CO_2 Inversion using Ensemble Kalman Filters and Reduced Order Models

Bart van Bloemen Waanders, Jaideep Ray, Sean McKenna, Paul Constantine

Sandia National Laboratories

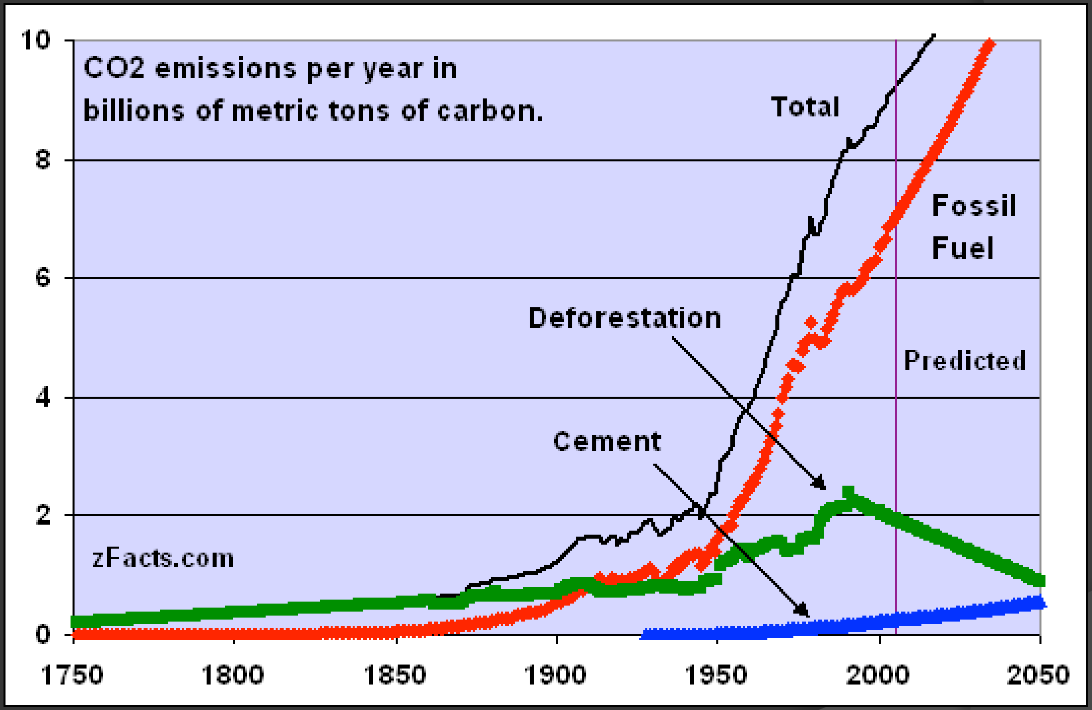
SIAM Conference on
Mathematical & Computational Issues in the Geosciences

March 21-24, 2011
Hilton Long Beach
& Executive Meeting Center
Long Beach, California USA

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

CO_2 Emissions

- CO2 responsible for global temperature increase
- Fossil fuel is the largest contributor
- Crtitical need to characterize sources globally
- Motivates a classic large scale inversion problem



Sandia National Laboratories

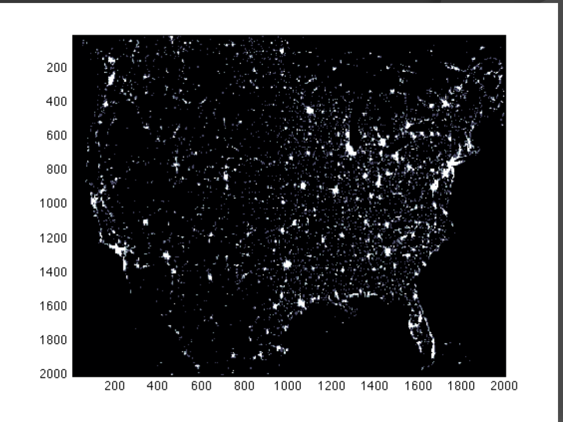
Research Challenges

- Different character associated with anthropogenic and biospheric sources
- Very large scale inversion problem
- Large scale simulation of dynamics (PCTM, GEOS, ECMWFC, etc)
- Different measurement type – point (flask), lines (plane), column (satellite)
- Model and measurement errors

Sandia National Laboratories

Our Strategy

- Ensemble Kalman Filters
- Prototype with 2D convection-diffusion
- Implement image (nightlights) based RHS
- An appropriate basis for sources (RHS)
- Reduced order modeling



Previous work on inference of Fossil Fuel (FF) emissions and CO/CO_2

- FF emissions predicted using population density, economic factors (“bottoms-up”) :
  - Doll et al, 2000: nightlight imagery for socio-econ. params
  - Rayner et al, 2010 : All variables are easily observed in a spatially resolved manner
  - Oda & Maksyutov, 2011: Nightlights give spatial distribution
- CO deterministic source inversion (“top-down”):
  - Palmer et al, 2006: Aircraft measurements
  - Petron et al, 2002: In-situ sensors
  - Wang et al, 2009; Kopacz et al, 2009; Kopacz et al, 2010: Satellites, different resolutions

Sandia National Laboratories

Outline of talk

- Ensemble Kalman Filter based inversion
  - Kalman filter $\longrightarrow$ Ensemble Kalman filter
  - Gaussian Kernel transform
  - Numerical results
- Inversion with Reduced Order Models
  - Least squares formulation
  - Karhunen and Loeve transform
  - Numerical results
- Conclusions

Ensemble Kalman Filters

- Deterministic

$$\min_{u,d} F(u.d) = \frac{1}{2}\sum_{j=1}^{N_r}\int_\Omega (u-u^*)^2\delta(x-x_j)\,dx\,dt + \frac{\beta}{2}\int_\Omega d^2\,dx$$

- Bayes Theory

$$\pi_{post} \propto exp(||d-d_{prior}||_{P^{-1}_{prior}} - ||u-u^*-e||_{P^{-1}_{noise}})$$

- Kalman Filters

$$\hat{u}_k = \hat{u}_k^- + K(z_k - H\hat{u}_k^-)$$

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

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

$$\downarrow$$

- Ensemble Kalman Filters

$$P = \overline{(u-\overline{u})(u-\overline{u})^T}$$

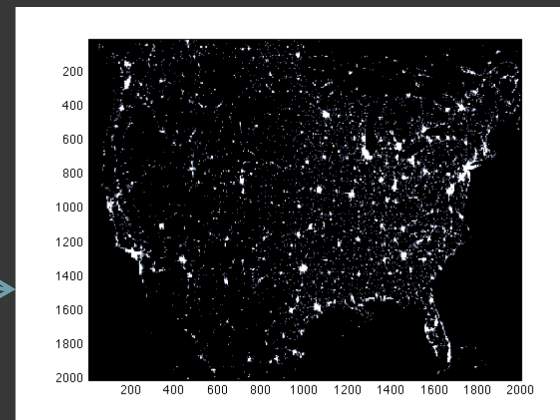
Sandia National Laboratories

Numerical Process

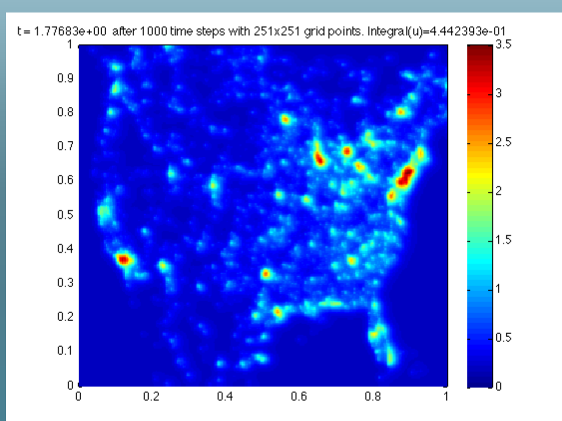
- 2D convection-diffusion with assume time varying velocity field
- Make use of satellite image of lights at night as a proxy for anthropogenic sources
- Simulate O(days) with reasonable Peclet numbers
- Continuous sources start at t=0
- Limit simulation to North America
- Parameterize source with Gaussian Kernels, Karhunen and Loeve

Convection-Diffusion with nightlights

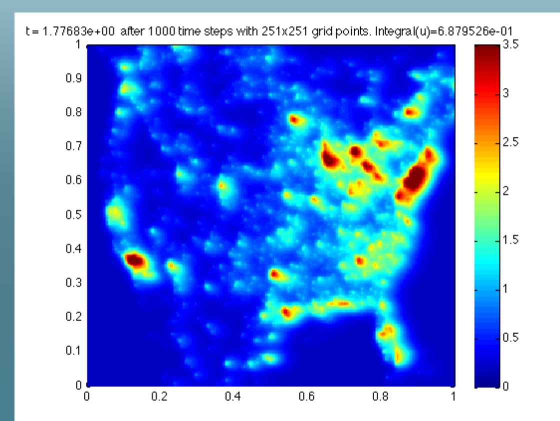
$$\frac{\partial c}{\partial t} + v\nabla c - D\Delta c = f$$



Simulation for two time periods:



Ts = 1000

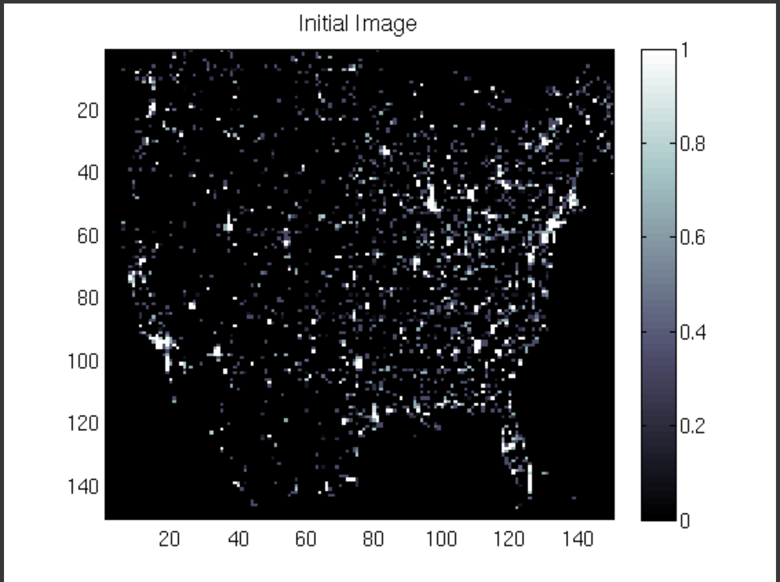


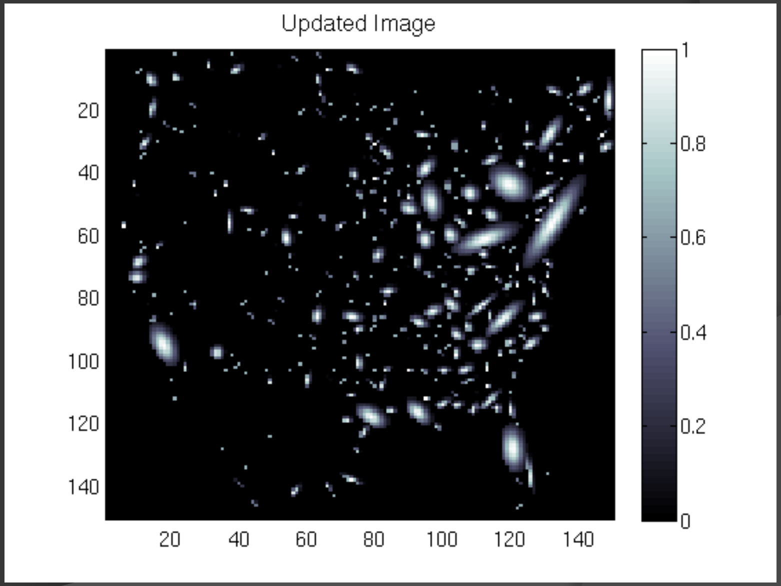
Ts = 2000

Sandia National Laboratories

Gaussian Kernel Transform

Capture pixels with a number of bilinear Gaussian kernels and set amplitudes to a constant value for an initial guess to the inversion process



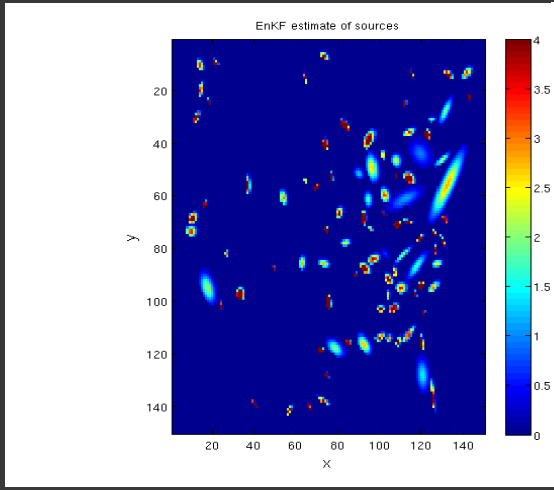


Sandia National Laboratories

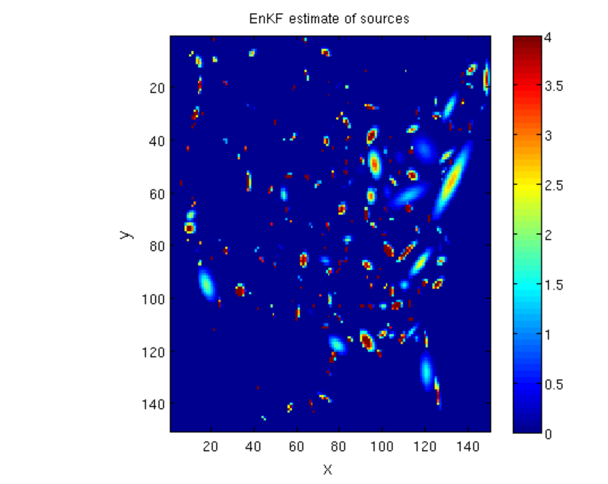
Gaussian Kernel Inversion with 150x150 grid

Inversrion Process:

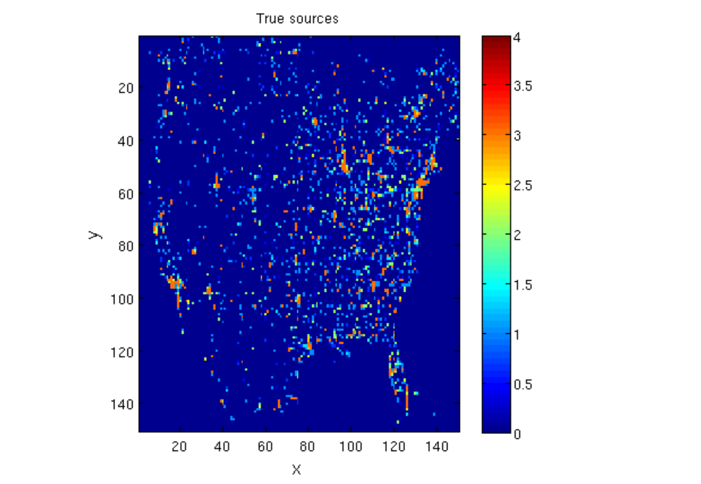
- Get concentration data at sparse locations (280) by running CD on truth model
- Set GK to a constant value of one
- At 4 time increments inject data in the EnKF routine
- Produce source and concentration predictions



100 kernels



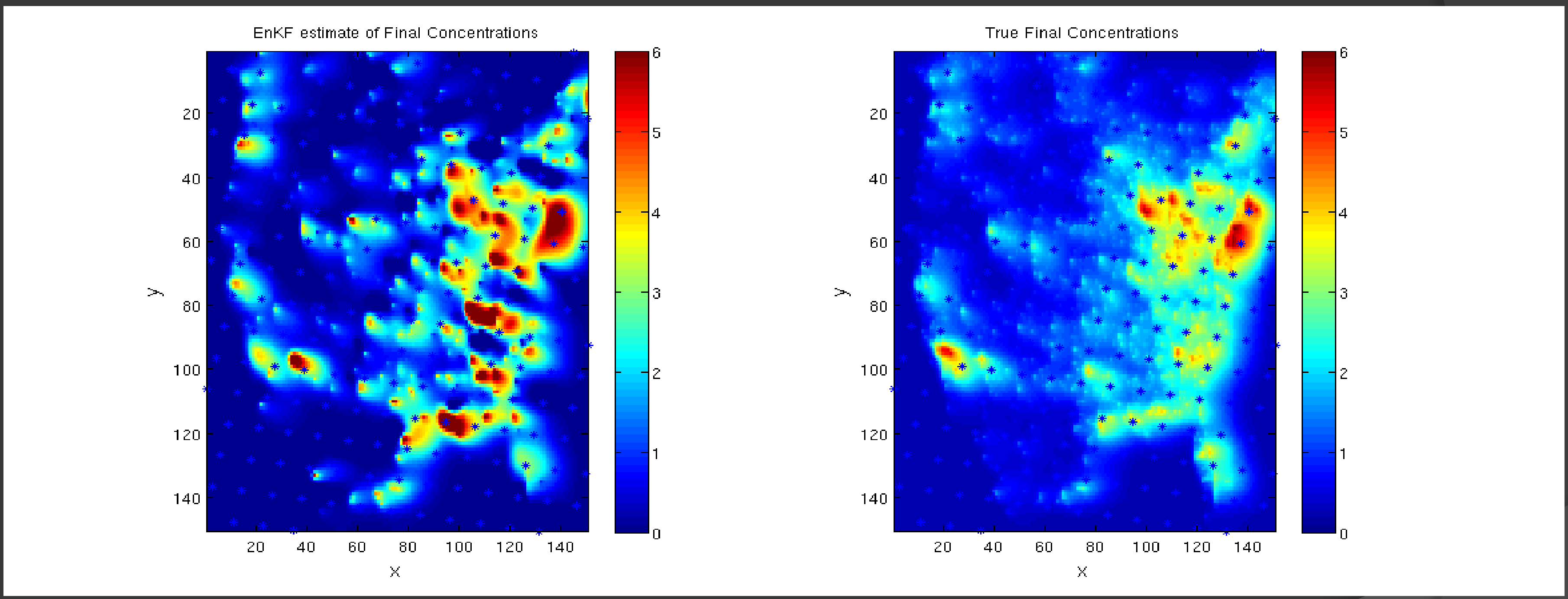
400 kernels



Truth

Sandia National Laboratories

Gaussian Kernel Inversion 150x150 grid concentration prediction



280 observations, 1E-5 noise, 100 ensembles, 400 kernels, 4000 timesteps

EnKF Inversion Summary

- Implemented EnKF
- Used 2D conv-diff with imaged-based RHS
- Parameterized image with Gaussian Kernels
- EnKF able to reconstruct sources and concentration dynamics
- In parallel, cost of inversion is equivalent to approximately one forward simulation
- Can ROM be considered to further reduce computational cost?

Outline of talk

- Ensemble Kalman Filter based inversion
  - Kalman filter $\longrightarrow$ Ensemble Kalman filter
  - Gaussian Kernel transform
  - Numerical results
- Inversion with Reduced Order Models
  - Least squares formulation
  - Karhunen and Loeve transform
  - Numerical results
- Conclusions

Sandia National Laboratories

Overview of Least Squares Approach to Reduced Order Modeling

- Assume a linear dynamical system
- Similar to Proper Othogonal Decomposition, create a snapshot matrix for variable forcings (RHS)
- Solve a least-squares minimization problem where the residual consists of affine combinations of the state vector
- In the linear case, this results in a simple transformation which allows for simple mat-vec to predict the state for a given forcing.

ROM-based CO_2 Source Inversion

The convection-diffusion equation is used to model CO2 transport.

$$\frac{\partial c}{\partial t} = \kappa \Delta c - \mathbf{u} \cdot \nabla c + f(s)$$

The forcing term is modeled with a scaled Gaussian random field.

$$f(s) = \gamma(1 + g(s))$$

Semi-discretized in space, and approximating the forcing term with a truncated Karhunen-Loeve expansion…

differential operators

forcing bases

$$x' = A(t)x + b + Ws$$

forcing parameters (independent Gaussians)

forcing mean

ROM-based CO_2 Source Inversion

Notice that *any* affine combination of solutions satisfies the convection diffusion equation for *some* forcing. The left hand side…

$$\sum_j a_j x_j' = \left(\sum_j a_j x_j\right)' \equiv \tilde{x}'$$

And the right hand side…

$$\sum_j a_j (A(t)x_j + b + W s_j) = A(t)\left(\sum_j a_j x_j\right) + b\left(\sum_j a_j\right) + W\left(\sum_j a_j s_j\right)$$
$$= A(t)\tilde{x} + b + W\tilde{s}$$

So…

$$\boxed{\tilde{x}' = A(t)\tilde{x} + b + W\tilde{s}}$$

Sandia National Laboratories

ROM-based CO_2 Source Inversion

The relationship between the forcing parameters of $\tilde{x}$ and the ROM coefficients is

$$\tilde{s} = \sum_j a_j s_j = Sa$$

We have total freedom in choosing S_j, so we choose $S_j = e_j$, the *j*th column of the identity. We can construct an invertible transformation by computing one extra basis with $S = 0$ to enforce the affine constraint.

$$\boxed{\begin{bmatrix} 0 & I \\ 1 & e^T \end{bmatrix} a = \begin{bmatrix} s \\ 1 \end{bmatrix}}$$

An invertible linear mapping between forcing parameters and ROM coefficients!!!

Sandia National Laboratories

ROM-based CO_2 Source Inversion

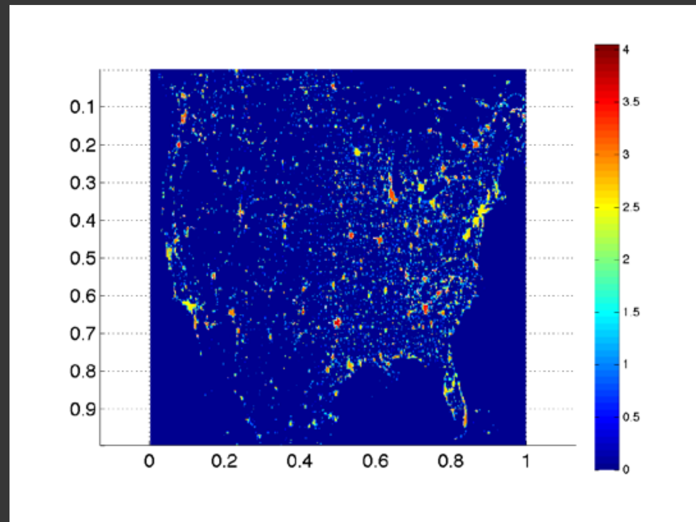
We state the inversion problem as: Given data d corresponding to CO_2 concentration at specified locations at the final time, solve

$$\underset{s}{\text{minimize}}\ \|d - Px(s)\|$$

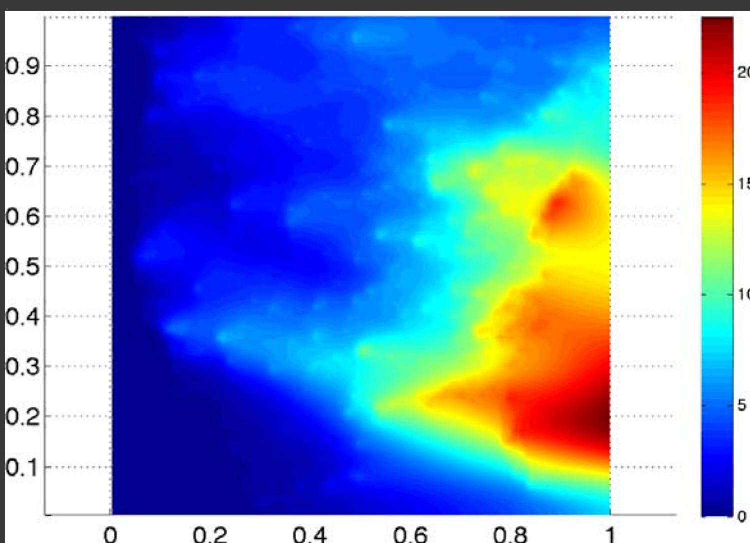
Using the invertible linear mapping,

$$\underset{a}{\text{minimize}}\ \|d - PXa\| + \text{reg}(a)$$

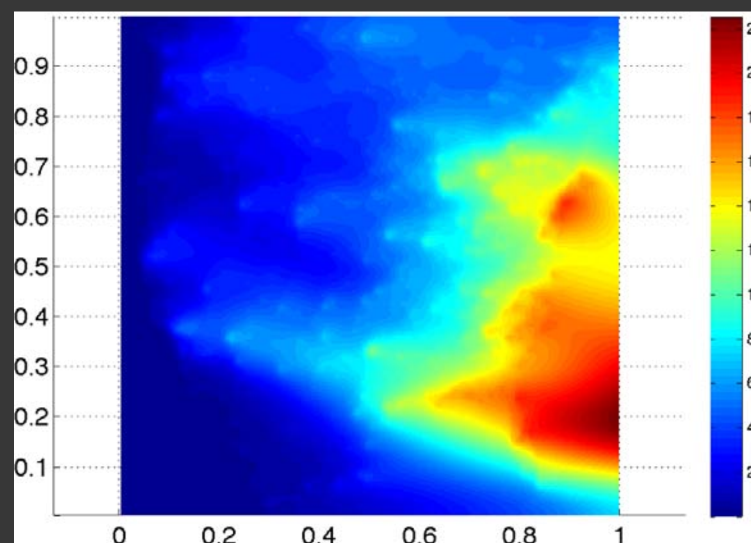
ROM-based CO_2 Source Inversion



Nightlights representation with a KL perturbation



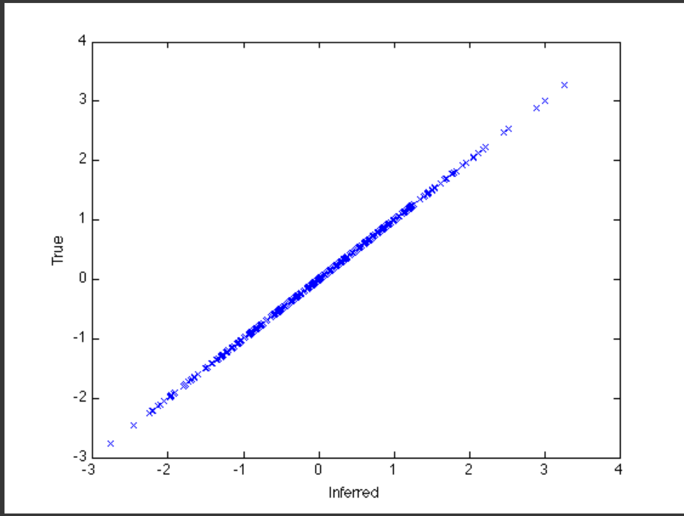
True CO2 Concentration



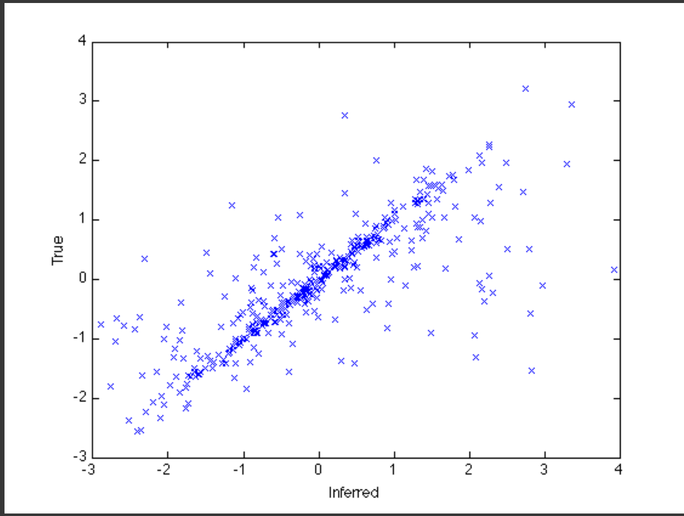
ROM CO2 Concentration

Inverted versus truth forcings Number of sensors

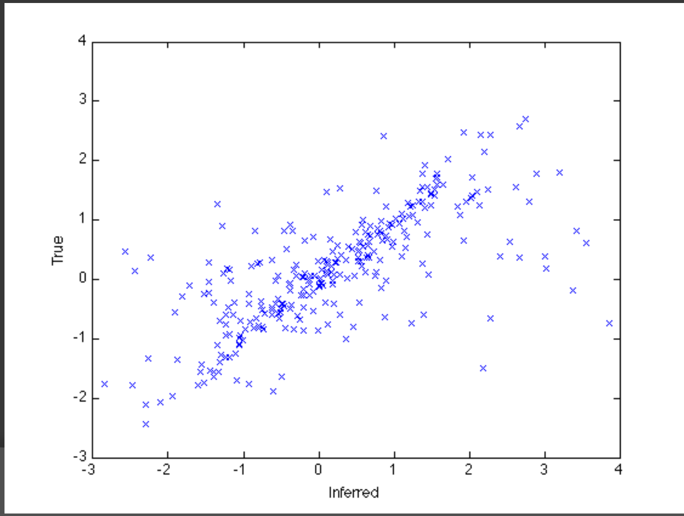
500



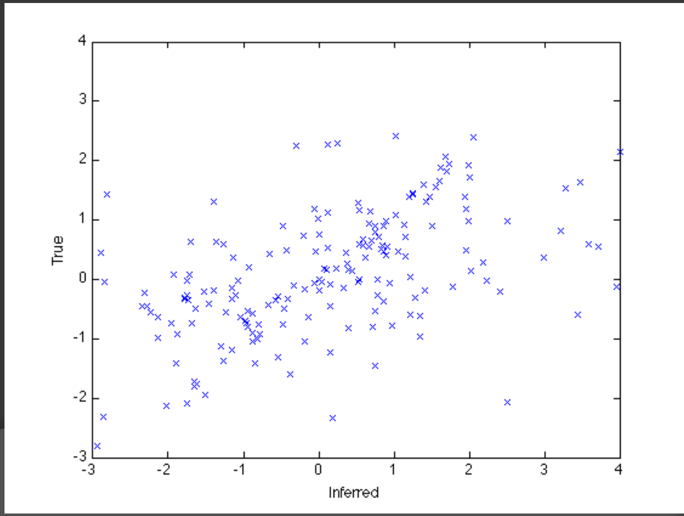
400



300



200

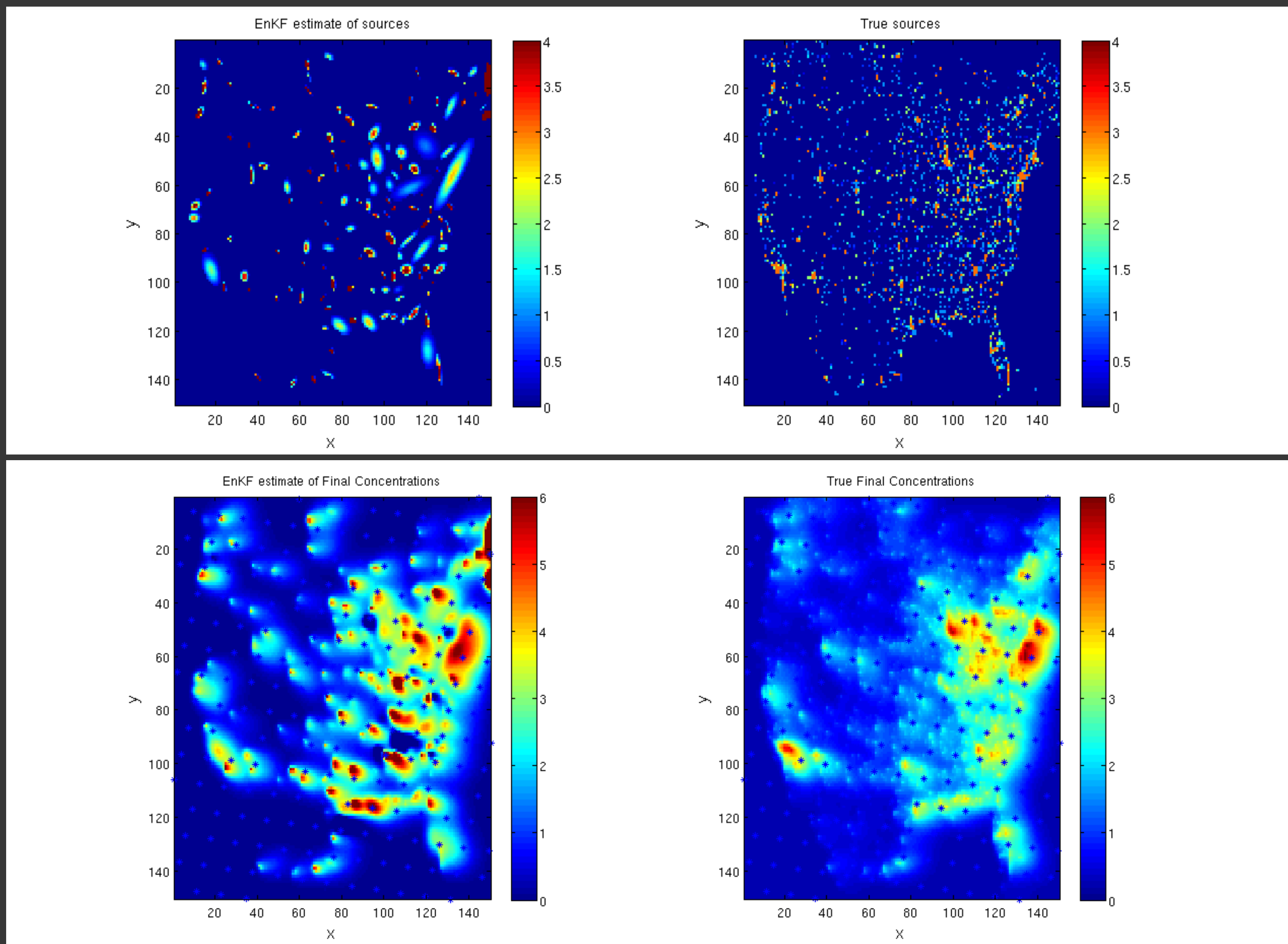


Conclusions

- Developed convection-diffusion prototype to test inversion scheme.
- Nightlight image provides reasonable proxy.
- Gaussian kernels and KL were considered as possible bases in the inversion.
- EnKF is able to invert for amplitudes of Gaussian Kernels.
- Developed an efficient ROM approach.
- Future work: consider other bases, extend ROM to 3D, extend to multiphysics

Sandia National Laboratories

Gaussian Kernel Inversion with 700 kernels on 150x150 grid



Possible Algorithmic strategies

- Deterministic – adjoint based
- MCMC algorithms
- EnKF
- Hybrid approaches
- Reduced order modeling

Sandia National Laboratories

Transformations

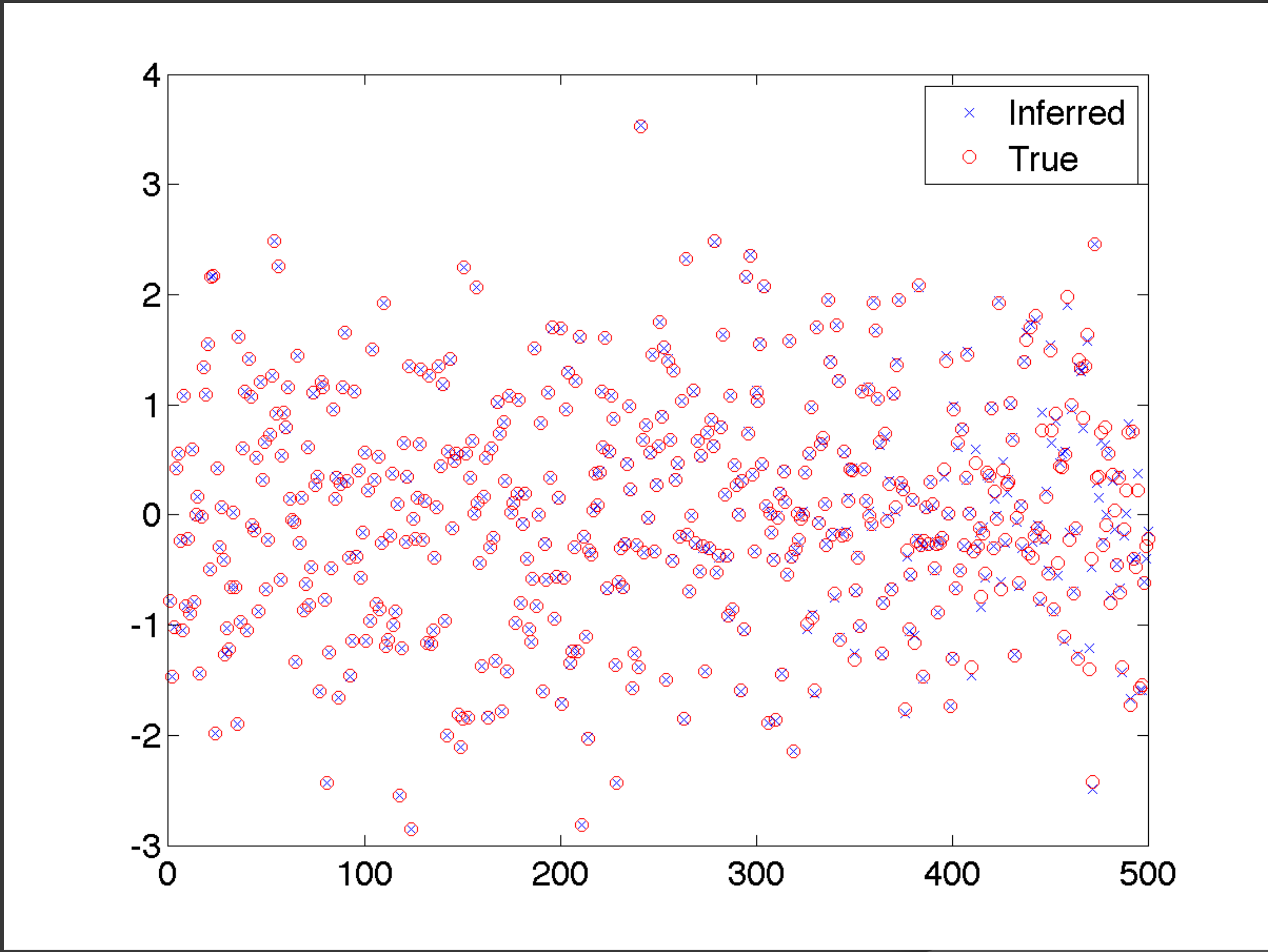
- Fourier

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-2j\pi f t} dt$$

- Karhunen and Loeve
- Wavelets $X(f) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} x(t) \cdot \xi^{*} \left(\frac{t-\tau}{s} \right) dt$
- Gaussian kernels

Sandia National Laboratories

ROM-based CO_2 Source Inversion



Inferred Forcing Parameters

Gaussian Kernel Inversion 150x150 grid Sensitivities

Sensors vs rmse