Los Alamos
NATIONAL LABORATORY

Sensitivity Analysis in the Presence of Intrinsic Stochasticity for Multifidelity Discrete Fracture Network Simulations

Alexander C. Murph
Statistical Sciences Group

February 27th, 2024

LA-UR-24-21689

Acknowledgments

Joint work with collaborators at LANL:

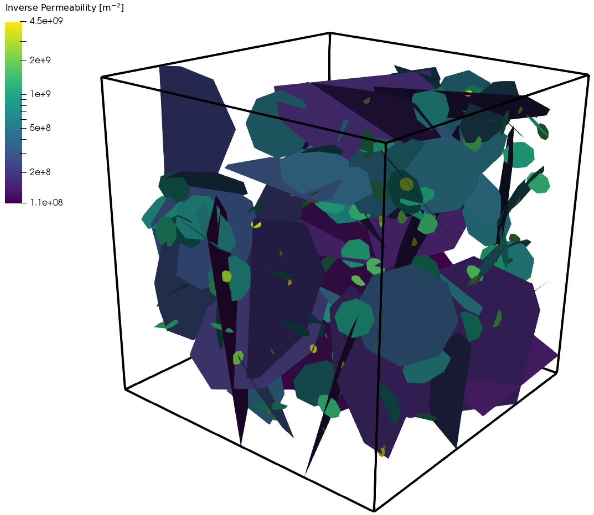
- Kelly Moran, Justin Strait – Statistical Sciences Group (CCS-6)
- Jeffrey Hyman, Philip Stauffer, Hari Viswanathan – Energy and National Resources Security (EES-16)

This talk is primarily focused on work from the following manuscript, which you can find on arXiv:

Murph et al. "Sensitivity Analysis in the Presence of Intrinsic Stochasticity for Discrete Fracture Network Simulations." *In review*, 2024.

Input Parameters of Interest

- Several parameters dictate the generation of a discrete fracture network (DFN) simulation.



Possible "Sources" of Variation

Semi-Correlation (α,β,σ):

$$\log(T) = \log(\alpha \cdot r^{\beta}) + \sigma_T \mathcal{N}(0,1),$$

Dictates relationship between fracture radius and transmissivity.

PDF of Truncated Power Law Dist'n (ɣ):

$$p_r(r, r_0, r_u) = \frac{\gamma}{r_0} \frac{(r/r_0)^{-1-\gamma}}{1-(r_u/r_0)^{-\gamma}}$$

Probability dist'n from which fracture radii are drawn

p32:

A measure of fracture density in the simulation

ε:

The 'seed variable' that accounts for random noise.

Input Parameters of Interest - Ranges

- Ranges according to information for crystalline rock and hydrogen gas.

Value	Min	Max
p32	5e-02	0.2
Semi-Correlation: α	1e-10	1e-8
Semi-Correlation: β	0.1	1.2
Semi-Correlation: σ	0.5	1.0
Radius TPL exponent	5e-2	0.2

Possible "Sources" of Variation

Semi-Correlation (α,β,σ):

$$\log(T) = \log(\alpha \cdot r^{\beta}) + \sigma_T \mathcal{N}(0,1),$$

Dictates relationship between fracture radius and transmissivity.

PDF of Truncated Power Law Dist'n (ɣ):

$$p_r(r, r_0, r_u) = \frac{\gamma}{r_0} \frac{(r/r_0)^{-1-\gamma}}{1 - (r_u/r_0)^{-\gamma}}$$

Probability dist'n from which fracture radii are drawn

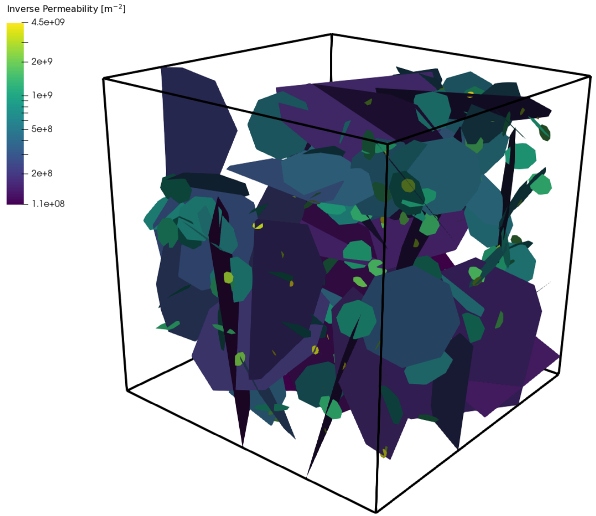
p32:

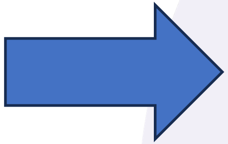
A measure of fracture density in the simulation

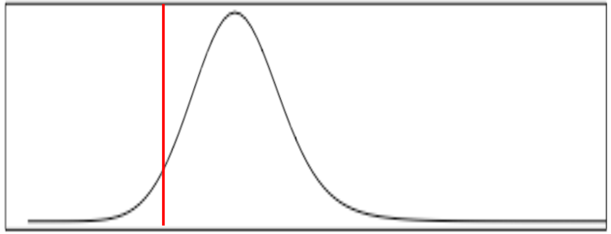
ε:

The 'seed variable' that accounts for random noise.

Output Parameters of Interest







- We are going to focus on the 10th percentile of the particle breakthrough curves.
- The variance observed for this Quantity of Interest (QoI) is relatively low compared to other quantiles.

Methods

- Naïve approach: make a finite change and measure the change in the response.
- Global SA: *Sobol Indices*

Functional ANOVA Decomposition

$$Var[f(\boldsymbol{x})] = \sum_{i=1}^{p} Var[f_i(x_i)] + \sum_{i=1}^{p} \sum_{j>i} Var[f_{ij}(x_i, x_j)] + \dots + Var[f_{1\dots p}(\boldsymbol{x})]$$

Total function variance

Variance explained by each main effect

Variance explained by each two-way interaction

Variance explained by other interactions

Sobol' Sensitivity: divide both sides by total variance

$$1 = \sum_{i=1}^{p} S_i + \sum_{i=1}^{p} \sum_{j>i} S_{ij} + \dots + S_{1\dots p}$$

Proportion variance explained by each main effect

Proportion variance explained by each two-way interaction

Proportion variance explained by other interactions

How to Decompose the Variance

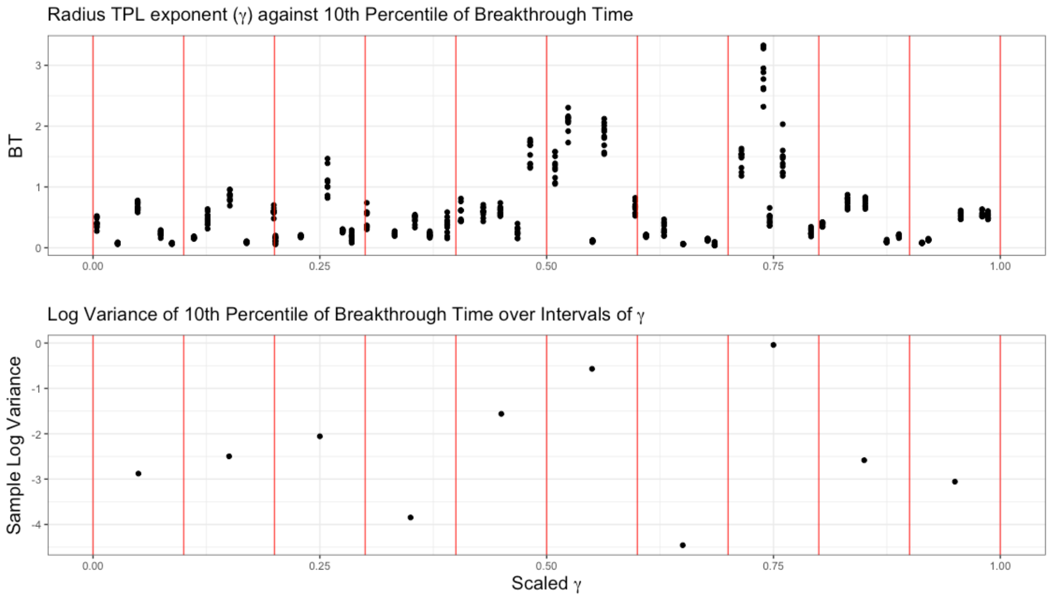
- A single run of a DFN simulation can take several days.
- Let Y be the observed breakthrough (BT) curve, x be the input parameters, and f(x) be dfnWorks (so, Y = f(x)). Beyond empirical estimates, it is difficult to calculate

$$Var[f_i(x_i)]$$

- So, we instead fit an emulator (or metamodel) $\hat{f}$(x) to substitute for f(x) in our variance calculations. For this project, we plan to fit a Gaussian Process (GP).

Quantifying Heteroskedastic & Aleatoric Noise

- **Heteroskedastic:** the *spread* of random noise changes across the input space.
- **Aleatoric:** two runs of dfnWorks at identical input values may lead to two completely different BT curves.



- Because of this, we build two Gaussian Process emulators to this process: one on the mean response term and one that explicitly models this noise.
- Known as a **heteroskedastic GP**.

To analyze how much noise is due to changing, innate stochasticity vs. how much is due to changes in the input variables, we will perform a Sensitivity Analysis using Sobol' indices

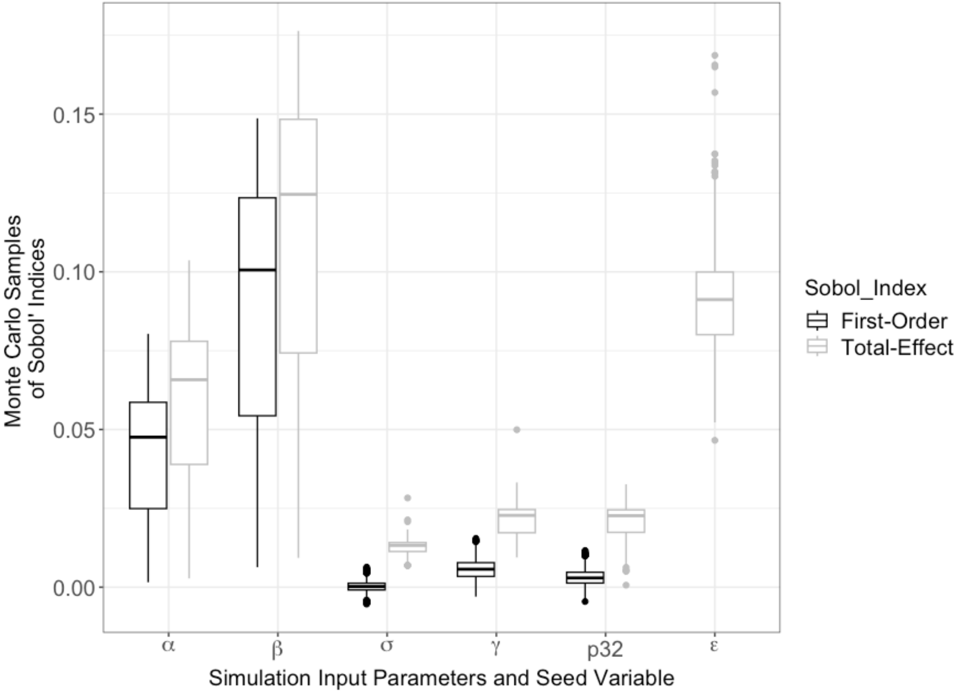
How to Estimate Sobol’ Indices

- Our problem is of small enough dimension that it’s feasible to Monte Carlo sample with our emulator. (see Saltelli et al.)
- Consequently, there are two sources of uncertainty in estimating the Sobol’ indices:
  1. Uncertainty due to using the GP to emulate the DFN simulation;
  2. Uncertainty due to the Monte Carlo approximation.
- To handle these, we repeat the following process:

Resample the data (i.e. bootstrap) → Fit the joint GP → MC Estimate Sobol’ Indices

Analyzing & Attributing Uncertainties via Sobol' Indices

Sobol' Indices calculate the proportion of variation in the output attributed to different sources



Possible "Sources" of Variation

Semi-Correlation (α,β,σ):

$$\log(T) = \log(\alpha \cdot r^{\beta}) + \sigma_T \mathcal{N}(0,1),$$

Dictates relationship between fracture radius and transmissivity.

PDF of Truncated Power Law Dist'n (γ):

$$p_r(r, r_0, r_u) = \frac{\gamma}{r_0} \frac{(r/r_0)^{-1-\gamma}}{1-(r_u/r_0)^{-\gamma}}$$

Probability dist'n from which fracture radii are drawn

p32:

A measure of fracture density in the simulation

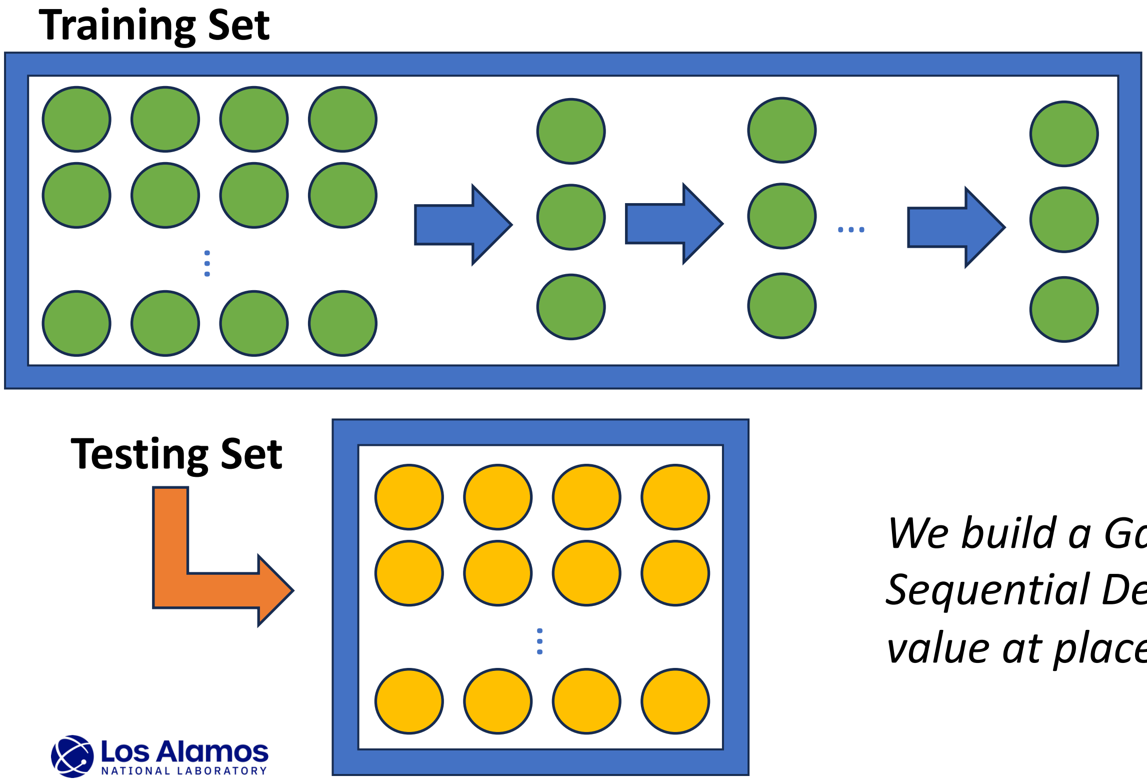
ε:

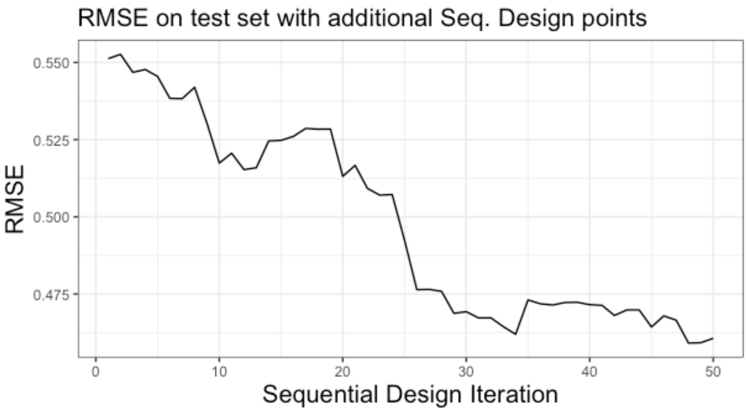
The 'seed variable' that accounts for random noise.

Big Takeaway: the variables that govern the relationship between fracture radii and transmissivity account for the greatest variation in the log breakthrough time.

Innovation: Sequential Design

Select training points according to a minimization criteria





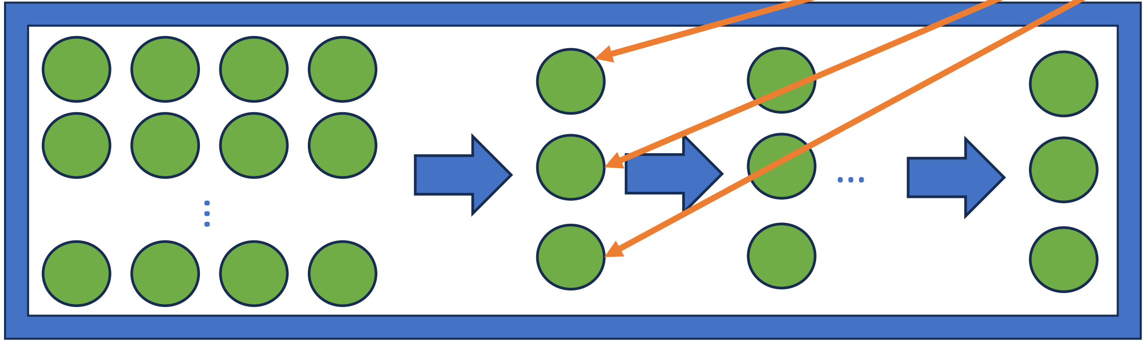
We build a Gaussian Process at each stage of the Sequential Design, then select the next sample value at places where we need more information

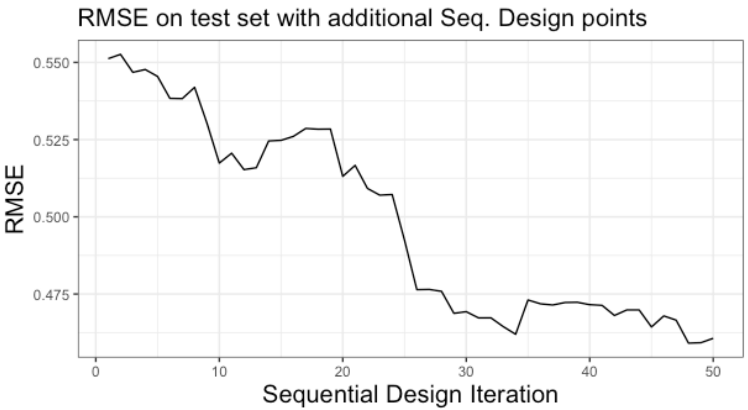
Innovation: Sequential Design

Select training points according to a minimization criteria

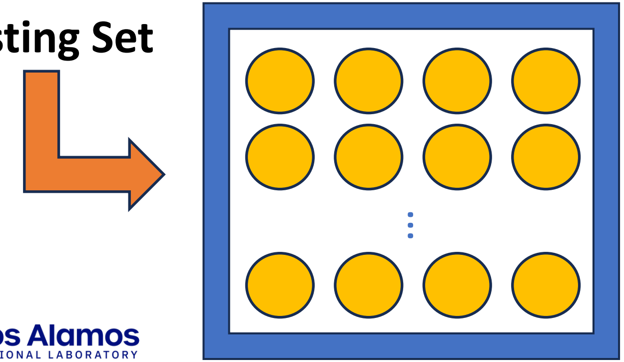
These are all at the same input location

Training Set





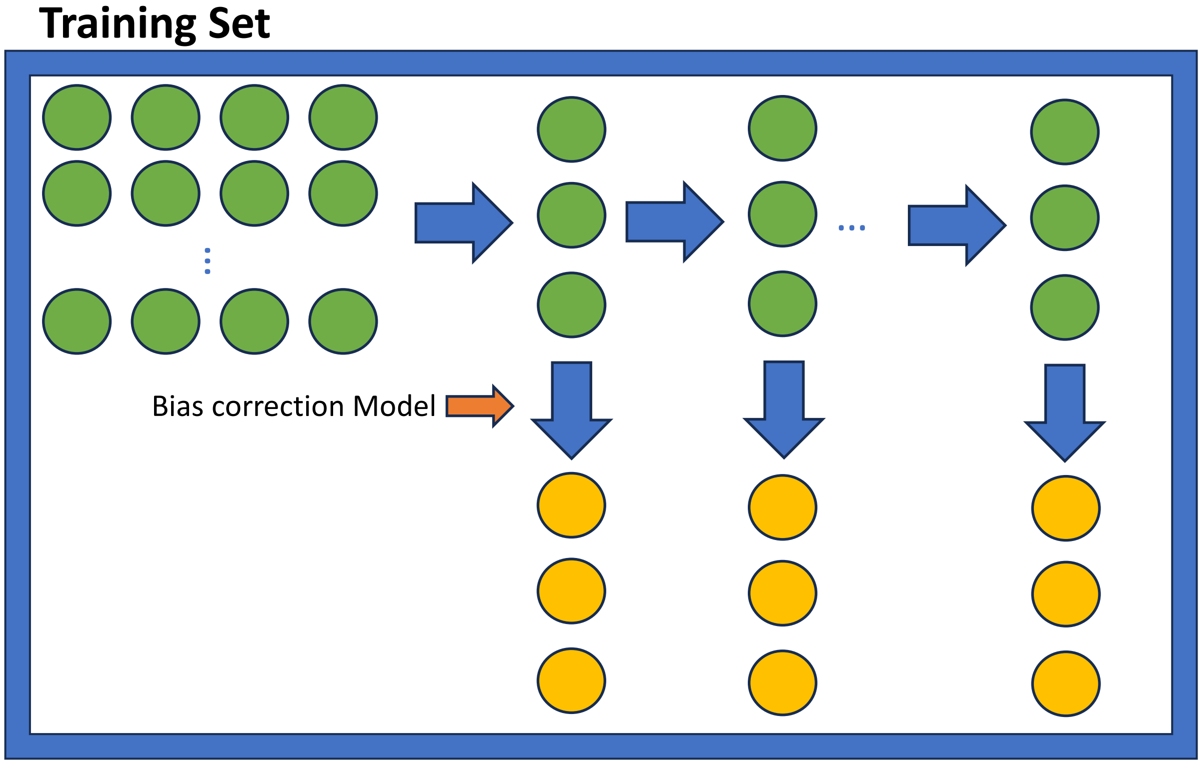
Testing Set



We build a Gaussian Process at each stage of the Sequential Design, then select the next sample value at places where we need more information

Innovation: Bias-correction for replications

Select training points according to a minimization criteria



The bias-correction model has been developed for multiple PDF data at a single input location.

At each stage of the sequential design, we can learn the bias and generate several strong data approximations.

Thank you!

Questions?
Email me: murph@lanl.gov

Paper is also available on arXiv:
https://arxiv.org/abs/2312.04722

manuscript submitted to *JGR: Solid Earth*

Sensitivity Analysis in the Presence of Intrinsic Stochasticity for Discrete Fracture Network Simulations

A.C. Murph[1], J.D. Strait[1], K.R. Moran[1], J.D. Hyman[2,3], H.S. Viswanathan[2], and P.H. Stauffer[2]

[1]Statistical Sciences (CCS-6), Computer, Computational, and Statistical Sciences Division, Los Alamos