Gaussian Process Regresion For Hydrometallurgy

Scott Koermer Aaron Noble PhD PE

Virginia Tech

02 March, 2022

Overview

1. Problem Statement
2. Gaussian Processes
3. Optimization
4. Example on synthetic data
5. Future application

The Ideal Model

Ideally a hydromet processes would:

- be simple to model accurately
- have a low cost for collecting real data

. . . This certainly is not *always* the case.

The Ideal Model

Ideally a hydromet processes would:

- be simple to model accurately
- have a low cost for collecting real data

. . . This certainly is not *always* the case.

To make modeling hydromet processes more affordable we need methods for:

- efficiently collecting real data

The Ideal Model

Ideally a hydromet processes would:

- be simple to model accurately
- have a low cost for collecting real data

. . . This certainly is not *always* the case.

To make modeling hydromet processes more affordable we need methods for:

- efficiently collecting real data
- modeling the *bias* between a simulation and reality

The Ideal Model

Ideally a hydromet processes would:

- be simple to model accurately
- have a low cost for collecting real data

. . . This certainly is not *always* the case.

To make modeling hydromet processes more affordable we need methods for:

- efficiently collecting real data
- modeling the *bias* between a simulation and reality
- efficient optimization

The Ideal Model

Ideally a hydromet processes would:

- be simple to model accurately
- have a low cost for collecting real data

. . . This certainly is not *always* the case.

To make modeling hydromet processes more affordable we need methods for:

- efficiently collecting real data
- modeling the *bias* between a simulation and reality
- efficient optimization
- flexible modeling which quantifies uncertainty of our data and predictions

The Ideal Model

Ideally a hydromet processes would:

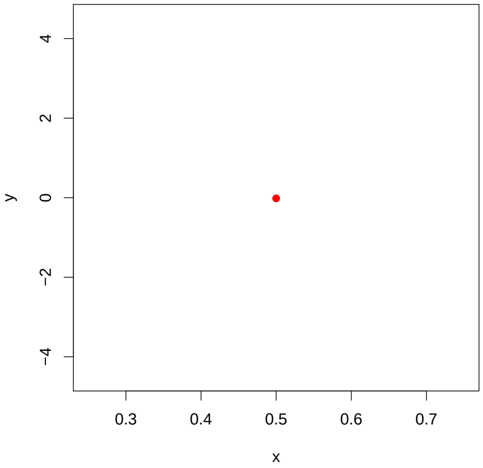
- be simple to model accurately
- have a low cost for collecting real data

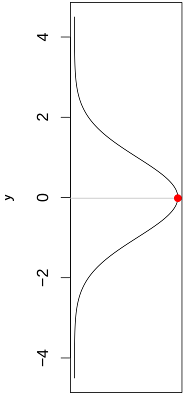
...This certainly is not *always* the case.

To make modeling hydromet processes more affordable we need methods for:

- efficiently collecting real data
- modeling the *bias* between a simulation and reality
- efficient optimization
- flexible modeling which quantifies uncertainty of our data and predictions
- uncertainty quantification

Uncertainty: A Gaussian Distribution



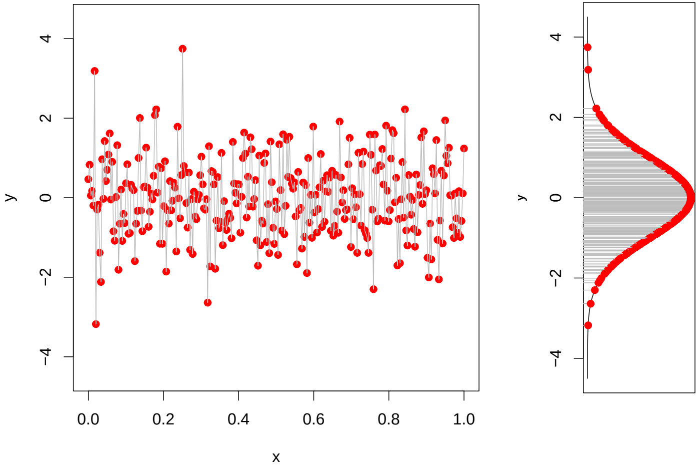


$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\frac{1}{2}(y-\mu)^2}{\sigma^2}}$$

Points independent but have a common:

- Mean (μ) of 0
- Standard deviation (σ) of 1

Multivariate Gaussian



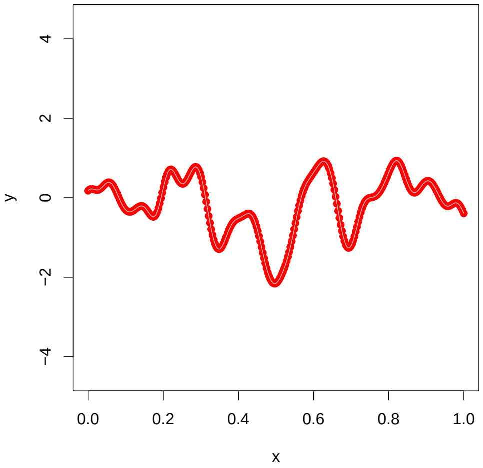
$$f(Y) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\frac{1}{2}(Y-\mu)^T(Y-\mu)}{\sigma^2}}$$

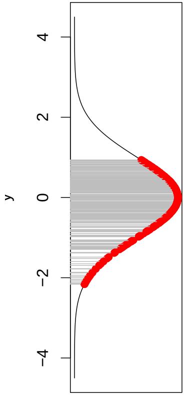
Points within a single vector *draw* are independent, vectors drawn are independent.

Points have a common:

- Mean (μ) of 0
- Standard deviation (σ) of 1

Multivariate Gaussian as a Random Function Generator

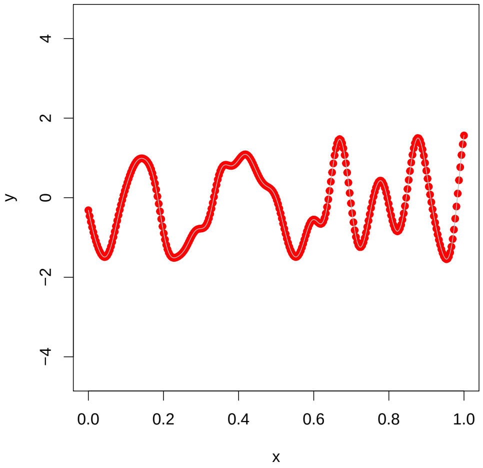


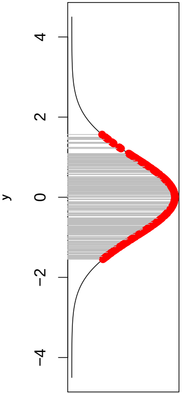


$$(2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}{\sigma^2}}$$

Covariance or Correlation Matrix Σ - distance related?

Multivariate Gaussian as a Random Function Generator

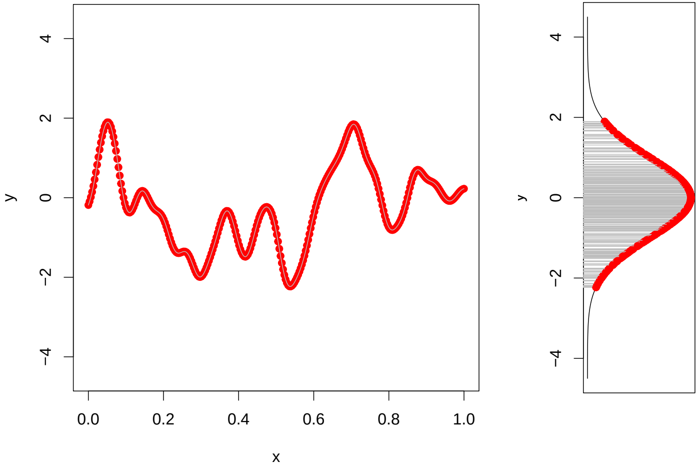




$$(2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}{\sigma^2}}$$

Covariance or Correlation Matrix Σ - distance related?

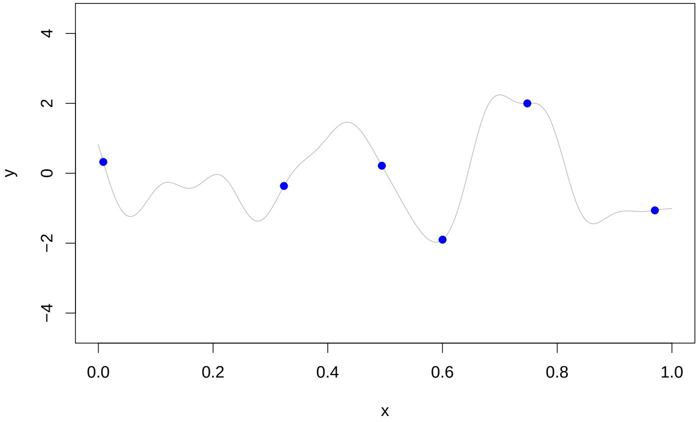
Multivariate Gaussian as a Random Function Generator



$$(2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}{\sigma^2}}$$

Covariance or Correlation Matrix Σ - distance related?

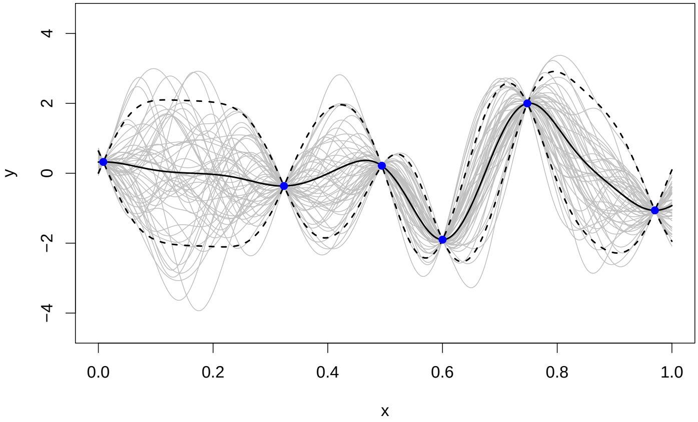
Incorporating Data



$$(2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}{\sigma^2}}$$

Covariance or Correlation Matrix Σ - distance related?

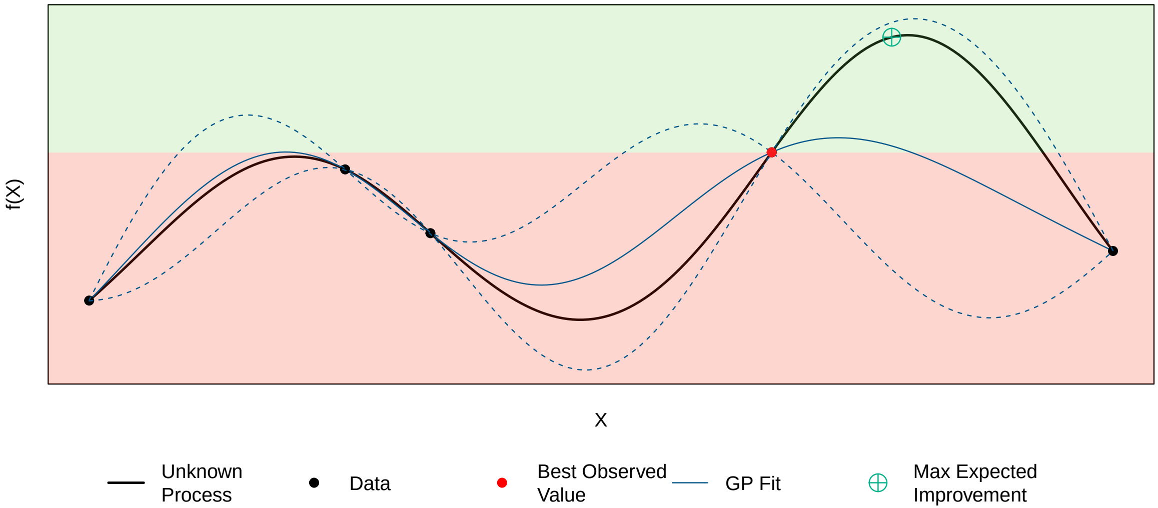
Incorporating Data



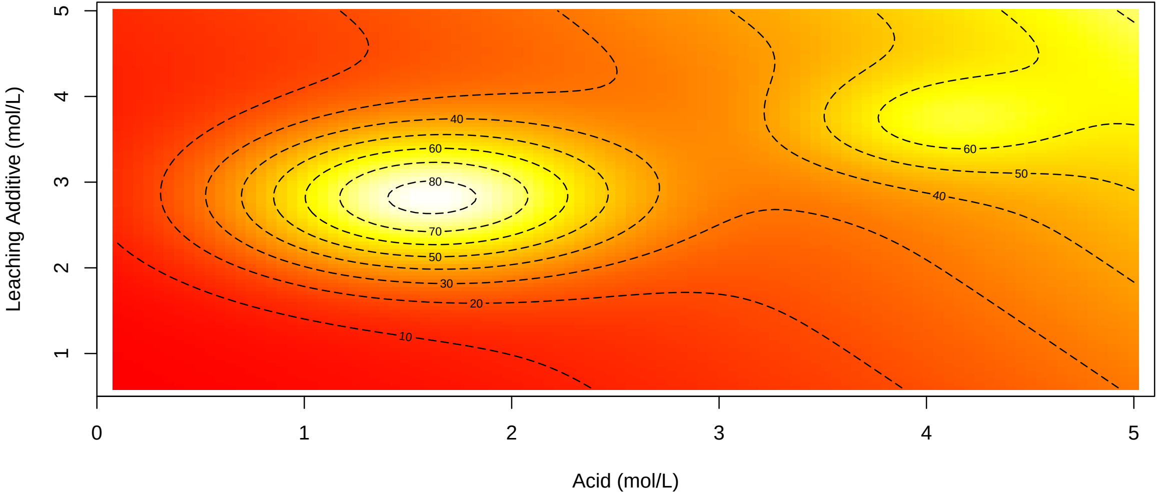
$$(2\pi\sigma^2)^{-\frac{n}{2}}|\Sigma|^{-\frac{1}{2}}e^{-\frac{\frac{1}{2}(Y-\mu)^T\Sigma^{-1}(Y-\mu)}{\sigma^2}}$$

Covariance or Correlation Matrix Σ - distance related?
Properties of this fit?

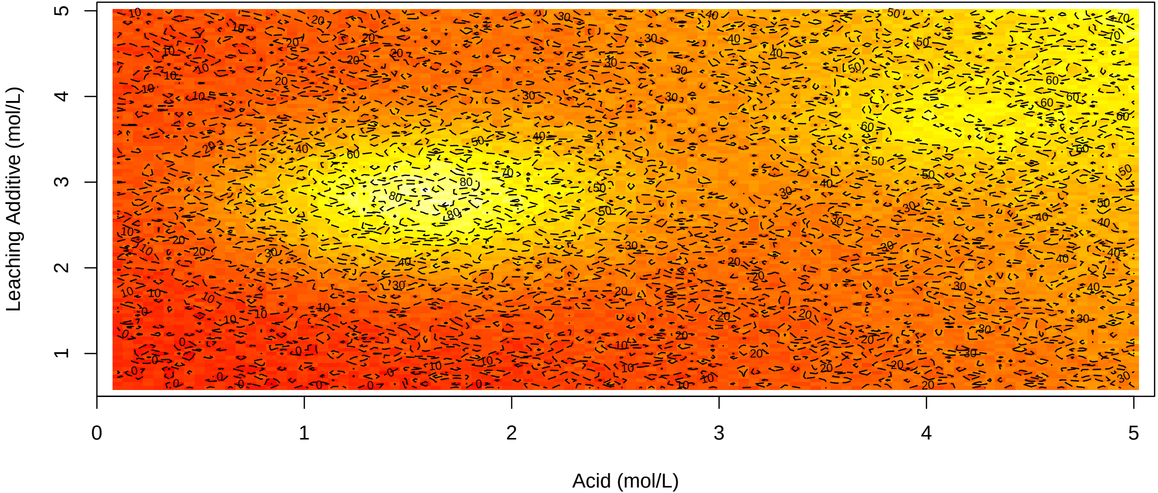
Optimization: Expected Improvement



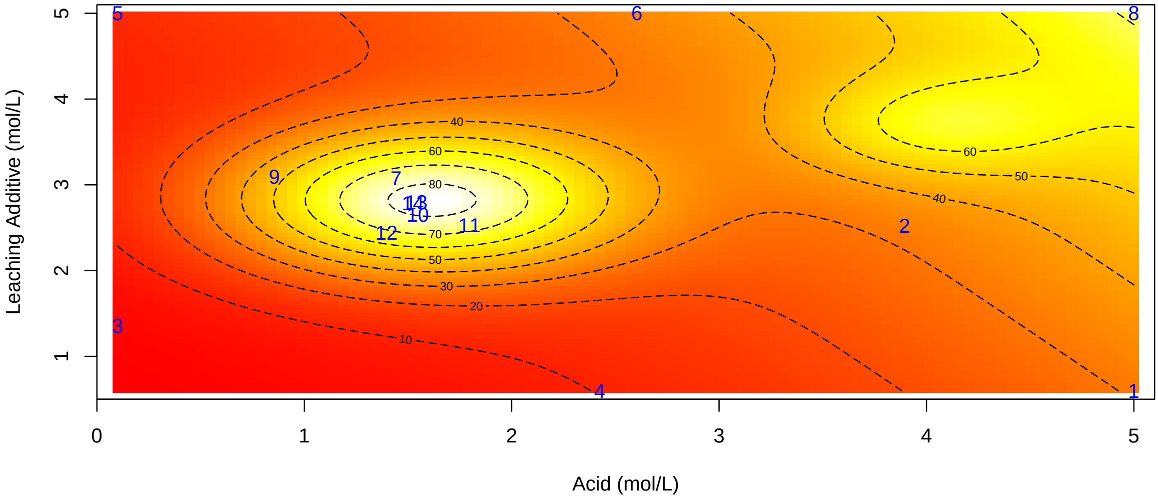
Application: TREE Recovery(%) for Simulated Leaching Process



Application: Simulated Leaching Process

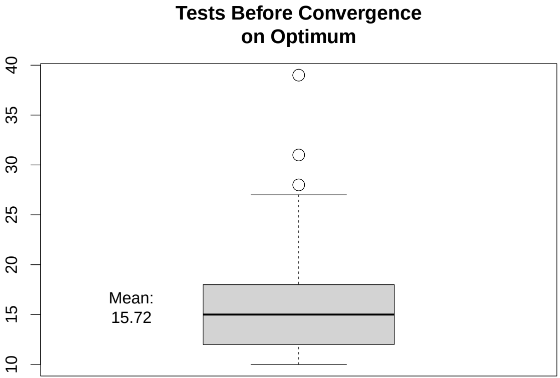


Optimization



Average Behavior

- Monte Carlo Experiment of 100 repitions
- Mean number of experiments to convergence: 15.72



Summary

- GPs are a flexible regression method
- Just like a normal distribution, GPs have an inherent quantification of uncertainty
- GP Functional uncertainty can be used for *active learning* experimental designs
- The complexity and uncertainty of hydrometallurgical processes provides a motivation for the utility of such methods.

"All models are wrong, but some are useful" -George Box

Additional Applications

GP flexibility and uncertainty quantification allows for

- Emulation and calibration of computer simulations with long computation time (dissertation work)
- Experimental design where points are specifically added to reduce prediction variance
- Modeling data where noise is a function of X

Thank You!!

Contact Info

email: skoermer@vt.edu

website: `skoermer.github.io`

Acknowledgement

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Federal Award Number: DE-FE0031834