Thesis Defence

Simon Fraser University

Parameter Estimation and Uncertainty Quantification Applied to Advection-Diffusion Problems Arising in Atmospheric Source Inversion

Juan Gabriel García

April 10 of 2018

Content

Problem of Interest

Given a function

$$F : A \times \Theta \to \mathbb{R},$$

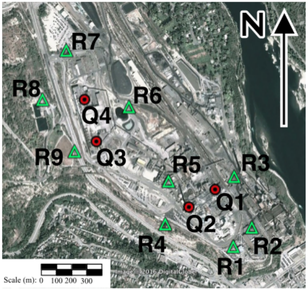
where $A \subset \mathbb{R}^n$ and Θ is a set of parameters. Given experimental (noisy) measures of F at known points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in A$. How to infer the values of the parameters in Θ and their uncertainties when F is computationally expensive?

Case Study

Consider the model of pollutant transport for the concentration c of a pollutant

$$\partial_t c + L(\theta)c = f$$

Goal: Estimate Q_i and θ using measurements of deposition in R_j.



Mathematical Model

$$\partial_t c(\mathbf{x}, t) + \nabla \cdot (\mathbf{u}(\mathbf{x}, t) c + \mathbf{S}(\mathbf{x}, t) \nabla c) = q(\mathbf{x}, t) \qquad \text{on } \mathbb{R}^2 \times \mathbb{R}_{\geq 0} \times (0, T).$$

- $\mathbf{u}(\mathbf{x}, t) = (u_x(z, t), u_y(z, t), u_{set})$ (Wind velocity field)
- $\|(u_x, u_y)\|_2 \propto \left(\frac{z}{z_r}\right)^{\gamma}$,
- $\mathbf{S} = diag(s_x, s_y, s_z)$ (Eddy diffusion matrix),
- $s_z = f(L, z_{cut})$,
- $s_x = s_y = g(z_i, L)$, with z_i Mixing layer height.

Boundary Conditions

- Far-field boundary condition

$$c(\mathbf{x}, t) \to 0 as \|x\| \to \infty$$

- Robin boundary conditions at $z = 0$

$$\left(u_{set} c + s_z \frac{\partial c}{\partial z} \right) |z = 0 = u_{dep} c|_{z=0}$$

- Concentration and deposition are related via

$$w(x, y, T) = \int_0^T c(x, y, 0, t) u_{set} dt. \tag{1}$$

$$F(x_i, y_i, T) = \int_{R_i} w(x, y, T) dxdy \approx w(x_i, y_i, T) \Delta A,$$

- ΔA is the cross-sectional area of the dust-fall jar
- T is taken to be one month.

- Numerical solution of the concentration c was obtained via a finite volume solver[1] using a 30x30 resolution grid on the domain.
- Solving for one set of parameters can take up to half an hour.

Conclusion: Finding the deposition F is computationally very expensive.

[1] Hosseini, Bamdad and Stockie, J.M. Estimating airborne particulate emissions using a finite-volume forward solver coupled with a Bayesian inversion approach.

Roadmap

1. Find a surrogate for F,
2. Locate the optimal points to evaluate F so that the surrogate is accurate,
3. See if it is possible to do dimensionality reduction,
4. Use the Bayesian framework to obtain the posterior distribution of the parameters in the light of the experimental data,
5. Perform inference on the posterior using numerical methods.

Content

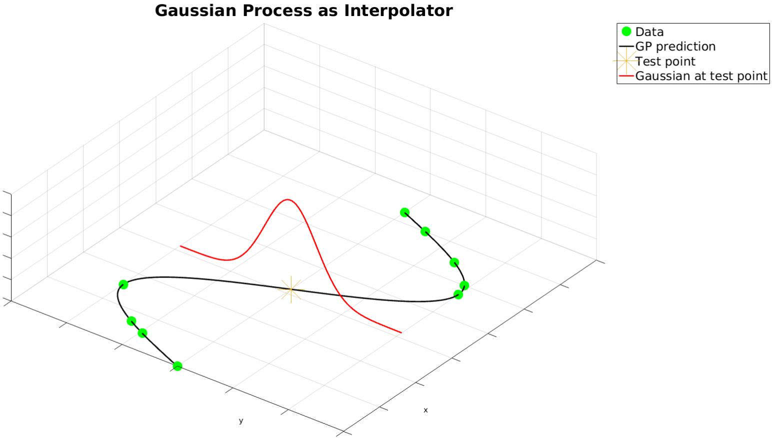
Gaussian Process

A Gaussian process (GP) is a collection of random variables $\{g(x)\}_{x\in A}$, for some set A, possibly uncountable, such that any finite subset of random variables $\{g(x_k)\}_{k=1}^N \subset \{g(x)\}_{x\in A}$ for $\{x_k\}_{k=1}^N \subset A$ are jointly Gaussian.
A GP is completely defined by its mean m(x) and covariance operator k(x,x'):

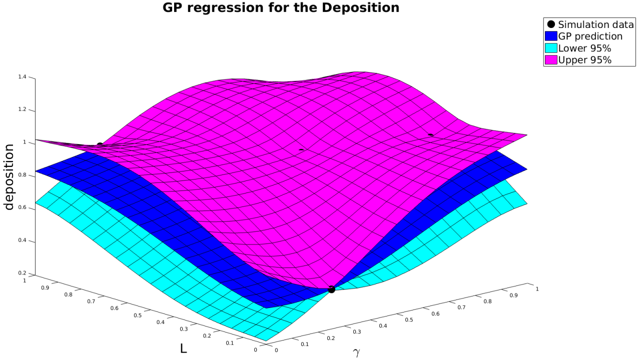
$$m(x) = \mathbb{E}(g(x)),$$

$$k(x,x') = \mathbb{E}\left((x - m(x))(x' - m(x'))\right).$$

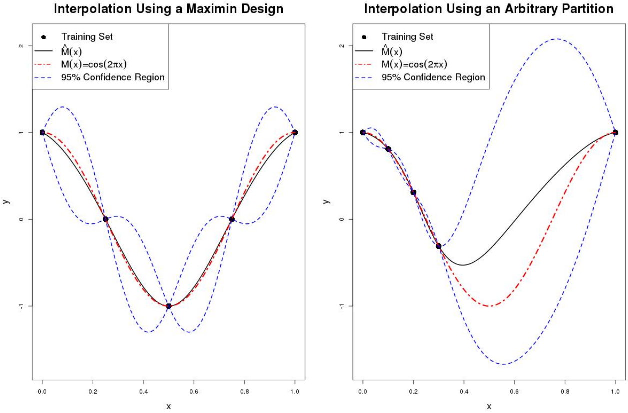
Gaussian Process as Interpolator



Interpolation with Real Data



Why Experimental Design

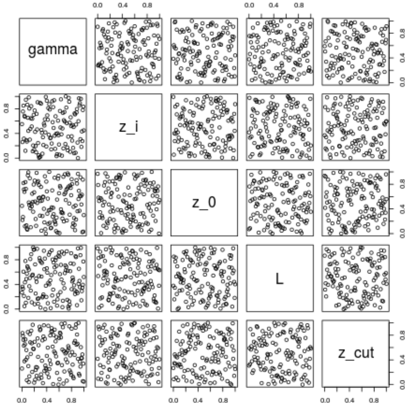


Maximin Design

Given $T \subset \mathbb{R}^n$ and a subset S of T, with finite (fixed) cardinality, say $|S| = n$. A maximin distance design S^o is a collection of points of T such that

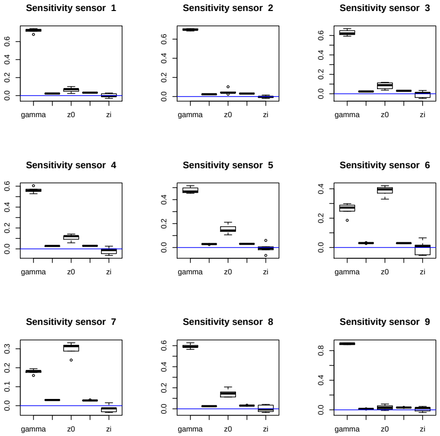
$$\max_{(S \subset T,\ |S|=n)} \min_{(s,s' \in S)} \|s - s'\| = \min_{s,s' \in S^o} \|s - s'\| = max!,$$

Example of a Design



Sensitivity Analysis

- γ: Fitting parameter for the z dependence of the velocity.
- z_0: Roughness length.
- z_i: Mixing layer height.
- L: Monin-Obukhov length.
- z_{cut}: cutoff height.



What About the Sources?

The deposition behaves linearly with respect to the values of the 4 sources, hence

$$\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_9 \end{bmatrix} = \mathcal{A}(\gamma, z_0, L) \begin{bmatrix} q_1 \\ \vdots \\ q_4 \end{bmatrix}$$

To evaluate the matrix $\mathcal{A}(\gamma, z_0, L)$ we approximate its entries by Gaussian process and create a surrogate $A(\gamma, z_0, L)$.

Content

Bayes' Rule

Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and two events $A, B \in \mathscr{F}$, with $\mathbb{P}(B) \neq 0$, we define the conditional probability of A given B by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

With the definitions above, we are now in a position to state Bayes' fomula as

$$\mathbb{P}_{post}(A|B) \propto \mathbb{P}_{like}(B|A)\mathbb{P}_{prior}(A). \tag{2}$$

Looking at the Stochastic Model

To account for the uncertainties in the model and in the experimental measurements we propose the model

$$\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_9 \end{bmatrix} = A(\gamma, z_0, L) \begin{bmatrix} q_1 \\ \vdots \\ q_4 \end{bmatrix} + \epsilon, \qquad \text{where } \epsilon \sim \mathcal{N}(0, \lambda_\epsilon I_{9\times 9})$$

Probabilistic Model

Our goal is to estimate the values of $\omega := (\gamma, z_0, L)$ and $q := (q_1, q_2, q_3, q_4)$ given the measurements $\vec{R}$. Mathematically we want to estimate:

$$\mathbb{P}_{post}(\omega, q|\vec{R}) \propto \underbrace{\mathbb{P}_{like}(\vec{R}|\omega, q)\mathbb{P}_{prior}(\omega)\mathbb{P}_{prior}(q)}_{\text{Assuming } p \text{ and } q \text{ independent}}$$

We assume $\omega \sim \mathit{Uniform}$ over the domain of definition of the parameters.

What About q?

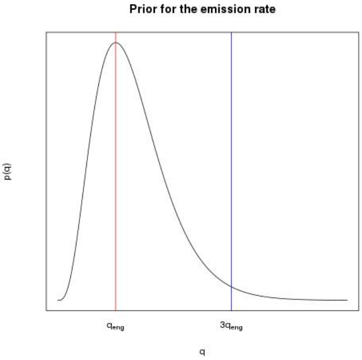
- $q_k > 0$ for $k = 1, 2, 3, 4$.
- If we trust the engineers the most likely value for q is the engineers estimate and the true value cannot be very far away from those estimates.

Source	Estimated Emission Rate [ton/yr]
q_1	35
q_2	80
q_3	5
q_4	5

Choosing a Prior for q

A consistent assumption is $q_k \sim Ga(\alpha_k, \beta_k)$,for $k = 1, 2, 3, 4$, the following conditions that define α_k and β_k for all k uniquely.

- $\beta_k(\alpha_k - 1) = q_{eng,k}$
- $qgamma(0.99, \alpha_k, \beta_k) = 3q_{eng,k}$



Likelihood

Recall that $\epsilon \sim \mathcal{N}(0, \lambda_\epsilon I_{9\times 9})$ then

- $\mathbb{P}(\epsilon) \propto \exp\left(-\frac{\|\epsilon\|_2^2}{2\lambda_\epsilon^2}\right)$
- $\mathbb{P}_{like}(\vec{R}|\omega, q) \propto \exp\left(-\frac{1}{2\lambda_\epsilon^2}\|\vec{R} - A(\omega)q\|_2^2\right)$

Calculating the Posterior

- $\mathbb{P}_{like}(\vec{R}|\omega, q) \propto \exp\left(-\frac{1}{2\lambda_\epsilon^2}\|\vec{R} - A(\omega)q\|_2^2\right)$
- $\mathbb{P}_{prior}(q_k) \propto q_k^{\alpha_k - 1}\exp(\beta_k q_k)$
- $\mathbb{P}_{prior}(\omega) \propto \mathbf{1}_{[0.1,0.6]\times[0,2]\times[-600,0]}$

Then

$$\mathbb{P}_{post}(\omega, q|\vec{R}) \propto \mathbb{P}_{like}(\vec{R}|\omega, q)\mathbb{P}_{prior}(\omega)\mathbb{P}_{prior}(q)$$

Content

Sampling From a Probability Distribution

Algorithm 1 Metropolis-Hastings Algorithm

```
pick a point $q_1$ in the support of the distribution
for j=2:N do
    Draw $u \sim U([0, \alpha])$
    $q_j \leftarrow q_{j-1} + u$
    $\beta \leftarrow \min(1, \frac{\mathbb{P}_{post}(q_j|D)}{\mathbb{P}_{post}(q_{j-1}|\mathbf{y})})$
    Draw $w \sim U([0, 1])$
    if $w < \beta$ then
        $q_{j-1} = q_j$        (Accept the move)
    else
        $q_{j-1} = q_{j-1}$        (Reject the move)
    end if
end for
```

Setting λ_ϵ

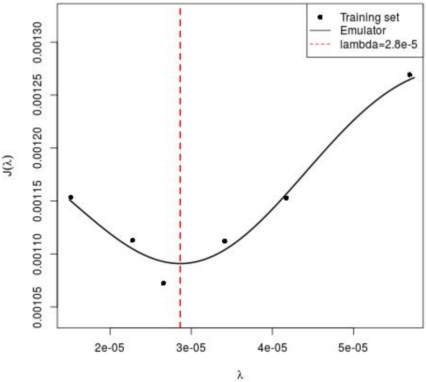
We define

$$J(\lambda_\epsilon) = \frac{1}{2}\int \left(\|A(p)q - \vec{R})\|_2 + \|q - q_{est}\|_2 \right) d\mathbb{P}_{post}^{\lambda_\epsilon},$$

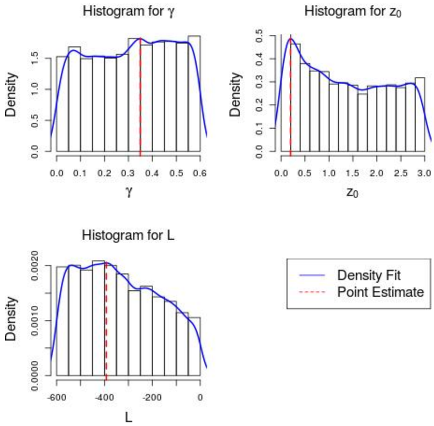
and choose a minimizer of J

$$\hat{\lambda}_\epsilon = argmin\ J(\lambda_\epsilon).$$

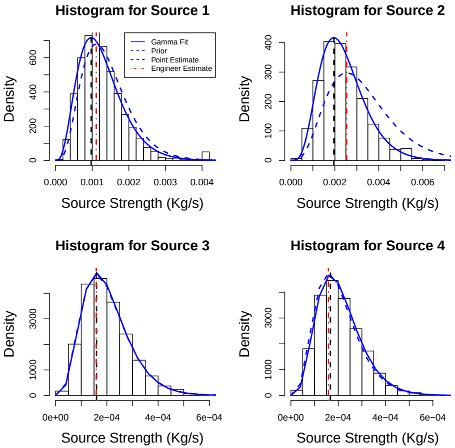
Calculating J



Histograms for the Parameters



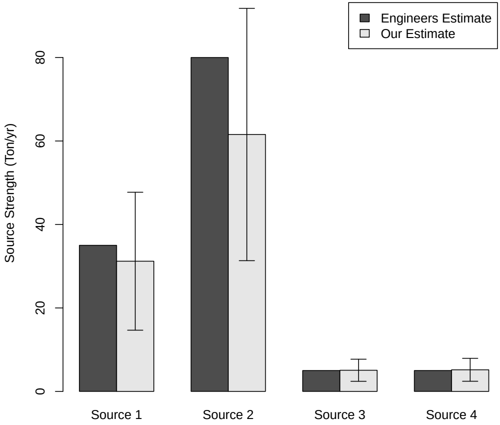
Histograms for the Sources



Results for the Parameters

Parameter	Point Estimate	68% Confidence Interval
γ	0.3478	$[0.1498, 0.5458]$
z_0	0.0811	$[0, 1.5781]$
L	-379.45	$[-195.86, -563.04]$

Results for the Sources



Comparison with Related Work

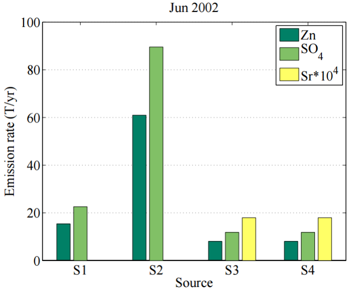


Figure: Hosseini, Bamdad and Stockie, J.M. Estimating airborne particulate emissions using a finite-volume forward solver coupled with a Bayesian inversion approach.

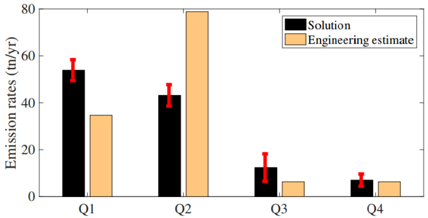


Figure: Lushi, E. and Stockie, J.M. An inverse Gaussian plume approach for estimating atmospheric pollutant emissions from multiple point sources

Conclusions

- We have developed a method to cheaply estimate parameters in computationally expensive models.
- Instead of trial and error, we propose a methodology that allows to estimate parameters in complex models using experimental data.
- Besides a point estimate we are able to obtain a confidence interval for it.
- Our results qualitatively agree with previous results.