Thesis Defence

Simon Fraser University

Parameter Estimation Using Gaussian Process Regression.

Juan Gabriel García

April 10 of 2018

Content

Problem of Interest

Given a function

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

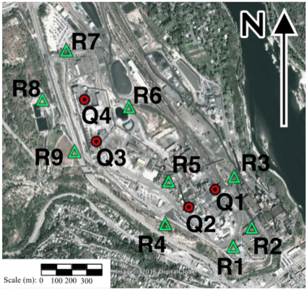
where $A \subset \mathbb{R}^n$ and Θ is a set of parameters. Given experimental (noisy) measures of u at known points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in A$. How to infer the values of the parameters in Θ and their uncertainties when u is computationally expensive?

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.



Case Study

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

- $\mathbf{u}(\mathbf{x}, t) = (u_x(z, t), u_y(z, t), u_{set})$ (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)
- $s_z = f(L, z_{cut})$, where $\frac{1}{L} = a + b \log_{10}(z_0)$.
- $s_x = s_y = g(z_i, L)$, with z_i Mixing layer height.

Case Study

In this slide expand a little on the model, like initial conditions and such

Steps

How to proceed?

1. Find a surrogate for u.
2. Locate the optimal points to evaluate u 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.

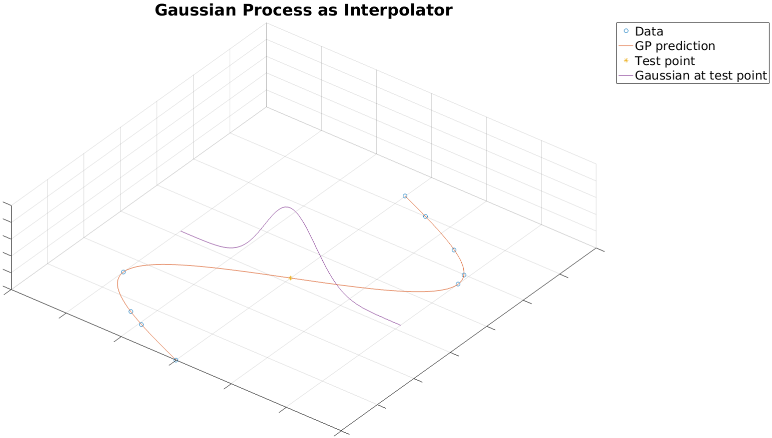
Multivariate Gaussian Distribution

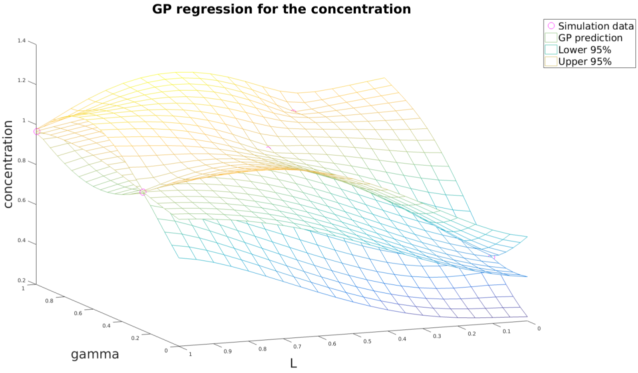
Two random vectors $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ are jointly Gaussian if their joint density function is given by

$$\frac{1}{(2\pi)^{\frac{m+n}{2}} |\Sigma|^{\frac{1}{2}}} \exp(-([x,y]-[\mu_x,\mu_y]) \underbrace{\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{bmatrix}}_{\Sigma} ([x,y]-\underbrace{[\mu_x,\mu_y]}_{\mu})^T)$$

- Marginal distribution is multivariate Gaussian.
  - $X \sim \mathcal{N}(\mu_x, \Sigma_{xx})$.
  - $Y \sim \mathcal{N}(\mu_y, \Sigma_{yy})$.
- Conditional distribution is multivariate Gaussian.
  - $X|Y = y \sim \mathcal{N}(\mu_x - \Sigma_{xy}\Sigma_{yy}^{-1}(y-\mu_y), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx})$.

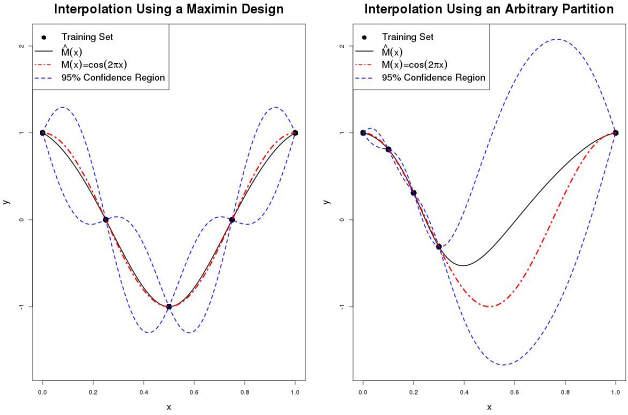
Gaussian Process as Interpolator




GP regression for the concentration
Simulation data
GP prediction
Lower 95%
Upper 95%
concentration
1.4
1.2
1
0.8
0.6
0.4
0.2
gamma
1
0.8
0.6
0.4
0.2
0
L
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0

Content

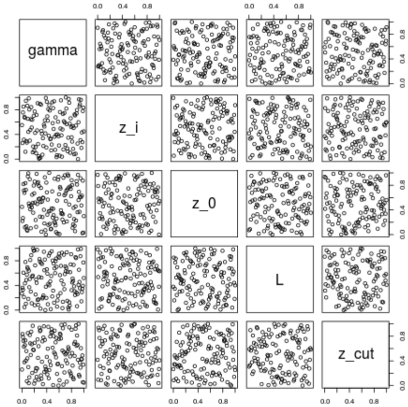
Why Experimental Design



Maximin Design

In this slide you talk about maximin design

Example of a design



Content

Sobol Indices

To assess the relevance of the parameters we do a Sensitivity analysis.

- Given a function of interest φ we decompose it as

$$\varphi(x_1,\ldots,x_n)=\varphi_0+\sum_{k=1}^{n}\varphi_k(x_k)+\sum_{1\le k<l\le n}\varphi_{kl}(x_k,x_l)+\ldots+\varphi_{1,2,\ldots,n}(x_1,\ldots,x_n).$$

- With the constraint

$$\int_{[0,1]}\varphi_{i_1,\ldots,i_j}dx_{i_k}=0 \qquad \text{if } i_k\in\{i_1,\ldots,i_j\} \tag{1}$$

- This condition allows to find each element in the decomposition recursively.

Sobol Indices

The total variance D of φ is defined as

$$D = \int_{\Omega^n} \varphi^2(x)dx - \varphi_0^2.$$

Similarly we can compute the partial variances as

$$D_{i_1,\ldots,i_s} = \int_{[0,1]^{n-1}} \varphi^2_{i_1,\ldots,i_s} dx_{i_1} \ldots dx_{i_s}.$$

With these variances we define the $s - th$ order Sobol index

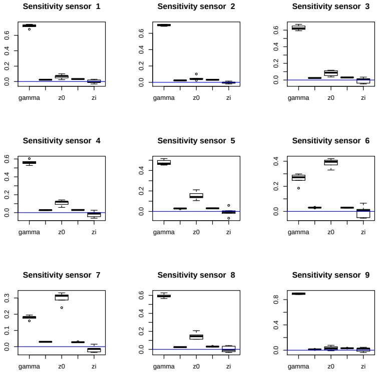
$$S_{i_1,\ldots,i_s} = \frac{D_{i_1,\ldots,i_s}}{D}.$$

To measure the relevance of the i-th parameter we calculate

$$S_i + S_{i1} + S_{i2} + \ldots + S_{i12} + S_{i13} + \ldots + S_{12\ldots,i,\ldots,n}. \text{ (Total Sobol Index)}$$

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.



In this frame it is necessary to explain why you add an ϵ for the noise and stuff

What About the Sources?

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

$$\begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_9 \end{bmatrix} = A(\gamma, z_0, L) \begin{bmatrix} q_1 \\ \vdots \\ q_4 \end{bmatrix} + \vec{\epsilon}, \qquad \text{where } \vec{\epsilon} \sim \mathcal{N}(0, \sigma^2 I_{9\times 9}).$$

To evaluate the matrix $A(\gamma, z_0, L)$ in points (γ, z_0, L) outside the experimental design we fit a Gaussian process in each of the 36 entries of $A(\gamma, z_0, L)$.

In this frame explain how, now everything is a random variable and proceed to introduce the bayesian model

Content

Probabilistic Model

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

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

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

What about q?

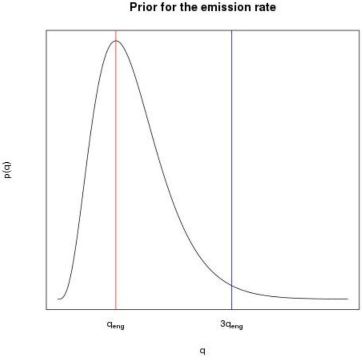
To set a prior for q we first need to check what do we know about q. This knowledge can be summarized as follows

- $q_k > 0$ for $k = 1, 2, 3, 4$.
- If we trust the engineers the most likely value for q is the engineers estimate.
- If we think they know what they are doing the true value for q cannot be very far away from their estimate.

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.

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



Getting the posterior

- We have $\mathbb{P}_{like}(\vec{m}|q,p) \propto \exp(-\frac{1}{2\sigma_\epsilon^2}\|m - A(p)q\|_2^2)$
- Also $\mathbb{P}_{prior}(q_k) \propto q_k^{\alpha_k - 1}\exp(\beta_k q_k)$.
- and $\mathbb{P}_{prior}(p) \propto \mathbf{1}_{[0.1,0.4]\times[10^{-3},2]\times[-500,-1]} := \mathbf{1}_B$

Then

$$\mathbb{P}_{post}(q,p|\vec{m}) \propto$$
$$\mathbf{1}_B \left(\prod_{k=1}^4 q_k\right)\exp(-\frac{1}{2\sigma_\epsilon^2}\|\vec{m} - A(p)q\|_2^2 + \sum_{k=1}^4 \beta_k q_k).$$

Content

In a couple of frames, talk about MCMC methods

Finally show the results you got