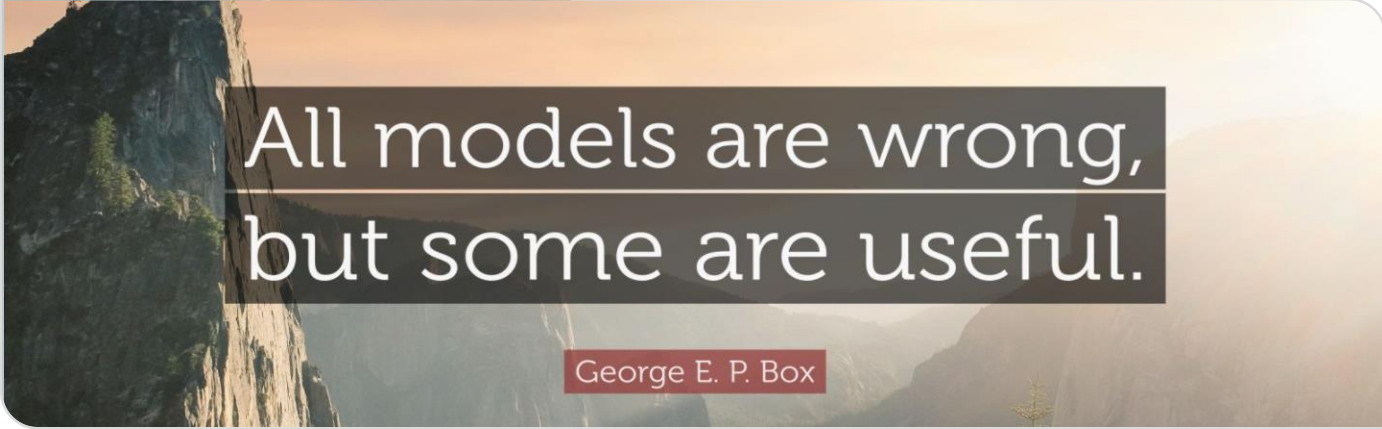


# Geodatenanalyse I: Sensitivitätsanalyse

Kathrin Menberg



All models are wrong,  
but some are useful.

George E. P. Box

# Stundenplan

Vorläufiger Stundenplan		
Datum	Thema	Dozent
20.10.2021	Einführung in die Programmierung mit <i>Python</i>	Gabriel Rau
25.10.2021	Univariate Statistik und statistisches Testen	Kathrin Menberg
01.11.2021	<i>Feiertag</i>	
08.11.2021	Umgang und Berechnung von Datensätzen	Gabriel Rau
15.11.2021	Bivariate und schließende Statistik	Kathrin Menberg
22.11.2021	Datenvisualisierung mit <u><i>matplotlib</i></u>	Gabriel Rau
29.11.2021	Multivariate Statistik	Kathrin Menberg
06.12.2021	Einführung in den Umgang mit Datensätzen	Gabriel Rau
13.12.2021	<i>ausgefallen</i>	Kathrin Menberg
20.12.2021	Weiterführender Umgang mit Datensätzen	Gabriel Rau
27.12.2021	<i>Weihnachtsferien</i>	
03.01.2022	<i>Weihnachtsferien</i>	
10.01.2022	Monte-Carlo Methoden	Kathrin Menberg
17.01.2022	Analyse und Visualisierung von Geodaten	Gabriel Rau
24.01.2022	Räumliche Interpolation	Kathrin Menberg
31.01.2022	Datenethik, Lizenzierung und Entwicklungstools	Gabriel Rau
07.02.2022	Regressionsanalyse	Kathrin Menberg

# Vorlesungsplan

Uhrzeit	Inhalt
10:00 – 10:20	<b>Grundlagen Sensitivitätsanalysen</b>
10:20 – 11:00	Übung
11:00 – 11:10	Diskussion und Reflexion
11:10 – 11:25	<u>Pause</u>
11:25 – 11:45	Fortgeschrittene Methoden
11:45 – 12:20	Übung
12:20 – 12:30	Diskussion und Reflexion

# Lernziele Teil 1: Grundlagen

Am Ende der Stunde werden die Teilnehmer:

- ▶ ... mit den Grundlagen der lokalen und globalen Sensitivitätsanalyse vertraut sein.
- ▶ ... einfache Methoden zur qualitativen Sensitivitätsanalyse in Python anwenden und die Ergebnisse graphisch darstellen können.

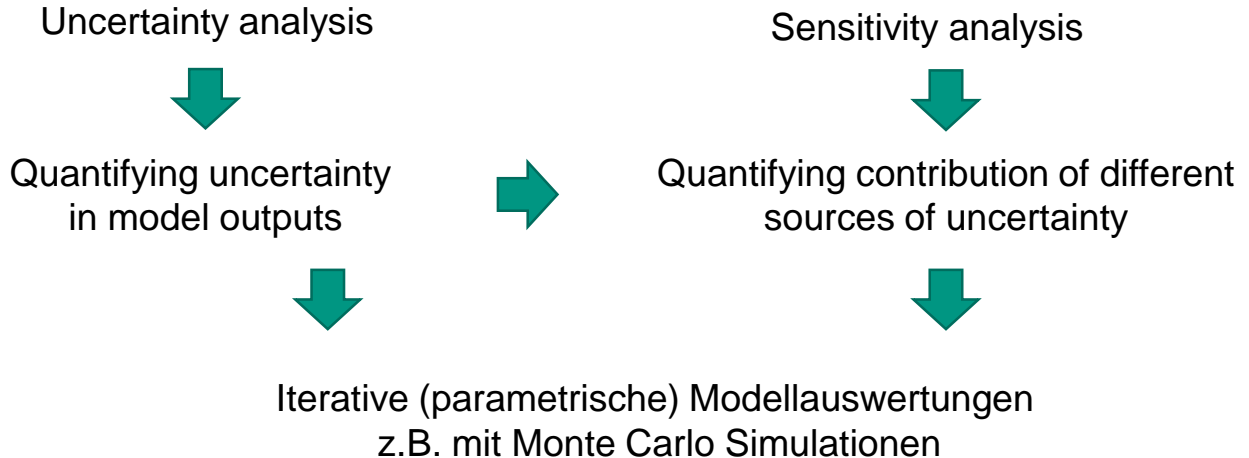
# Wozu Sensitivitätsanalyse

*The study of how uncertainty in the output of a model can be apportioned to different sources of uncertainty in the model input.*

*(Saltelli et al. 2004)*

# Was kann eine Sensitivitätsanalyse

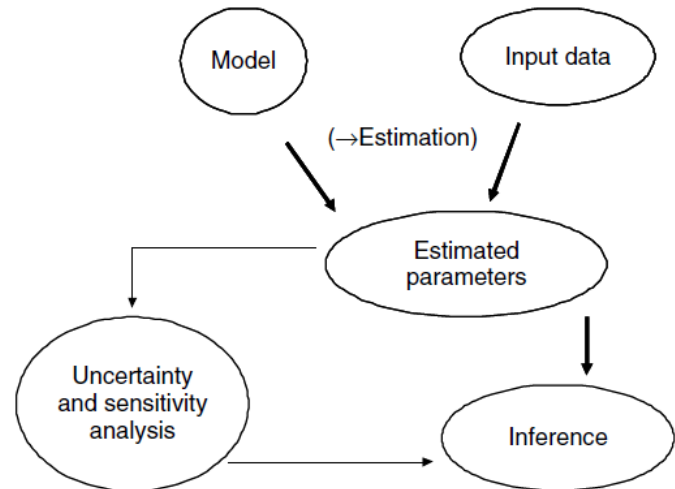
- ▶ Identifizierung wichtiger Parameter und Faktoren
- ▶ Kritische Regionen im Input Parameterraum aufdecken
  - ▶ Schwerpunkte für weitere Forschung, Messungen usw. identifizieren
- ▶ Technische Fehler in einem Modell aufdecken
  - ▶ z.B. bei unerwarteten Zusammenhängen
- ▶ Möglichkeiten zur Vereinfachung von Modellen aufzeigen
  - ▶ Parameter oder Faktoren ohne Einfluss fixieren
- ▶ Erweiterung von System-, bzw. Modellverständnis
- ▶ Robustheit der Modellergebnisse unter Unsicherheit zeigen
- ▶ ... und vieles mehr.



- ▶ Bewertung der Aussagekraft eines Modells, bzw. dessen Outputs
- ▶ „Alle Modelle sind falsch, aber manche sind nützlich“ (George E.P. Box)

# Input, Output und Model

- ▶ Input:
  - ▶ Daten, Parameter, ...
  - ▶ Annahmen, ...
  
- ▶ Model:
  - ▶ Diagnostisch oder prognostisch
  - ▶ Data-driven oder law-driven
  
- ▶ Output:
  - ▶ Grundlage für weitere Entscheidungen usw.



Saltelli et al. (2008)



# Workflow zur Sensitivitätsanalyse

1. Model-Inputs (Parameter) und deren Stichprobenbereich festlegen
2. Model-Inputs anhand einer Sampling-Strategie generieren
  - ▶ z.B. Zufallswerte, Latin-Hypercube, o.ä.
3. Das Model für die generierten Parameterwerte auswerten
  - ▶ Iterative Modellläufe z.B. mit Monte Carlo Simulation
4. Model-Inputs und Outputs in Bezug auf ihren Zusammenhang analysieren
  - ▶ Graphische Darstellung, Korrelationskoeffizienten, ...
  - ▶ Varianzbasierte Methoden, ...

# Typen von Sensitivitätsanalyse

- ▶ Lokale und globale Sensitivitätsanalyse
  - ▶ **Lokal:** Einfluss von Variation eines Inputs in einem beschränkten Bereich
    - ▶ z.B. durch Analyse von Gradienten oder partielle Ableitungen
  - ▶ **Global:** Betrachtung der Unsicherheit in allen Faktoren über ihre gesamten Wertebereiche
    - ▶ z.B. mit Hilfe von Monte Carlo Simulationen
- ▶ Qualitative und quantitative Sensitivitätsanalyse
  - ▶ **Qualitativ:** relative Einordnung der Parameter nach ihrem Einfluss
    - ▶ sog. Parameter Ranking Methoden
  - ▶ **Quantitative:** absolute Bestimmung des Einflusses aller Parameter
    - ▶ z.B. über die Analysen von Varianzen, lineare Regression, usw.

# Mathematische Grundlagen

## ► Ableitungen vs. Scatterplots

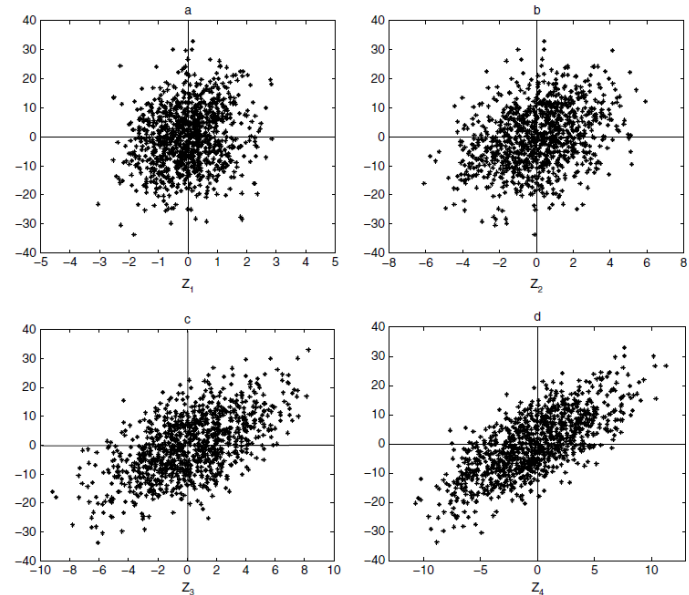
- $\frac{\partial Y_j}{\partial X_i} \rightarrow$  Sensitivität Output  $Y_j$   
gegenüber Input  $X_i$

## ► Beispiel Modell:

- $Y = \sum_{i=1}^r \Omega_i Z_i$
- $Z_i$  normalverteilt mit  $\mu_{Z_i} = 0$ ,  
und  $\sigma_{Z_1} < \sigma_{Z_2} < \dots < \sigma_{Z_n}$
- Kein Unterschied in Ableitungen

$$\frac{\partial Y}{\partial Z_i}$$

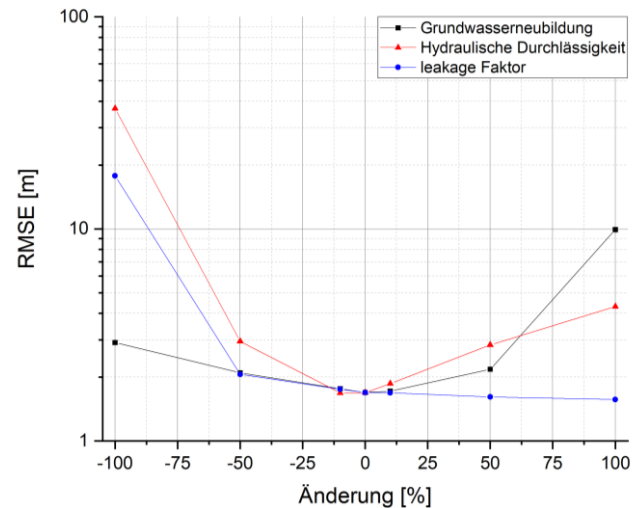
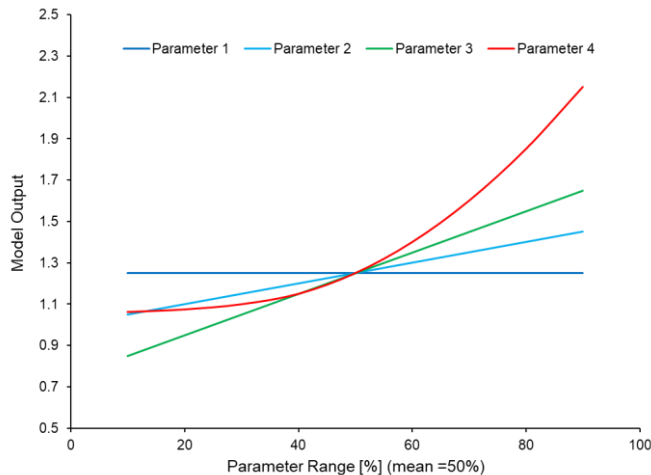
Saltelli et al. (2008)



Welcher Parameter hat den größten Einfluss?

# One-at-a-time (OAT) Vorgehen

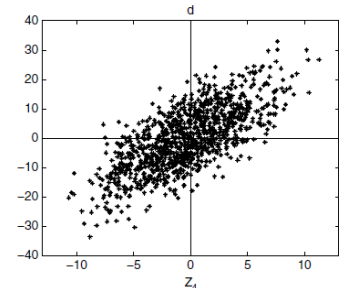
- ▶ Parameter nacheinander innerhalb ihrer Wertebereiche variieren
- ▶ Ergebnisse normieren und gegen den Output plotten



Ohmer et al. (2021)

# Contribution to Variance (CoV)

- ▶ Methode aus der Risikoanalyse und der Ökonomie
- ▶ Beitrag der Unsicherheit (oder  $\sigma_{x_i}^2$ ) eines Postens, zur Unsicherheit des Gesamtportfolios ( $\sigma_Y^2$ )
- ▶ Bestimmung der Korrelation zwischen  $x_i$  und  $Y$
- ▶ bzw. der Kovarianz
- ▶ Oft normiert in % von  $\sigma_Y^2$  angegeben
- ▶ Problematisch bei stark korrelierenden Input Parametern
- ▶ ... und komplexen, nicht-linearen Modellen



Saltelli et al. (2008)

# Contribution to Variance (CoV)

- ▶ Zwei Beispiele zur Visualisierung:
- ▶ Tornadoplot

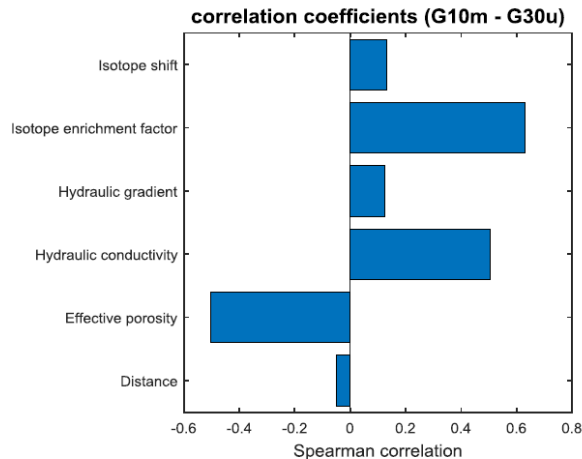
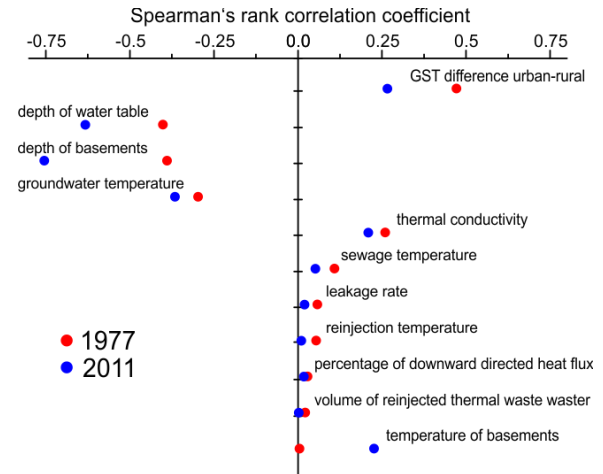


Figure 22: Statistics for sensitivity analysis of  $\lambda$  for G10m-G30u illustrated in tornado plots. SRC are given on the x-axis (MATLAB 2018b).



Menberg et al. (2013)

Würth, BSc. Thesis (2019)

# Mögliche Fallstricke und Schwierigkeiten

- ▶ zu viele Inputparameter für genaue Analyse
  - ▶ Screening-Methoden (nächste Stunde)
- ▶ alle wichtigen Parameter, bzw. Faktoren berücksichtigt?
- ▶ zu wenig Information um Wahrscheinlichkeitsverteilung für Parameter aufzustellen
- ▶ (Un)abhängigkeit der betrachteten Parameter
  - ▶ Spezielle Methoden für gruppierte Parameter
- ▶ Modellläufe dauern zu lange
  - ▶ Emulatoren (vereinfachte statistische Modelle) benutzen

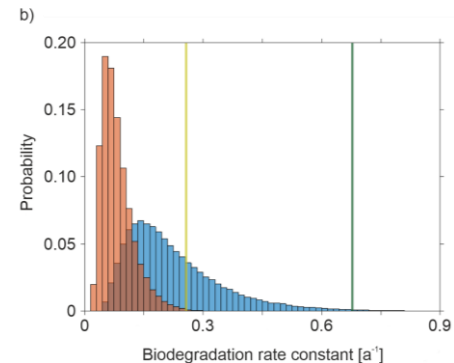
# Übung 7: Sensitivitätsanalyse 1

- ▶ Basierend auf MC Simulation aus Übung 6 grundlegende Methoden zur Sensitivitätsanalyse
  - ▶ Contribution-to-Variance
  - ▶ Visualisierung

$$\lambda = \frac{\Delta \delta^{13}C \cdot k_f \cdot i}{\varepsilon \cdot S \cdot n_e}$$

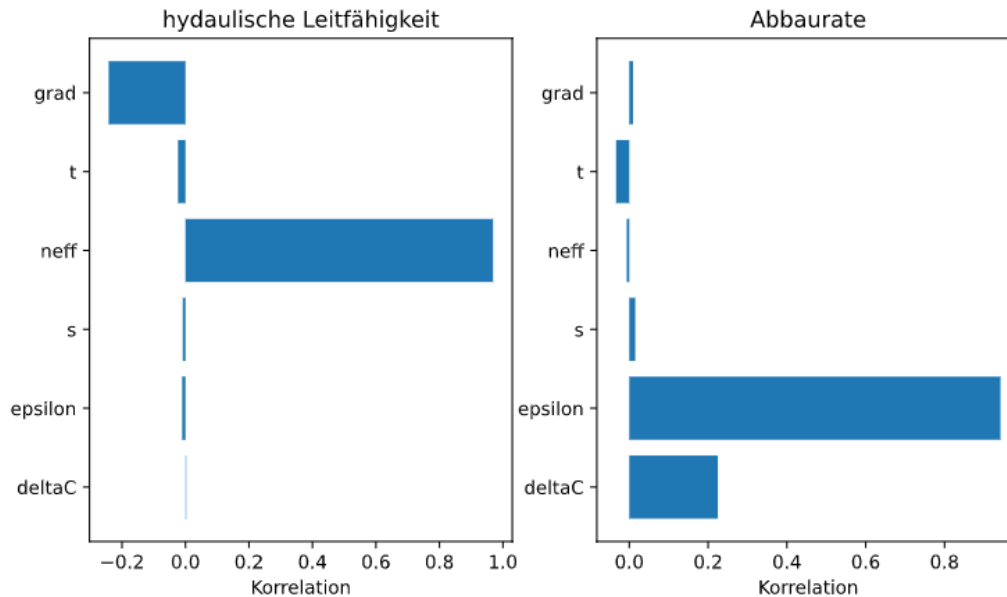


- ▶ Aufgaben in Jupyter Notebook:  
07\_Sensitivitätsanalyse\_1





## ► Contribution-to-Variance



# Pause

... bis 11:25 Uhr



# Vorlesungsplan

Uhrzeit	Inhalt
10:00 – 10:20	Grundlagen Sensitivitätsanalysen
10:20 – 11:00	Übung
11:00 – 11:10	Diskussion und Reflexion
11:10 – 11:25	<u>Pause</u>
11:25 – 11:45	<b>Fortgeschrittene Methoden</b>
11:45 – 12:20	Übung
12:20 – 12:30	Diskussion und Reflexion

## Lernziele Teil 2: Fortgeschrittene Methoden

Am Ende der Stunde werden die Teilnehmer:

- ▶ ... mit den mathematischen Grundlagen der Varianzdekomposition vertraut sein.
- ▶ ... verschiedene quantitative Methoden zur Sensitivitätsanalyse und deren Einsatzgebiete kennen.
- ▶ ... in Python Sobol Indizes bestimmen und interpretieren können.

# Varianzdekomposition

- Generisches Modell  $Y = f(X_1, X_2, \dots, X_k)$

$$Y = f_0 + \sum_{i=1}^d f_i(X_i) + \sum_{j<i}^d f_{ij}(X_i, X_j) + \dots$$

$$V(Y) = \sum_i V_i + \sum_i \sum_{j<i} V_{ij} + \dots$$

- Ableitung von verschiedenen Indizes zur Sensitivitätsanalyse für nicht-lineare Systeme, bzw. Modelle
- Sobol Indizes (nach Ilya Sobol):
  - Sobol Effekte erster Ordnung, Effekte höherer Ordnung, Totale Effekte

# Sobol Effekte erster Ordnung

## Bedingte Varianzen (1. Option)

- ▶ Generisches Modell  $Y = f(X_1, X_2, \dots, X_k)$
- ▶ Jedes  $X$  variiert innerhalb eines bestimmten Wertebereichs
  - ▶ Quantifizieren über Varianz  $V_{X_i}$
- ▶ Sensitivität definieren als Effekt den das Fixieren von  $X_i$  auf die Varianz in  $Y$  hat
  - ▶  $E_{X \sim i}(Y|X_i)$
- ▶ Für gesamten Wertebereich von  $X_i$  und normiert auf die Gesamtvarianz

$$S_i = \frac{V_{X_i}(E_{X \sim i}(Y|X_i))}{V(Y)}$$

# Sobol Effekte erster Ordnung

- ▶ Sampling Strategie über Monte Carlo Simulation mit Zufallswerten für Modell-Inputs
- ▶ Referenzmatrix und eine Vergleichs-Matrix für jedes  $X_i$

$$X^{ref} = \begin{bmatrix} X_{11}^{ref} & \dots & X_{1i}^{ref} & \dots & X_{1k}^{ref} \\ X_{21}^{ref} & \dots & X_{2i}^{ref} & \dots & X_{2k}^{ref} \\ \vdots & & \dots & & \vdots \\ X_{N1}^{ref} & \dots & X_{Ni}^{ref} & \dots & X_{Nk}^{ref} \end{bmatrix}$$

$$X^{(i)} = \begin{bmatrix} X_{11}^{(i)} & \dots & X_{1i}^{ref} & \dots & X_{1k}^{(i)} \\ X_{21}^{(i)} & \dots & X_{2i}^{ref} & \dots & X_{2k}^{(i)} \\ \vdots & & \dots & & \vdots \\ X_{N1}^{(i)} & \dots & X_{Ni}^{ref} & \dots & X_{Nk}^{(i)} \end{bmatrix}$$

→ die Werte von  $X_i$  bleiben gleich,  
alle anderen werden variiert

- ▶ Benötigte Modelldurchläufe:  $(1 + \text{Anzahl Paratemeter}) * N$

# Sobol Totale Effekte

## Bedingte Varianzen (2. Option)

- ▶ Generisches Modell  $Y = f(X_1, X_2, \dots, X_n)$
- ▶ Jedes  $X$  variiert innerhalb eines bestimmten Wertebereichs
  - ▶ Quantifizieren über Varianz  $V_{X_i}$
- ▶ Sensitivität als Effekt den Variieren von  $X_i$  auf die Varianz in  $Y$  hat, wenn alle  $X_{\sim i}$  fixiert sind
  - ▶  $V_{X_i}(Y|X_{\sim i})$
- ▶ Für gesamten Wertebereich von  $X_i$  und normiert auf die Gesamtvarianz

$$S_{T_i} = \frac{E_{X_{\sim i}}(V_{X_i}(Y|X_{\sim i}))}{V(Y)} = S_i + S_{ij} + S_{ij\dots k}$$



# Sobol totale Effekte

- ▶ Sampling Strategie über Monte Carlo Simulation mit Zufallswerten für Modell-Inputs
- ▶ Referenzmatrix und eine Vergleichs-Matrix für jedes  $X_i$

$$X^{ref} = \begin{bmatrix} X_{11}^{ref} & \dots & X_{1i}^{ref} & \dots & X_{1k}^{ref} \\ X_{21}^{ref} & \dots & X_{2i}^{ref} & \dots & X_{2k}^{ref} \\ \vdots & & \dots & & \vdots \\ X_{N1}^{ref} & \dots & X_{Ni}^{ref} & \dots & X_{Nk}^{ref} \end{bmatrix}$$

$$X^{(i)} = \begin{bmatrix} X_{11}^{ref} & \dots & X_{1i}^i & \dots & X_{1k}^{ref} \\ X_{21}^{ref} & \dots & X_{2i}^i & \dots & X_{2k}^{ref} \\ \vdots & & \dots & & \vdots \\ X_{N1}^{ref} & \dots & X_{Ni}^i & \dots & X_{Nk}^{ref} \end{bmatrix}$$

→ Nur die Werte von  $X_i$  werden variiert, alle anderen bleiben gleich

- ▶ Benötigte Modelldurchläufe:  $(1 + \text{Anzahl Parameter}) * N$

# Effekte erster Ordnung vs. Totale Effekte

- ▶ Definition der „Wichtigkeit“ von Parametern
- ▶ Ziel: Reduzierung von Unsicherheiten
  - ▶ Parameter der bei Fixierung Varianz in  $Y$  verringert
  - ▶ Effekt erster Ordnung
- ▶ Ziel: Modellvereinfachung
  - ▶ Parameter, der bei Variation möglichst viel der Varianz in  $Y$  erhält
  - ▶ Totale Effekte
- ▶ für additive Modellen ergibt die Summe aller Effekte erster Ordnung 1

# Fazit Sensitivitätsanalyse

- ▶ Methoden mit unterschiedlichem Aufwand für verschiedene Zwecke
- ▶ Sobol Indizes: **die** Methode für quantitative und globale Sensitivitätsanalyse
- ▶ Rechenkosten für Effekte erster Ordnung und totale Effekte:  
 $(1 + \text{Anzahl Parameter}) * N * 2$ 
  - ▶ ... schnell auf mehrere 10.000 iterative Simulationen
- ▶ Parameter Reihenfolge mit Screening Methoden (z.B. Morris Method) verlässlich und effizient zu bestimmen

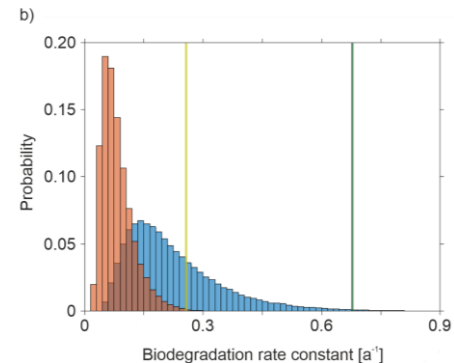
# Übung 8: Sensitivitätsanalyse II

- ▶ Basierend auf MC Simulation aus 6 fortgeschrittene Methoden zur Sensitivitätsanalyse
  - ▶ Sobol Indizes
  - ▶ Visualisierung

$$\lambda = \frac{\Delta \delta^{13}C \cdot k_f \cdot i}{\varepsilon \cdot S \cdot n_e}$$



- ▶ Aufgaben in Jupyter Notebook:  
08\_Sensitivitätsanalyse\_2



# Literatur

- ▶ Saltelli et al. (2008): Global Sensitivity Analysis. The Primer, John Wiley & Sons.
- ▶ Menberg et al. (2016): Sensitivity analysis methods for building energy models: Comparing computational costs and extractable information, Energy and Buildings 133, 433-445.
- ▶ Würth et al. (2021): Quantifying biodegradation rate constants of o-xylene by combining compound-specific isotope analysis and groundwater dating. Journal of Contaminant Hydrology, 238, 103757

