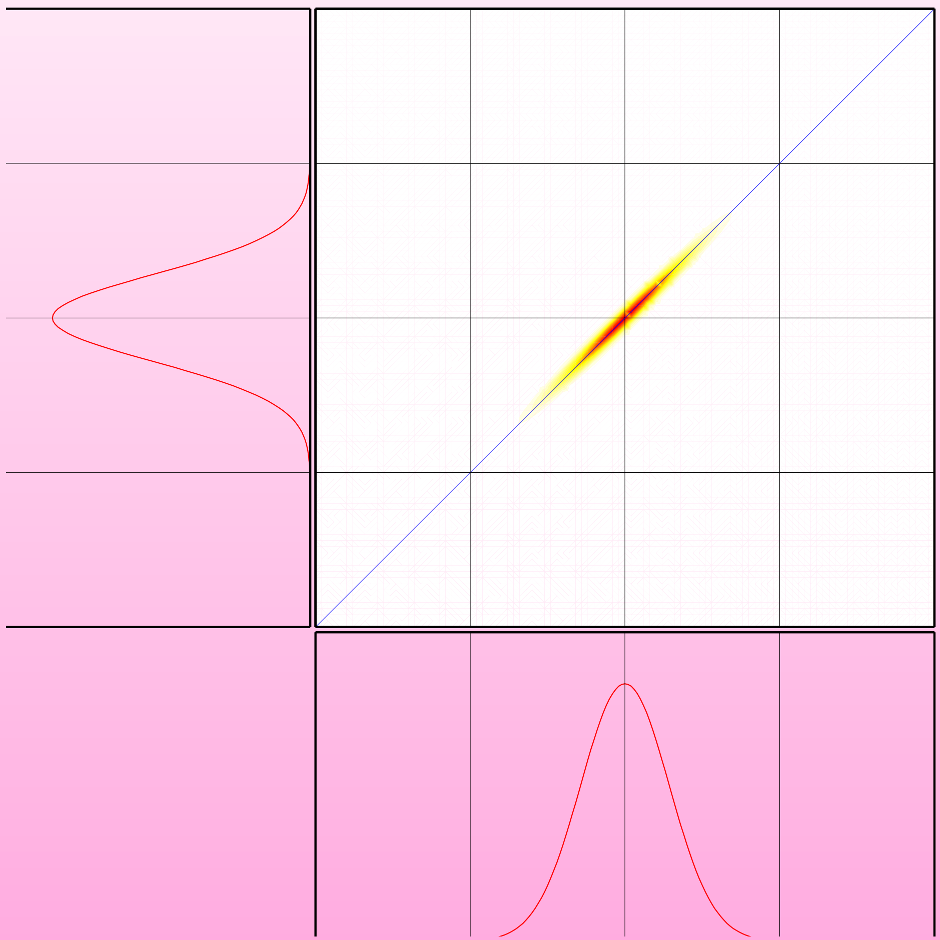
When and why we need to use this approach ???

- nonlinear forward problems
- error analysis
- resolution analysis
- theory verification

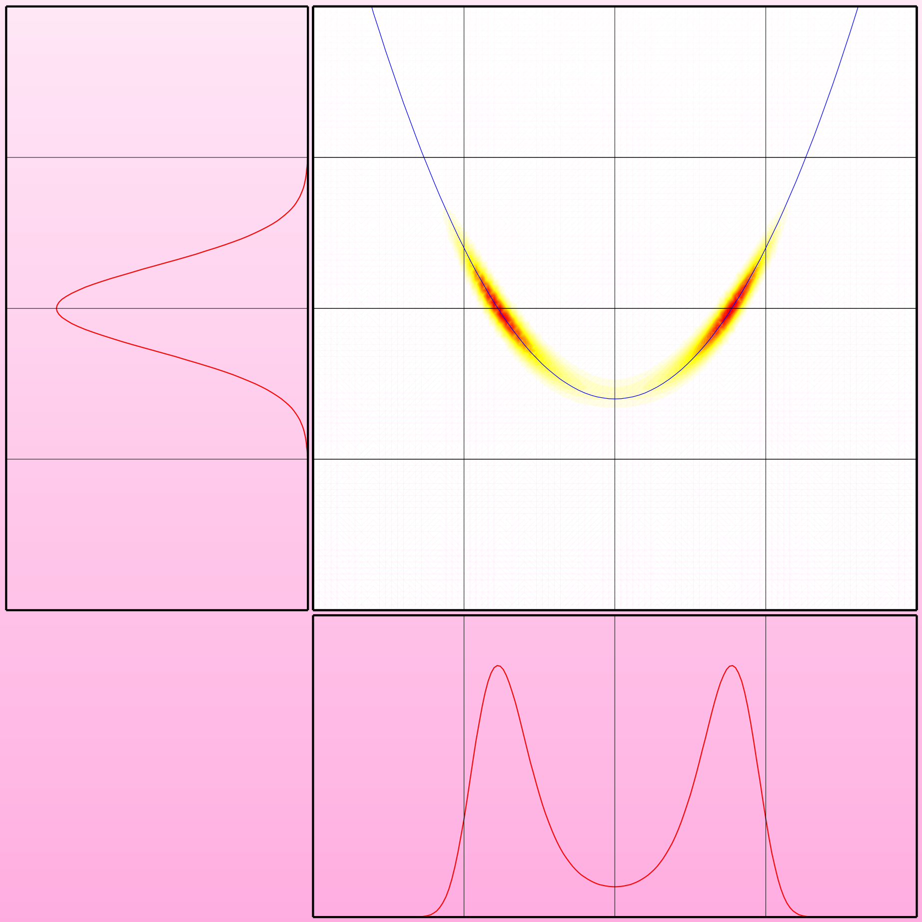
Probabilistic approach is very intuitive!



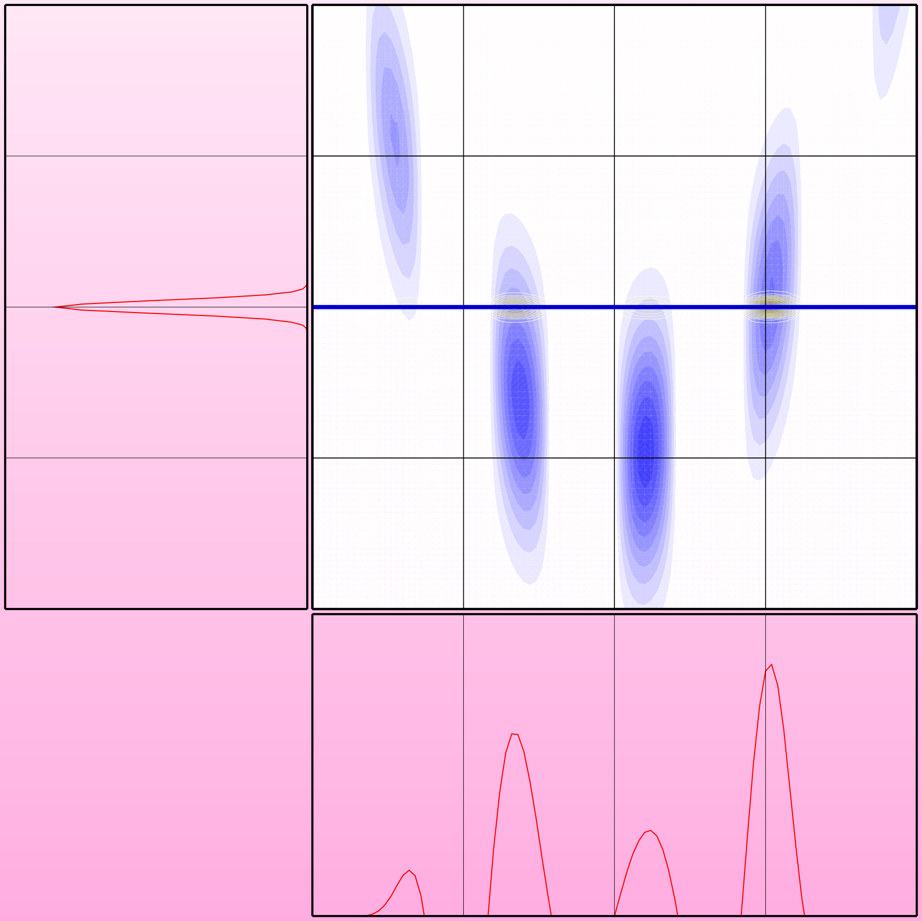
Exact theory: linear case



Exact theory: non-linear case



“Exact” data: uncertain theory



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

- Searching *Maximum Likelihood* solution
  - Gradient methods
  - Global Optimization
- Point Estimators
- Sampling
  - Sampling over regular grid
  - Monte Carlo Sampling

A posteriori estimators

- Point Estimators
  - $\mathbf{m}^{ml} = \mathbf{m}_i$
    $\sigma_M(\mathbf{m}_i) = \max$
  - $\mathbf{m}^{avr} = \sum_{alpha} \mathbf{m}_\alpha$
  - $\delta^{avr} = \sum_\alpha (\mathbf{m}^{avr} - \mathbf{m}_\alpha)^2$

- Marginal pdf

$$\sigma(\mathbf{m}_i) = \int_{\mathbf{m}_j \neq \mathbf{m}_i} \sigma(\mathbf{m})\, \mathrm{d}\mathbf{m}$$

$$\sigma(\mathbf{m}_i) \approx \frac{1}{N} \sum_\alpha N^\alpha_{\mathbf{m} \in (\mathbf{m}_i \pm \epsilon)}$$

A posteriori pdf

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

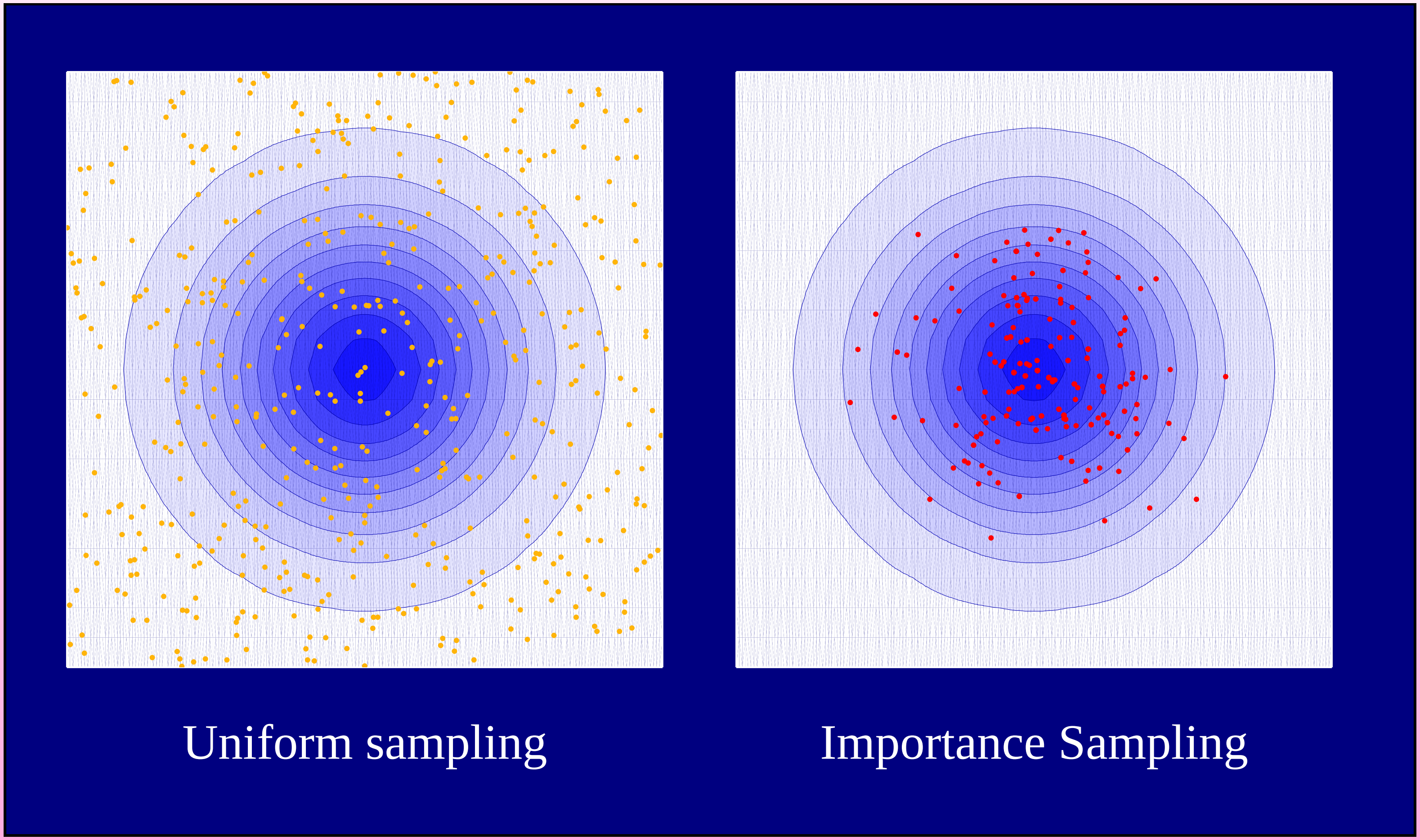
$$L(\mathbf{m}, \mathbf{d}^{obs}) = e^{-||\mathbf{d}^{obs} - \mathbf{G}(\mathbf{m})||}$$

Monte Carlo

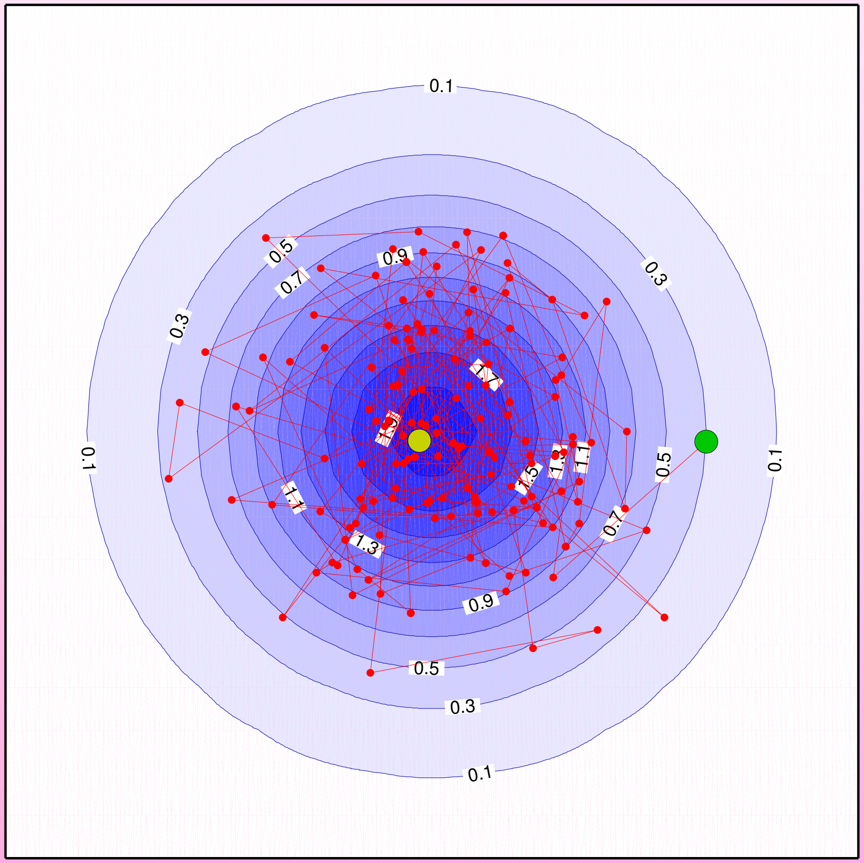
Monte Carlo

Monte Carlo

Monte Carlo sampling



Monte Carlo as Random Walk



Metropolis Algorithm

$$p_{o \to n} = \frac{\sigma(\mathbf{m}_n)}{\sigma(\mathbf{m}_o)}$$

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

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

Travel time tomograpphy - Specification

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

Travel time tomograpphy - Classical approach

Forward modelling:

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

Parameterization:

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

Numerical optimization

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

Travel time tomograpphy - MC approach

A posteriori pdf:

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

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

Monte Carlo sampling

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

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

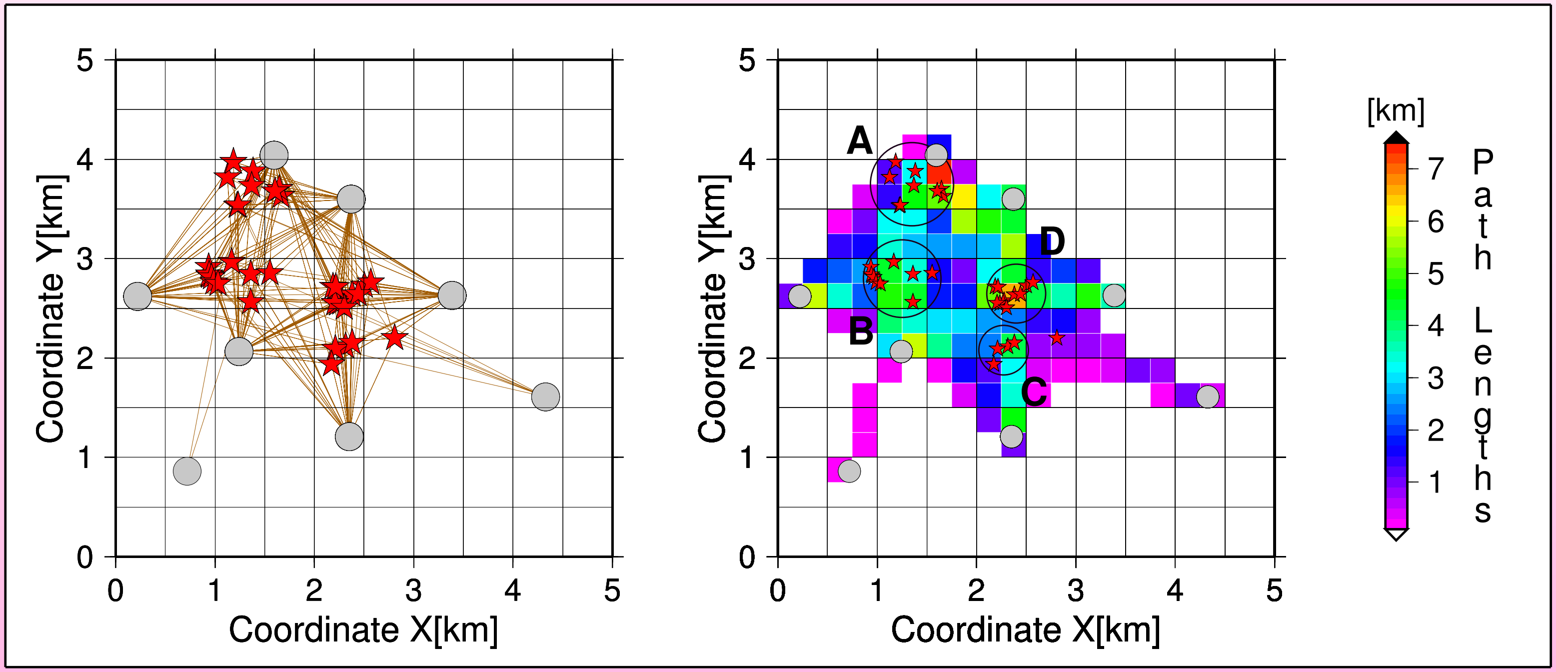
Statistical Estimators

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

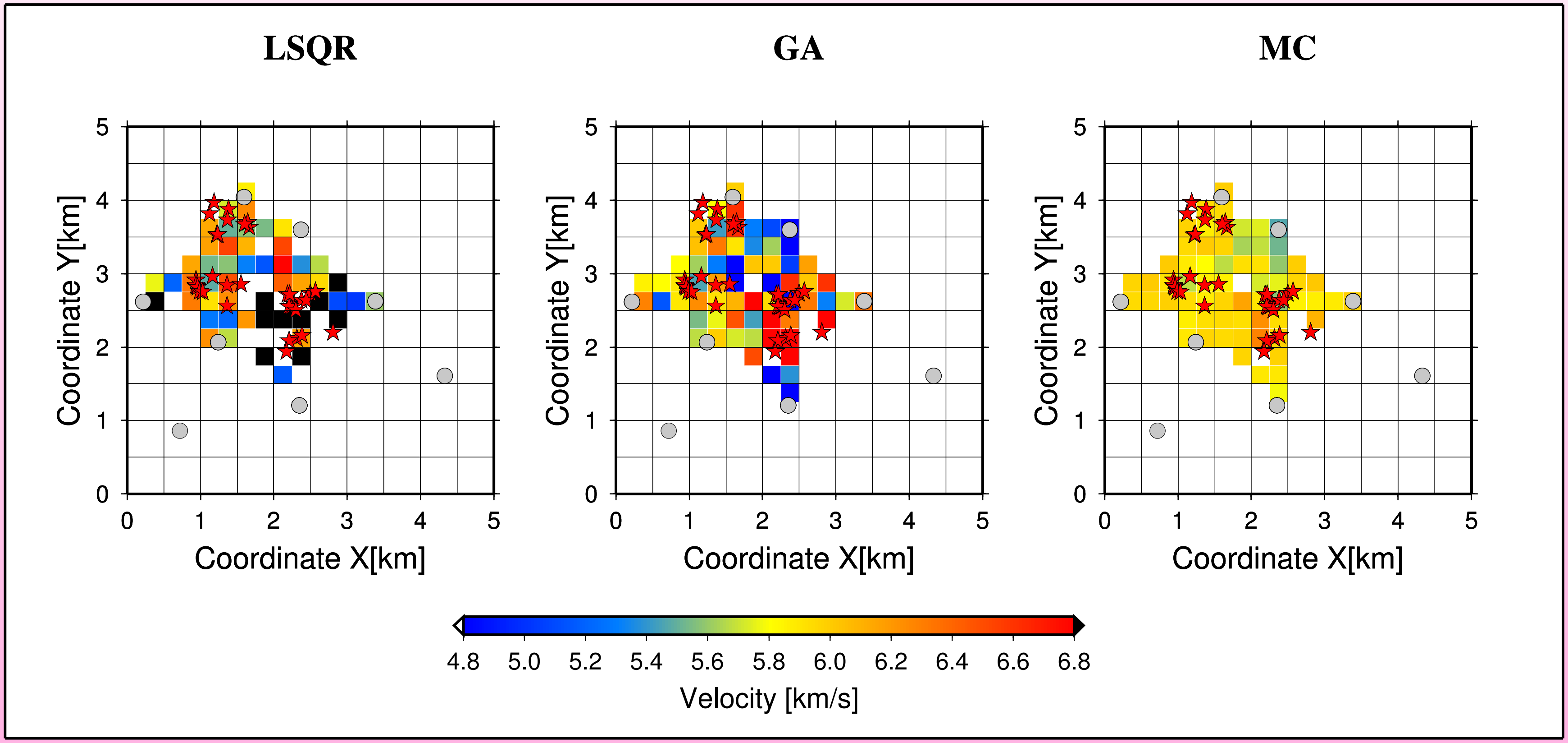
Tomography studies

- Effects of damping and *a priori* model choice
  - dependence on γ
  - dependence on $\mathbf{m}^{apr}$
- Comparison different estimators
  - $\mathbf{m}^{ml}$ - Maximum Likelihood Model
  - $\mathbf{m}^{apr}$ - Average Model
  - $\mathbf{m}^{ml}_{1D}$ ML Model for Marginal PDF

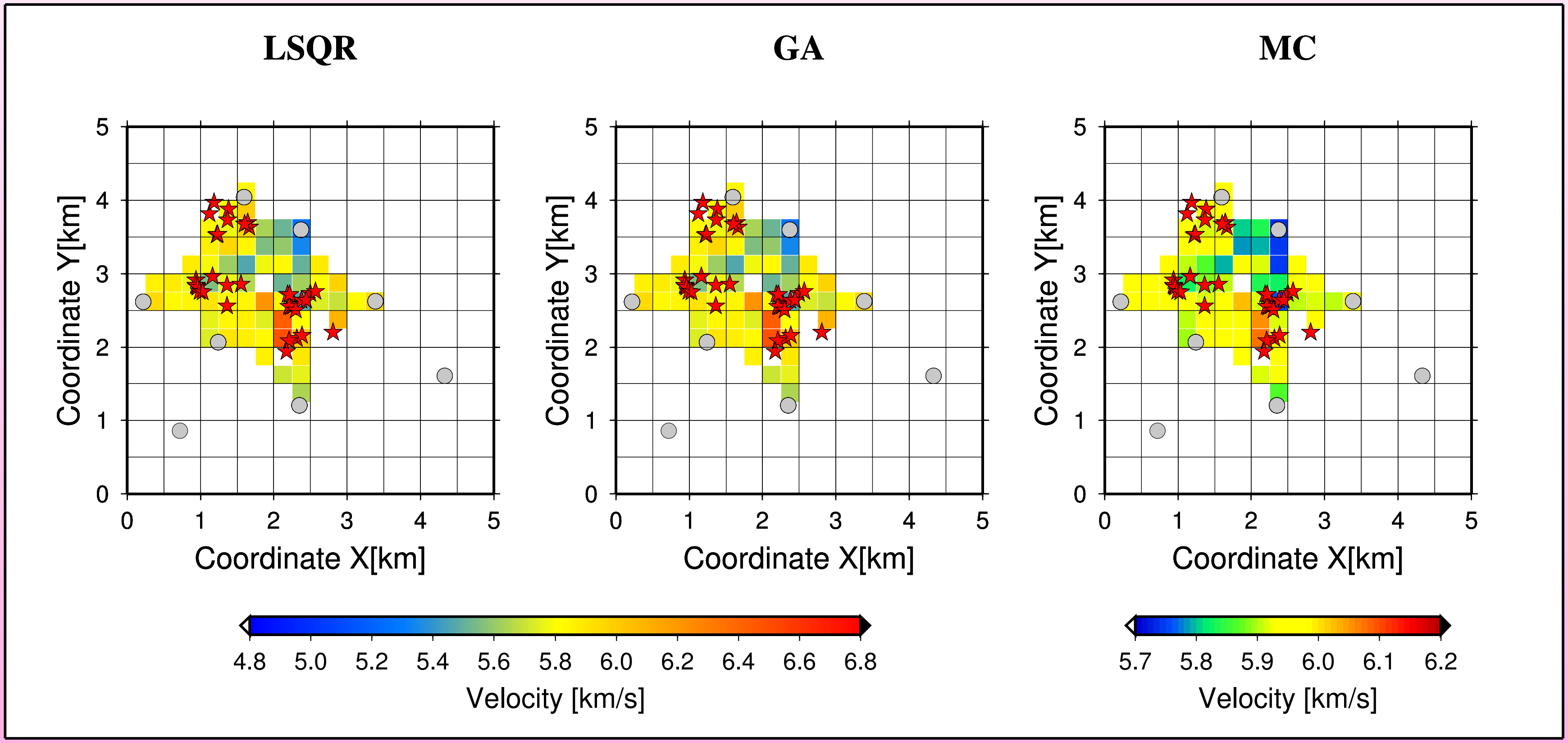
E1: Rudna (Poland) copper mine



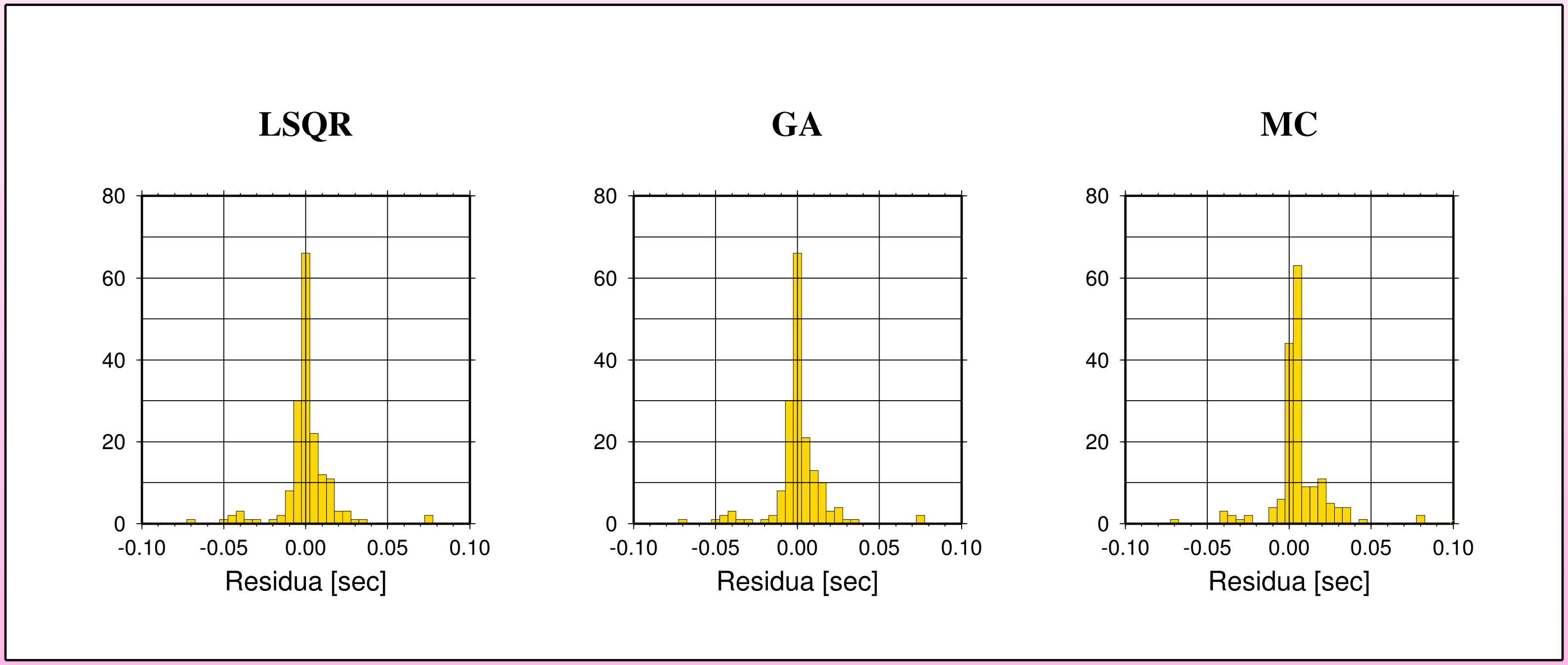
E1: Rudna copper mine - large γ



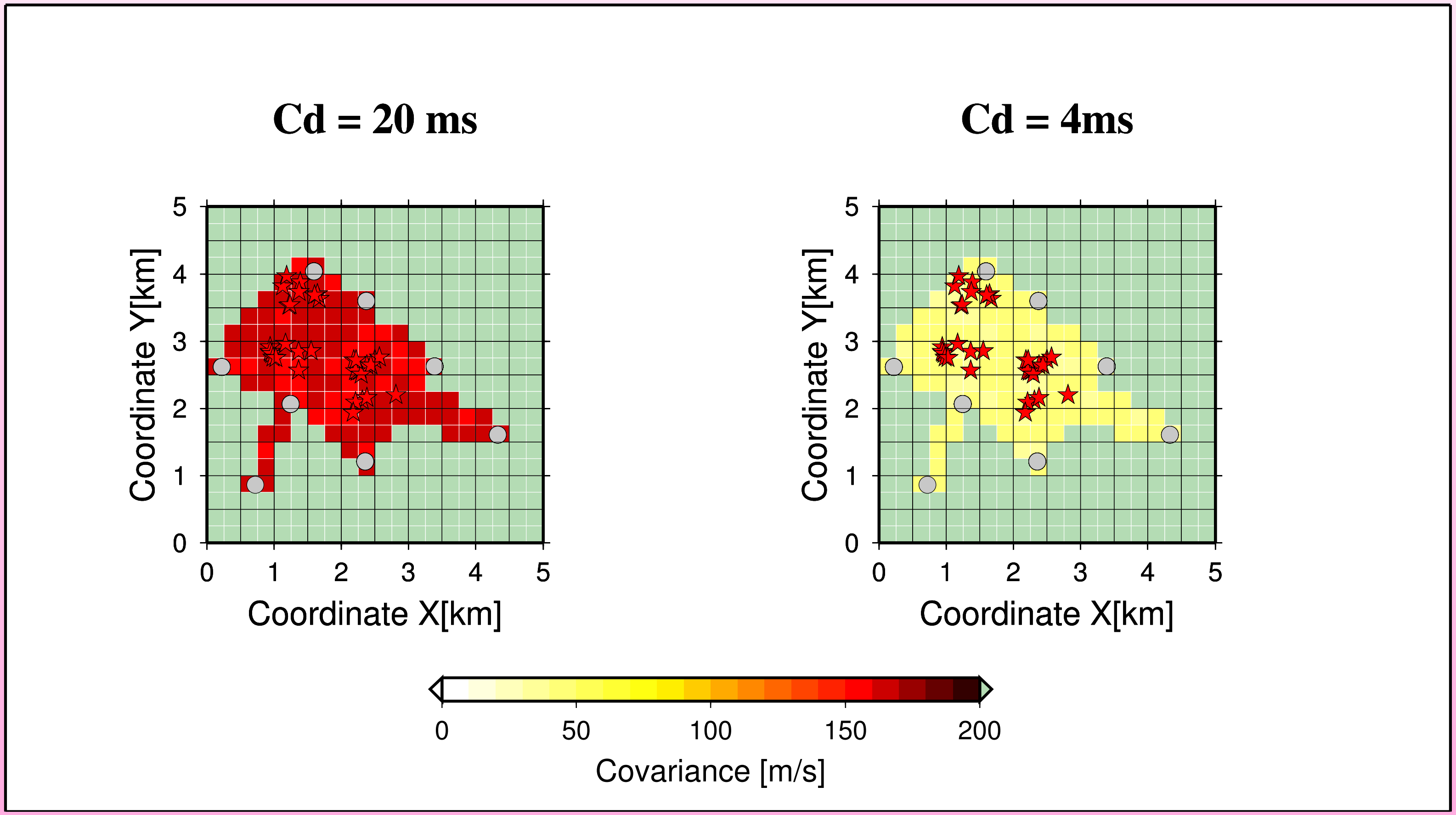
E1: Rudna copper mine - small γ



E1: Travel Time residua - large γ

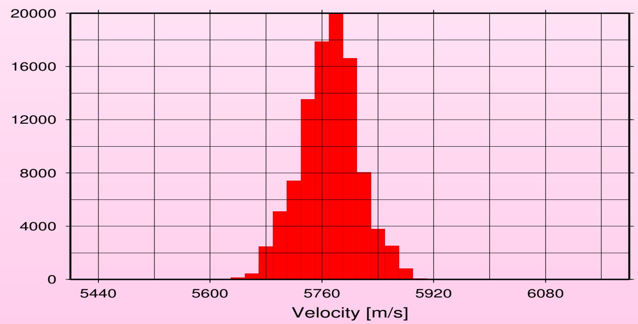


E1: Variance (error) distribution

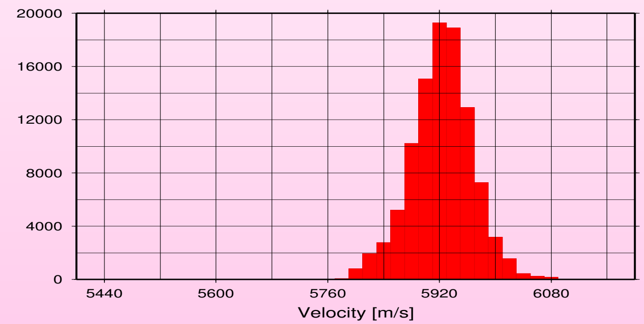


E1: Marginal distributions - small γ

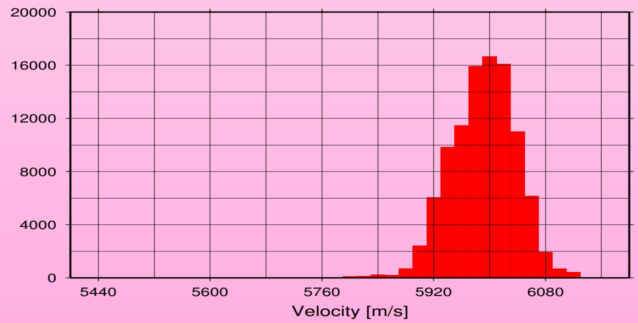
Cd=0.004 Cm=400 Cs=267



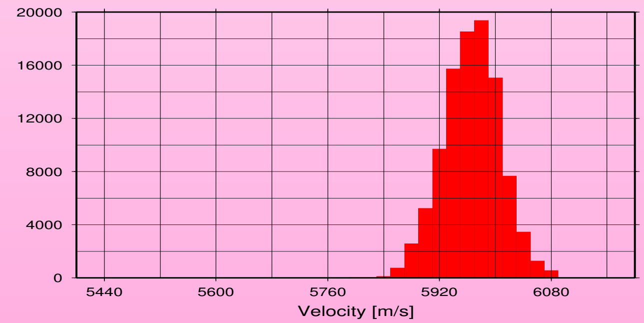
Cd=0.004 Cm=400 Cs=226



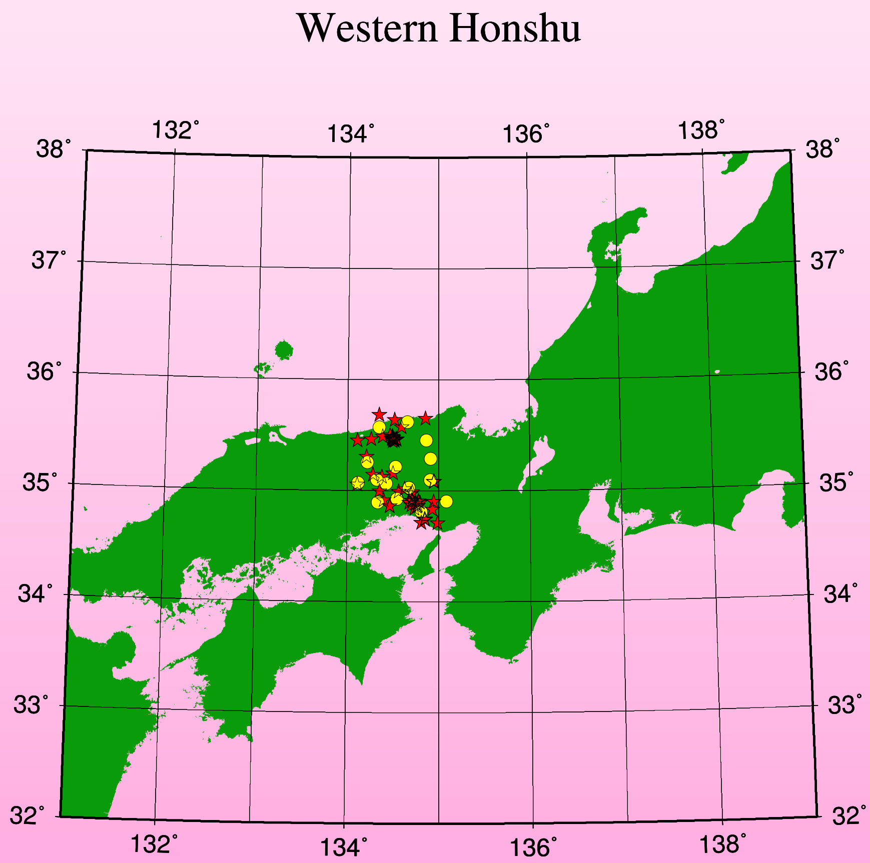
Cd=0.004 Cm=400 Cs=169



Cd=0.004 Cm=400 Cs=201

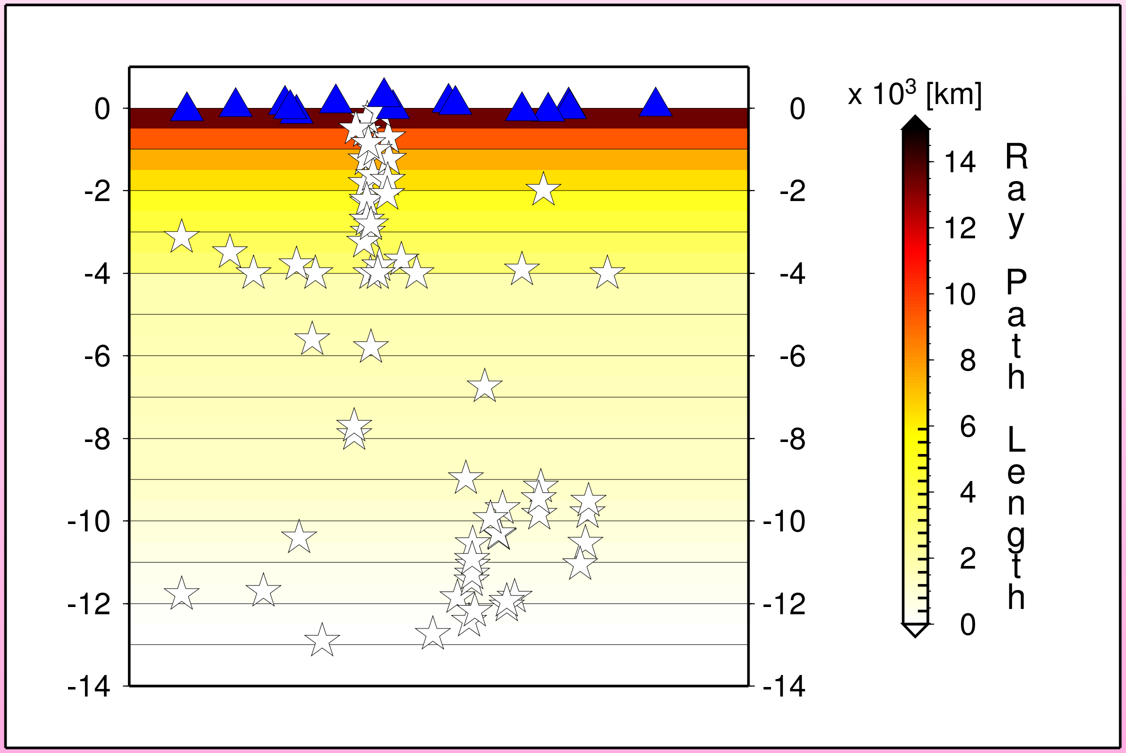


E2: Eastern Chugoku

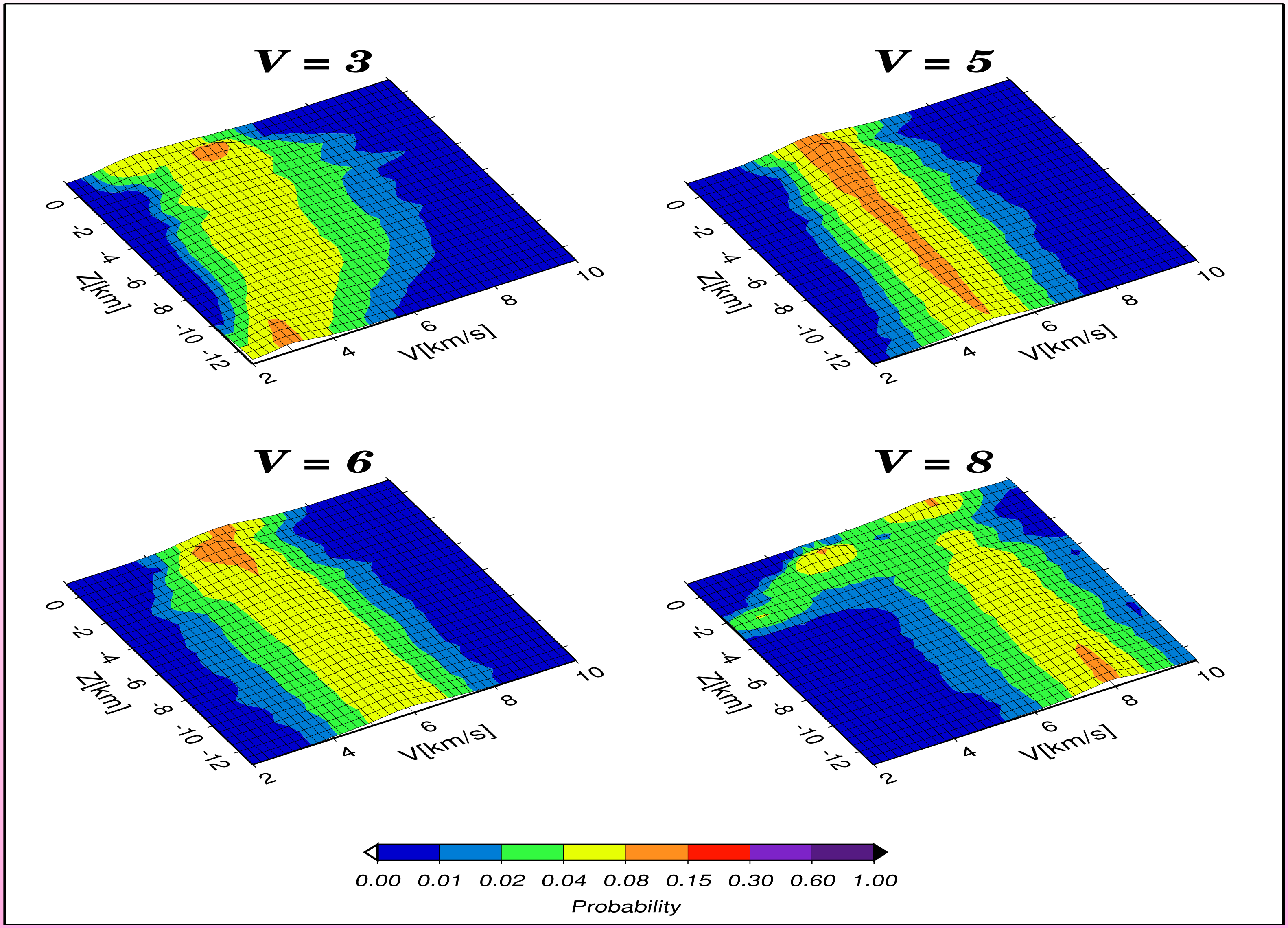


E2: Ray paths

Ray path Coverage

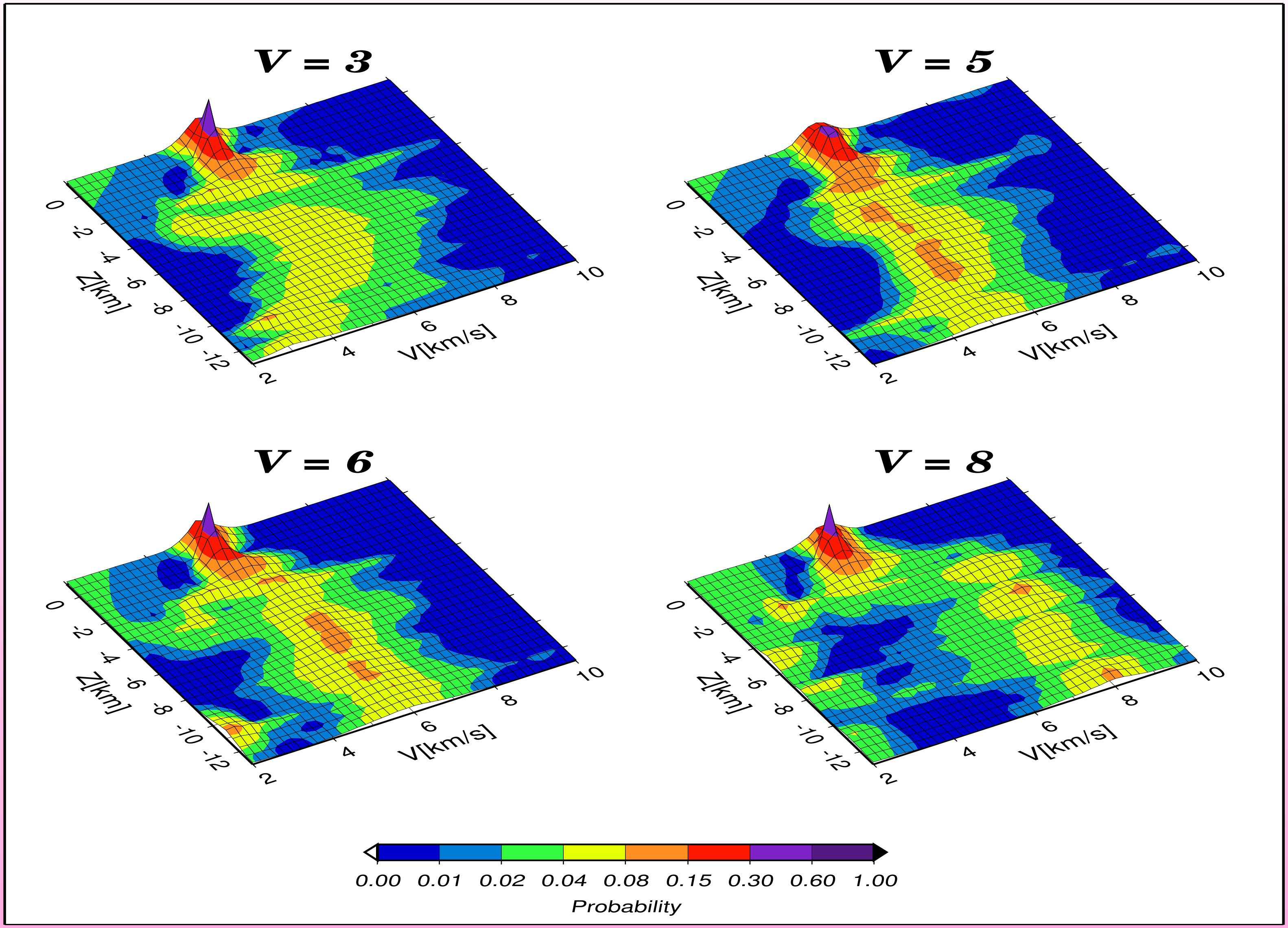


Case A



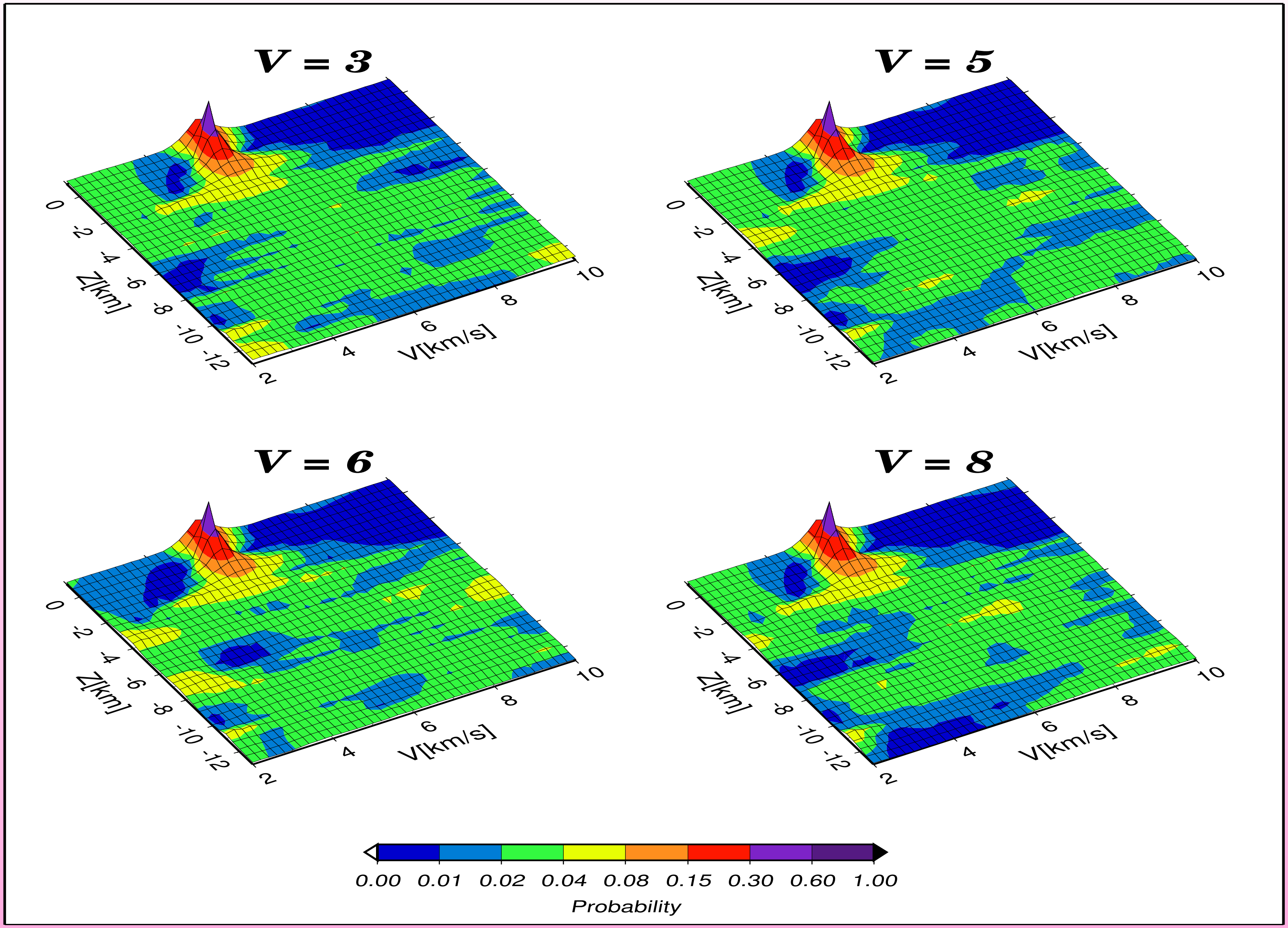
V = 3
V = 5
V = 6
V = 8
Z[km]
V[km/s]
0 -2 -4 -6 -8 -10 -12
2 4 6 8 10
0.00 0.01 0.02 0.04 0.08 0.15 0.30 0.60 1.00
Probability

Case B



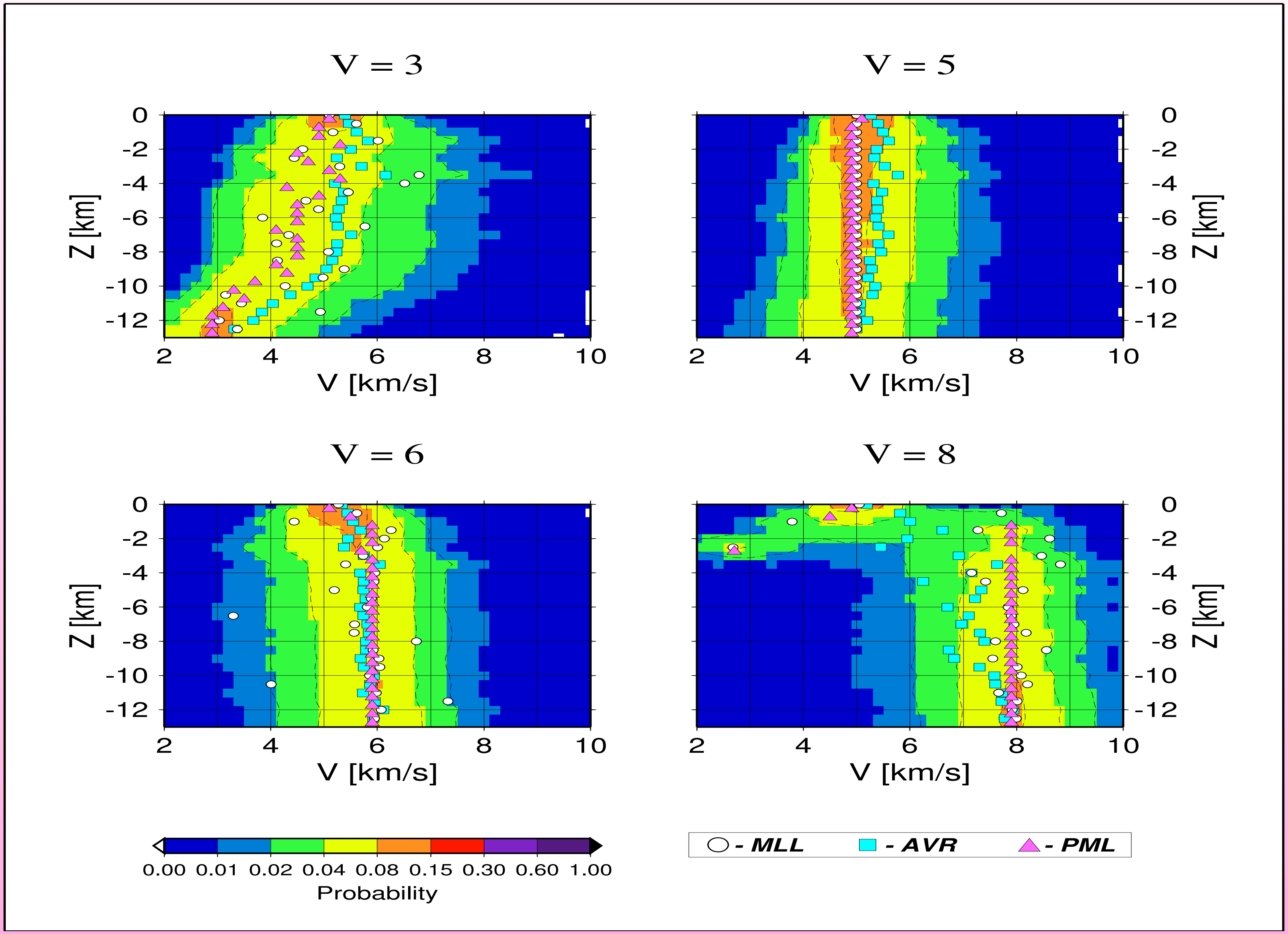
V = 3
V = 5
V = 6
V = 8
Z[km]
V[km/s]
0
-2
-4
-6
-8
-10
-12
2
4
6
8
10
0.00
0.01
0.02
0.04
0.08
0.15
0.30
0.60
1.00
Probability

Case C

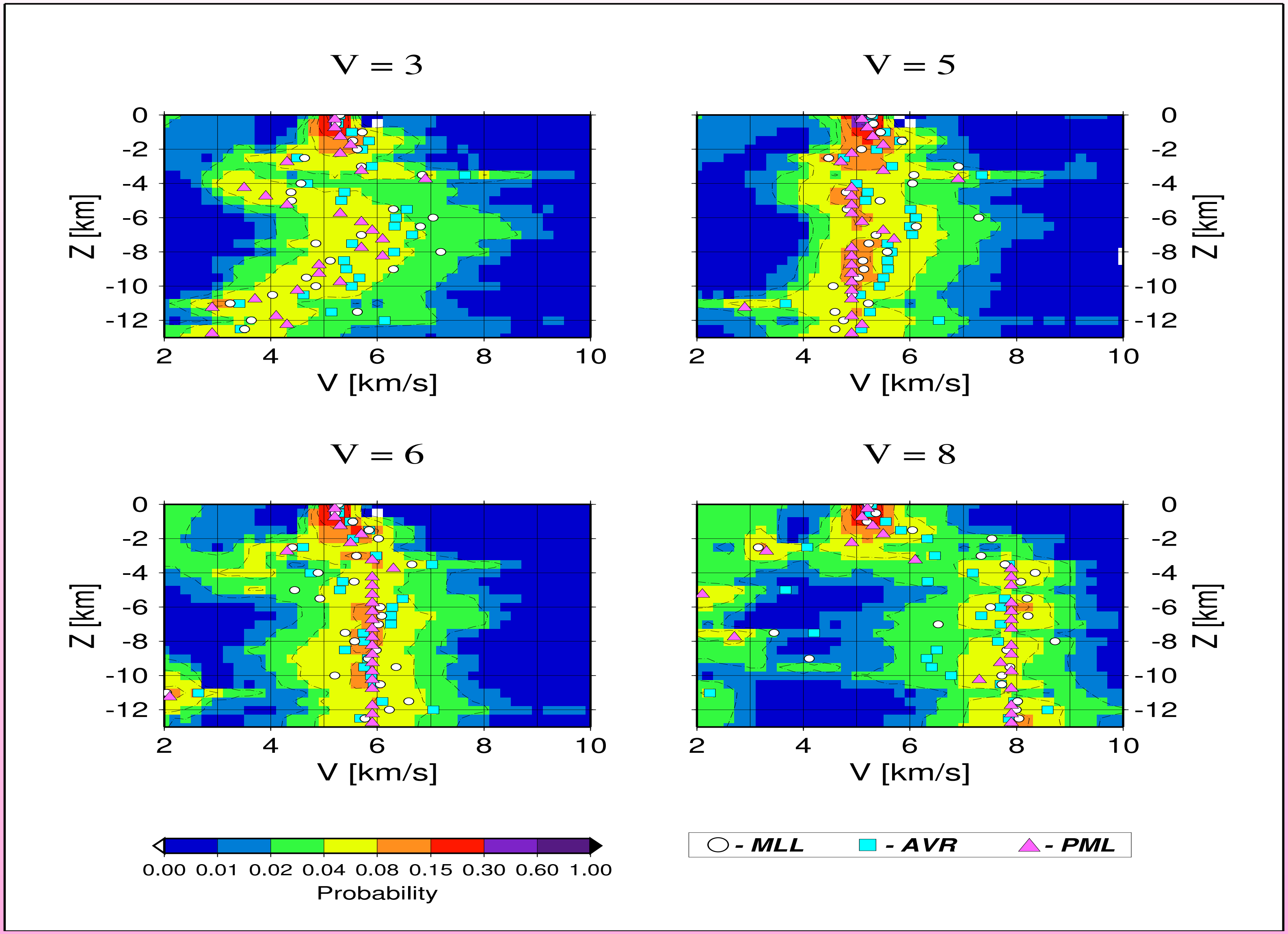


V = 3
V = 5
V = 6
V = 8
Z[km]
V[km/s]
0
-2
-4
-6
-8
-10
-12
2
4
6
8
10
0.00
0.01
0.02
0.04
0.08
0.15
0.30
0.60
1.00
Probability

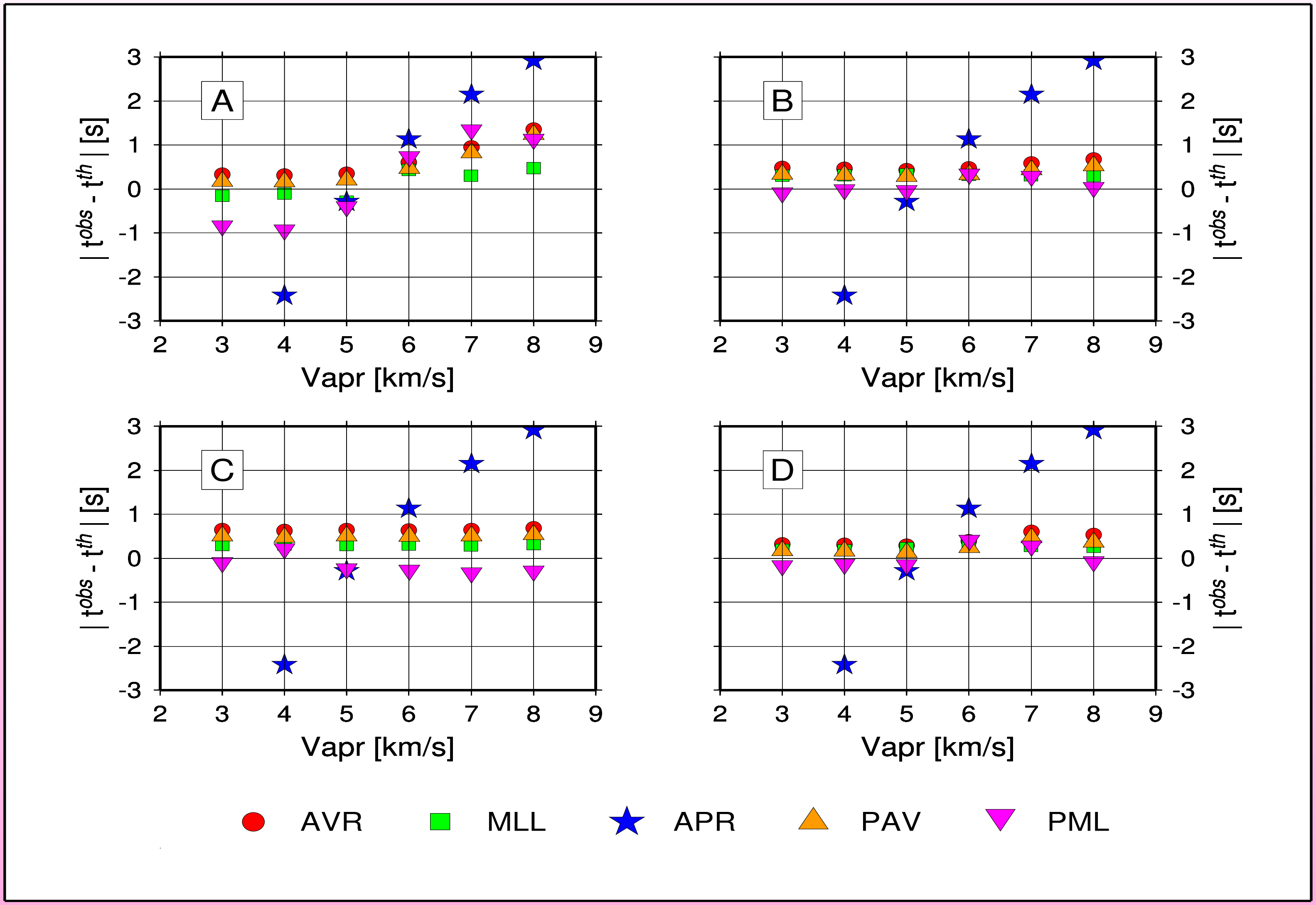
Case A


V = 3
V = 5
V = 6
V = 8
Z [km]
V [km/s]
0
-2
-4
-6
-8
-10
-12
2
4
6
8
10
0.00 0.01 0.02 0.04 0.08 0.15 0.30 0.60 1.00
Probability
- MLL
- AVR
- PML

Case B


V = 3
V = 5
V = 6
V = 8
Z [km]
V [km/s]
0
-2
-4
-6
-8
-10
-12
2
4
6
8
10
0.00 0.01 0.02 0.04 0.08 0.15 0.30 0.60 1.00
Probability
- MLL
- AVR
- PML

Average Residua



Waveform inversion

$$u(t) = G(t) * S(t)$$

- "Tomography": given S(t)

$$u(t) \longrightarrow G(t)$$

- Source tomography: given G(t)

$$u(t) \longrightarrow S(t)$$

Empirical Green Function

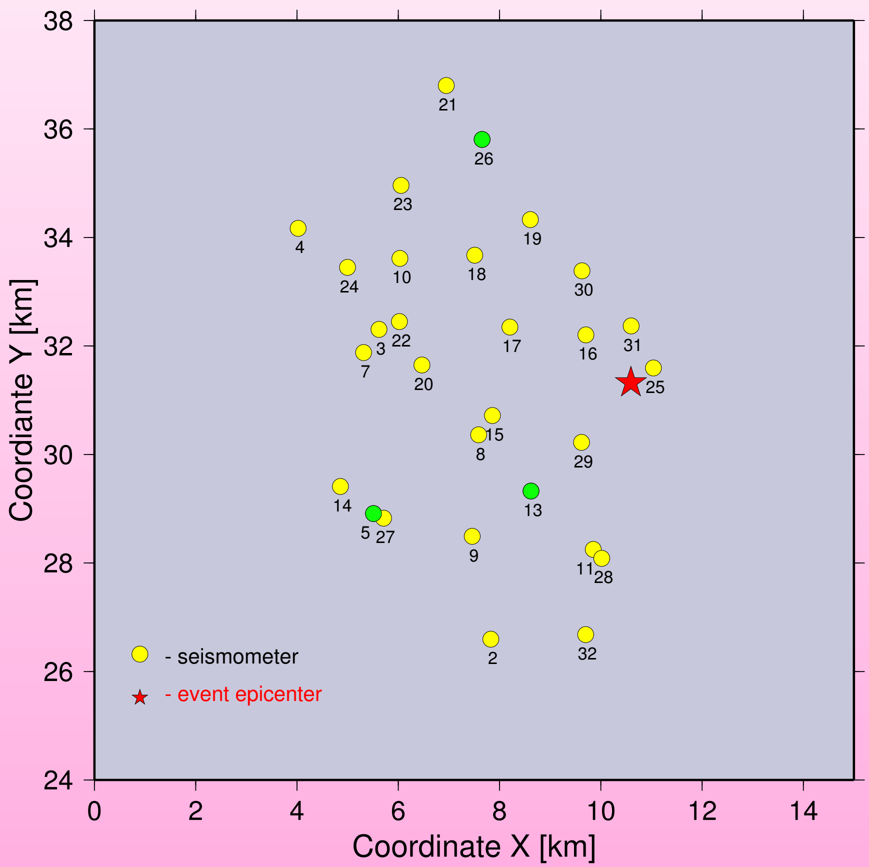
$$u(t) = G(t) * S(t)$$

$$G(t) = ???$$

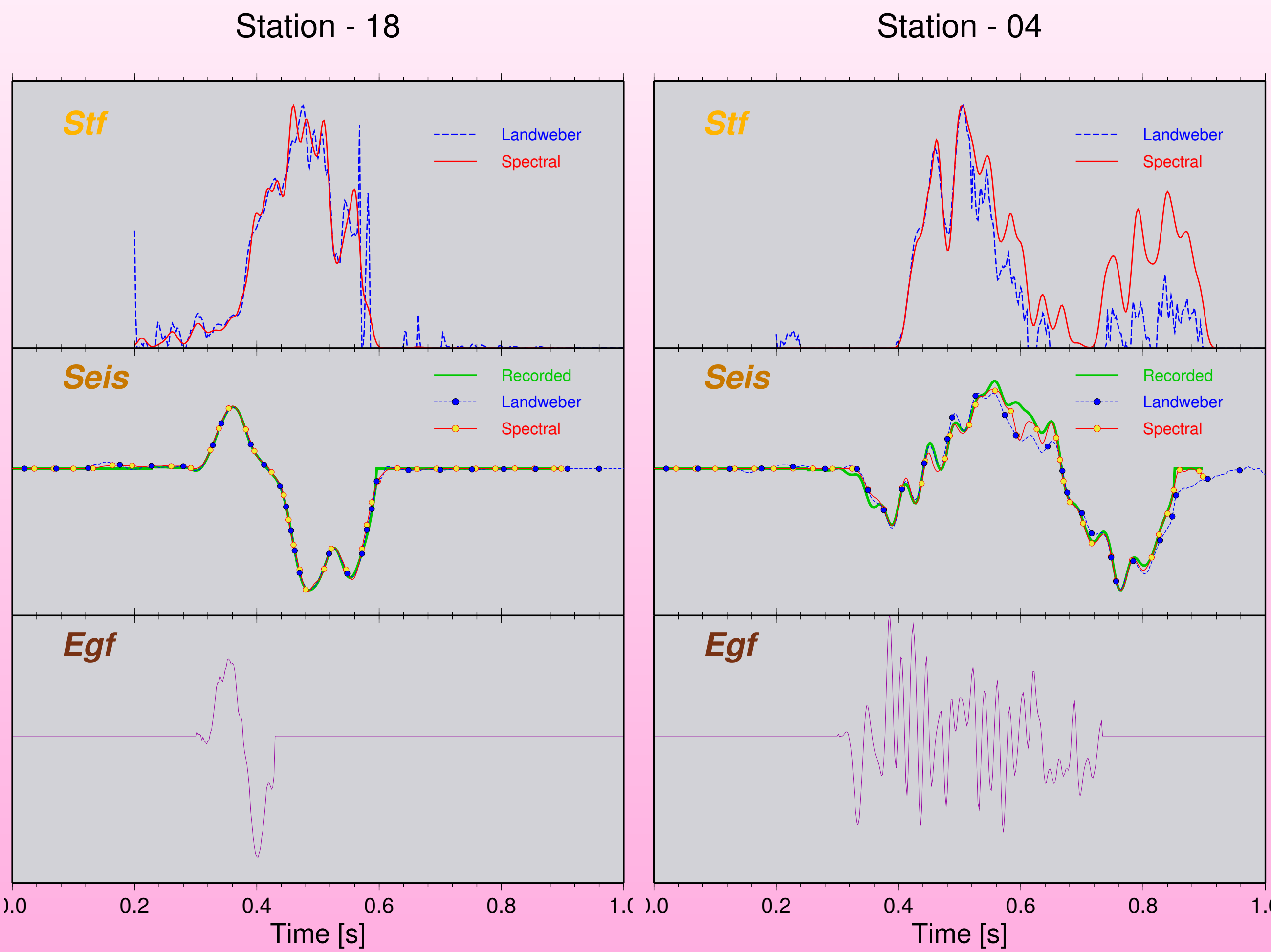
$$G(t) \approx u_1(t) \qquad f \leq f_{c2}$$

$$u_2(t) \approx u_1(t) * S(t)$$

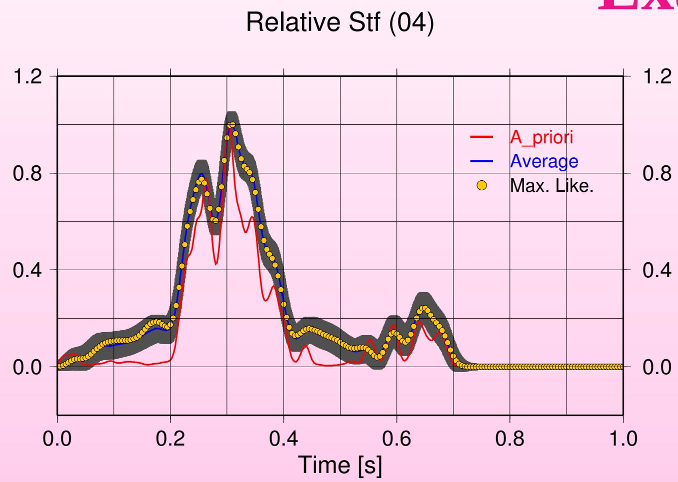
E3: Lubin mine seismic event - ($M_L \approx 3.1$), July, 1998

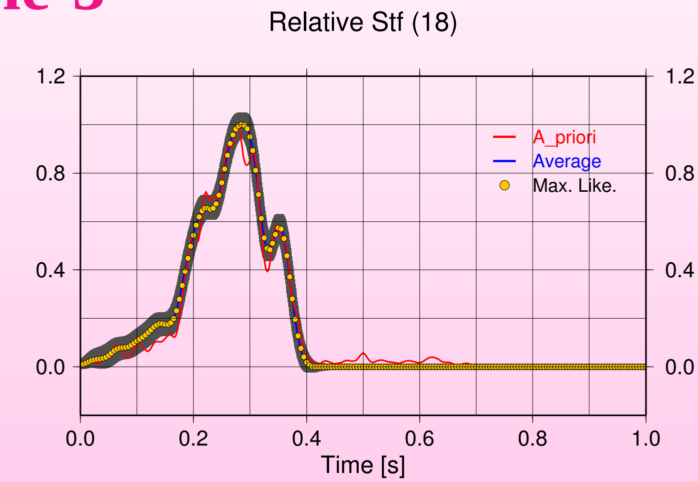


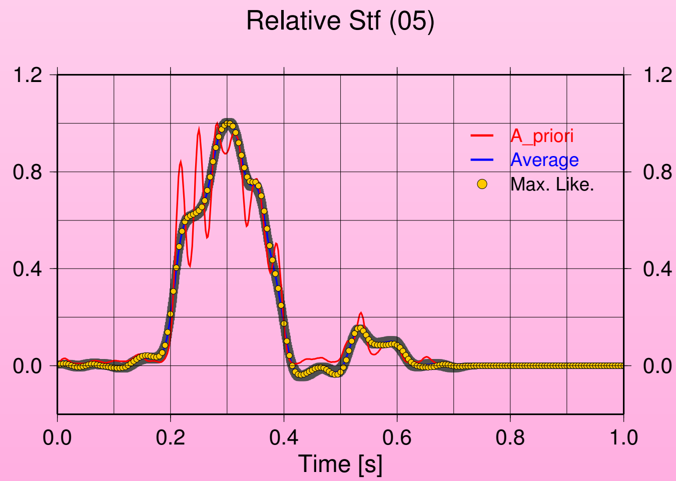
Example-3

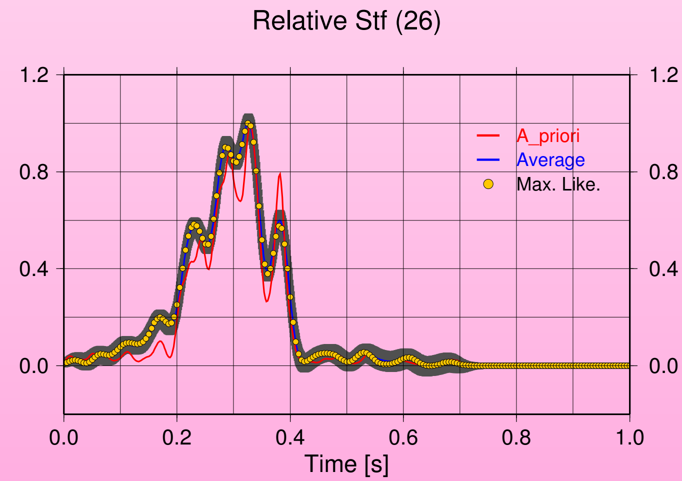

Station - 18
Station - 04
Stf
Landweber
Spectral
Seis
Recorded
Landweber
Spectral
Egf
0.0
0.2
0.4
0.6
0.8
1.0
Time [s]
Stf
Landweber
Spectral
Seis
Recorded
Landweber
Spectral
Egf
0.0
0.2
0.4
0.6
0.8
1.0
Time [s]

Example-3

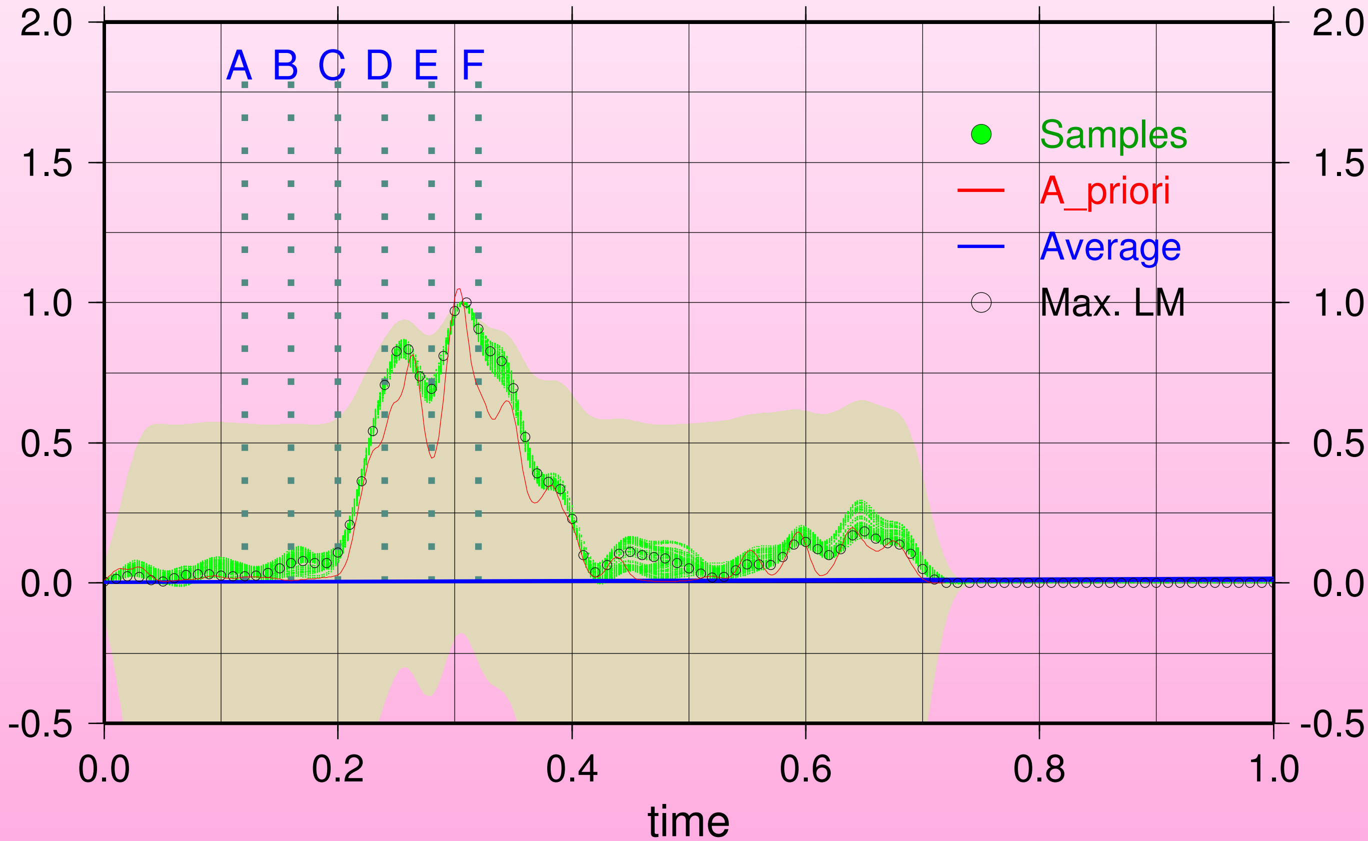

Relative Stf (04)
A_priori
Average
Max. Like.
Time [s]
0.0
0.2
0.4
0.6
0.8
1.0
1.2


Relative Stf (18)
A_priori
Average
Max. Like.
Time [s]
0.0
0.2
0.4
0.6
0.8
1.0
1.2

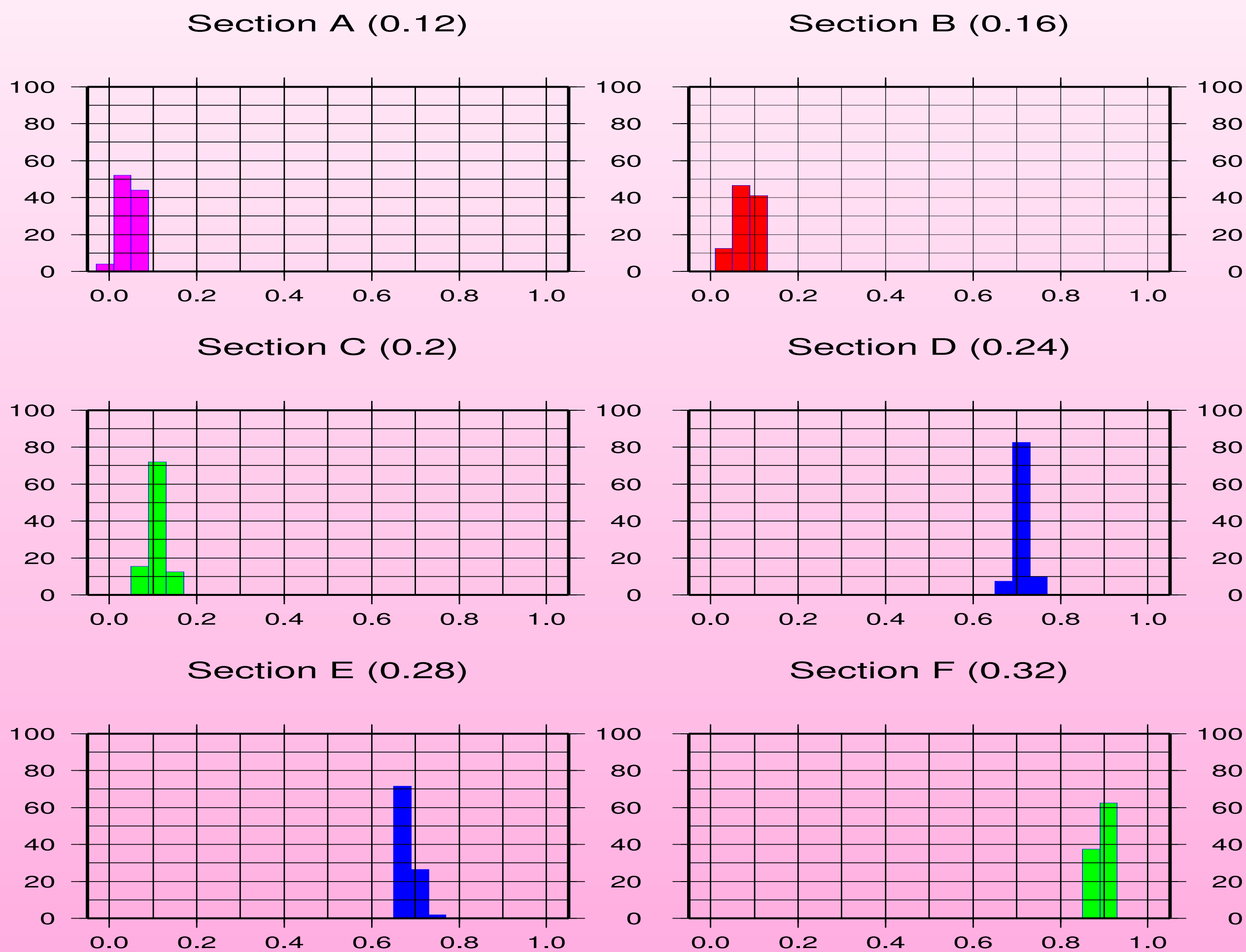

Relative Stf (05)
A_priori
Average
Max. Like.
Time [s]
0.0
0.2
0.4
0.6
0.8
1.0
1.2


Relative Stf (26)
A_priori
Average
Max. Like.
Time [s]
0.0
0.2
0.4
0.6
0.8
1.0
1.2

STF estimation - channel 04



Marginal distributions



Features of the technique

- Stochastic (Monte Carlo) method
- Full error analysis
- Variety of statistical estimators
- Analysis of the *a priori* constraints
- Possible advanced resolution analysis