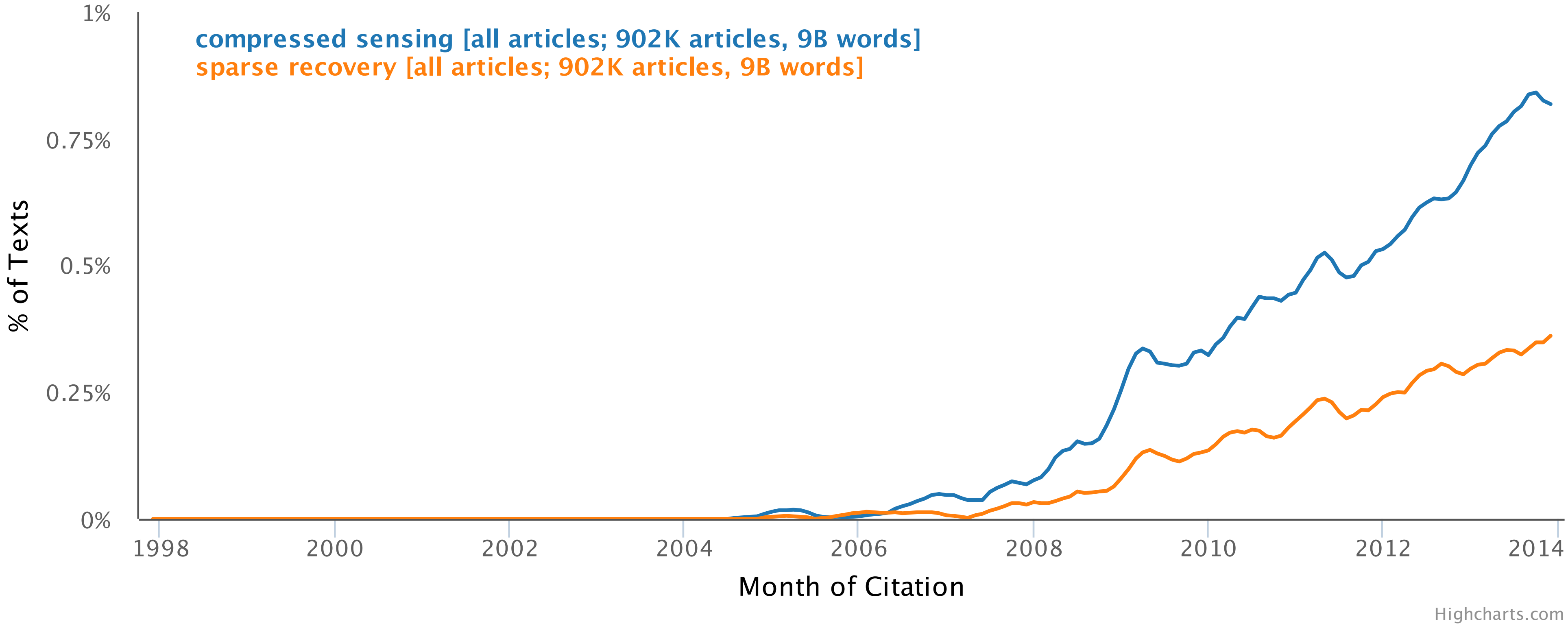
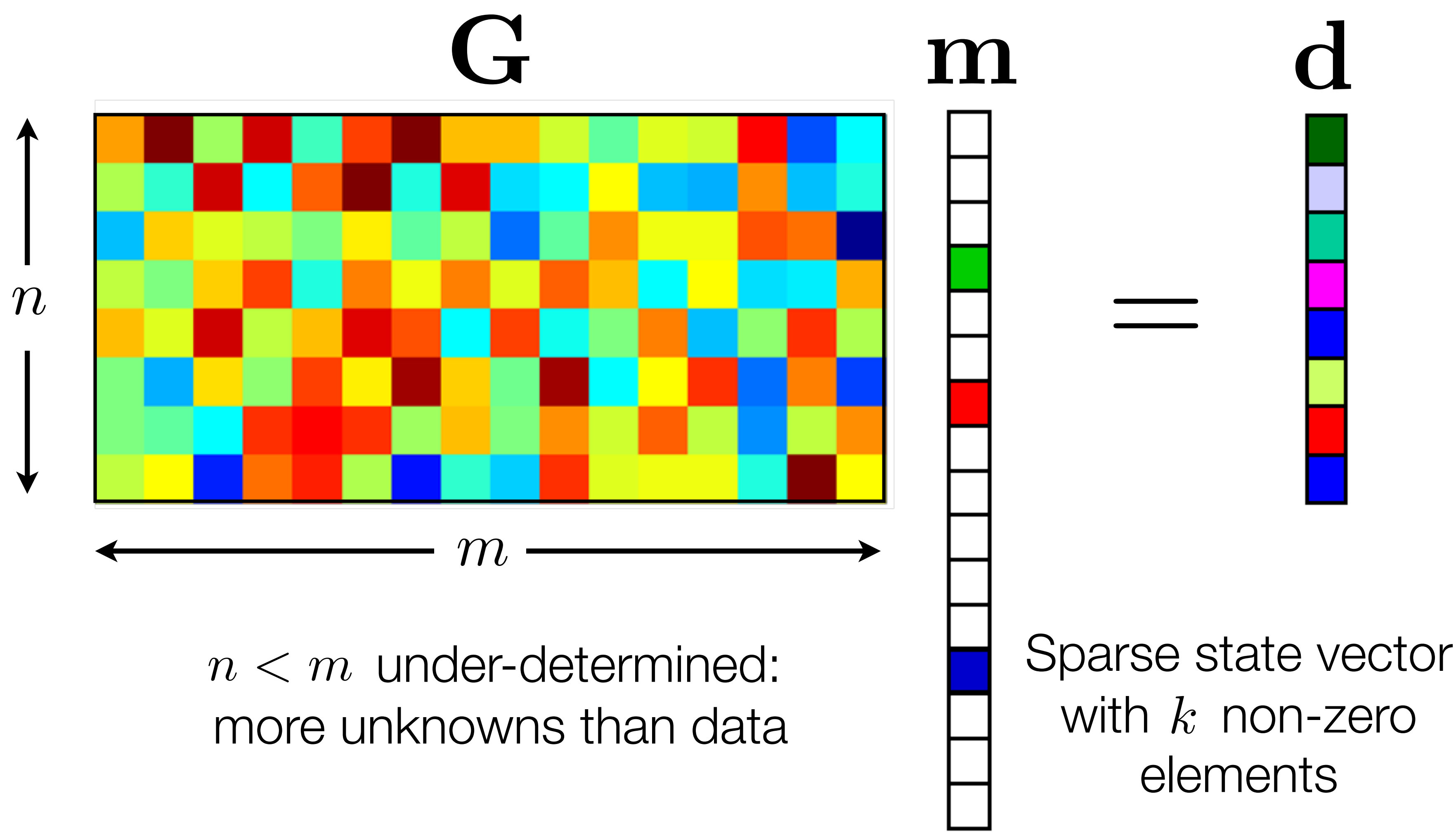
arXiv.org N-gram



Note to self (delete before presentation):

Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes
Don’t talk about earthquakes

Sparse signal recovery (compressed sensing)



Variations on state vector regularization

Damped least squares: L_2 regularization

$$\min ||\mathbf{Gm} - \mathbf{d}||_2 + \boxed{\lambda ||\mathbf{m}||_2}$$ damp oscillations

Solutions vary smoothly in space (common in various formulations)

Classical approach with exact single step solution

Sparsity promoting methods: L_0 regularization

$$\min ||\mathbf{Gm} - \mathbf{d}||_2 + \boxed{\lambda ||\mathbf{m}||_0}$$ promote sparsity

The L_0 pseudo-norm simply counts the number of non-zero elements

This is combinatorial and seemingly unfeasible to solve in reasonable time (<years) for any large system (>100 elements)

Variations on state vector regularization

Damped least squares: L_2 regularization

$$\min ||\mathbf{Gm} - \mathbf{d}||_2 + \boxed{\lambda ||\mathbf{m}||_2}$$ damp oscillations

Solutions vary smoothly in space (common in various formulations)

Classical approach with exact single step solution

Sparsity promoting methods: L_1 regularization

$$\min ||\mathbf{Gm} - \mathbf{d}||_2 + \boxed{\lambda ||\mathbf{m}||_1}$$ promote sparsity

The L_1 norm can often be used to recover the L_0 pseudo-norm solution

Global minimum can be found by convex optimization (e.g., quadratic programming) and many new algorithms

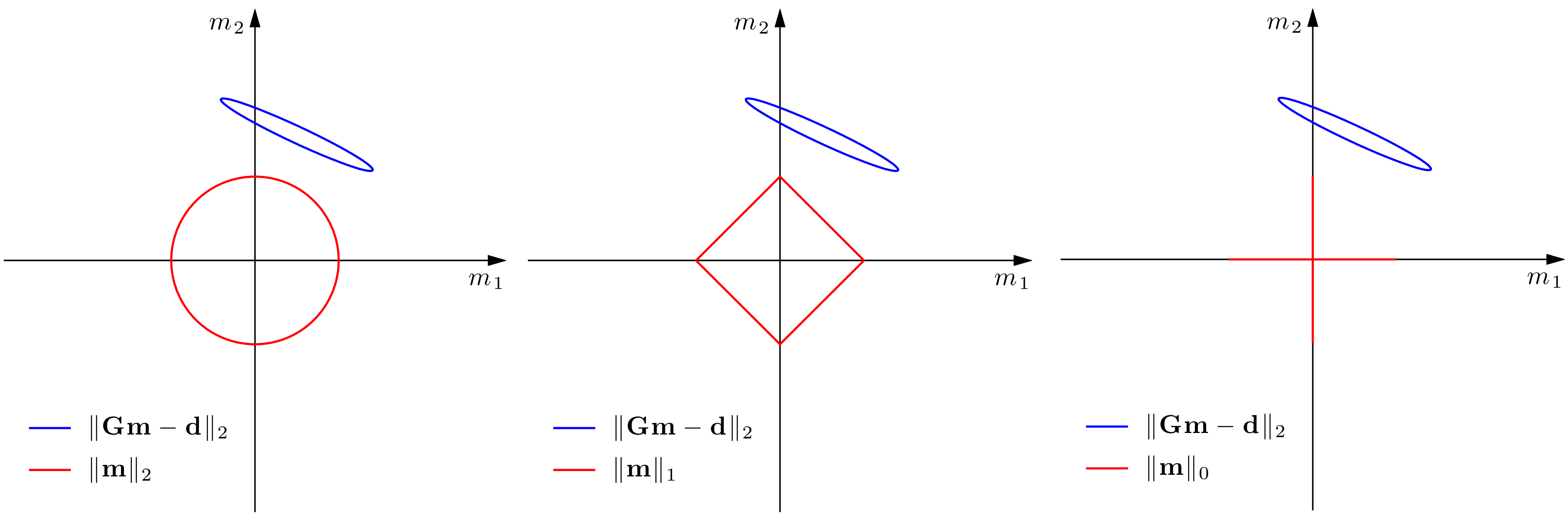
A geometric view of compressed sensing

Minimize data misfit and *p*-norm of state vector

$$f = \|\mathbf{Gm} - \mathbf{d}\|_2 + \lambda\|\mathbf{m}\|_p \qquad \|\mathbf{m}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$$

fit to data

model regularization



A geometric view of compressed sensing

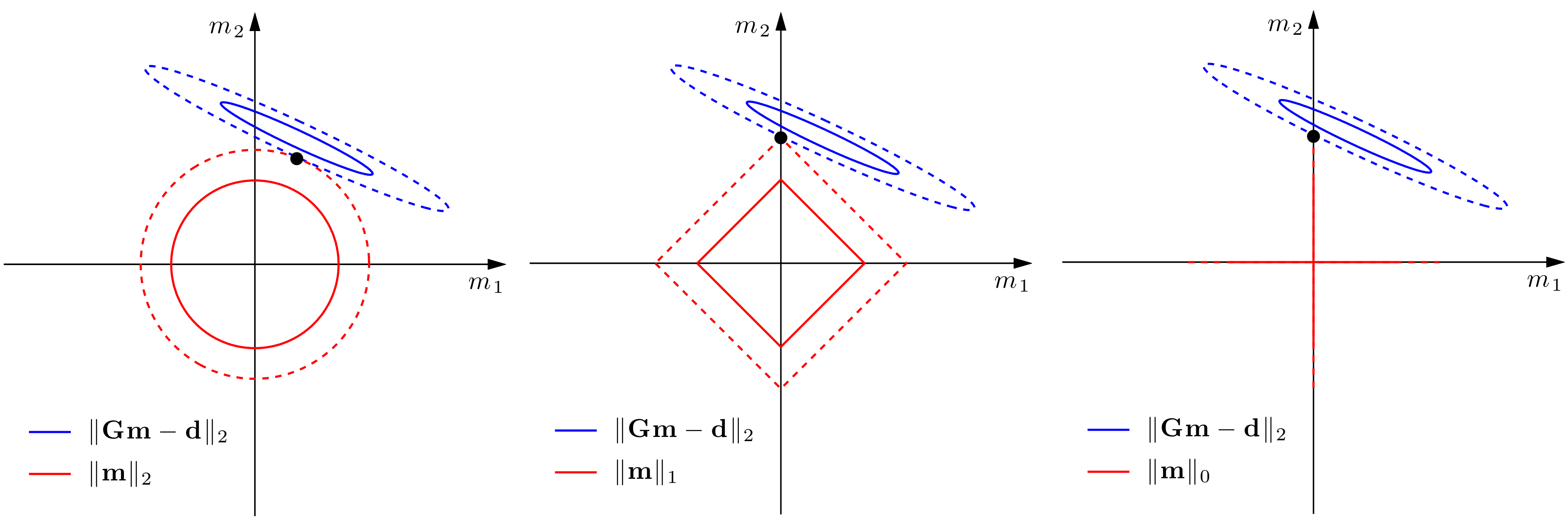
Minimize data misfit and *p*-norm of state vector

$$f = \|\mathbf{Gm} - \mathbf{d}\|_2 + \lambda \|\mathbf{m}\|_p$$

fit to data

model regularization

$$\|\mathbf{m}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$



A geometric view of compressed sensing

Minimize data misfit and *p*-norm of state vector

$$f = \|\mathbf{Gm} - \mathbf{d}\|_2 + \lambda \|\mathbf{m}\|_p$$

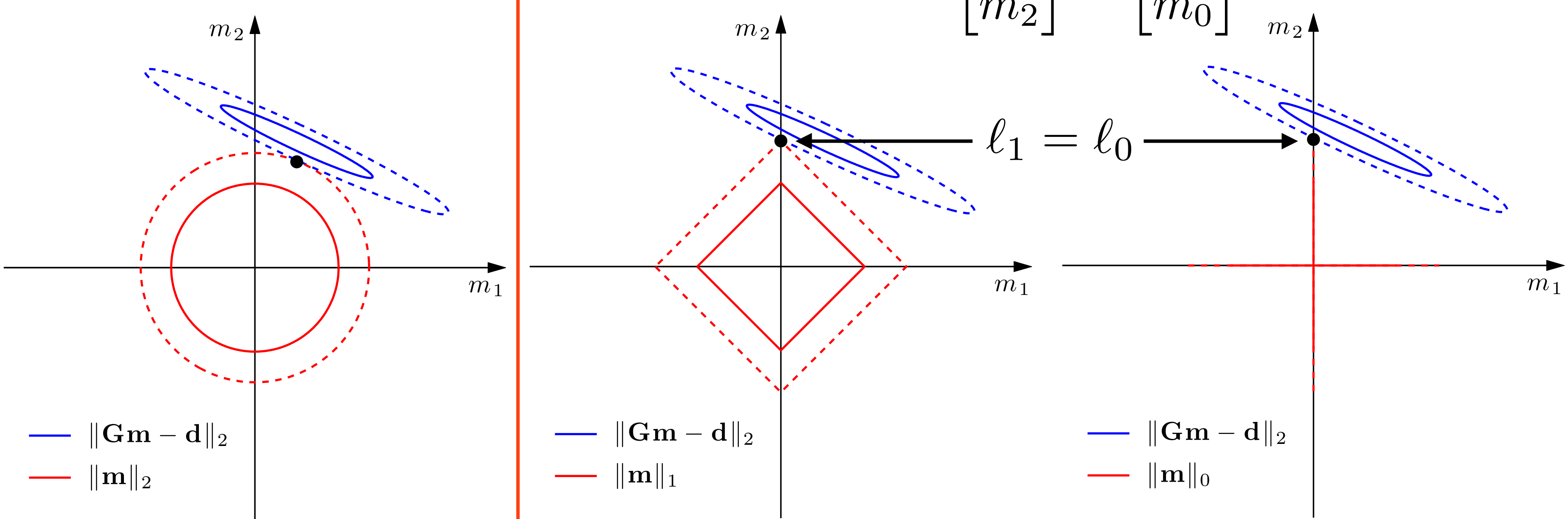
fit to data

model regularization

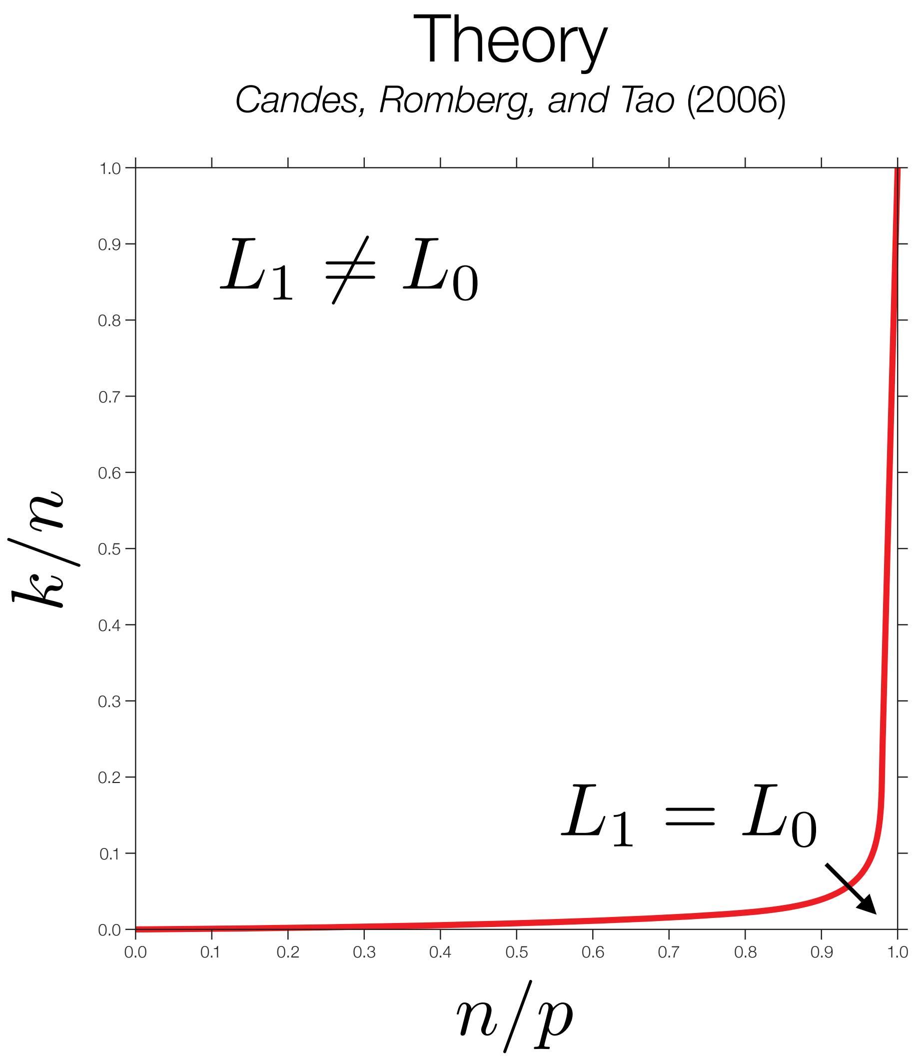
$$\|\mathbf{m}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

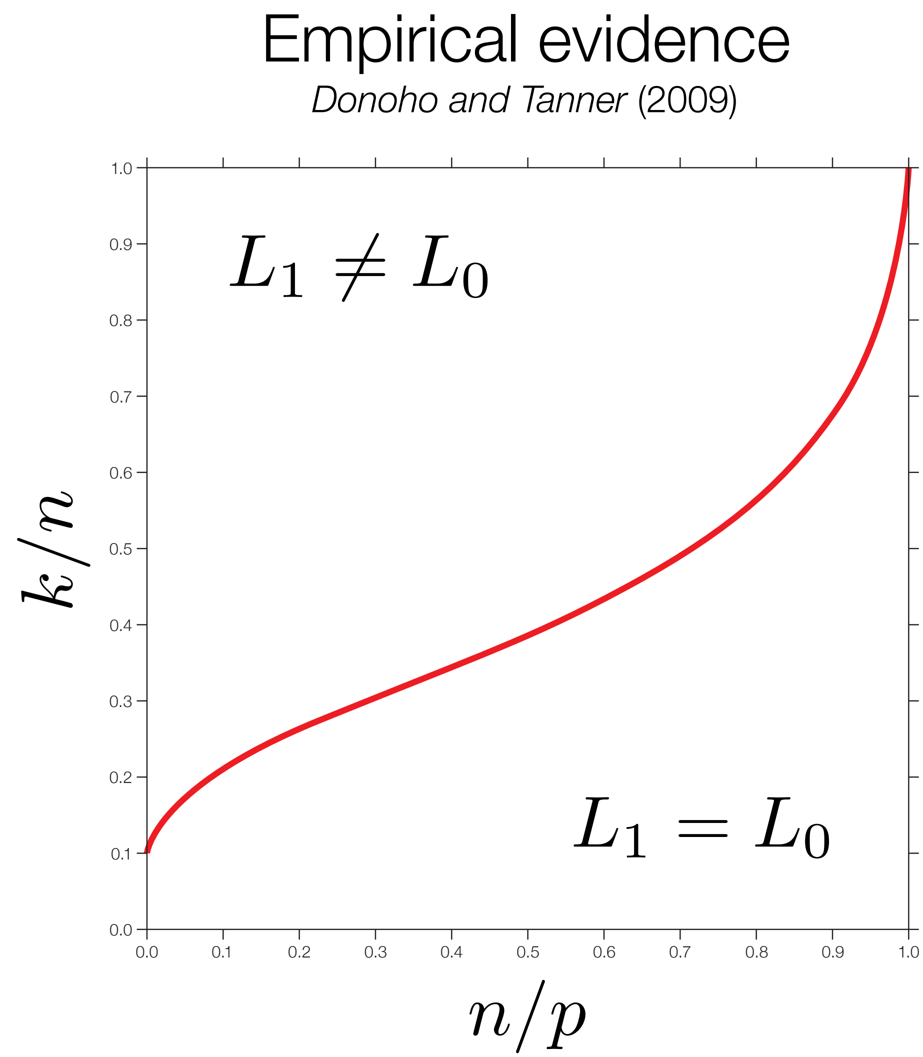
Same solution for *p*=1, *p*=0

$$\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} 0 \\ m_0 \end{bmatrix}$$



When does this work?





Frequency distribution of operator elements (*Davenport et al*., 2011)

Good: Power-law, Gaussian, …

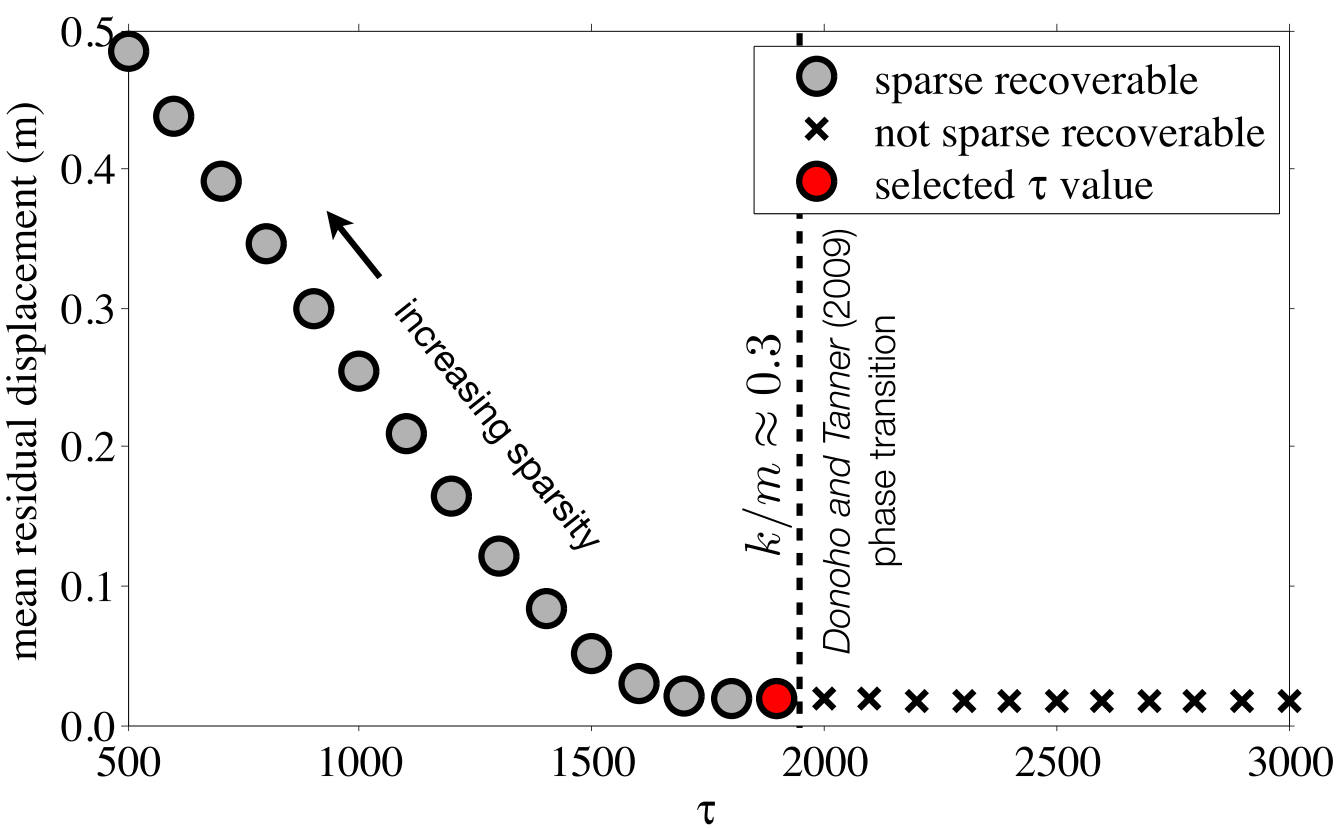
Bad: Uniform, anything with any negative eigenvalues

How sparse?

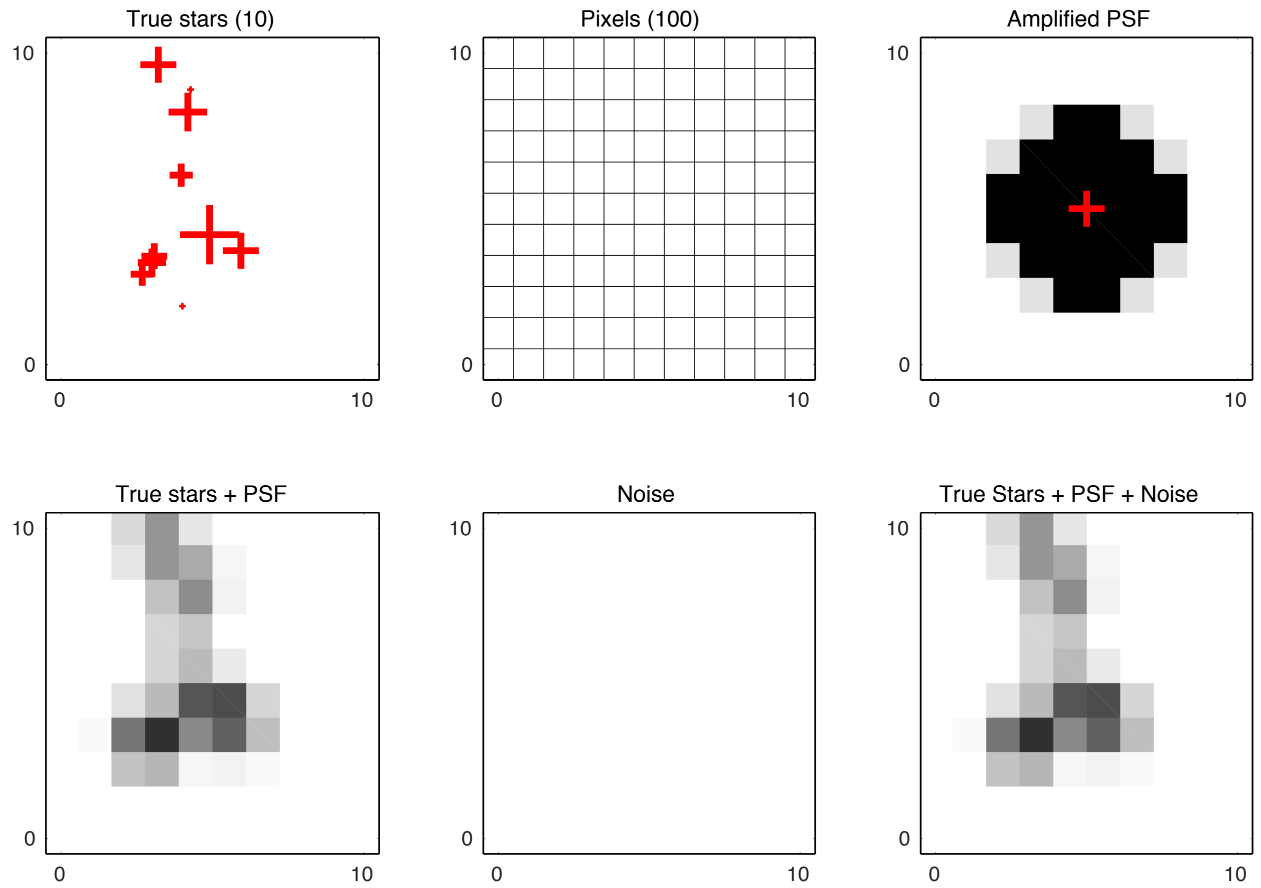
Replace unconstrained problem with an equivalent constrained problem (*Tibshirani*, 1996):

$$\min \|\mathbf{Gm} - \mathbf{d}\|_2 \text{ subject to } \|\mathbf{m}\|_1 \leq \tau$$

τ -selection: How sparse?



Some stuff Doug made me do

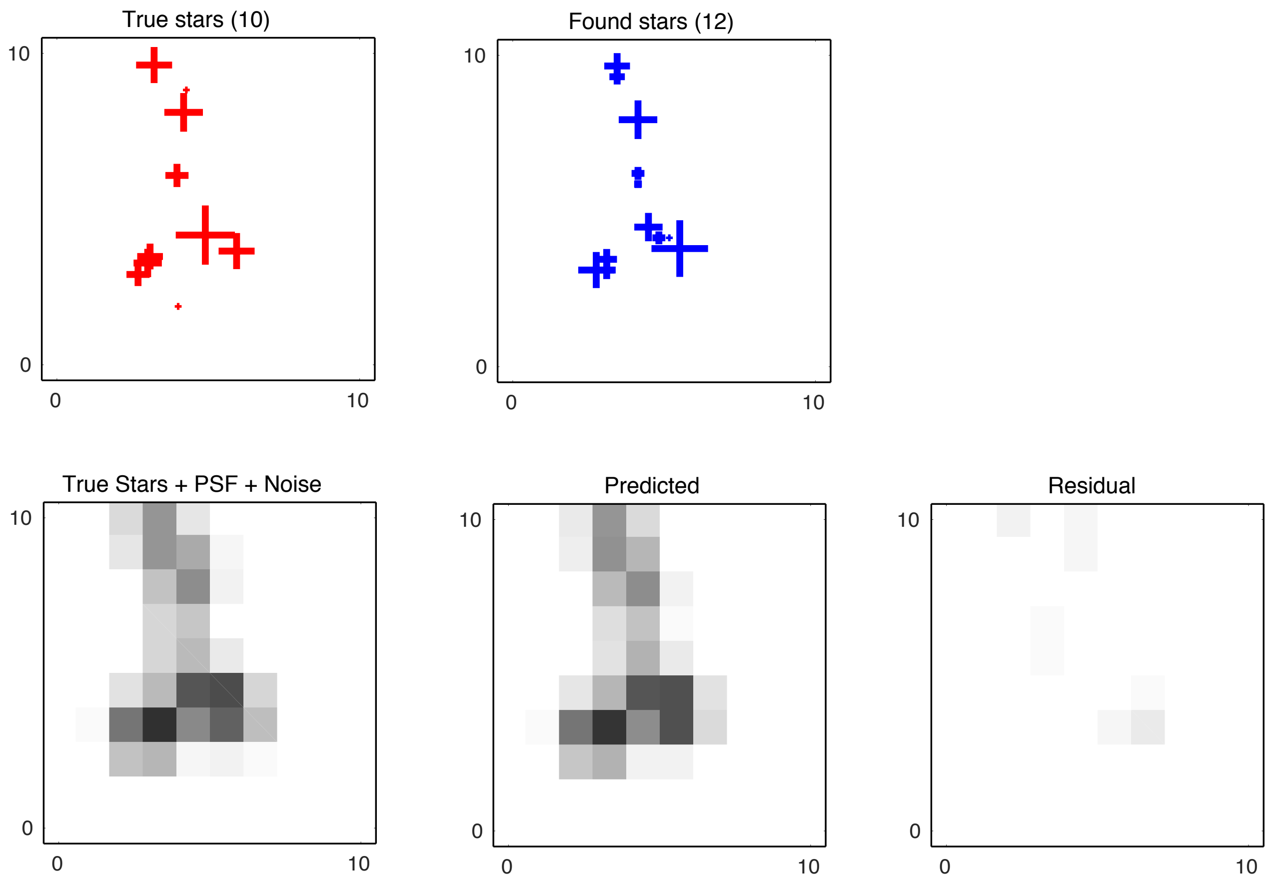


Over parameterize geometry (100 grid points per pixel)
- 100 x 10,000 linear flux & quantized position operator

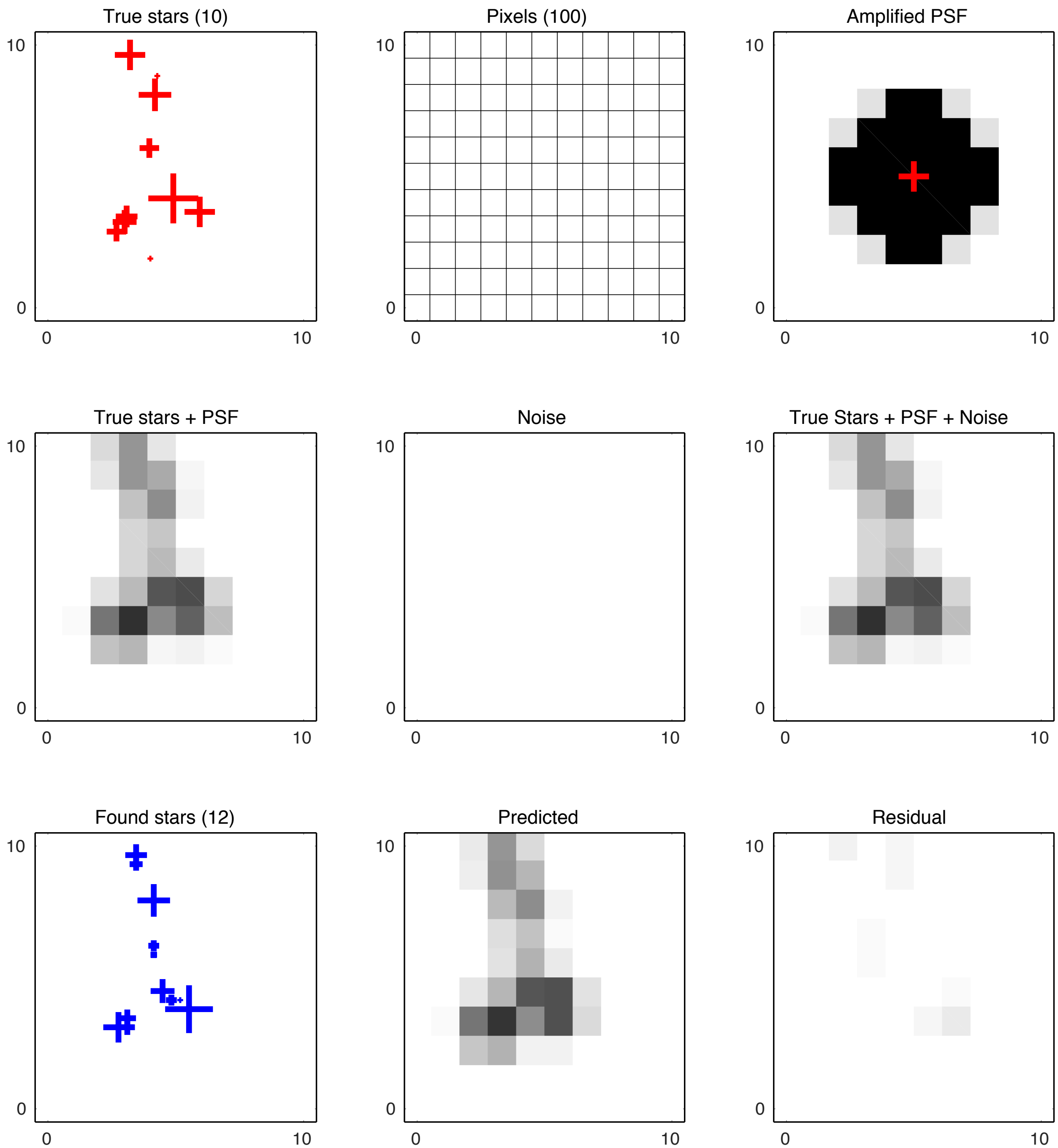
Gaussian point spread function

Solve in < 0.2 seconds (*van den Berg and Friedlander*, 2008)

More stuff that Doug made me do

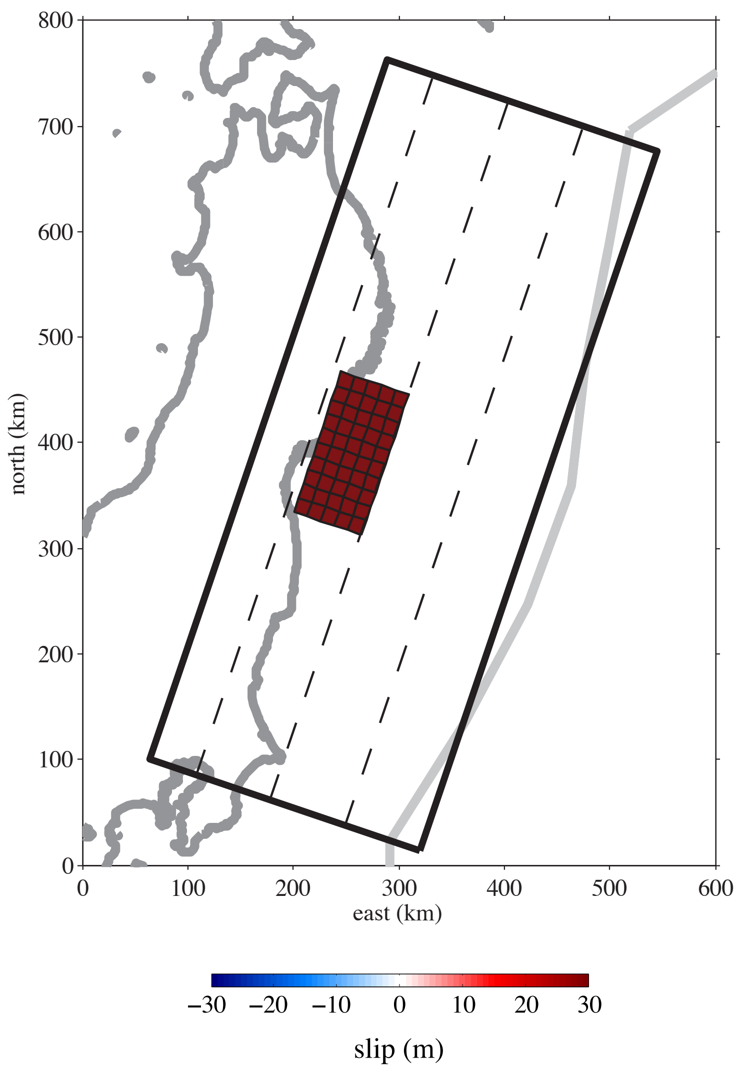


To do:
Constrain flux distributions
How many stars can we recover? vs. How many are there?
If PSFs are localized in space we can go very fast

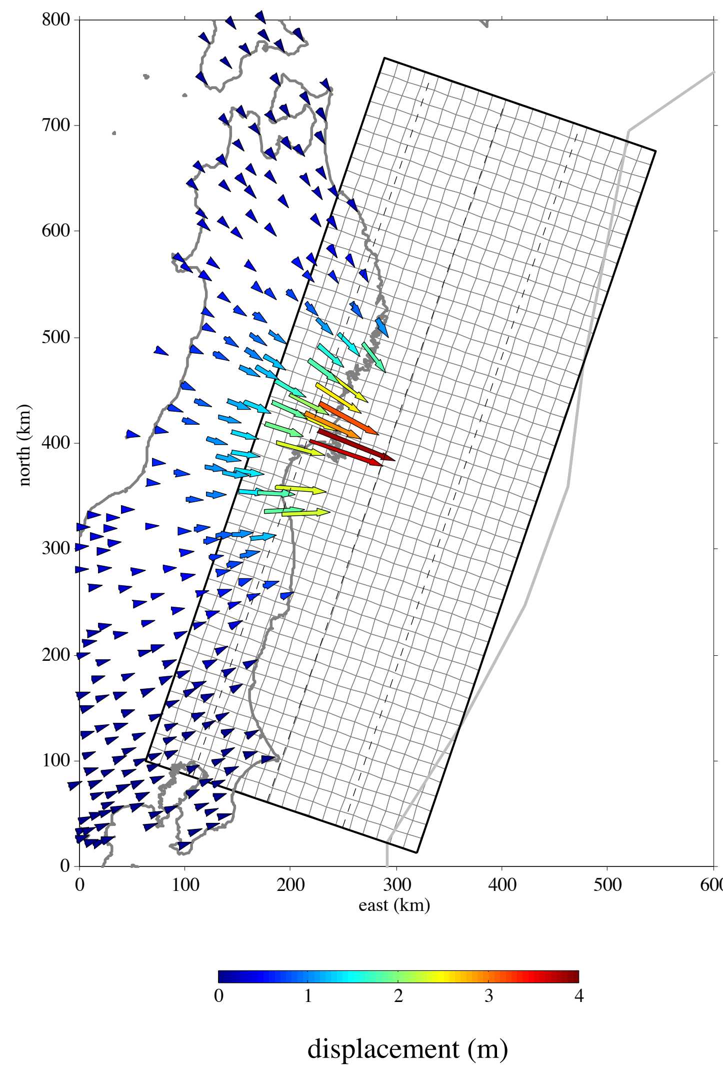

True stars (10)
Pixels (100)
Amplified PSF
True stars + PSF
Noise
True Stars + PSF + Noise
Found stars (12)
Predicted
Residual

Imaging fault behavior…how well can we do?

Synthetic slip

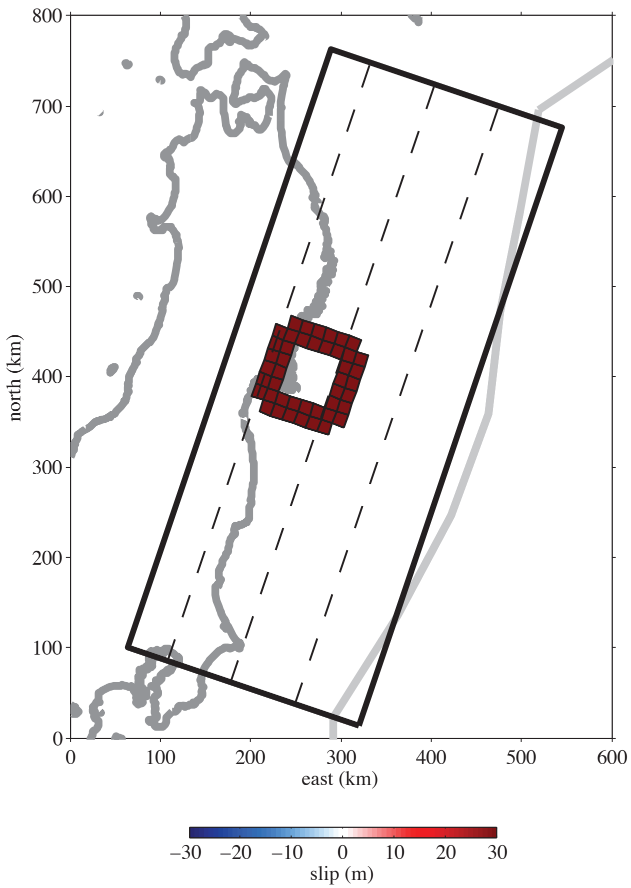


Synthetic observations



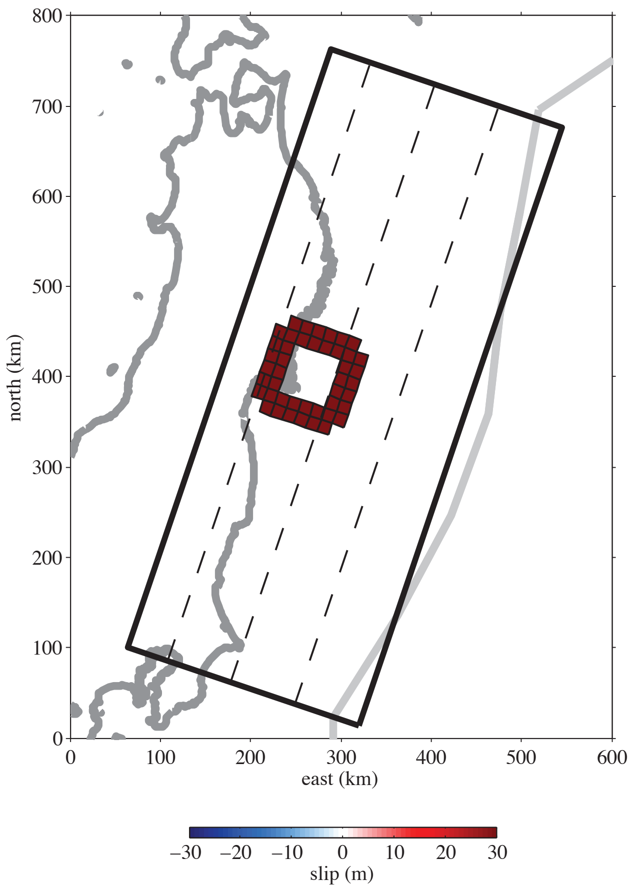
Resolution test - ring

Synthetic slip

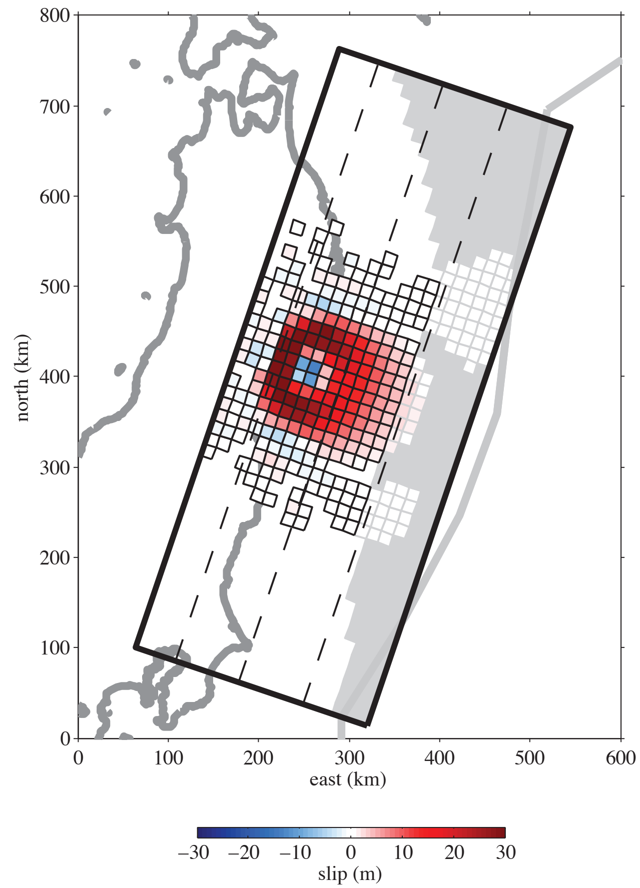


Resolution test - ring

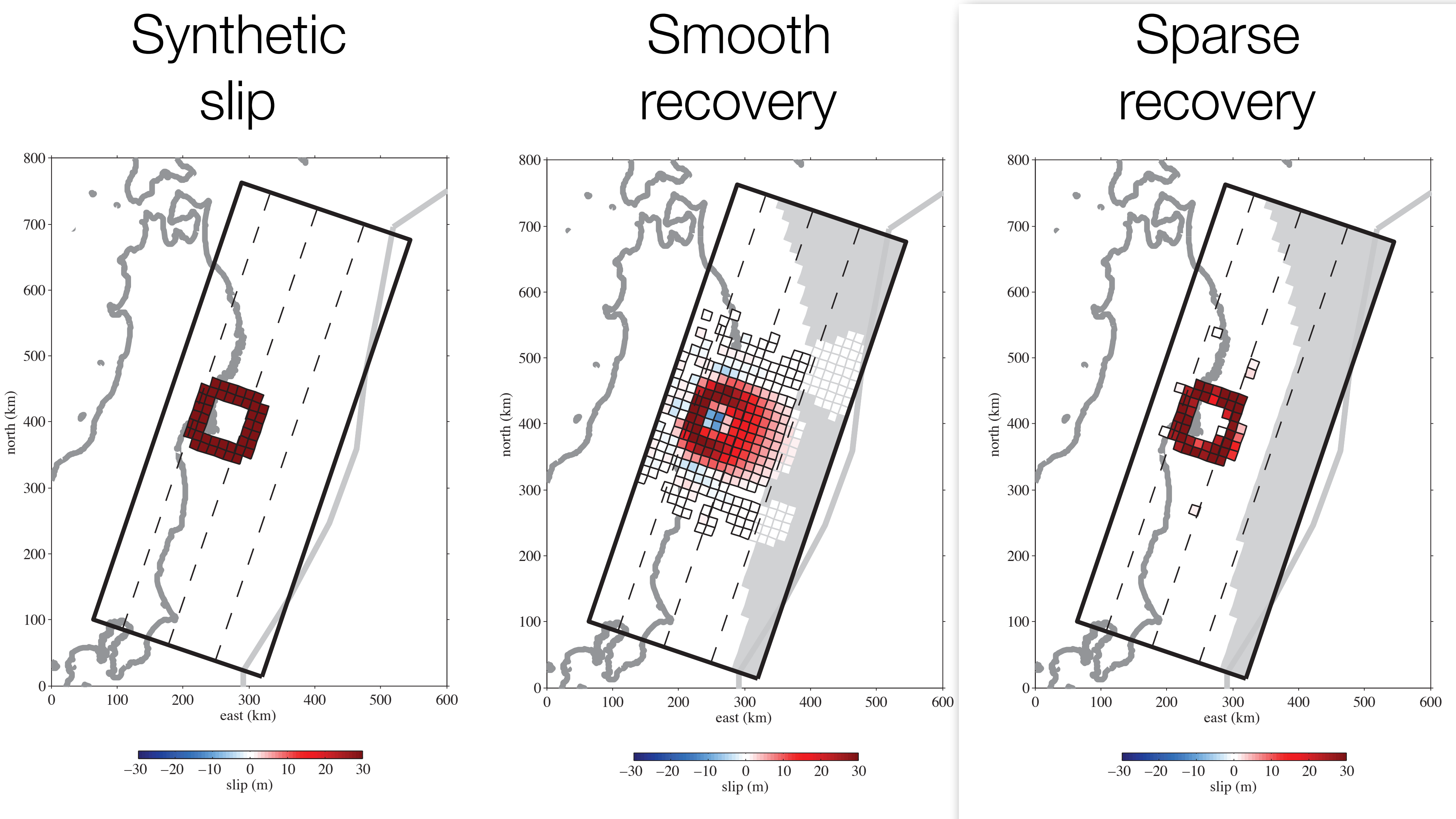
Synthetic slip



Smooth recovery



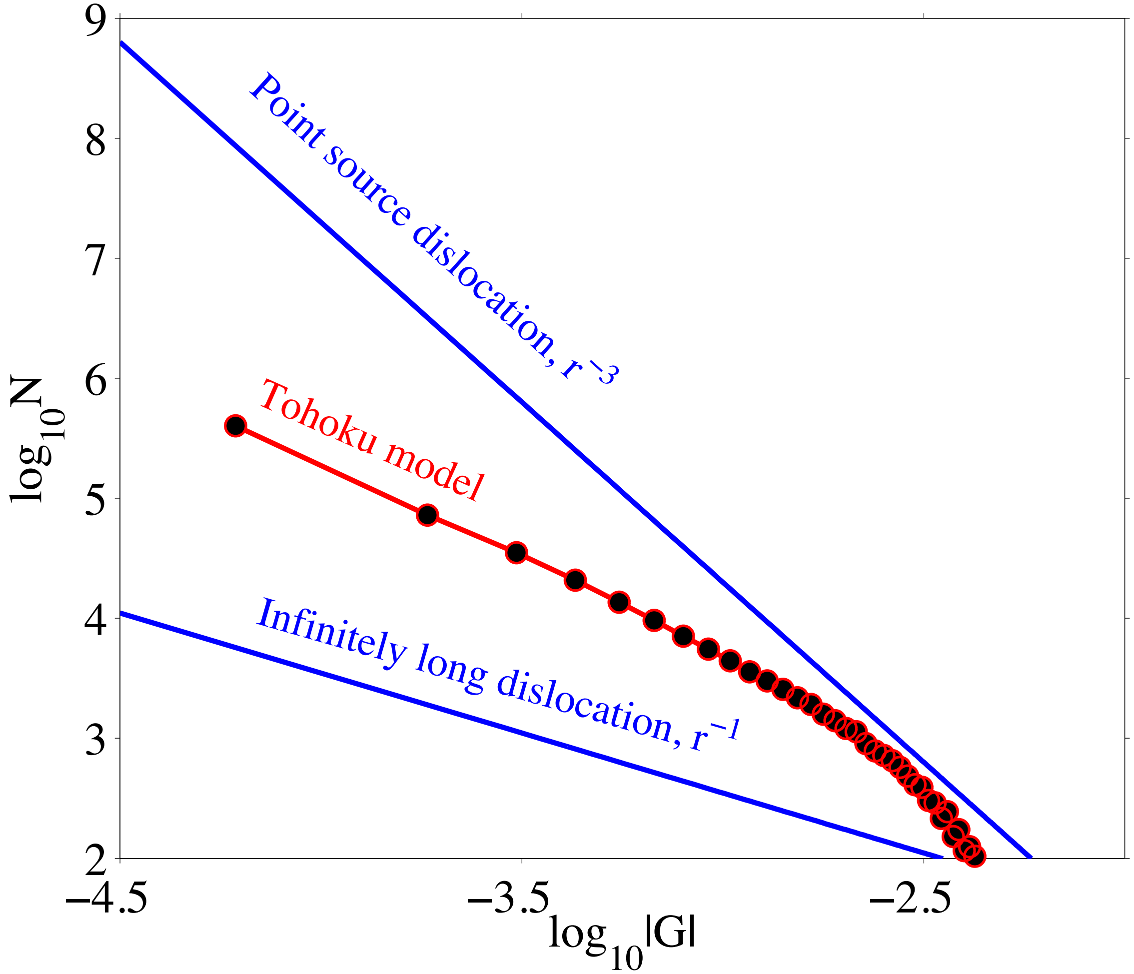
Resolution test - ring



Sparsity promoting recovery methods are not perfect at this density and may exhibit low magnitude outliers

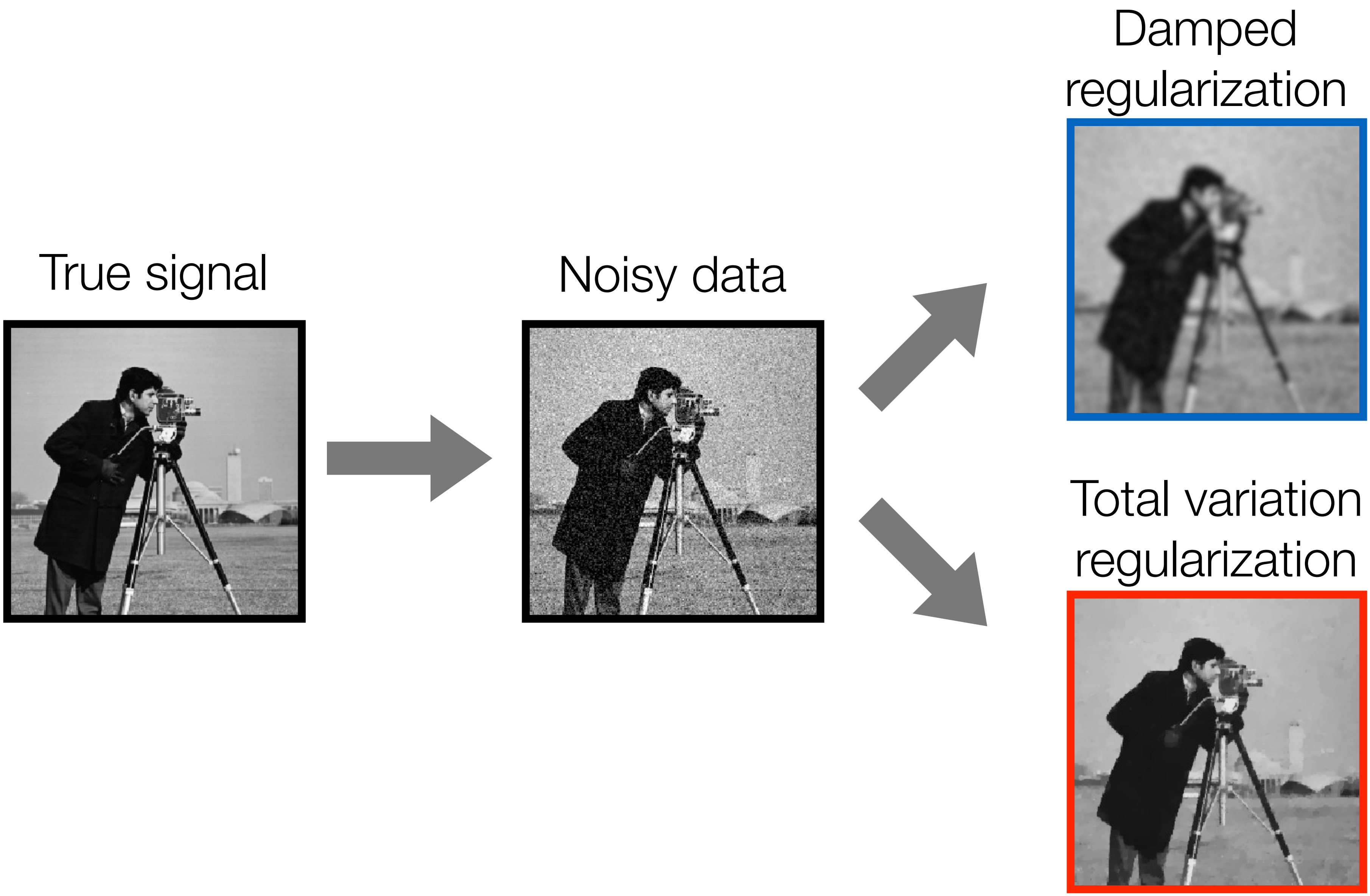
Evans and Meade, (2012)

Why does sparsity work for this problem?: Elasticity

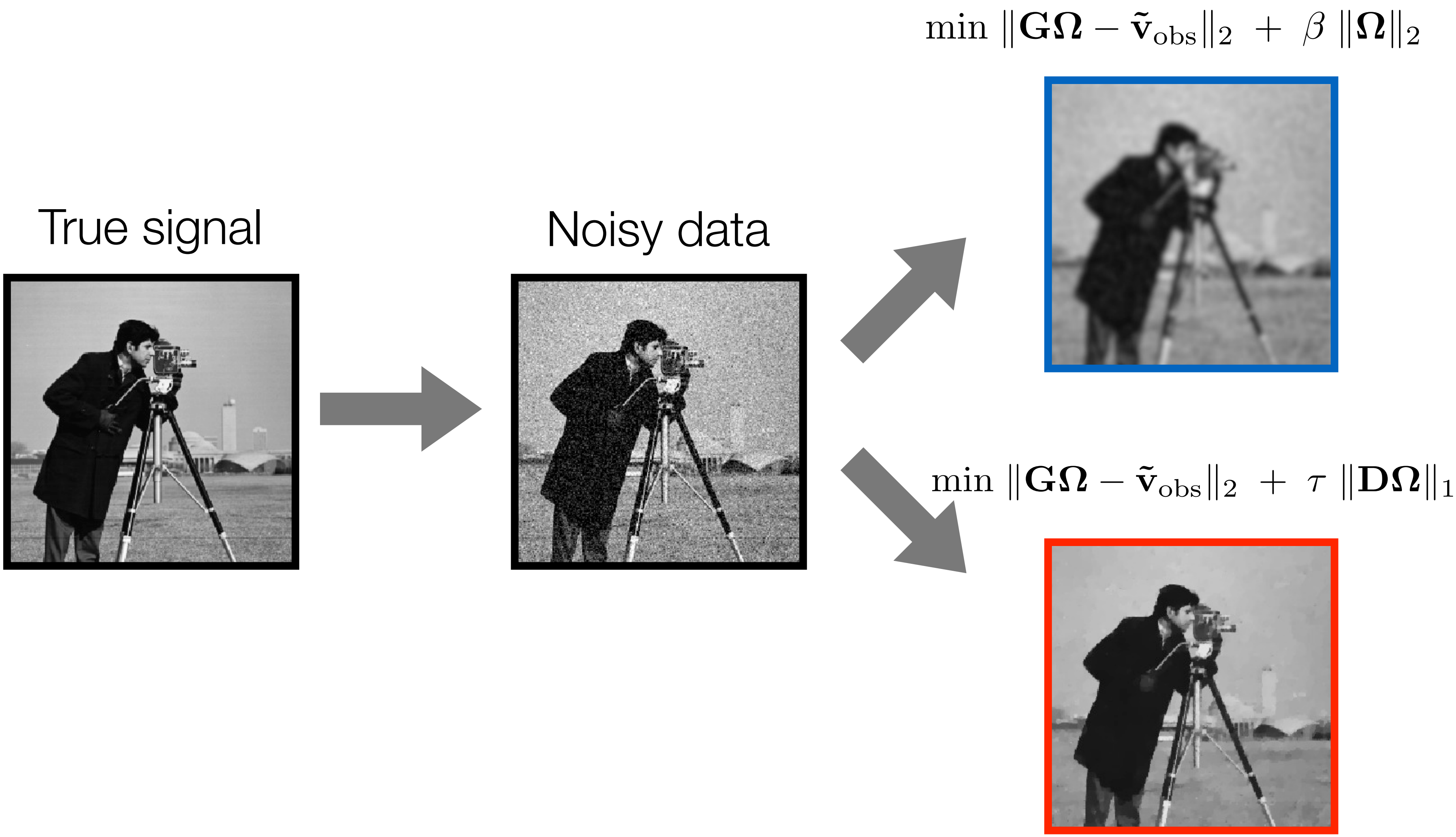


The combination of elasticity and effectively random GPS locations gives rise to a power law frequency distribution of partial derivatives. This distribution is known to support sparse solutions (*Davenport et al*., 2011)

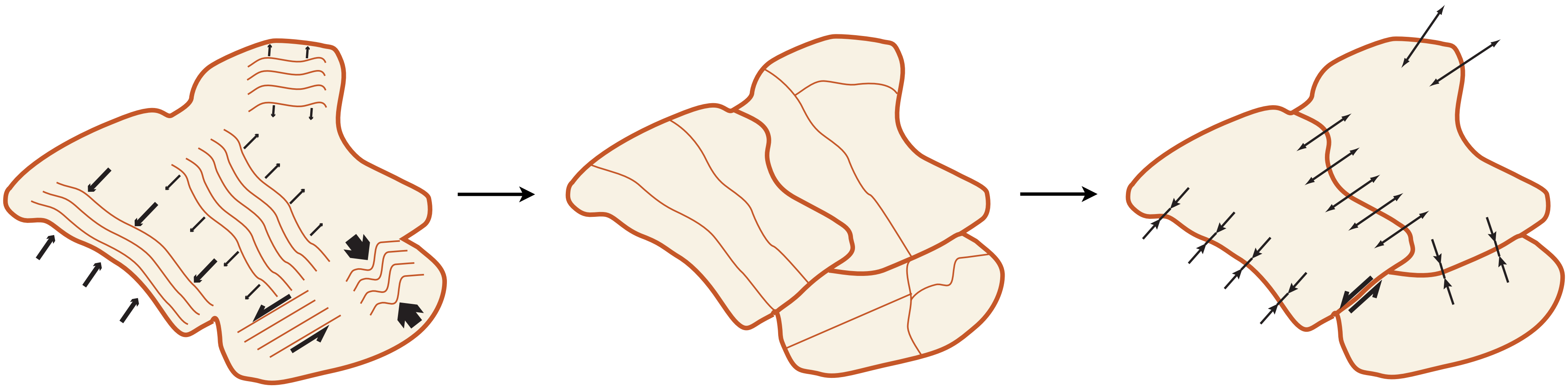
Selection by a modified version of sparse recovery



Selection by a modified version of sparse recovery



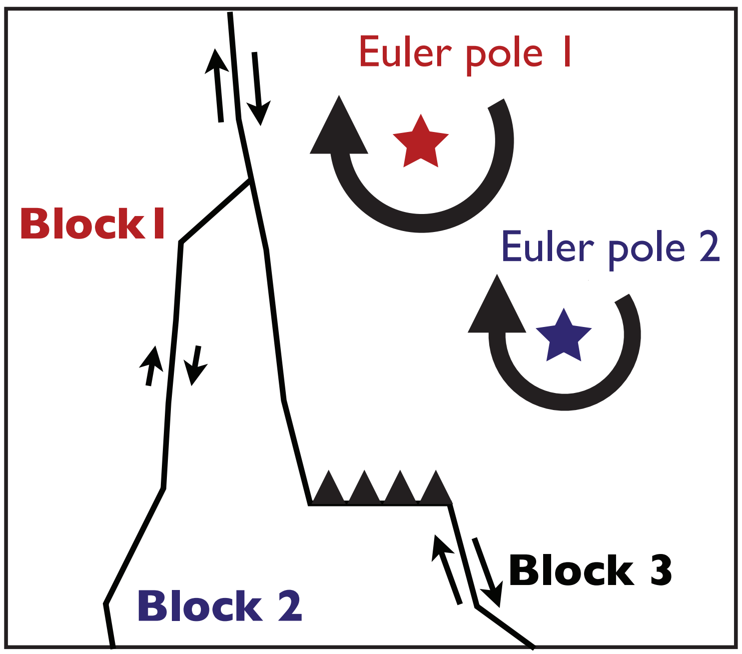
How kinematically complex are plate boundaries?



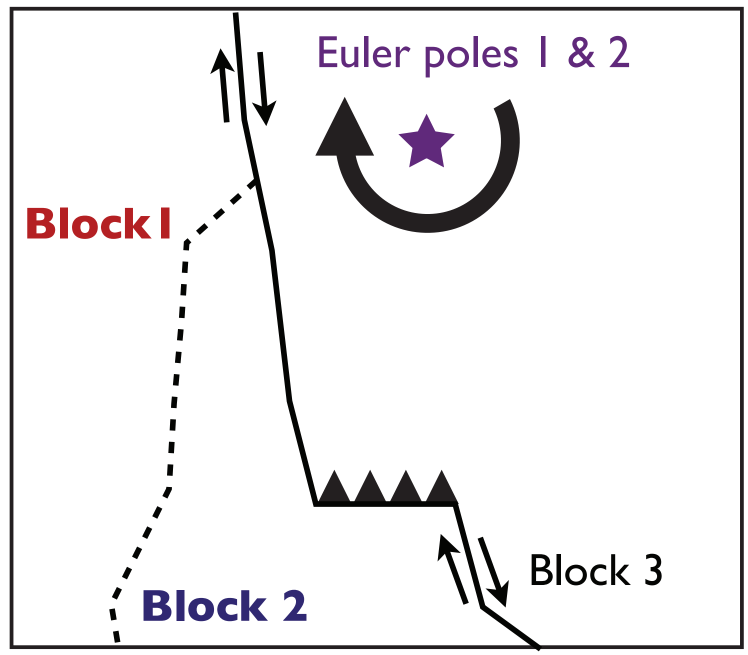
Thatcher (2009)

How kinematically complex are plate boundaries?

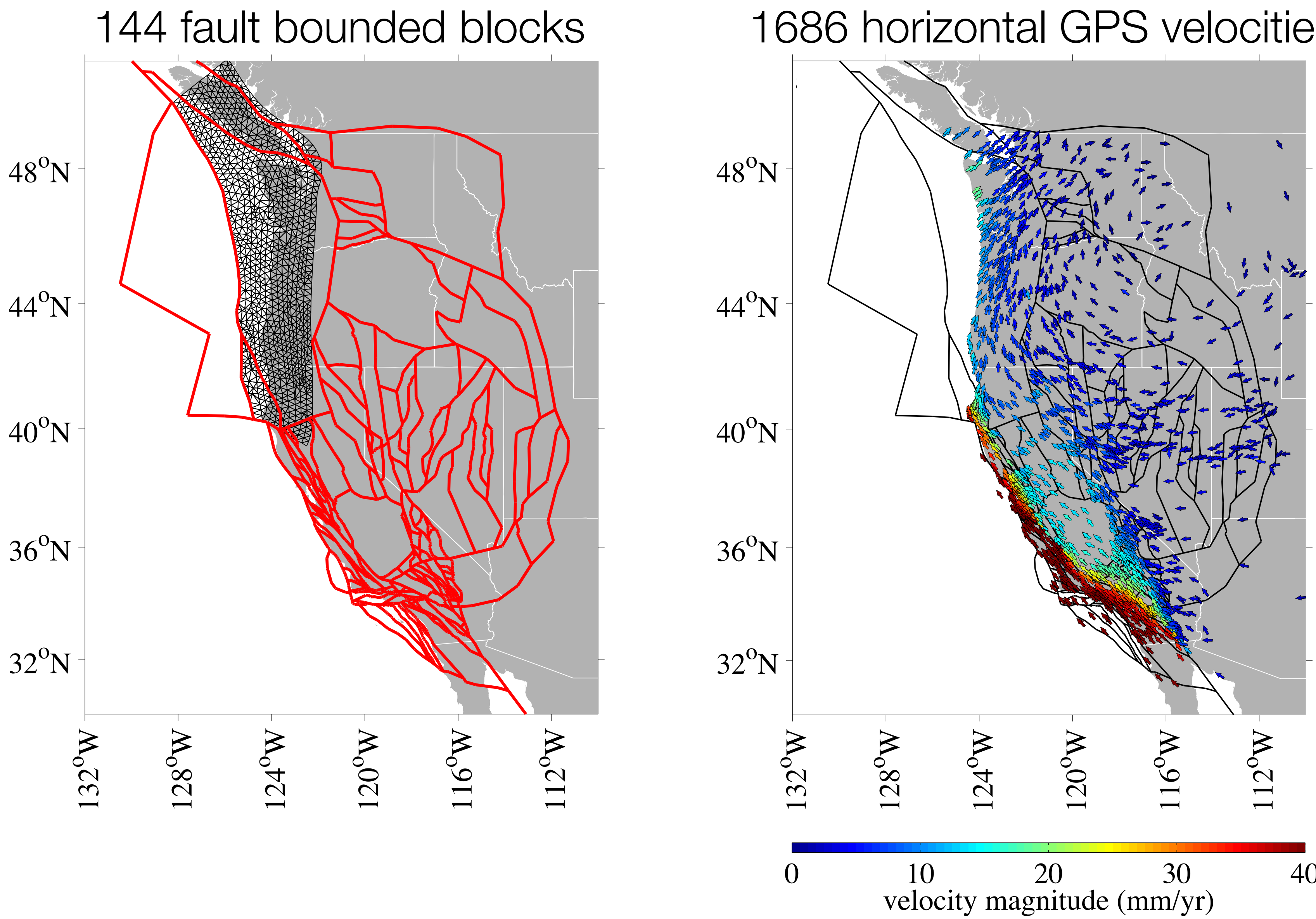
More complex
3 fault bounded blocks
3 Euler poles



Less complex
3 fault bounded blocks
2 Euler poles

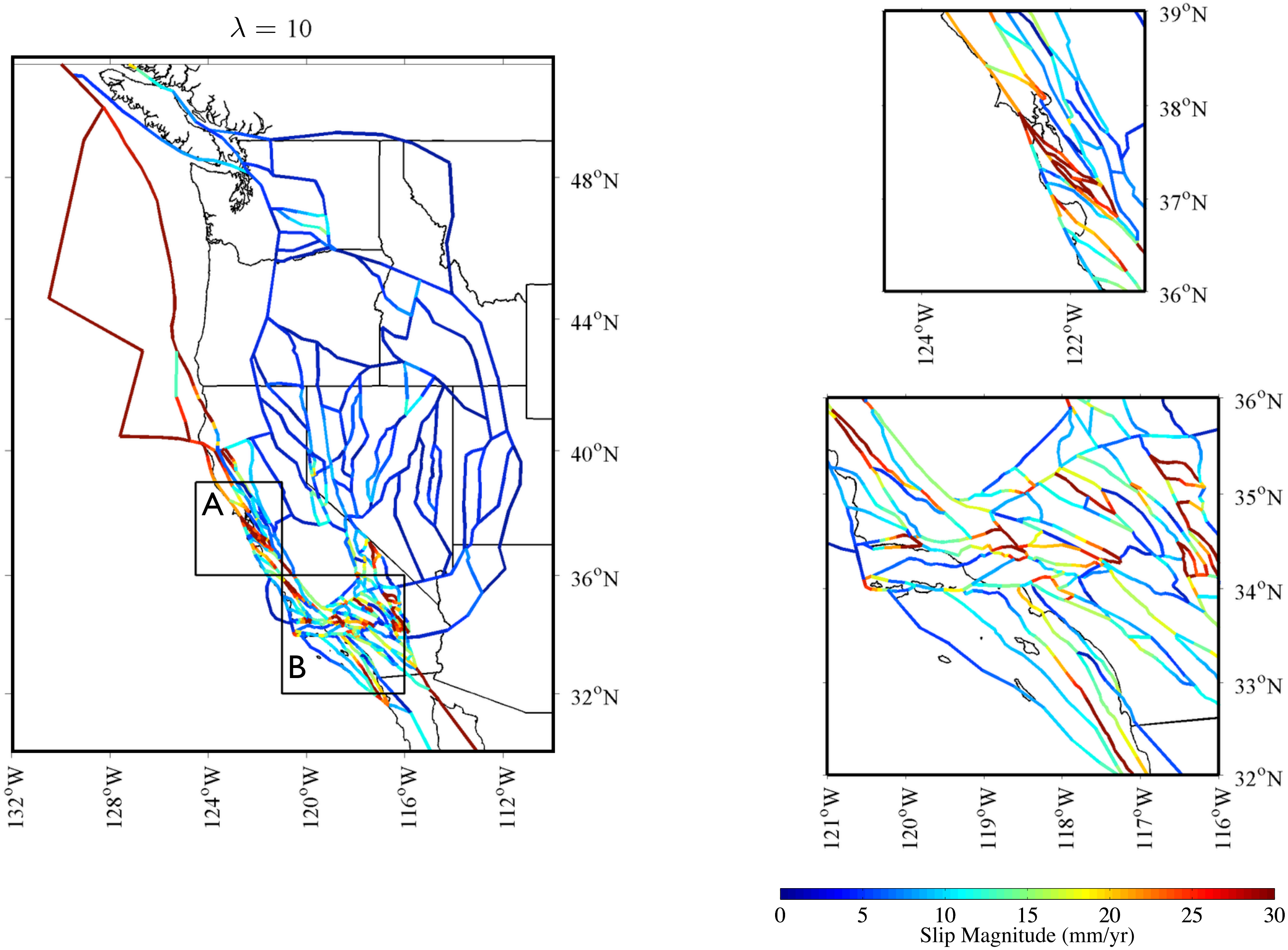


Selection by a modified version of sparse recovery

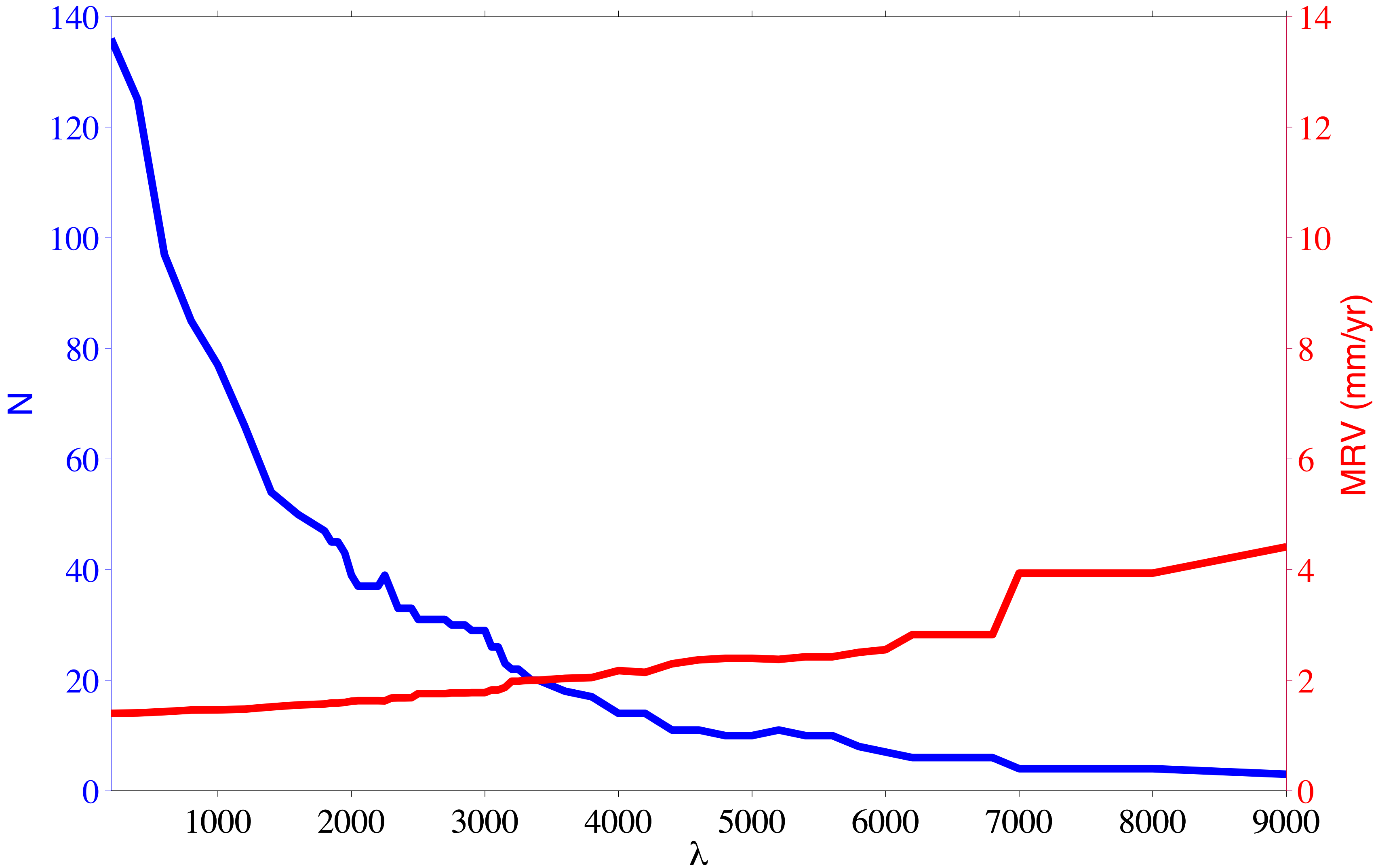


McClusky et al., (2001); *Shen et al*., (2003); *Hammond and Thatcher*, (2005); *Williams et al.*, (2006); *McCaffrey et al.*, (2007); and PBO

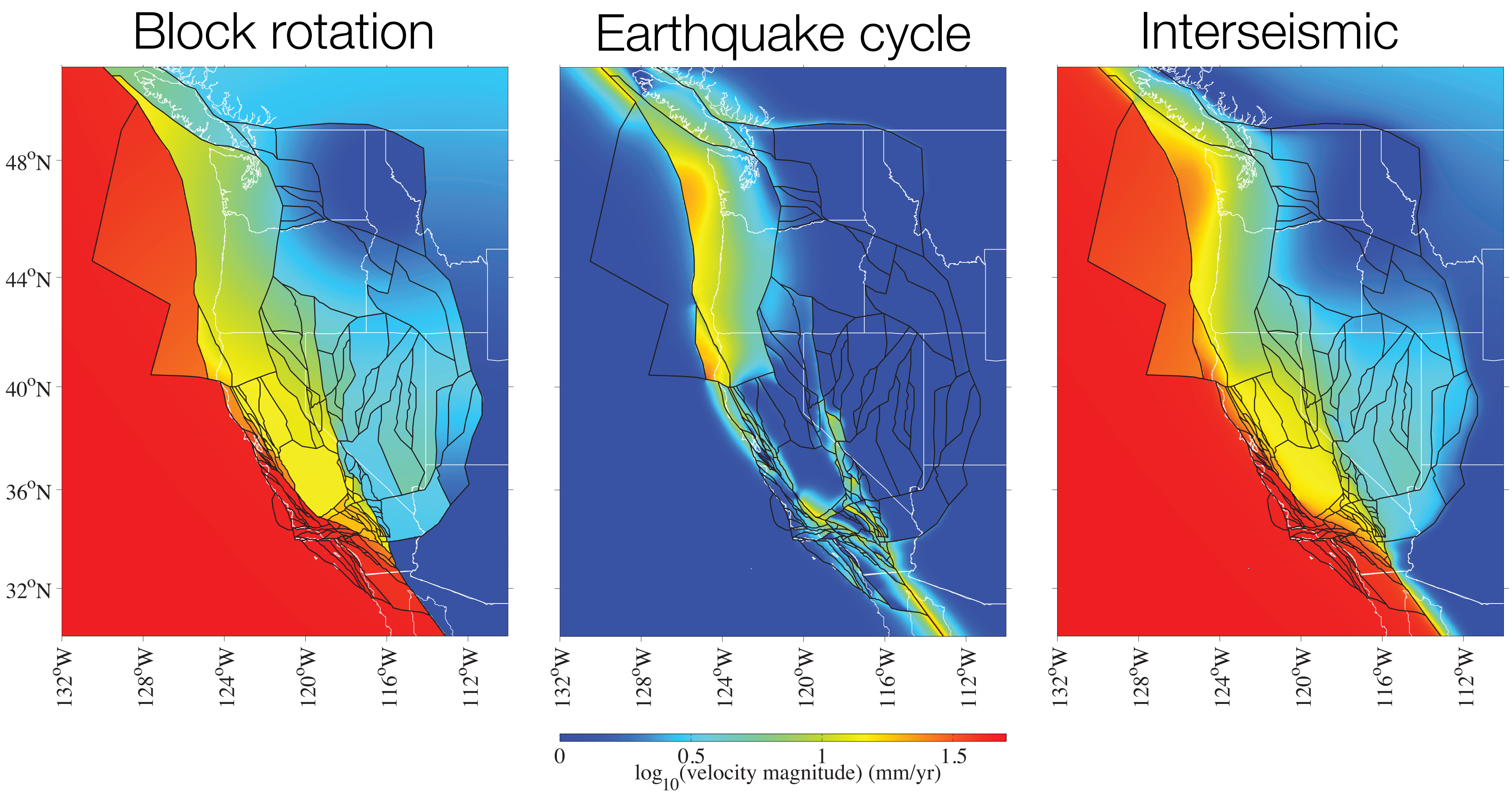
Selection by a modified version of sparse recovery



How many active plates are required?



The active fault system of the western US

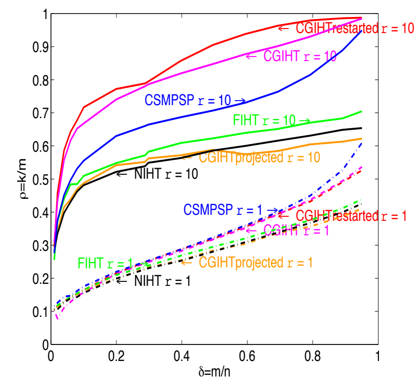


Evans and Meade, (submitted)

Where are we?

1) Sparse recovery algorithms can perform some model selection and recovery many of the things that we've always been interested in

2) Algorithm development is very rapid. Dantzig selector (*Candes et al.*, 2007) 1000 times slower than spectral projection algorithm just one year later (*Friedlander and van den Berg*, 2008)!

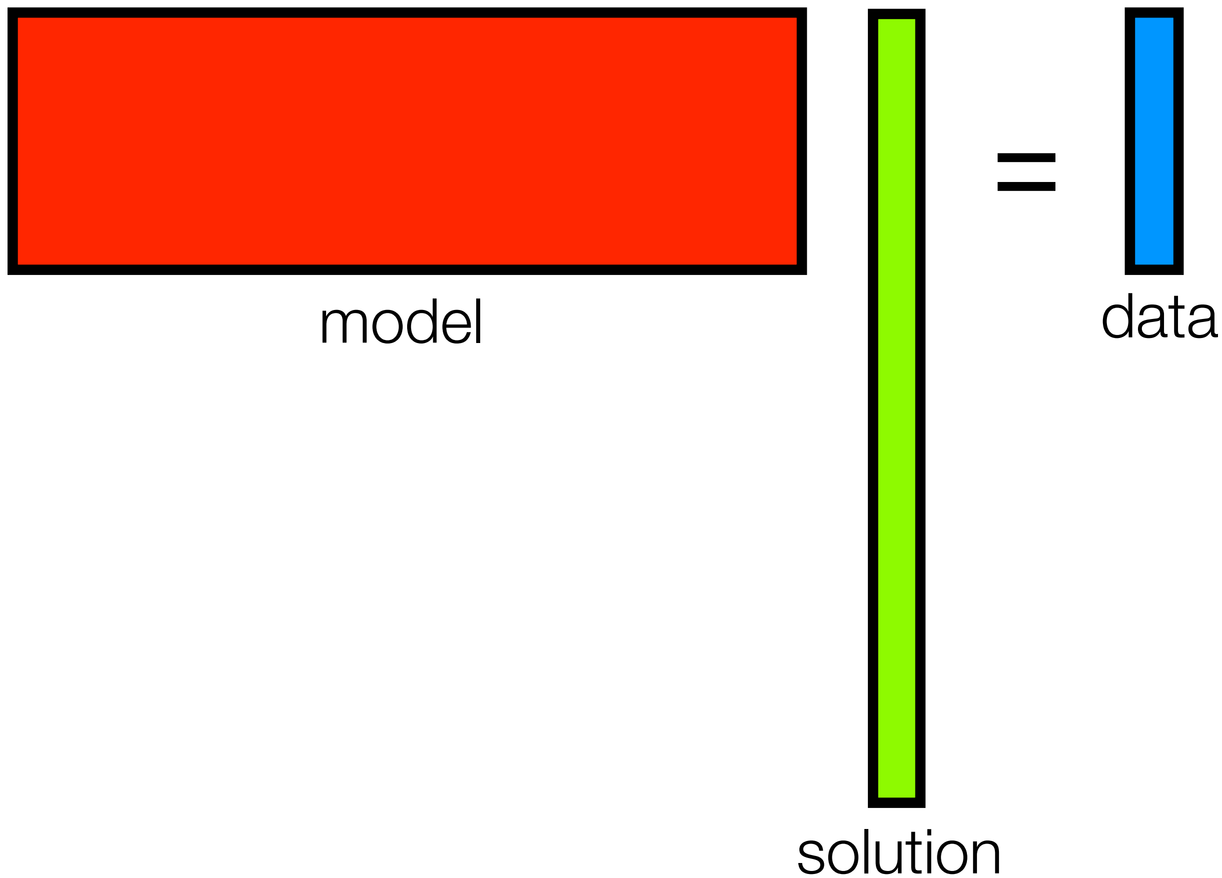


3) Empirical conditions for successful recovery rapidly evolving too.

Dimension	Time	
	SDPT3	TFOCS
128	0.3 s	0.3 s
512	2.2 s	0.3 s
1024	16.0 s	0.5 s
2048	145.0 s	0.7 s
4096	N/A	1.0 s
16384	N/A	2.9 s
131072	N/A	40.2 s
1048576	N/A	838.5 s

4) Large problems now becoming possible as Prony style issues are overcome.

We almost always solve underdetermined problems



Prony had a precursor >200 years ago…

Approximate signals with exponentially damped cosines (1795)

$$\hat{f}(t) = \sum_{k=1}^{N} a_k e^{b_k t} \cos(2\pi c_k t + d_k)$$

Proposed recovery algorithm only stable up to N = 25 and, curiously, returned estimates that were 50% zeros and 50% non-zeros.

Prony had a precursor >200 years ago…

Approximate signals with exponentially damped cosines (1795)

$$\hat{f}(t) = \sum_{k=1}^{N} a_k e^{b_k t} \cos(2\pi c_k t + d_k)$$

Proposed recovery algorithm only stable up to N = 25 and, curiously, returned estimates that were 50% zeros and 50% non-zeros

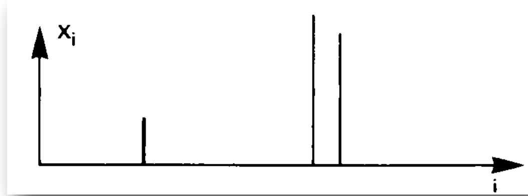
Legendre (1795) Clear statement of least squares, turned out to be somewhat popular

Early CS developments

1948 - *Dantzig*
Simplex algorithm for linear programming

$$\begin{array}{ll} \text{maximize} & \mathbf{c}^T\mathbf{x} \\ \text{subject to} & A\mathbf{x} \leq \mathbf{b} \\ \text{and} & \mathbf{x} \geq \mathbf{0} \end{array}$$

1973 - *Claerbout and Muir*
Linear programming for sparse state vectors



1984 - *Karmarkar*
Interior point methods make linear and quadratic programming fast

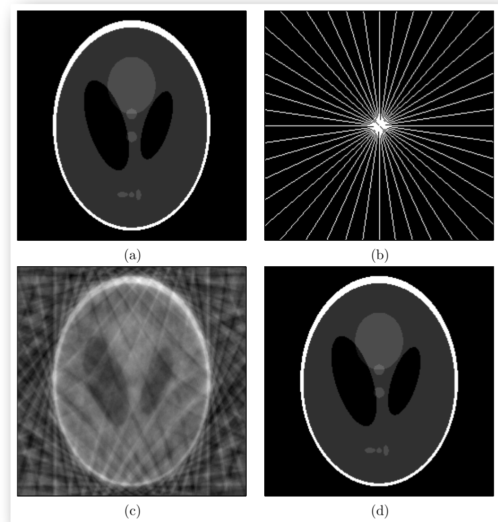
1995 - *Chen et al*.
Mathematicians start to take notice

- *Sparsity.* We should obtain the sparsest possible representation of the object — the one with the fewest significant coefficients.
- *Superresolution.* We should obtain a resolution of sparse objects that is much higher-resolution than that possible with traditional non-adaptive approaches.

Growth of CS theory

1996 - *Tibshirani*
Connection to quadratic programming

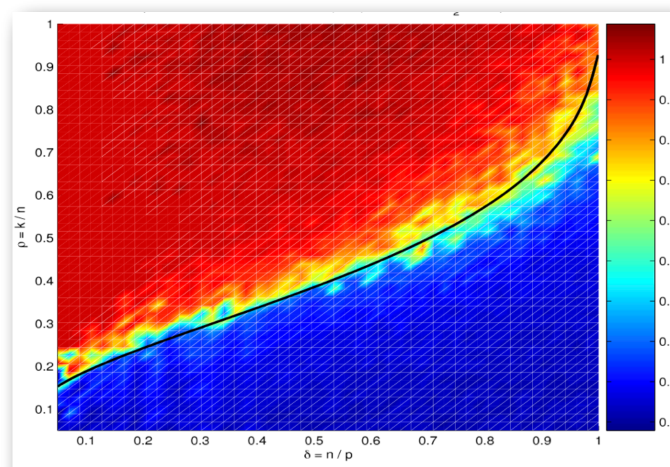
$$\min \|\mathbf{Gm} - \mathbf{d}\|_2 \quad \text{subject to } \|\mathbf{m}\|_1 \leq \tau$$



2005 - *Candes et al*.
Conditions for exact reconstruction, $k/n < 0.01$

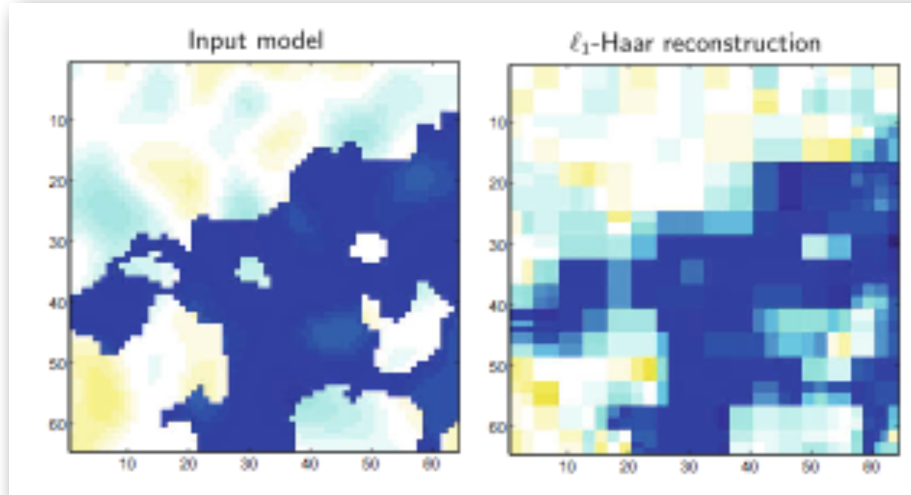
2008 - *van den Berg and Friedlander*
Fast & robust spectral gradient methods

2009 - *Donoho and Tanner*
Broader recovery conditions, $k/n < 0.30$

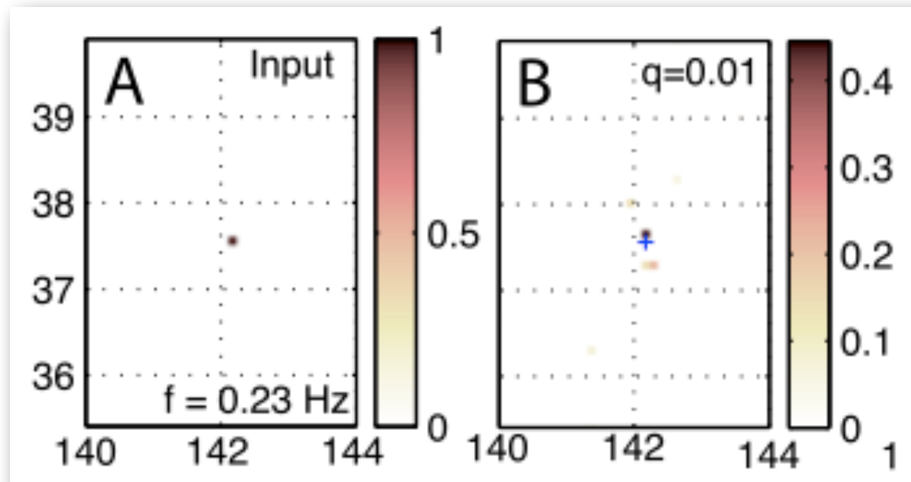


Some recent intentional CS in solid Earth geophysics

2011 - *Loris et al*.
Synthetic tomography



2011 - *Simons et al*.
Setup for global tomography

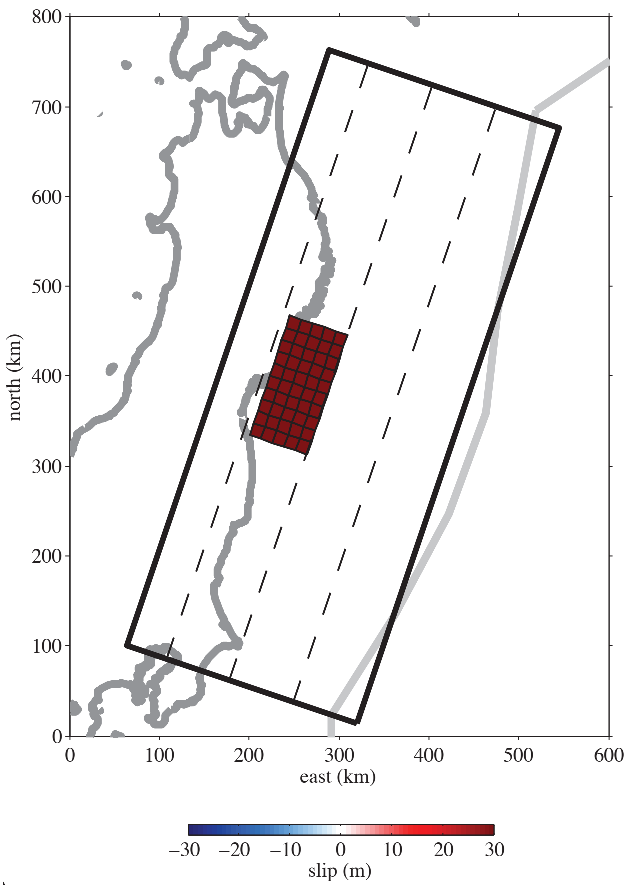


2011 - *Yao et al*.
Tohoku dominant frequencies

2012 - *Evans and Meade*
Tohoku coseismic and postseismic slip

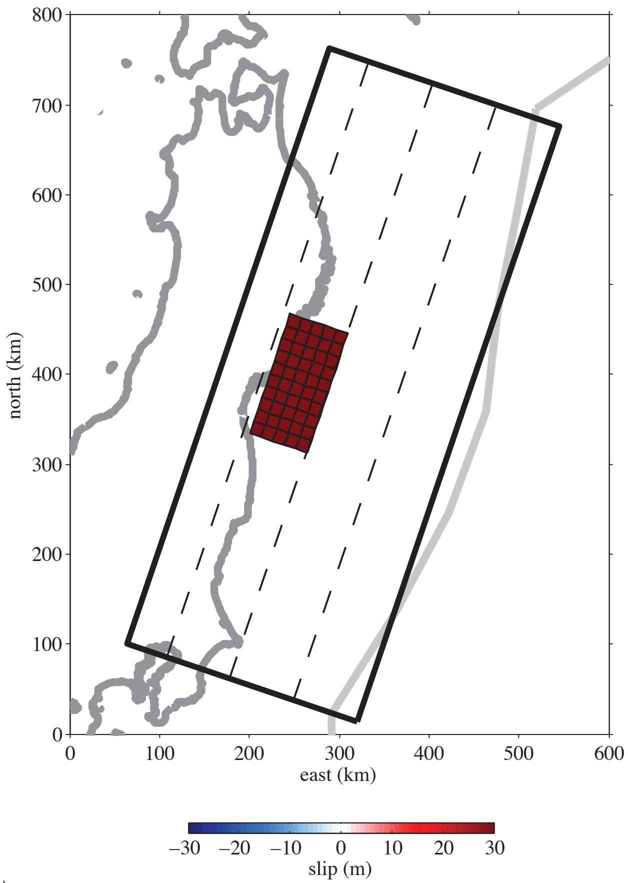
Resolution test - block

Synthetic slip

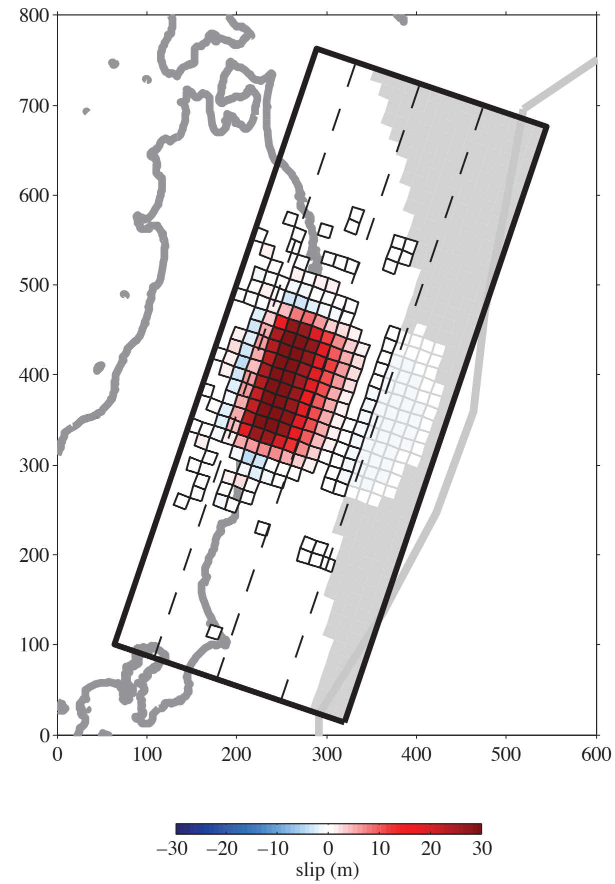


Resolution test - block

Synthetic slip



Smooth recovery



The role of state vector regularization

Damped least squares: L_2 regularization

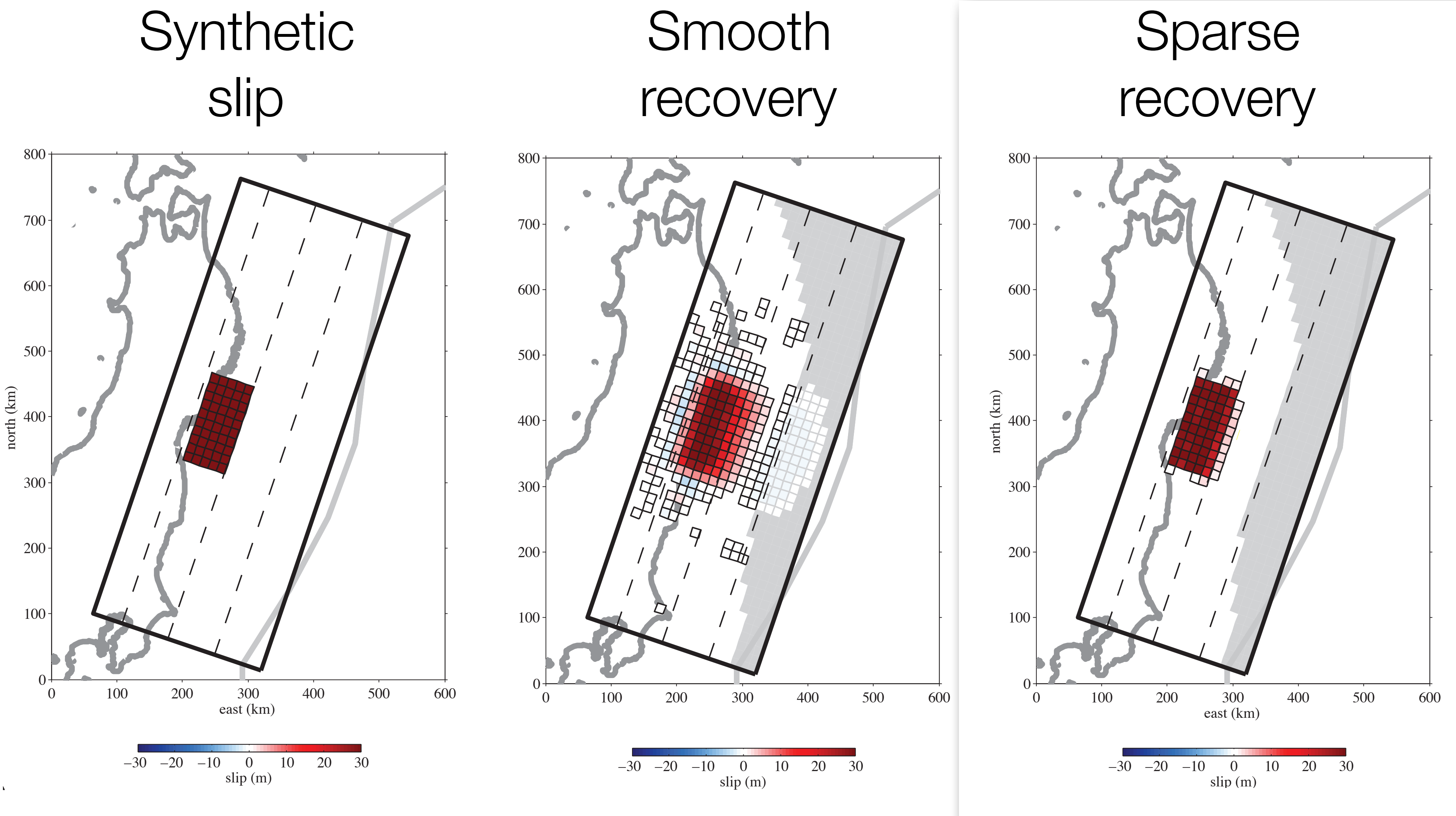
$$\min \|\mathbf{Gm} - \mathbf{d}\|_2 + \boxed{\lambda\|\mathbf{m}\|_2}$$ damp oscillations

Solutions vary smoothly in space (common in various formulations)

Classical approach with exact single step solution

So what about the other extreme; a not necessarily smooth and sparse/ compact solution?

Resolution test - block



Recovery of localized signals is possible with current GPS station spacing if signal is sufficiently sparse

Evans and Meade, (2012)