Sensitivity Analysis with SALib (Python)

DMDU Tutorial - Nov. 13, 2018

Jon Herman

Civil & Environmental Engineering, UC Davis

Which uncertain inputs have the most influence on system performance?

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

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

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

Model output, Y

SA: The general idea

- 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)

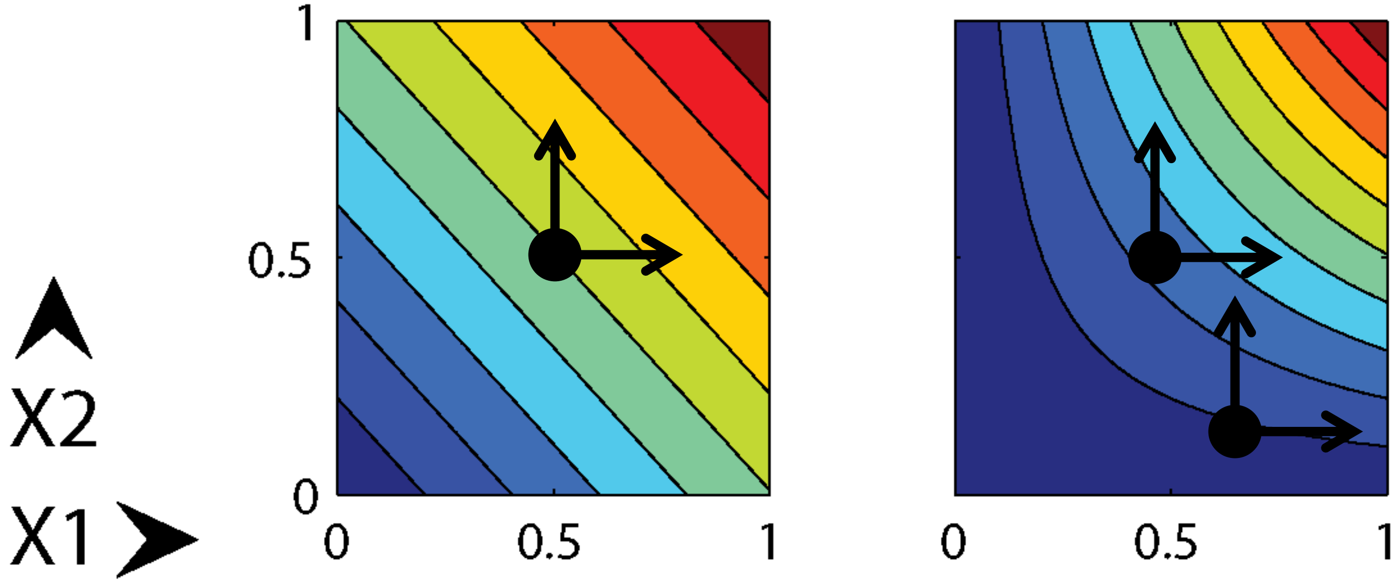
Interpret the results to figure out:

- 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.

UCDAVIS
UNIVERSITY OF CALIFORNIA

Global SA: Sample throughout the space

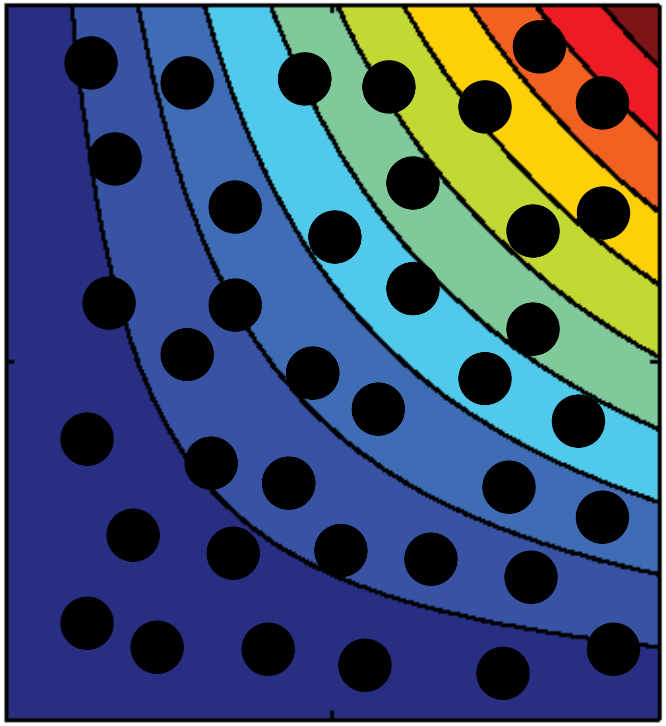
Variance Decomposition

$$D(f) = \sum_i D_i + \sum_{i<j} D_{ij} + \sum_{i<j<k} D_{ijk} + D_{12\ldots p},$$

$$\text{First-Order Index:} \quad S_i = \frac{D_i}{D},$$

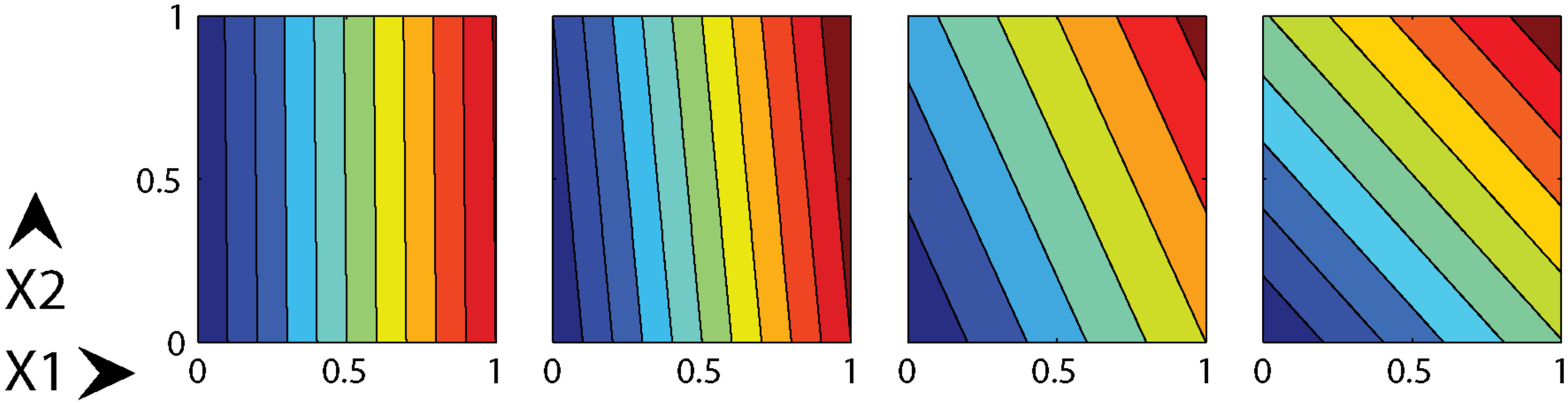
$$\text{Total-Order Index:} \quad S_{T_i} = 1 - \frac{D_{\sim i}}{D}.$$

These can be estimated with numerical integration of the global sample



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

Example Sobol sensitivity indices for linear (separable) functions

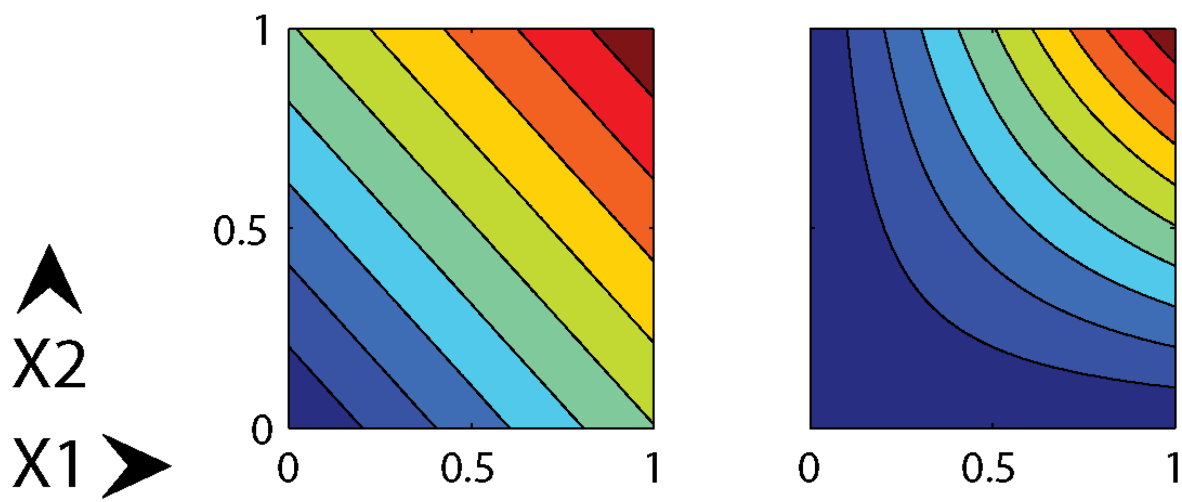


Sobol SI (Total Order)	X1 = 1.0 X2 = 0.0	X1 = 0.99 X2 = 0.01	X1 = 0.8 X2 = 0.2	X1 = 0.5 X2 = 0.5

No interactions: total-order indices sum to 1

UNIVERSITY OF CALIFORNIA

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

UNIVERSITY OF CALIFORNIA

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

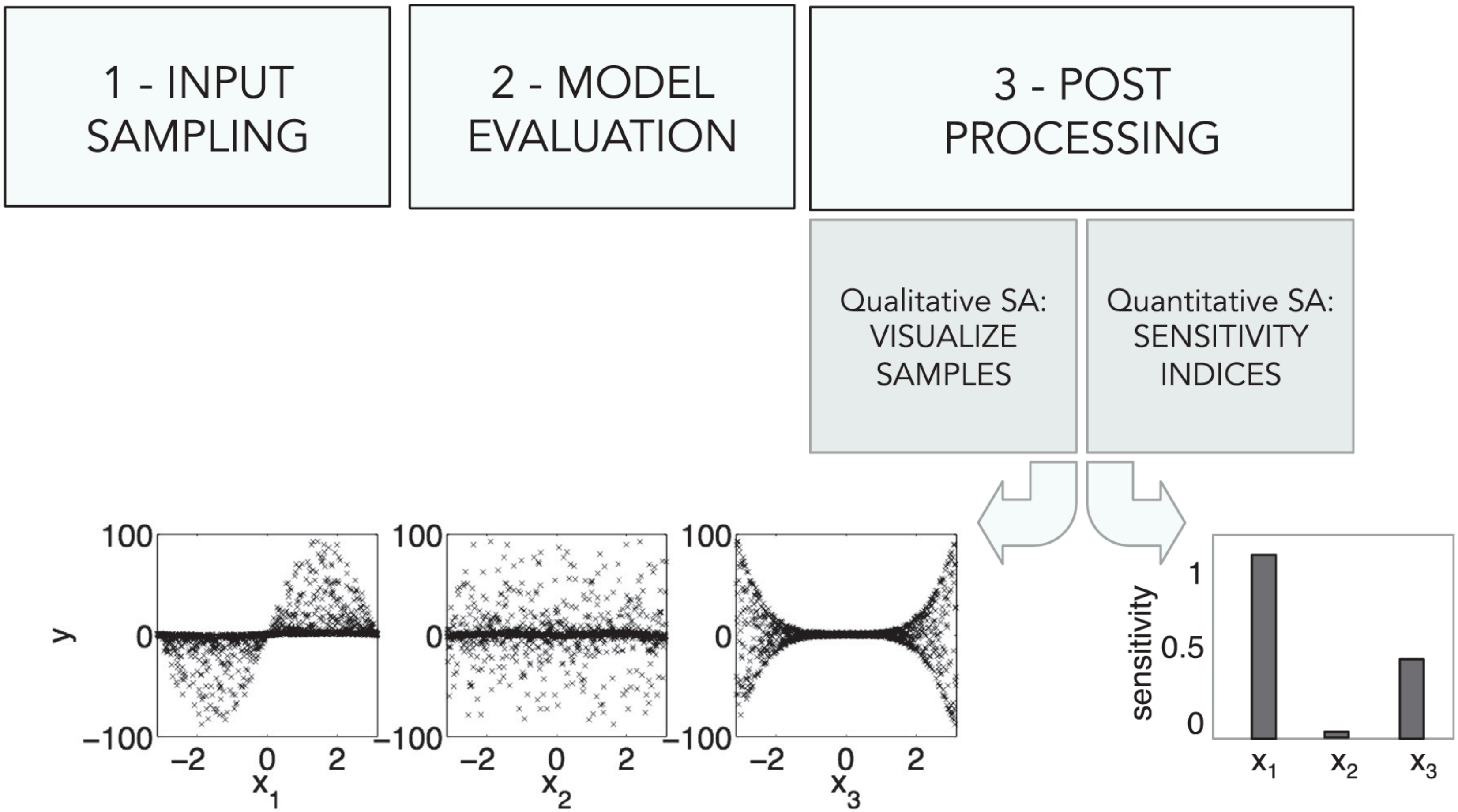


Fig. 2. The three basic steps in sampling-based Sensitivity Analysis, with an example of qualitative or quantitative results produced by the post-processing step.

- **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: $S_{T_1} = 1 - \frac{V_{\sim 1}}{V} = 1 - \frac{V_2 + V_3 + V_{23}}{V}$

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

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

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.

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

$$f_0 = \frac{1}{n}\sum_{s=1}^{n} Y_s{}^A \qquad V(Y) = \frac{1}{n}\sum_{s=1}^{n}(Y_s{}^A)^2 - f_0{}^2$$

Mean & variance of model output Y

$$V[E(Y|x_1)] = \frac{1}{n}\sum_{s=1}^{n} Y_s{}^A\, YsC^1 - f_0{}^2$$

Examples of conditional variances for parameter x_1

$$V[E(Y|{\sim}x_1)] = \frac{1}{n}\sum_{s=1}^{n} Y_s{}^B\, YsC^1 - f_0{}^2$$

Conditional variances are scalar products.

10

- Different estimators have been developed, analyzed based on whether they converge faster with N
- Confidence intervals for S_i are estimated by bootstrapping (default 95% CI)

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`
- Tutorial materials: https://github.com/jdherman/DMDU-2018-SALib-Tutorial
- Requirements: Python, NumPy, SciPy

ANACONDA

https://www.continuum.io/downloads

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.

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.

UCDAVIS
UNIVERSITY OF CALIFORNIA

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

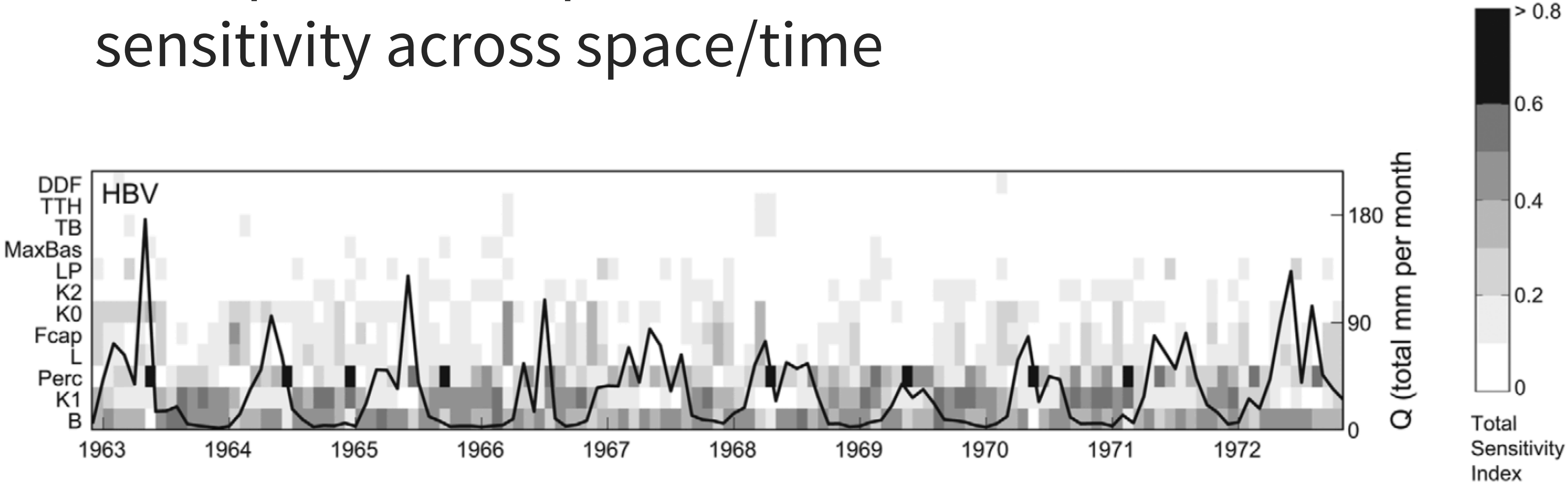
Why are there negative S_i values?

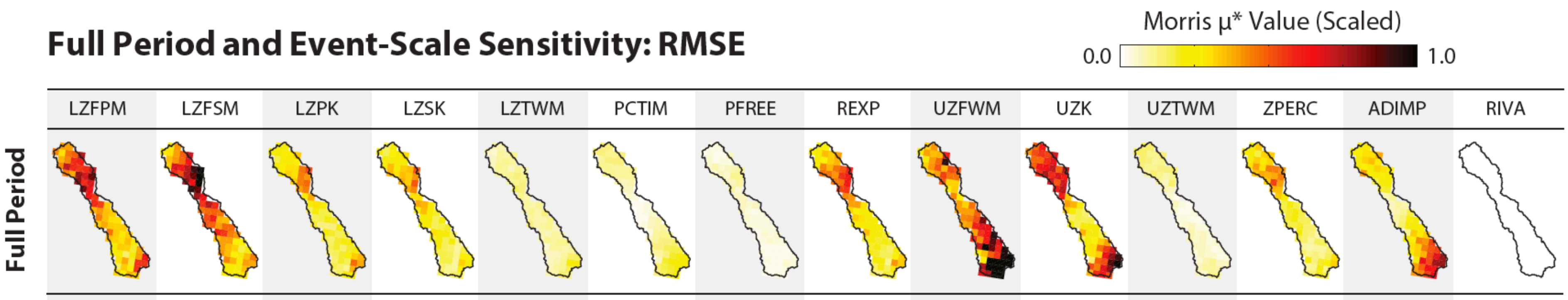
- This shouldn't happen – check CIs, probably $S_i = 0$

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)

Example results: parameter sensitivity across space/time





UC DAVIS
UNIVERSITY OF CALIFORNIA

Role of SA in decision support

- As a model diagnostic: which assumptions need to be refined?
- Some inputs are uncertain because they can't be measured perfectly (e.g. parameters measured within +/- 20%)
- Others are uncertain because we simply can't know them perfectly (e.g. avg.water availability in 50 years)
- Relationship to Scenario Discovery methods