SINBAD Consortium 2010

# Robust FWI Using Student's t-distribution

**Aleksandr Aravkin, Tristan van Leeuwen, Felix Herrmann**

SLIM
University of British Columbia

SLIM

# Full Waveform Inversion

- **The Full Waveform Inversion (FWI) problem is to find solutions to the Helmholtz PDE that match data from source experiments on the surface**
- **Problems are typically very large: billions of variables and terabytes of data.**
- **Typically formulated as a Nonlinear Least Squares (NLLS) problem:**

$$\min_{\mathbf{m}} \quad \left\{ f(\mathbf{m}) := \|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]\|_F^2 \right\}$$

$$\begin{aligned}
\mathbf{D} &:= \text{data} \\
\mathbf{m} &:= \text{model parameters (speed or slowness squared)} \\
\mathbf{Q} &:= \text{multiple source experiments} \\
\boldsymbol{\mathcal{F}} &:= \text{solution operator of Helmholtz eqn. with absorbing boundary}
\end{aligned}$$

# Statistical Implications

- The NLLS formulation is equivalent to the following statistical model:

$$\begin{aligned} \mathbf{D} &= \mathcal{F}[\mathbf{m};\mathbf{Q}] + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \mathbf{N}(0, I) \end{aligned}$$

- Equivalence follows from maximum likelihood estimate for model parameters:

$$\mathcal{L}(\mathbf{m}) \quad \propto \quad \exp\left(-\frac{1}{2}\left\|\mathbf{D} - \mathcal{F}[\mathbf{m};\mathbf{Q}]\right\|_F^2\right)$$

- Minimizing the negative log likelihood is exactly the FWI problem.

- Q: So what?

SLIM

# Practical Consequences

- **Large deviations from the mean are VERY unlikely in the Gaussian model:**

| | Gaussian |
|---|---|
| $p(\|x-\mu\| > 4\sigma)$ | $6.3 \times 10^{-5}$ |
| $p(\|x-\mu\| > 8\sigma)$ | $1.3 \times 10^{-15}$ |

- **Observations more than 4 standard deviations away from the mean occur less than .006 percent of the time.**
- **As we get further away, the likelihood shrinks astronomically.**
- **Low likelihood values correspond to HIGH penalties for outliers.**

# 'Outliers' in FWI??

- Mathematical model cannot distinguish 'artifacts' from 'outliers'. Any unexplained events in the residual will have a strong effect on the final image.

- Examples:
  1) Modeling Inelastic/Anisotropic data with Acoustic PDE
  2) Ignoring Acquisition Models

  [Brossier '10]

- Key point: models are improving all the time, but are never perfect. It is worthwhile to have methods that still perform well when models are wrong.

- Q: How do we design such methods?

# Statistical Modeling

- **We can tweak the assumptions on the NOISE in the model:**

$$\begin{aligned} \mathbf{D} &= \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}] + \boldsymbol{\epsilon} \\ \boldsymbol{\epsilon} &\sim \text{Heavy Tailed Distribution} \end{aligned}$$

- **The parametric form of the distribution then determines the optimization formulation:**

$$\min_{m} \quad -\log(\boldsymbol{\mathcal{L}}(\mathbf{m})) := -\log\Big(f\left(\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]\right)\Big)$$

- **Q: Which distribution do we choose, and how do we solve the problem?**

# A Simple Comparison

- **We present a comparison with two other distributions:**

| | Gaussian | $L(\lambda = 1)$ | $T(df = 3)$ |
|---|---|---|---|
| $p(\|x - \mu\| > 4\sigma)$ | $6.3 \times 10^{-5}$ | $1.8 \times 10^{-2}$ | $0.6 \times 10^{-2}$ |
| $p(\|x - \mu\| > 8\sigma)$ | $1.3 \times 10^{-15}$ | $3.3 \times 10^{-4}$ | $8.1 \times 10^{-4}$ |

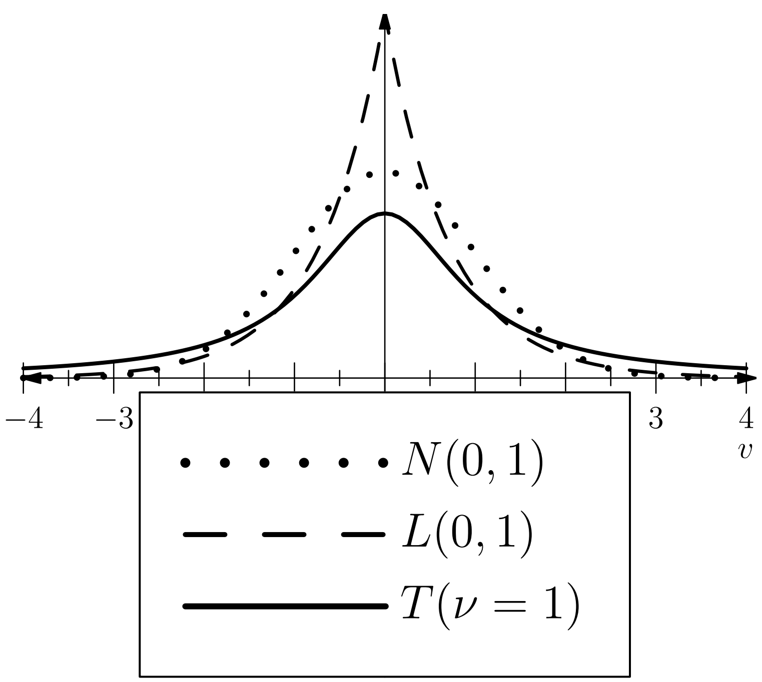
- **The Laplace distribution corresponds to the L1 penalty on the misfit:**

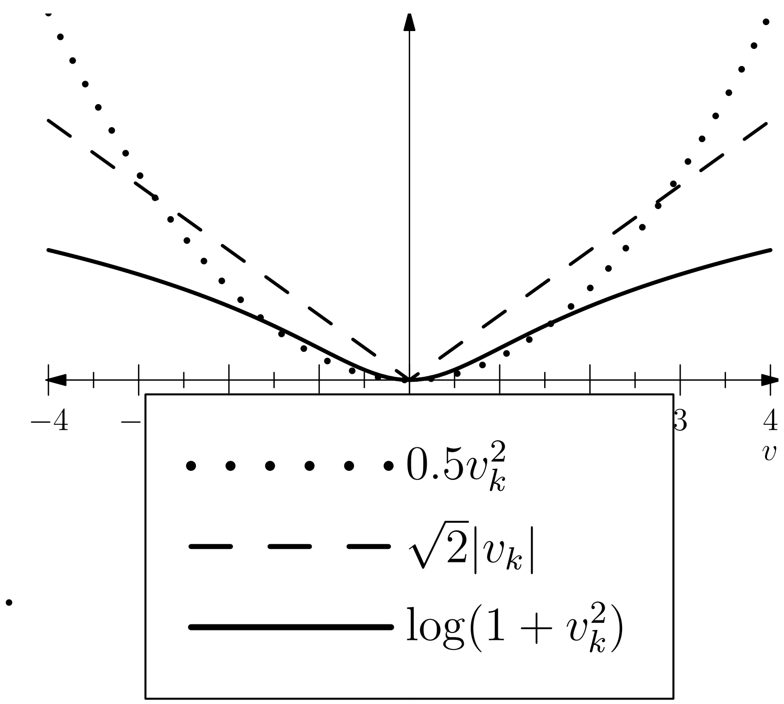
$$\|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m}; \mathbf{Q}]\|_1$$

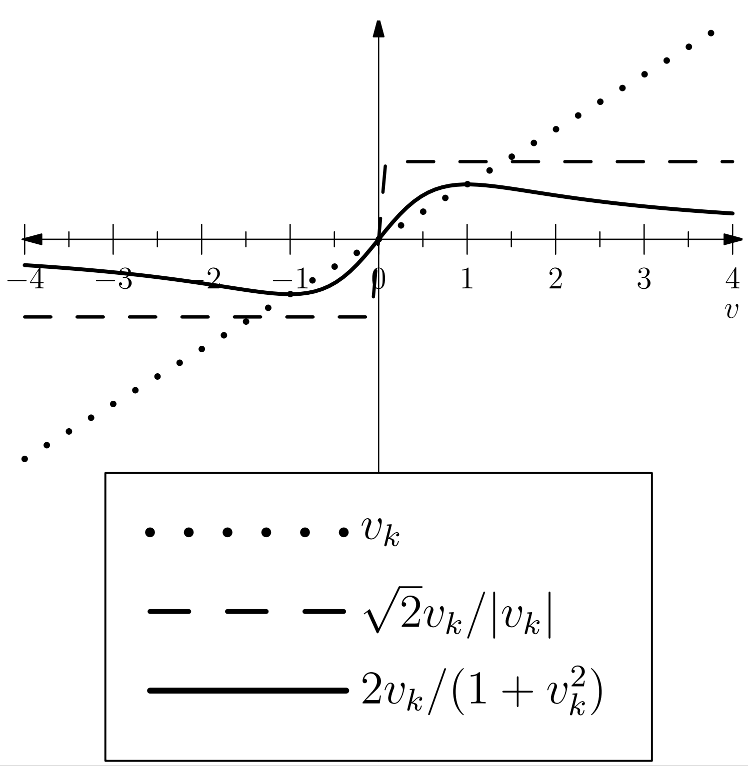
- **In the class of CONVEX negative log likelihoods, it has the heaviest tail (as does the distribution corresponding to the Huber misfit).**

- **But the full problem is non-convex anyway, so let's consider Student's t!**

# Densities, Penalties, and Gradients



# Student's t-Distribution

- **Student's t-density**

$$\mathbf{p}(\epsilon|\mu, M) \quad = \quad \frac{\Gamma(\frac{s+l}{2})}{\Gamma(\frac{s}{2})\det[\pi s M]^{1/2}}\left(1 + \frac{\|\epsilon - \mu\|^2_{M^{-1}}}{s}\right)^{\frac{-(s+l)}{2}}$$

- **Student's t-objective for Seismic case (negative log likelihood)**

$$\min_m \quad \phi_T(m) := \sum_i \frac{s+l}{2} \log\left(s + (D - PH[m]^{-1}Q)^2_i\right)$$

- **Student's t-gradient**

$$\nabla_m \phi_T(m) = \sum_i \frac{s+l}{2} \frac{\nabla_m F[m;Q]^T_i (F[m;Q] - D)_i}{s + (D - F[m;Q])^2_i} \, .$$

# Gradient Comparison:

NLLS

$$\sum_i \frac{1}{2} \nabla_m F[m;Q]_i^T (F[m;Q]-D)_i$$

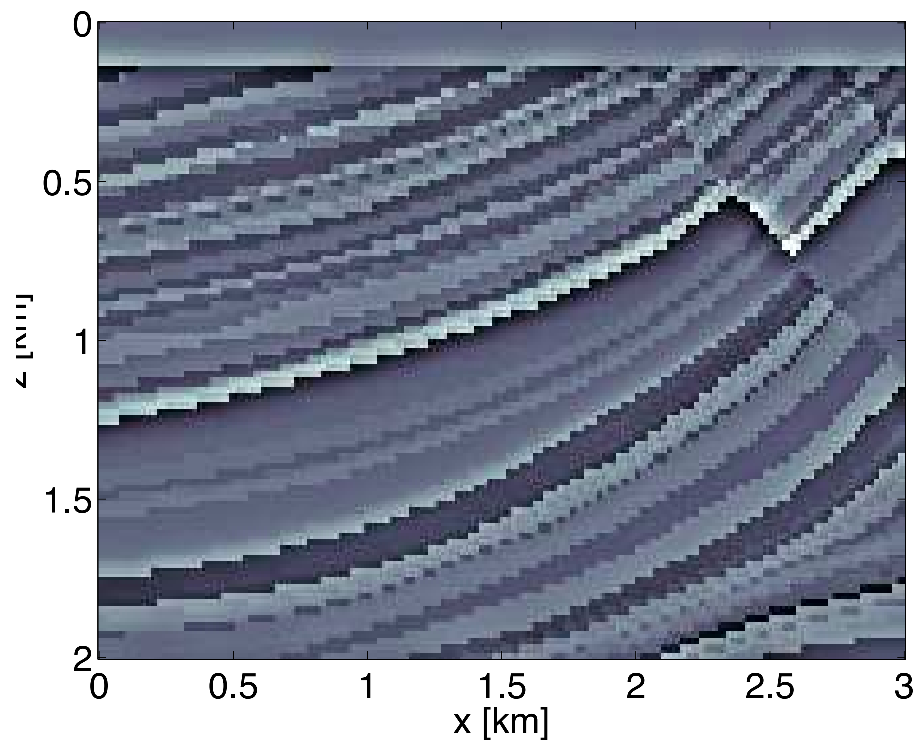
Student

$$\sum_i \frac{s+l}{2} \frac{\nabla_m F[m;Q]_i^T (F[m;Q]-D)_i}{s+(D-F[m;Q])_i^2} \, .$$

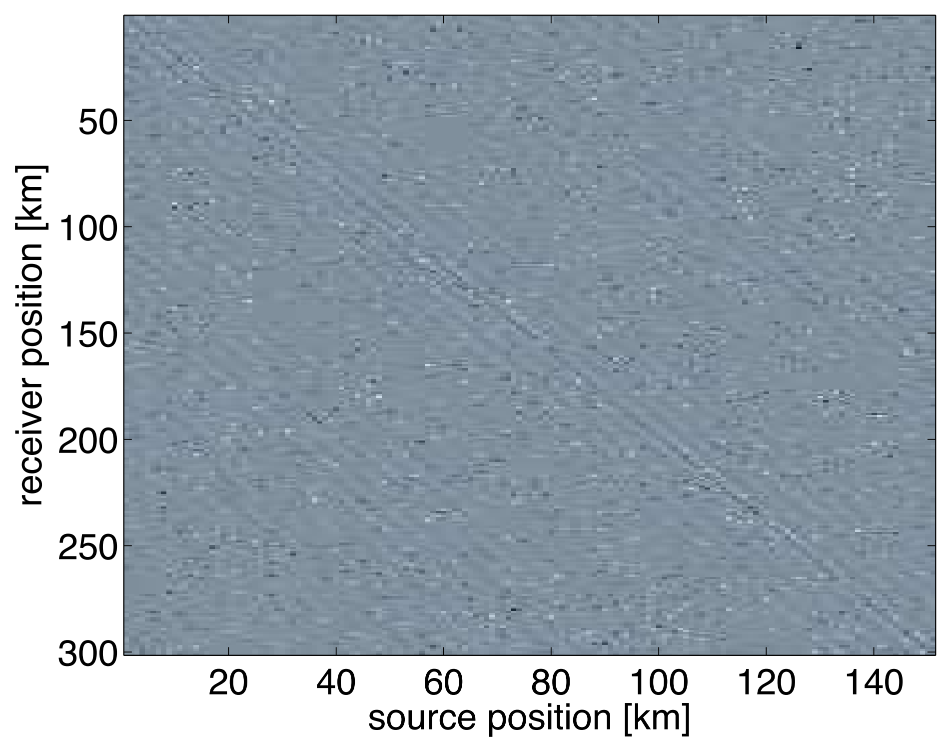
# Marmoussi Example

- We consider a subset of the Marmoussi model
- 151 shots, 301 receivers
- 9 pt. discretization of Helmholtz operator with absorbing boundary; 10 m. spacing on grid
- Sample of Frequencies [5.0, 6.0, 11.5, 14.0, 15.5, 17.5, 23.5] Hz

# Experiment I

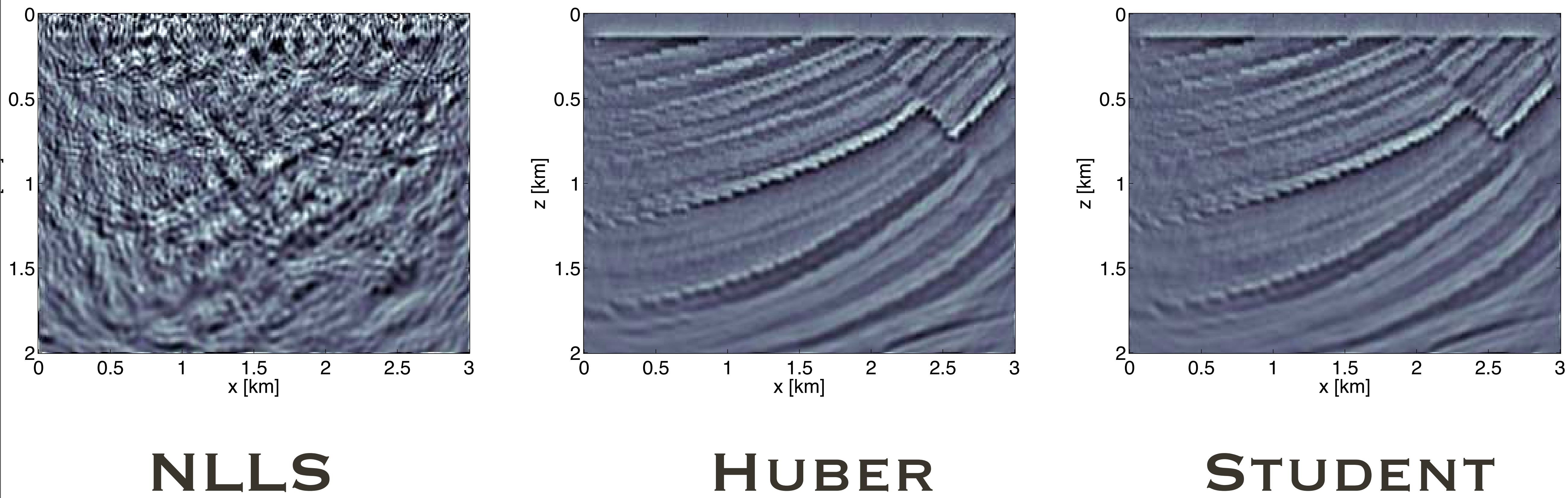



True Reflectivity

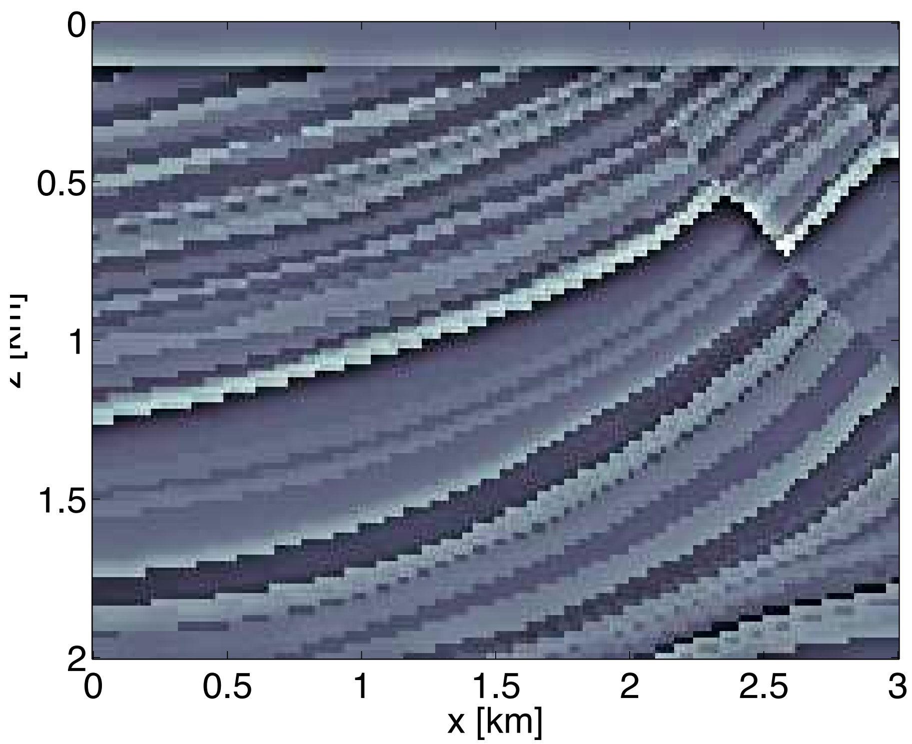



15 hz data slice
with spiky noise

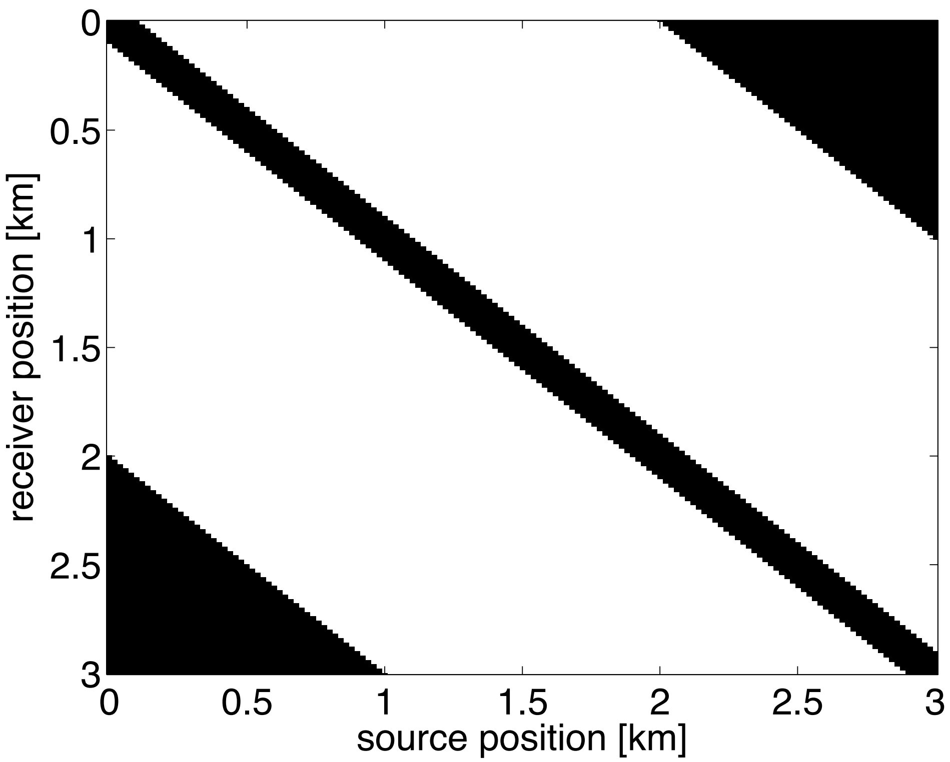
# Results I



# Experiment II

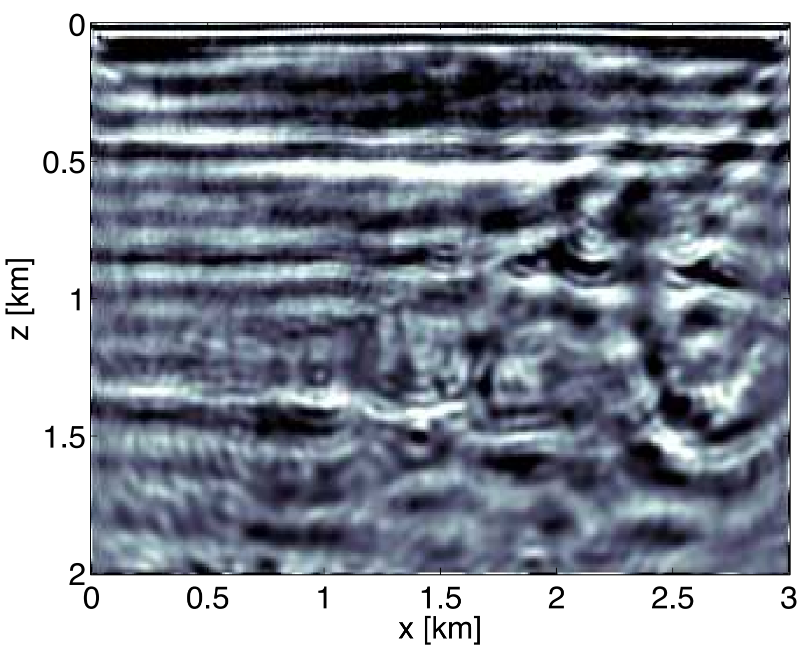



True Reflectivity

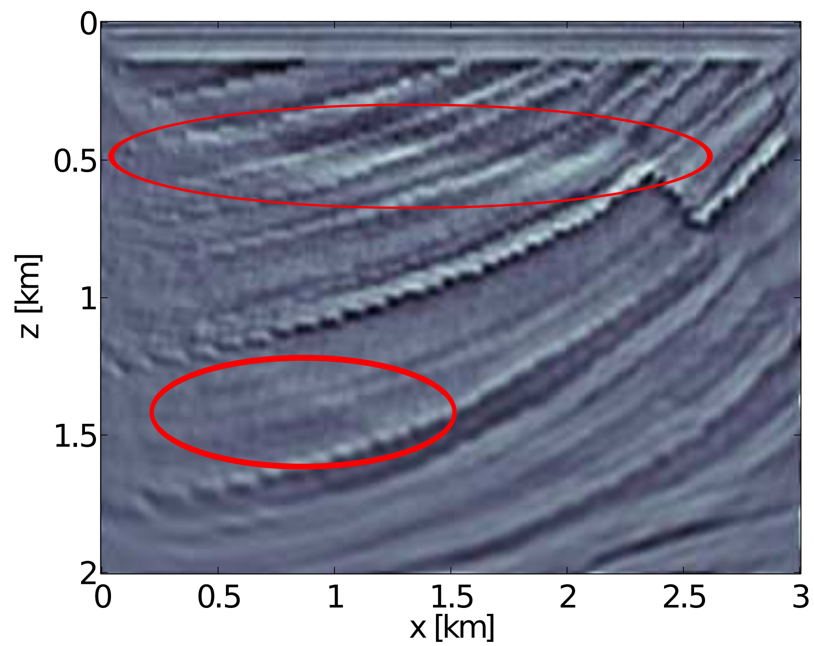



Marine Acquisition Mask

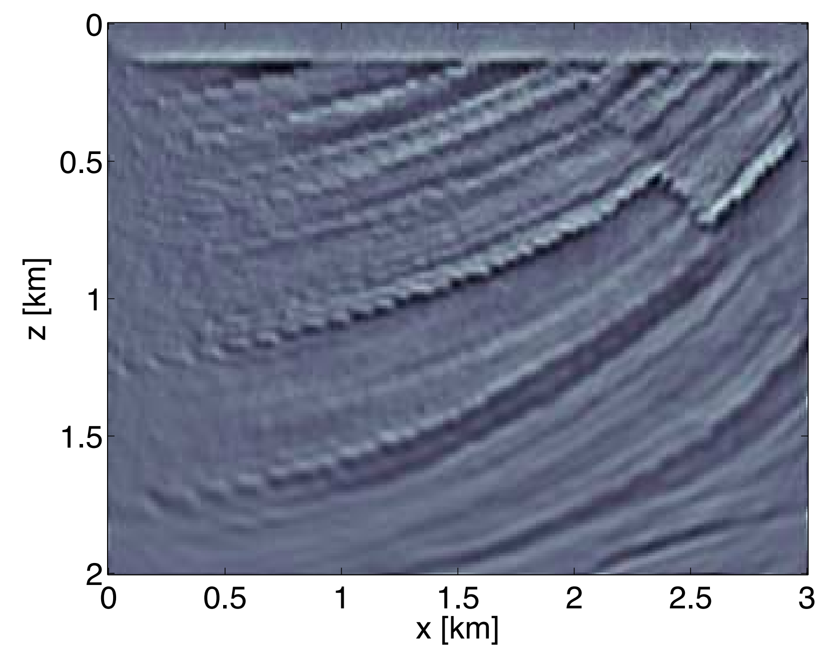
# Results II




NLLS




Huber




Student

SLIM

# Conclusions

- Robust formulations allow good recovery even with poor modeling
- 'Mistakes' are typically thought of as 'outliers' in the data, but can also be events unexplained (or ignored) by the modeling
- Since the FWI problem is non-convex, we can feel free to exploit distributions with non-convex negative log likelihoods
- Future direction: combining robust and sparse recovery.

# Acknowledgements

SINBAD

NSERC CRSNG

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.