CO$_2$ Inversion using
Ensemble Kalman Filters
and Reduced Order Models

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

Sandia National Laboratories
CO₂ Emissions

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

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

- Ensemble Kalman Filter based inversion
  - Kalman filter → 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

\[
\min F(u,d) = \frac{1}{2} \sum_{j=1}^{N_r} \int_{\Omega} (u - u^*)^2 \delta(x - x_j) dxdt + \frac{\beta}{2} \int_{\Omega} d^2 dx
\]

• Bayes Theory

\[
\pi_{post} \propto \exp\left(||d - d_{prior}||_{P_{prior}}^{-1} - ||u - u^* - e||_{P_{noise}}^{-1}\right)
\]

• Kalman Filters

\[
\hat{u}_k = \hat{u}_{k-1} + K(z_k - H\hat{u}_{k-1})
\]

\[
K_k = P_k^{-1}H^T(HP_k^{-1}H^T + R)^{-1}
\]

\[
P_k = AP_{k-1}A^T + Q
\]

• Ensemble Kalman Filters

\[
P = (u - \bar{u})(u - \bar{u})^T
\]
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

\[ \frac{\partial c}{\partial t} + \nu \nabla c - D \Delta c = f \]

Simulation for two time periods:

Ts = 1000

Ts = 2000
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.
Gaussian Kernel Inversion with 150x150 grid

Inversion 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
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 → Ensemble Kalman filter
  - Gaussian Kernel transform
  - Numerical results

- Inversion with Reduced Order Models
  - Least squares formulation
  - Karhunen and Loeve transform
  - Numerical results

- Conclusions
Overview of Least Squares Approach to Reduced Order Modeling

- Assume a linear dynamical system
- Similar to Proper Orthogonal 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.
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...
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...

\[ \sum_j a_j (A(t)x_j + b + Ws_j) = A(t) \left( \sum_j a_j x_j \right) + b \left( \sum_j a_j \right) + W \left( \sum_j a_j s_j \right) \]

\[ = A(t)\tilde{x} + b + W\tilde{s} \]

So...

\[ \tilde{x}' = A(t)\tilde{x} + b + W\tilde{s} \]
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.

An invertible linear mapping between forcing parameters and ROM coefficients!!!
We state the inversion problem as: Given data $d$ corresponding to CO$_2$ concentration at specified locations at the final time, solve

$$\min_{s} \| d - Px(s) \|$$

Using the invertible linear mapping,

$$\min_{a} \| d - PXa \| + \text{reg}(a)$$
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
Gaussian Kernel Inversion with 700 kernels on 150x150 grid
Possible Algorithmic strategies

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

- Fourier
  \[ X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-2j\pi ft} \, dt \]

- 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
ROM-based CO₂ Source Inversion

Inferred Forcing Parameters
Gaussian Kernel Inversion 150x150 grid
Sensitivities

Sensors vs rmse