

CORTEX Training Course

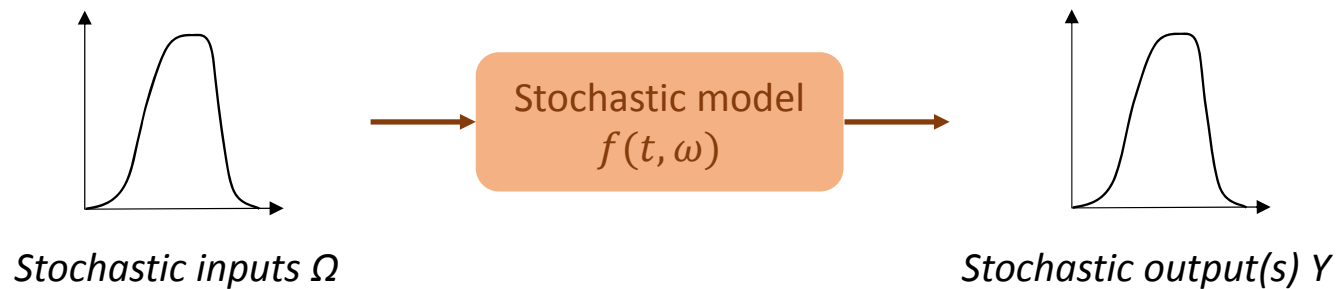
Introduction to the Methodology of Sensitivity Analysis

SOOBEEN YUM

Technical University of Munich

Dec 05, 2019

What is Sensitivity Analysis?



Question

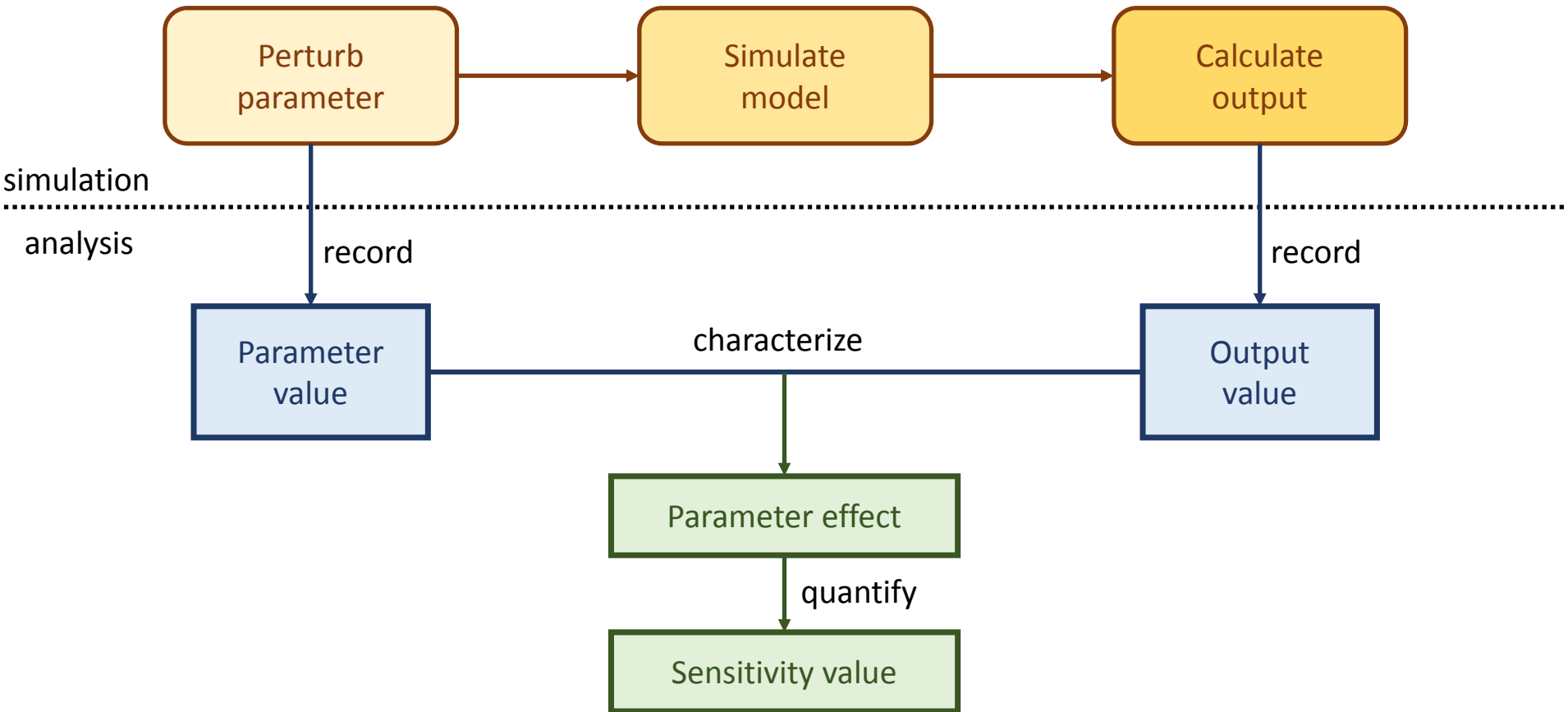
- How sensitive is Y to changes in $\omega \in \Omega$?
- What is the relative contribution of $\omega_i, i = 1, \dots, d$ to the output uncertainty?

Expectation

- Compute "sensitivities" at a very low computational cost: Which uncertain parameters contribute most to the stochastic output Y ?
- Reasonable "measure" of the output uncertainty

What is Sensitivity Analysis?

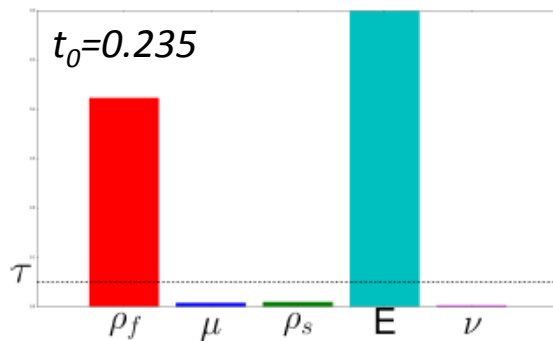
Sensitivity analysis can be regarded as a systematic “perturbation analysis”



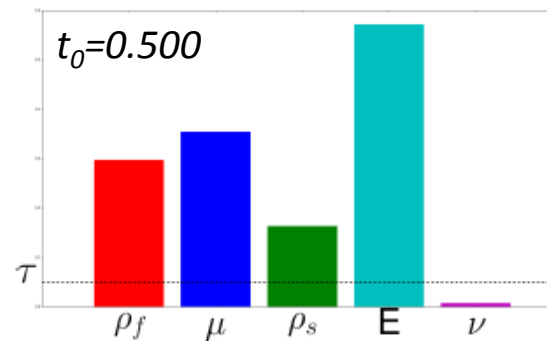
Why do we need a Sensitivity Analysis?

- Ascertain robustness of underlying model w.r.t. various parameters
- Stochastic dimensionality reduction → Can model be simplified by fixing insensitive parameters to deterministic value?
- Guide experimental design to determine measurement regimes that have greatest impact on parameter or response sensitivity

Example : Total Sobol's index (at different times t) ¹⁾



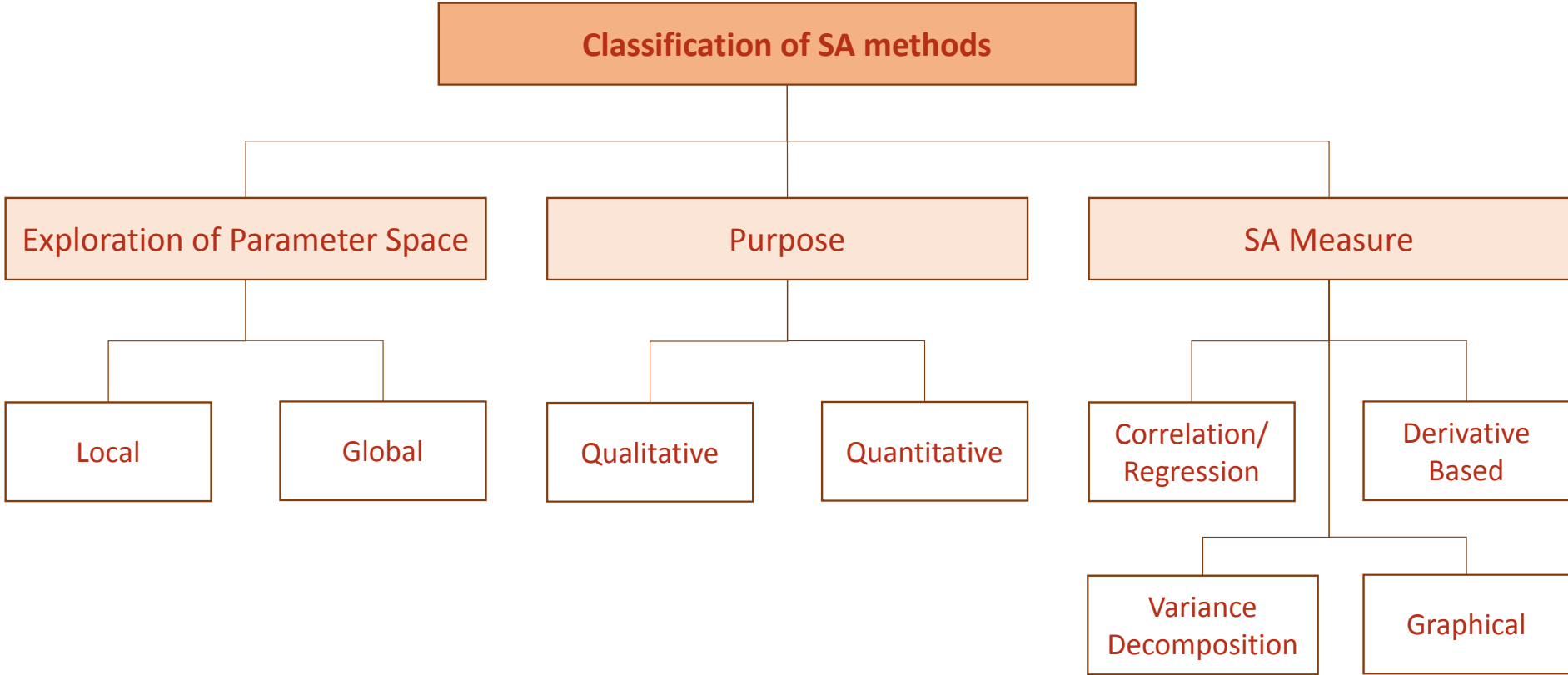
Dimension reduction $5 \rightarrow 2$
∴ 50h compute time saved



Dimension reduction $5 \rightarrow 4$
∴ 20h compute time saved

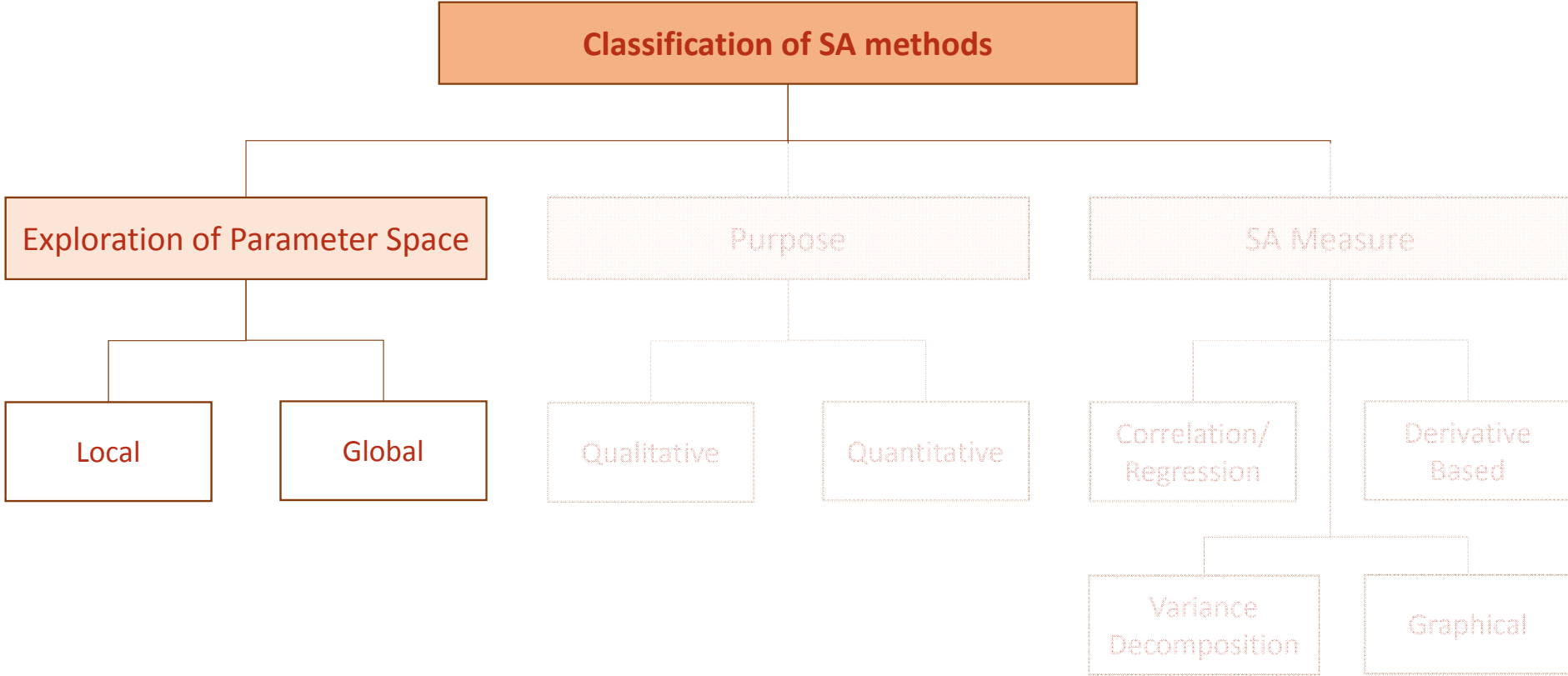
¹⁾ Tobias Neckel, Lecture material, Scientific Computing in Computer Science TUM, 2018.

Which methods are available for Sensitivity Analysis?



Yogesh Khare, Hydrologic and water quality models: Sensitivity, American Society of Agricultural and Biological Engineers, Vol.58(6), pp. 1721-1744, 2015.

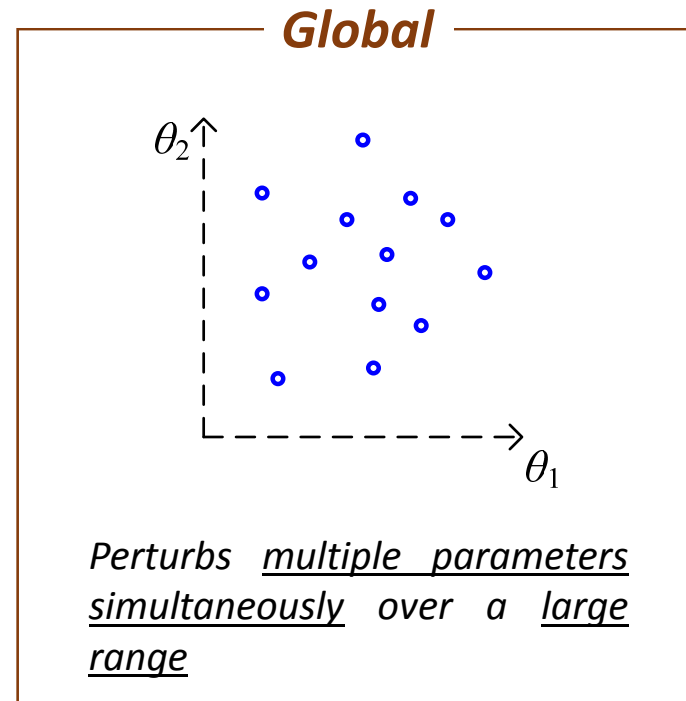
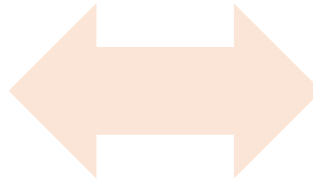
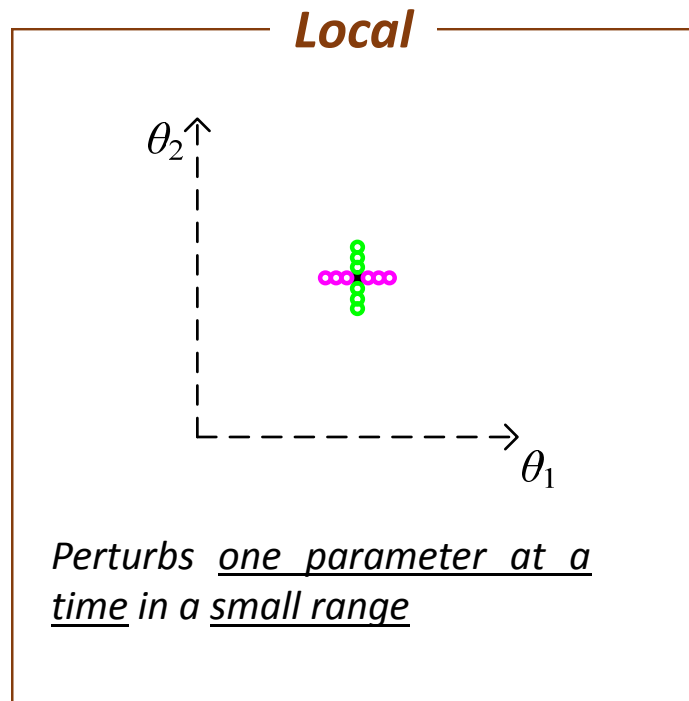
Which methods are available for Sensitivity Analysis?



Yogesh Khare, Hydrologic and water quality models: Sensitivity, American Society of Agricultural and Biological Engineers, Vol.58(6), pp. 1721-1744, 2015.

Which methods are available for Sensitivity Analysis?

Depending upon how parameter values are perturbed



Which methods are available for Sensitivity Analysis?

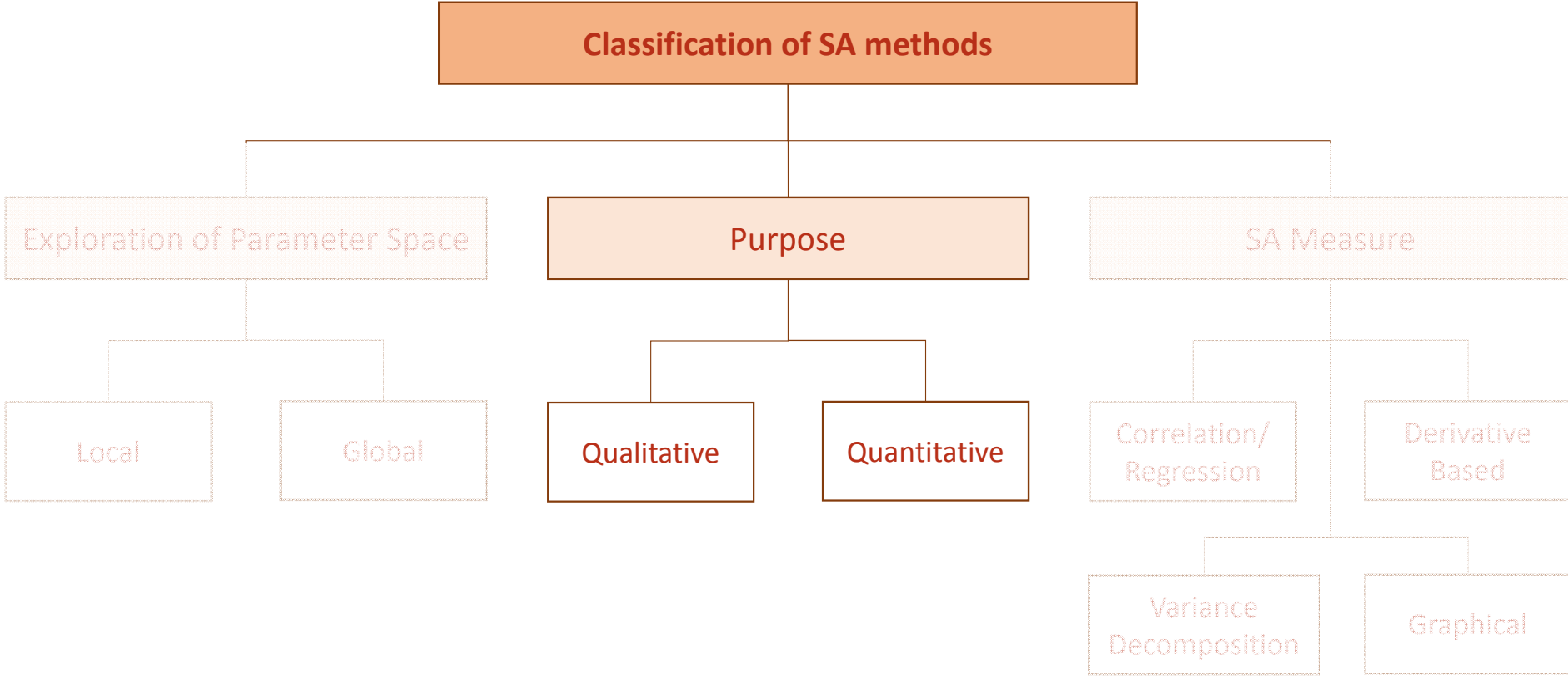
Local sensitivity analysis

- *Assess the “Sensitivity” of output with respect to inputs perturbed about a nominal value*
- *Generally uses gradients of output with respect to inputs*
- *Can be used to “screen” out the “insensitive” uncertain inputs → one input at a time (OAT)*
- *Tool for stochastic dimensionality reduction before doing the actual forward uncertainty propagation*

Global sensitivity analysis

- *Based on analyzing a suitable “measure” of uncertainty, e.g. the variance*
- *Quantifies the contribution of each input to the output uncertainty*

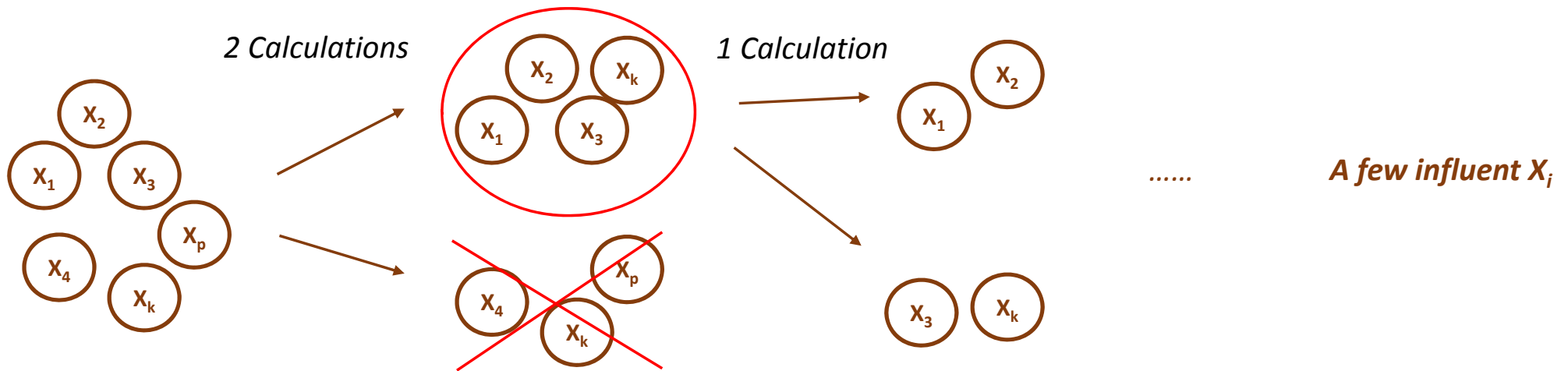
Which methods are available for Sensitivity Analysis?



Which methods are available for Sensitivity Analysis?

Qualitative sensitivity analysis

- Many inputs ($p \gg 10$) and CPU time costly computer code
- Objective: less computations than number of inputs
- Hypotheses: Number of influent inputs \ll total number of inputs
Monotony of the model, no interaction between inputs
- Example: method of sequential bifurcations



Which methods are available for Sensitivity Analysis?

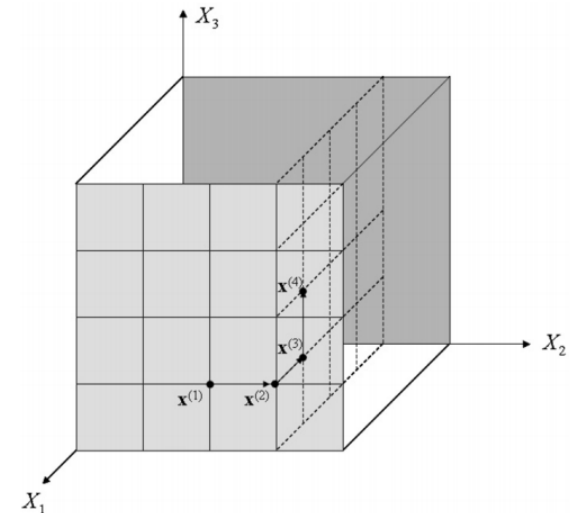
Qualitative sensitivity analysis

- Identifies the influential parameters → reduces the burden on further model analysis
- Low computational cost
- ~ screening method
- E.g. Elementary Effects (EE) Method: Extension of OAT approach but a global method

$$EE_i = \frac{Y(x_1, \dots, x_{i-1}, x_i + \Delta i, x_{i+1}, \dots, x_k) - Y(x_1, \dots, x_k)}{\Delta i}$$

$$\mu_i = \frac{\sum_{i=1}^r EE_i}{r}$$

- μ_i is an indicator of total or overall importance



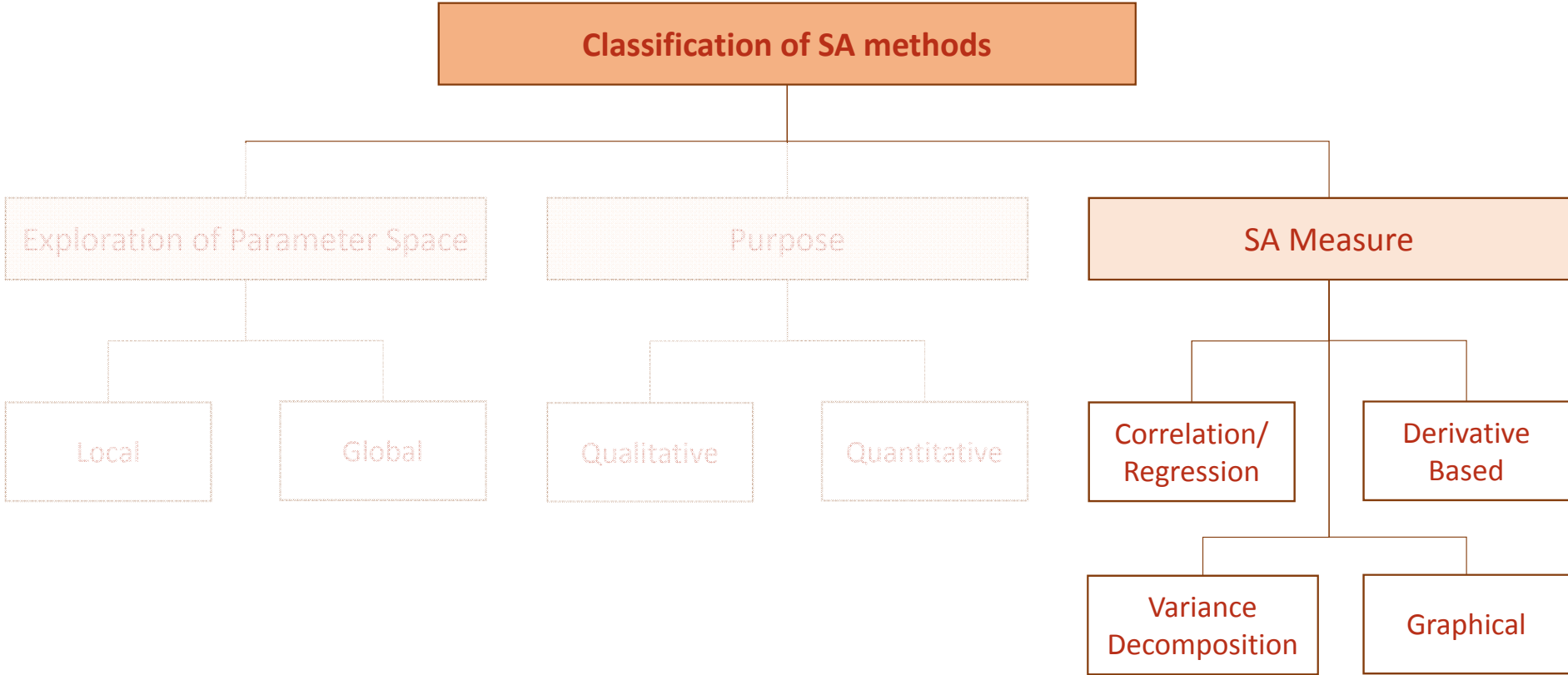
F. Campolongo et al, From screening to quantitative sensitivity analysis. A unified approach, Computer Physics Communications. vol.182, 978-988, 2011.

Which methods are available for Sensitivity Analysis?

Quantitative sensitivity analysis

- *Quantifies the contributions of individual parameters to output variance*
- *Estimates which percentage of output variance each factor is responsible for*
- *Obtaining good estimates of quantitative sensitivity indices → High number of model evaluation is needed!*

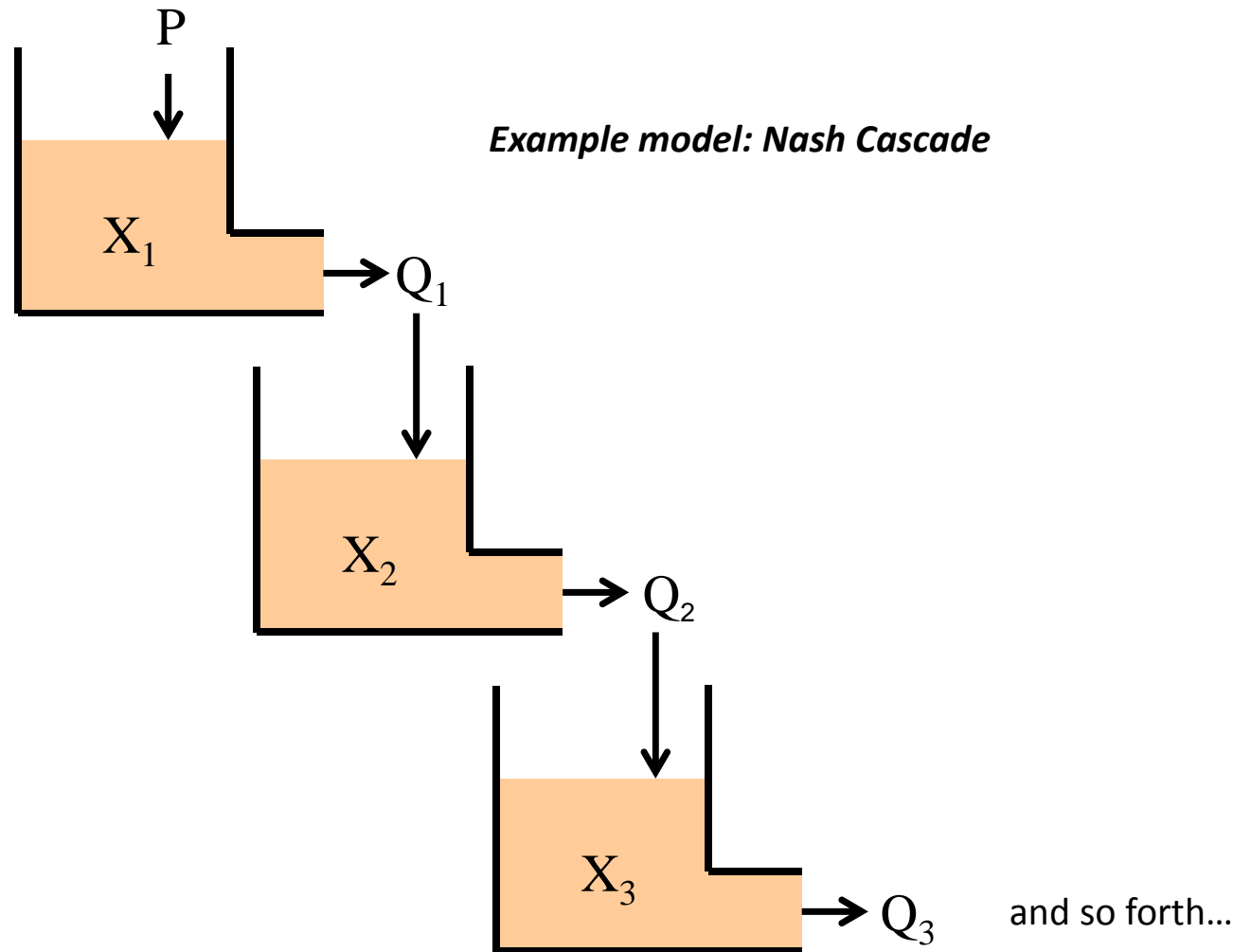
Which methods are available for Sensitivity Analysis?



Which methods are available for Sensitivity Analysis?

Graphical

- Provides representations of sensitivity in the form of graphs, charts or surfaces
- Typically not the full solution, rather in support of a more detailed analysis

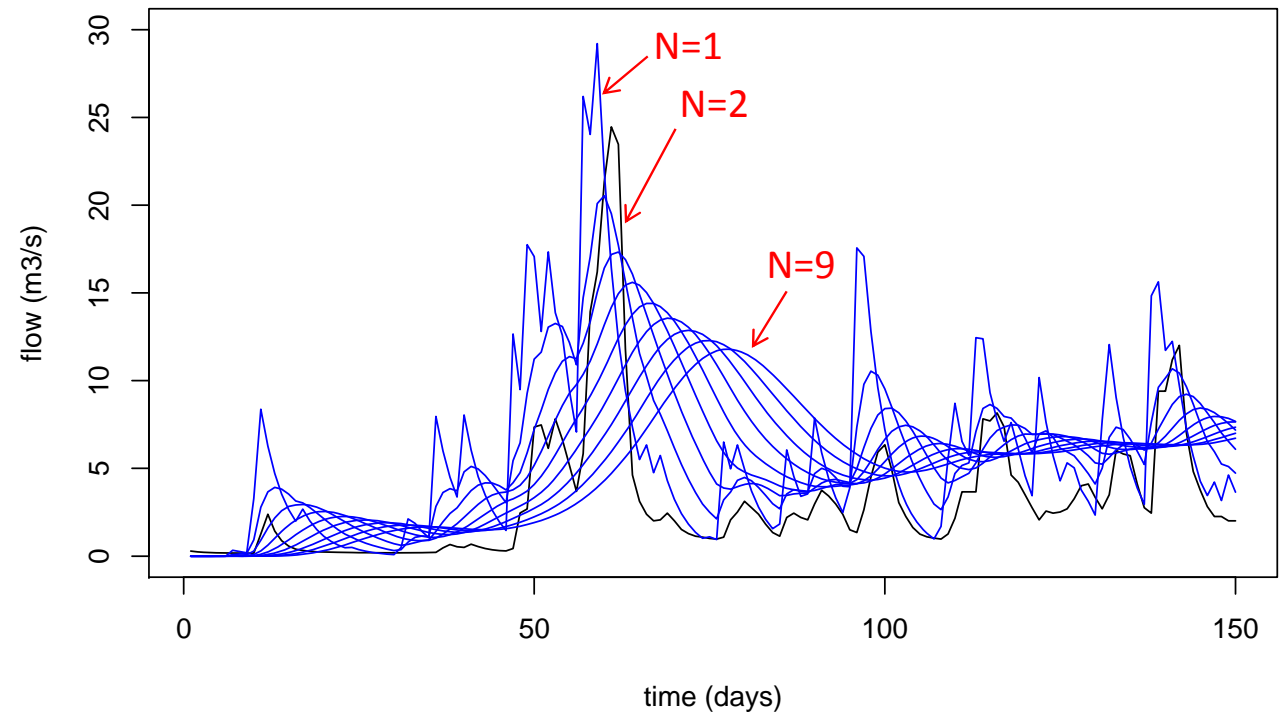


Which methods are available for Sensitivity Analysis?

Graphical

- Provides representations of sensitivity in the form of graphs, charts or surfaces
- Typically not the full solution, rather in support of a more detailed analysis

Effect of perturbation of N from 1 to 9

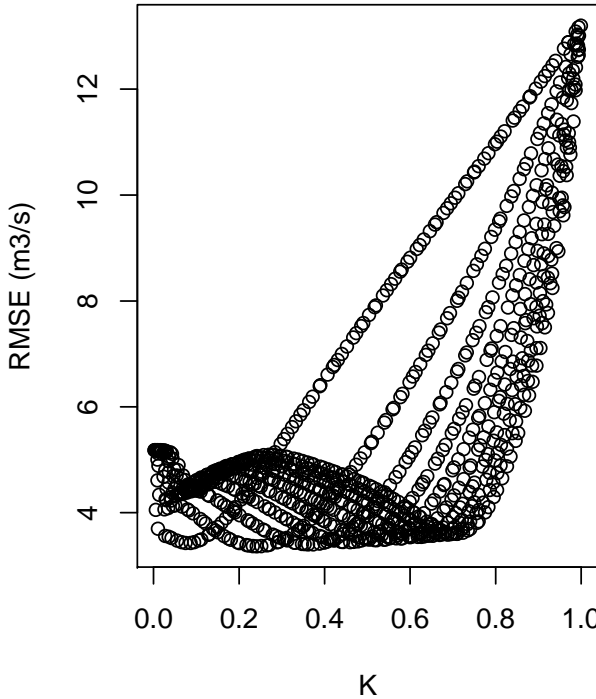
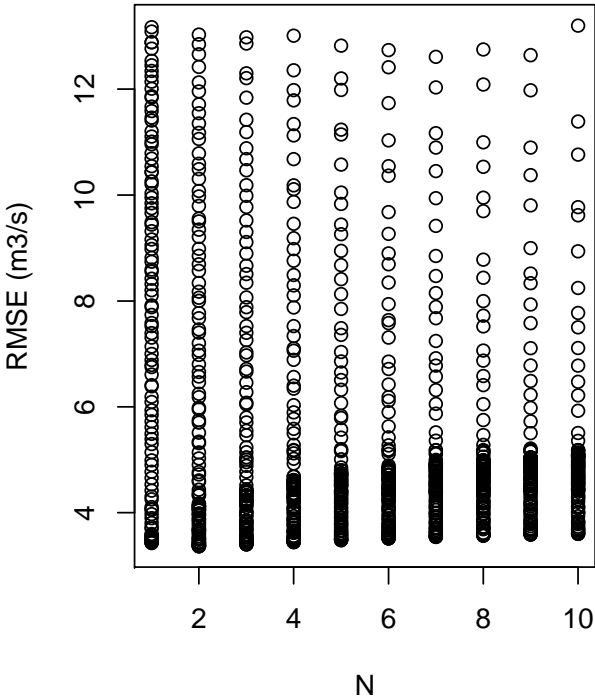


Which methods are available for Sensitivity Analysis?

Graphical

- Provides representations of sensitivity in the form of graphs, charts or surfaces
- Typically not the full solution, rather in support of a more detailed analysis

Scatter plots (factor vs. output-metric)



K : lag time in the reservoir

Which methods are available for Sensitivity Analysis?

Correlation/Regression

- **Correlation and Regression**
- Mathematically, they are identical.
- Conceptually, very different.

Correlation

- Co-variation
- Relationship or association
- No direction or causation is implied
- $Y \leftrightarrow X$ $X_1 \leftrightarrow X_2$



Regression

- Prediction of Y from X
- Implies, but does not prove, causation
- X (independent variable)
→ Y (dependent variable)

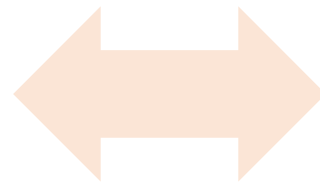
Which methods are available for Sensitivity Analysis?

Correlation/Regression

- **Correlation and Regression**
- Mathematically, they are identical.
- Conceptually, very different.

Correlation

- Co-variation
- Relationship or association
- No direction or causation is implied
- $Y \leftrightarrow X$ $X_1 \leftrightarrow X_2$



Regression

- Prediction of Y from X
- Implies, but does not prove, causation
- X (independent variable)
→ Y (dependent variable)

Which methods are available for Sensitivity Analysis?

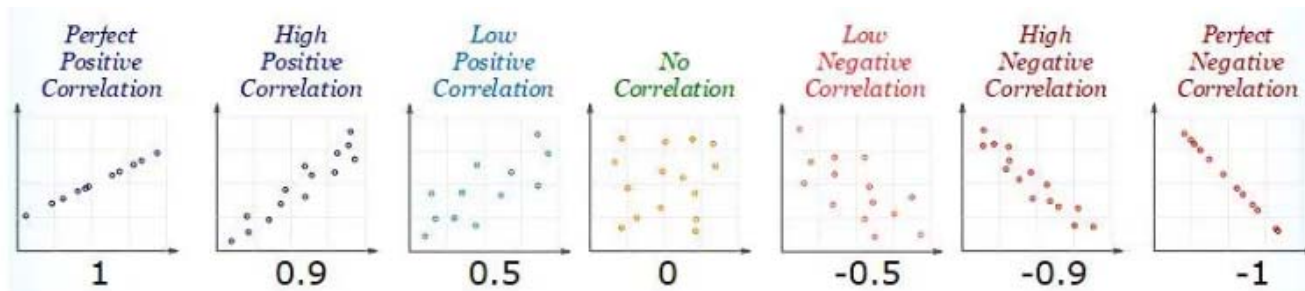
Correlation Coefficient

- The most common statistic in all of science
- Measures the **strength of the relation** (or “association”) between two variables
- Varies on a scale from -1 through 0 to +1

+1 implies a perfect positive association (e.g. income and education)

0 implies no association

-1 implies perfect negative association (e.g. price and quantity purchased)



- **Pearson Correlation Coefficient (PCC) & Spearman Rank Correlation Coefficient**

Which methods are available for Sensitivity Analysis?

Example 1: Pearson Correlation Coefficient

- A measure of the linear correlation between two variables x and y
- The correlation coefficient $r(x,y)$, between two variable x and y is given by,

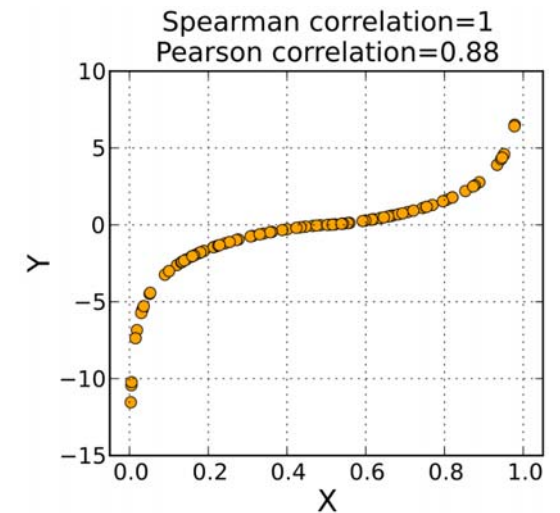
$$\begin{aligned} \bullet \quad r(x, y) &= \frac{\text{Cov}(x,y)}{\sqrt{\text{Var}(x)}\sqrt{\text{Var}(y)}} \text{ or } \frac{\text{Cov}(x,y)}{\sigma_x\sigma_y} \\ &= \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x\sigma_y} \end{aligned}$$

Example 2: Spearman's Rank Correlation Coefficient

- A measure of the rank correlation (Monotonic relationship, whether linear or not)
- Large coefficient when observations have a similar rank
- The correlation coefficient $\rho(x,y)$, between two variable x and y is given by,

$$\bullet \quad \rho_{rg_x,rg_y} = \frac{\text{Cov}(rg_x,rg_y)}{\sigma_{rg_x}\sigma_{rg_y}}$$

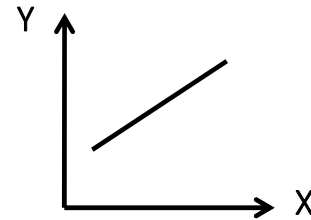
- rg_x and rg_y are the rank variables



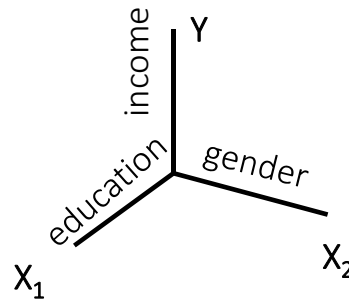
Which methods are available for Sensitivity Analysis?

Regression Coefficient

- *Simple regression : Between two variables*
 - One dependent variable (Y)
 - One independent variable (X)



- *Multiple regression : Between three or more variables*
 - One dependent variable (Y)
 - Two or more independent variables (X_1, X_2, \dots)



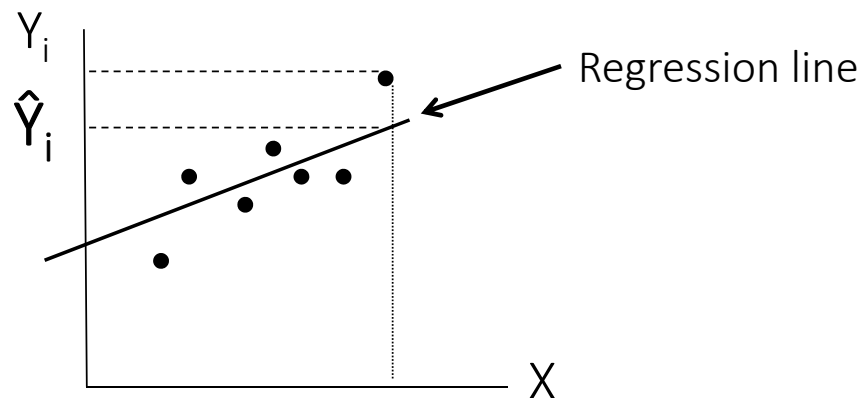
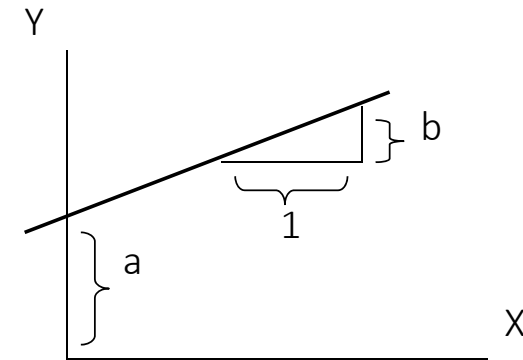
Which methods are available for Sensitivity Analysis?

Example 1: Simple Linear Regression Coefficient

- Concerned with “Predicting” one variable (Y) from another variable (X)

$$Y = a + bX + \varepsilon$$

- a is the intercept, the value of Y when $X=0$
- b is the regression coefficient or slope of the line
- ε is a residual (error), $Y_i - \hat{Y}_i = \text{Actual}(Y_i) - \text{Predicted}(\hat{Y}_i)$



Which methods are available for Sensitivity Analysis?

Example 1: Simple Linear Regression Coefficient

- Coefficient determination (r^2)
 - The coefficient measures the proportion of the variance in $Y \rightarrow$ which can be predicted by X

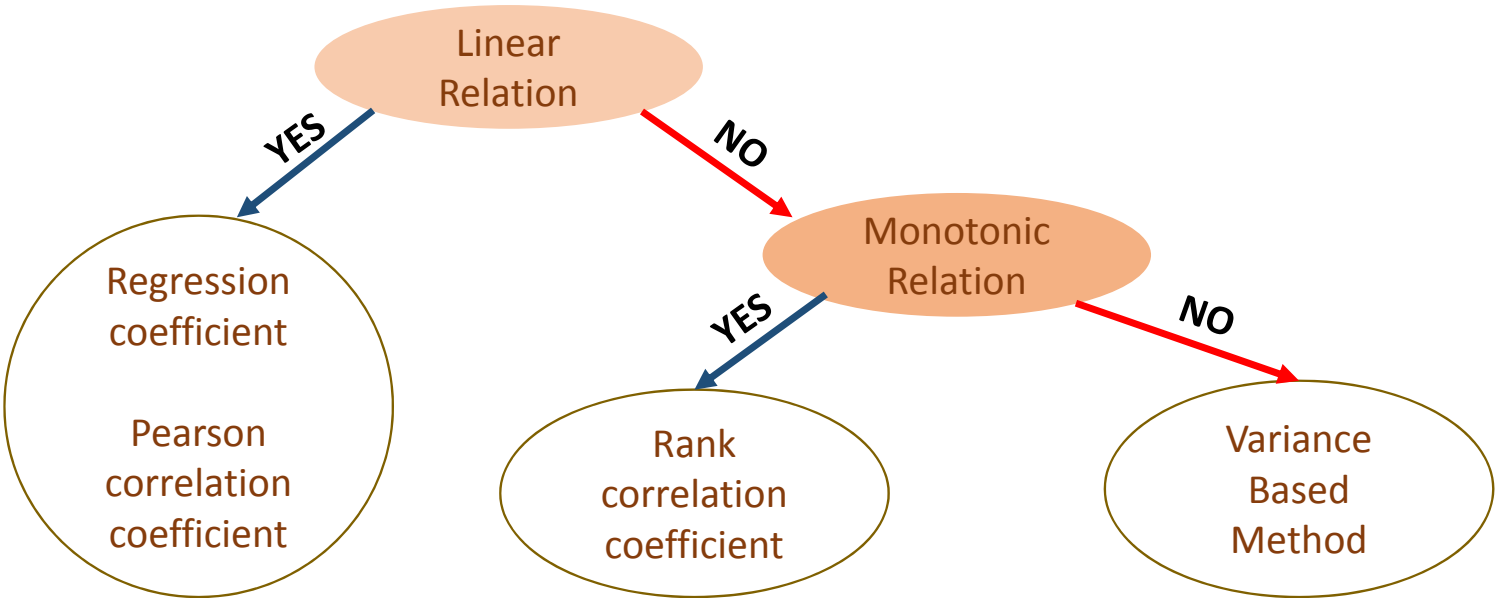
- $r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$ ← Treatment sum of squares (SST)
← Total SS

- $\frac{\sum(Y_i - \bar{Y})^2}{\text{Total SS}} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\text{SST}} + \frac{\sum(Y_i - \hat{Y}_i)^2}{\text{SS of residual error}}$ (ANOVA test)

- Treatment SS : Tells how much of the variation in the dependent variable model explained
- SS of residual error: Tells how much of the dependent variable's variation model did not explain

Which methods are available for Sensitivity Analysis?

Variance Based Method



- *Quantifies sensitivity by decomposing the variance of model outputs into factors related components*

Which methods are available for Sensitivity Analysis?

Example 1: Sobol' index

- The attribution of total output variance

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

- $D(f)$: total variance of the output metric f
- D_i : the first-order variance contribution of the i^{th} parameter
- D_{ij} : the second-order contribution of the interaction between parameters i and j
- $D_{12\dots p}$: all interactions higher than third-order, up to p total parameters

- First-Order Index

$$S_i = \frac{D_i}{D}$$

- Total-Order Index

$$S_{T_i} = 1 - \frac{D_{\sim i}}{D}$$

Which methods are available for Sensitivity Analysis?

Example 1: Sobol' index

- How to interpret?

First-order index

$$S_i = \frac{D_i}{D}$$

- The main effect
- Measures the contribution to the output variance from i^{th} factor alone
- The higher $S_i \rightarrow$ The higher influence of i^{th} factor
- $S_i = 0 \rightarrow$ No direct influence
- $\sum S_i < 1$
- What if $\sum S_i = 1$?

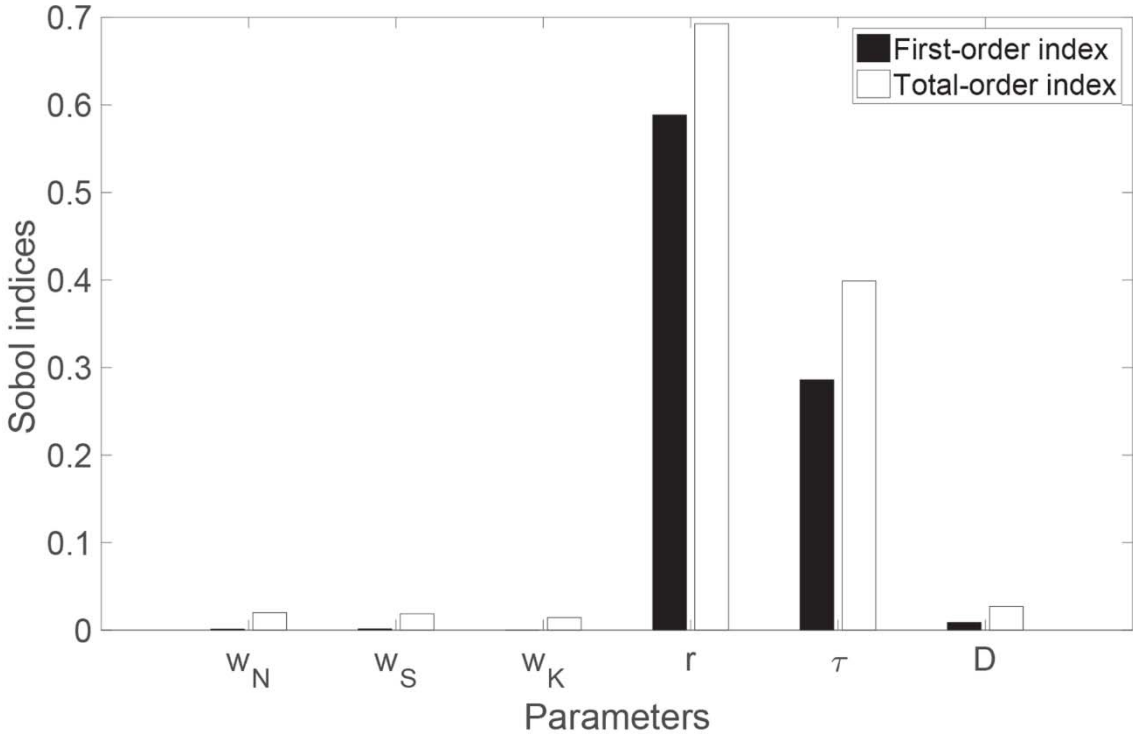
Total index

$$S_{T_i} = 1 - \frac{D_{\sim i}}{D}$$

- Measures the total contribution to the output variance from i^{th} factor alone + interactions with other factors
- $S_{T_i} > S_i$
- What if $S_{T_i} = S_i$?
- $\sum S_{T_i} \geq 1$

Which methods are available for Sensitivity Analysis?

Example 1: Sobol' index



M. Chen et al, Calibration and analysis of the uncertainty in downscaling global land use and land cover projections from GCAM using Demeter (v1.0.0), Geoscientific Model Development, vol.12, 1753-1764, 2019.

Which methods are available for Sensitivity Analysis?

Variance Based Method

Advantages

- *Extremely robust → working with any type of discontinuous mapping between input and output*
- *Computing main effects and total effects for each factor → giving a fairly instructive description of the system*

Disadvantages

- *Very high computational cost → impossible for computationally expensive models*
- *Required number of model evaluations = $N \times (2p + 1)$*

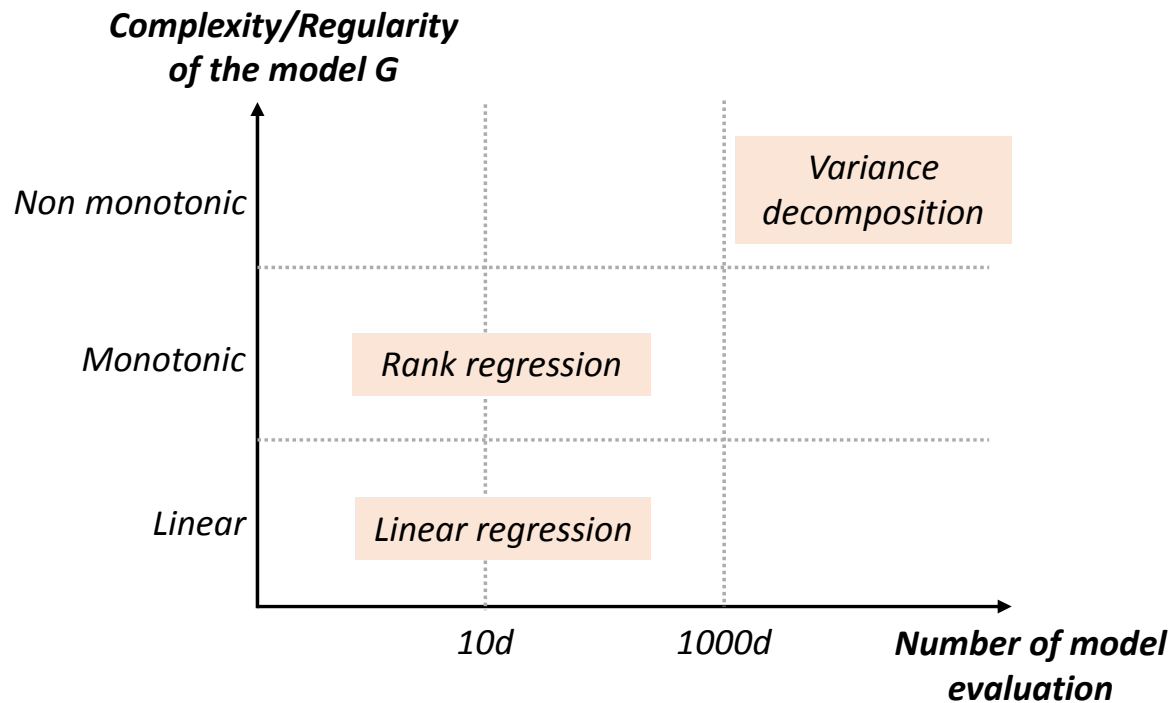
Which methods are available for Sensitivity Analysis?

Variance Based Method

Advantages

- Extremely robust → working with any type of discontinuous mapping between input and output
- Computing main effects and total effects for each factor → giving a fairly instructive description of the system

Disadvantages



(d : number of input parameters)

Which methods are available for Sensitivity Analysis?

Variance Based Method

Advantages

- *Extremely robust → working with any type of discontinuous mapping between input and output*
- *Computing main effects and total effects for each factor → giving a fairly instructive description of the system*

Disadvantages

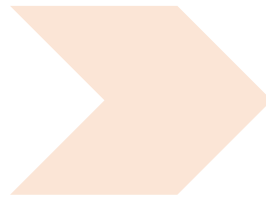
- *Very high computational cost → impossible for computationally expensive models*
- *Required number of model evaluations = $N \times (2p + 1)$*

→ Necessity of surrogate modelling!

What is surrogate modelling?

- Metamodels
- Surrogates
- Response surface models
- Approximation models
- Simulation models
- Data driven models
- Emulators
-

→ Depending on the field of research considered

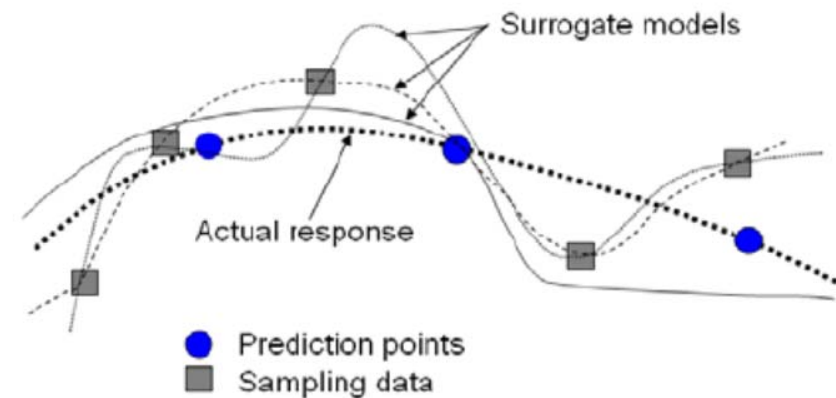


Surrogate model \tilde{M}



Approximation of the original computational model M

Builds from a limited set of runs of the original model M called the experimental design $X = \{x^{(i)}, i = 1, \dots, n\}$



A. Samed et al, Blade Optimization of a Transonic Compressor Using a Multiple Surrogate Model, Transactions of the Korean Society of Mechanical Engineers B, vol.32(4), 317-326, 2008.

What is surrogate modelling?

Assumes some regularity of the model M and some general functional shape

Name	Shape
Polynomial chaos expansions	$\tilde{M}(x) = \sum_{\alpha \in A} a_{\alpha} \Psi_{\alpha}(x)$
Low-rank tensor approximations	$\tilde{M}(x) = \sum_{l=1}^R b_l \left(\prod_{i=1}^M v_l^{(i)}(x_i) \right)$
Kriging (a.k.a. Gaussian processes)	$\tilde{M}(x) = \beta^T \cdot f(x) + Z(x, \omega)$
Support vector machines	$\tilde{M}(x) = \sum_{i=1}^m a_i K(x_i, x) + b$