SAND2011-9145C

Efficient Source Inversion Methodologies using Regional Transport Models

Cosmin Safta, Khachik Sargsyan, Ray Bambha, Hope Michelsen
Bert Debusschere, Habib Najm, Bart van Bloemen Waanders
Sandia National Laboratories

AGU Fall Meeting 2011 Poster ID: GC33A-1047 Corresponding email: csafta@sandia.gov

Motivation and Challenges

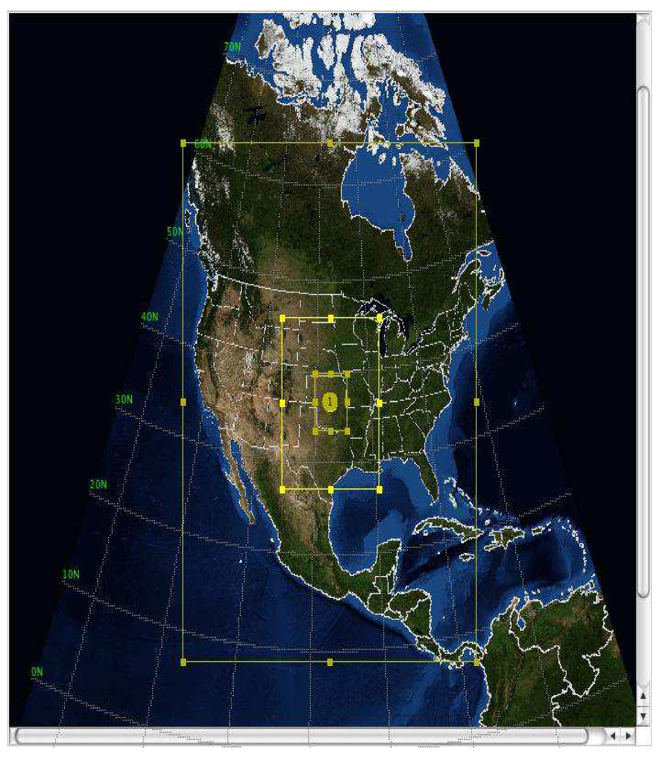
- Accurate regional estimates of Greenhouse Gas (GHG) emissions are necessary to evaluate the effectiveness of country, state, and municipal-level strategies for reducing the anthropogenic component of these emissions.
- Existing attribution methods based on CO2 concentrations alone provide little or no information on source types, although this additional information would be valuable for policy support.
- There are substantial discrepancies between reported bottom-up GHG emissions and measured top-down accumulations of these emissions in the atmosphere.
- The biogenic component dominates the GHG measurements. Isolating the anthropogenic component is challenging.

Case Study - Inference of Source Parameters

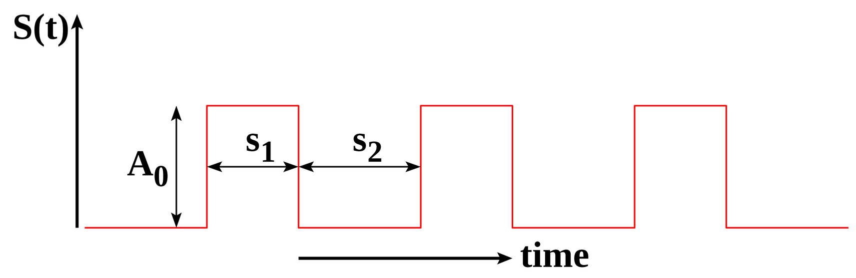
- Weather Research and Forecasting (WRF) framework used to model the atmospheric transport
  - Nested computational domain - 3 layers - using 30km, 10km, and 3.3 km grid sizes, respectively.
  - Initial and boundary conditions are based on NCEP Final Analyses (FNL) data.
- The velocity field at 10m is used to drive the 2D advection-diffusion of a passive scalar.

$$\frac{\partial c}{\partial t} + (\boldsymbol{u}\cdot c) = D\nabla^2 c + S_{\boldsymbol{x}}(\boldsymbol{x})S_t(t)$$

  - The spatial profile component $S_{\boldsymbol{x}}(\boldsymbol{x})$ of the source is designed to have the scalar emitted from one particular cell in the finest grid of the computational domain
  - The time profile $S_t(t)$ consists of a sequence of periodic puffs (see figure). The amplitude A_0 of the puffs is assumed to be known, while the duration, s_1, and the interval between them, s_2, are not.
  - The scalar transport is simulated only in the finest computational grid, using convective boundary conditions.



- Based on the measurements recorded at several locations away from the source we will attempt to infer the source parameters s_1 and s_2.



- *WRF-simulated winds are used in the 2D scalar dispersion to demonstrate inference in the presence of natural variations in wind velocity. The duration of the "source puffs" were chosen as the subject of the inference to ensure a non-linear system response.*

Bayesian Inference of the Source Parameters

Bayes formula

$$p(\boldsymbol{s}|\mathcal{D}) \propto L_{\mathcal{D}}(\boldsymbol{s})p(\boldsymbol{s})$$

relates the prior distribution $p(\boldsymbol{s})$ of source parameters $\boldsymbol{s}$ to the posterior $p(\boldsymbol{s}|\mathcal{D})$, where the data $\mathcal{D}$ is the set of measurements at various sites around the source.
The likelihood accounts for the discrepancy between the data $\mathcal{D}$ and the model $f(\boldsymbol{s})$.

$$L_{\mathcal{D}}(\boldsymbol{s}) \propto \exp\left(-\sum_{i=1}^{N}\frac{(f(\boldsymbol{s})-\boldsymbol{y}_i)^2}{2\sigma^2}\right)$$

- N is the number of measurement sites,
- $f(\boldsymbol{s})$ are pollutant concentration at the measurement site i computed using a transport model of choice (e.g. WRF+scalar transport)
- $\boldsymbol{y}_i$ are the experimental values.

The standard deviation σ includes both the instrument error as well as any model discrepancy error (initial and boundary conditions, sub-grid models, numerical approximations) introduced by f.

Posterior Distribution Sampled via Markov Chain Monte Carlo

Given the likelihood $L_{\mathcal{D}}(\boldsymbol{s})$ and the prior $p(\boldsymbol{s})$, we then draw samples from the posterior distribution $p(\boldsymbol{s}|\mathcal{D})$ via Markov Chain Monte Carlo (MCMC) sampling. MCMC is a class of techniques that allows sampling from a posterior distribution by constructing a Markov Chain that has the posterior as its stationary distribution [2].

Surrogate Model Construction: Polynomial chaos spectral representation

- The computational expense of the WRF and scalar transport simulations, typically associated with a large number of MCMC samples, will be circumvented by employing surrogate models, that are used instead of the forward model $f(\boldsymbol{s})$ in the MCMC.
- These surrogate models are based on polynomial chaos (PC) expansions [1, 3] and are used to represent quantities of interest, e.g., scalar concentration at specific locations, as functions of source and model parameterizations.

Interpret input parameters $\boldsymbol{s}$ as random variables, which can be represented via their cumulative distribution function (CDF) $F(\cdot)$, such that, with $\xi_i \sim$ Uniform$[-1,1]$, we have:

$$s_i = F_{s_i}^{-1}\left(\frac{\xi_i+1}{2}\right), \qquad \text{for } i=1,2,\ldots.$$

The forward model output for the scalar dispersion given by $f(\cdot)$ can be represented as a PC expansion:

$$f(\boldsymbol{s}) = Z \approx \sum_{k=0}^{K} Z_k\Psi_k(\xi)$$

$\Psi_k(\cdot)$ are standard Legendre polynomials of independent, random variables ξ, orthogonal w.r.t. uniform pdf $p_\xi(\xi)$, i.e.

$$\langle\Psi_i(\xi)\Psi_j(\xi)\rangle \equiv \int \Psi_i(\xi)\Psi_j(\xi)p_\xi(\xi)d\xi = \delta_{ij}\langle\Psi_i(\xi)^2\rangle$$

The coefficients Z_k are computed by Galerkin (orthogonal) projection

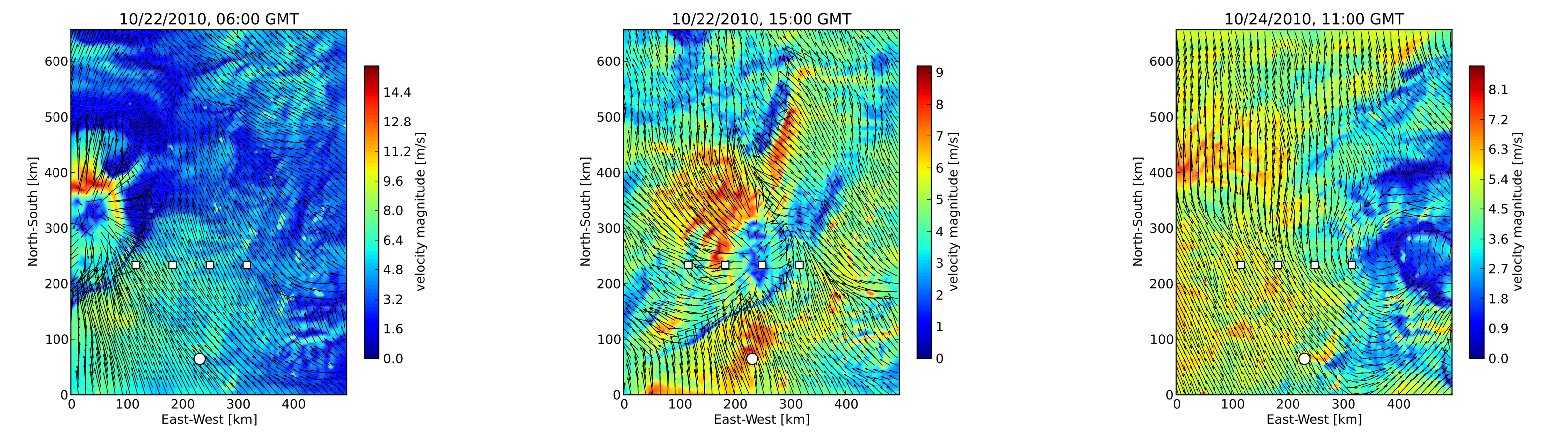
$$Z_k = \frac{\langle f(\boldsymbol{s}(\xi))\Psi_k(\xi)\rangle}{\langle\Psi_k^2(\xi)\rangle}$$

Here, the projection integrals are computed by quadrature

$$\langle f(\boldsymbol{s}(\xi))\Psi_k(\xi)\rangle = \sum_{l=1}^{N_{quad}} w_l f(\boldsymbol{s}(\xi_l))\Psi_k(\xi_l)$$

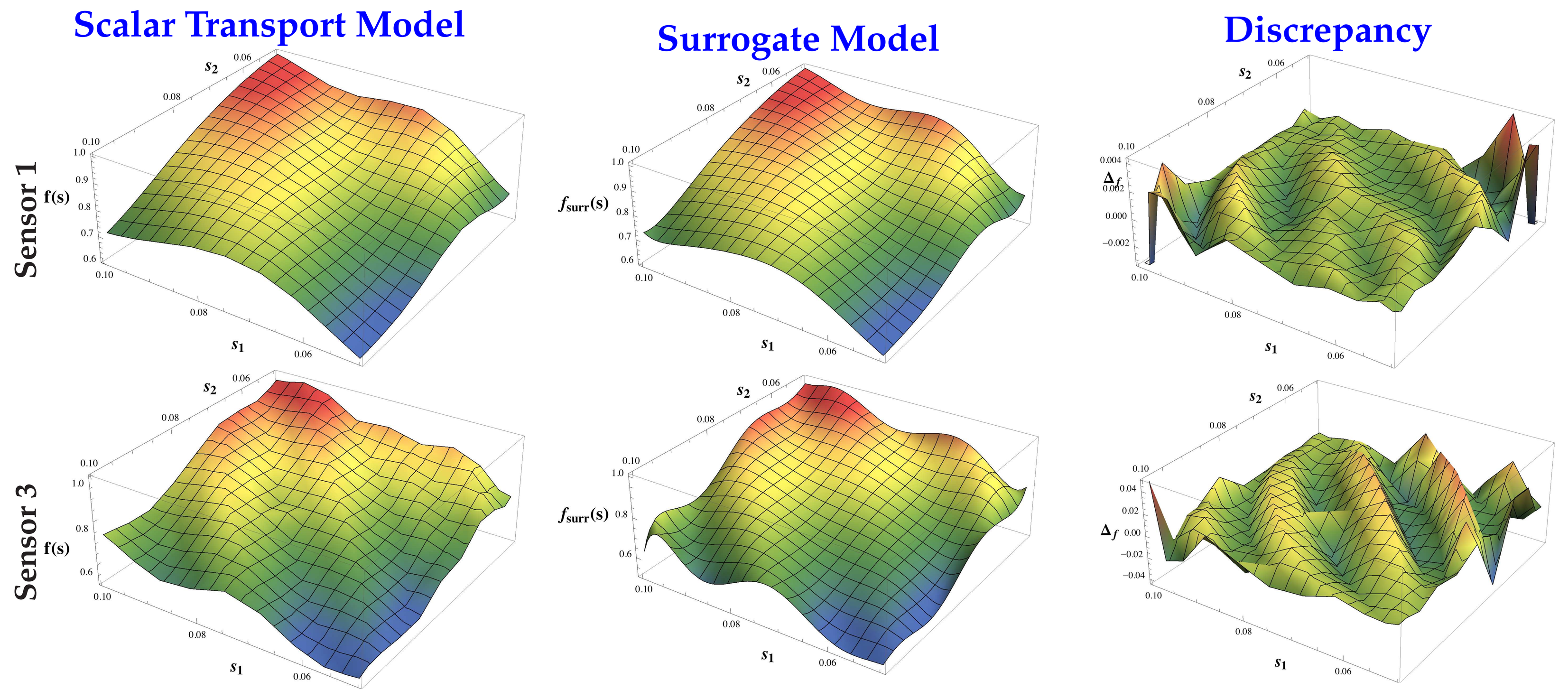
This work was supported by the Sandia National Laboratories LDRD program. Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract No. DE-AC04-94AL85000.

Atmospheric Transport using WRF



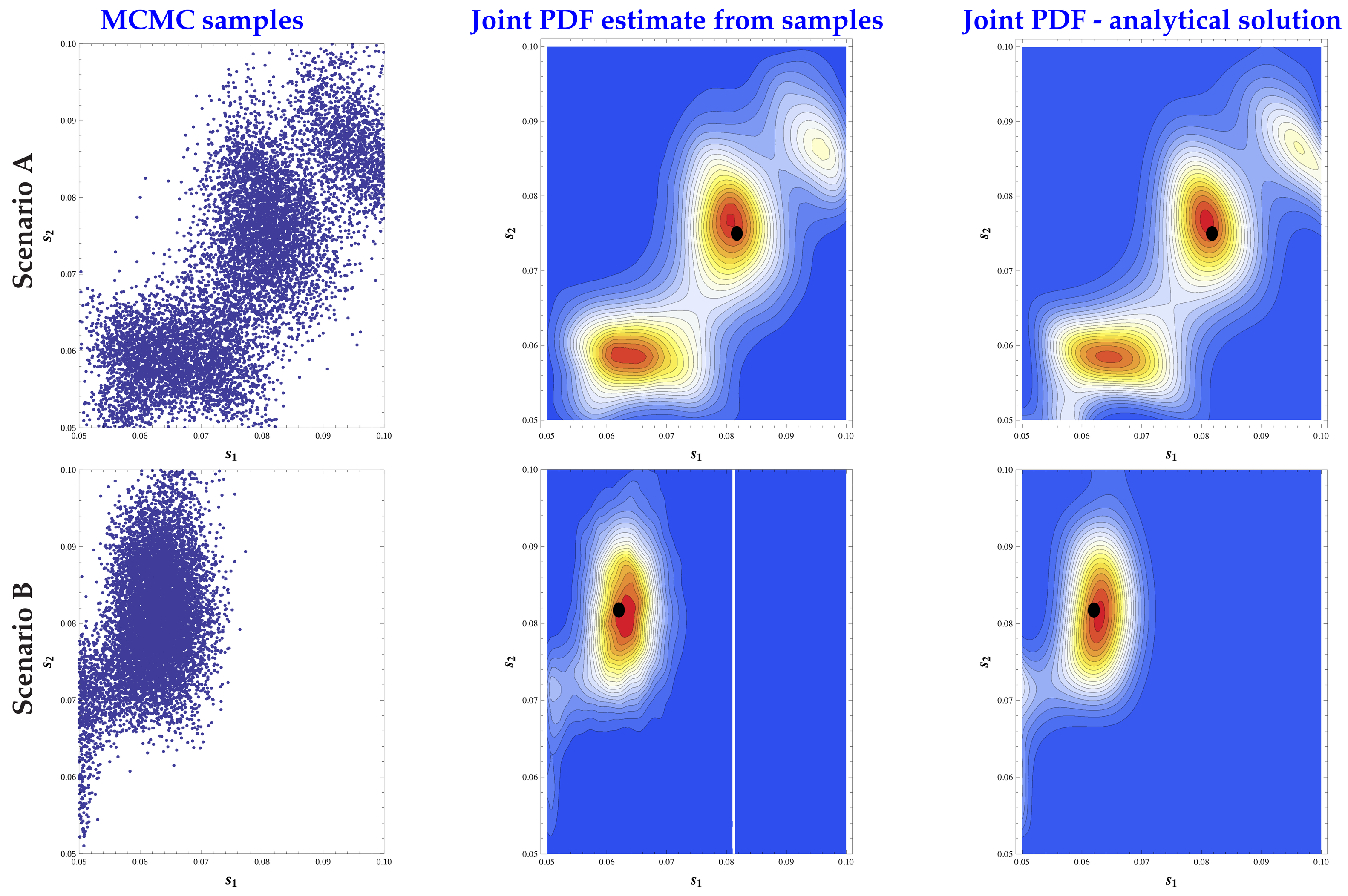
- The computational domain covers Oklahoma and Kansas - see the inner rectangle in the figure to the left.
- The source location is shown with concentric circles, and the sensors are shown with white squares.
- Typical wind patterns around an existing source were considered during sensor placement.

Surrogate Model Construction



- The cumulative scalar concentrations measure at several sensor locations over two days is represented as a function $f(\boldsymbol{s})$ of source parameters $\boldsymbol{s} = (s_1, s_2)$.
- The surrogate model values are based on 7-th order polynomials using Legendre basis functions. The polynomial coefficients were constructed using 121 model simulations.
- The surrogate models show a maximum discrepancy of about 4% compared to results from full model simulations.

Inference of Source Parameters



- Two synthetic source scenarios were considered: (a) $(s_1, s_2) = (0.082, 0.075)$ and (b) $(s_1, s_2) = (0.062, 0.082)$. Here the time values were normalized with respect to the total measurement time (2 days).
- The surrogate model approach reduced the computational cost for the MCMC samples by approximately 300 times.
- In the first scenario the inference process detected a multi-modal distribution with one of the modes centered around the "truth". For this scenario the information available from the measurements is not sufficient to pinpoint the source characteristics. In the second scenario the joint PDF is centered around the expected values.
- In both cases the joint PDF obtained through MCMC sampling agrees well with the analytical values.

Future Work

- Extend this methodology to high-dimensional parameter dependencies. Incorporate dimensionality reduction methodologies to make the inversion tractable.
- Take into account GHG's biogenic background and expand the type of sources (e.g. spatially distributed and line sources).
- Include tracer chemistry in the inversion to detect sector information.
- Incorporate experimental design methodologies to optimally place sensors given historical information on regional transport patterns.

References

[1] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach.* Springer Verlag, New York, 1991.

[2] H. Haario, E. Saksman, and J. Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7:223–242, 2001.

[3] O.P. Le Maître and O.M. Knio. *Spectral Methods for Uncertainty Quantification.* Springer, New York, NY, 2010.