Sensitivity Analysis with SALib (Python)

GRA Short Course - Jan. 13, 2020

Jon Herman

Assistant Professor, Civil & Environmental Engineering

UC Davis

Materials: https://github.com/jdherman/GRA-2020-SALib

Which uncertain inputs have the most influence on model outputs?

Inputs
$\boldsymbol{x} = [x_1, \dots, x_k]$

(climate, land use, uncertain system parameters, etc.)

System Model
$Y = f(\boldsymbol{x})$

Model output, Y

Sensitivity analysis in general

- For a model with K uncertain parameters, i=1,…,K
- Calculate a sensitivity index $\boldsymbol{S_i}$ for each one
- There are many different methods to do this (see Pianosi et al. 2016 for a review)

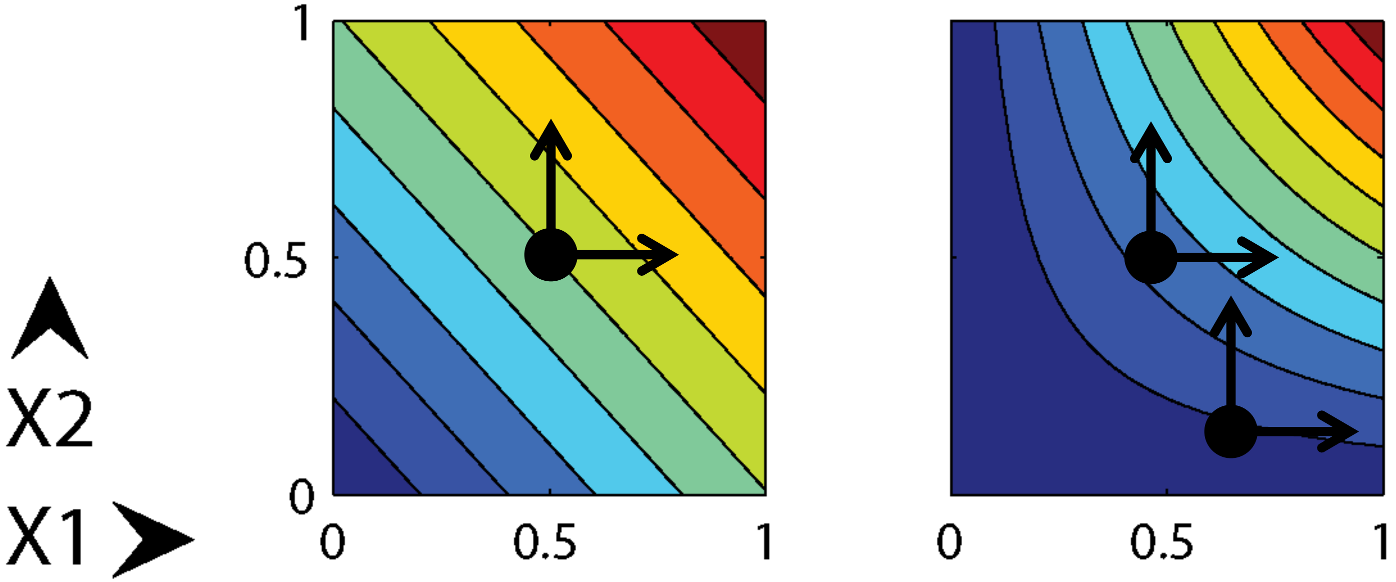
Interpret the results to figure out:

- Which parameters are most important (we should devote more effort to estimating these accurately)
- Which parameters can be ignored and fixed

UCDAVIS
UNIVERSITY OF CALIFORNIA

Local SA: Derivatives at a point

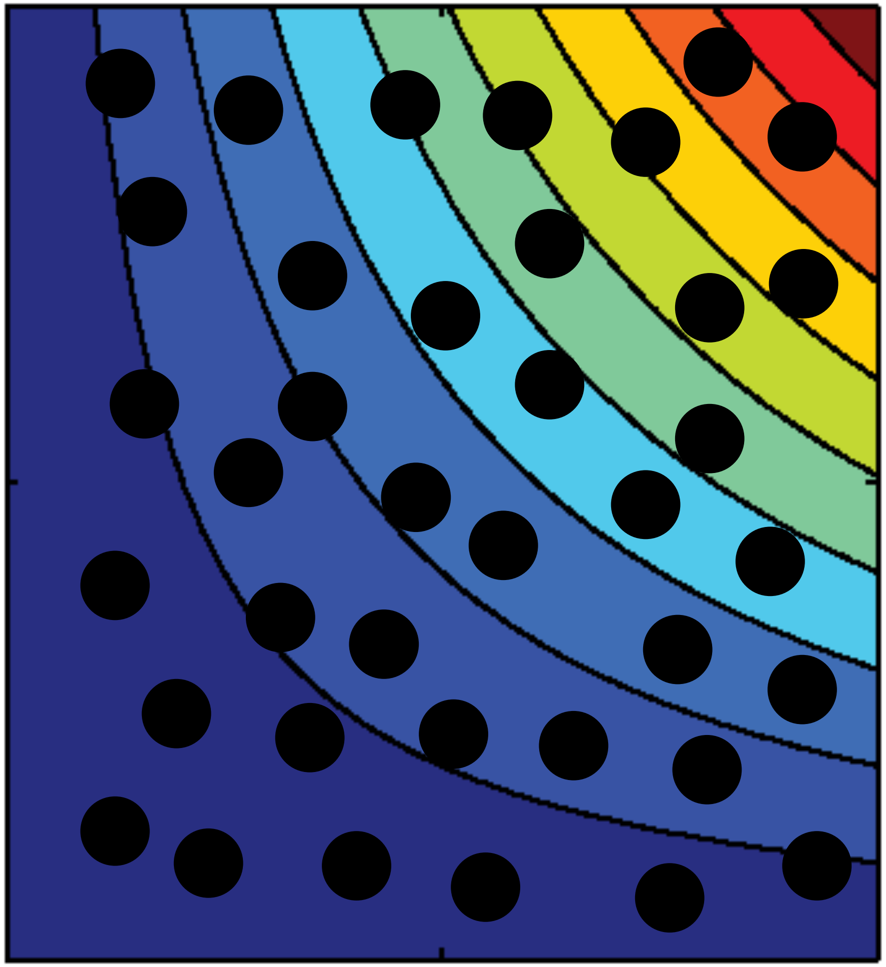
$$Y = f(x_1, x_2)\,;\;\; S_i = \partial Y / \partial x_i$$



Problem: Which point to use? Misses interactions.

UC DAVIS
UNIVERSITY OF CALIFORNIA

Global SA: Sample throughout the space



UCDAVIS
UNIVERSITY OF CALIFORNIA

Sobol Variance Decomposition

Variance of model output $V[Y]$ with K input parameters can be broken up into:

$$V[Y] = \sum_{i} V_i + \sum_{i<j} V_{ij} + \sum_{i<j<k} V_{ijk} + \ \ldots + V_{123\ldots K}$$

- V_i is the independent contribution of parameter x_i
- V_{ij} is the contribution of the second-order interaction between the pair of parameters, x_i and x_j

These can be estimated with numerical integration of the global sample

→ Saltelli et al. 2008 "Global SA: The Primer"

UCDAVIS UNIVERSITY OF CALIFORNIA

- **First-order index:** the fraction of total variance that a parameter is responsible for by itself
- **Total-order index:** the fraction of total variance that a parameter is responsible for, including interactions with other parameters

For a simple example, with three uncertain parameters:

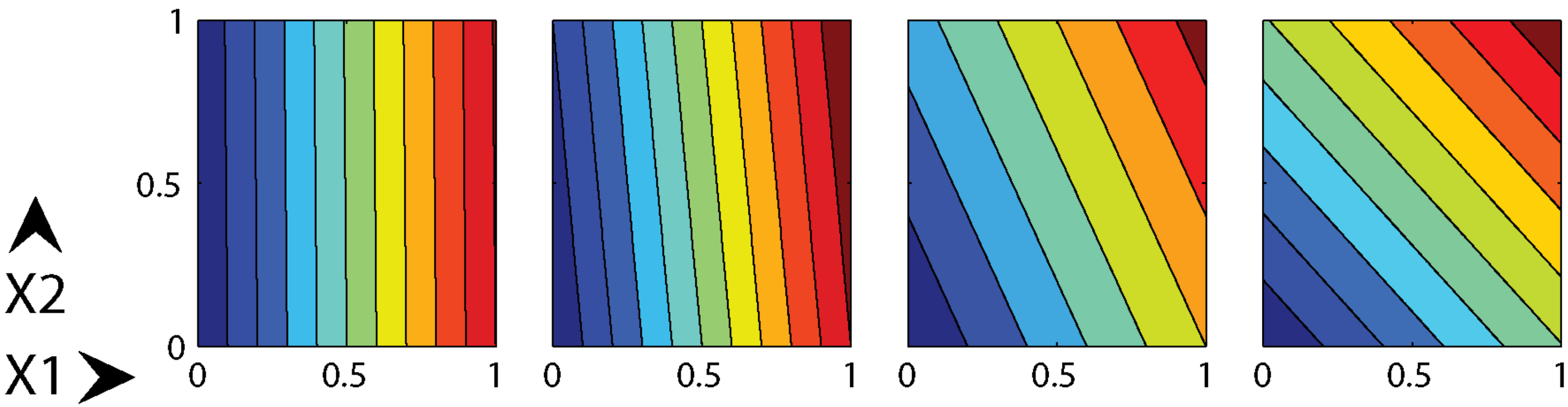
Total variance: $V(Y) = V_1 + V_2 + V_3 + V_{12} + V_{23} + V_{13} + V_{123}$

First order sensitivity index for Parameter 1: $S_1 = \frac{V_1}{V}$

Total order sensitivity index for Parameter 1: $S_{T_1} = 1 - \frac{V_{\sim 1}}{V} = 1 - \frac{V_2 + V_3 + V_{23}}{V}$

UCDAVIS
UNIVERSITY OF CALIFORNIA

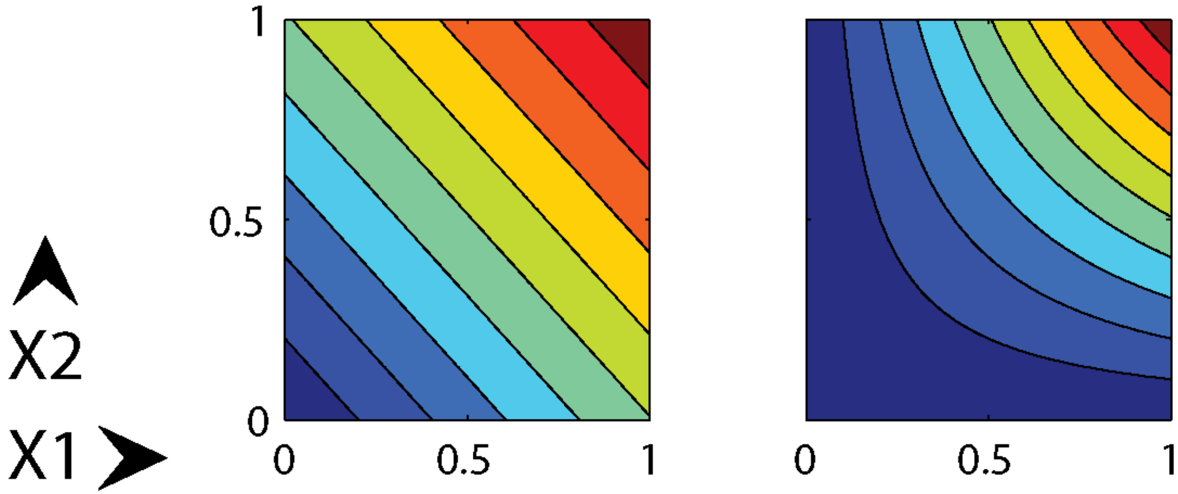
Example Sobol sensitivity indices for linear (separable) functions



Sobol SI (Total Order)	X1 = 1.0 X2 = 0.0	X1 = 0.99 X2 = 0.01	X1 = 0.8 X2 = 0.2	X1 = 0.5 X2 = 0.5

No interactions: total-order indices sum to 1

Example Sobol sensitivity indices for separable and non-separable functions



Sobol SI (Total Order)	X1 = 0.5 X2 = 0.5	X1 = 0.58 X2 = 0.58

With interactions, sum > 1 because interactions are double-counted

UCDAVIS
UNIVERSITY OF CALIFORNIA

SA: three main steps (Pianosi et al. 2016)

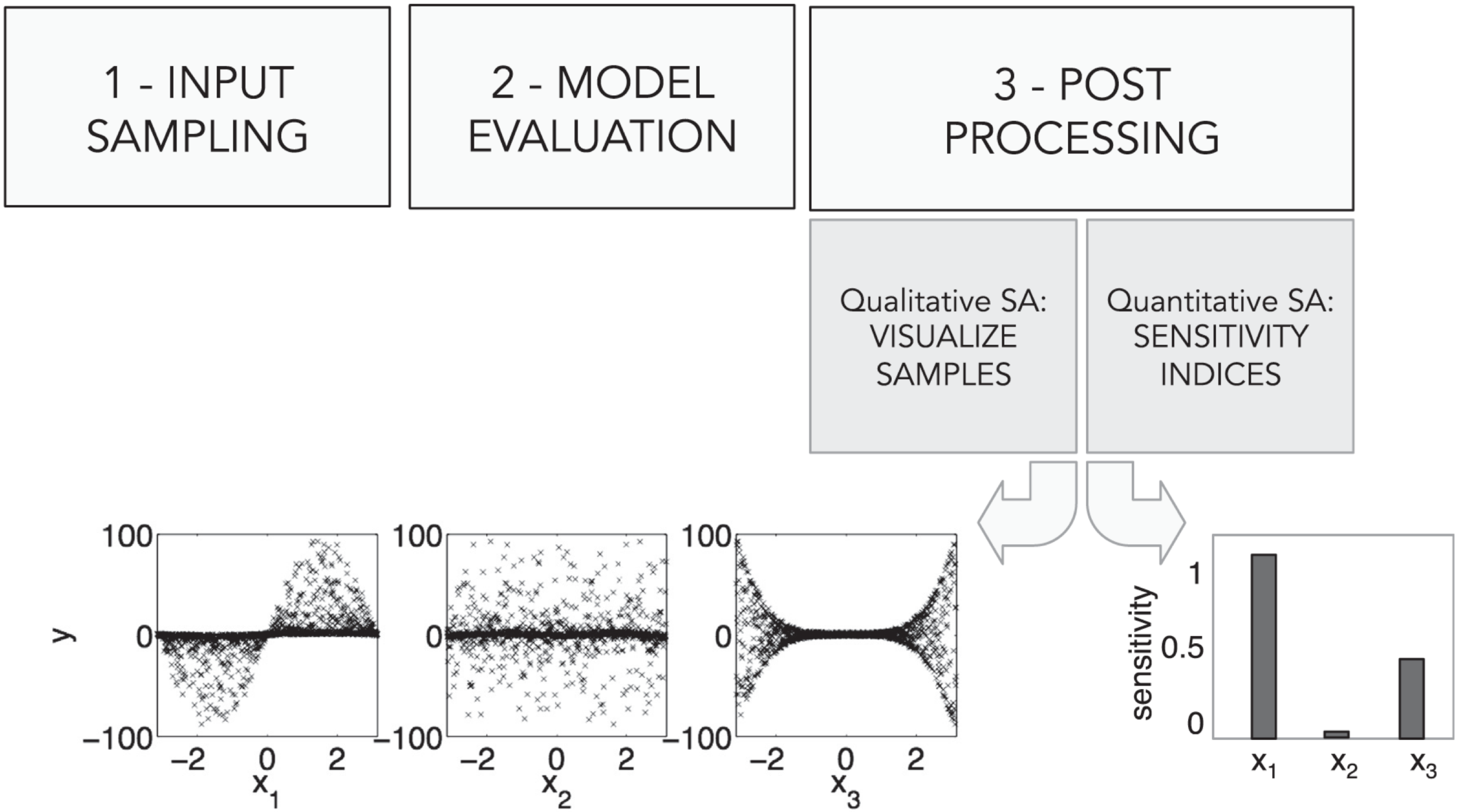


Fig. 2. The three basic steps in sampling-based Sensitivity Analysis, with an example of qualitative or quantitative results produced by the post-processing step.

Step 1: Sample parameters (Sobol method)

- Need to define upper and lower bounds for each uncertain parameter. Then, uniform sample N sets
- Cross samples, holding one param. fixed at a time
- This creates in $N(k+2)$ parameter sets to run through the model

Matrix A

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Matrix B

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Sample A and B;
From A and B,
construct a C matrix
for each parameter.

Matrix C_1

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

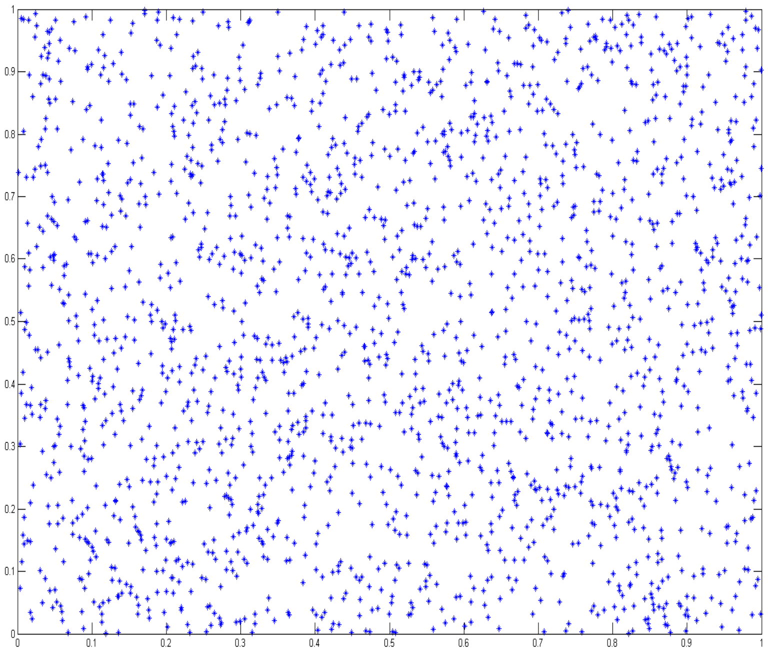
Matrix C_2

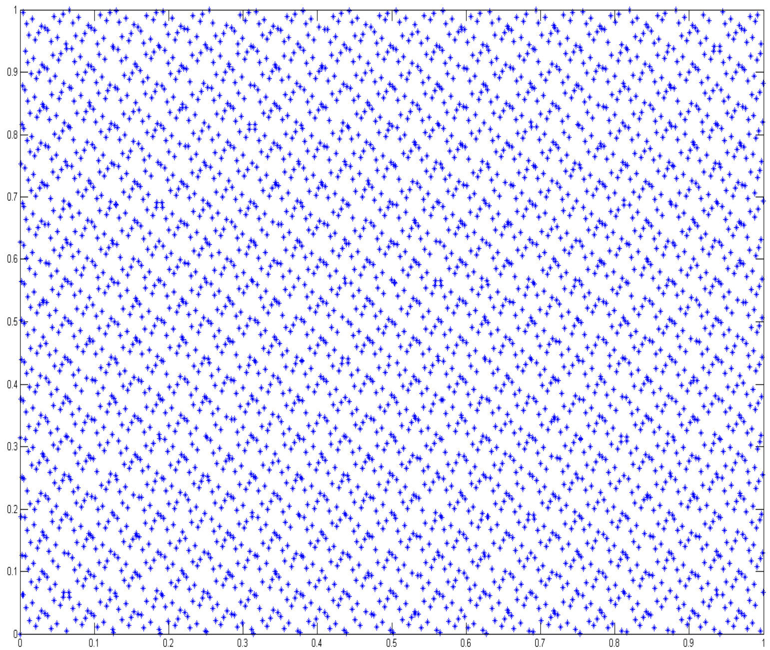
$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Matrix C_3

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Uniform Random Monte Carlo Sampling vs. Sobol Sequence Sampling ("quasi-random")





$1/\sqrt{N}$

$1/N$

Error Growth Rate
for estimating variance

UC DAVIS
UNIVERSITY OF CALIFORNIA

Step 2: Run model for all samples in the matrices A, B, and C. Save the output Y.

This step is user-specific and decoupled from everything else. Could even be in a different language, or using a GUI. Just save the output.

Step 3: Use the model output Y to *estimate* conditional variances

First calculate sample mean and variance of Y:

$$\hat{\mu}_Y = \frac{1}{N}\sum_{s=1}^{N} Y_s^A \quad , \quad \hat{\sigma}_Y^2 = \frac{1}{N}\sum_{s=1}^{N}(Y_s^A)^2 - \hat{\mu}_Y$$

Matrix A

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Matrix B

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

First-order (only parameter x_i fixed):

$$V[E(Y|X_i)] = V_i = \frac{1}{N}\sum_{s=1}^{N} Y_s^A Y_s^{C_i} - \hat{\mu}_Y$$

Matrix C_1

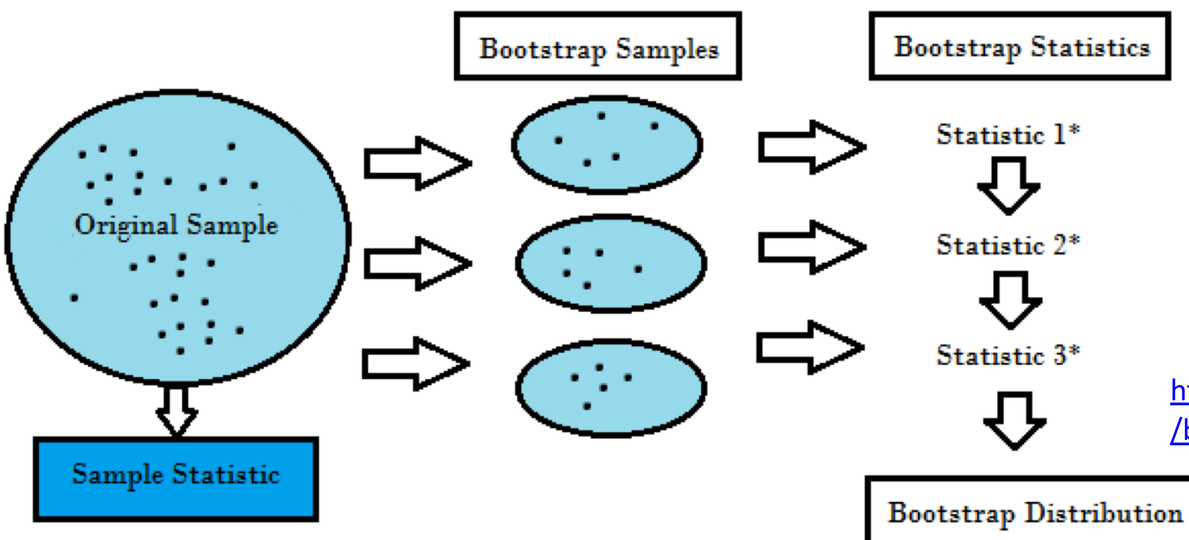
$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}$$

Total-order (everything except parameter x_i fixed):

$$V[E(Y|{\sim}X_i)] = V_{\sim i} = \frac{1}{N}\sum_{s=1}^{N} Y_s^B Y_s^{C_i} - \hat{\mu}_Y$$

Bootstrapping

- We are estimating sensitivity indices from a sample
- The accuracy of this estimate depends on N
- How to create a confidence interval for S_i, S_{T_i} without running more model evaluations?
- Bootstrapping: Resample many times from the output Y with replacement, and calculate S_i, S_{T_i}



http://www.statisticshowto.com/bootstrap-sample/

UCDAVIS UNIVERSITY OF CALIFORNIA

Sensitivity Analysis Library (SALib)

```
Herman, J. and Usher, W. (2017) SALib: An open-source Python library for sensitivity analysis.
Journal of Open Source Software, 2(9).
```

- Library: https://github.com/SALib/SALib
- Installation: `pip install SALib`
- Requirements: Python, NumPy, SciPy

ANACONDA®

https://www.continuum.io/downloads

UCDAVIS
UNIVERSITY OF CALIFORNIA

Methods included in SALib

- Sobol Sensitivity Analysis (Sobol 2001, Saltelli 2002, Saltelli et al. 2010)
- Method of Morris, including groups and optimal trajectories (Morris 1991, Campolongo et al. 2007)
- Fourier Amplitude Sensitivity Test (FAST) (Cukier et al. 1973, Saltelli et al. 1999)
- Random Balance Designs - Fourier Amplitude Sensitivity Test (RBD-FAST) (Tarantola et al. 2006, Plischke 2010, Tissot et al. 2012)
- Delta Moment-Independent Measure (Borgonovo 2007, Plischke et al. 2013)
- Derivative-based Global Sensitivity Measure (DGSM) (Sobol and Kucherenko 2009)
- Fractional Factorial Sensitivity Analysis (Saltelli et al. 2008)

- We'll focus on Sobol, but the examples/ folder on Github contains all of these

UCDAVIS
UNIVERSITY OF CALIFORNIA

Example: Ishigami function

- This is a test function used for SA method benchmarking, because we know what the answer should be.

ISHIGAMI FUNCTION

$$f(\mathbf{x}) = \sin(x_1) + a\sin^2(x_2) + b\,x_3^4\sin(x_1)$$

Description:

Dimensions: 3

The Ishigami function of Ishigami & Homma (1990) is used as an example for uncertainty and sensitivity analysis methods, because it exhibits strong nonlinearity and nonmonotonicity. It also has a peculiar dependence on x_3, as described by Sobol' & Levitan (1999).

The values of a and b used by Crestaux et al. (2007) and Marrel et al. (2009) are: a = 7 and b = 0.1. Sobol' & Levitan (1999) use a = 7 and b = 0.05.

Example

- First “pip install SALib” on the command line. Then:

```
from SALib.sample import saltelli
from SALib.analyze import sobol
from SALib.test_functions import Ishigami
import numpy as np

problem = {
  'num_vars': 3,
  'names': ['x1', 'x2', 'x3'],
  'bounds': [[-np.pi, np.pi]]*3
}

# Generate samples
param_values = saltelli.sample(problem, 1000)

# Run model (example)
Y = Ishigami.evaluate(param_values)

# Perform analysis
Si = sobol.analyze(problem, Y, print_to_console=True)
```

Results: reading the tea leaves

```
Parameter S1 S1_conf ST ST_conf
x1 0.307975 0.057222 0.560137 0.104099
x2 0.447767 0.058065 0.438722 0.038235
x3 -0.004255 0.062414 0.242845 0.026439

Parameter_1 Parameter_2 S2 S2_conf
x1 x2 0.012205 0.081241
x1 x3 0.251526 0.106296
x2 x3 -0.009954 0.069359
```

- X1 and X3 interact (second-order)
- This is reflected in the difference between their respective first- and total-order indices
- Confidence intervals should shrink as *N* increases
- Negative values are not possible – they are zero.

Frequently asked questions

Did I run enough samples?

- Check confidence intervals roughly < 10% of the S_i value

Are the parameter ranges justified?

- Subjective and very important

Why are there negative S_i values?

- This shouldn't happen – check CIs, probably $S_i = 0$

How to separate "sensitive" vs. "not sensitive" params?

- Again a subjective choice, depends on the number of parameters. But can eliminate any $S_i = 0$ (within the CI)

UCDAVIS UNIVERSITY OF CALIFORNIA

Example: rainfall-runoff model

Important questions:

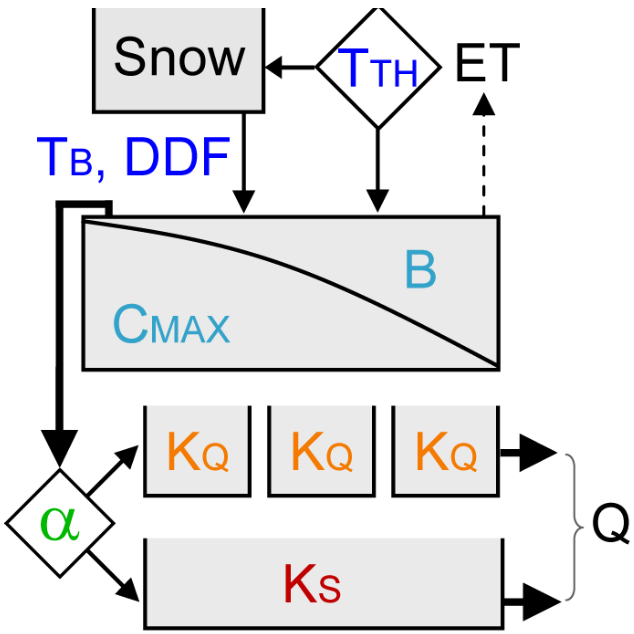
- Which model output should we use?
- What are the input parameter ranges?
- Do any of them vary over orders of magnitude? (if so, consider sampling in log space)

Uncertain parameters (6 of them, see code)

Science happens in here
$$Y = f(\boldsymbol{x})$$

Model output, Y

HyMod rainfall-runoff model (Moore 1985)



- Ignore the snow part for now
- Soil moisture bucket: 2 parameters C_{max} and β

```
Peff = P*(1 - max(1-Sm[t-1]/Sm_max,0)**B)
```

- Parameter $\alpha \in [0,1]$ is the fraction of runoff that becomes "quick flow"
- K_Q and K_S are rate constants for quick/slow flow (time^{-1})

UC DAVIS
UNIVERSITY OF CALIFORNIA

Parameter	S1	S1_conf	ST	ST_conf
Cmax	0.101145	0.045531	0.169630	0.030530
B	0.072793	0.037894	0.111594	0.023332
alpha	0.311073	0.056014	0.584154	0.106339
Kq	0.209991	0.068669	0.484375	0.091337
Ks	0.004130	0.012505	0.010668	0.002426

Parameter_1	Parameter_2	S2	S2_conf
Cmax	B	-0.027794	0.070163
Cmax	alpha	-0.000625	0.085452
Cmax	Kq	0.009421	0.086298
Cmax	Ks	-0.026197	0.066005
B	alpha	-0.007812	0.060587
B	Kq	-0.004674	0.058545
B	Ks	-0.017494	0.041870
alpha	Kq	0.190793	0.164978
alpha	Ks	0.004986	0.091721
Kq	Ks	-0.014286	0.106138

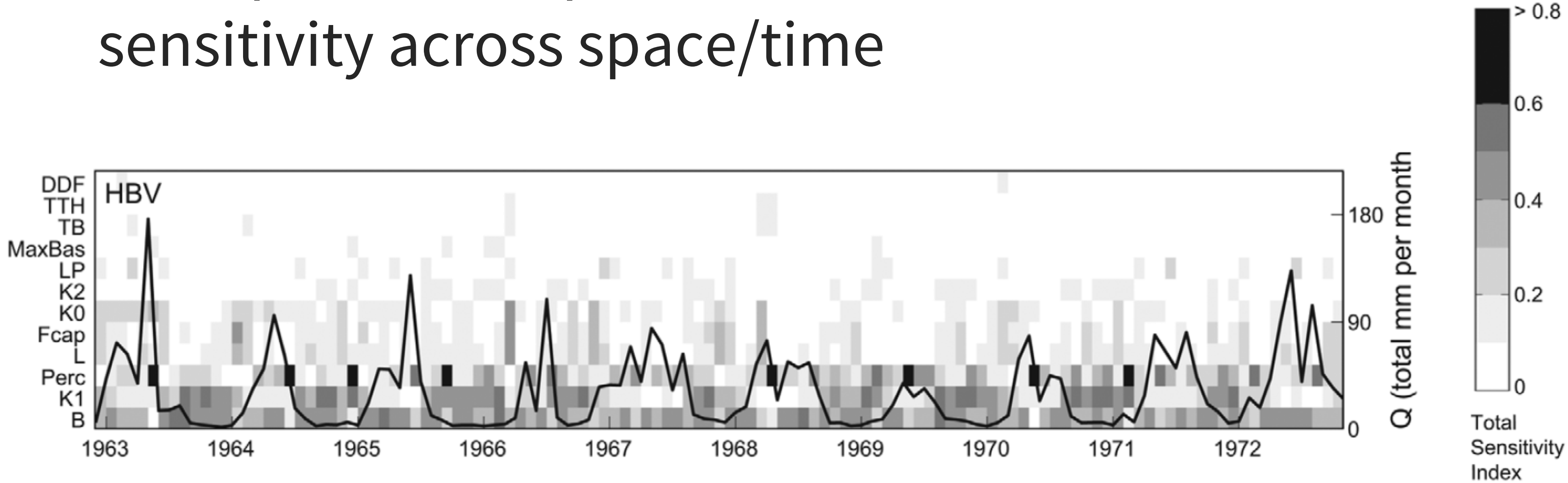
Limitations

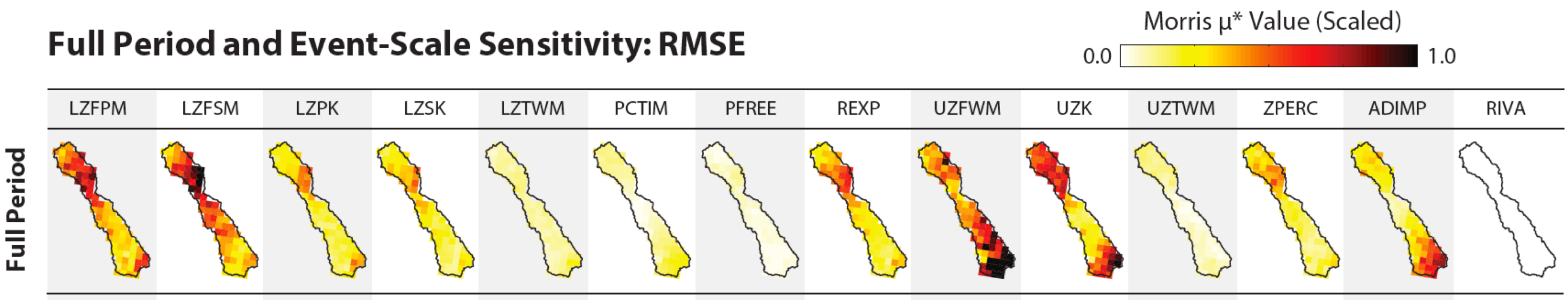
- These results are specific to the error metric (RMSE vs. some other metric)
- Also specific to the time period (precip/temp)
- … and, specific to the parameter ranges that were chosen - very important for nonlinear models

Considerations for groundwater models

- Parameter sensitivity varies across space and time
- Model runtime generally long
  - Screening methods in SALib (Morris, FAST): effective parameter ranking with fewer samples
  - Surrogate models
- Mix of uncertain environmental and human parameters
  - Implications for interpreting SA results

Example results: parameter sensitivity across space/time





UCDAVIS
UNIVERSITY OF CALIFORNIA

Sensitivity analysis: Summary

- A way of analyzing simulation models as a black box
- Which parameters matter most?
  - Which should we invest in measuring better?
  - Which ones can we ignore?
- Local vs. global SA
- 3 decoupled steps: sample parameters, run model, analyze results

UCDAVIS
UNIVERSITY OF CALIFORNIA