AVA analysis as an unsupervised machine learning problem

Ben Bougher, October 26th 2016

SLIM
University of British Columbia

Why machine learning?

Machine learning should be applied to a problem when:

- There exists an underlying but unknown relationship between input and output.
- There is no known physical model to describe the relationship, or there are too many unrealistic assumptions and approximations.

Supervised learning:

- Most high profile applications are supervised learning.
- Requires large database of truth data for training.

Unsupervised learning:

- Learn latent relationships directly from the data.

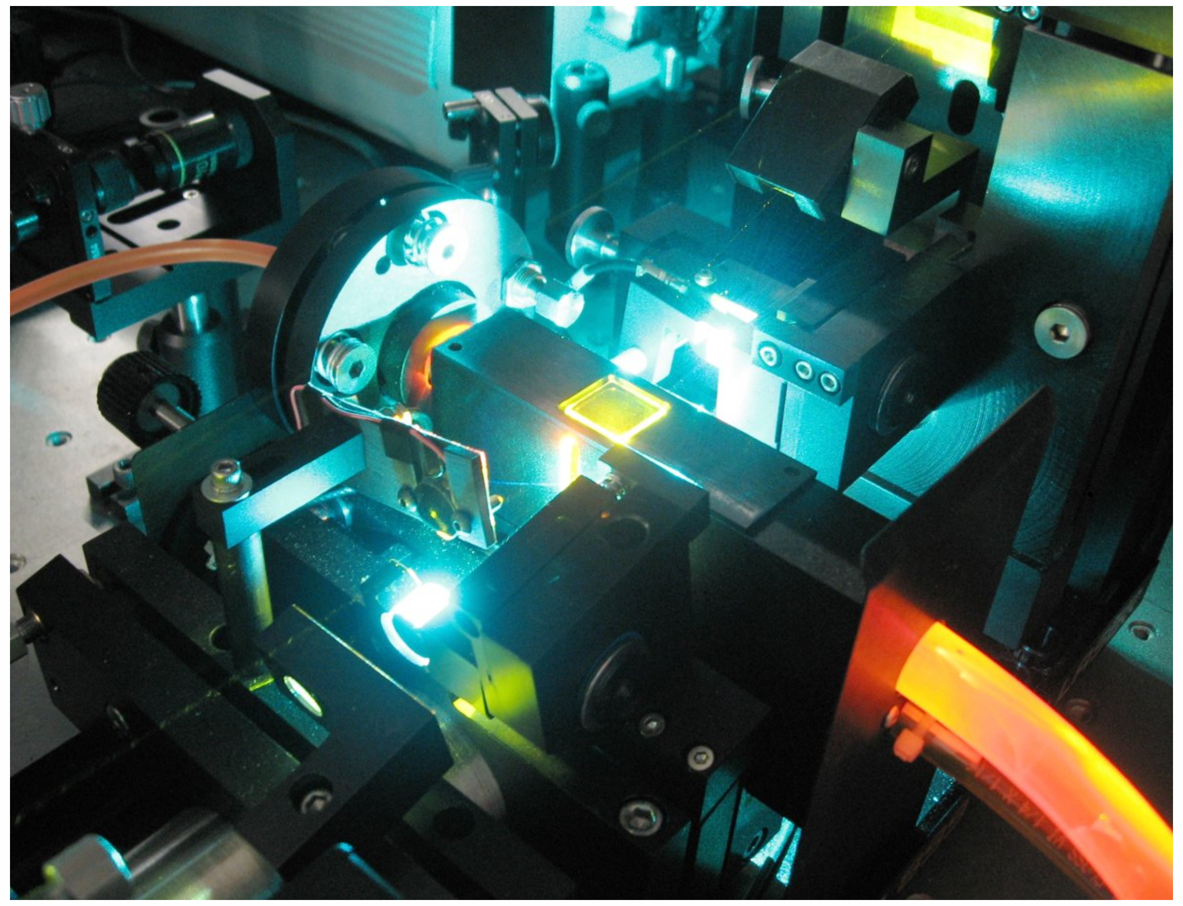
Scattering physics

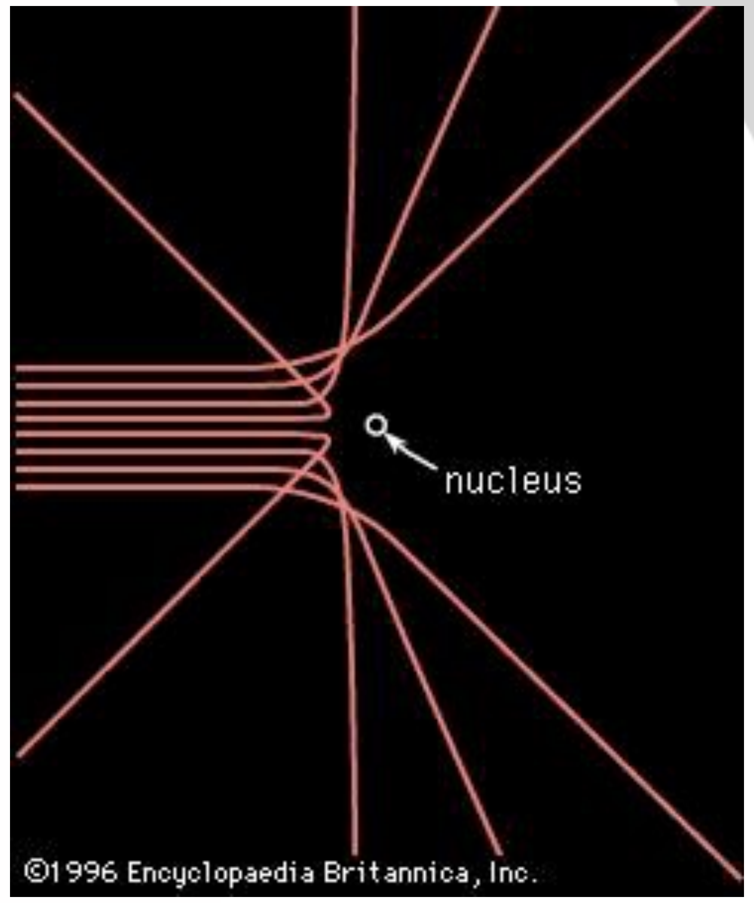
Ubiquitous in experimental physics.

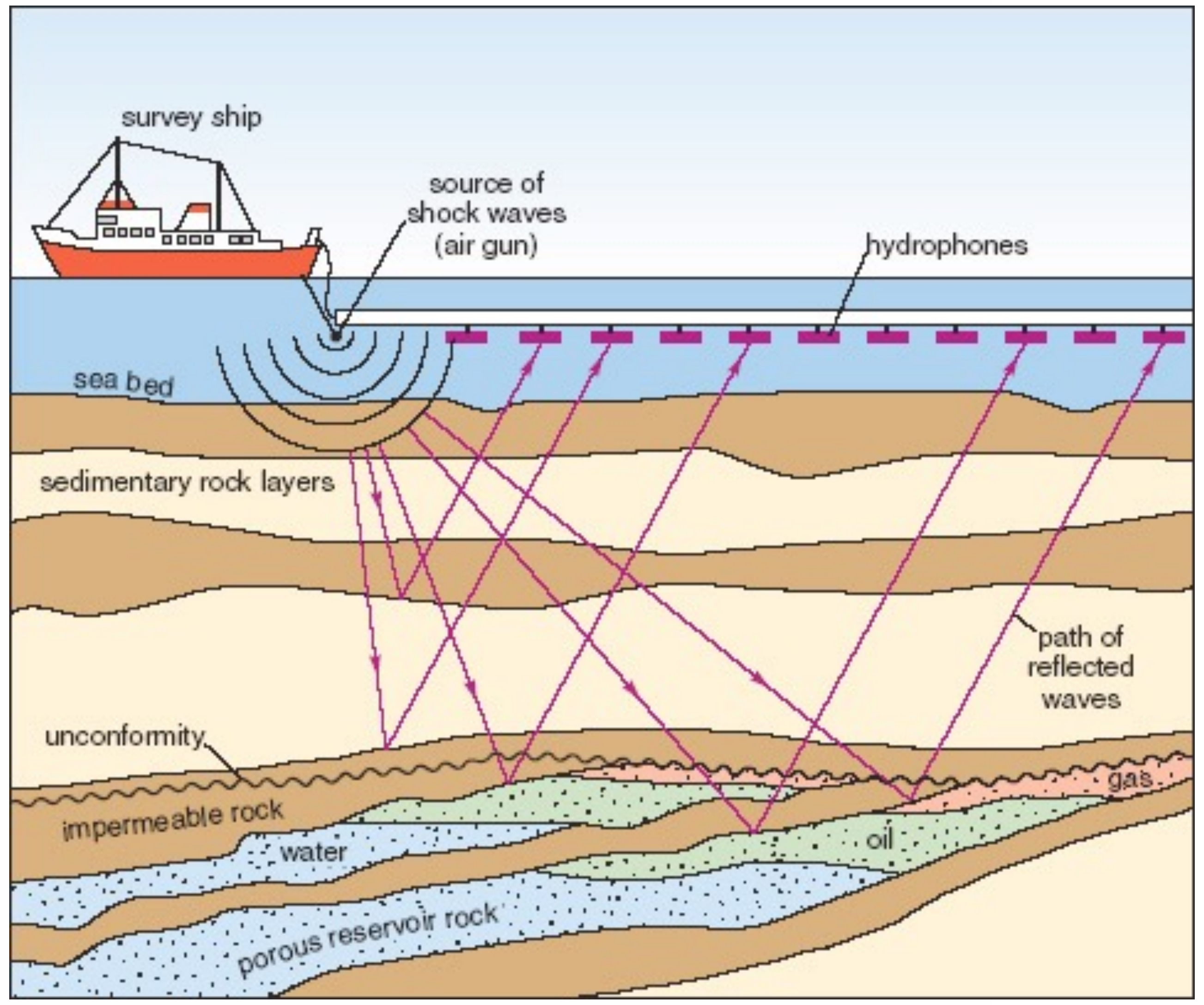
Measure the scattering pattern from a known source incident on a material.

Performed in highly controlled and calibrated laboratories (laser sources, temperature controlled, vacuums, etc...).

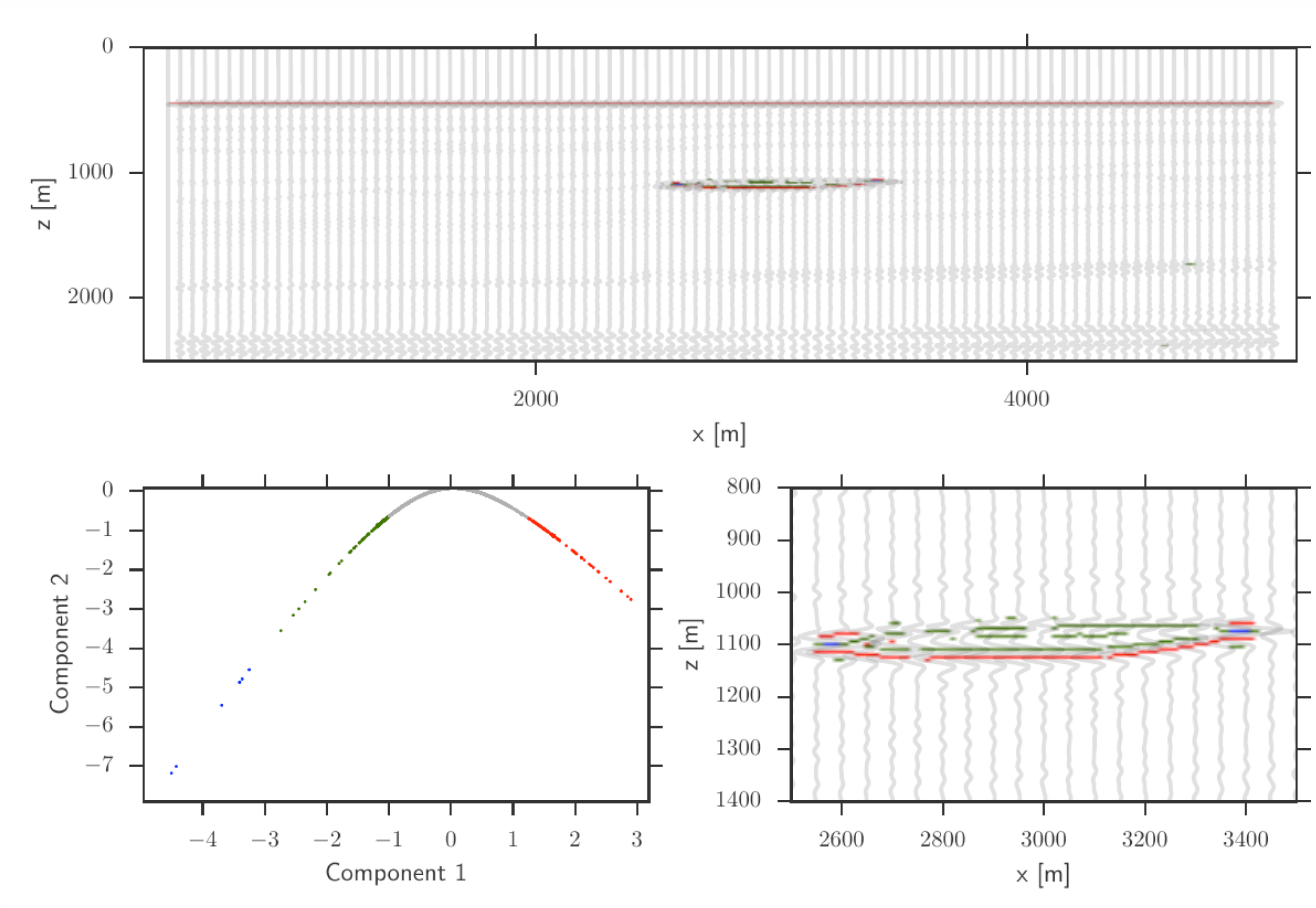
Reflection seismology is a scattering experiment in an uncontrolled environment.







Reflection seismology as an unsupervised learning problem



Problem:

Automatically segment potential hydrocarbon reserves from seismic images

Approaches:

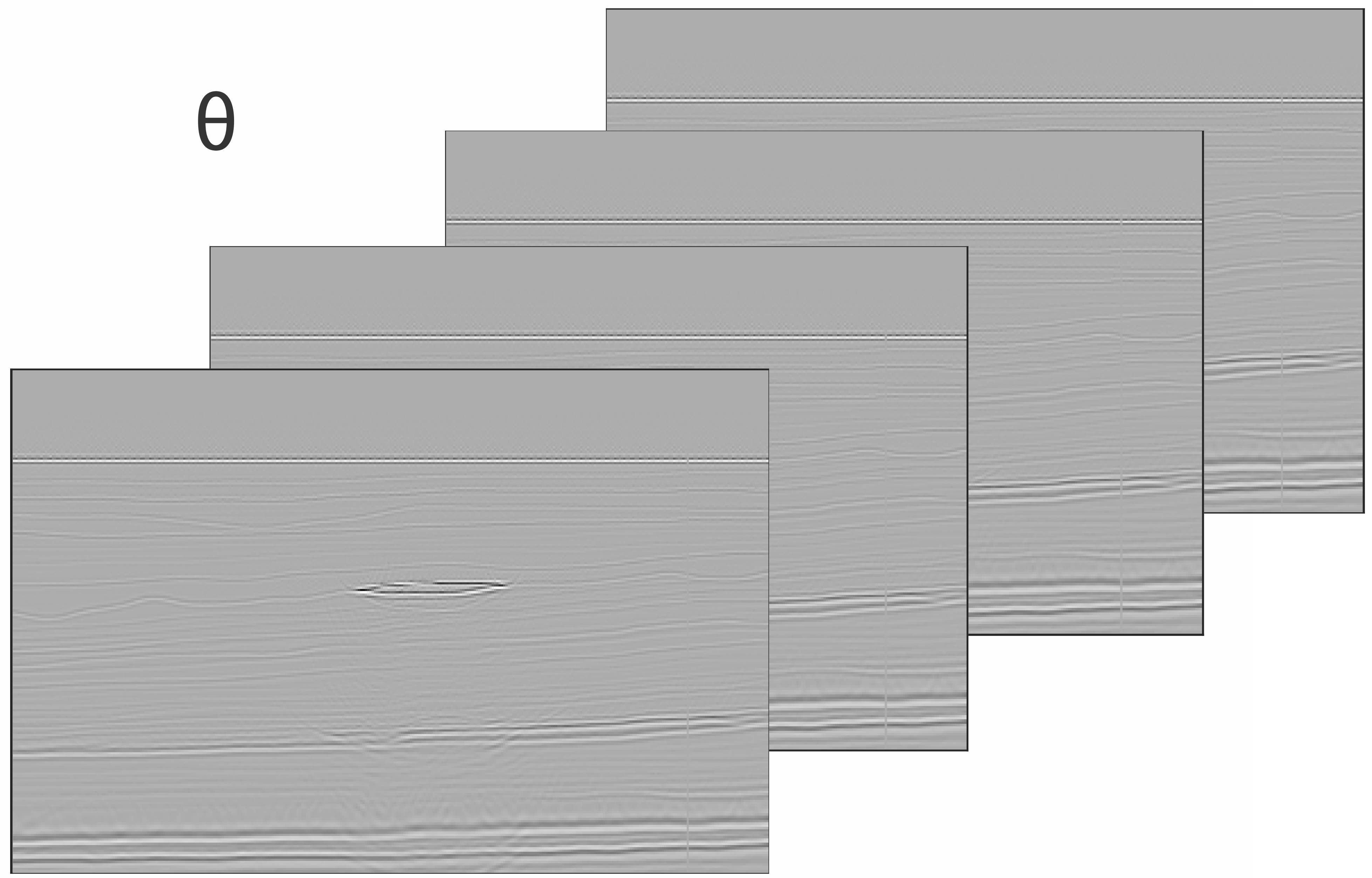
Physics driven (conventional)

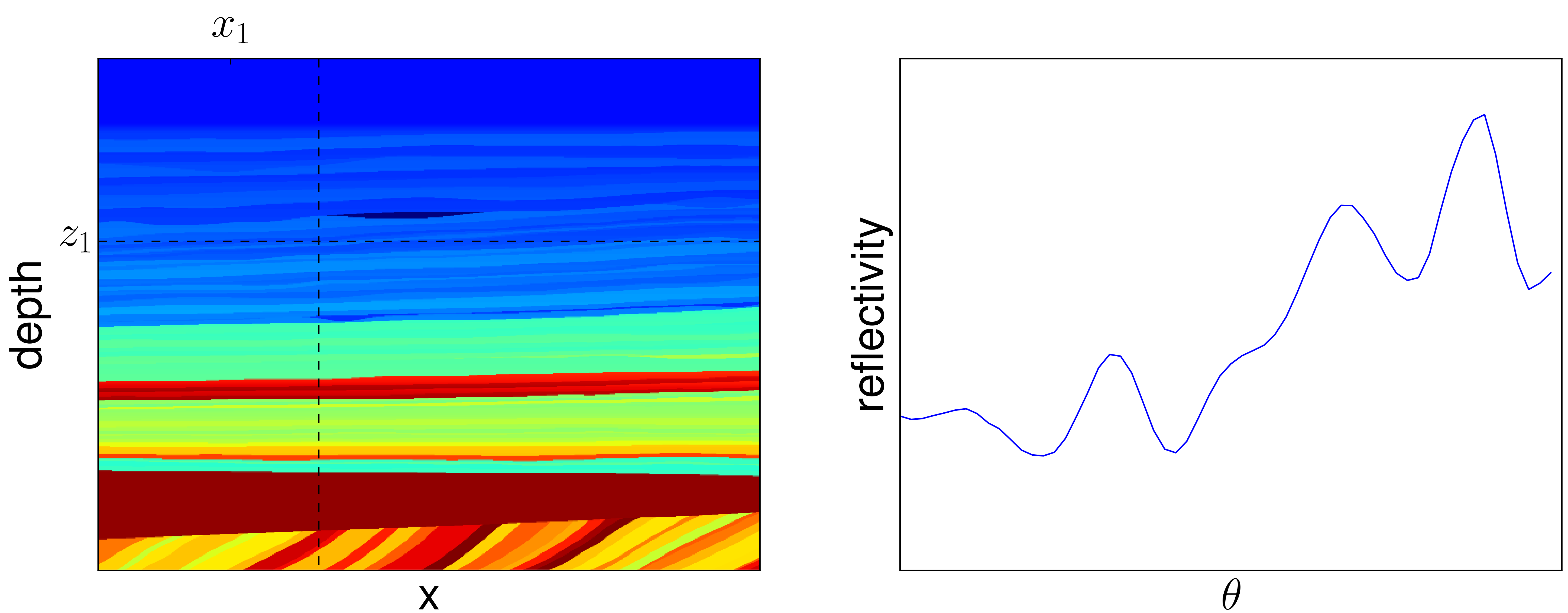
Data driven (thesis contributions)

Angle domain common image gathers

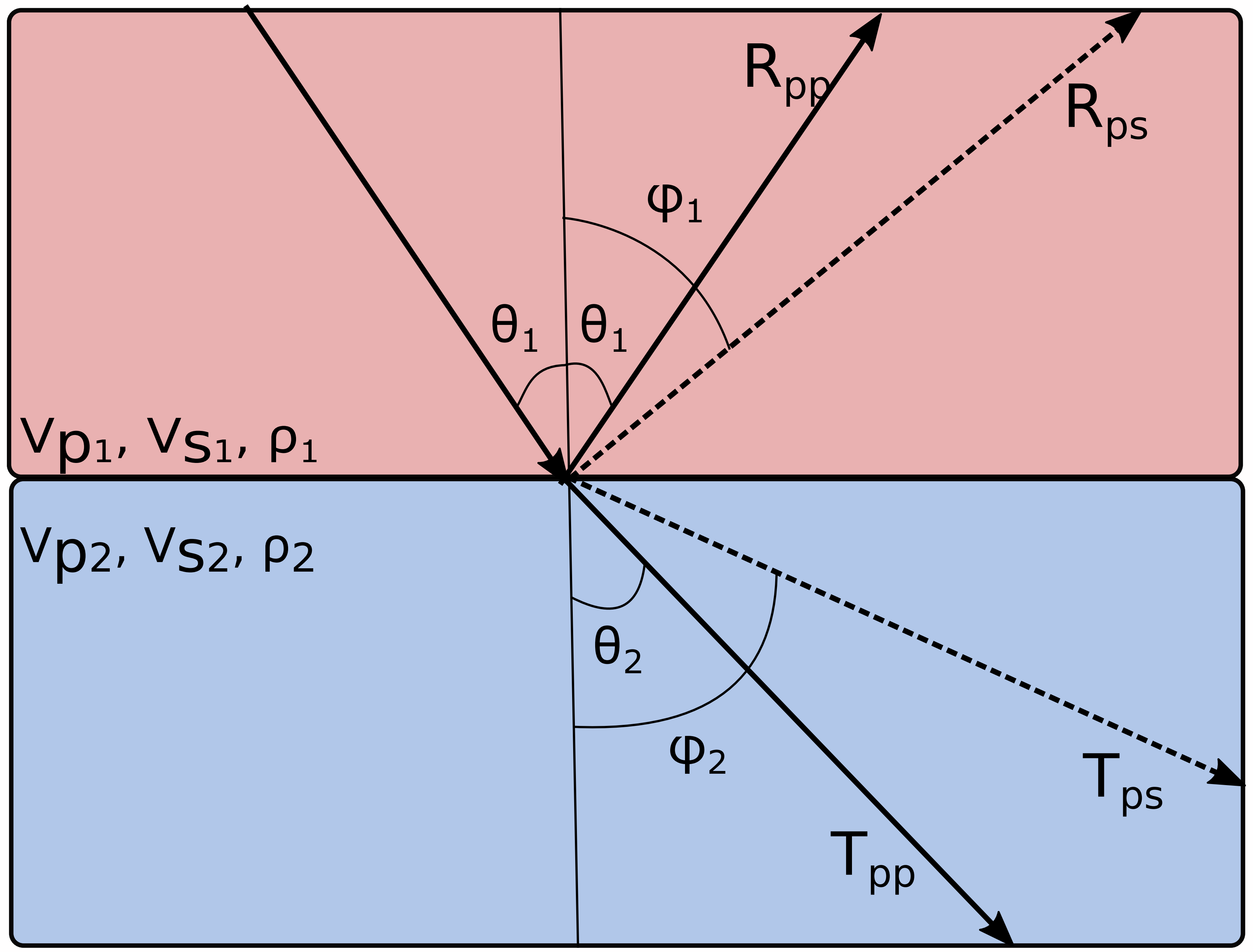
Problem:
Need angle dependent reflectivity responses

Solution:
Angle domain common image gather migration





Scattering theory (Zoeppritz)



Problem:
Relate angle dependent reflectivity to rock physics.

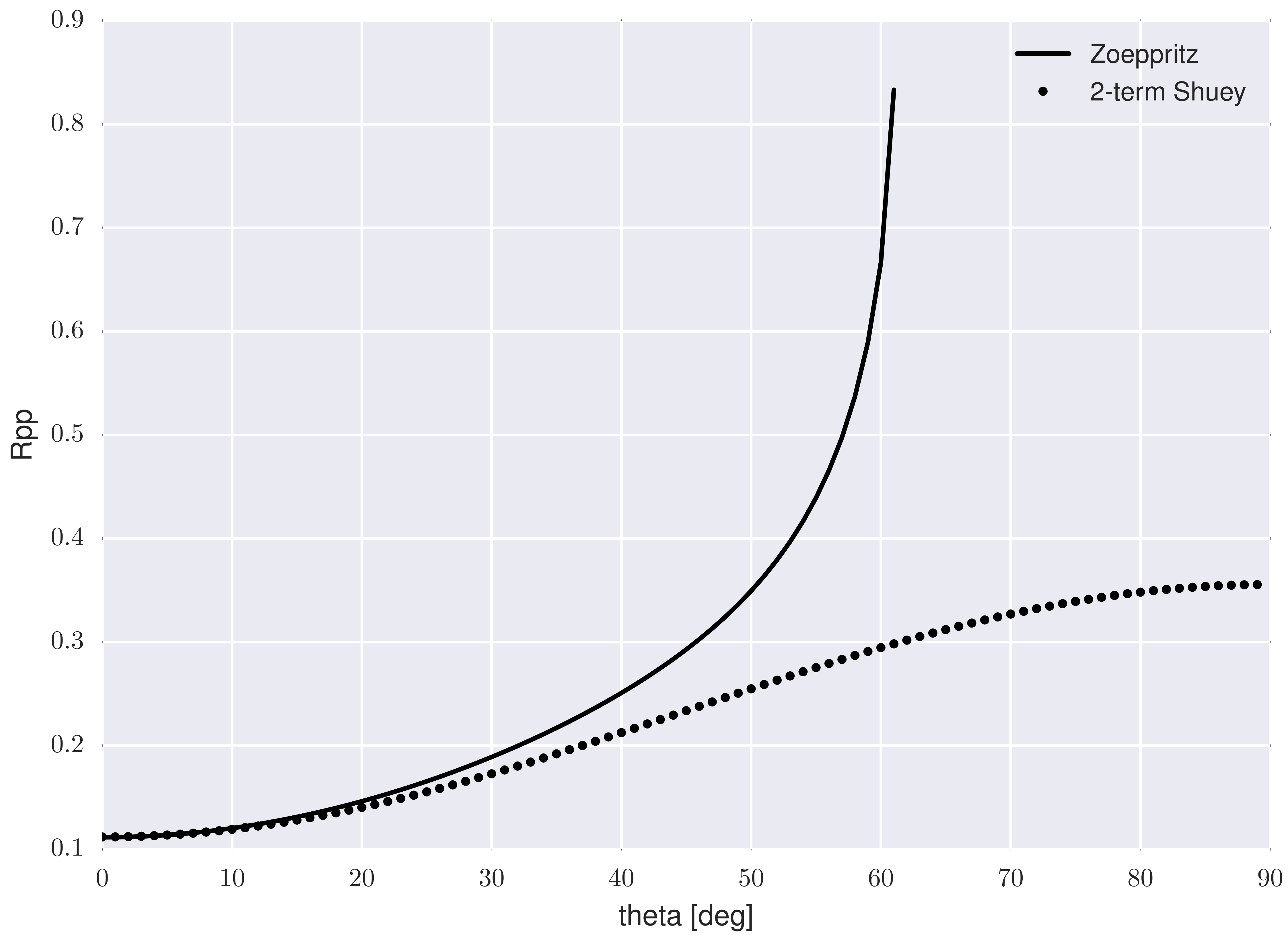
Assumption: Ray theory approximation.

Solution: $R(\theta) \propto V_P, V_S, \rho$

Problem: Non-linear, not useful for inversion.

$$\begin{bmatrix} R_{PP} \\ R_{PS} \\ T_{PP} \\ T_{PS} \end{bmatrix} = \begin{bmatrix} -\sin\theta_1 & -\cos\phi_1 & \sin\theta_2 & \cos\phi_2 \\ \cos\theta_1 & -\sin\phi_1 & \cos\theta_2 & -\sin\phi_2 \\ \sin 2\theta & \frac{V_{P1}}{V_{S1}} cos 2\phi_1 & \frac{\rho_2 V_{S2} 2 V_{P1}}{\rho_1 V_{S1}^2 V_P 2} \cos 2\theta_1 & \frac{\rho_2 V_{S2} V_{P1}}{\rho_1 V_{S1}^2} \cos 2\phi_2 \\ -cos\phi_2 & \frac{V_{S1}}{V_{P1}} \sin 2\phi_1 & \frac{\rho_2 V_{P2}}{\rho_1 V_{P1}} & \frac{\rho_2 V_{S2}}{\rho_1 V_{P1}} \sin 2\phi_2 \end{bmatrix}^{-1} \begin{bmatrix} \sin\theta_1 \\ \cos\theta_1 \\ \sin 2\theta_1 \\ \cos 2\phi_1 \end{bmatrix}$$

Scattering theory (Shuey)



$$R_{pp}(\theta) = i(\Delta V_P, \Delta\rho) + g(\Delta V_P, \Delta V_S, \Delta\rho)\sin^2\theta$$

$$i(\Delta V_P, \Delta\rho) = \frac{1}{2}\Big(\frac{\Delta V_P}{\langle V_P\rangle} + \frac{\Delta\rho}{\langle\rho\rangle}\Big)$$

$$g(\Delta V_P, \Delta V_S, \Delta\rho) = \frac{1}{2}\frac{\Delta V_P}{\langle V_P\rangle} - 2\frac{\langle V_S\rangle^2}{\langle V_P\rangle^2}\Big(\frac{\Delta\rho}{\langle\rho\rangle} + 2\frac{\Delta V_S}{\langle V_S\rangle}\Big)$$

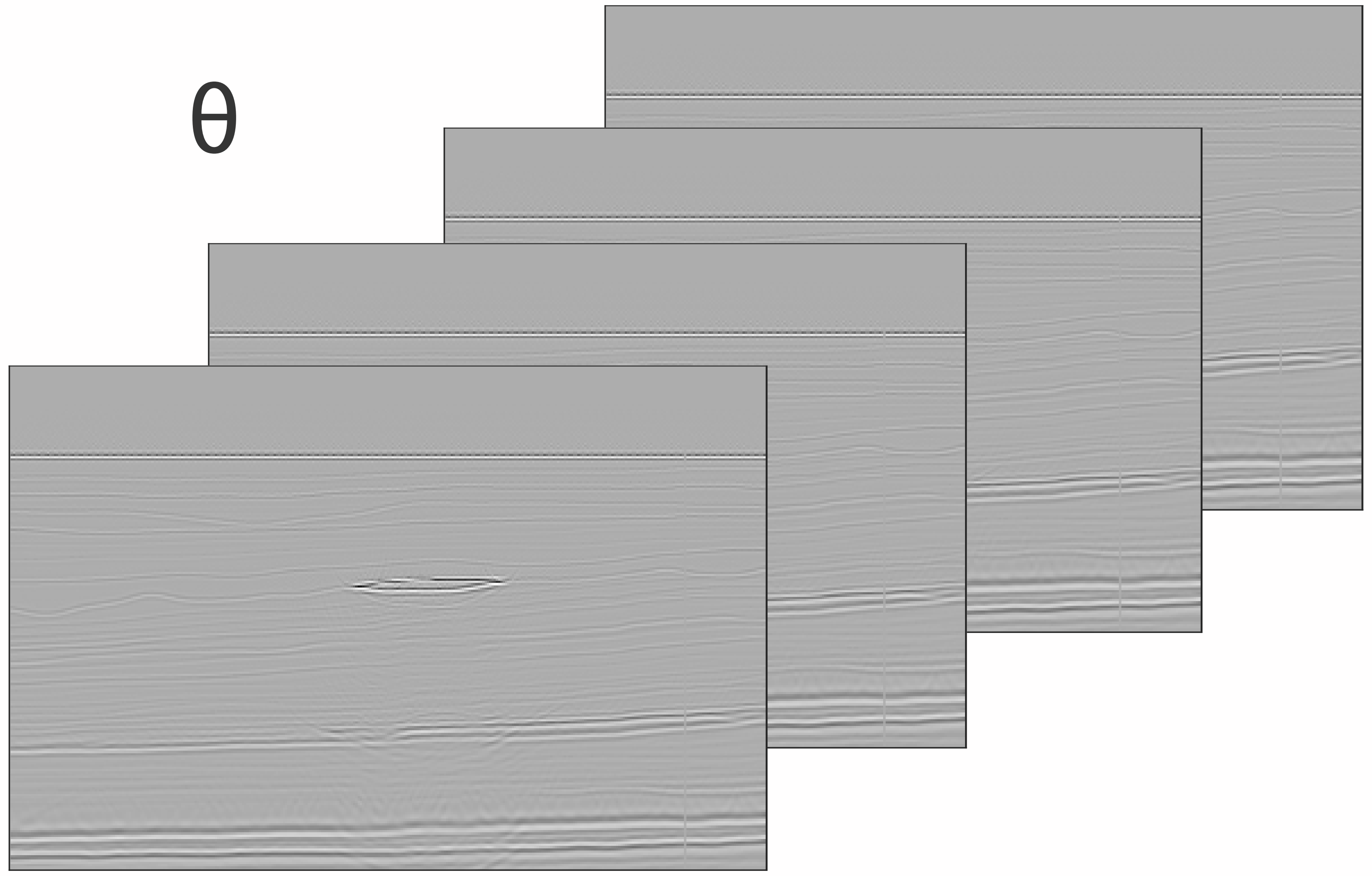
Limitations:

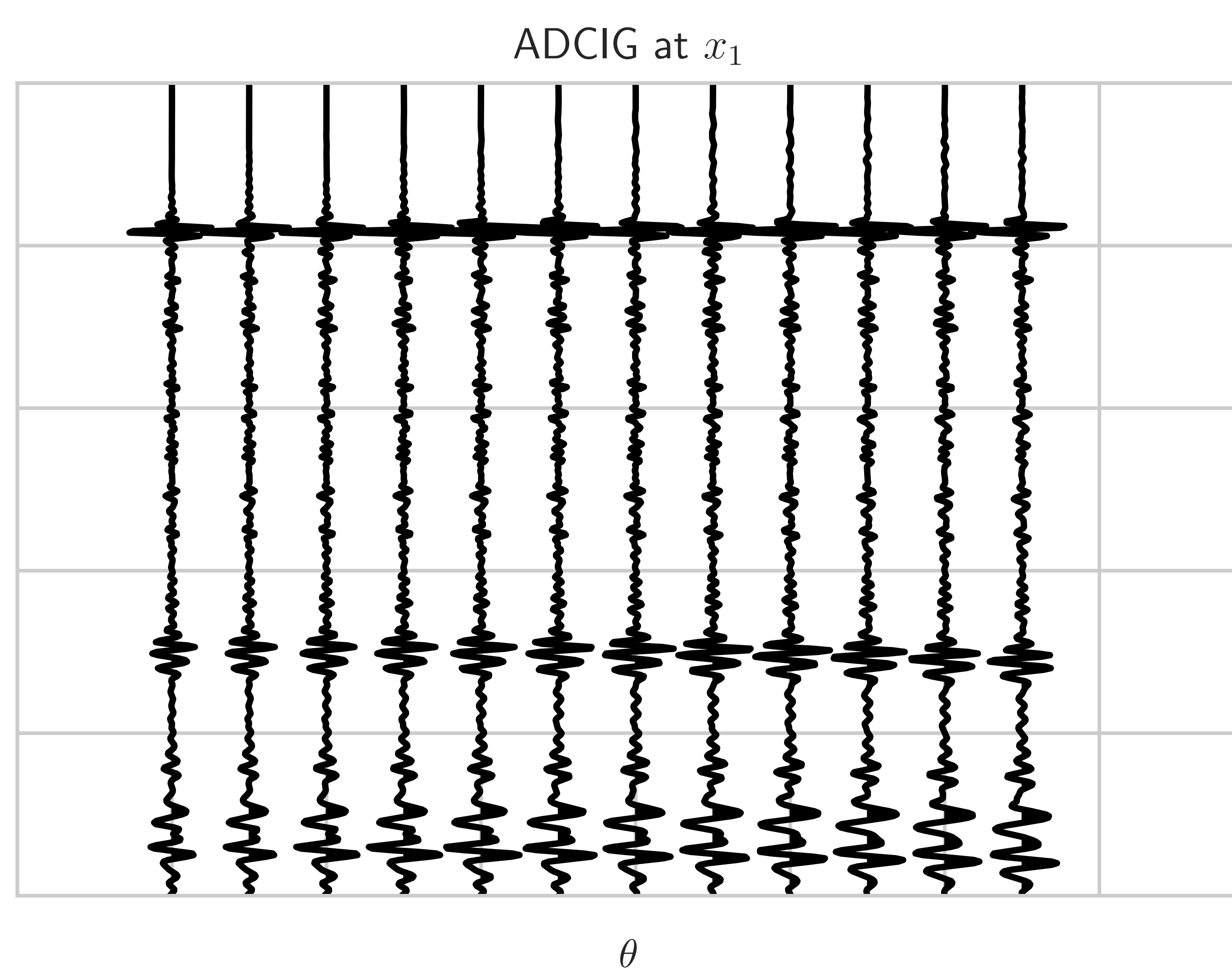
Small perturbations over a background trend, valid < 30 degrees

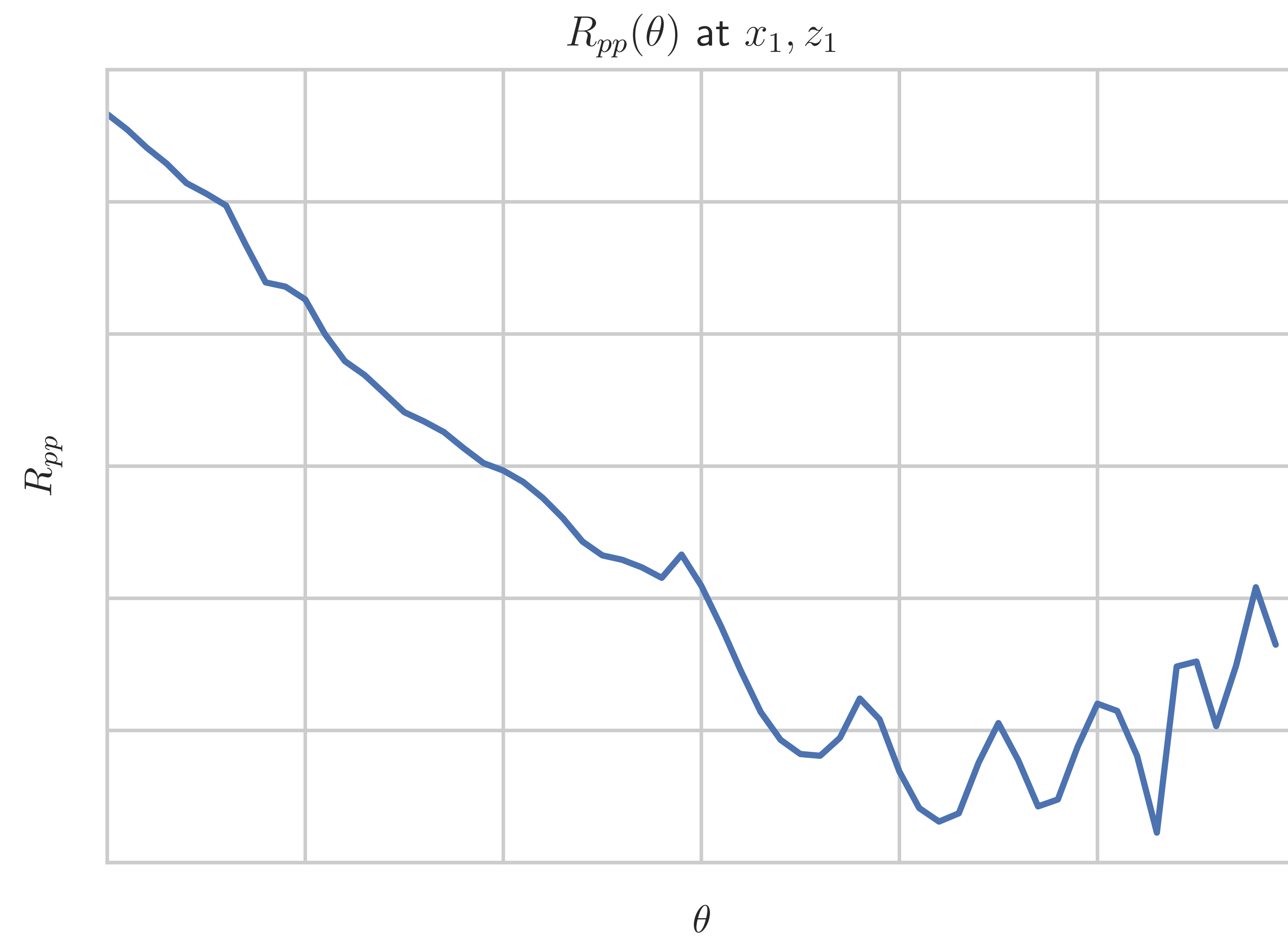
Benefits:

Linear for i and g, invert using simple least squares

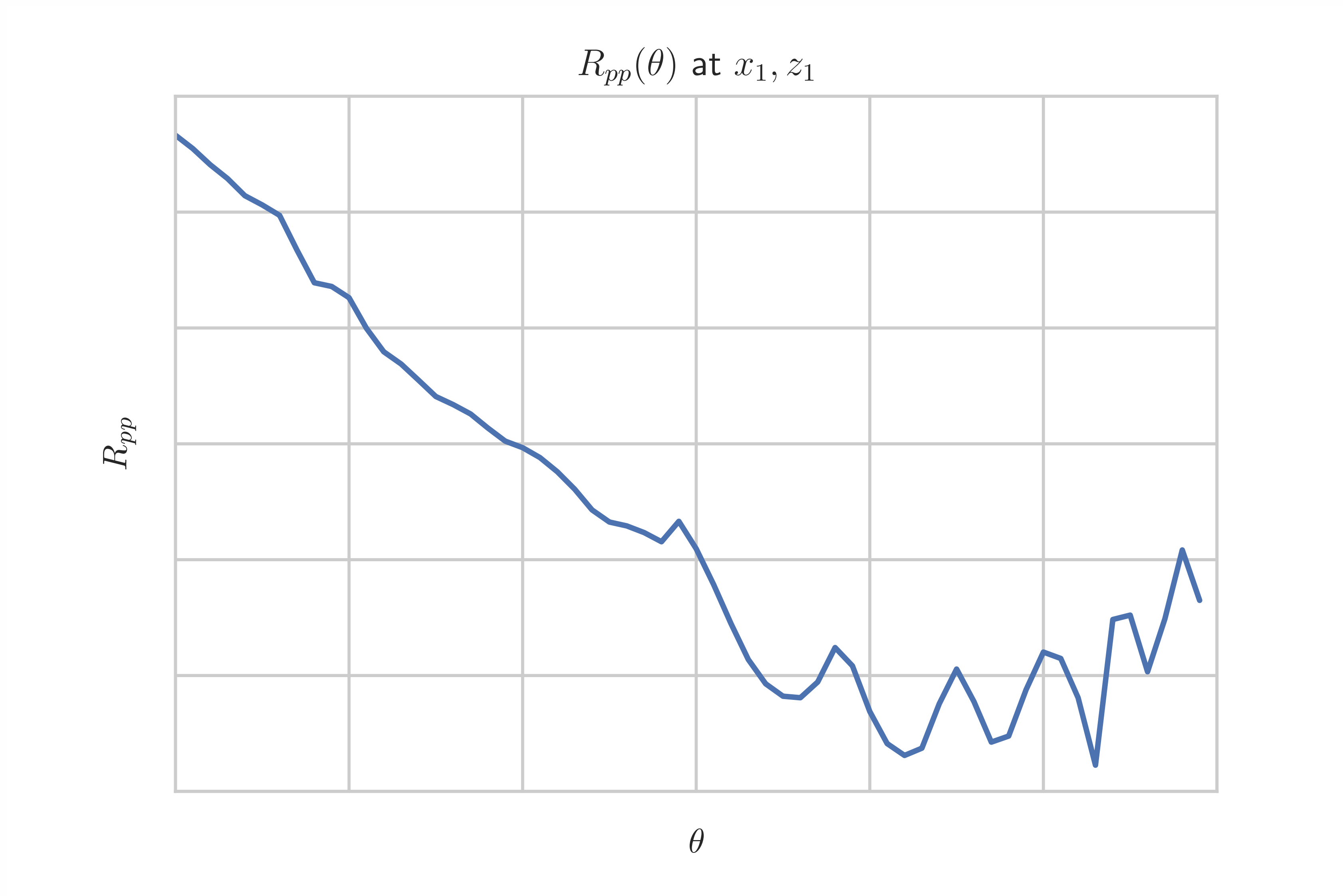
Shuey term inversion as a projection

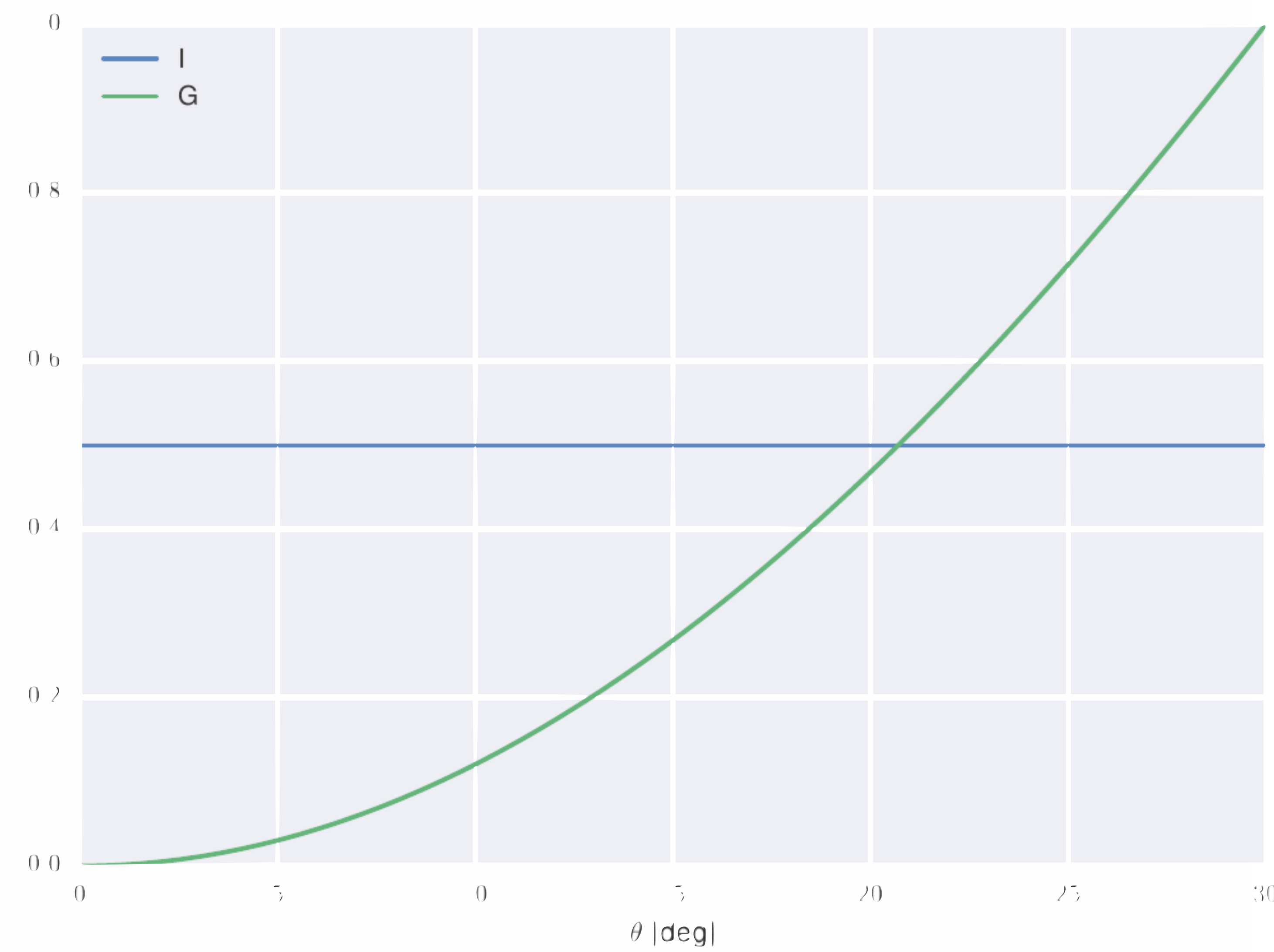






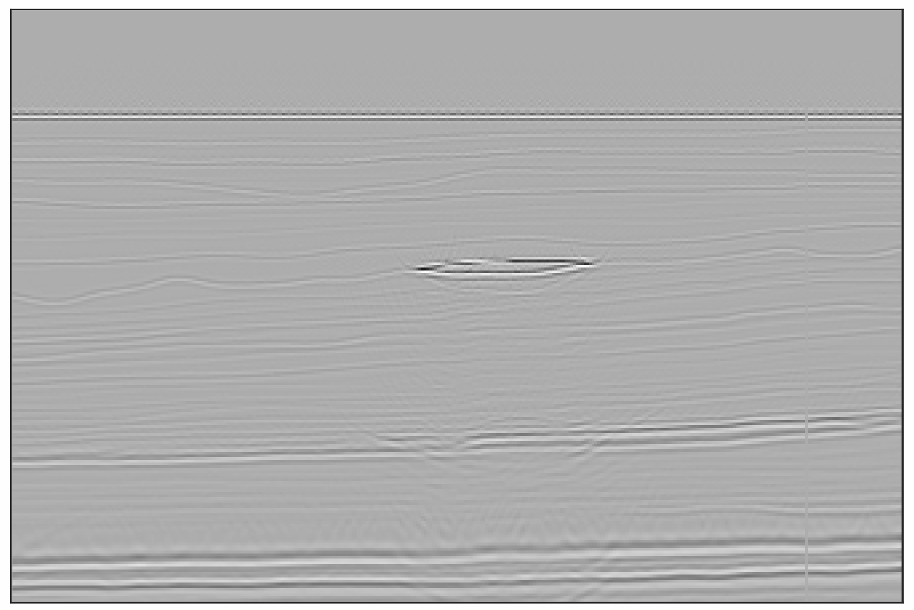
Shuey term inversion as a projection



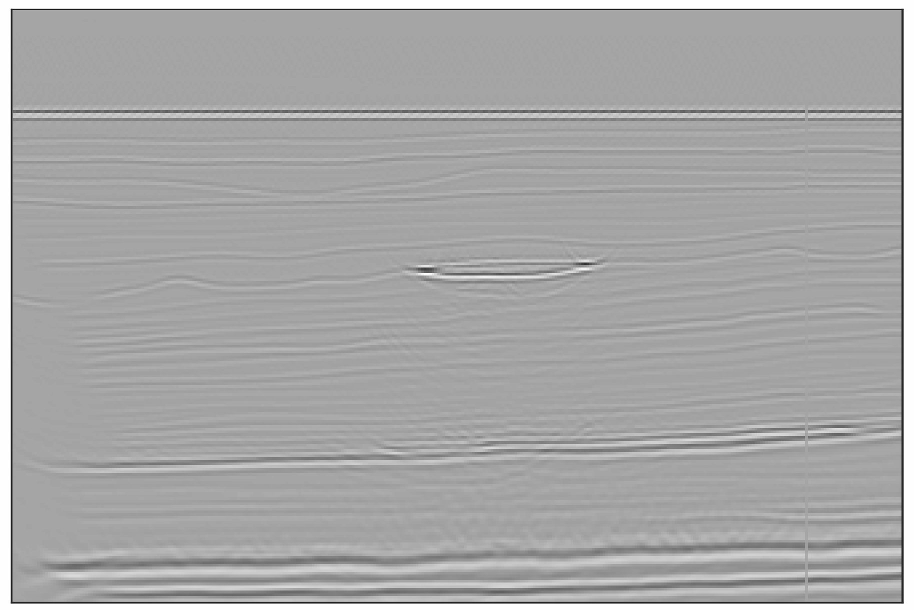


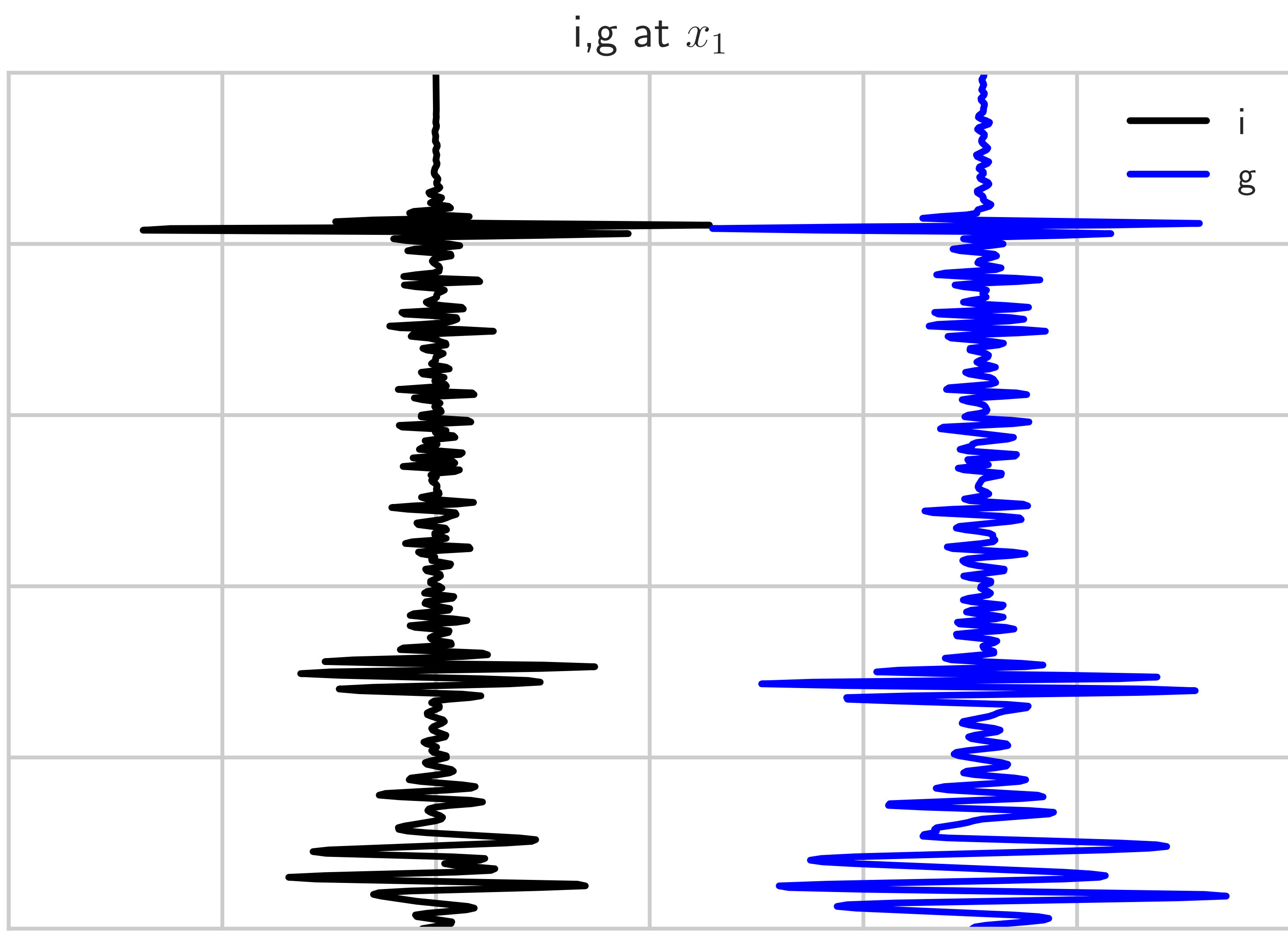
Shuey term inversion as a projection

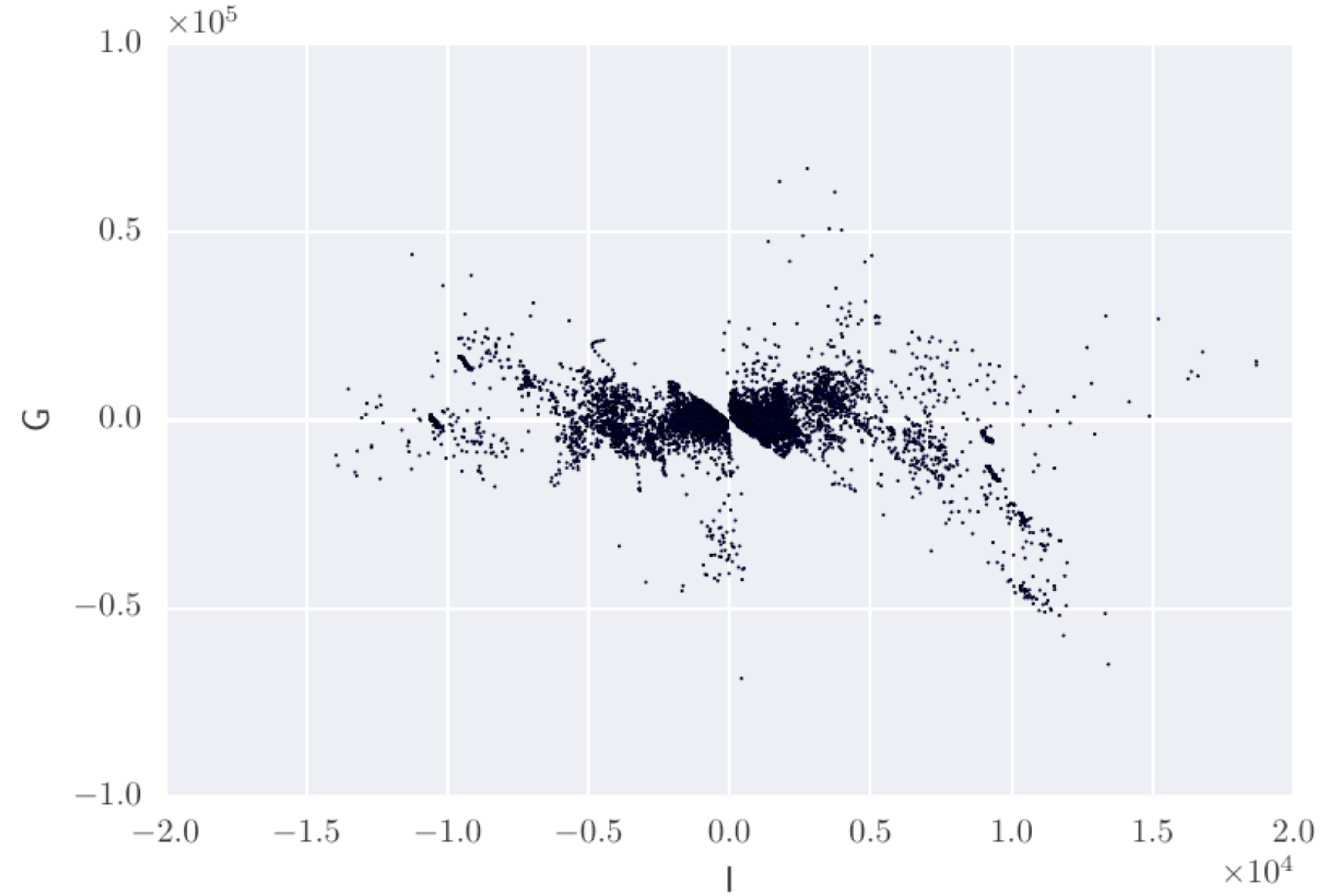
i



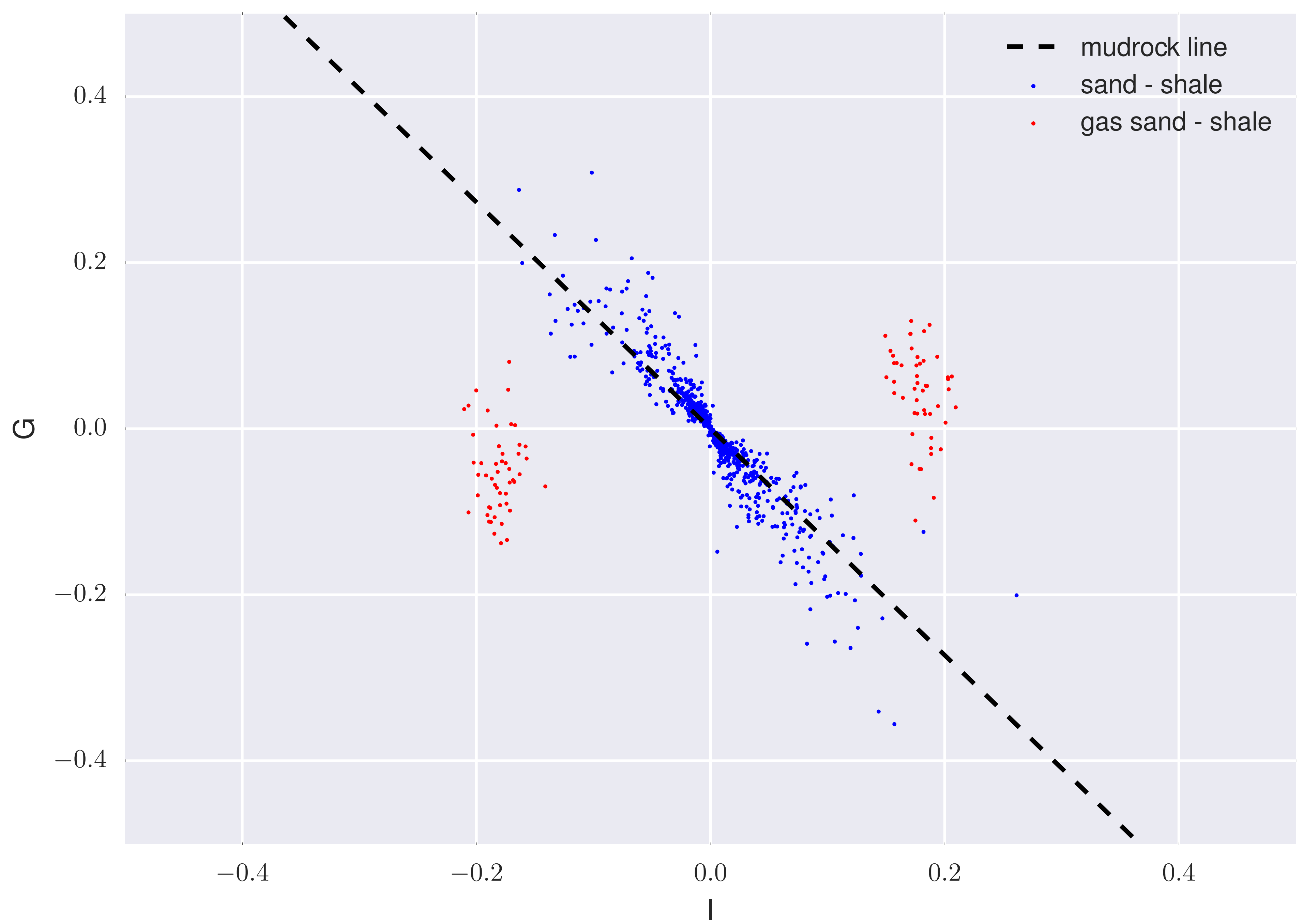
g







Relation to hydrocarbons



$$\frac{\Delta\rho}{\langle\rho\rangle} \propto k\frac{\Delta V_P}{\langle V_P\rangle}$$

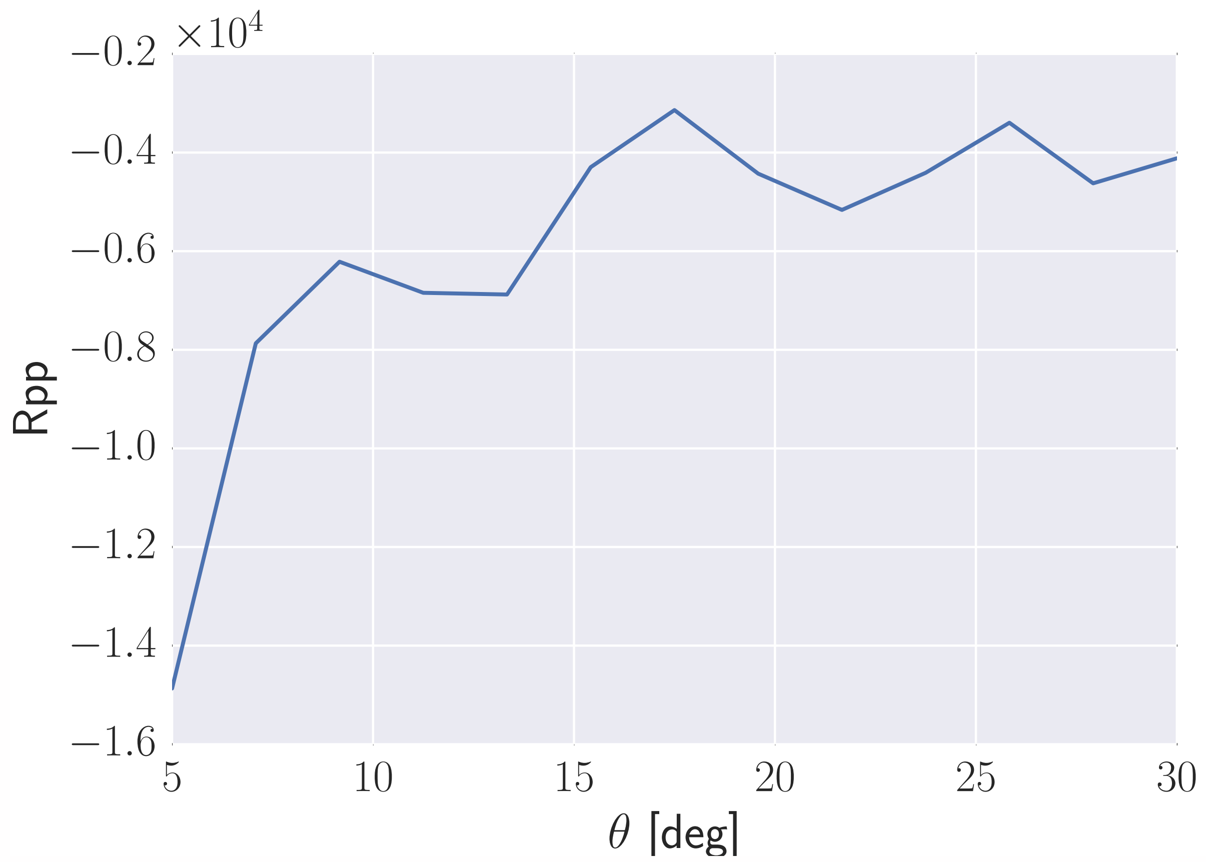
$$V_P = mV_S + c$$

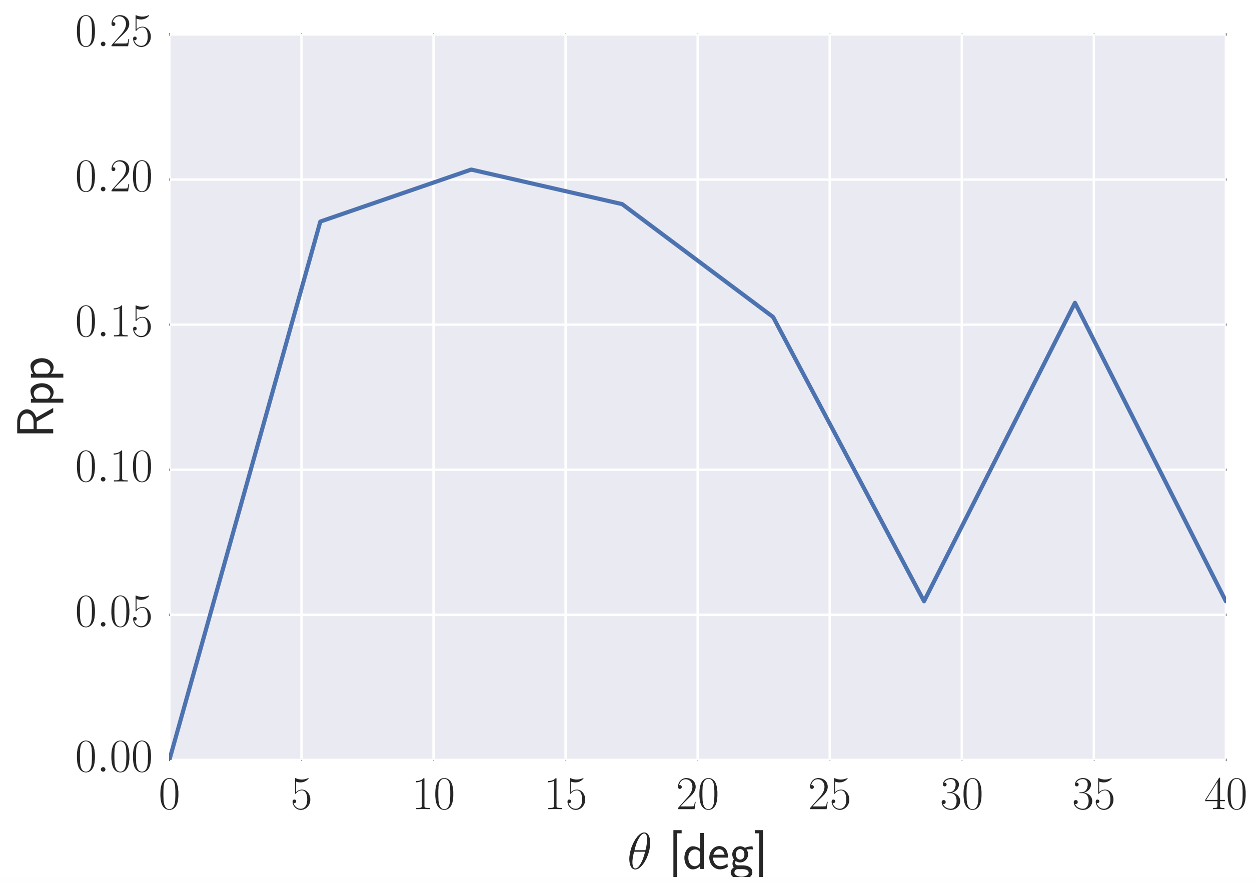
$$g = \frac{i}{1+k}\left[1 - 4\frac{\langle V_S\rangle}{\langle V_P\rangle}\left(\frac{2}{m} + k\frac{\langle V_S\rangle}{\langle V_P\rangle}\right)\right]$$

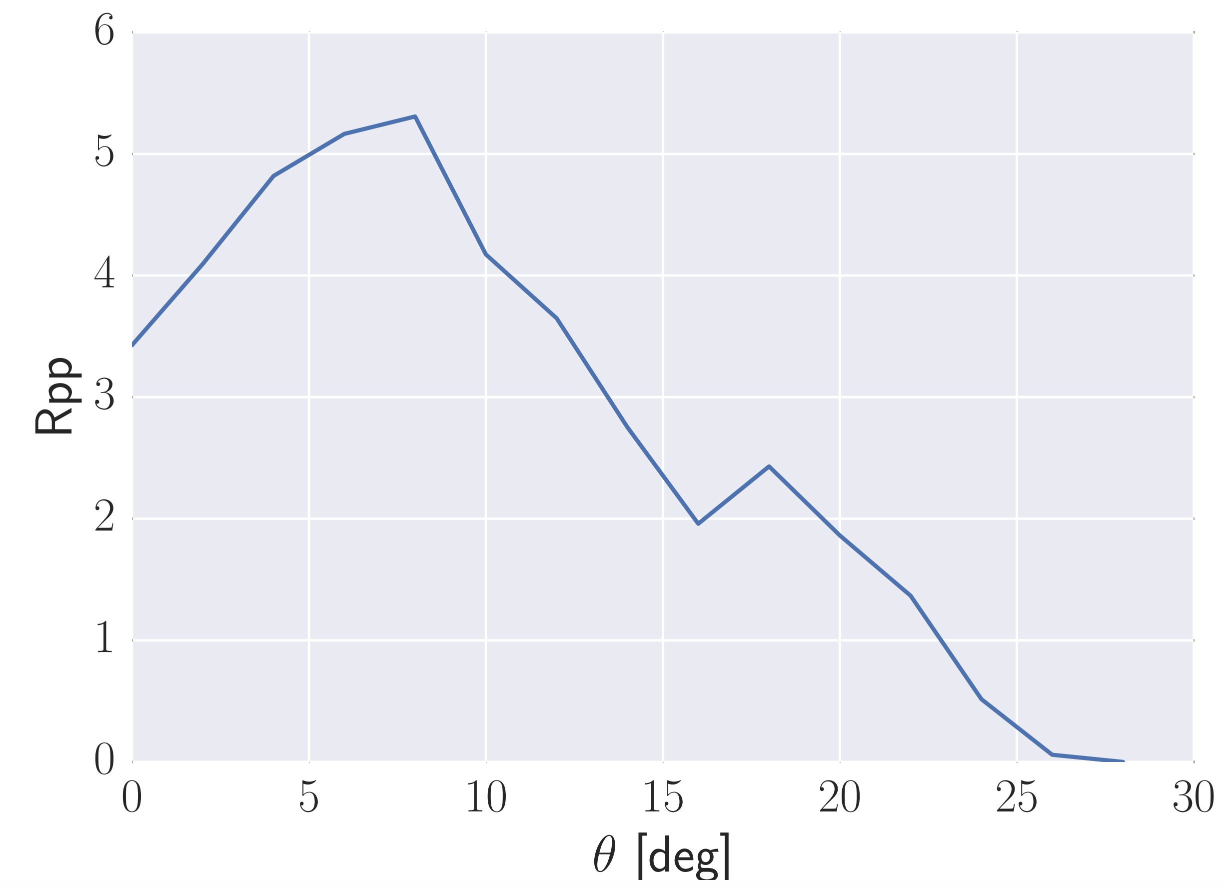
Hydrocarbon reserves are found from outliers of a crossplot!

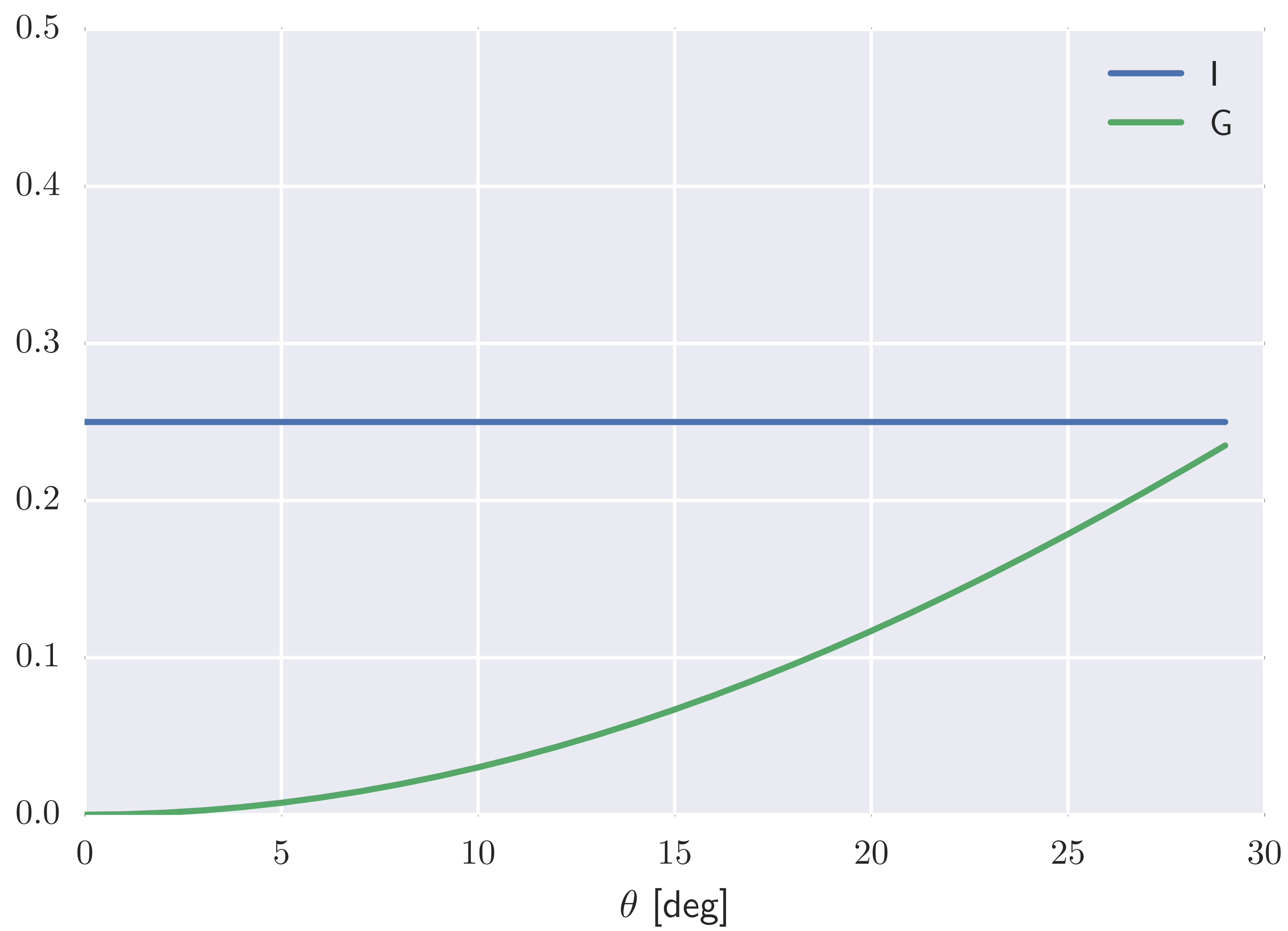
*m, c, k are geological parameters determined empirically from well logs/laboratory measurements

Reality bites









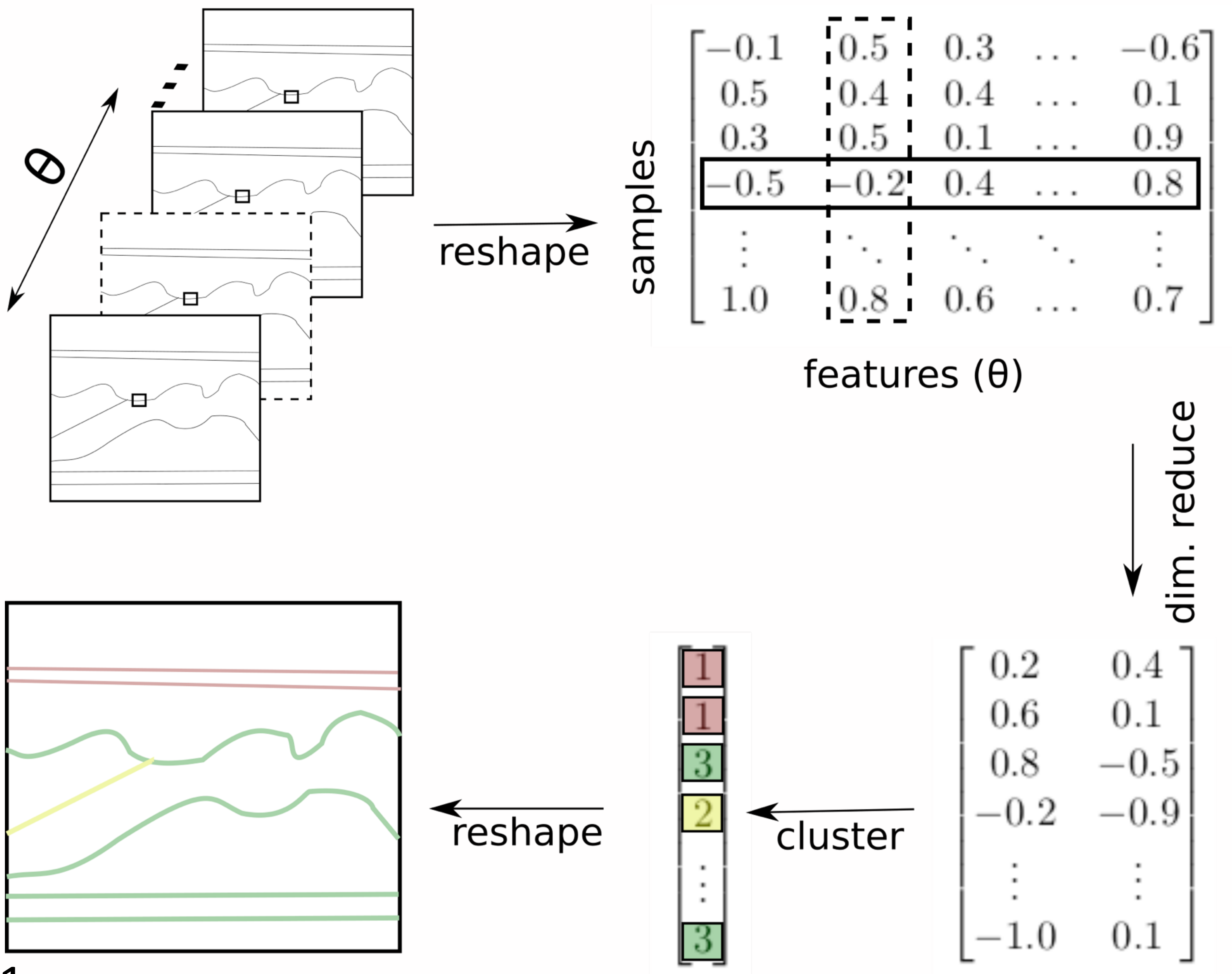
Problem:

Shuey components can't explain all the features in real data!

Solution:

Use unsupervised machine learning to find better projections.

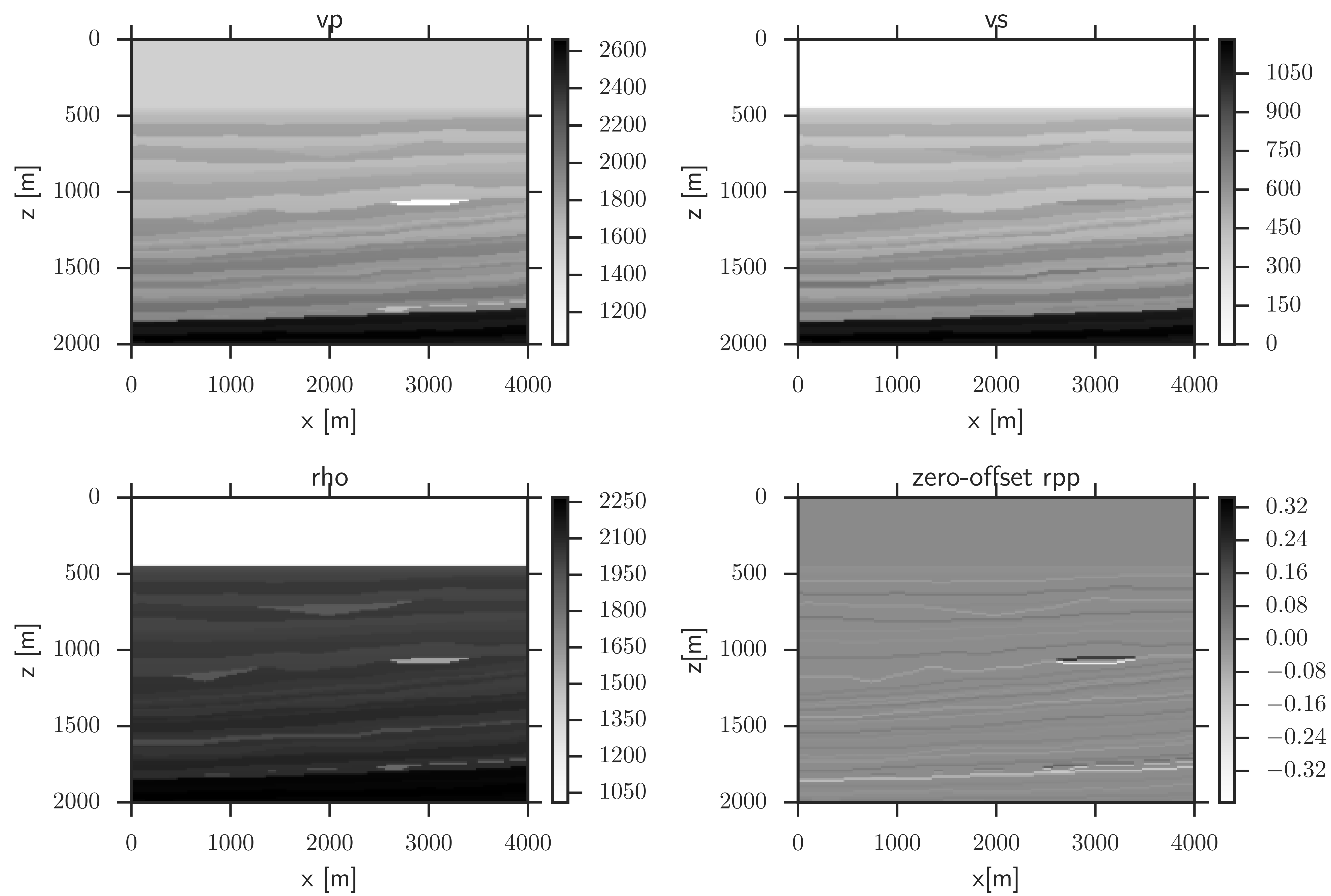
Unsupervised learning problem



$$X \in \mathbb{R}^{n \times d}$$

n is number of samples in the image,
d is the number of angles

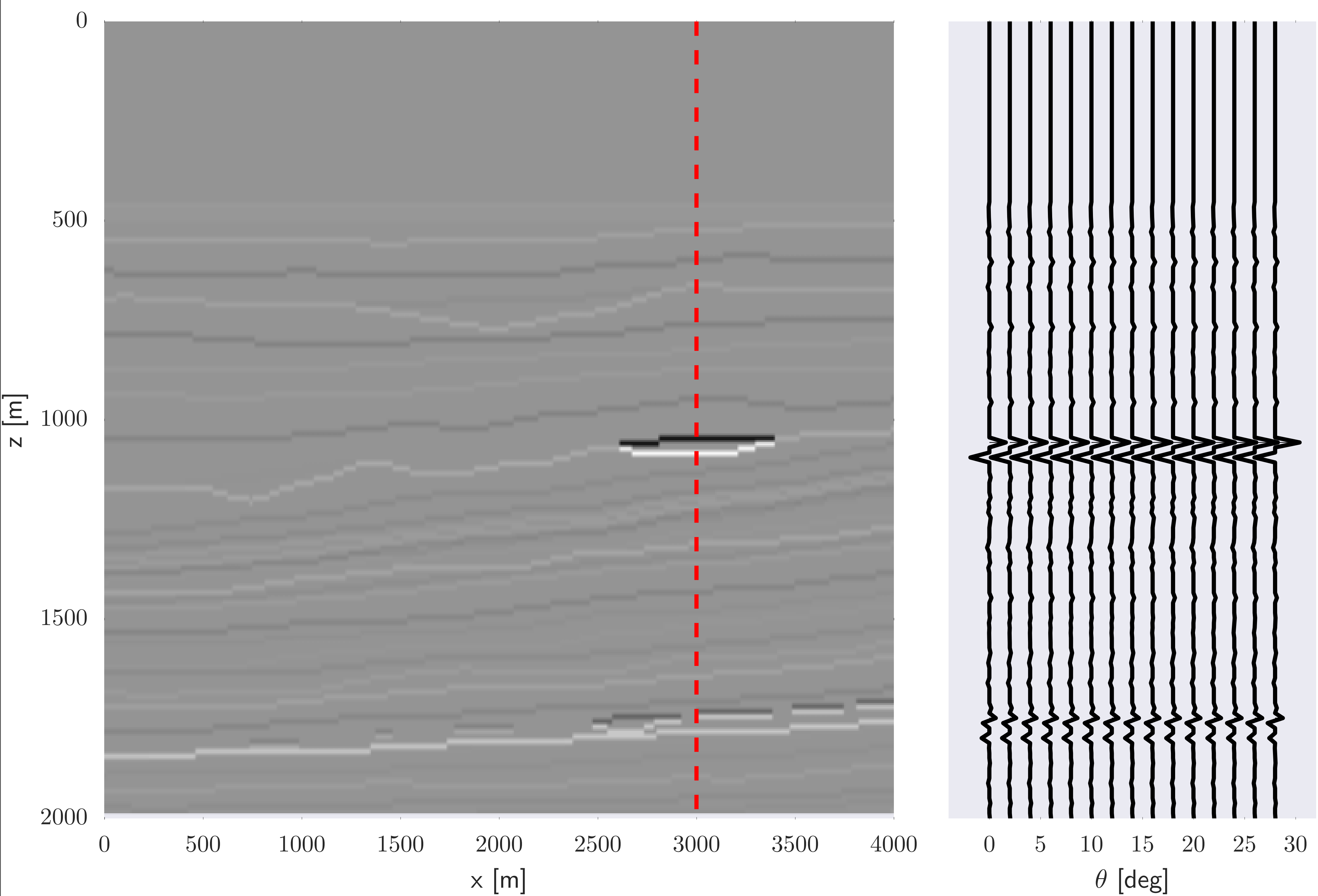
Marmousi II data



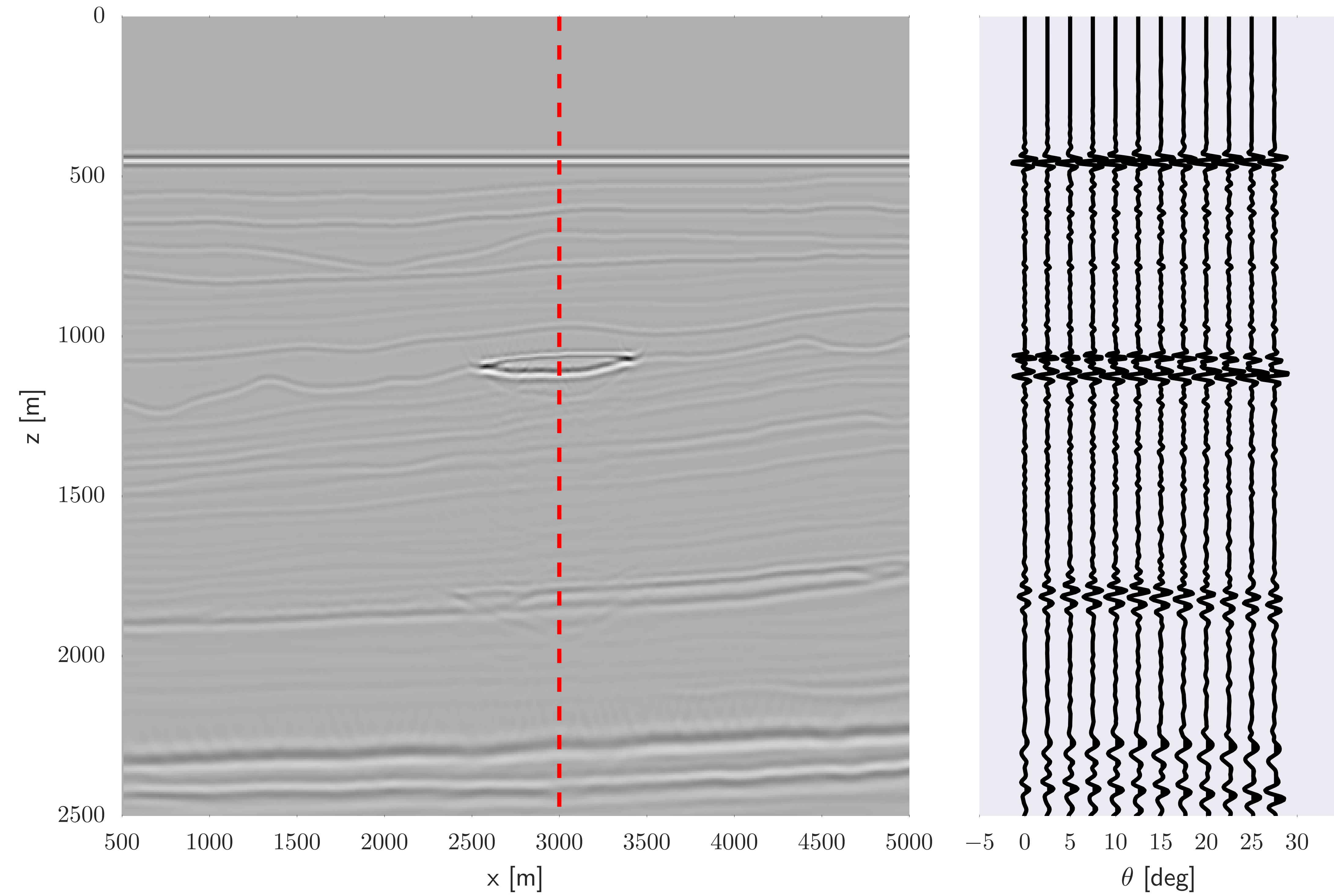
Specifically made for testing amplitude vs. offset analysis

Contains gas-saturated sand

Seismic modeling



Physically consistent with Zoeppritz



Migrated visco-acoustic survey

Principle component analysis (PCA)

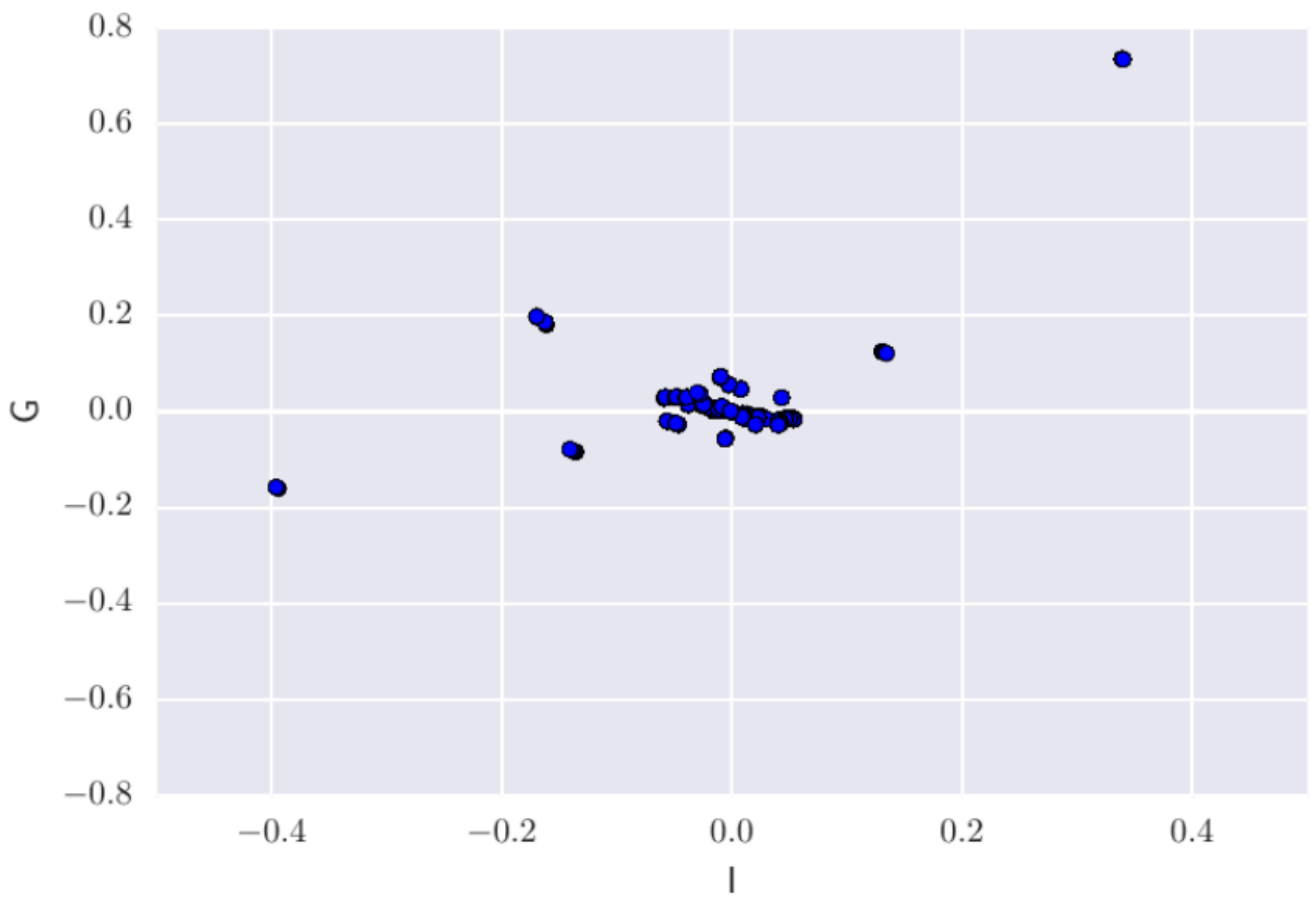
Eigendecomposition of the covariance matrix:

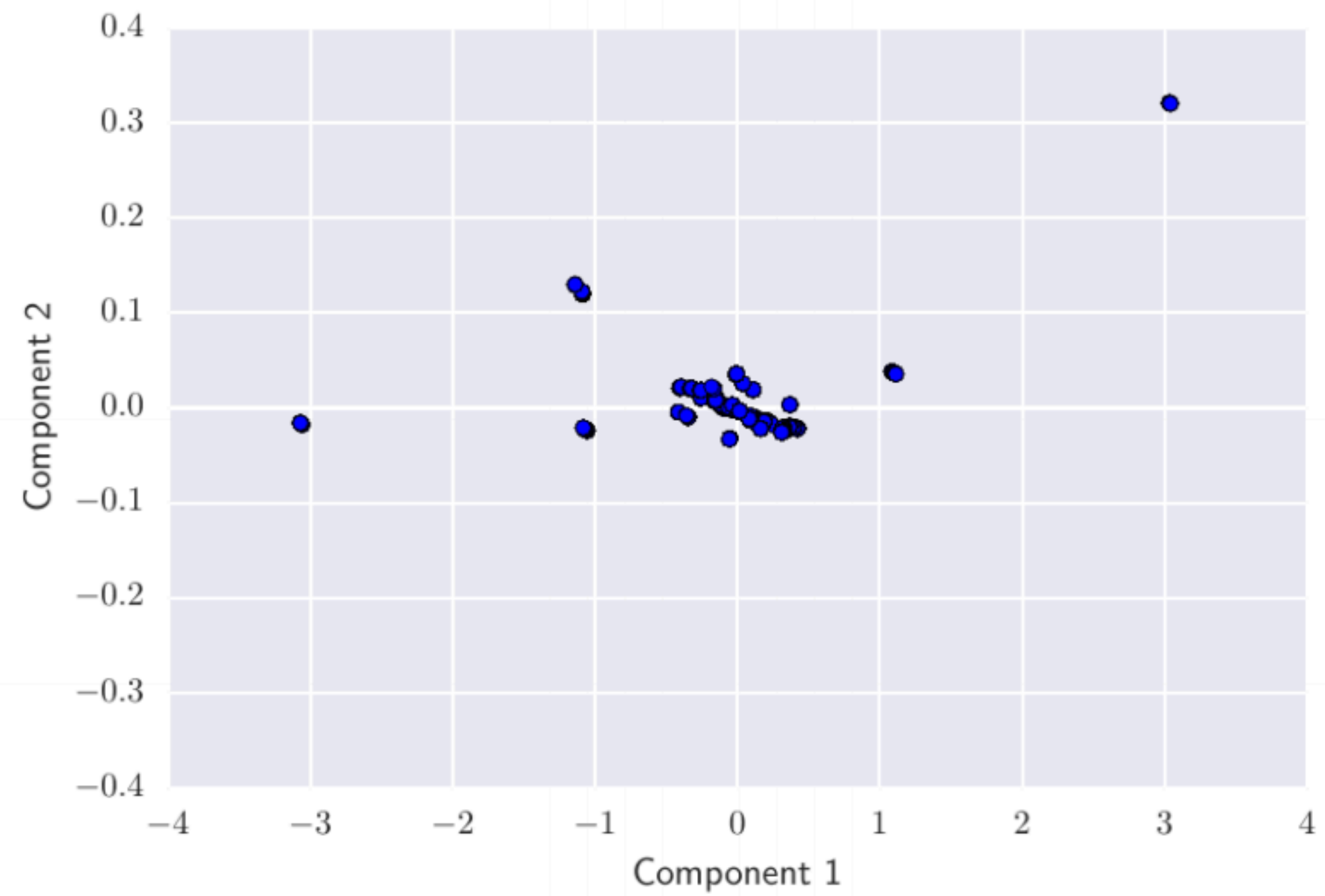
$$C = X^T X = \frac{1}{n} \sum_{i}^{n} \mathbf{x}_i \mathbf{x}_i^T$$

Project onto the eigenvectors with the two largest eigenvalues.

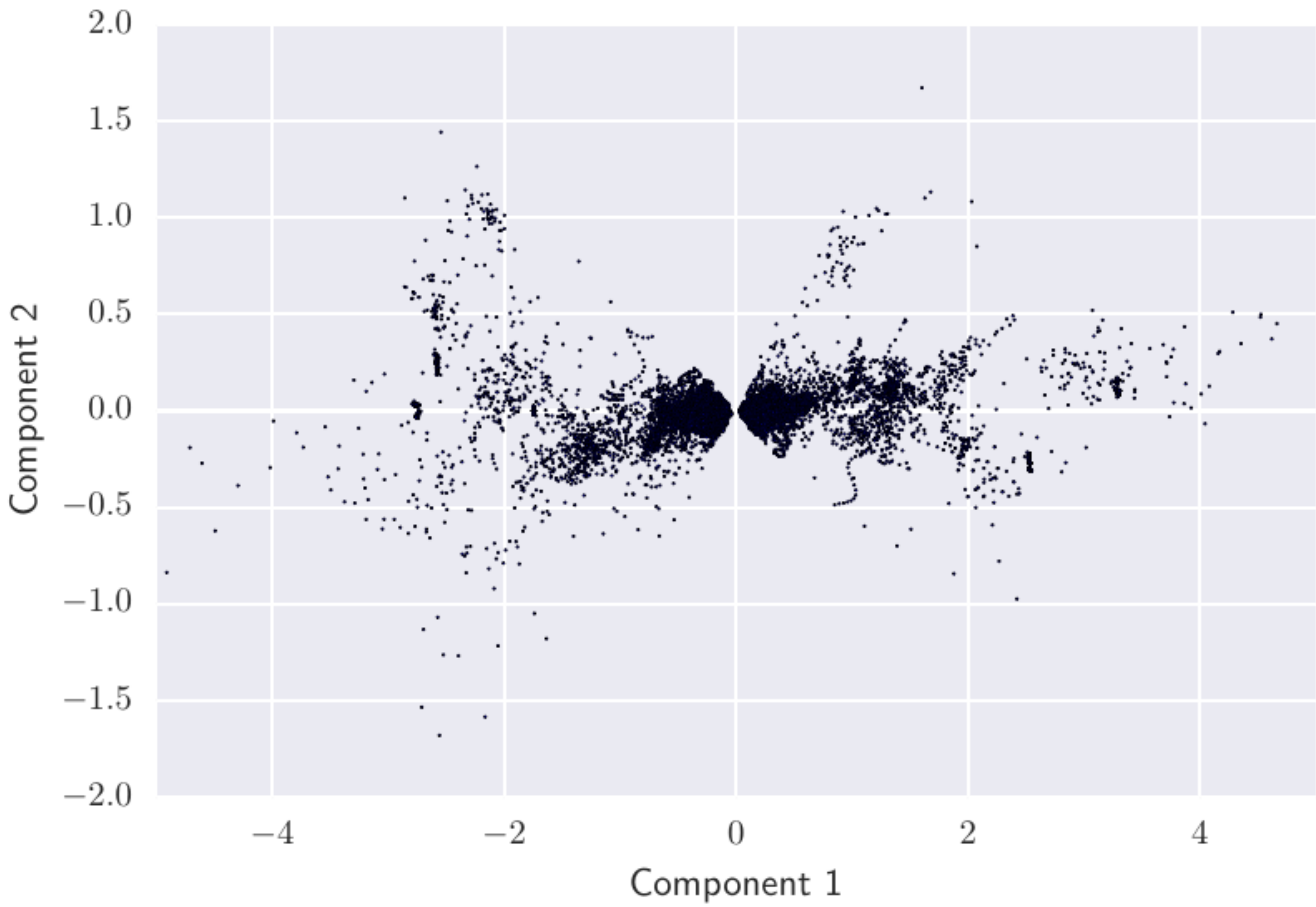
Maximizes the variance (a measure of information).

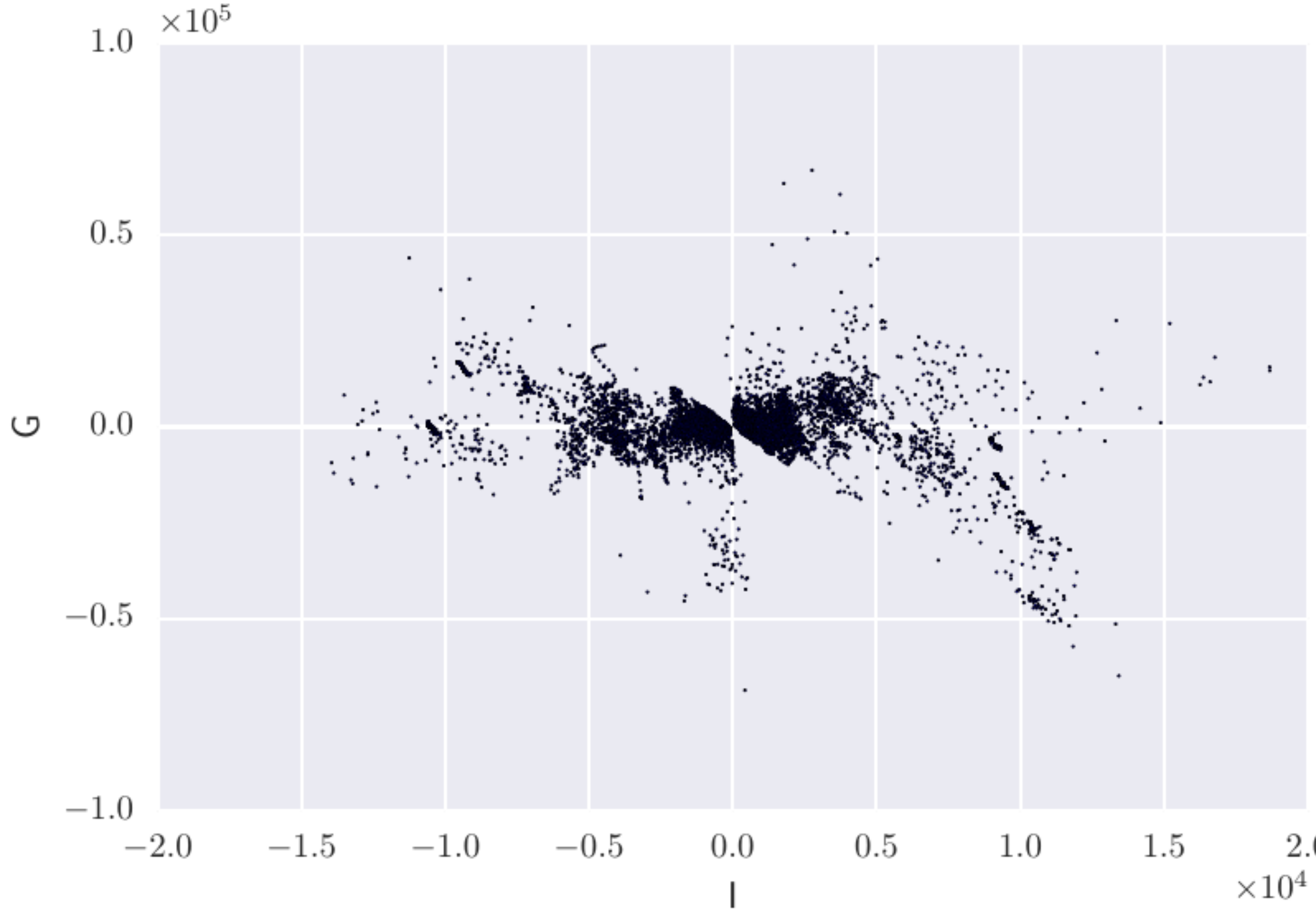
Crossplots of physically consistent data





Crossplots of migrated data





Kernel PCA

Problem:

Find a non-linear projection that provides better discrimination of trends and outliers.

Solution:

Use the “kernel-trick” to compute PCA in a high-dimensional non-linear feature space

Kernel trick

PCA can be calculated from the Gramian inner product matrix:

$$XX^T = \begin{pmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \dots & \langle \mathbf{x}_1, \mathbf{x}_n \rangle \\ \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \dots & \langle \mathbf{x}_2, \mathbf{x}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{x}_n, \mathbf{x}_1 \rangle & \langle \mathbf{x}_n, \mathbf{x}_2 \rangle & \dots & \langle \mathbf{x}_n, \mathbf{x}_n \rangle \end{pmatrix}$$

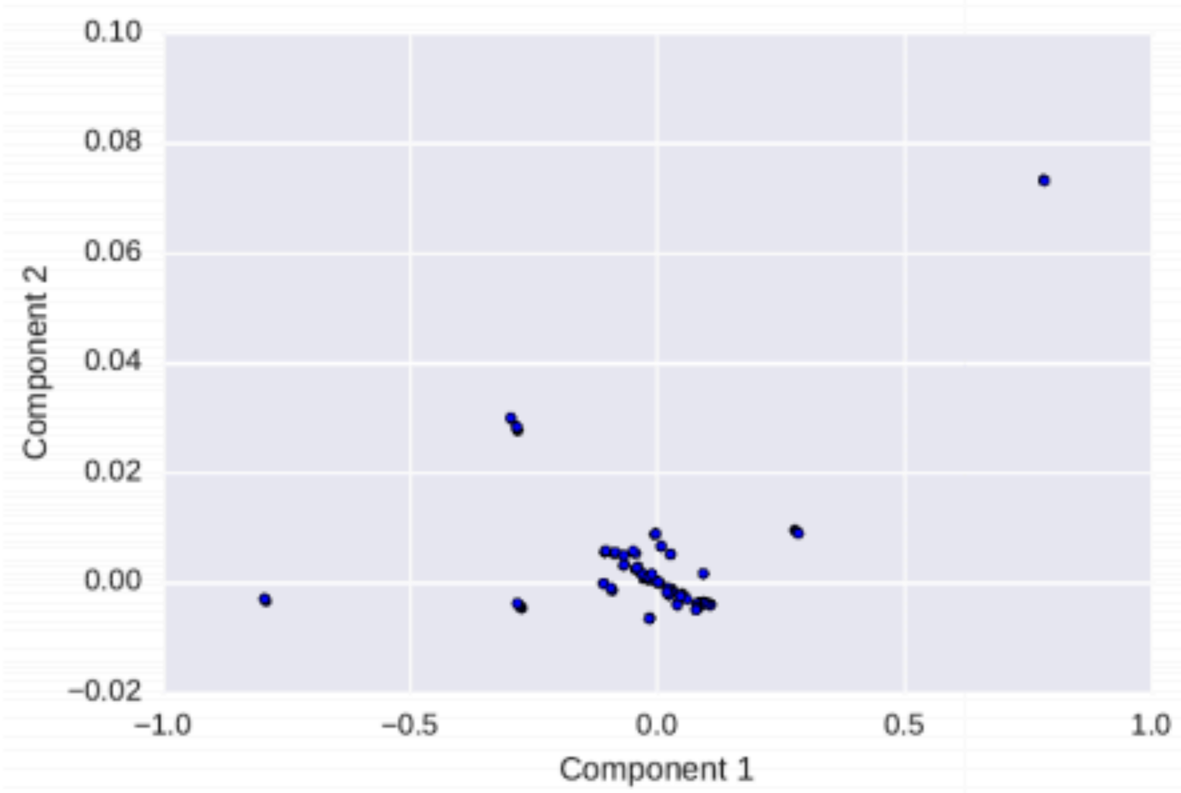
Replace $\langle \mathbf{x}_i, \mathbf{x}_j \rangle$ with a kernel $\kappa(\mathbf{x}_i, \mathbf{x}_j)$

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + b)^c = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$$

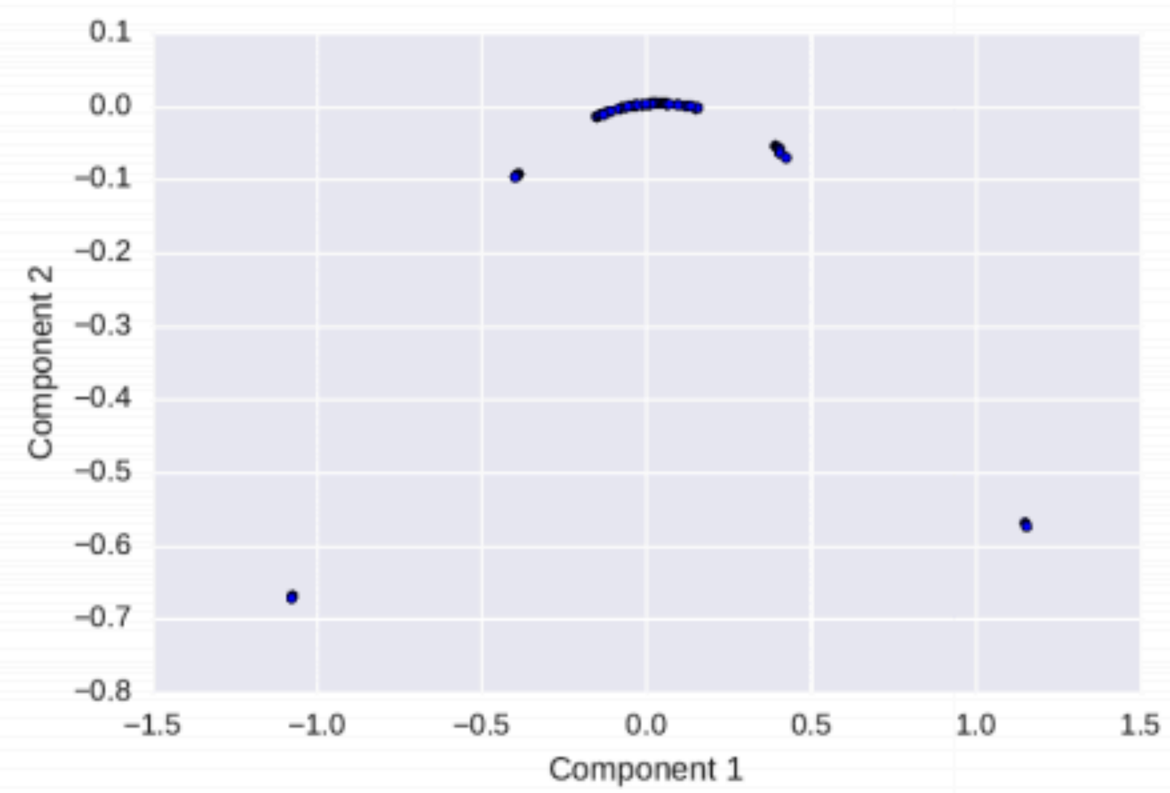
Example c=2, b=1

$$\phi(\mathbf{x}) = [1, \sqrt{2}x_1, \sqrt{2}x_2, x_1^2, x_2^2, \sqrt{2}x_1x_2]$$

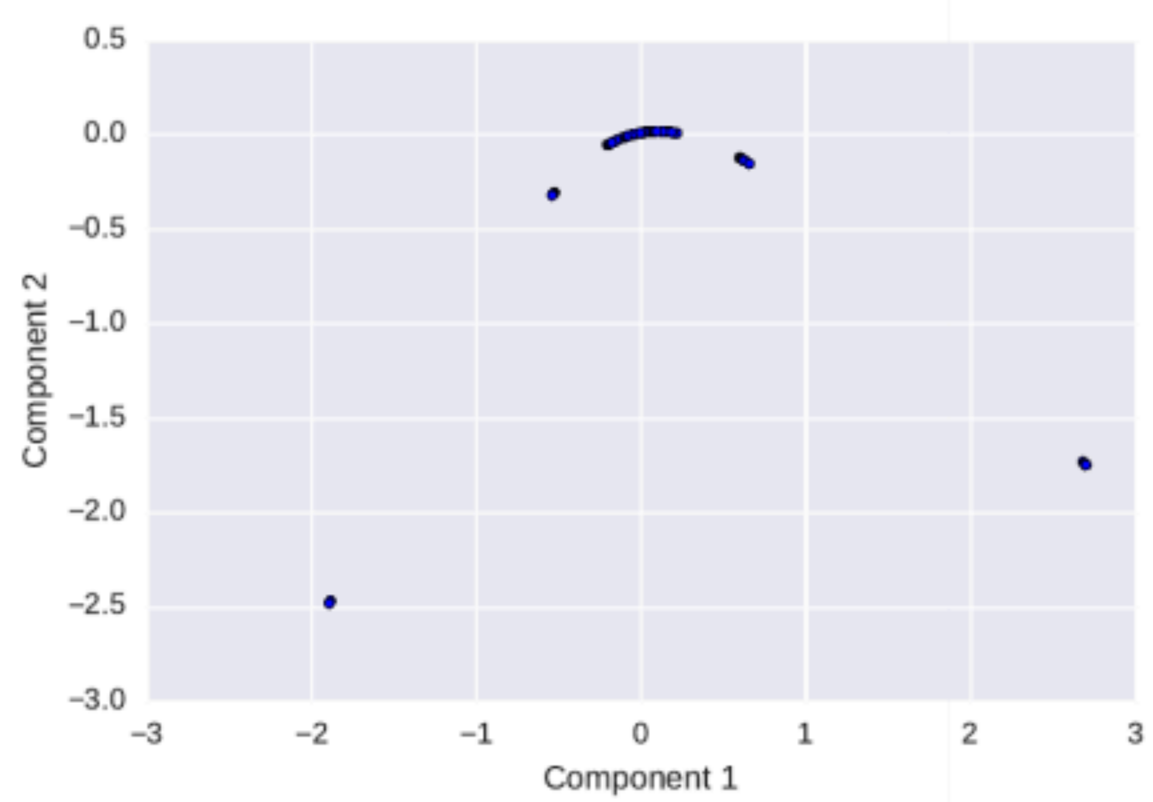
Kernel PCA projections for consistent data



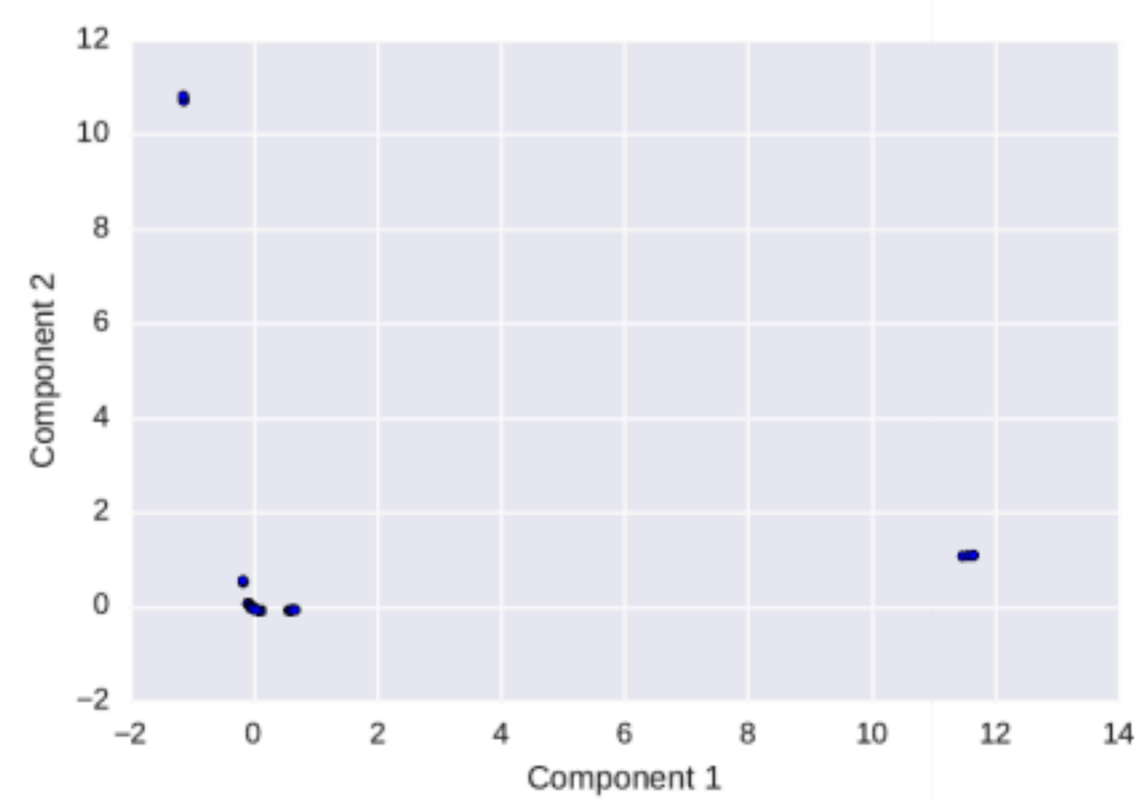
$c = 1$



$c = 2$

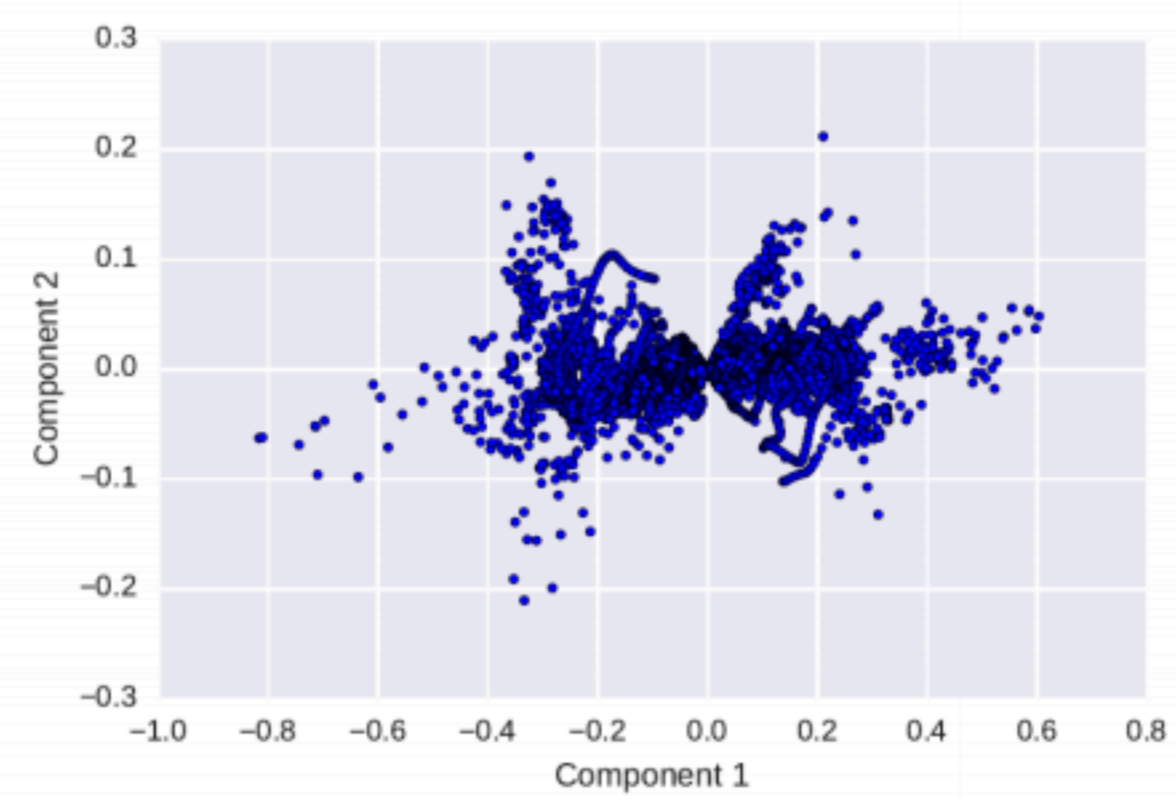


$c = 5$

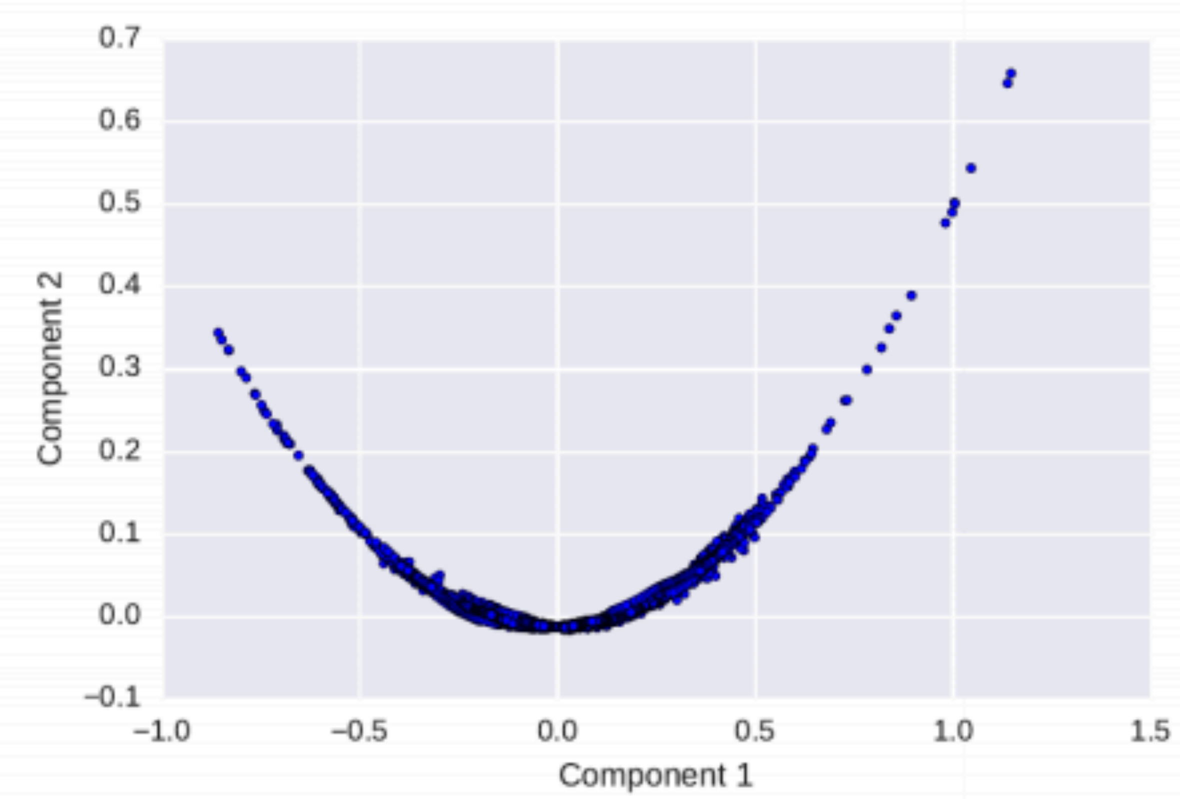


$c = 10$

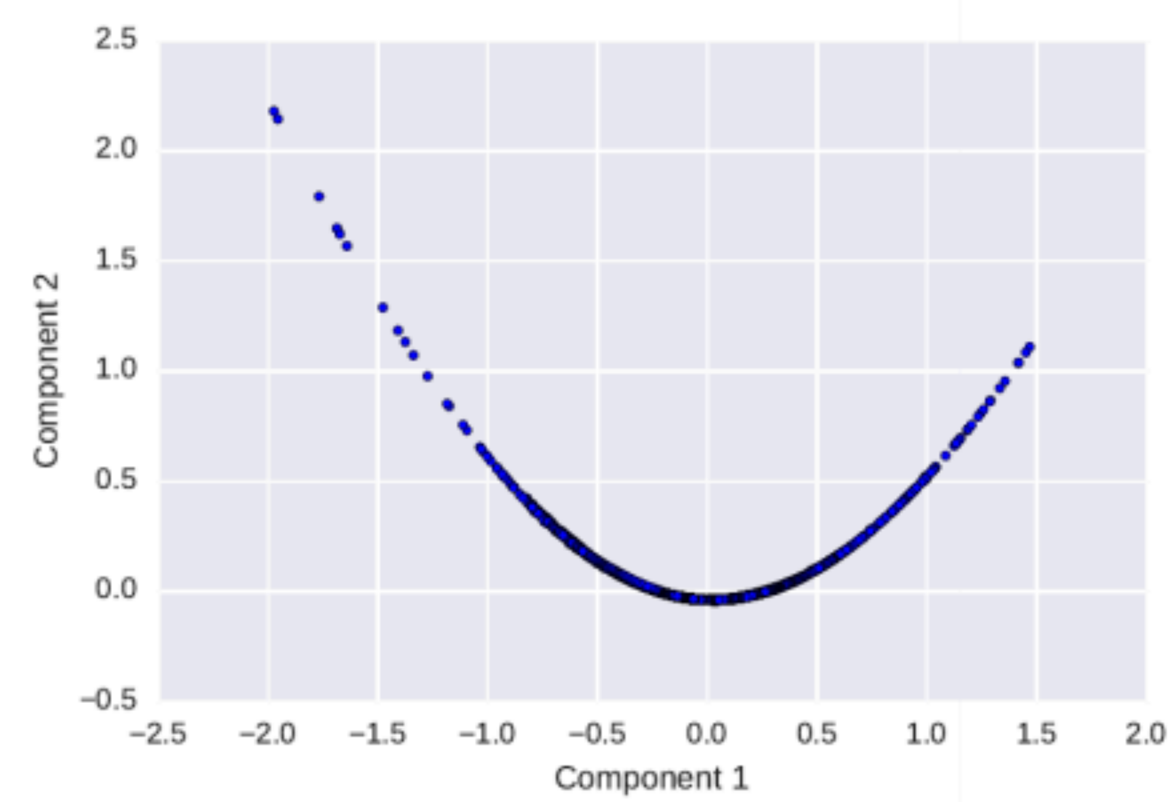
Kernel PCA projections for migrated data



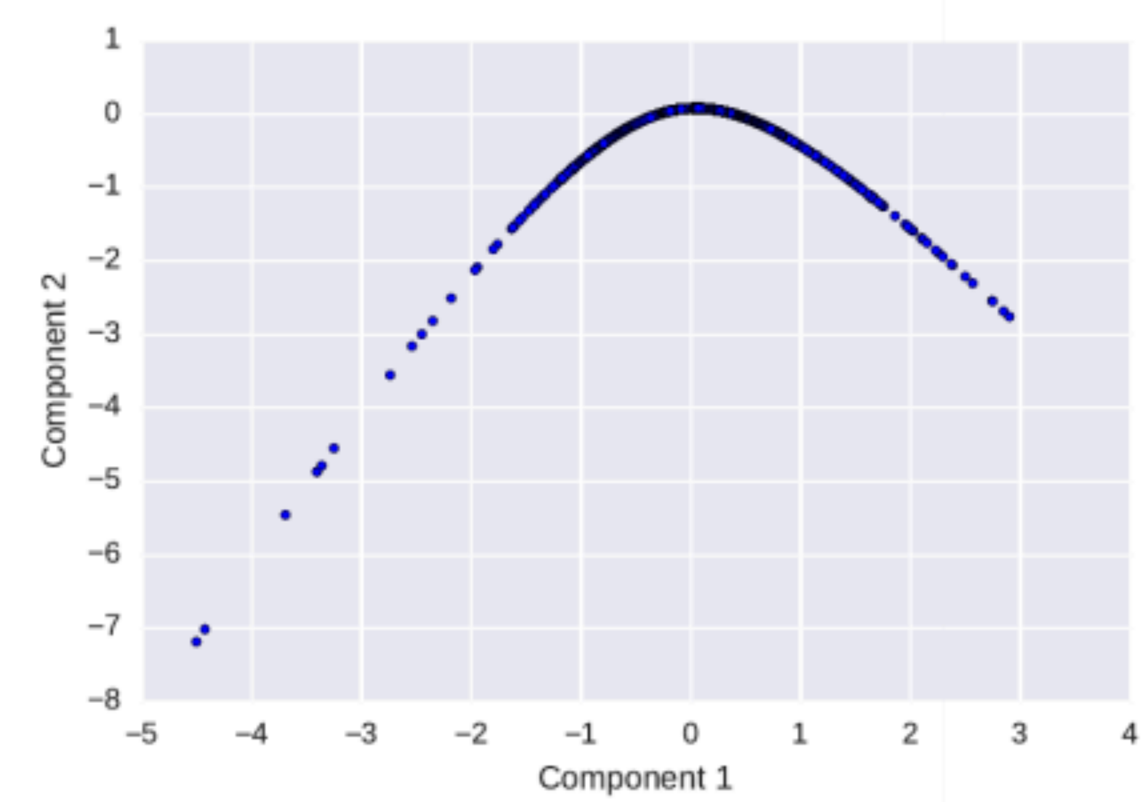
$c = 1$



$c = 2$



$c = 5$



$c = 10$

Recap

Situation: Finding outlying responses in projected seismic data.

Problem: Physical projection can not explain the features in real data.

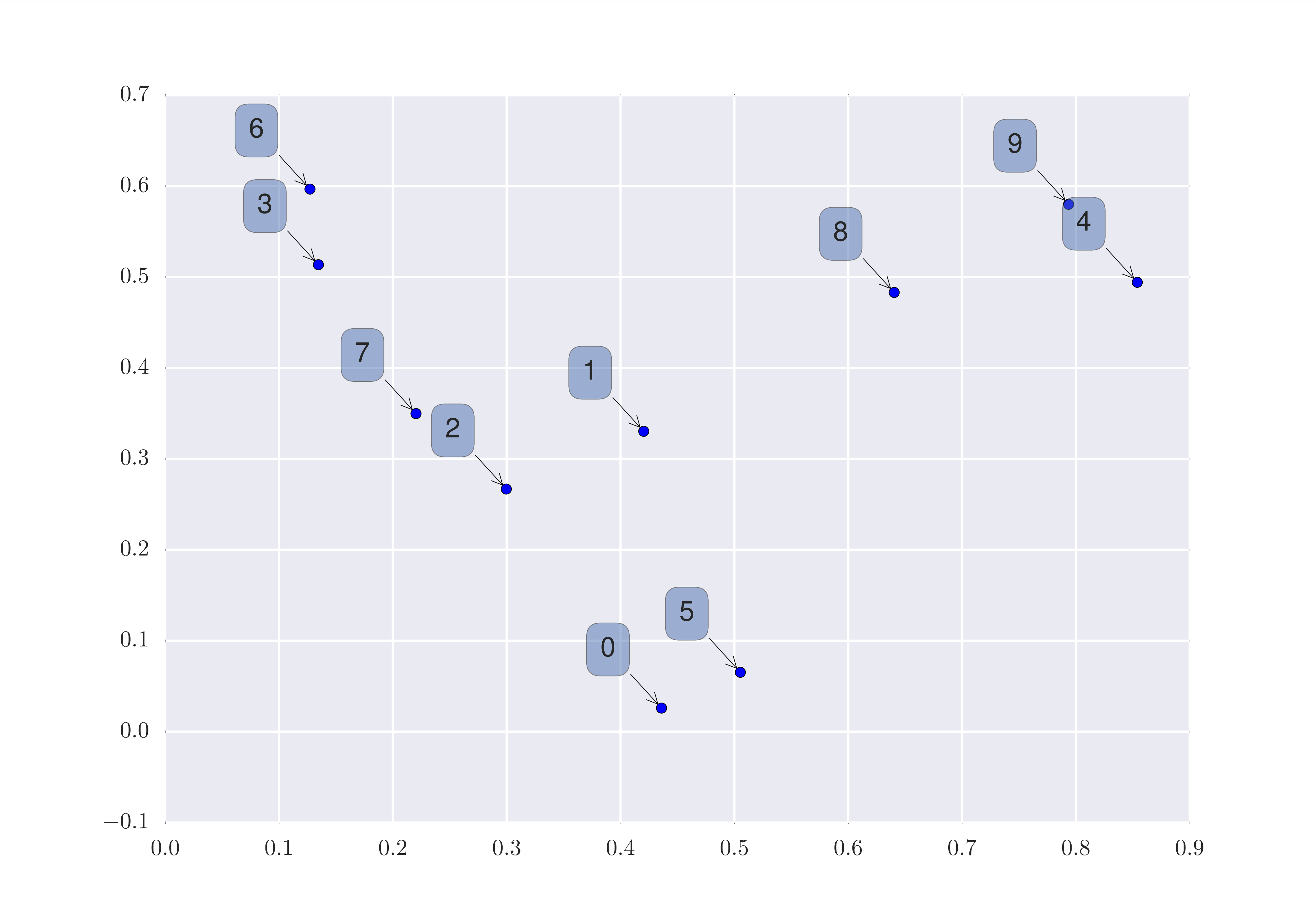
Solution: Use principle component-based projections.

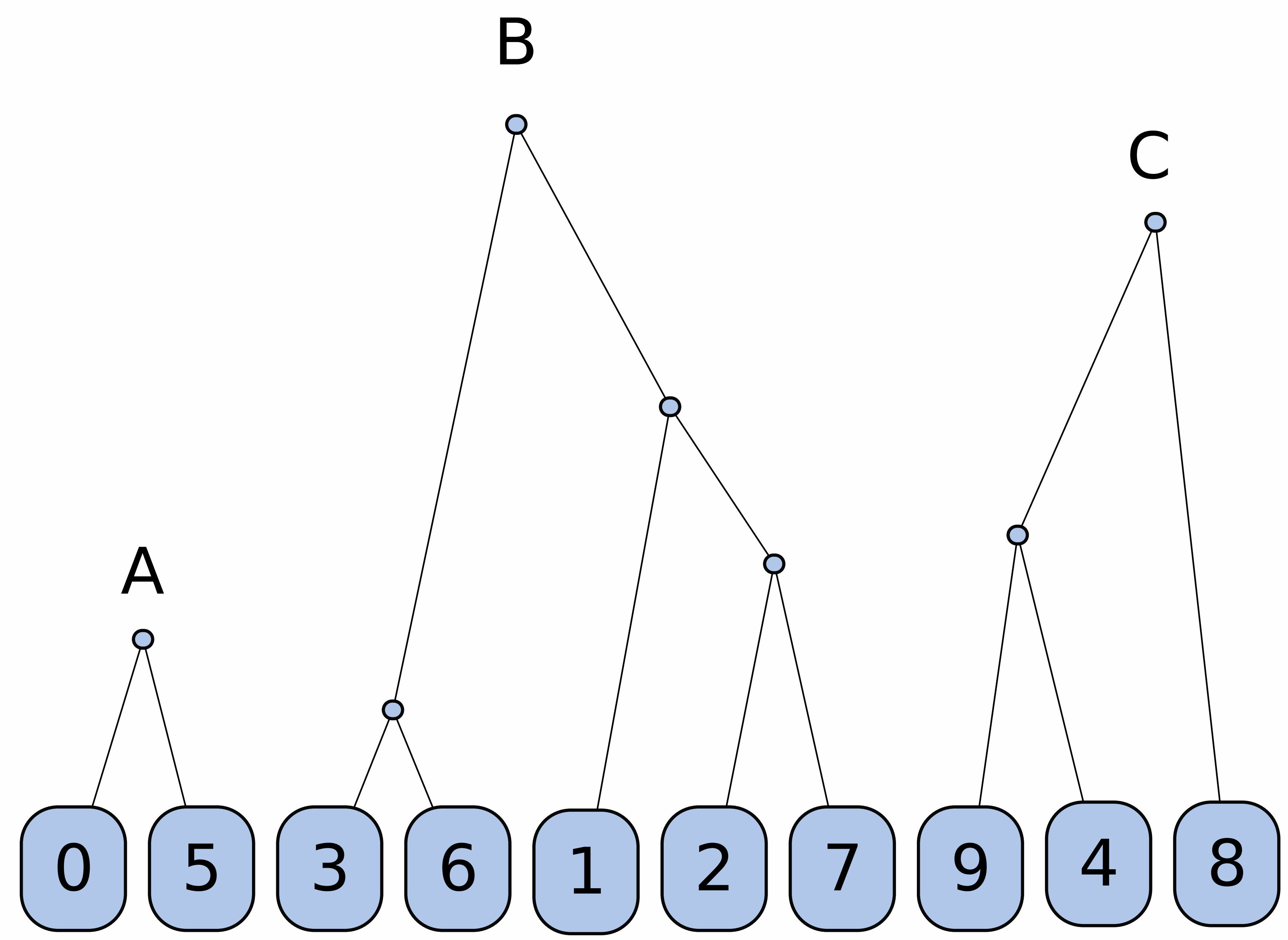
Assessment:

- PCA is equivalent for physically consistent data.
- PCA is more robust to processing/acquisition artifacts.
- Kernel PCA makes outliers linearly separable from the background.
- Lost direct link to rock physics.

Next: Automatic segmentation (clustering)

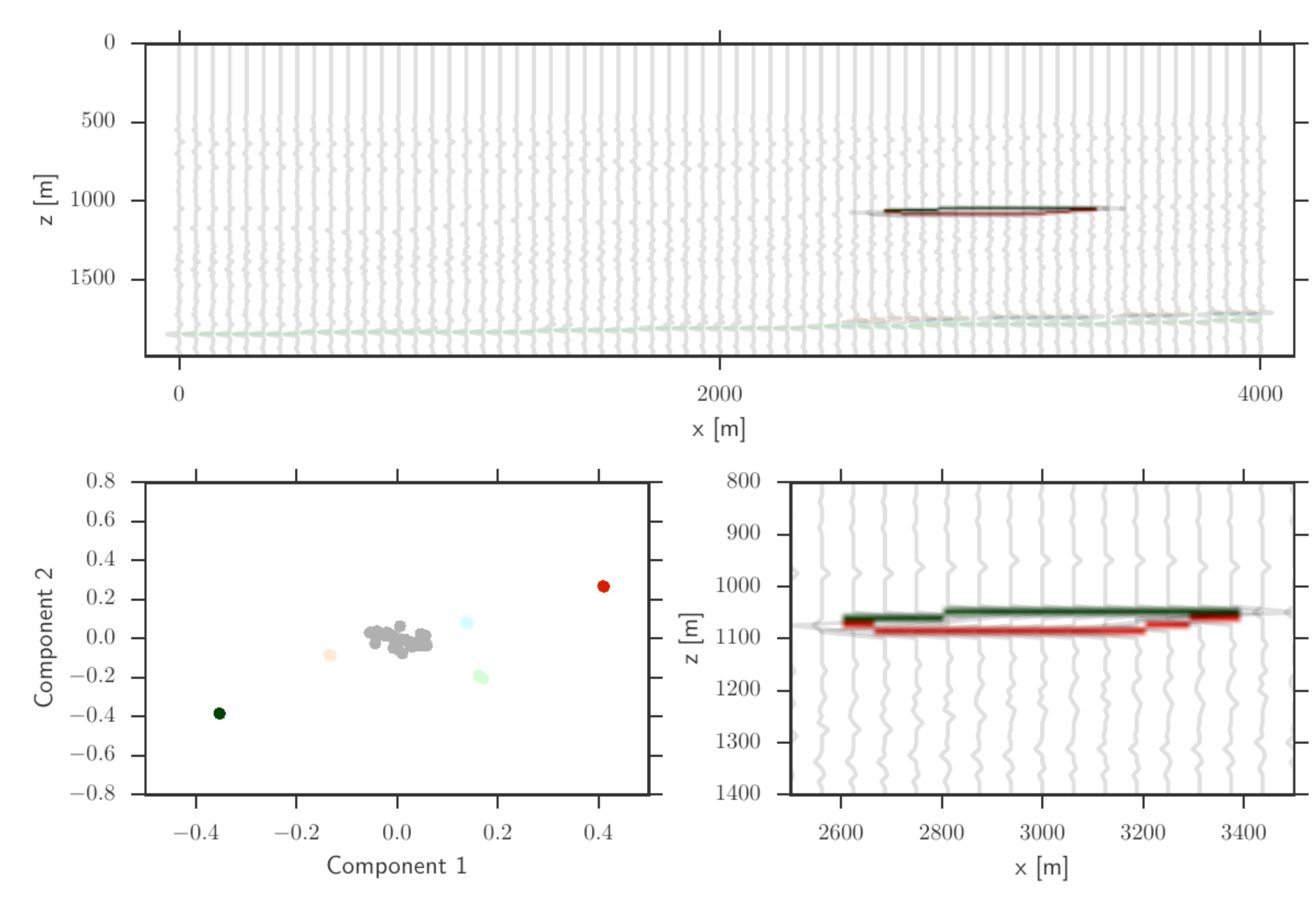
Hierarchical clustering

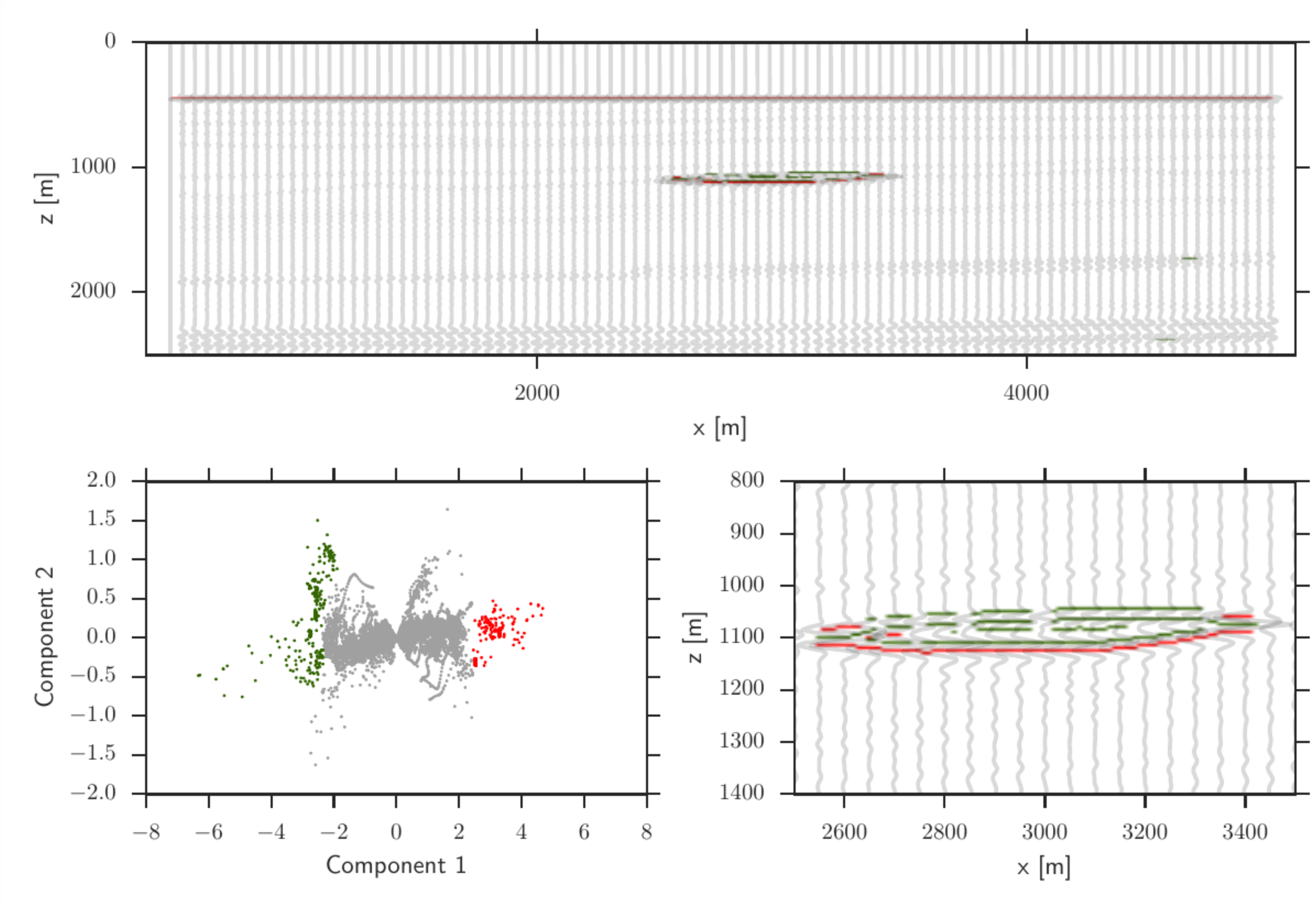




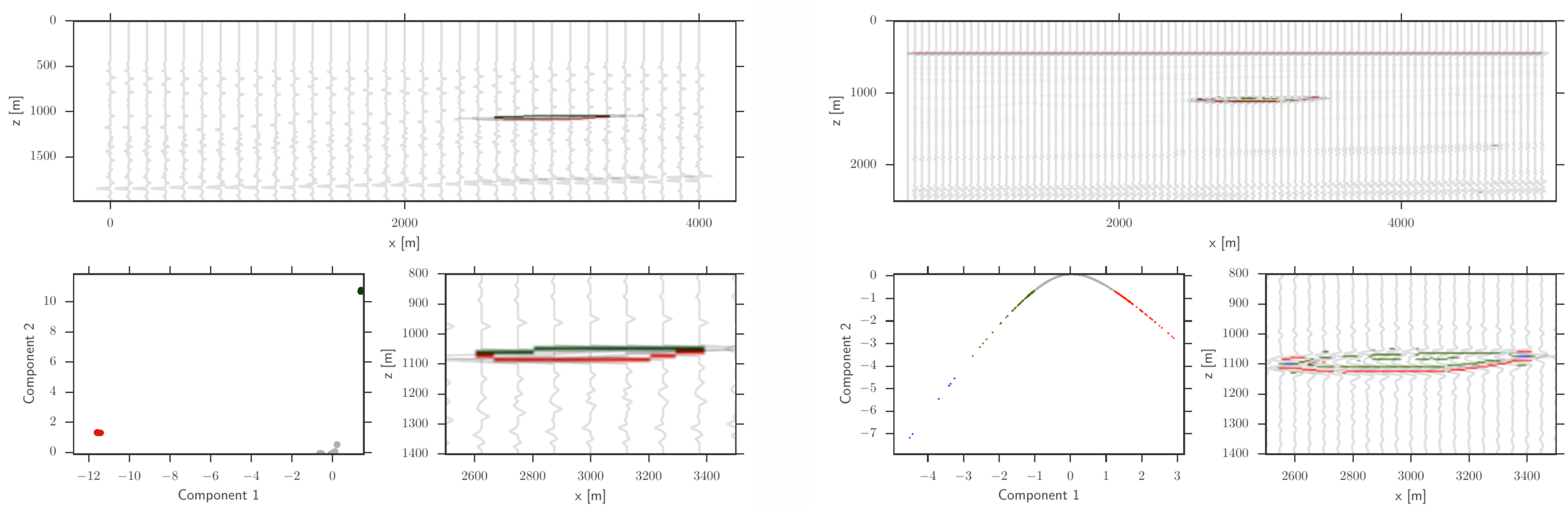
BIRCH clustering, developed specifically for clustering large databases

Results - PCA


z [m]
x [m]
Component 2
Component 1
z [m]
x [m]


z [m]
x [m]
Component 2
Component 1
z [m]
x [m]

Results - Kernel PCA



Summary

Successes:

Each projection could segment the reservoir.

Kernel PCA provided advantageous multivariate geometries (linearly separable).

Challenges:

Manual tuning of clustering parameters.

Kernel PCA is expensive and lacks interpretation.

Robust PCA

Problem:

Find a sparse set of outlying reflectivity responses against a background trend.

Assumption:

The background trend of similar curves is highly redundant, which forms a low rank matrix.

Solution:

$$L + S = X$$

Convex relaxation

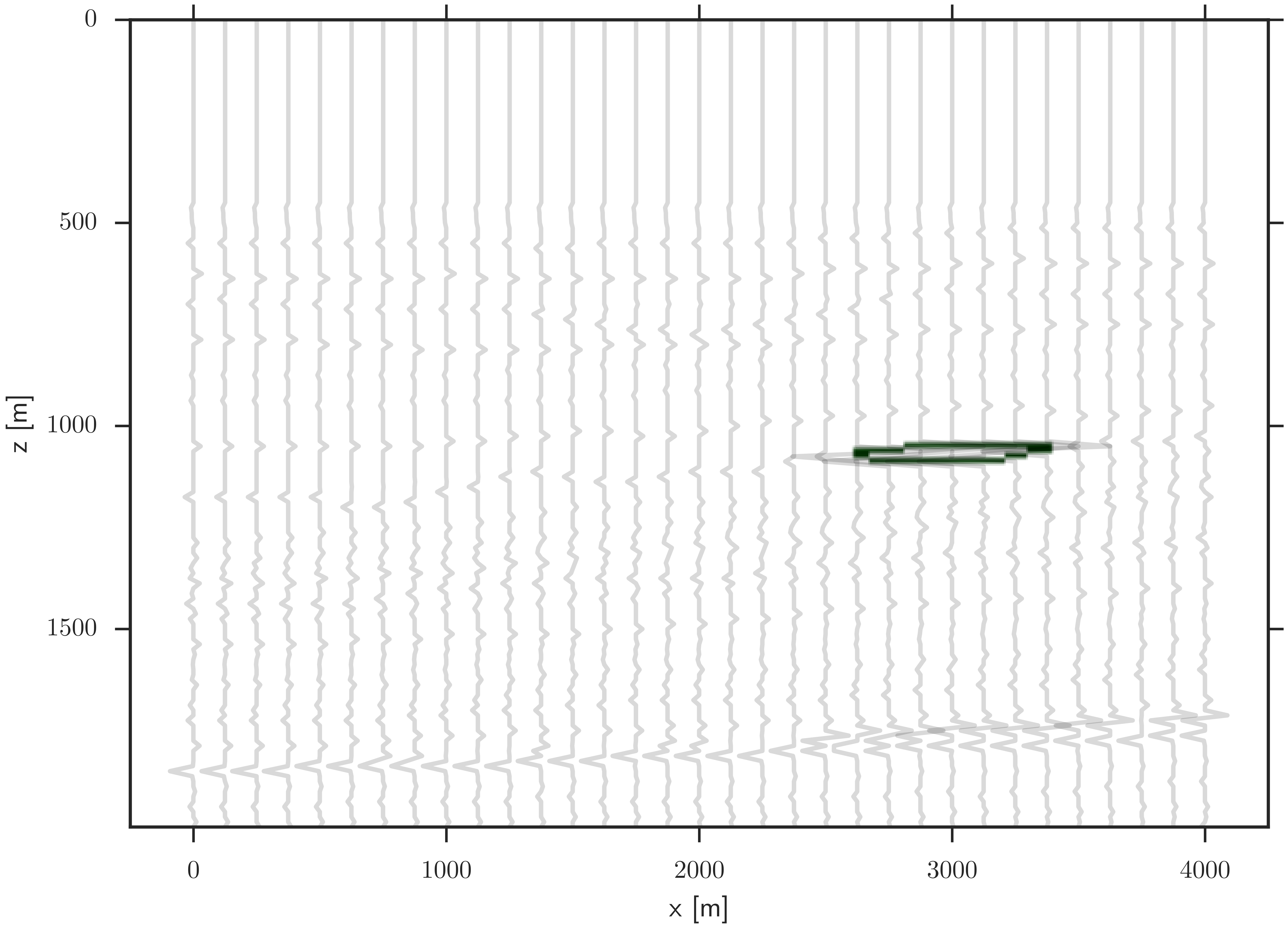
Rewrite with sparsity promoting convex terms:

$$\min_{L,S} \quad \|L\|_* + \lambda \|S\|_{1,\infty} \text{ s.t. } \quad L + S = X$$

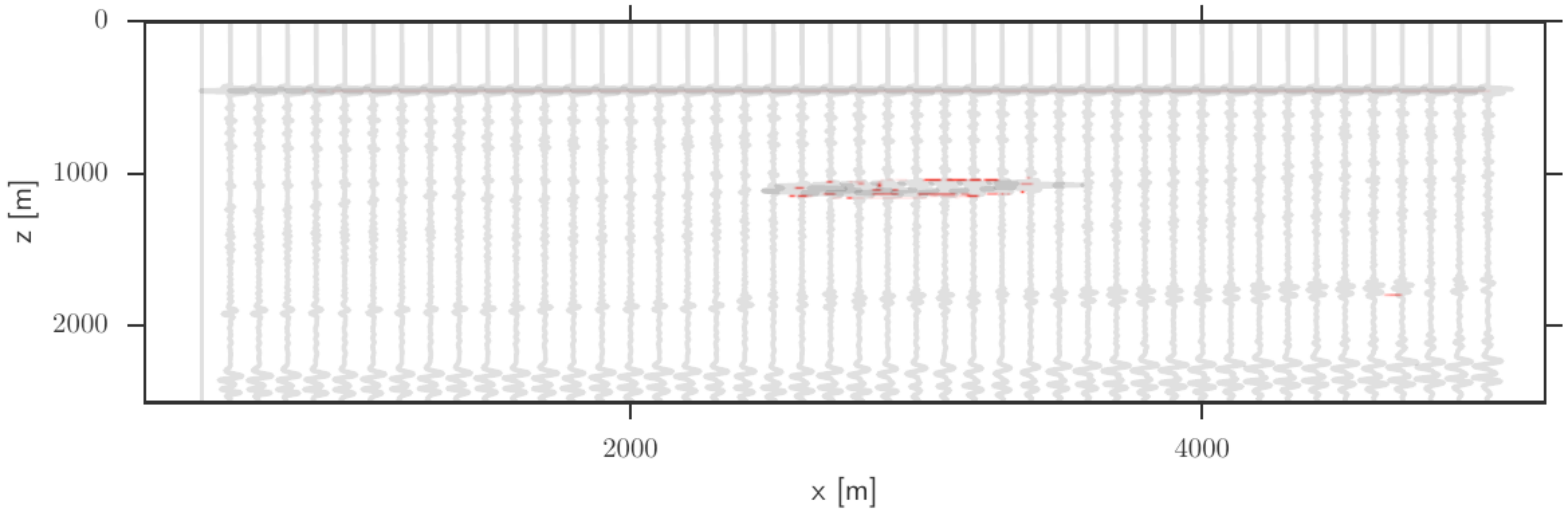
$$\|L\|_* = \text{trace}(\sqrt{L^* L}) = \sum_i \sigma_i = \|\sigma\|_1$$

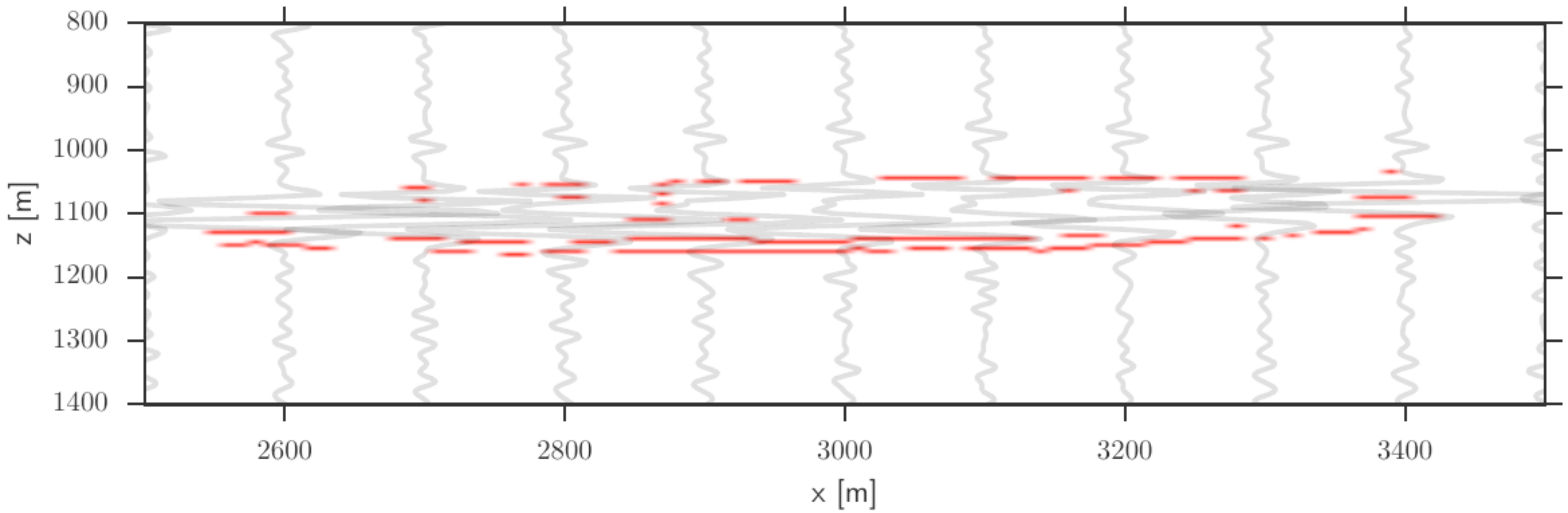
Solved using alternating direction method of multipliers

Results - physically consistent data



Results - migrated seismic





Summary

Successes:

Segmentation of reservoir in both images.

Physically interpretable segmentation without clustering.

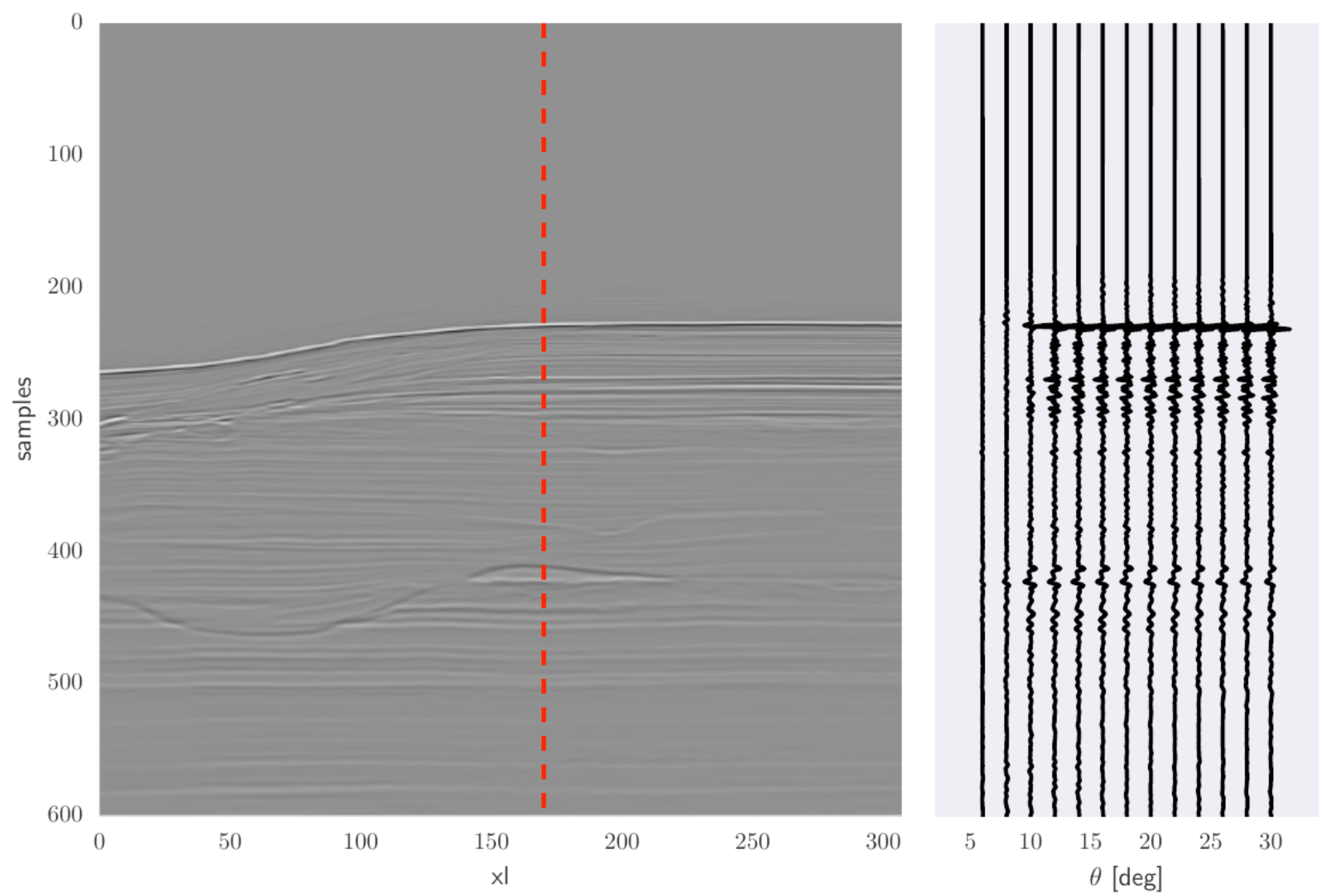
Challenges:

Requires tuning of one optimization trade off parameter.

Convergence sensitive to the rank of outliers, not well understood.

Comparison on field data

Migrated data provided by BG group to compare algorithms



Dynamic Intercept Gradient Inversion(DIGI)

Previous clustering based approaches were not successful.....

Compare robust PCA with BG's DIGI
Extend DIGI to use principle components

DIGI-inverse problem

$$\begin{bmatrix} \mathbf{d} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} W & W\sin^2\theta \\ \lambda\nabla & \lambda\nabla \\ W(\theta_{me})\cos(\chi_{me}) & W(\theta_{me})\sin(\chi_{me}) \end{bmatrix} \begin{bmatrix} \mathbf{i} \\ \mathbf{g} \end{bmatrix}$$

Convolution forward model: $d(x,t,\theta) = w(x,t,\theta) * r(x,t,\theta)$

- *ill-posed,* $\lambda\nabla$ term forces a smooth answer

Further augmented by extended elastic reflectivity (EER) term:

$$EER(\chi_{me}) = i\cos(\chi_{me}) + g\sin(\chi_{me})$$

- promotes correlation between *i* and *g*
- χ_{me} is related a priori geological information

System is solved using the conjugate gradient based algorithm LSQR.

Minimum energy projection

$$EER(\chi_{me}) = i\cos(\chi_{me}) + g\sin(\chi_{me})$$

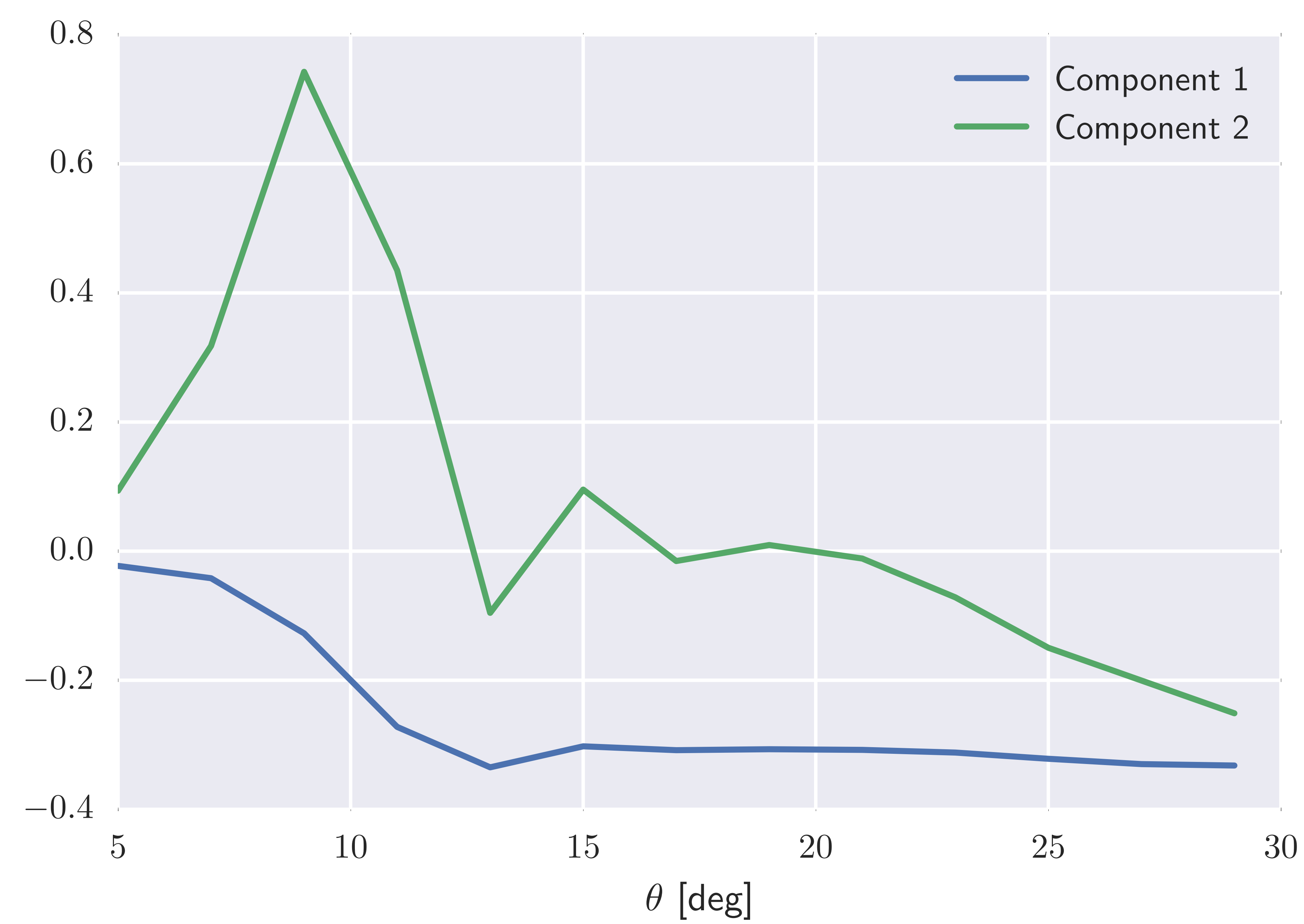
forms an image, where large values correspond to points where *i* and *g* are not correlated (outliers) along the projection χ_{me}.

Thresholding this image results in a segmented image.

PCA extended DIGI

$$\begin{bmatrix} \mathbf{d} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} W\mathbf{c_1} & W\mathbf{c_2} \\ \lambda\nabla & \lambda\nabla \\ W(\theta_{me})\cos(\chi_{me}) & W(\theta_{me})\sin(\chi_{me}) \end{bmatrix} \begin{bmatrix} \mathbf{i} \\ \mathbf{g} \end{bmatrix}$$

Exact same algorithm, but use the principle components extracted directly from the data.



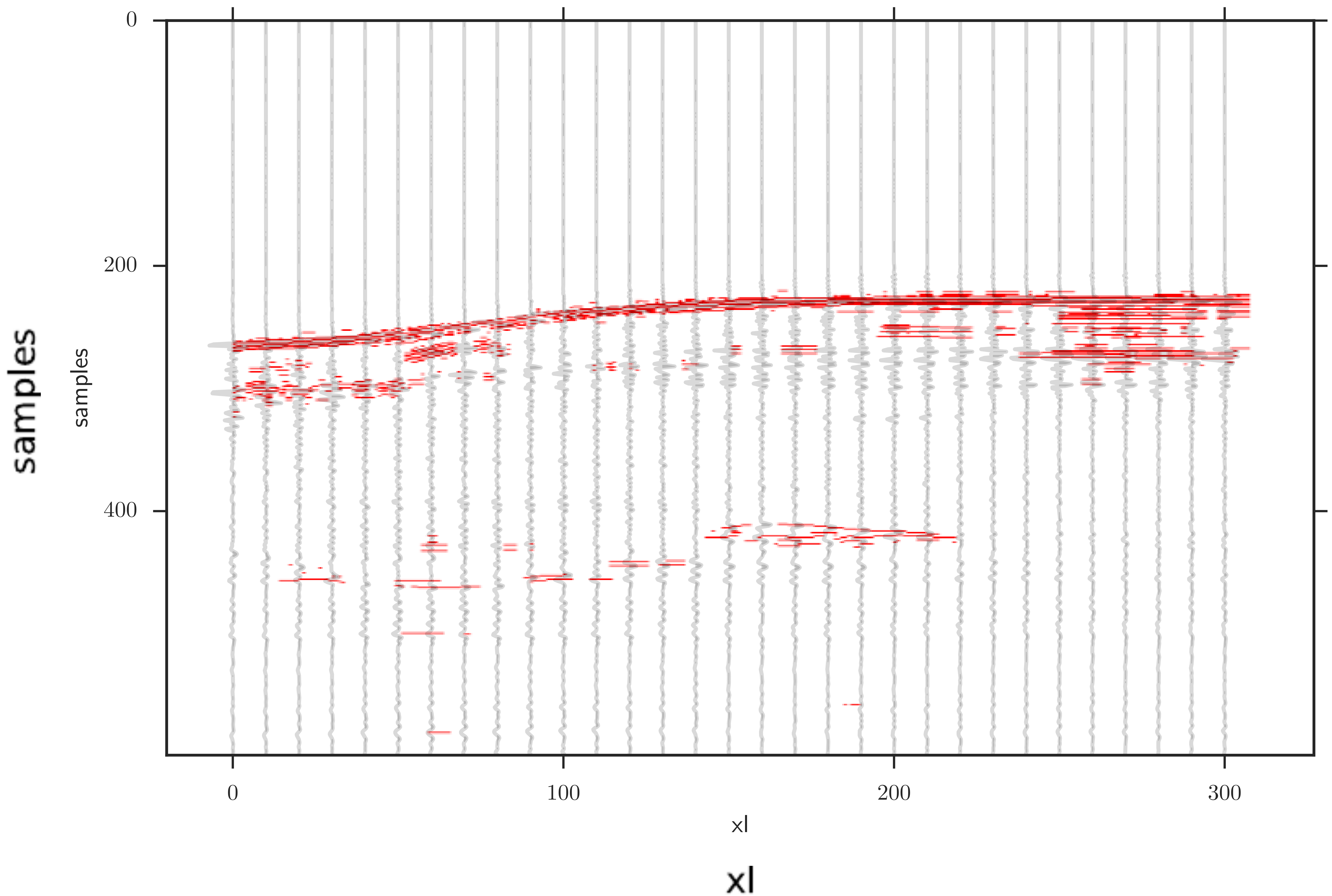
Comparison method

Manually threshold the image to segment the potential hydrocarbon reserve while maintaining the least amount of spurious segmentation.

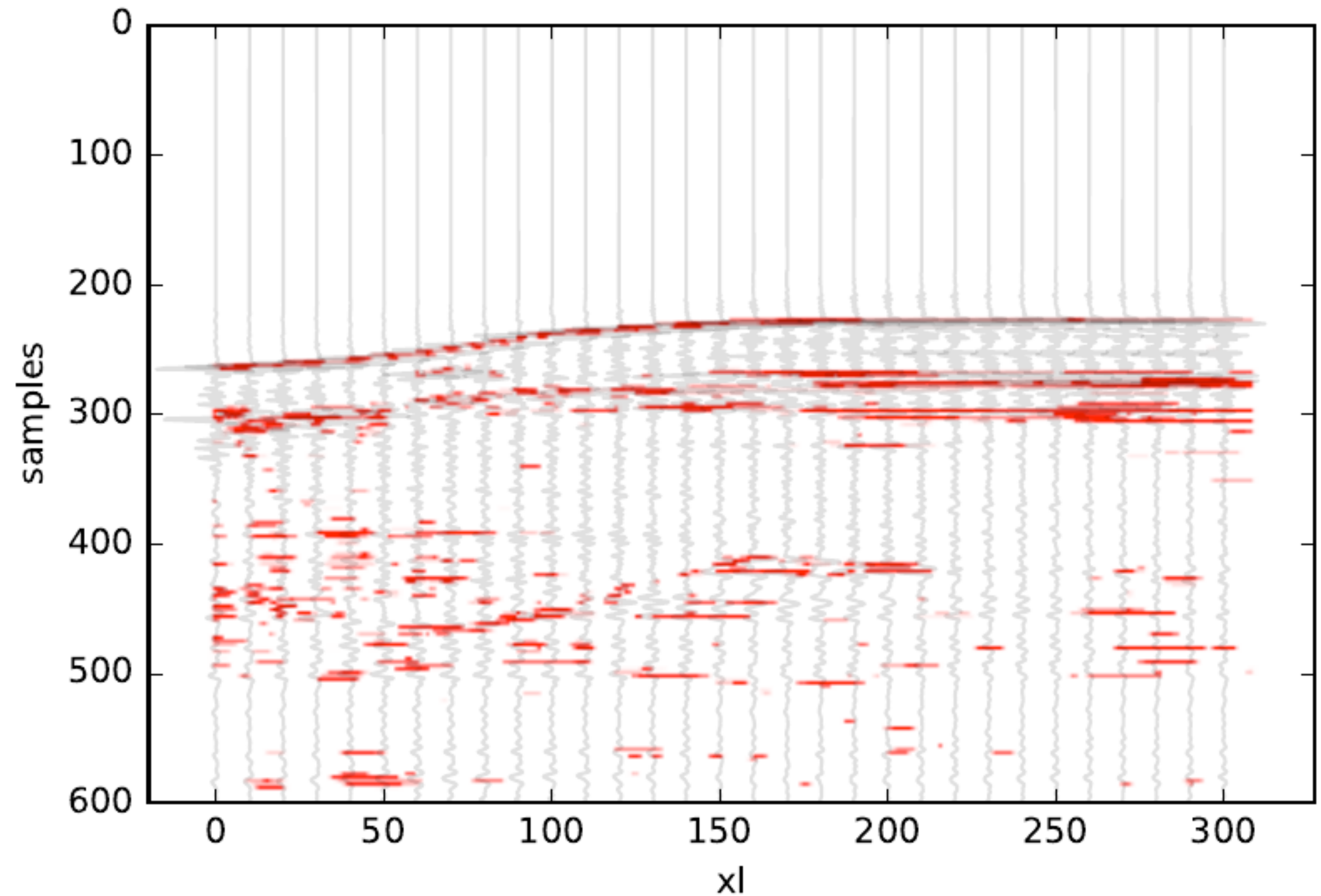
Compare:

- Robust PCA
- DIGI
- PCA extended

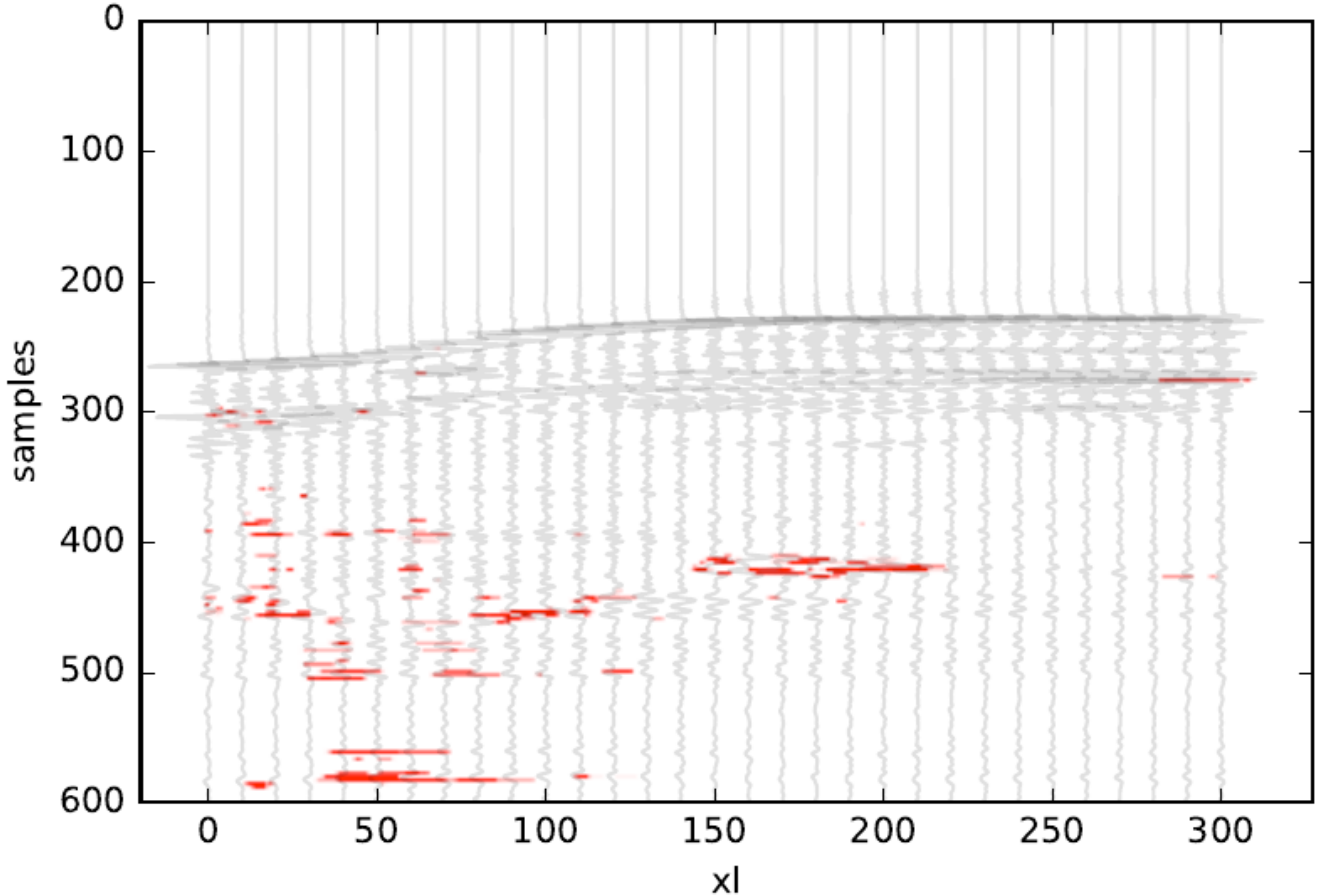
Results - robust PCA



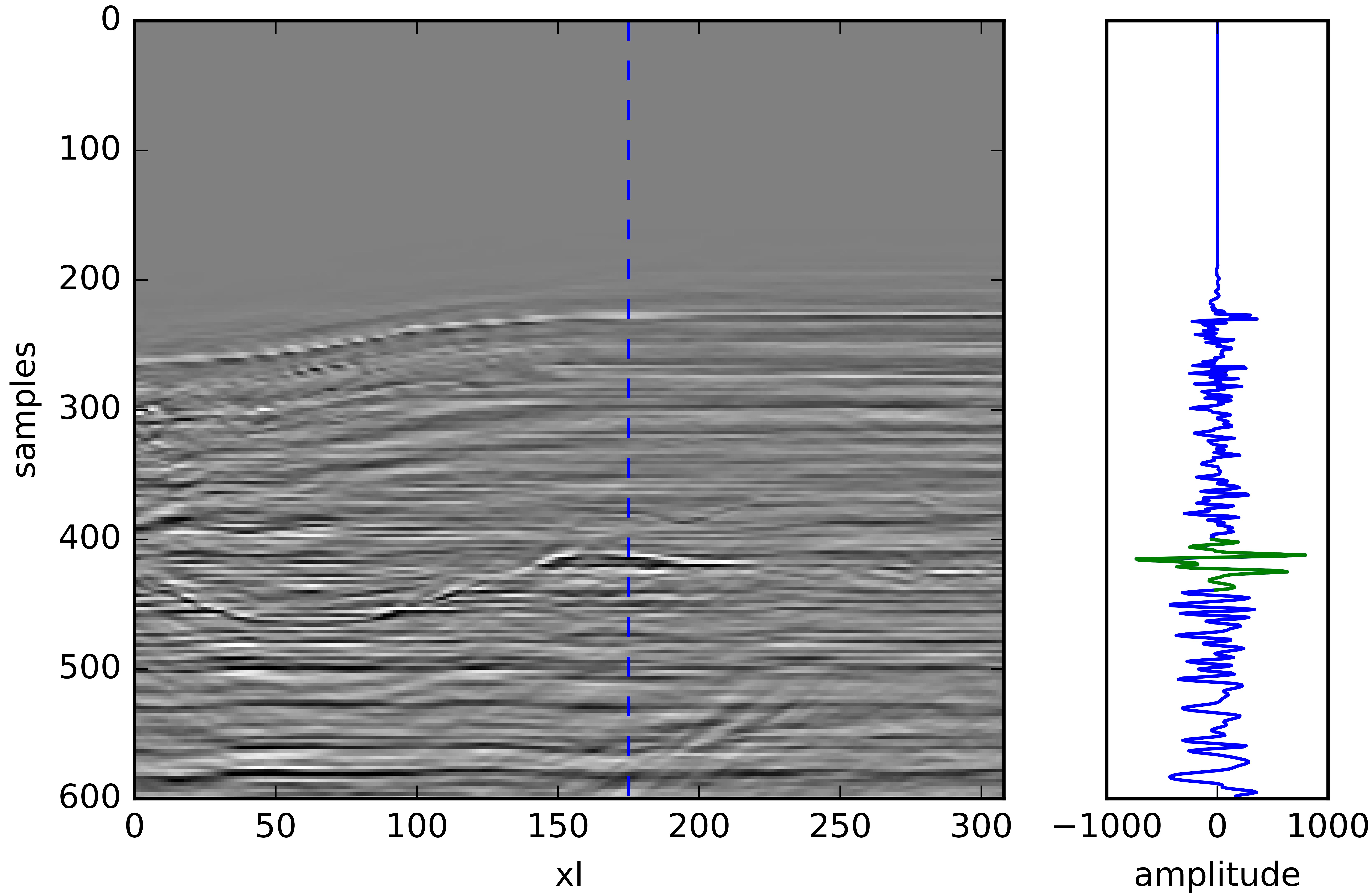
Results - DIGI



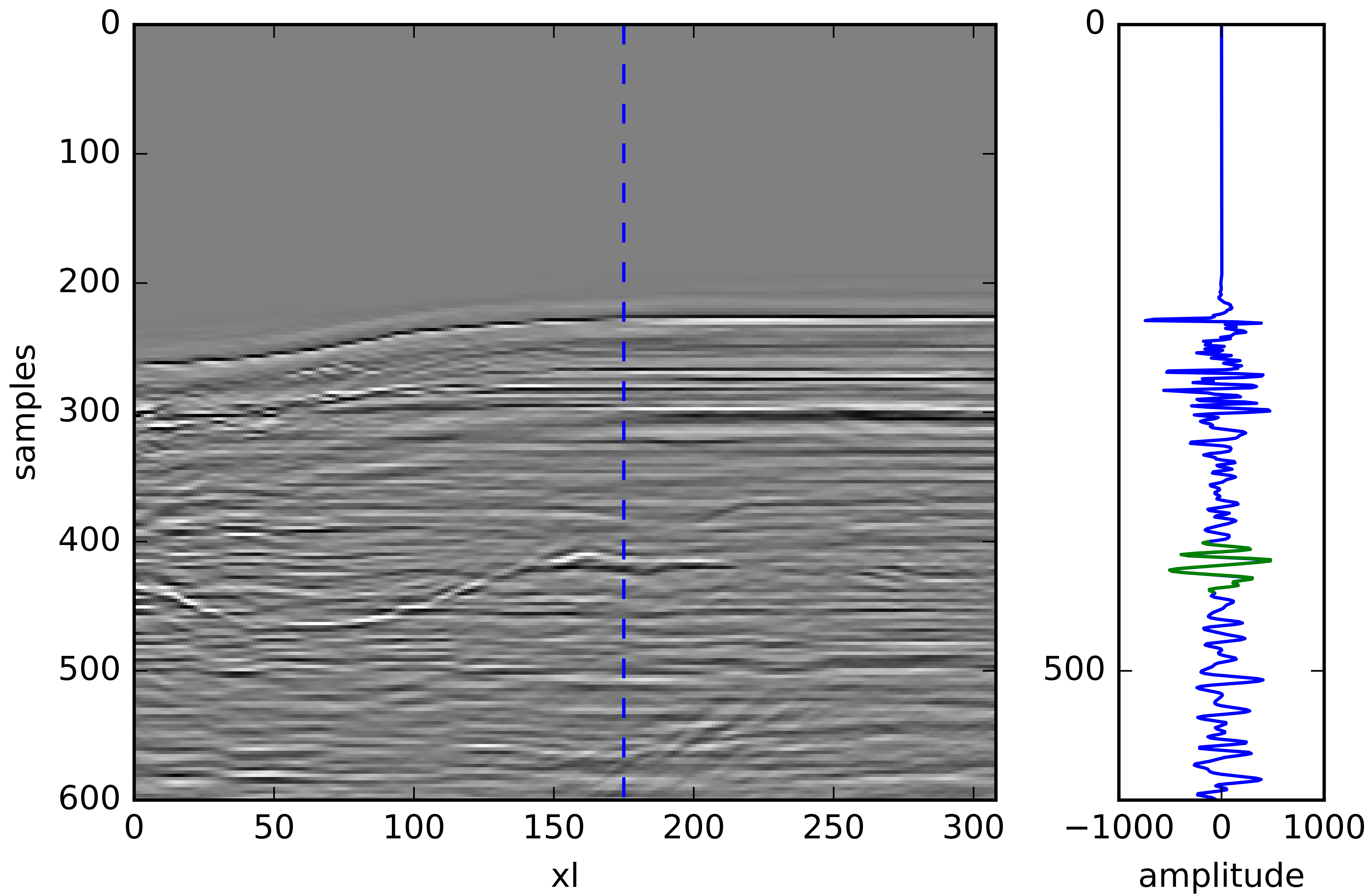
Results - PCA extend DIGI



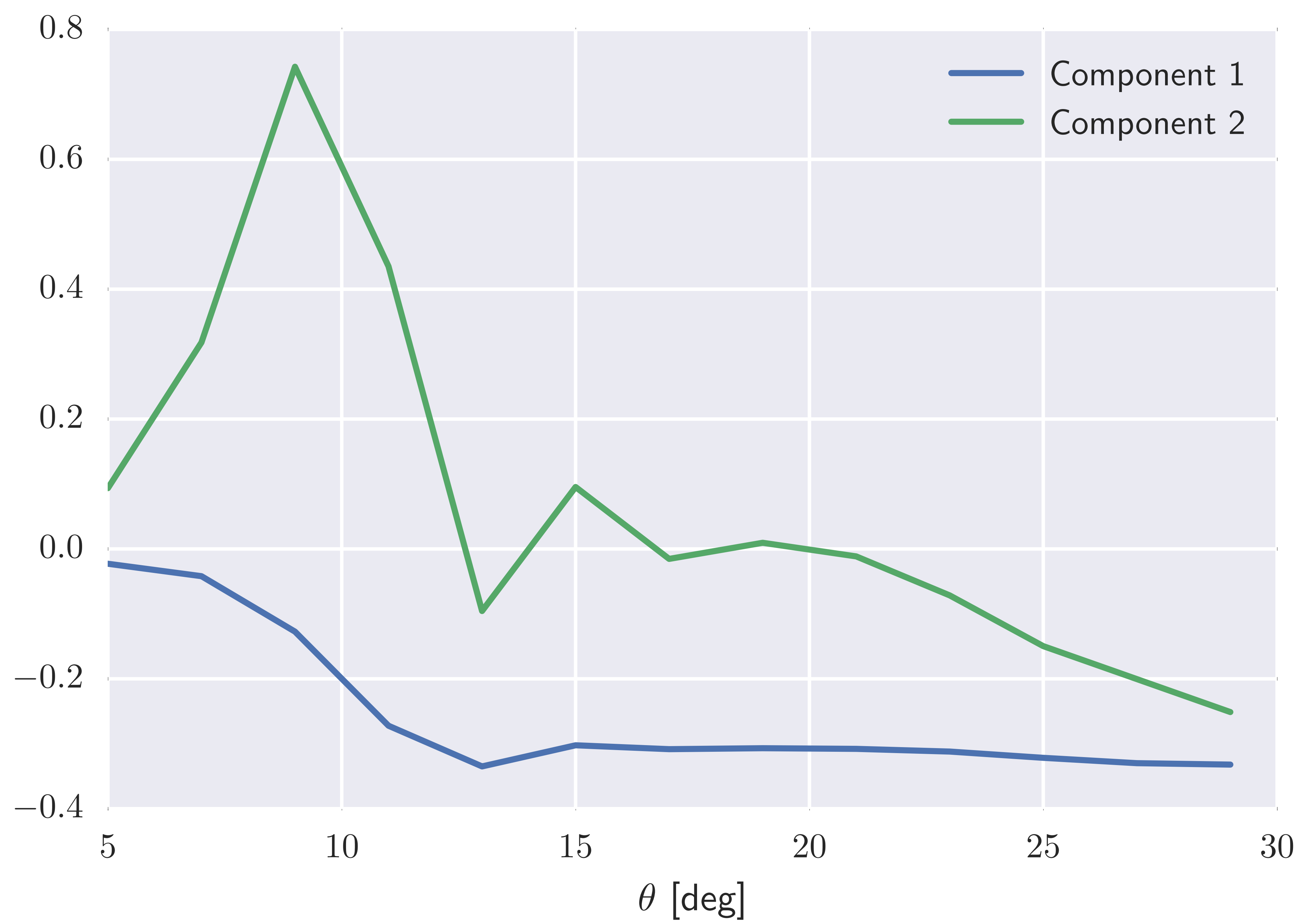
Results - PCA extended DIGI

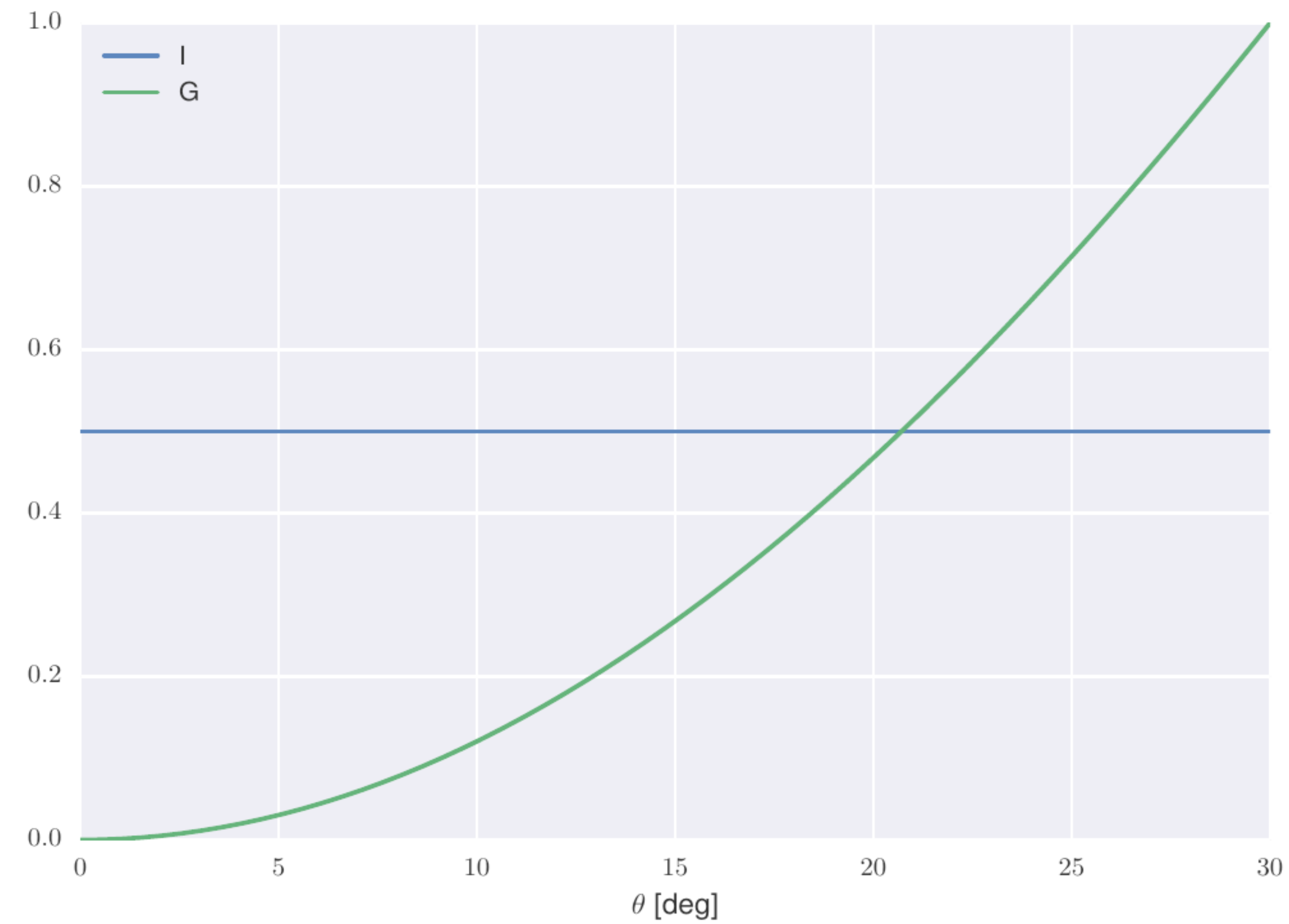


Results - DIGI



Why the difference?





Summary

- Robust PCA provided the best image segmentation.
- PCA extended DIGI better separated the potential reservoir from the background trend.
- The extracted principle components showed significantly different shapes than the Shuey vectors.

Epilogue

Outcome:

Generalized a conventional analysis approach using unsupervised learning models.

Successful in segmented potential hydrocarbon reserves from seismic data

Future:

More data, standardized datasets

Quantitative benchmarks

References

Zoeppritz, K., 1919, VII b. Über Reflexion und Durchgang seismischer Wellen durch Unstetigkeitsflächen: Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1919, 66–84

Shuey, R. T., 1985, A simplification of the Zoeppritz equations: Geophysics, 50, 609–614.

Gardner, G. H. F., L. W. Gardner, and A. R. Gregory, 1974, Formation velocity and density; the diagnostic basics for stratigraphic traps: Geophysics, 39, 770–780.

Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985, Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks: Geophysics, 50, 571–581.

Castagna, J. P., H. W. Swan, and D. J. Foster, 1998, Framework for AVO gradient and intercept interpretation: Geophysics, 63, 948–956.

Edgar, J., and J. Selvage, 2013, Dynamic Intercept-gradient Inversion

Acknowledgements

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.