Robust full waveform inversion:
In which domain should we measure the misfit?

F.J Herrmann, T. van Leeuwen, A.Y. Aravkin - UBC
H. Calandra - Total



SLIM
University of British Columbia

Overview

- MAP estimation
- Outliers
- Students T
- Results
- Conclusions

SLIM

MAP estimation

measurement model:

$$\mathbf{d}_i = F_i(\mathbf{m}) + \mathbf{n}_i$$

posterior likelihood:

$$\pi_{\text{post}}(\mathbf{m}) \sim \prod_{i=1}^{K} \pi_{\text{noise}}(F_i(\mathbf{m}) - \mathbf{d}_i)\pi_{\mathsf{prior}}(\mathbf{m})$$

MAP estimation

Maximization of the likelihood

$$\max_{\mathbf{m}} \pi_{\text{post}}(\mathbf{m})$$

is equivalent to

$$\min_{\mathbf{m}} -\log\left(\pi_{\text{post}}(\mathbf{m})\right)$$

MAP estimation

For Gaussian noise we have

$$\pi_{\text{noise}}(\mathbf{r}) \sim \exp\left(-||\mathbf{r}||_2^2\right)$$

which leads to the usual least-squares formulation

$$\min_{\mathbf{m}} \sum_i ||F_i(\mathbf{m}) - \mathbf{d}_i||_2^2$$

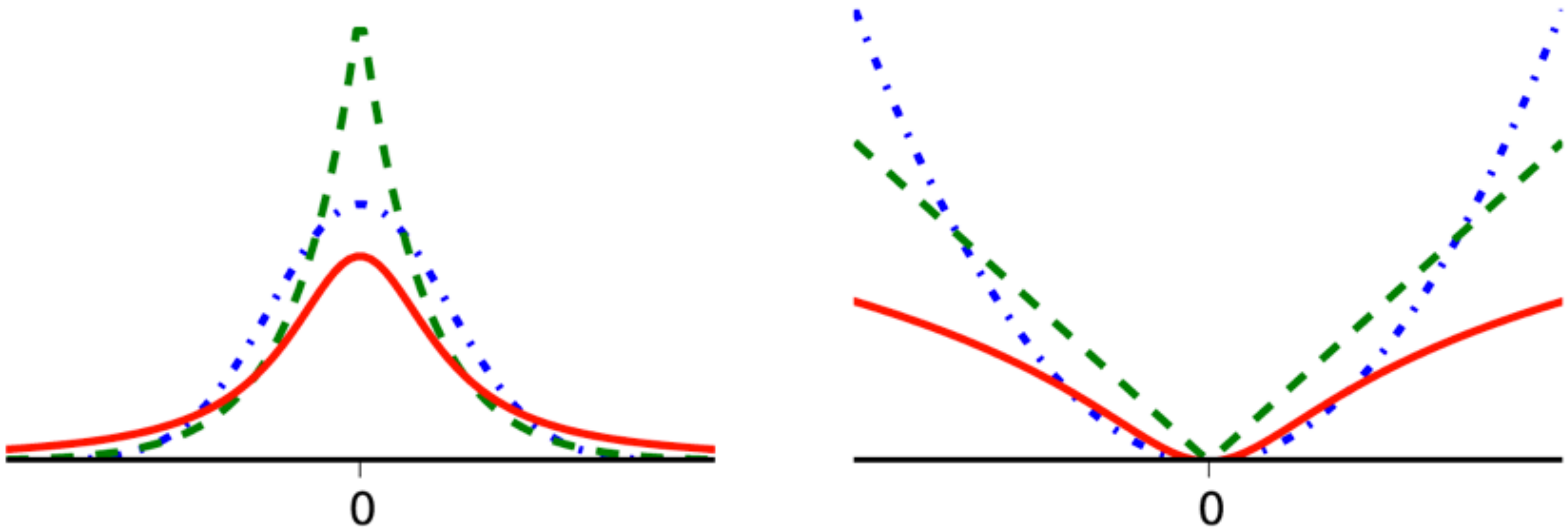
MAP estimation

The use of alternative penalties can be interpreted as using a different noise model

$$\min_{\mathbf{m}} \sum_i \rho\left(F_i(\mathbf{m}) - \mathbf{d}_i\right)$$

MAP estimation

densities & penalties

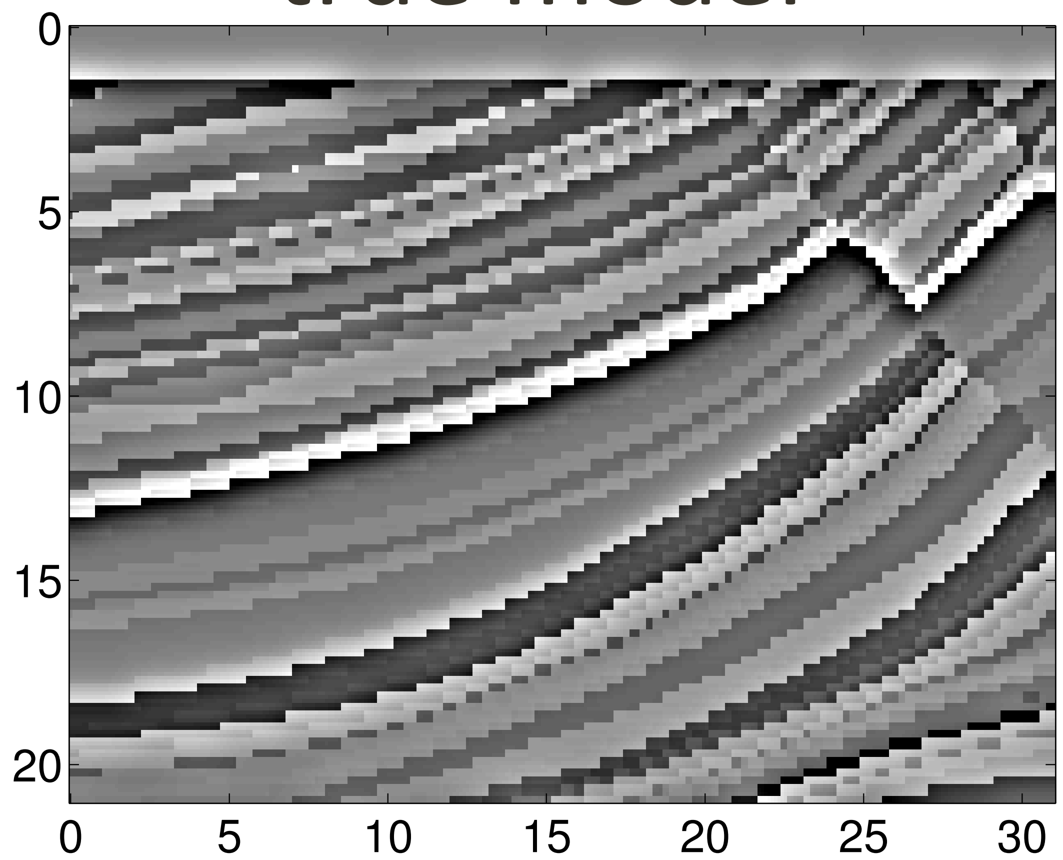


Gaussian, Laplace and Students T

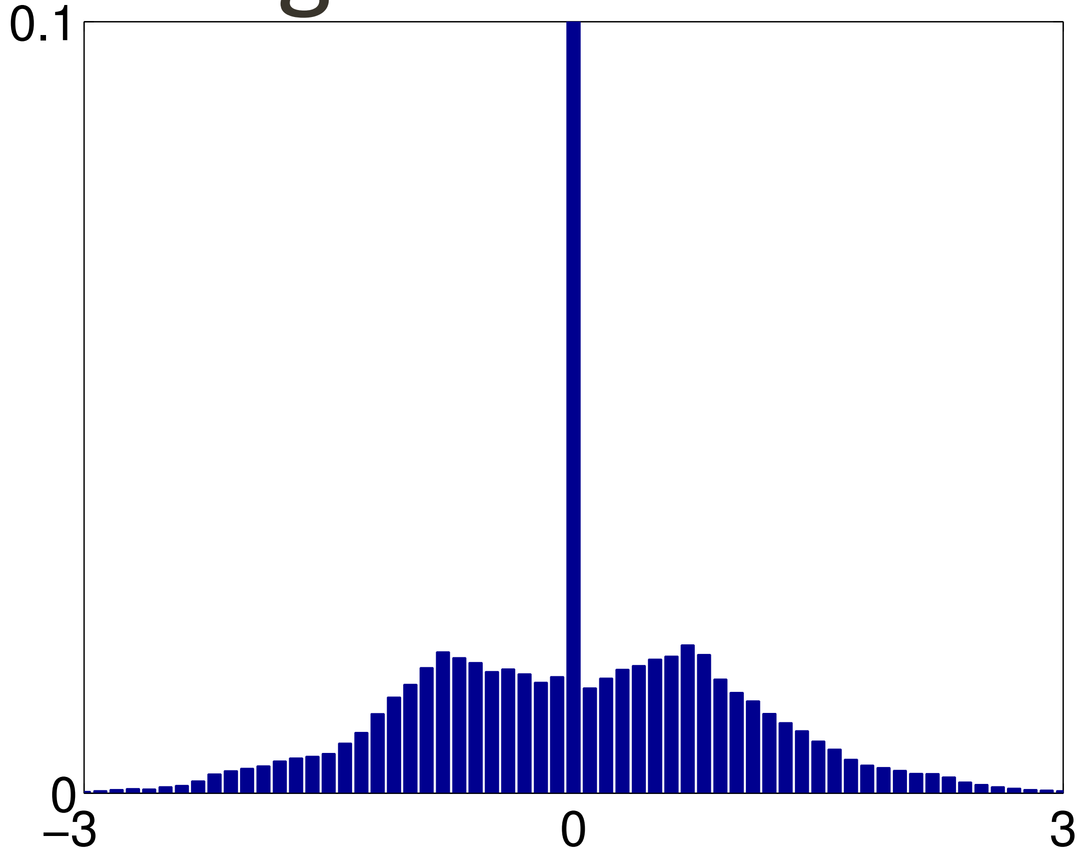
MAP estimation

data with 50% “bad traces”

true model



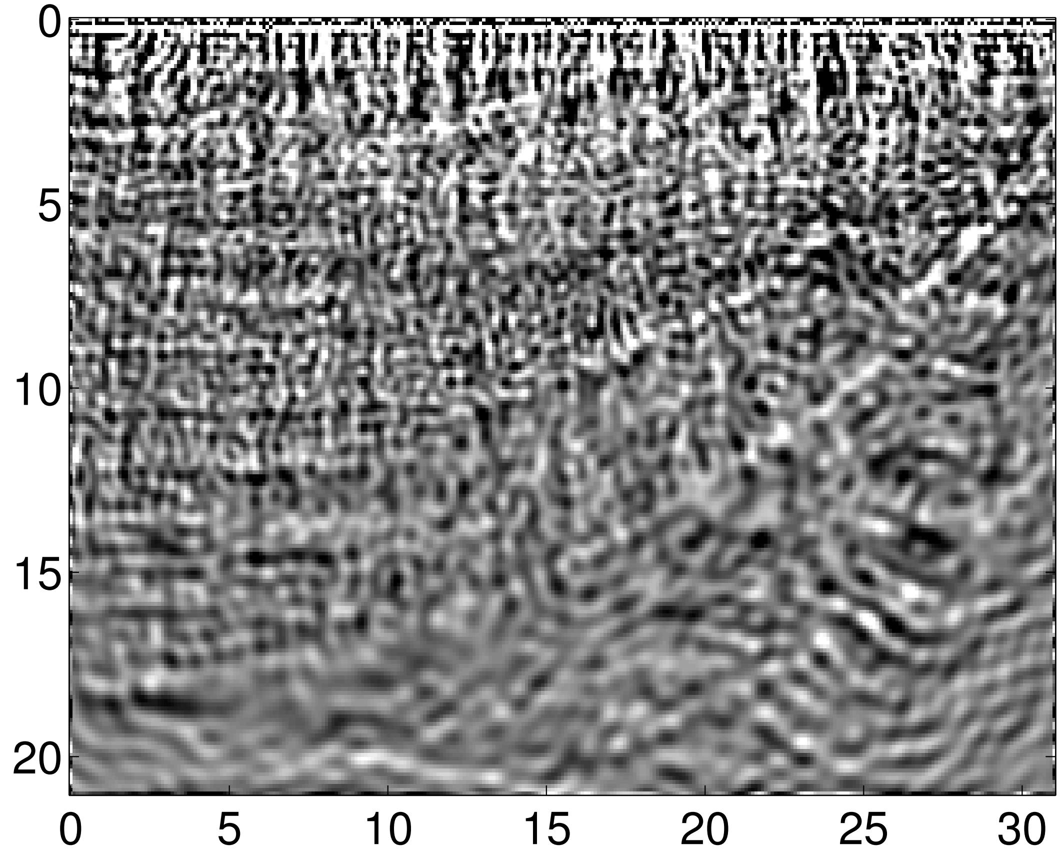
histogram of true noise



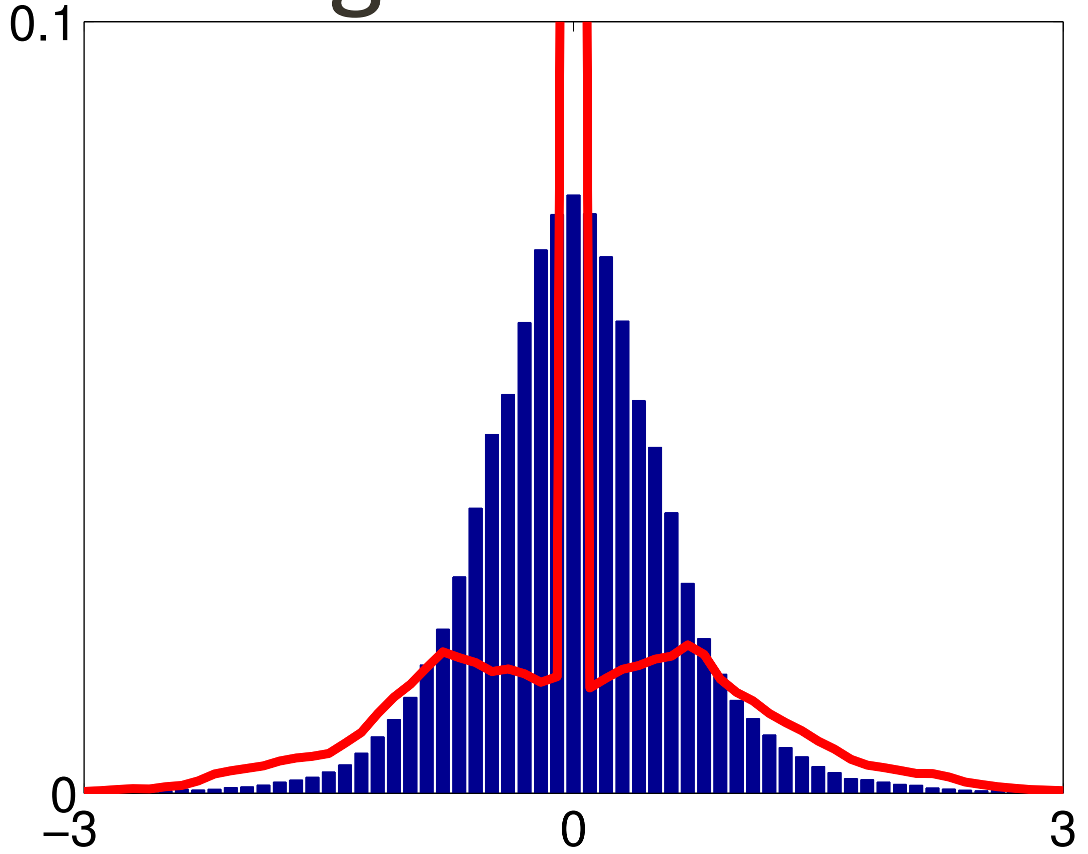
MAP estimation

least-squares penalty

recovered model



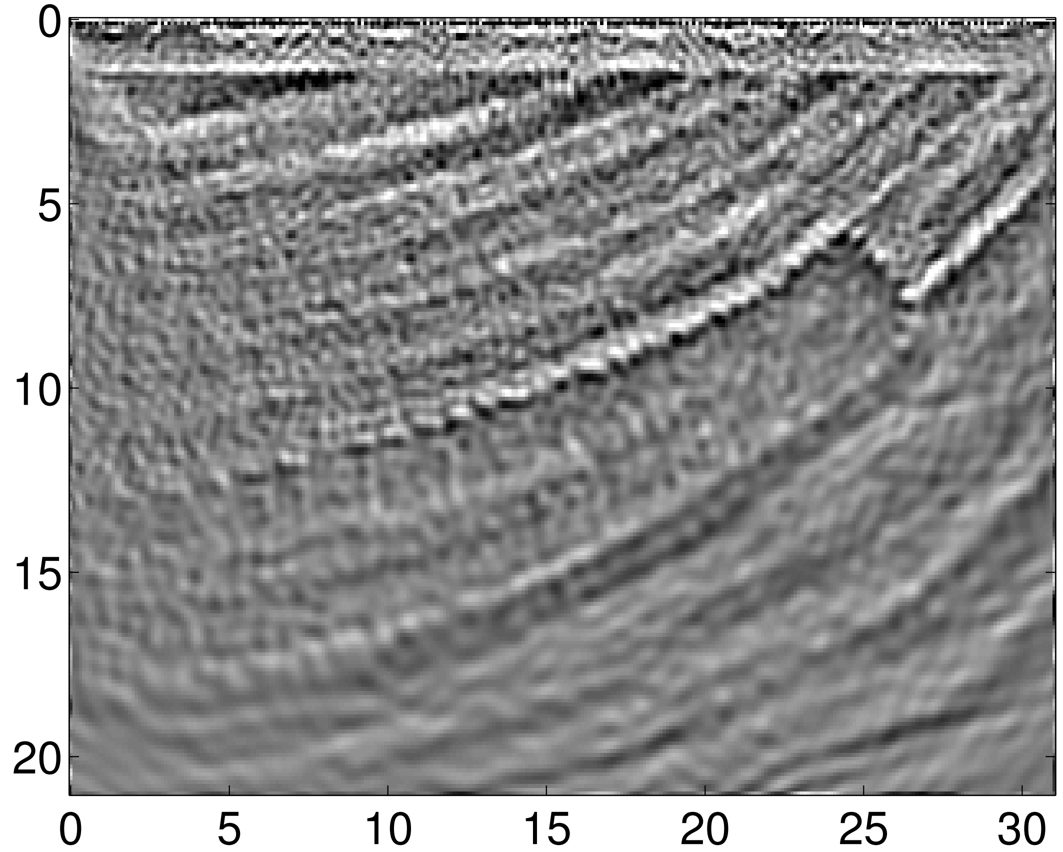
histogram of residual



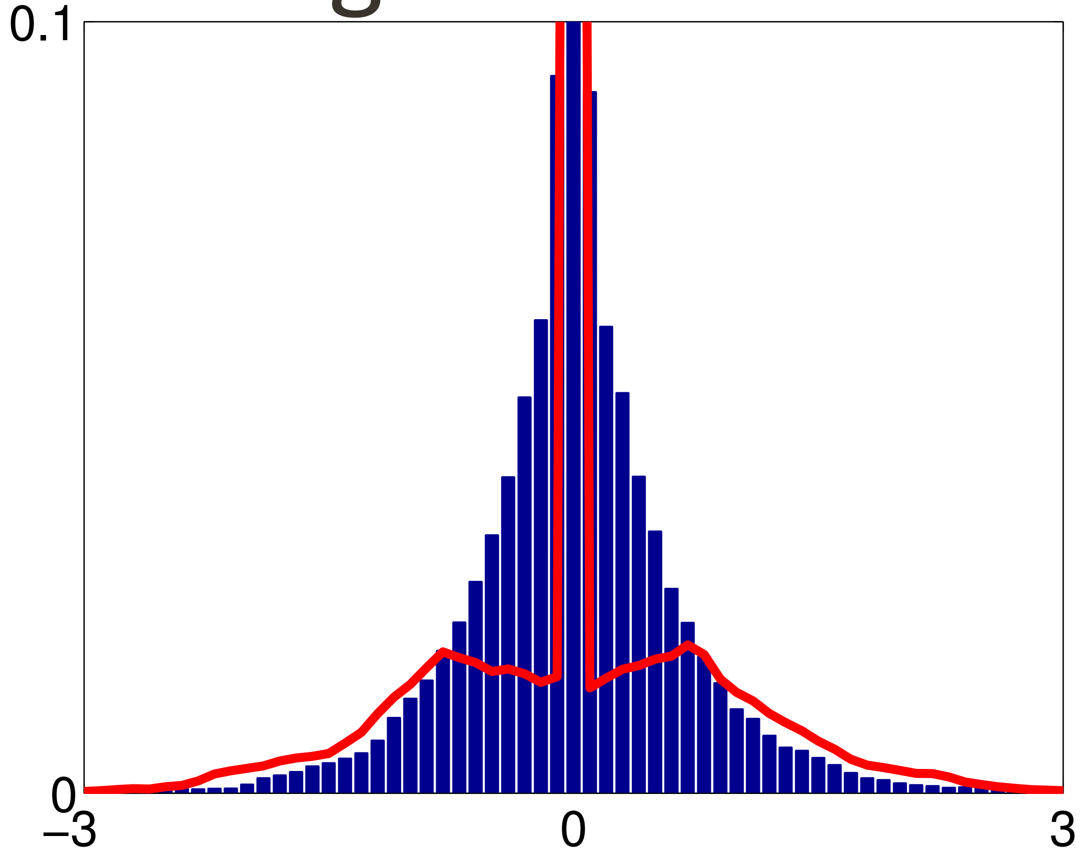
MAP estimation

Huber penalty

recovered model



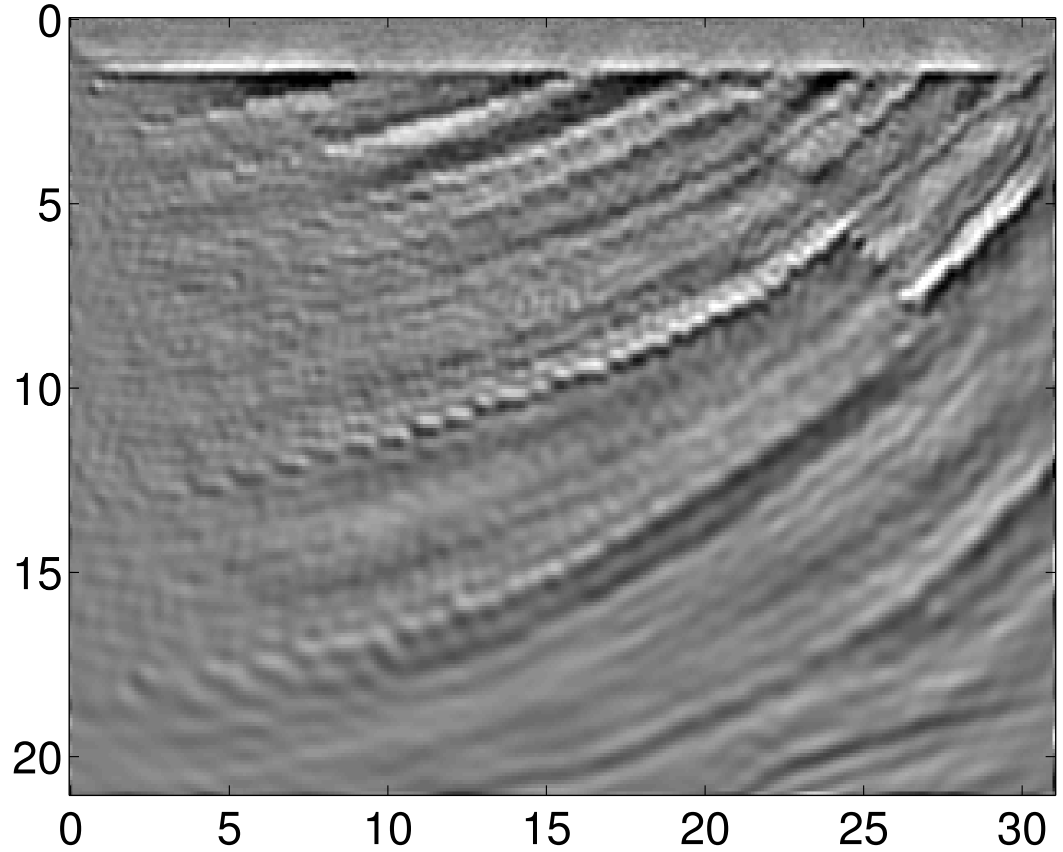
histogram of residual



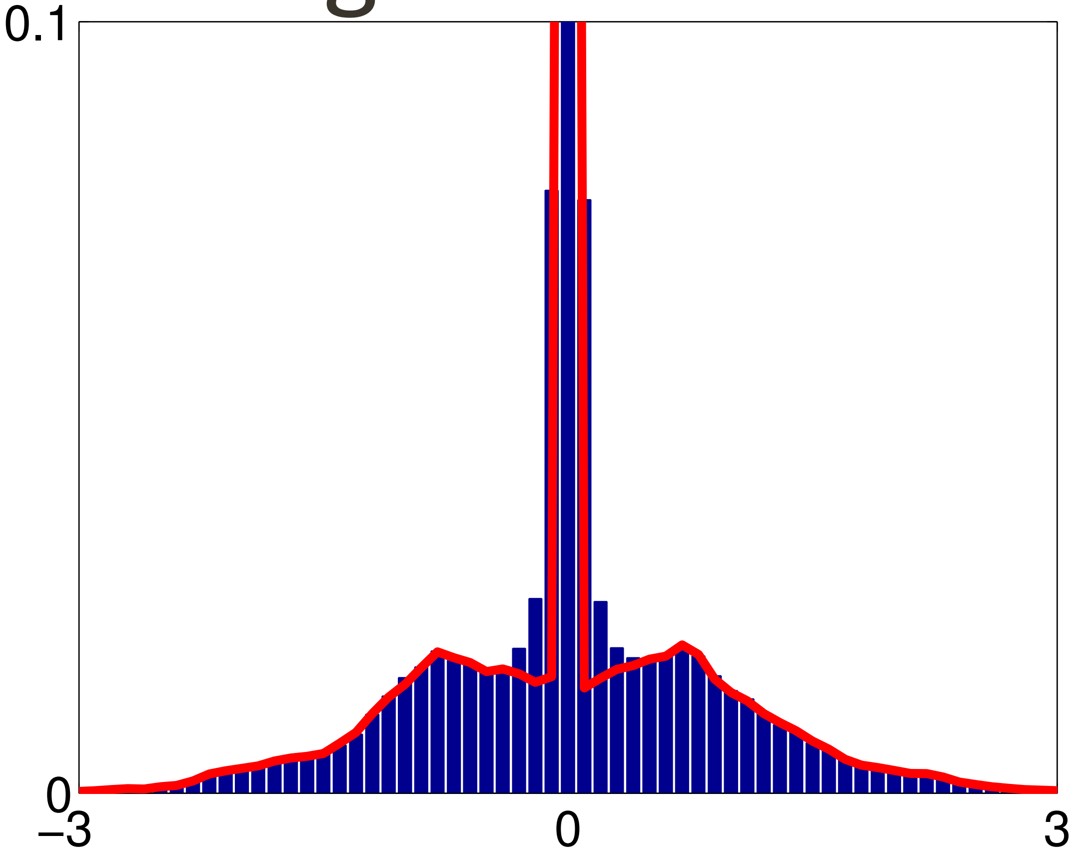
MAP estimation

Students T penalty

recovered model



histogram of residual

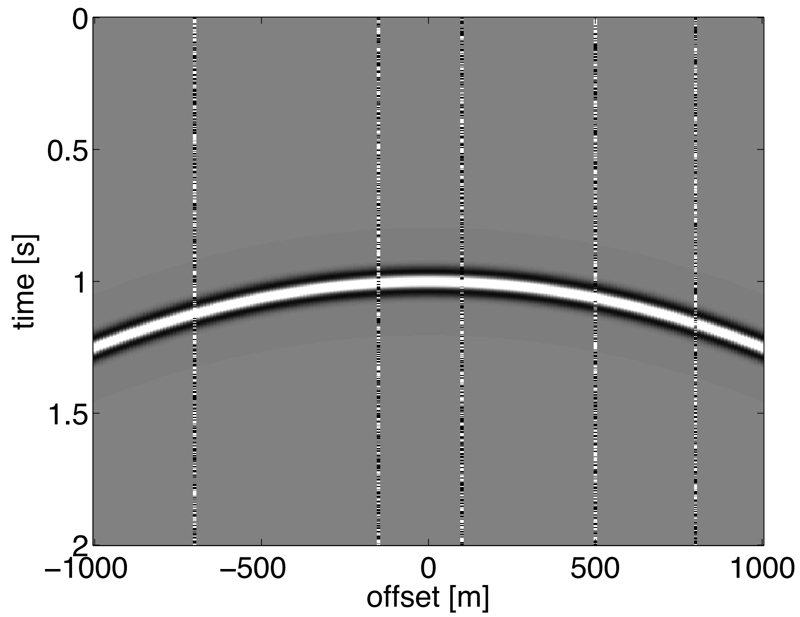


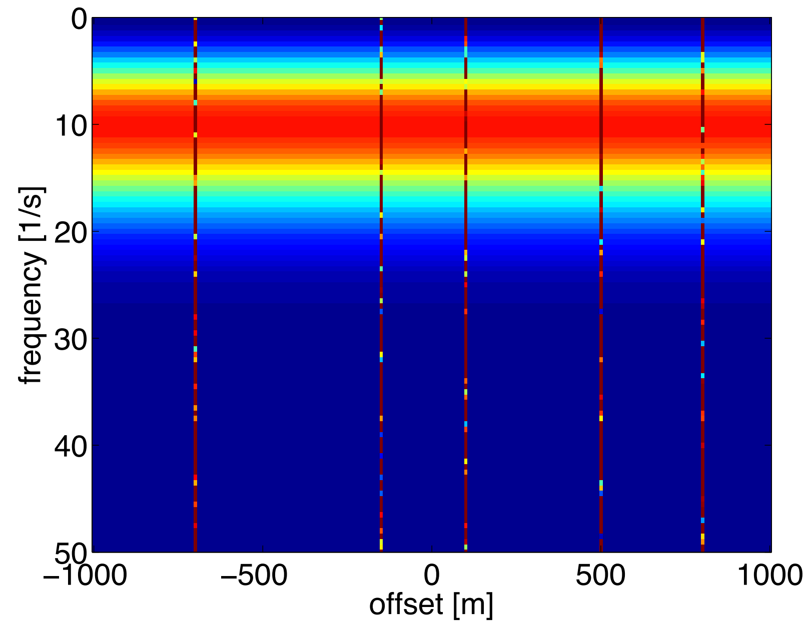
MAP estimation

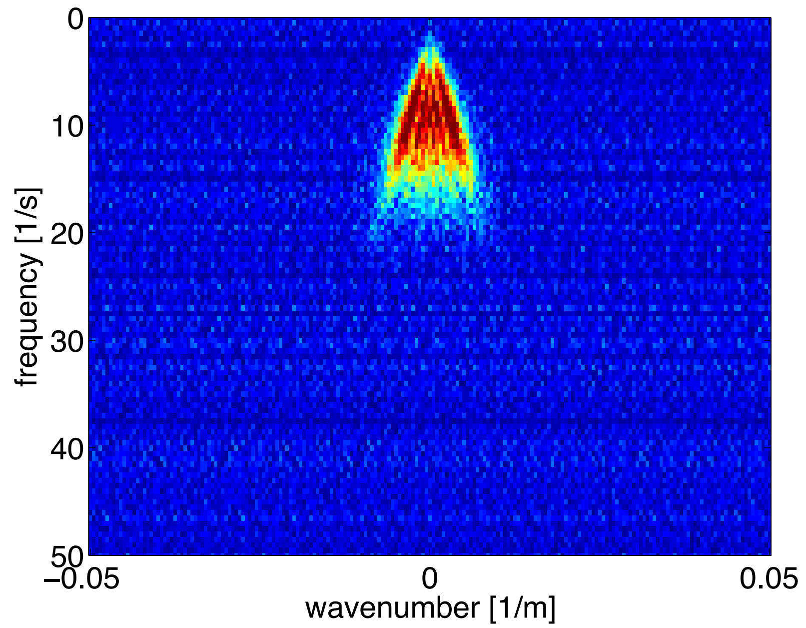
- Noise does *not* come from Students T distribution
- Use of Students T penalty may still be beneficial
- Noise has to be *spiky*

Outliers

What *is* an outlier?



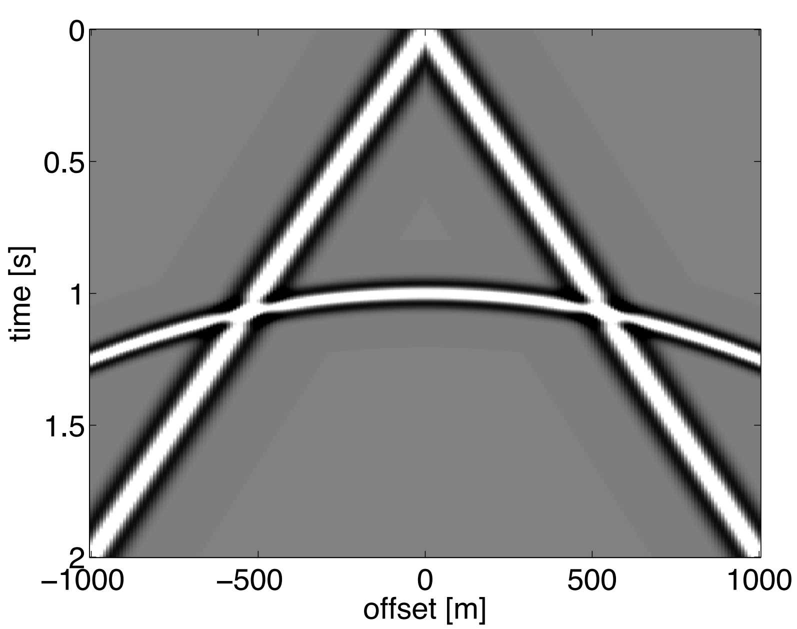




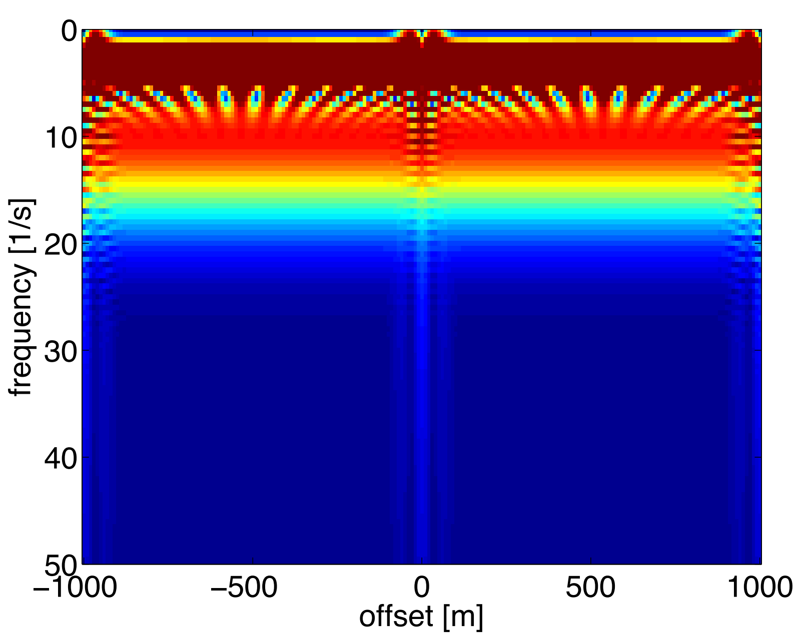
t,x ✓ f,x ✓ f,k ✗

Outliers

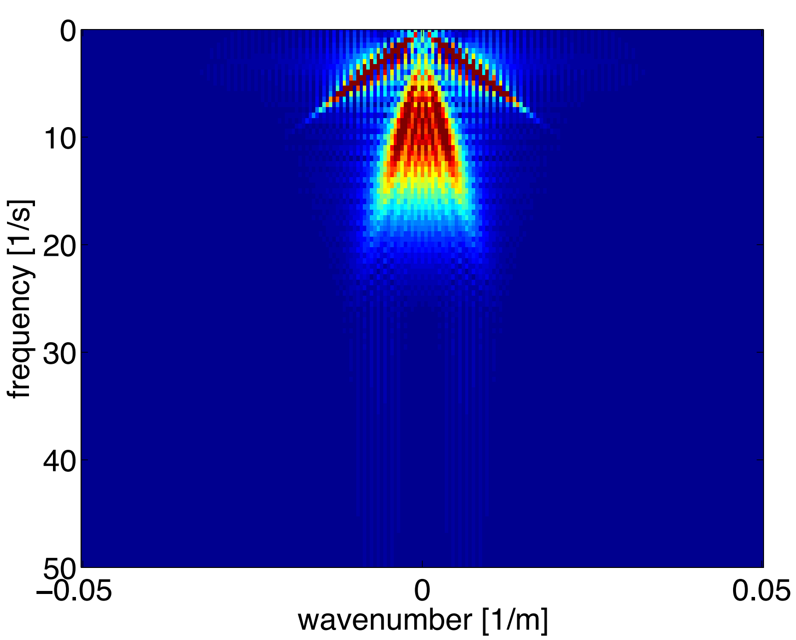
What *is* an outlier?



t,x ✓



f,x ✗



f,k ✓

Outliers

Measure the misfit in a domain that *sparsifies* the noise

$$\min_{\mathbf{m}} \sum_i \rho \left(B \left(F_i(\mathbf{m}) - \mathbf{d}_i \right) \right)$$

e.g., Fourier, Radon, Curvelets,…

Students T

The penalty is given by

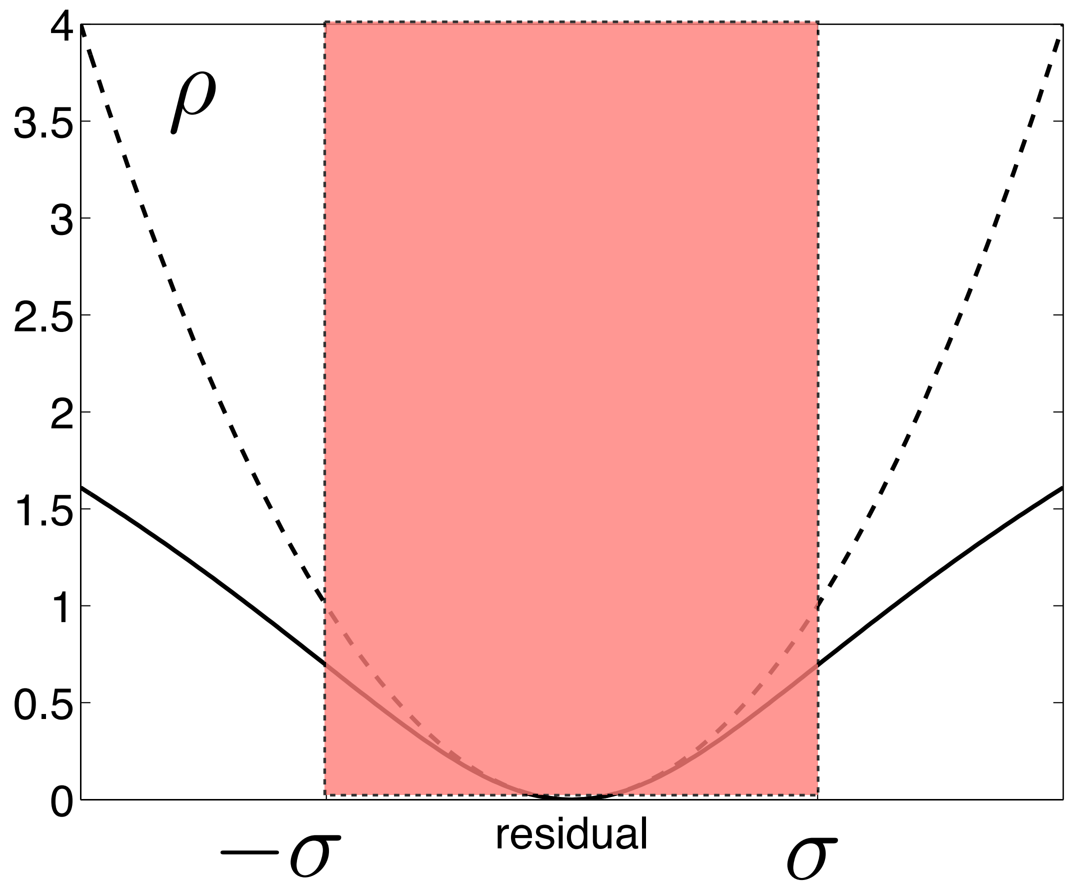
$$\rho(\mathbf{r}) = \sum_j \log(1 + |r_j|^2/\sigma^2)$$

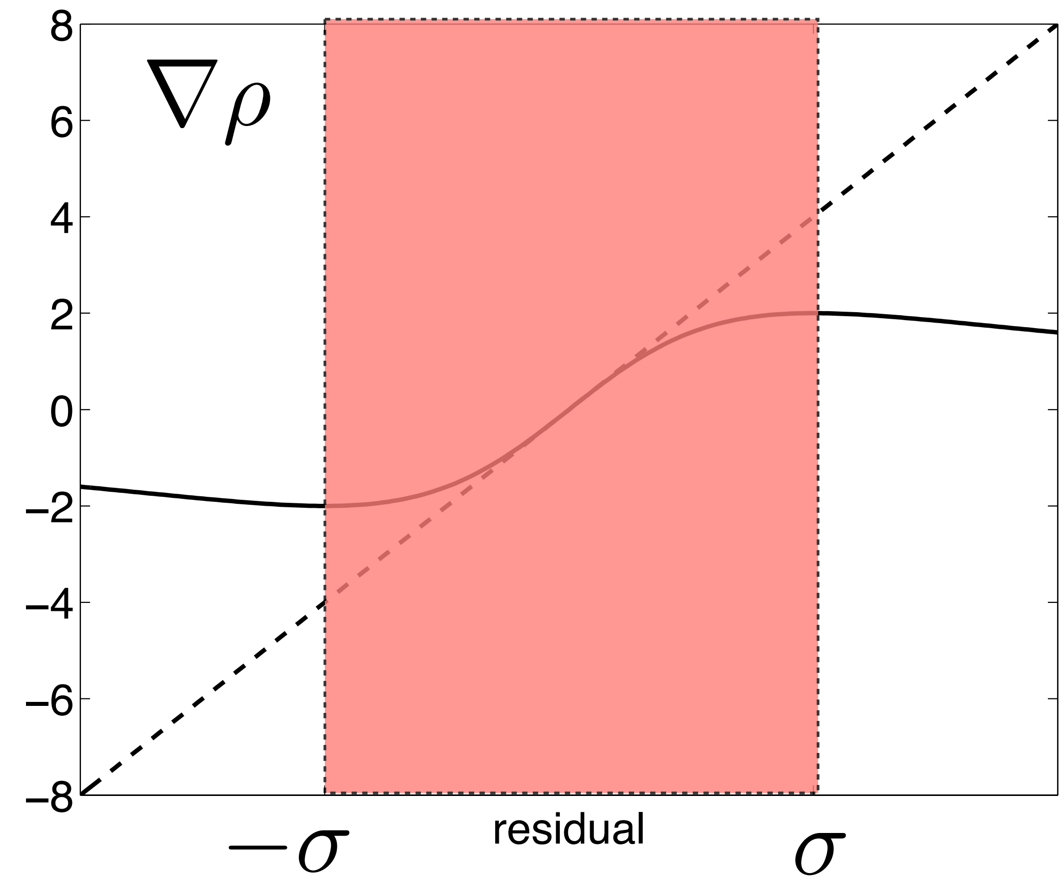
where σ is a scale parameter. The corresponding adjoint source is given by

$$(\nabla\rho)_j = \frac{2r_j}{|r_j|^2 + \sigma^2}$$

Students T

Scale parameter is used to separate outliers from good data





Students T

- *scale* parameter controls which residuals are ignored
- similar to a *weighted* least-squares approach
- how should we *choose* σ ?
- what about *source* estimation?

Source estimation

Use *variable* projection approach on

$$\min_{\mathbf{m},\mathbf{w}} \sum_i \rho\left(B\left(w_i F_i(\mathbf{m}) - \mathbf{d}_i\right)\right)$$

solve *source*-weights as

$$\min_{w_i} \rho\left(B\left(w_i F_i(\mathbf{m}) - \mathbf{d}_i\right)\right)$$

Auto-tuning

Extended Students T penalty:

$$\rho_\sigma(\mathbf{r}) = -N\log\left(\frac{\Gamma(\frac{\sigma^2+1}{2})}{\Gamma(\frac{\sigma^2}{2})\sqrt{\pi\sigma^2}}\right) + \frac{\sigma^2+1}{2}\sum_{j=1}^{N}\log(1+r_j^2/\sigma^2)$$

find *optimal* σ for a given *residual* by solving

$$\min_\sigma \rho_\sigma(\mathbf{r})$$

Workflow

1. Forward modeling $\mathbf{d}_i^{\text{pred}} = F_i(\mathbf{m}_k)$
2. Estimate source weight (scalar optimization)
3. Compute residual $\mathbf{r}_i = w_i \mathbf{d}_i^{\text{pred}} - \mathbf{d}_i$

4. Estimate scale (scalar optimization)
5. Compute adjoint source $\widetilde{\mathbf{r}_i} = B^* w_i^* \nabla \rho(B\mathbf{r}_i)$

7. Compute gradient $\mathbf{g} = \sum_i \nabla F_i(\mathbf{m}_k)^* \widetilde{\mathbf{r}_i}$

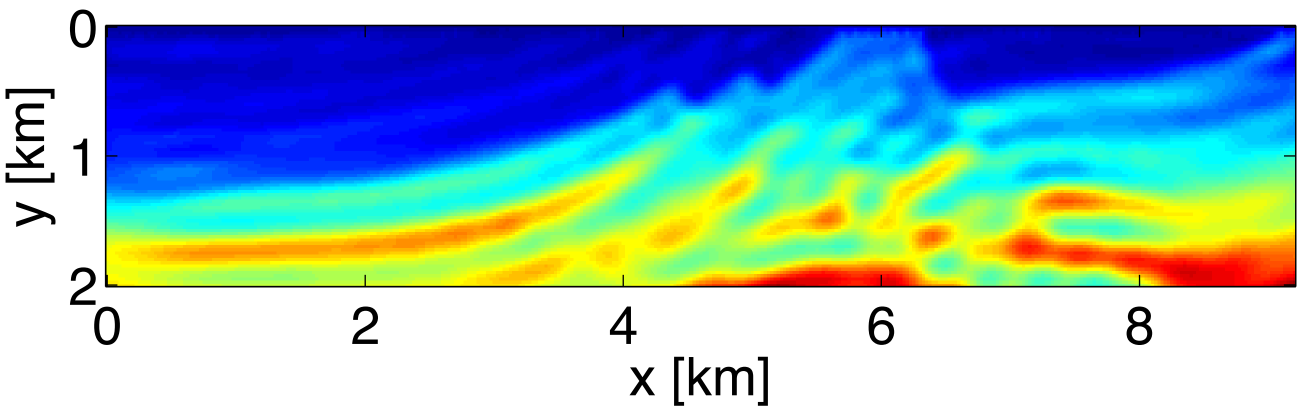
9. update $\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda \mathbf{g}$

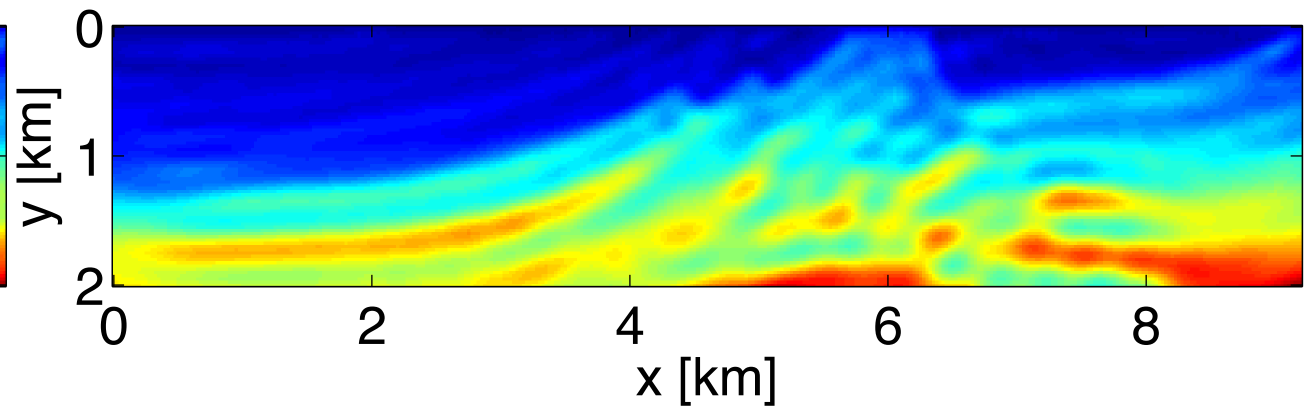
Results 1

- Marmousi model with *periodic* noise.
- inversion of *single* frequency (4 Hz) with 20 iterations
- Misfit measured in *(f,x)* or *(f,k).*

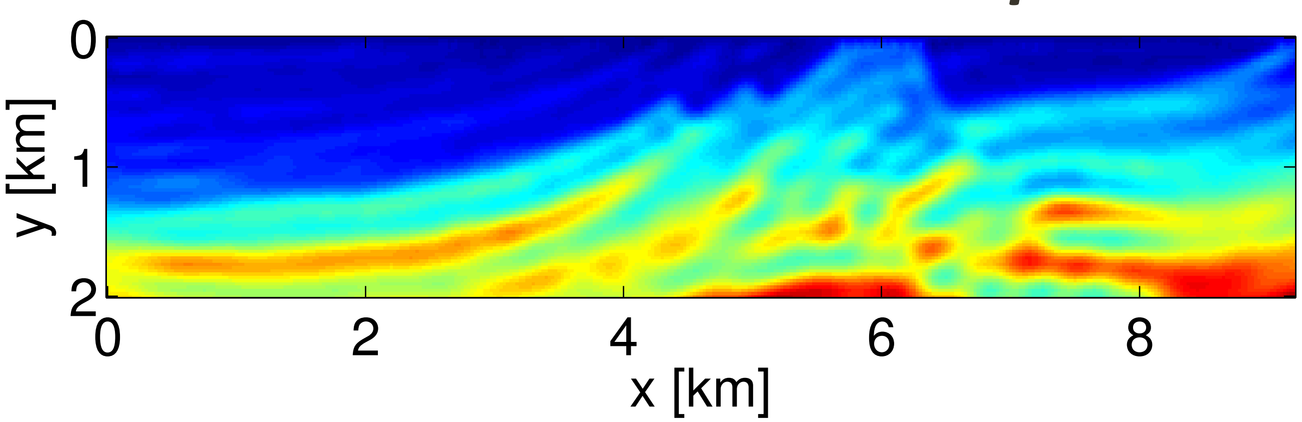
Results 1

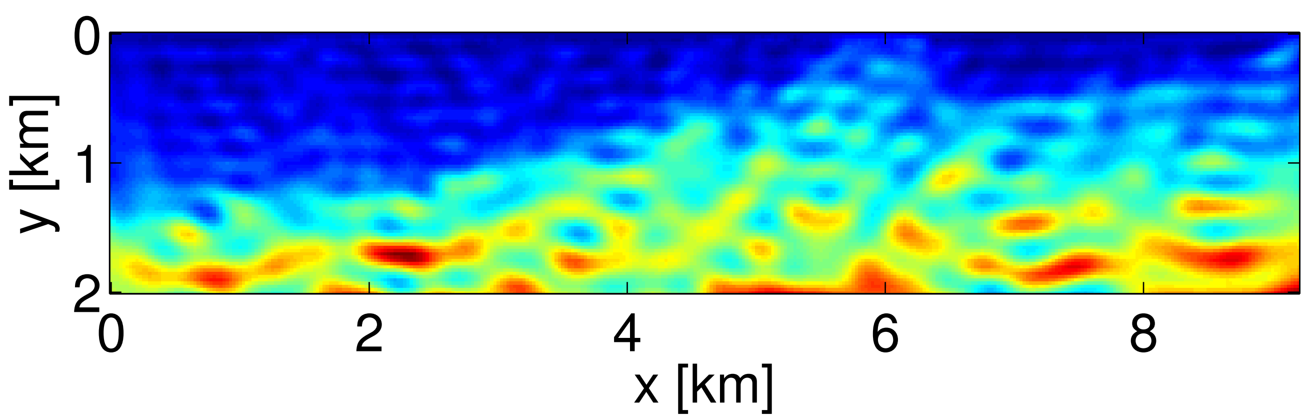
no noise:





periodic noise:



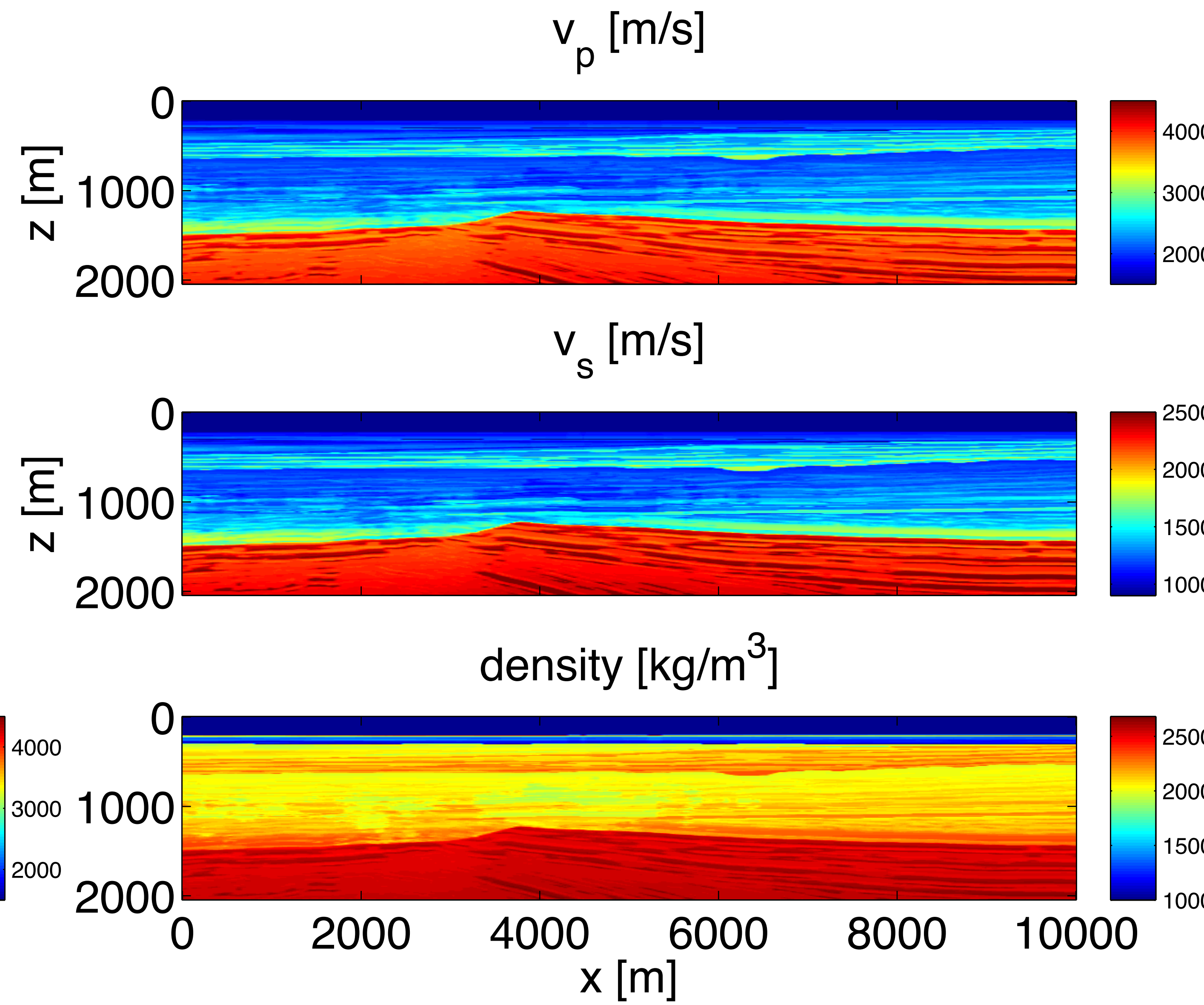


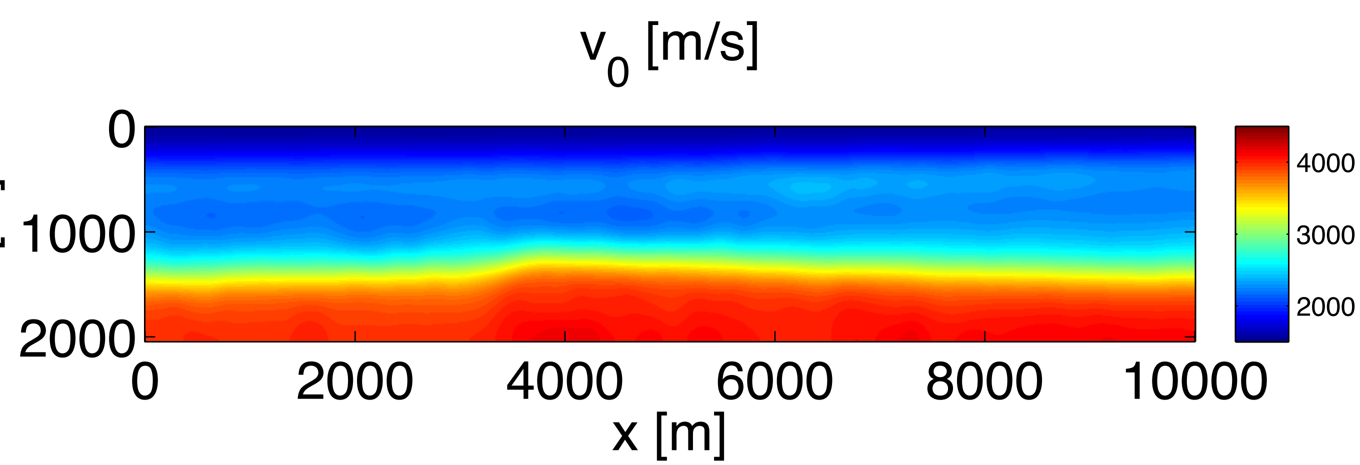
f-k

f-x

Results 2

Acoustic inversion

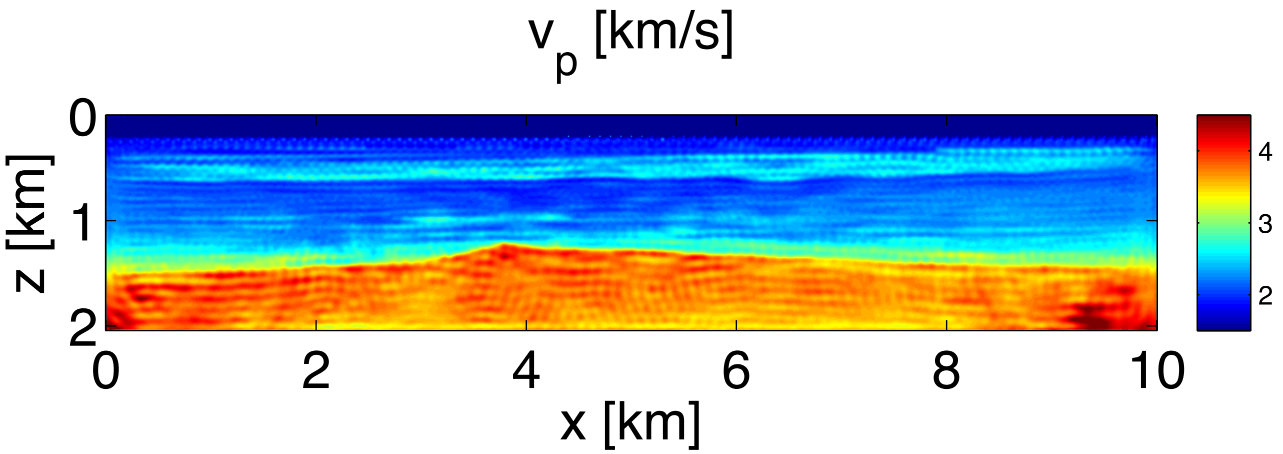




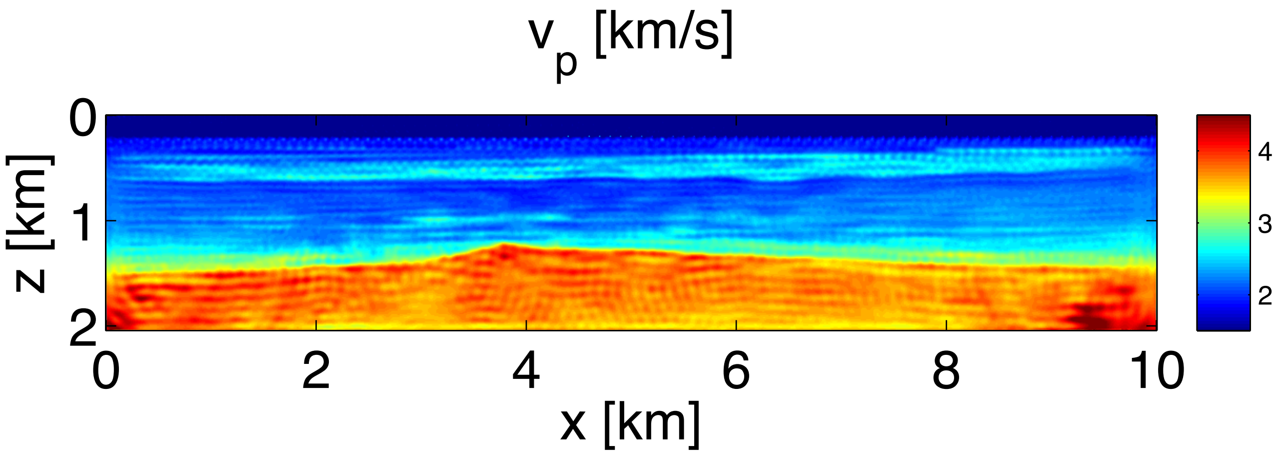
Results 2

Variable density data, no noise

least-squares



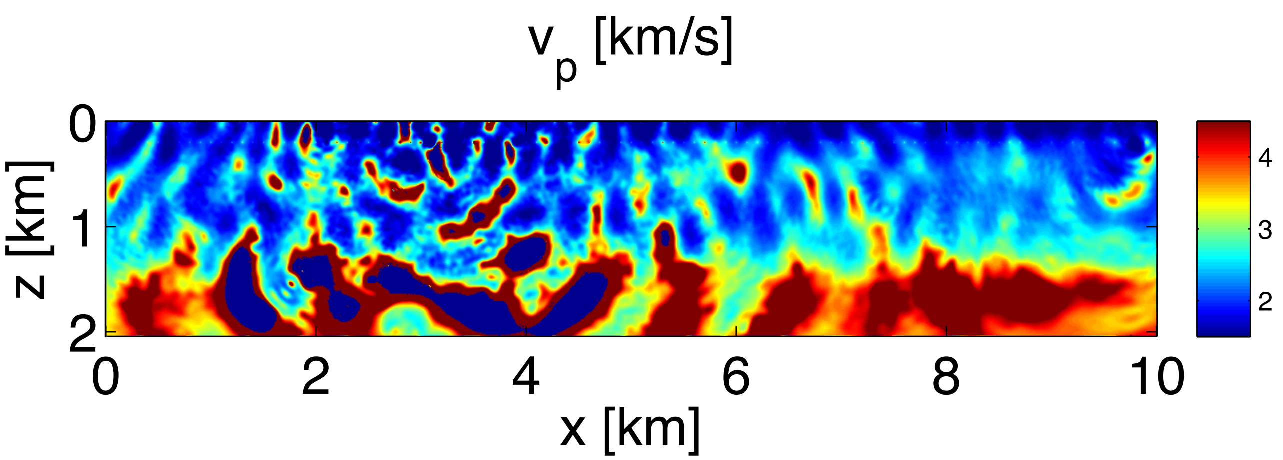
Students T (f,x)



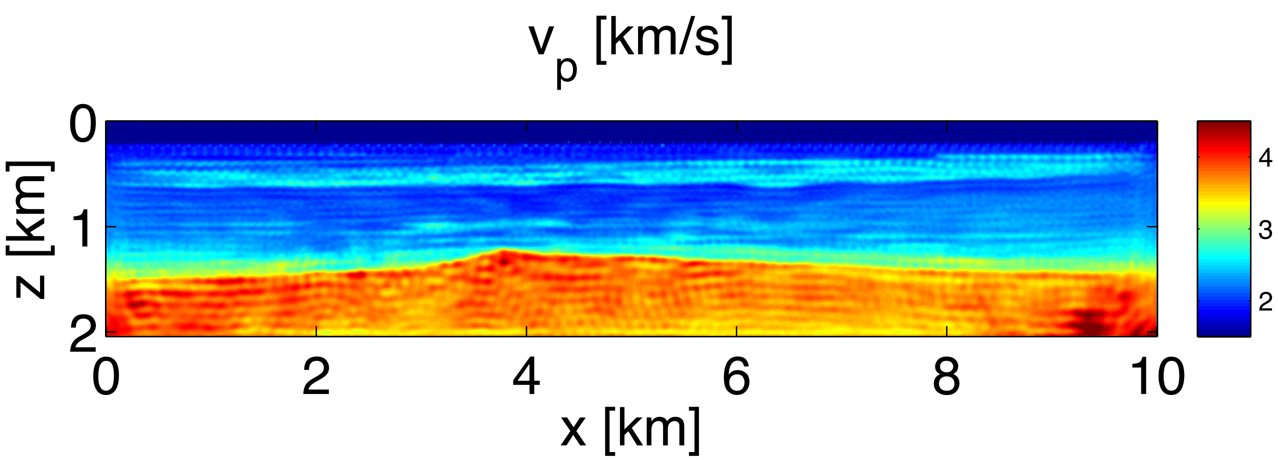
Results 2

Data with bad traces

least-squares



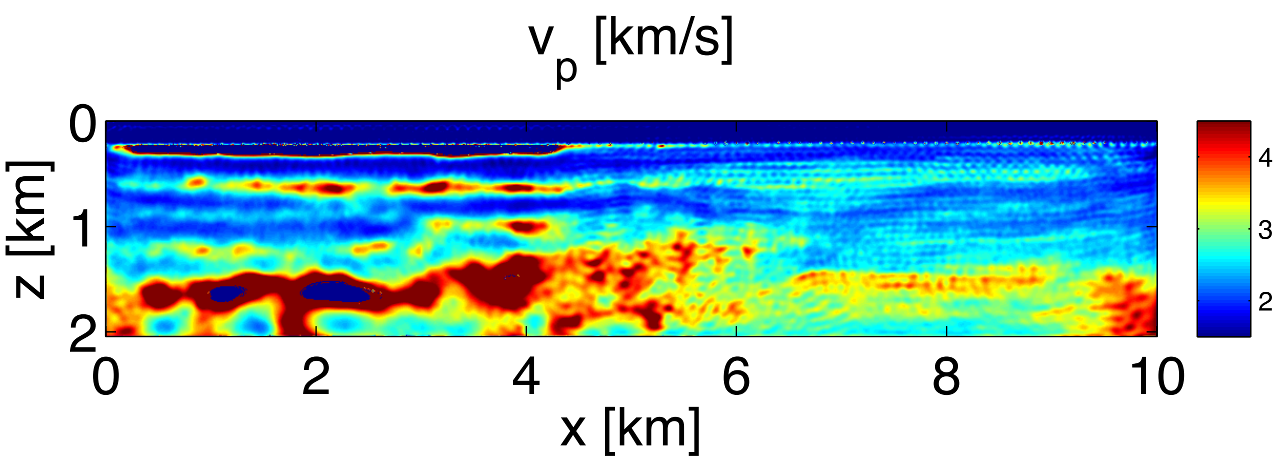
Students T (f,x)



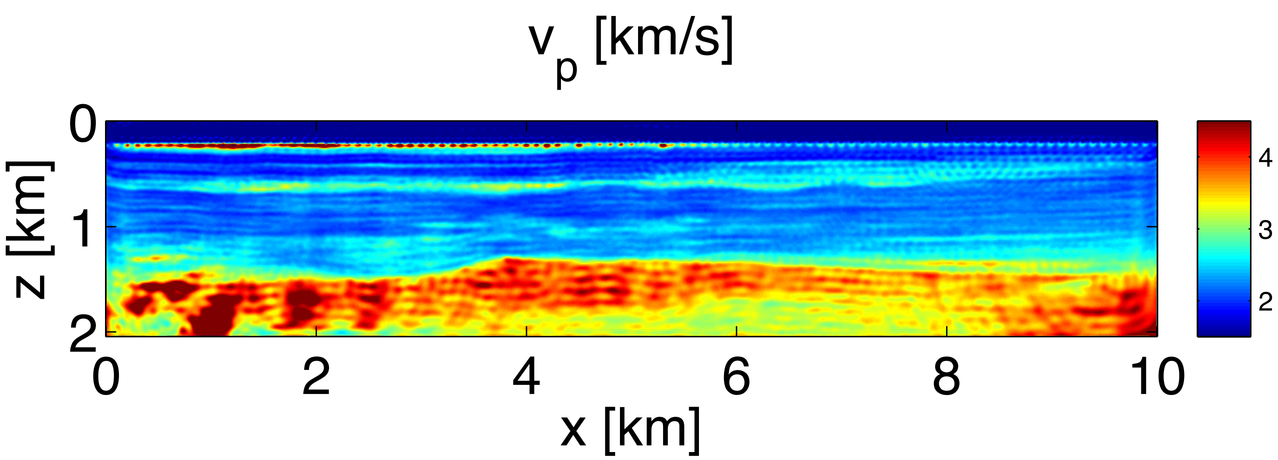
Results 2

Elastic data

least-squares



Students T (f,k)



Conclusions

- *Robust* inversion works best when noise is *localized*
- Measure misfit in domain in which noise is *sparse*
- *Source* and *scale* estimation can be done *automatically*
- ...