SAND2017-7357C

A Bayesian Framework for the Estimation of Regional Methane Fluxes

Cosmin Safta, Ray Bambha, Hope Michelsen

csafta@sandia.gov

Sandia National Laboratories
Livermore, CA, USA

14th US National Congress on Computational Mechanics
Montréal, CA
July 18, 2017

Acknowledgments

- Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

Outline

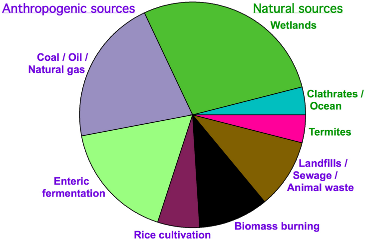
- Motivation
- Parameter Estimation Framework
  - Experimental Observations
  - Atmospheric Transport Framework (equations, discretization, assumptions)
  - Modeling the Discrepancy between Model Preditions and Experiments
- Results
- Future Work

Motivation - Methane (CH_4) Emissions

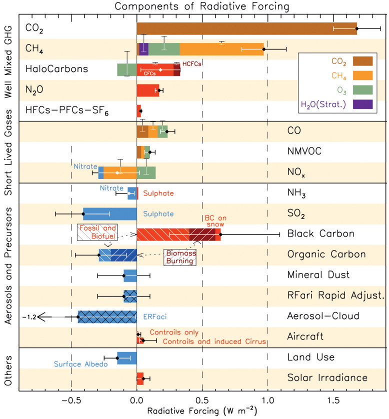
CH_4 is a significant contributor to radiative forcing

- anthropogenic sources are ∼ 70% of current total
- wetlands account for ∼ 25% of total emissions
- coal/oil/natural gas are ∼ 30% of anthropogenic
- short atmospheric lifetime: ∼ 12 yrs (>1000 years for CO_2)

current increase in global CH_4 levels are not explained

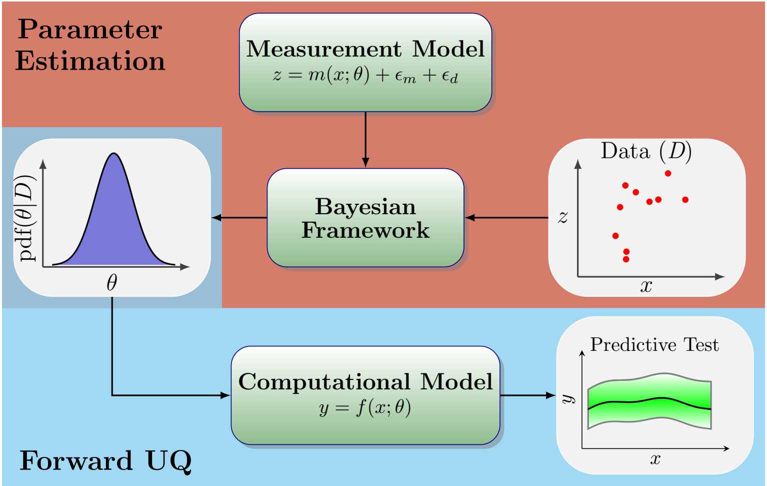


Dlugokencky *et. al.* 2011 (10.1098/rsta.2010.0341)



IPCC AR5 (2013)

Uncertainty Quantification in Computational Science



Bayes formula for Parameter Inference

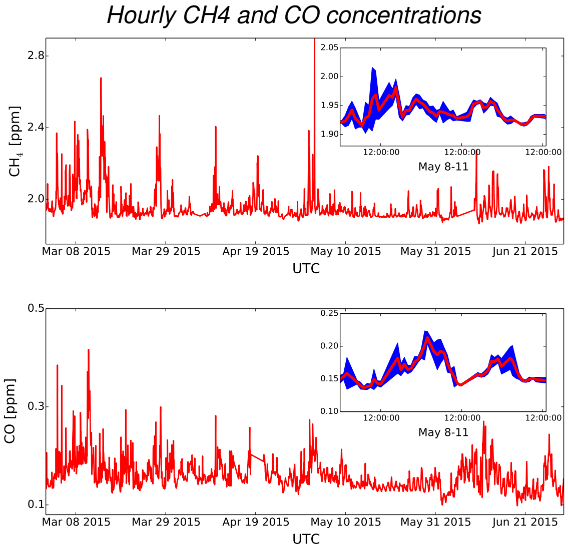
- Data Model (fit model + noise): $z = m(x;\theta) + \epsilon$
  - Collectively $D = \{z_1, z_2, \ldots\}$
- Bayes Formula:

$$p(\theta, D|M) = p(\theta|D, M)p(D|M) = p(D|\theta, M)p(\theta|M)$$

$$\underbrace{p(\theta|D,M)}_{\text{Posterior}} = \underbrace{p(D|\theta,M)}_{\text{Likelihood}}\ \underbrace{p(\theta|M)}_{\text{Prior}} \Big/ \underbrace{p(D|M)}_{\text{Evidence}}$$

- Prior: knowledge of θ prior to data
- Likelihood: forward model (M) and measurement noise
- Posterior: combines information from prior and data
- Evidence: normalizing constant, used for model comparison

Experimental Observations - Livermore, CA

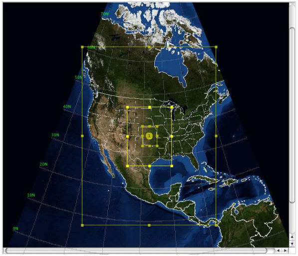


Livermore, CA: ∼150m above sea level and 27m above ground level

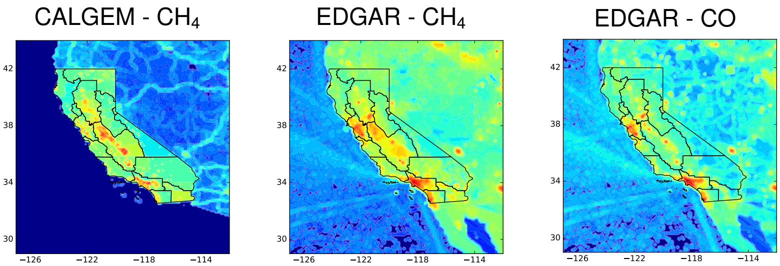
Model - Atmospheric Transport

- The **W**eather **R**esearch & **F**orecasting (WRF) Model
  - transport of mass, momentum, energy
  - vertical coordinate: hydrostatic pressure - terrain following coordinate
  - parameterizations, boundary conditions (*surface fluxes*)

- **S**tochastic **T**ime-**I**nverted **L**agrangian **T**ransport (STILT) Model
  - Lagrangian particle dispersion model
  - Derive the upstream influence region on atmospheric measurement locations



Model - Surface Fluxes



- Given footprint $f(x_r, t_r|x, t)$, the concentration y at receptor point (x_r, t_r) is written as

$$y(x_r, t_r) = \int_{t_0}^{t_r} \int_S f(x_r, t_r|x, t) E(x, t) dx\, dt + y_0(x_r, t_r)$$

- Emission databases: $E(x, t) \to E(x)$

$$y(x_r, t_r) = \int_S \left(\int_{t_0}^{t_r} f(x_r, t_r|x, t) dt \right) E(x, t) dx + y_0(x_r, t_r)$$

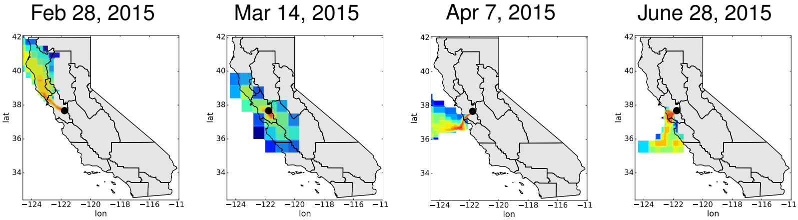
Model - Footprints

- Numerical discretization in space and time

$$\int_{t_0}^{t_r} f(x_r, t_r|x, t)dt \rightarrow \sum_i f_i(x) \rightarrow \sum_i f_{i,j}; \quad i : \text{time}, \; j : \text{space}$$

- The concentration at the measurement site

$$y = F \times E + y_0$$



Footprints representative for Livermore, CA

Statistical Discrepancy: Model - Data

$$y = F \times E + y_0 + \underbrace{\epsilon_F + \epsilon_E}_{\text{transp,surface error}} + \underbrace{\epsilon_0}_{\text{bgrd error}} + \underbrace{\epsilon_D}_{\text{exper. error}}$$

- Hard to disambiguate between transport and surface flux error
- Simultaneously infer CH_4 and CO fluxes.
  - surface CO fluxes are well understood, leading to the combined parameter estimation problem

$$\begin{aligned} y_{CH4} &= F \times E_{CH4} + y_{0,CH4} + \epsilon_F + \epsilon_{E,CH4} + \epsilon_{0,CH4} + \epsilon_{D,CH4} \\ y_{CO} &= F \times E_{CO} + y_{0,CO} + \epsilon_F + \epsilon_{0,CO} + \epsilon_{D,CO} \end{aligned}$$

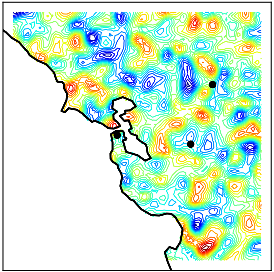
- CO observations will inform on ϵ_F, while CH_4 will inform on both ϵ_F and $\epsilon_{E,CH4}$.

Representation of Model Error - ϵ_F & ϵ_E

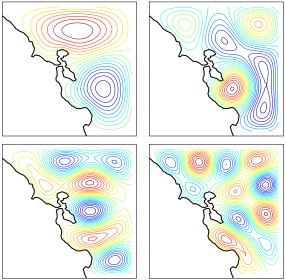
Augment the footprint F and emission fluxes E with additional terms to account for model imperfections/limitations

- Global bias terms
$$F \to F(1+\lambda_F)\ \&\ E \to E(1+\lambda_E)$$
- Region- and sector-dependent bias terms
$$F_{j,k} \to F_{j,k}(1+\lambda_{F,j,k})\ \&\ E_{j,k} \to E_{j,k}(1+\lambda_{E,j,k})$$
- High-dimensional bias terms, represented as random fields (RF) via Karhunen-Loeve Expansions

Gaussian RF over the Bay Area

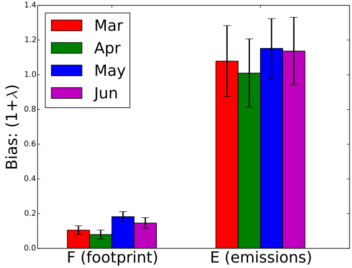


RF Modes

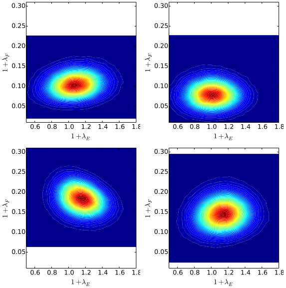


Results - Transport and Fluxes for 2015 (Mar-June)

Magnitude of bias terms for F and E



2D Marginal PDFs for 2015, March-June



- CO data induce a significant correction for the magnitude of F
  - Currently exploring ways to validate these results, possibly via additional observations of trace species
- Limited dependence between F and E - distance correlation between λ_F and λ_E are small (< 0.2)
  - Predictive studies can employ corrected F values for other species

Summary

- Assembled a framework for the assessment of biases in atmospheric transport models and emission databases
  - Multiple data streams are used to inform on different model components (transport vs boundary conditions)
  - Transport bias was significant; currently looking at other sources that can impact the results, e.g. background model.
- Moving forward to estimate corrections as random fields.
  - Determine the appropriate model parsimony using Bayesian model evidence