SAND2012-2561 C
SAND2012-2561C

Estimation of multiscale fields representing anthropogenic CO_2 emissions from sparse observations

J. Ray[1], B. van BloemenWaanders[2] and S. A. McKenna[2]
jairay [at] sandia [dot] gov

[1]Sandia National Laboratories, Livermore, CA
[2]Sandia National Laboratories, Albuquerque, NM

Acknowledgements: The work was funded by the LDRD program in Sandia National Labs

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.

Problem statement

- **Aim:** Develop a technique to estimate anthropogenic CO_2 emissions from sparse observations
- **Motivations:**
  - An alternative to estimating CO_2 emission using bottom-up (economic model) techniques
  - Can provide independent verification in case of CO_2 abatement treaties
- **How is it done?**
  - Measure CO_2 concentrations in flasks at measurement sites; also column-averaged satellite measurements
  - Use an atmospheric transport model to invert for source locations

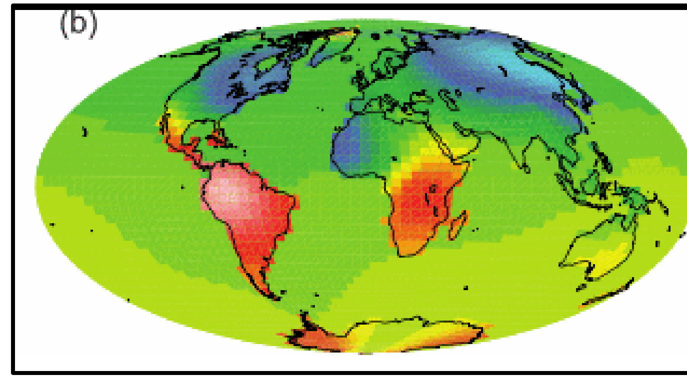
Technical challenges

- Atmospheric transport model - largest source of uncertainty
- Limited measurements - second-largest contribution to uncertainty
- Spatial models for anthropogenic CO_2
  - Non-stationary distribution in space
  - No spatial models exist to date – but need one is emissions are to be estimated from sparse observations
  - Impact of choice of spatial model on emission estimates?
- Discriminating between anthropogenic and biogenic CO_2 (biogenic is 10x larger)
  - But anthropogenic and biogenic CO_2 and different (and known) proportions of $^{12}CO_2$ and $^{14}CO_2$

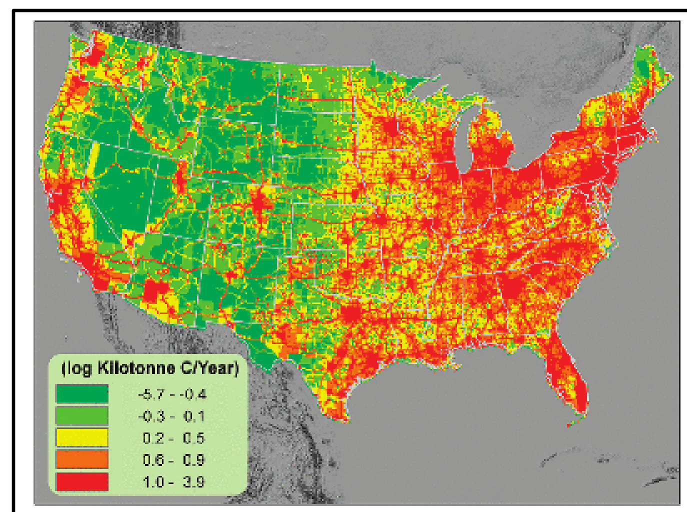
Sandia National Laboratories

Differences in spatial characteristics

- Biogenic CO_2 fluxes:
  - Smoothly variable in space
  - Modeled using multivariate Gaussian
  - Separate correlation lengths over land and oceans
- Anthropogenic (fossil fuel) emissions
  - Currently, only bottom-up estimates exist
  - A few databases – Vulcan (US-only, 2002); EDGAR (world)
  - Gaussian process will probably not work
    - What non-stationary covariance model to use?



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



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

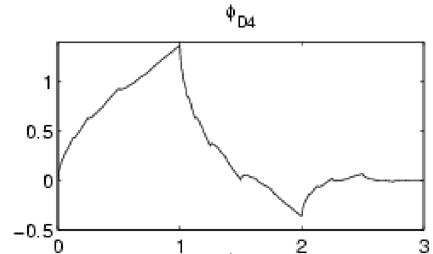
Sandia National Laboratories

Outline of the talk

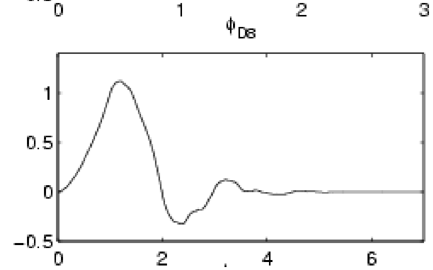
- Choosing a spatial model
  - Our hypothesis: *wavelets*
  - Study spatial and temporal characteristic of CO_2 emissions
    - Use Vulcan as source of emissions
  - Search for a good wavelet model - and what makes it good
- Demonstrate the spatial model in an OSS (observing system simulation)
  - Estimate CO_2 emissions from synthetic CO_2 observations
  - Using Ensemble Kalman Filters
    - Can handle large number of unknowns; estimates uncertainty in them
  - Using sparsity-enforcing methods used in compressive-sensing
    - If a parameter makes no difference to outputs, identifies and zeros it out

How does one represent emissions with wavelets?

- Propose $$E(x) = \sum_{s,l} w_{s,l}\,\phi_{s,l}(x)$$
  - $\phi_{s,l}(x)$ is a wavelet basis; *s, l* are its *scale* and *location* indices
  - $w_{s,l}$ are weights
- So what are wavelets?
  - Basis set with compact support
  - Belong to different families
  - Within a family, can have different orders (high order ~ smoother)
  - One chooses a family and an order, to expand E(x)
  - The expansion consists of varying
    - s, to get different frequency content
    - l, to shift in space (location)

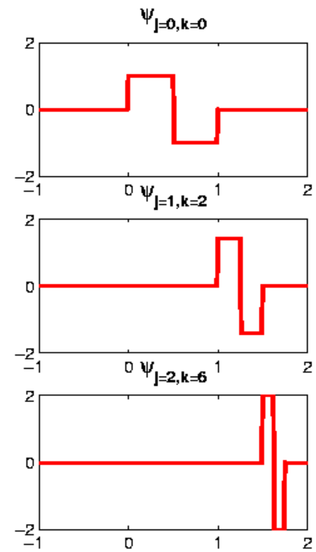


Daubechies, order 4



Daubechies, order 6

Symlet, order 6



Haars at different scales and locations

Posing the problem

- An emission field on 2^N x 2^N pixels
  - Can be decomposed on a wavelet basis, N deep
  - Each level s has 2^s x 2^s – (2^{s-1} x 2^{s-1}) weights
- Emissions

$$\boldsymbol{E}(\boldsymbol{x}) = \sum_{\boldsymbol{s}=1}^{\boldsymbol{N}} \sum_{\boldsymbol{i}=1}^{2^{\boldsymbol{s}}} \sum_{\boldsymbol{j}=1}^{2^{\boldsymbol{s}}} \boldsymbol{w}_{\boldsymbol{s},\boldsymbol{i},\boldsymbol{j}} \phi_{\boldsymbol{s},\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{x})$$

- Conjecture
  - $w_{s,i,j}$ are mostly zero (i.e., is sparse)
  - $w_{s,i,j}$ and $w_{s+1,i,j}$ are correlated – parent-child relationship
- Conjecture checked
  - Using CO_2 emissions from Vulcan (SIAM GeoSc, 2011)
  - Checked Haars, Daubechies of different orders
  - Found that Haar wavelets provided the sparsest representation
  - Reconstruction error was also small

Dimensionality reduction

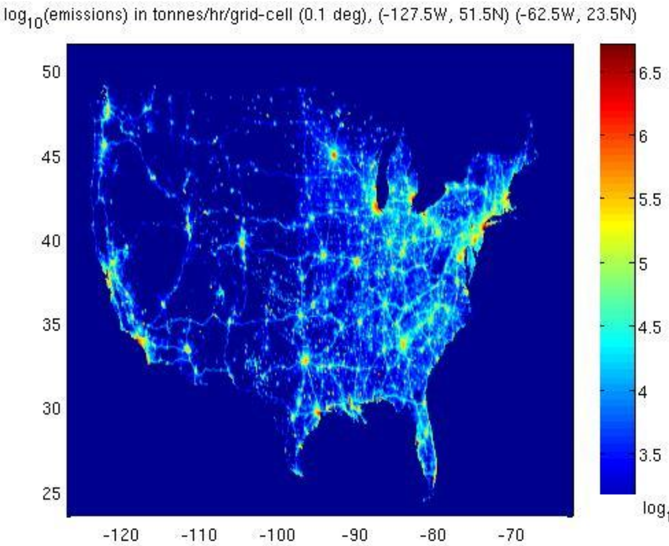
- Nightlights are a good proxy for FF emissions
  - Except emissions from electricity generation and cement production
  - Nightlights easily observed – DoD's DMSP-OLS
- Use thresholded radiance-calibrated nightlights from 1997-98 to mask out unpopulated regions

Random field model using nightlights

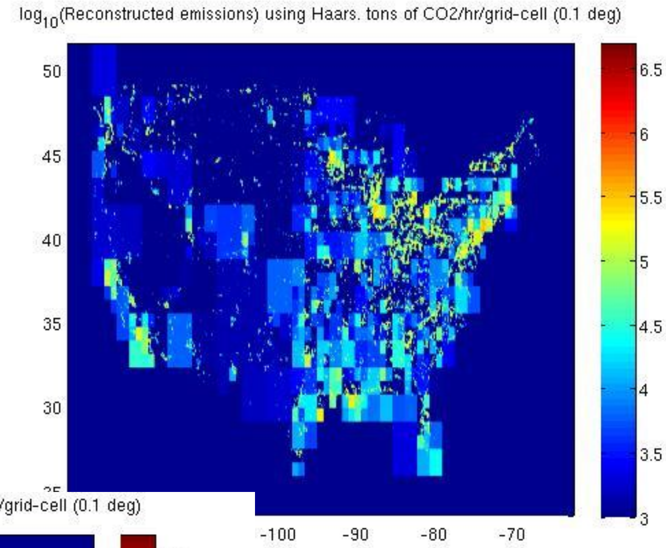
- Threshold nightlights at radiance R_{min}
  - Removes low-population regions of the US
  - Make a nightlight "mask"
- Mask EDGAR fluxes (1^{o} resolution; annual average for 2002)
  - Project to a Haar wavelet basis set & retain non-zero wavelet coefficients
- Wavelet-based Random Field model
  - With 635 coefficients ("mid-complexity")
  - Remove wavelet coefficients at finest level too – 253 parameter model ("low" complexity)
- Errors introduced by this approximation
  - We lost some emissions due to nightlight masking
  - We lost spatial fidelity due to coarsening

Sandia National Laboratories

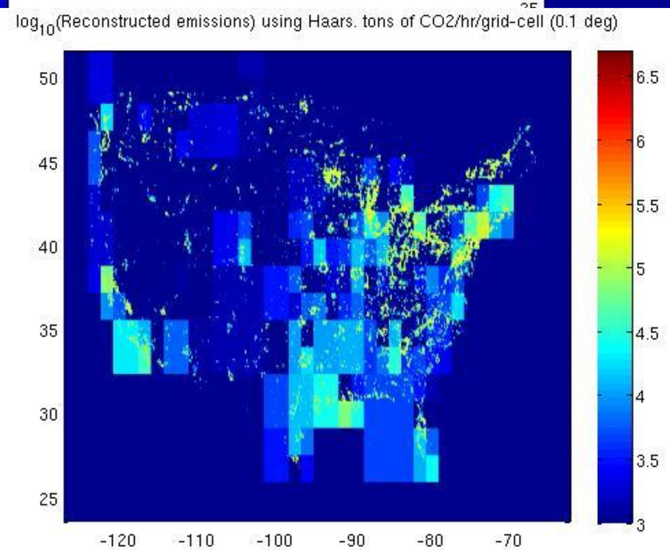
Emission reconstruction comparison



Original from Vulcan database



635 parameter model



253 parameter model

- Lost 10% of total emissions due to masking
- 253 or 635 parameters may be still too many to estimate

Sandia National Laboratories

Emission estimation problem

- Use the 2 wavelet RF models to fit to synthetic CO_2 concentration data
  - Is the dimensionality of the model high enough?
- Synthetic data generation
  - Choose location of 35 towers (NOAA's network)
  - Use Vulcan emissions, coarsened to 1^o resolution, to generate time-dependent CO_2 concentration "observations"
    - Concentration measurements at every 3 hours
  - Atmospheric transport simulated using WRF
- Inverse problem is linear
  - $y = H x$, where $y = CO_2$ concentrations; x = emissions over ~ 1 year
    - We estimate emissions averaged over 8-day periods ("Period")
  - $x = \Phi w$, where F = wavelet bases, w = basis weights
  - H constructed using WRF and 2008 wind fields

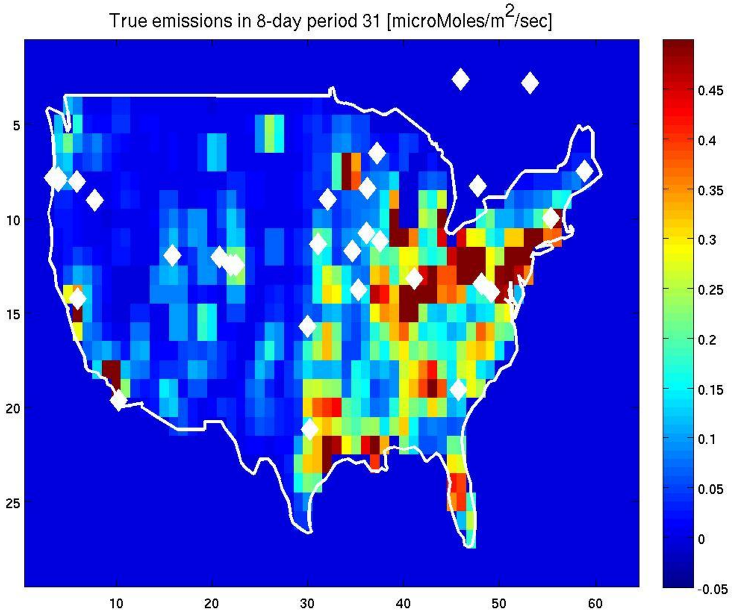
Sandia National Laboratories

Sparsity-enforced estimation

- The observations may not be sufficient to estimate 253 (or 635) parameters per "Period"
  - Atmospheric transport is diffusive – destroys information
  - If RF model parameters w cannot be estimated, set to zero
- Fitting procedure
  - Minimize $\| y - H \Phi w \|_2 + \|w\|_1$
  - Uses a greedy, orthogonal matching pursuit algorithm called StOMP (Donoho & Tsaig, 2006)
- The basic idea is borrowed from compressive sensing
  - H is the "sampling" matrix, but is neither random, nor maximally incoherent with Φ, nor does it satisfy Restricted Isometry
  - But will prevent overfitting

Sandia National Laboratories

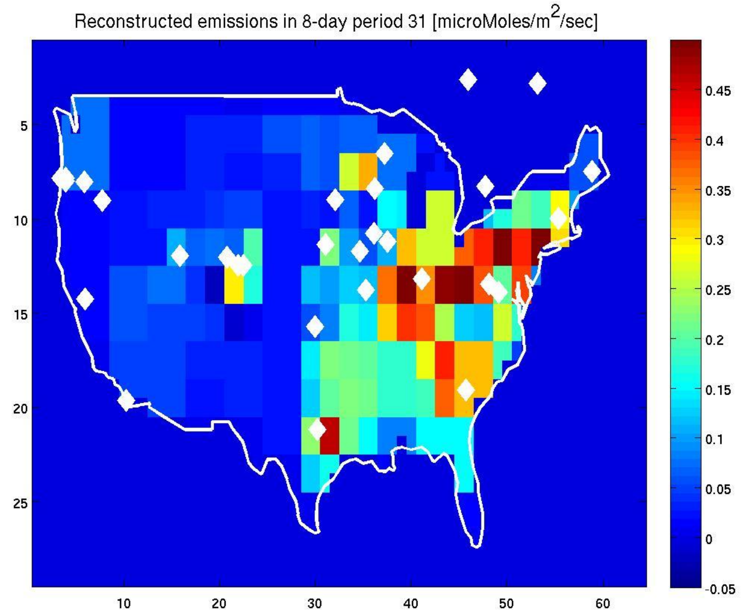
Estimated emissions

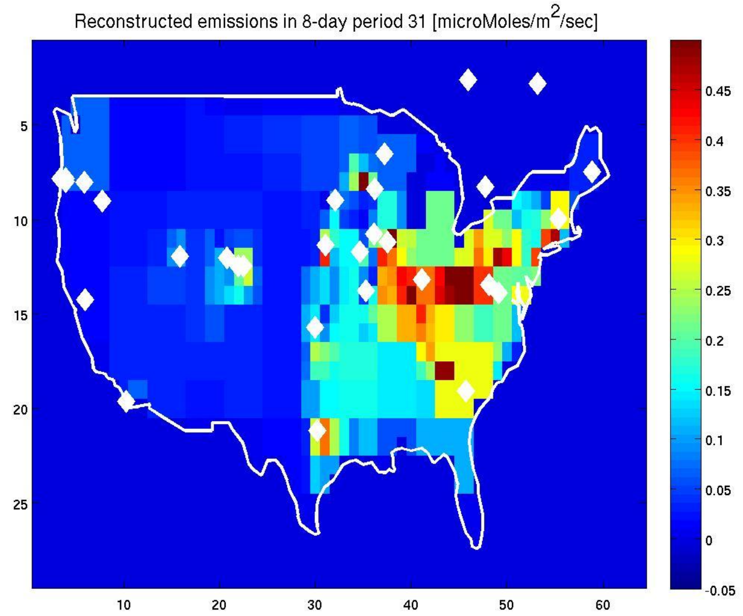


Vulcan emissions; coarsened to 1º

- Reconstructions look similar for Period # 31 (~August 2008)
- Mid-complexity model (635 parameters) has more spatial fidelity
  - Significant, or just artifact?

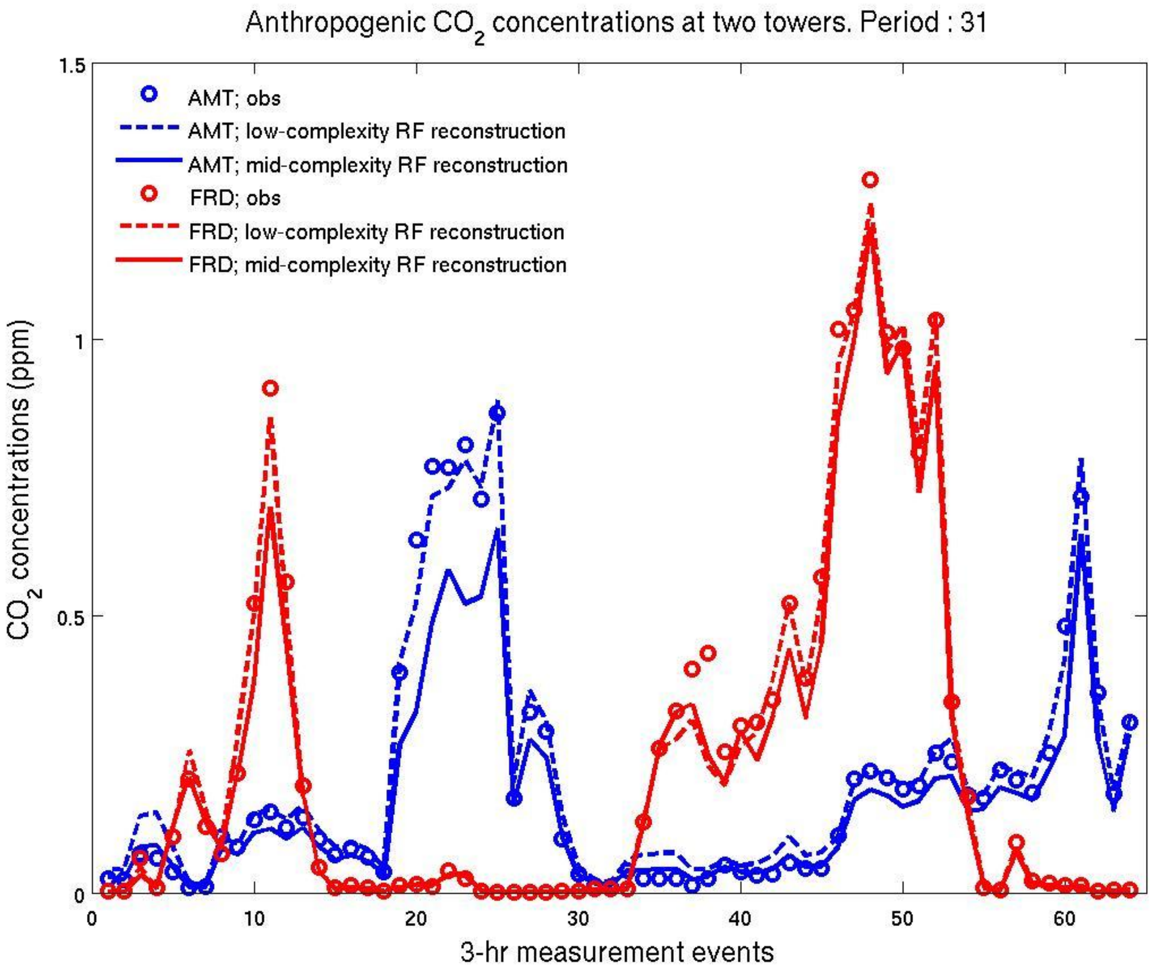
Reconstruction; mid-complexity RF model





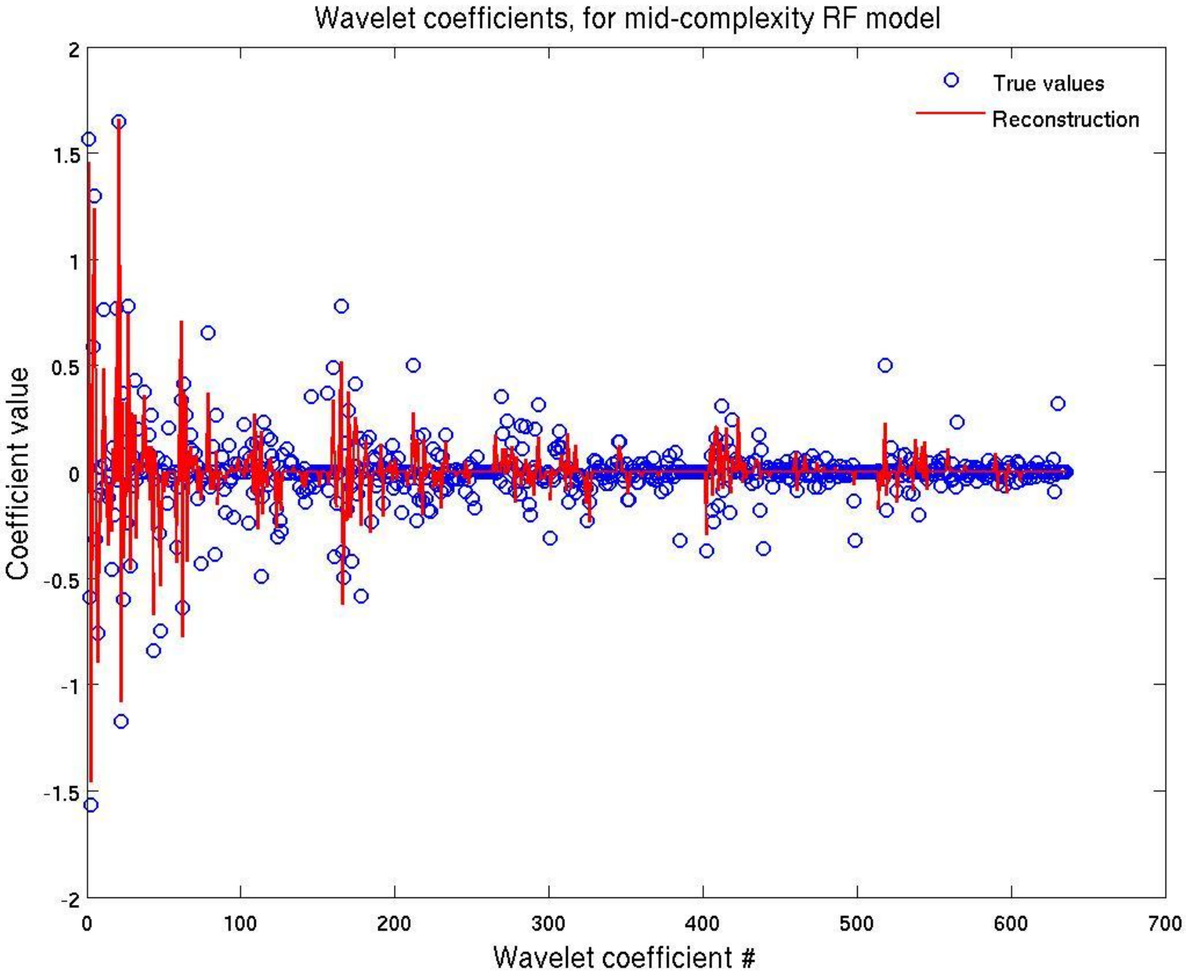
Reconstruction; low-complexity RF model

Predictive capacity



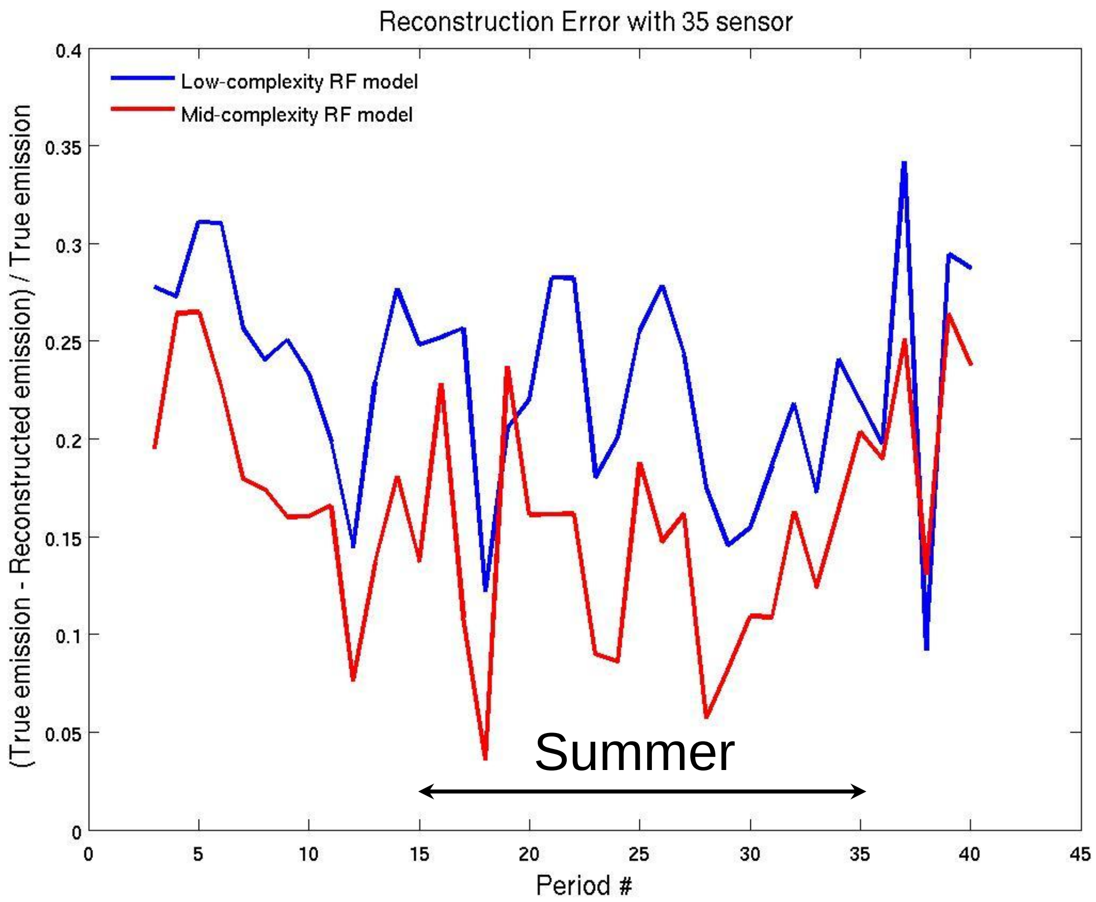
- Predicted CO_2 concentrations at 2 towers
  - Basically, not much difference between 2 RF models
- Results shown for Period 31

Did sparsification work?



- In the mid-complexity model (635 parameters), about 57% of the parameters are set to zero
- In the low-complexity, about 30%
- Lesson learnt: Our RF models are still too high-dimensional
  - But perhaps we’re not over-fitting

Accuracy of reconstruction



- The mid-complexity RF model has lower errors
  - But the errors are uncomfortably high
- Requirements
  - Need UQ of parameters
  - Need finer spatial resolution, but
    - With sparsity enforcement
- More sensors would be nice

Sandia National Laboratories

Estimating emission

- Aim: Estimate emissions, given time-variant CO_2 concentrations
  - Use a wavelet-based RF model
  - Quantify uncertainty in estimates
  - Use EnKF (scalable; also captures uncertainty in estimates)
- Basically:
  - Can wavelet-based RF models be used in estimation with UQ?
  - How large are the uncertainties if no model-reduction is done (CS or a priori)
- Data – CO_2 concentrations at sensor locations
  - Generated synthetically, using a transport model
  - Domain: Lower 48 states of USA (51.5N, -127.5W) to (23.5N, 62.5W)

Modeling and numerical details

- Transport model: Simple advection-diffusion

$$\frac{\partial c}{\partial t} - D\Delta c + v\nabla = f \quad \in \omega \times (0,T)$$

$$\nabla c = 0 \quad \in \Gamma \times (0,T)$$

- c = concentration
- v = velocity,
- f = CO2 source
- D = diffusion coefficient.

- Ensemble Kalman filters

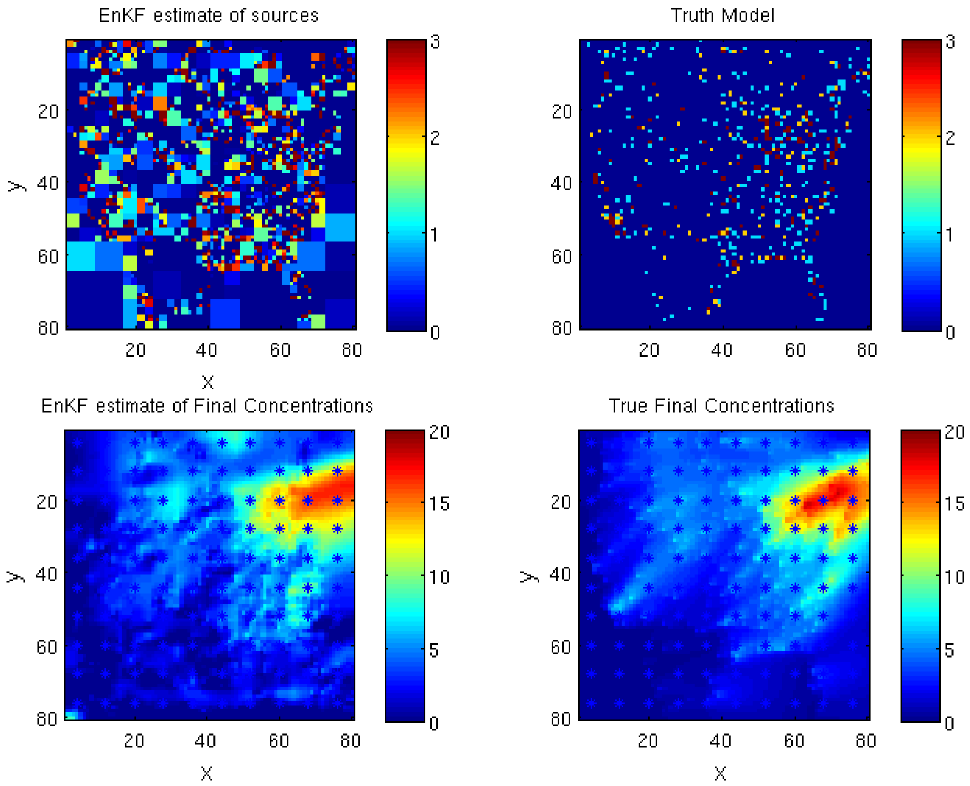
$$u_k = u_{k-1} + K(z_k - Hu_k)$$

$$K = P_k H^T (H P_k H^T + R)^{-1}$$

$$P_k = A P_{k-1} A^T + Q$$

- Spatial models
  - Used Haars and Debauchies (order = 8)

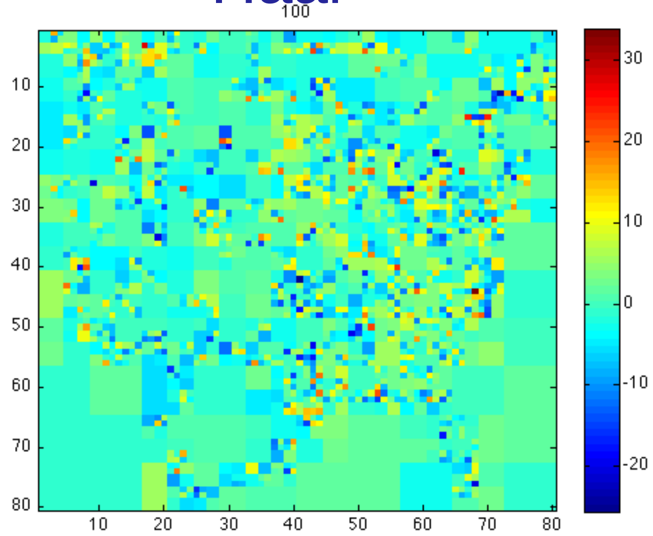
Emission estimation (MAP estimates)



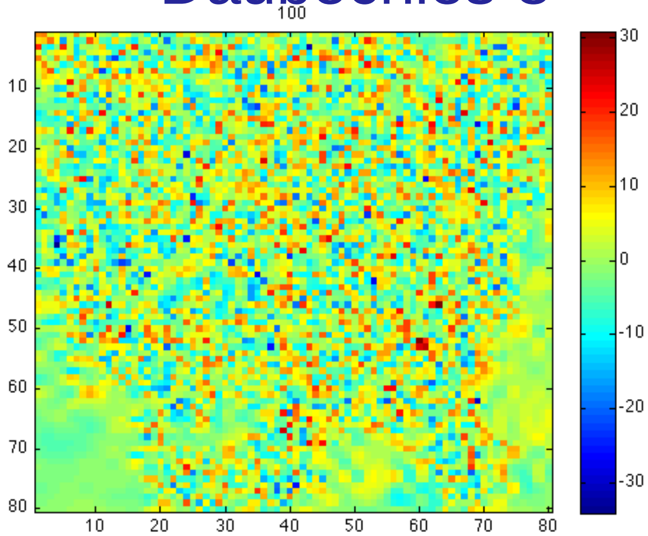
- Emissions with Haars (wavelets on all levels)
- 80x80 grid resolution; sensor grid = 10 x 10

Emission reconstruction – impact of wavelet model

Haar



Daubechies 8



RMSE = 1.44

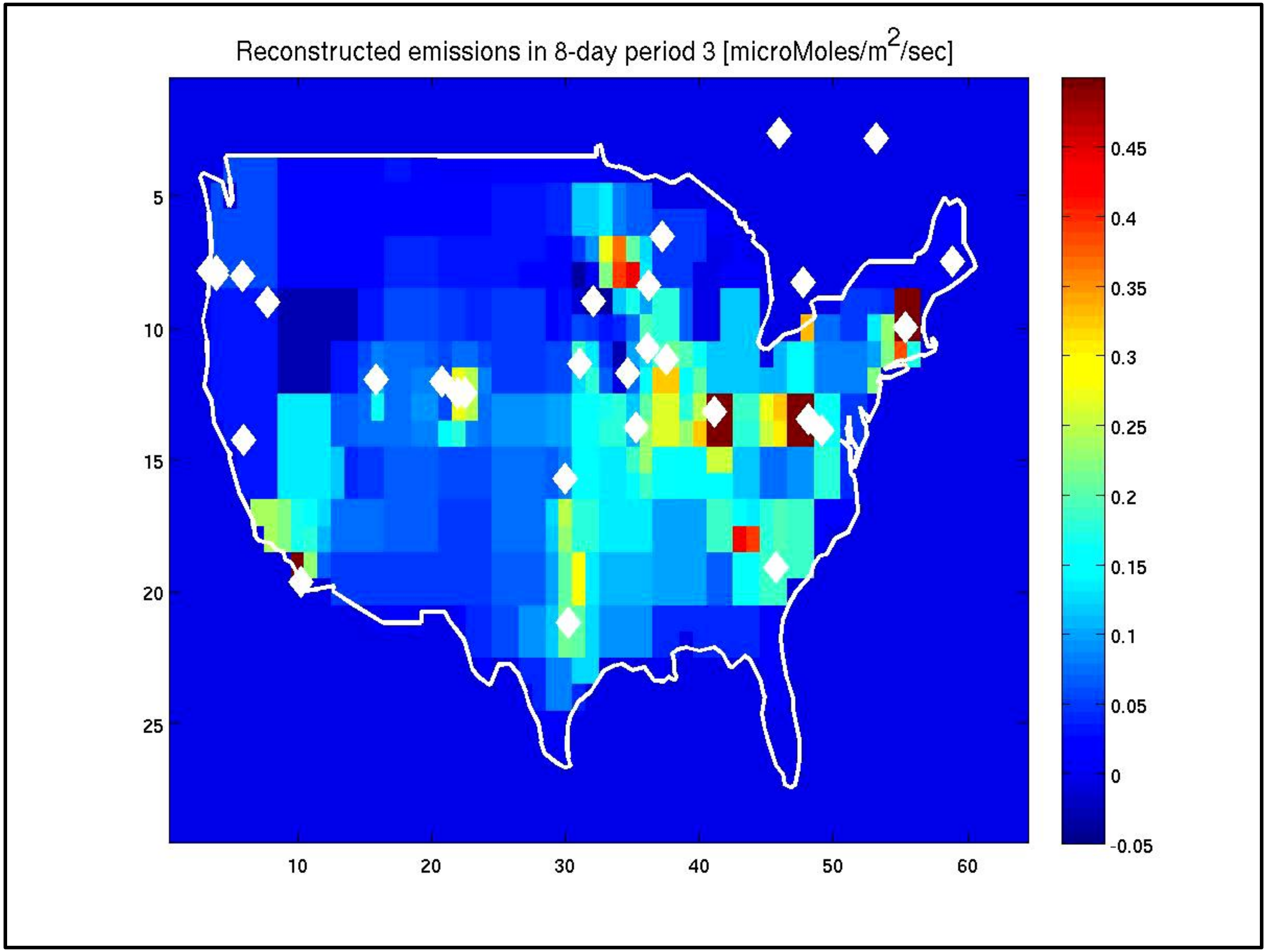
RMSE = 1.61

Conclusions

- We have created a multiresolution random field (RF) model for CO_2 emissions
- RF model can be fitted to data by enforcing sparsity
  - No uncertainty quantified, by > 50% of the coefficients were identified and inactivated
  - The 35 sensors that we have can estimate anthropogenic emissions if it were an inert tracer
    - 20-30% errors are observed
    - But the sensors were placed for biospheric, not anthropogenic fluxes
    - Unknown when joint anthropogenic and biospheric inversion can be done
- RF model also tested with EnKF, but with simplified transport
  - Both CO2 concentrations and sources can be estimated

Sandia National Laboratories

Questions?


Reconstructed emissions in 8-day period 3 [microMoles/m2/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