Coherence metric for optimal compressive sensing

J. Ray[1] , J. Lee[1] , S. Lefantzi[1] and S. A. McKenna[2]

{jairay, jlee3, slefant} [at] sandia [dot] gov

[1]Sandia National Laboratories, Livermore, CA

[2]IBM Research, Ireland

Funded by the LDRD program in Sandia National Labs

SAND2013-1591C

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Aim of the talk

- Aim: Interpret a physics-based linear inverse problem in terms of compressive sensing
  - The inverse problem involves estimating a high-dimensional field, not parameters
  - To explain why it works and the degree of inefficiency
  - Define metrics that help quantify efficiency of reconstruction
- Motivation
  - Compressive Sensing (CS) and associated sparse reconstruction techniques are very efficient and practical means of sampling random fields
    - Impose no pre-conditions (like smoothness etc.) on the field being sampled
  - Since all measurements are approximate, a CS interpretations may allow us to
    - Define the degree of approximation
    - Impact on the accuracy of inversion

Sandia National Laboratories

Outline of the talk

- What is compressive sensing, some basic concepts and terms
- Explanation of the physics-based linear inverse problem
  - Estimation of fossil-fuel CO_2 (ffCO2) emissions in the US
  - Temporally & spatially varying field
- Demonstration of inversion using an adaptation of CS techniques
- Explanation of why it worked
  - And why it could not have worked with the adaptation

What is compressive sensing?

- CS – a way of measuring a signal **e** very efficiently and then reconstructing it
  - Requires far fewer samples that Nyquist sampling
- If **e** is a signal/image of size N, and can be represented sparsely in some orthogonal basis set Φ
  - **e** = Φ w, where on K << N elements of w are non-zero
  - then the # of compressive samples needed is
    - $M = C\ K \log_2(N/K),\ C \sim 4$
- Compressive samples y^{obs} are given by
  - $y^{obs} = \Psi\ e = \Psi\ \Phi\ w = \mathbf{G}\ w$
  - each row $\psi_{r,.}$ is a random unit vector

Sandia National Laboratories

Sparse reconstruction

- How does one recover **e** from y^{obs}?
  - By exploiting the fact that **e** is sparse representable in Φ
- One minimizes, w.r.t. w
  - $|y^{obs} - \Psi \Phi w|_2 + \lambda |w|_1 = |y^{obs} - \mathbf{G}w|_2 + \lambda |w|_1$
  - Reduces the model – observation mismatch while penalizing for non-zero w using $|w|_1$
- Alternatively, w.r.t. w
  - min $|w|_1$ under the constraint $|y^{obs} - \mathbf{G}w|_2 < \varepsilon$
- Many convex optimization methods do this, not necessarily fast
  - Basis pursuit, orthogonal matching pursuit etc.
- Reconstruction uses no "crutch" / prior / regularization in the estimation problem, beyond sparsity
  - The observations really have to be informative to do this ….

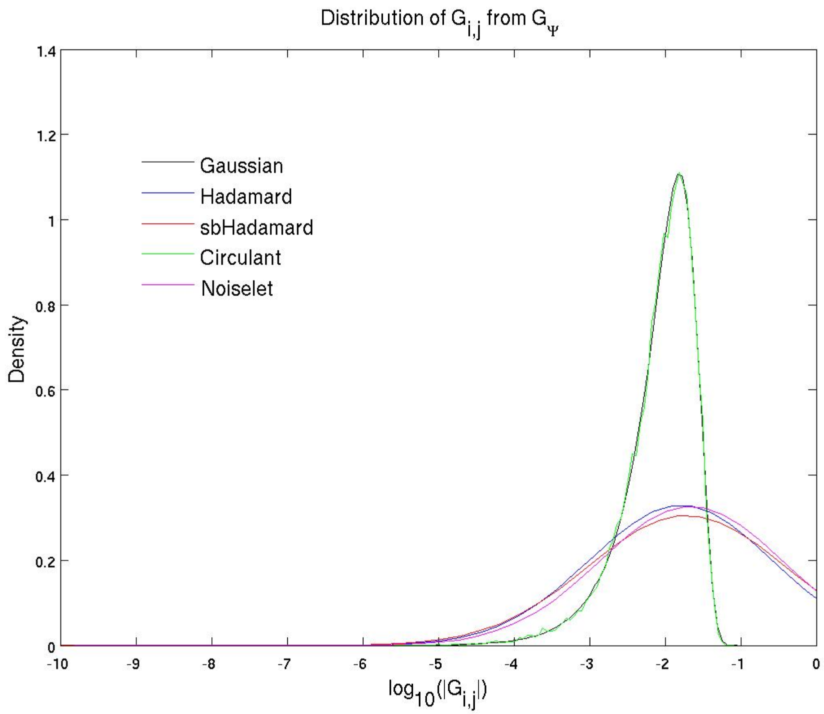
Why is CS so efficient in sampling?

- $y^{obs} = \Psi\, e = \Psi\, \Phi\, w = \mathbf{G}\, \mathbf{w}$
  - Each row $\psi_{r,.}$ collects information on all columns $\phi_{.,c}$
  - If $\psi_{r,.}$ is random, it will be non-aligned with all $\phi_{.,c}$
- Called incoherence

$$\mu(\Psi,\Phi) = N^{1/2} \quad \max\left|\left\langle \psi_{r,.}\phi_{.,c}\right\rangle\right|$$

$$= N^{1/2} \quad \max\left|G_{r,c}\right|$$

  - $1 < \mu(\Psi, \Phi) < N^{1/2}$



- In image processing, Ψ are "standard" random matrices like Bernoulli, Toeplitz etc., or noiselets
  - $\mu \sim 4$

Stability of sparse reconstruction

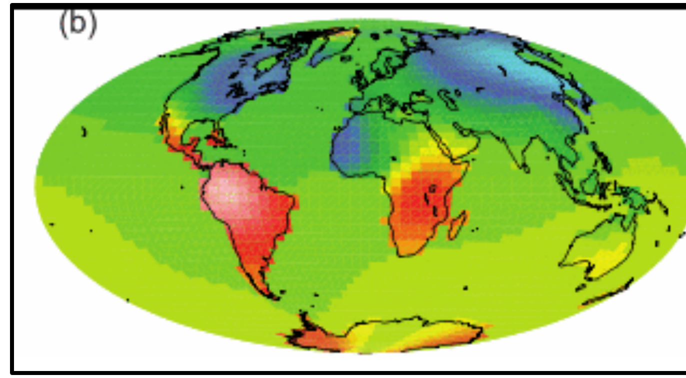
- Given so few measurements, why is the sparse reconstruction stable?
  - $\mathbf{w} = \mathbf{w}_{true} + \mathbf{n}$, both $\mathbf{w}_{true}$ and $\mathbf{n}$ are K-sparse
  - $y^{obs} = G(\mathbf{w}_{true} + \mathbf{n}) = \mathbf{G}\mathbf{w}_{true} + \mathbf{G}\mathbf{n}$
- For stability
  - $(1 - \delta)\ |x|_2 < |Gx|_2 < (1 + \delta)\ |x|_2$
  - Called Restricted Isometry Property (RIP) of $G = \Psi\Phi$
- A more conservative definition
  - If the columns of G are nearly orthogonal to each other, we have RIP
  - Alternatively, non-diagonal elements of G^TG are far away from 1
    - Max (G^TG) is around 0.25

Matrix	Max($\|G^TG\|$)
Gaussian	0.27
Hadamard	0.28
Scrambled-block Hadamard	0.489
Circulant	0.478
noiselets	0.3

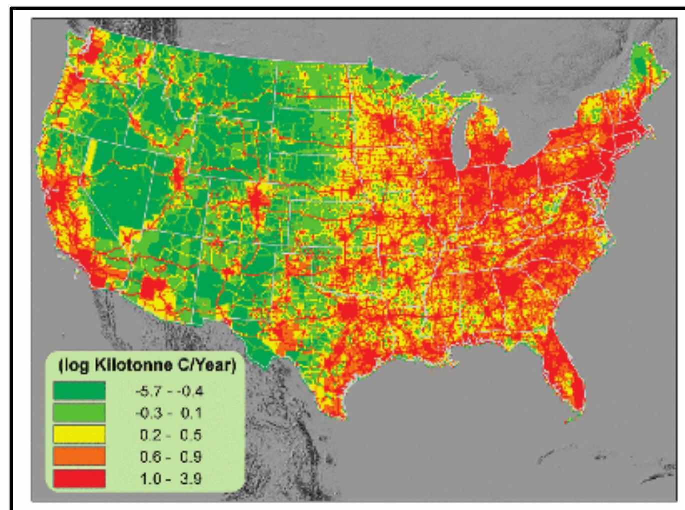
Sandia National Laboratories

Using CS ideas in a physics inversion

- Aim: Devise a method to estimate fossil-fuel CO_2 (ff-CO2) emissions
  - Data: measurements of ff-CO2 concentrations at a sparse set of sensors
- Motivation
  - Monitoring emission & cap-and-trade treaties
  - Updating global process-based inventories of ff-CO2 emissions
- Technical challenge
  - ff-CO2 emissions have a rough, non-stationary spatial distribution
    - Current smooth models (for estimating biogenic CO2 fluxes) don't work



Biogenic emissions: Mueller et al, *JGR*, 2008



Anthropogenic emissions: Gurney et al, *EST*, 2009

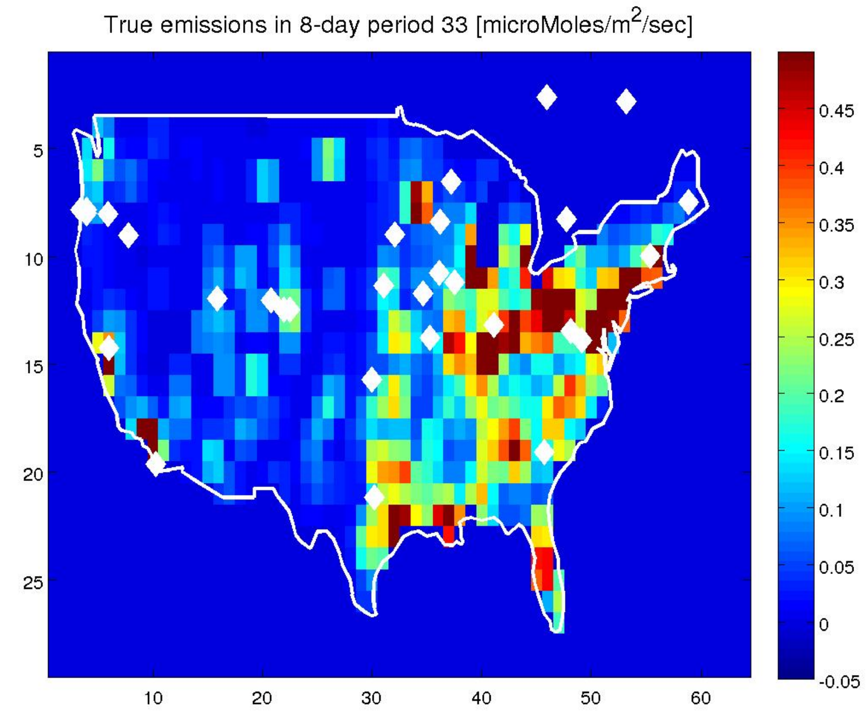
Sandia National Laboratories

Characteristic of the ff-CO2 estimation problem

- Linear inverse problem
  - $y^{obs} = \mathbf{H}\,\mathbf{e}(x, t)$, $\mathbf{H}$ = sensitivity matrix, $\mathbf{e}(x, t)$ = emissions
  - H – determined using atmospheric dispersion models
  - y^{obs} – measured at a set of CO2 measurement towers
- $\mathbf{e}(x, t)$ is non-stationary and non-smooth
  - Could be expressed with wavelets i.e. $\mathbf{e}(x, t) = \Phi\, w(t)$
  - $w(t)$ will be sparse; $\mathbf{e}(x, t)$ exists only where humans live (+ electricity generators)
- Could we solve $y^{obs} = \mathbf{H}\,\mathbf{e}$ using CS arguments?
  - H collects info from all emission sources / grid-cells; functions like Ψ
  - $y^{obs} = \mathbf{H}\,\mathbf{e} = \mathbf{H}\,\Phi\, w$ has the same formulation as CS
    - But what is the incoherence $\mu(\mathbf{H}, \Phi)$?
  - Sparse reconstruction could work

Posing the synthetic data inversion

- Aim: Estimate ff-CO_2 emissions in US
  - On a 1^o resolution, 64 x 64 mesh; assume zero non-US emissions
  - Weekly-averaged emissions
- Synthetic data, CO_2 concentrations @ 35 sensors, every 3 hours
  - True emissions – Vulcan database for US, 2002
  - Sensor measurements simulated using WRF
- Spatial model for emissions
  - $y^{obs} = \mathbf{H}\,\mathbf{e} = \mathbf{H}\,\mathbf{\Phi}\,w$
  - Φ modeled with Haar wavelets



Emissions for a week in August 2002 (Vulcan database, 1 degree resolution)

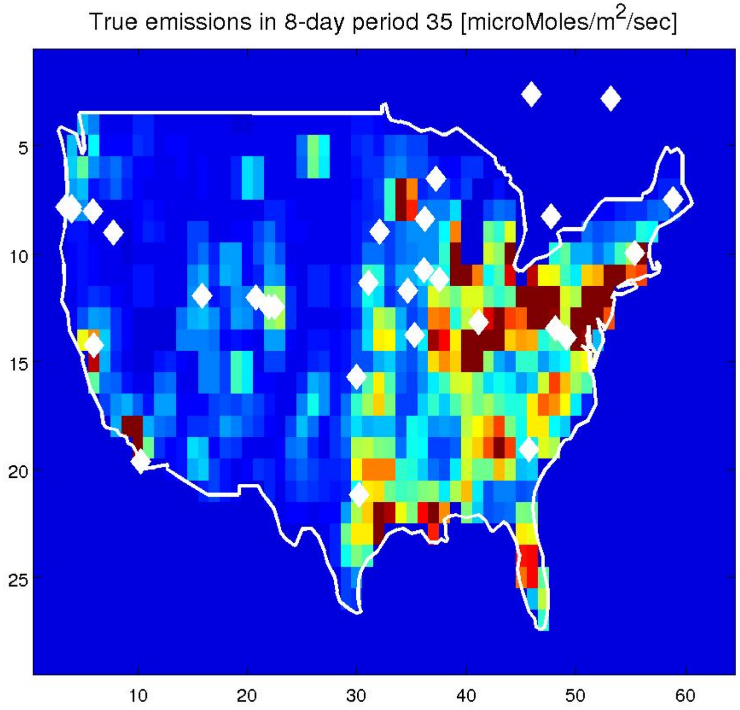
Dimensionality reduction

- A Haar wavelet model does not reduce dimensionality
  - 4096 coefficients to be estimated to model 1 week's emissions
- Nightlights (DoD's DMSP-OLS) are a good proxy for FF emissions
  - Except emissions from electricity generation and cement production
- Use thresholded radiance-calibrated nightlights from 1997-98 to mask out unpopulated regions
  - Reduce dimensionality from 4096 to 1031

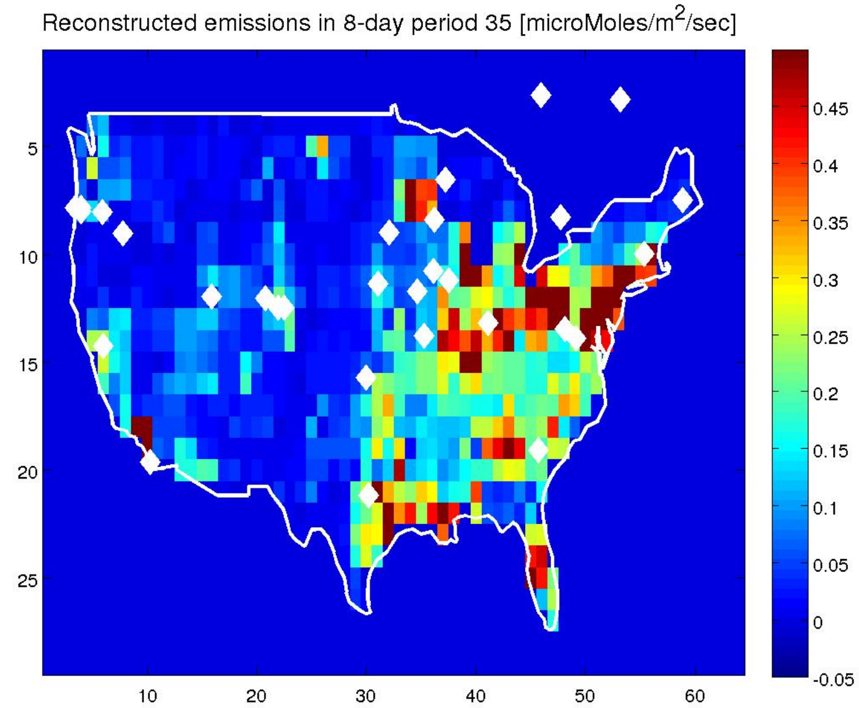
Introducing priors / regularization

- CS of images is done with nothing more than sparsity priors
  - ff-CO2 inversion failed with that
- Original inversion :
  - min $|w|_1$ such that $|y^{obs} - \mathbf{H}\,\boldsymbol{\Phi}\, w|_2 < \varepsilon$ (failed)
- Assume we have a model for emissions $e_{model} = \Phi\, w_{model}$
  - Easily made by scaling lights-at-night with a constant to match annual US emissions
  - $y^{obs} = \mathbf{H}\,\boldsymbol{\Phi}'\, w'$, $\Phi' = \mathrm{diag}(w_{model})\,\Phi$, $w' = w / w_{model}$
  - min $|w'|_1$ such that $|y^{obs} - \mathbf{H}\,\boldsymbol{\Phi}'\, w'|_2 < \varepsilon$
- The normalization of w with the model ensures that the estimated values do not deviate very much from w_{model}
  - Unless observations say otherwise
  - Basically, a prior or regularization

How good is the reconstruction?



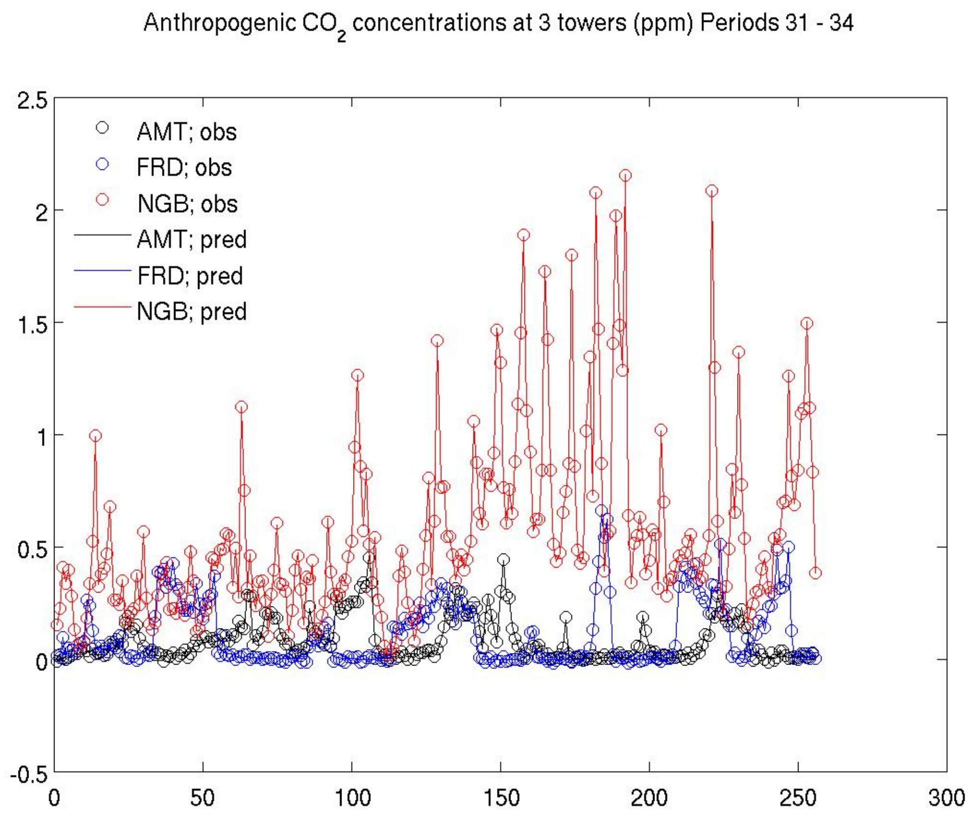
True emissions



Reconstructed emissions

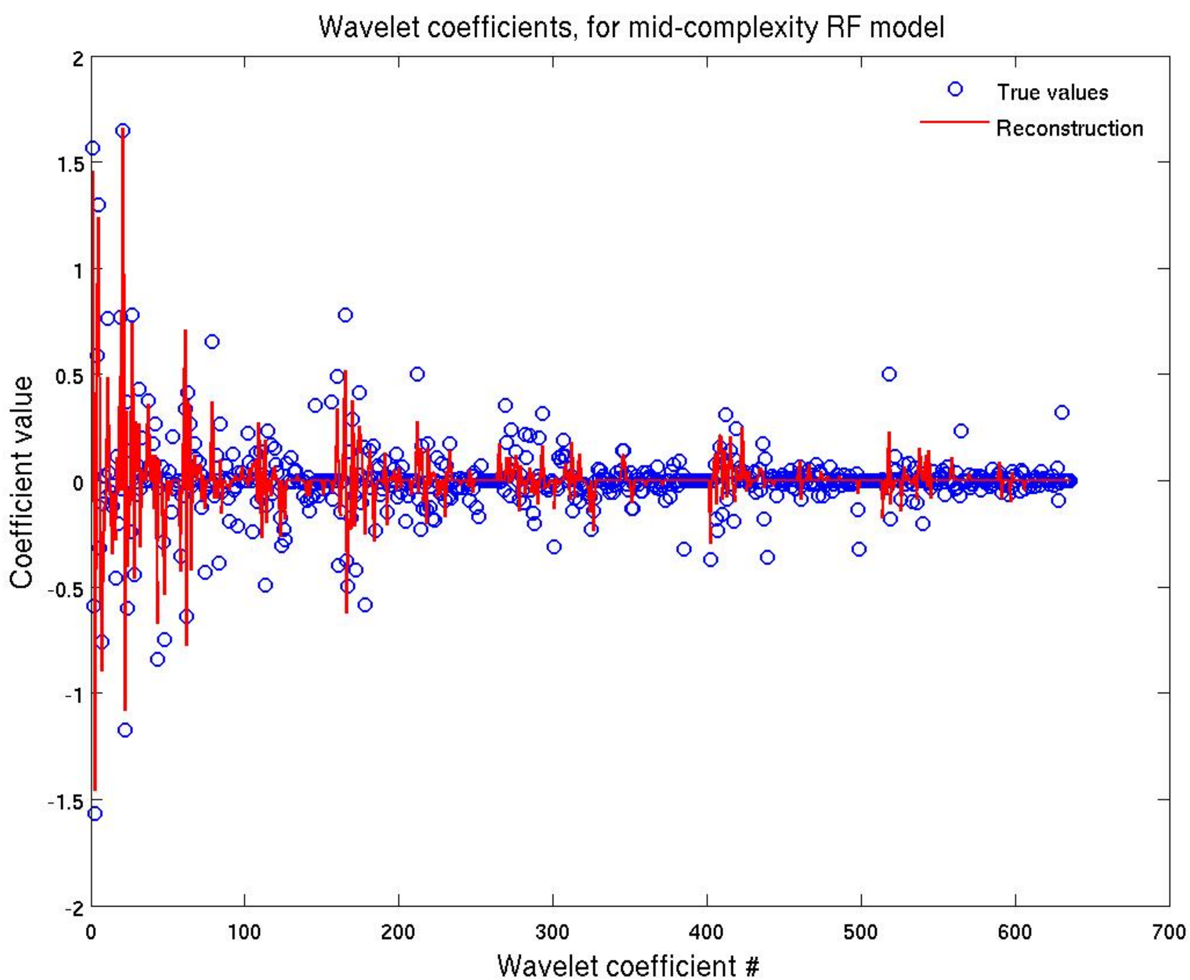
- A week in September 2002

Can we reproduce tower observations?



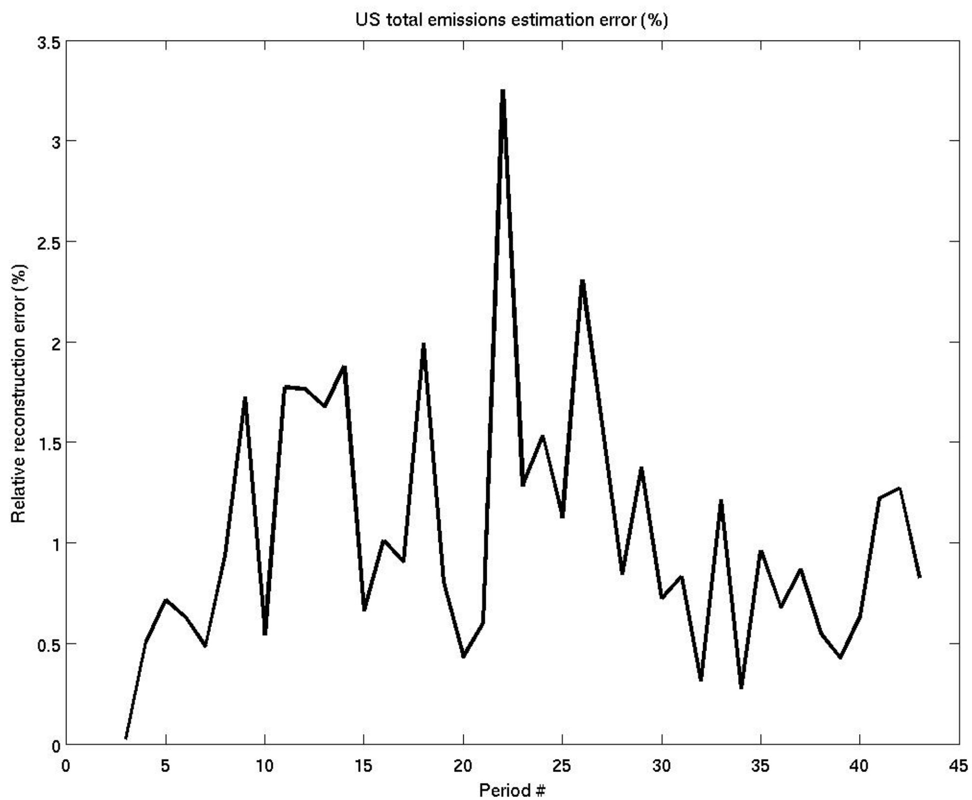
- Tower concentration predictions with reconstructed fluxes (only 3 weeks)
  - Symbols : observations used in the inverse problem.

Did sparsification work?



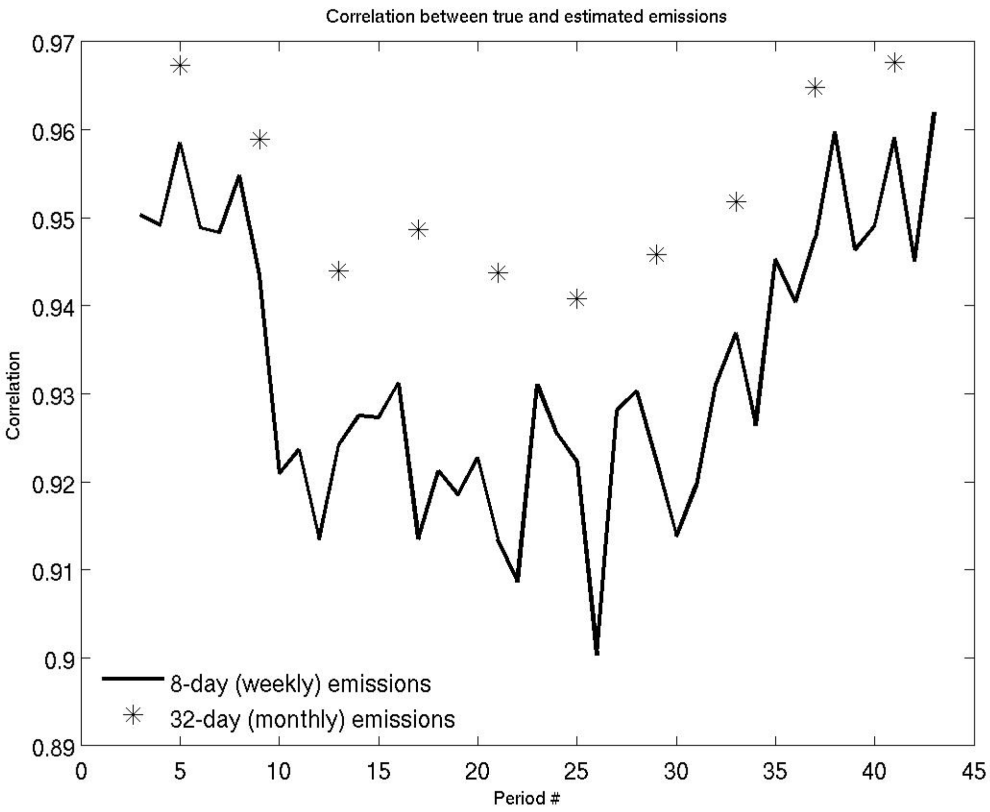
- Only about half the wavelets could be estimated
- We are probably not over-fitting the problem
  - Data-driven sparsification works

Reconstruction error in total US emission



- We get about 3.5% error, worst case

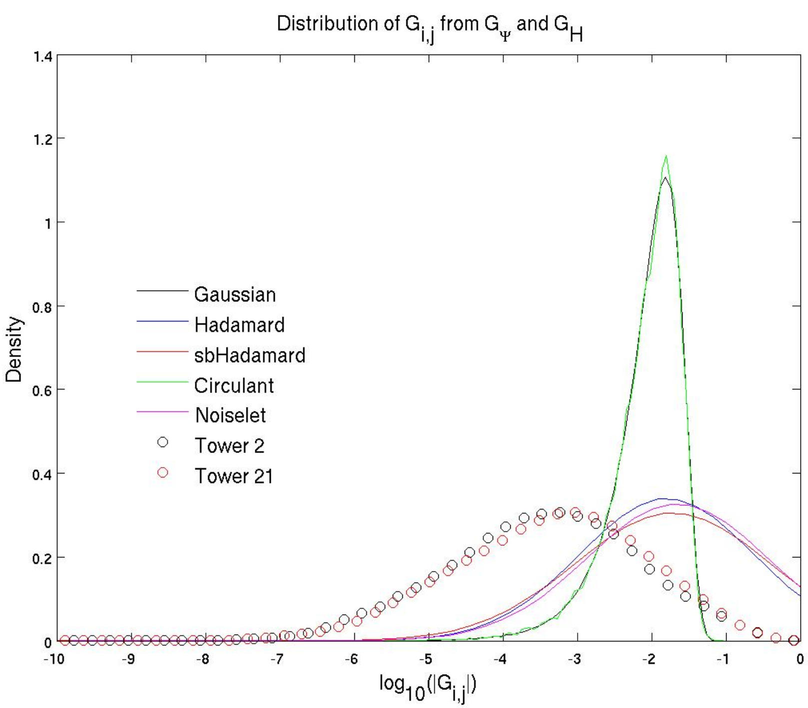
Is the spatial distribution correct?



- The spatial distribution of emissions is very close to truth
- Especially, if considering monthly fluxes

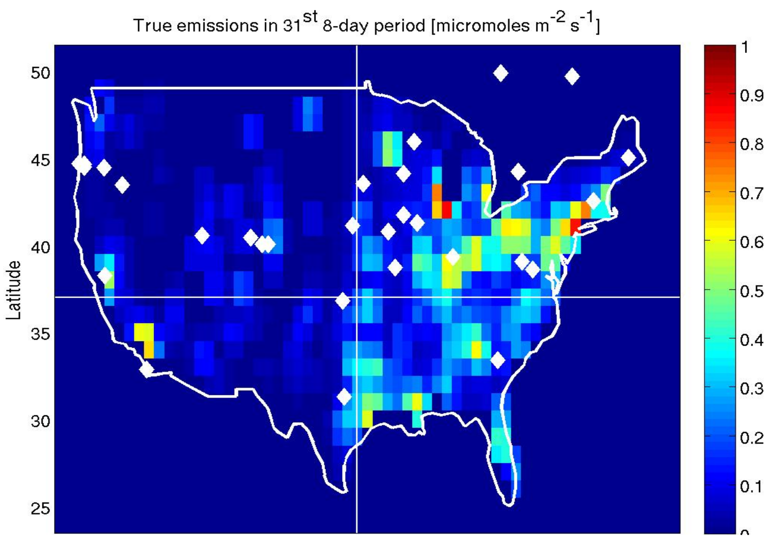
Why did this work?

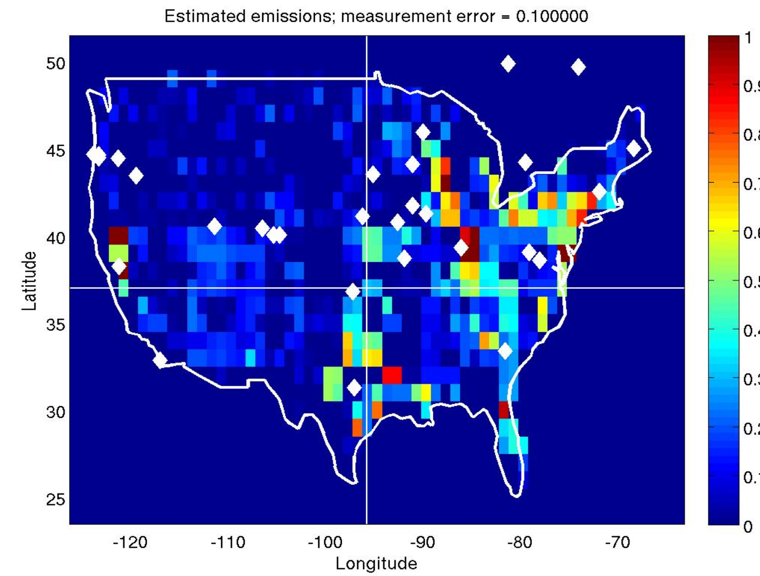
- Meteorology is not aligned with Haar wavelets; **H** and Φ should be incoherent.
  - Plot $|\mathbf{G}_{rc}|$, $\mathbf{G_H} = \mathbf{H}\,\Phi$
  - Compare against $G_\Psi = \Psi\,\Phi$
- There are some large $|\mathbf{G}_{rc}| \sim 1$
  - Sensor footprint is about 1500km, but very affected by the closest 30 km
    - Coherent with the Haar wavelet around the sensor
    - But only 1 wavelet / sensor
    - And only for sensors in the nightlight-bright regions

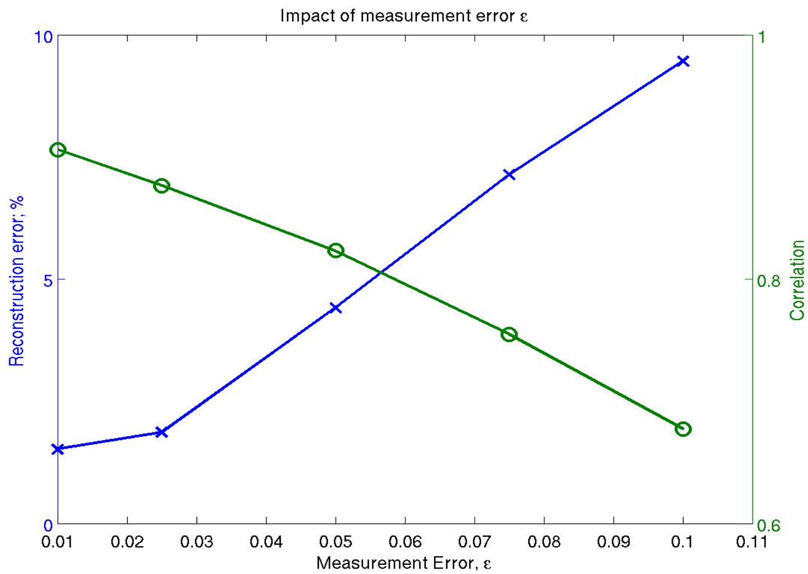


On the whole, (H, Φ) are incoherent

Impact of noise in y^{obs}







- Increasing observation noise 10x did not lead to obvious corruption
  - Only loss of detail & correlation
- Inversion seems stable – why?

Is the inversion stable to noise?

Matrix, Ψ	1st percentile	Median	75th percentile	99th percentile
Gaussian	6.4e-4	3.5e-3	5.9e-2	1.3e-1
Hadamard	0	3.3e-2	5.6e-2	1.25e-1
Circulant	6.4e-4	3.4e-2	5.9e-2	1.3e-1
Noiselet	0	2.1e-2	5.2e-2	1.5e-1
H, tower # 1	0	0	0	7.0e-2
H, tower # 21	0	0	0	1.9e-2

Statistics of $|G^T G|$ from different sampling matrices

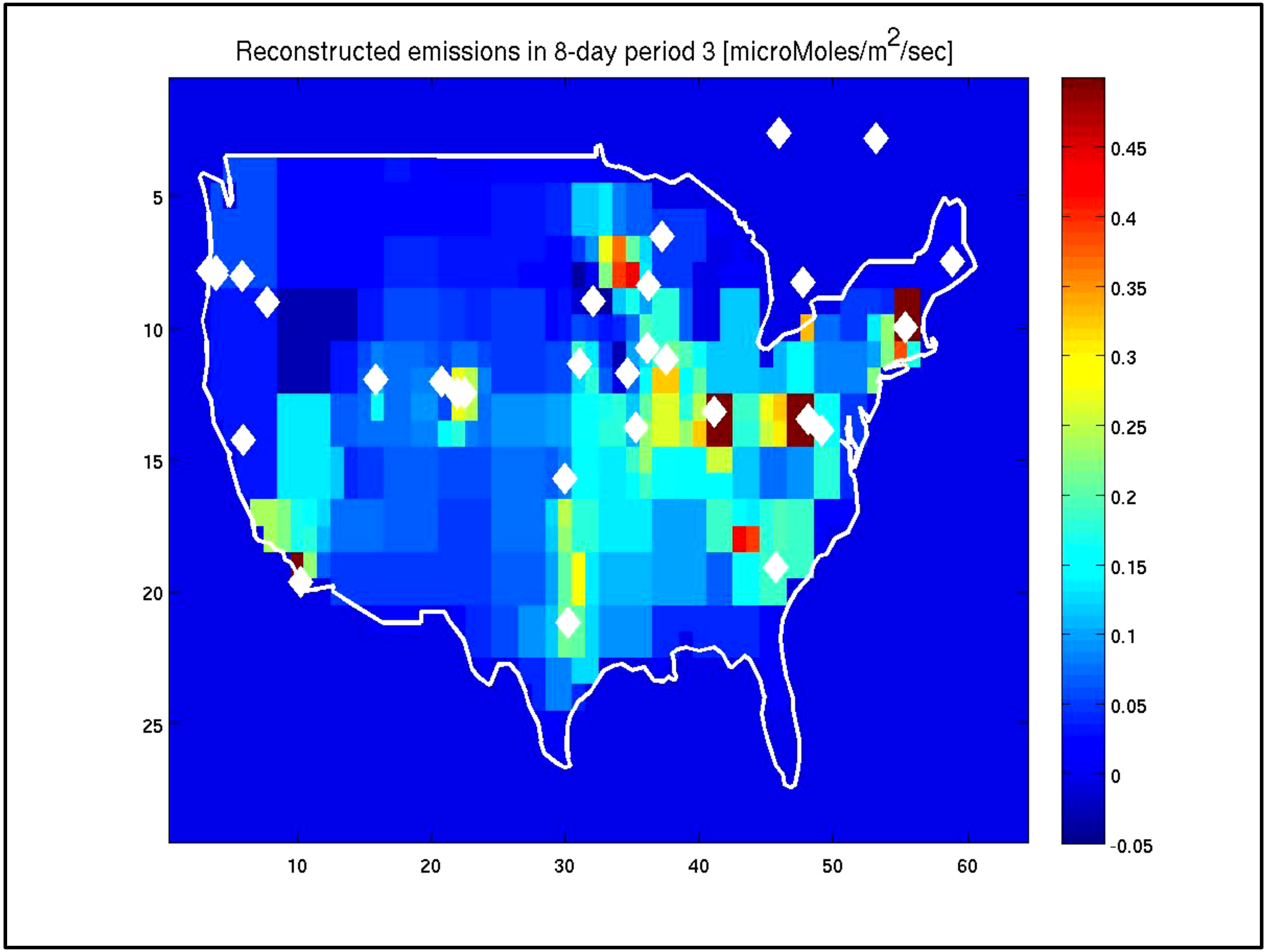
- RIP implies non-diagonal elements of $\mathbf{G}^T\mathbf{G}$, $\mathbf{G} = \mathbf{H}\,\Phi$, are small (away from 1)
  - Max value for CS samplers around 0.25
- RIP of **H** matrix is weak (max value of non-diagonal term ~ 0.99)
- That's why we needed a prior – sparsity-only was not sufficient

Sandia National Laboratories

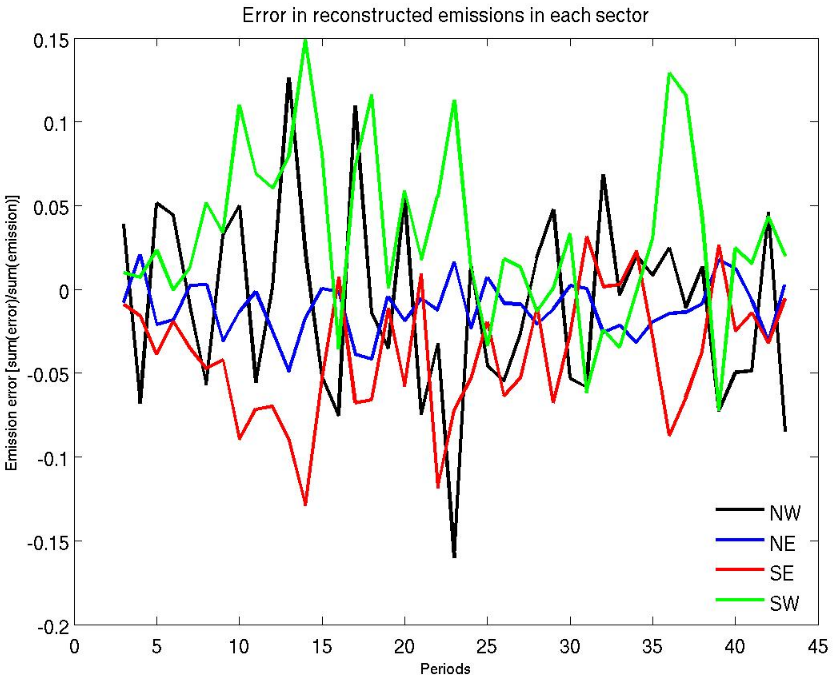
Conclusions

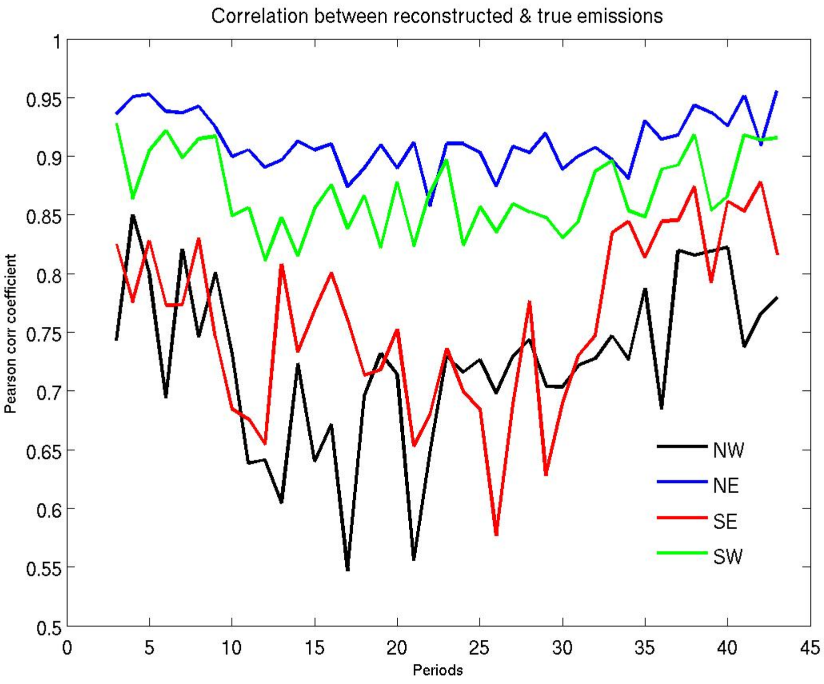
- CS provides a framework for interpreting physical inverse problems
  - Provides metrics for the efficiency of inversion
  - Metrics tend to be conservative
  - Can be considered an “ideal” inversion situation
- CS metrics, RIP and incoherence, provide a measure of deviation from ideal. In particular
  - How coherent is the sampling strategy (are the measurements only locally informative?)
  - How many samples needed to reconstruct fields with just a sparsity prior?
  - Can the inversion go unstable because of measurement noise?

Questions?


Reconstructed emissions in 8-day period 3 [microMoles/m^2/sec]
5
10
15
20
25
10
20
30
40
50
60
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05

Which parts of US are well estimated?





- The NE has the lowest errors and best correlations
- The NW is generally the worst estimated

Sandia National Laboratories