Global Sensitivity Analysis of Complex Systems
implications for natural resources

Jef Caers
Geological Sciences
Stanford University, USA
Analysis of (Subsurface) Systems

- The model: description of the world as we see it as defined by choices of model variables
- The prior model: the stated probability distributions on the model variables jointly: a multi-variate distribution
- The data forward model and the prediction forward model
Characteristics

- High-dimensional
- Non-linear
- Spatially distributed
- Uncertain
- Costly
Simple Example

Unsaturated groundwater flow model

- Infiltration basin
- Well
- Contaminated area between 5 m and 6 m

Bloc <- list(x = c(0, 200), y = c(0, 300), z = c(0, 10), nx = 100, ny = 150, nz = 10)

Vertical section at y = 150

Infiltration basin

Flow velocity along x
## Simple Example

<table>
<thead>
<tr>
<th>Parameter code</th>
<th>Description</th>
<th>Variable Type</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hydraulic Conductivity Representation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kmean</td>
<td>Mean value of hydraulic conductivity $K$ (m/s)</td>
<td>Continuous</td>
<td>$U(7e-4, 10^{-3})$</td>
</tr>
<tr>
<td>Ksd</td>
<td>Standard deviation of log($K$) (m/s)</td>
<td>Continuous</td>
<td>$U(0.05, 0.3)$</td>
</tr>
<tr>
<td>KCov</td>
<td>Type of covariance model for simulation of $K$</td>
<td>Discrete</td>
<td>Gaussian or Spherical</td>
</tr>
<tr>
<td>Kangle</td>
<td>Horizontal anisotropy angle for $K$ (degree)</td>
<td>Continuous</td>
<td>$U(110, 150)$</td>
</tr>
<tr>
<td>Krangle</td>
<td>Correlation length along the principle direction for $K$ (m)</td>
<td>Continuous</td>
<td>$U(10, 100)$</td>
</tr>
<tr>
<td>Kanixy_ratio</td>
<td>Anisotropy - horizontal stretching ratio</td>
<td>Continuous</td>
<td>$U(1/20, 1/2)$</td>
</tr>
<tr>
<td>Kaniz_ratio</td>
<td>Anisotropy - vertical stretching ratio</td>
<td>Continuous</td>
<td>$U(15, 30)$</td>
</tr>
<tr>
<td>Knugget</td>
<td>Nugget for $K$ (m/s)</td>
<td>Continuous</td>
<td>$U(0, 0.1)$</td>
</tr>
<tr>
<td><strong>Boundary Conditions Representation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hsd</td>
<td>STD of the Matérn covariance model for simulation of boundary conditions</td>
<td>Continuous</td>
<td>$U(0.01, 0.1)$</td>
</tr>
<tr>
<td>Hrange</td>
<td>Correlation length of the Matérn covariance model for simulation of boundary condition</td>
<td>Continuous</td>
<td>$U(20,40)$</td>
</tr>
<tr>
<td>Hnu</td>
<td>Smoothness of the Matérn covariance model for simulation of boundary conditions</td>
<td>Continuous</td>
<td>$U(1.5, 3.5)$</td>
</tr>
<tr>
<td>HrivGrad</td>
<td>Gradient of the river</td>
<td>Continuous</td>
<td>$N(-0.0015, 0.0001)$</td>
</tr>
<tr>
<td><strong>Measurement error</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HrivRef</td>
<td>River hydraulic head (meters)</td>
<td>Continuous</td>
<td>$N(7, 0.05)$</td>
</tr>
<tr>
<td>Hnugget</td>
<td>measurement error groundwater hydraulic heads (meters)</td>
<td>Continuous</td>
<td>$U(0.02, 0.1)$</td>
</tr>
</tbody>
</table>
Response of interest

Graphs showing concentration over time for Gaussian and Spherical models. The left graph displays multiple curves, while the right graph shows a histogram for the arrival time of models with a pollutant.
Local vs Global

- **Local**: Assesses the effect for a single (deterministic) set of input parameters

- **Global**: Assesses the effect of joint changes in input parameters
**Local SA**

OAT sensitivities on pollutant arrival time*

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Arrival Time (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kmean</td>
<td>20</td>
</tr>
<tr>
<td>Ksd</td>
<td>22</td>
</tr>
<tr>
<td>HrivGrad</td>
<td>24</td>
</tr>
<tr>
<td>Krange</td>
<td>26</td>
</tr>
<tr>
<td>Hrange</td>
<td>28</td>
</tr>
</tbody>
</table>

- Change the spatial distribution

* Covariance is fixed to Gaussian, no spatial uncertainty
Variance-based methods

- Evaluates the part of the total variance of the response that can be attributed to input parameters

- In Sobol’, two measures can be obtained for each parameter:
  - **First order index**: contribution (without interaction) of a parameter to the response variance
  - **Total effect index**: total contribution (including interaction) of a parameter to the response variance
### Example case: reservoir study in Libya

<table>
<thead>
<tr>
<th>Number</th>
<th>Parameters</th>
<th>Abbreviation</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Oil-water contact</td>
<td>owc</td>
<td>U[-1076, -1061]</td>
</tr>
<tr>
<td>2</td>
<td>Transmissibility multiplier of fault 1</td>
<td>mflt1</td>
<td>U[0, 1]</td>
</tr>
<tr>
<td>3</td>
<td>Transmissibility multiplier of fault 2</td>
<td>mflt2</td>
<td>U[0, 1]</td>
</tr>
<tr>
<td>4</td>
<td>Transmissibility multiplier of fault 3</td>
<td>mflt3</td>
<td>U[0, 1]</td>
</tr>
<tr>
<td>5</td>
<td>Transmissibility multiplier of fault 4</td>
<td>mflt4</td>
<td>U[0, 1]</td>
</tr>
<tr>
<td>6</td>
<td>Residual oil saturation</td>
<td>sor</td>
<td>N[0.2, 0.05²]</td>
</tr>
<tr>
<td>7</td>
<td>Connate water saturation</td>
<td>swc</td>
<td>N[0.2, 0.05²]</td>
</tr>
<tr>
<td>8</td>
<td>Oil viscosity</td>
<td>oilvis</td>
<td>N[10,²]</td>
</tr>
<tr>
<td>9</td>
<td>Corey exponent of oil</td>
<td>oilexp</td>
<td>N[3,0.25²]</td>
</tr>
<tr>
<td>10</td>
<td>Corey exponent of water</td>
<td>watexp</td>
<td>N[2,0.1²]</td>
</tr>
<tr>
<td>11</td>
<td>Scenario</td>
<td>scen</td>
<td>Channel, bar, Channel with bar</td>
</tr>
<tr>
<td>12</td>
<td>Size</td>
<td>size</td>
<td>U[100, 300]</td>
</tr>
<tr>
<td>13</td>
<td>Proportion</td>
<td>prop</td>
<td>U[30, 50]</td>
</tr>
<tr>
<td>14</td>
<td>Ratio of vertical and horizontal permeability</td>
<td>kvkh</td>
<td>[0.001, 0.01, 0.1]</td>
</tr>
<tr>
<td>15</td>
<td>Maximum time step</td>
<td>dtmax</td>
<td>[91.25, 182.5, 365]</td>
</tr>
</tbody>
</table>
Application of Sobol’ to “Libyan” case

- 12 uncertain parameters ➔ a total of 14,000 simulations
- Response: total field water production at a given time
  ➔ for time-varying response, indices must be re-computed for each time step

Sobol’ sensitivity indices as a function of time
Distance-based Sensitivity analysis (DGSA)

Sampling of input parameters:
Structure, Rock, Fluid

Measure of sensitivity:
difference between the frequency distributions of input parameters per each class

Flow model

Dim. Reduction classification

CDF

Distance-based Sensitivity analysis (DGSA)
Bootstrap Procedure to assess sensitivity

L1-norm values when $x$ is not influential on the response
DGSA – Application to the DNAPL example

Note: entire concentration curve was used, as well as spatial uncertainty on the hydraulic conductivity field and the boundary conditions
Regression trees
Building Regression trees

Split Quality:

\[ Q_{5.5, X1}^{R_m} = G(R_m) - \{G(R_{mleft}) + G(R_{mright})\} \]

Where:

\[ G(R_m) = \sum_{y_i \in R_m} (y_i - \mu_m)^2 \]

Different outputs = different cost function

Splitting can be performed on any type of input (scalar, category, function)

Elements of Statistical Learning, Hastie 10th ed., Tibshirani, Freidman
Variable Importance

On every split we compute and save the quality of the best split on each predictor

\[ S_{m,p} = \frac{1}{N_m} \max(Q_{s,p}^{R_m}, \forall s) \]

We then compute variable importance index by summing over all regions/splits

\[ S_p = \sum_{m=1}^{T} S_{m,p} \]

Where \( T \) is the number of splits in a tree

This sensitivity index works with any cost function
Building regression trees with functional output?

Scalar Cost Function

\[ G(R_m) = \sum_{y_i \in R_m} (y_i - \mu_m)^2 \]

Functional Cost Function

\[ G(R_m) = \sum_{q_i(t) \in R_m} \|q(t) - \mu(t)\|^2 \]

Split Quality:

\[ Q_{s,p} = G(R_m) - \{G(R_{mL}) + G(R_{mR})\} \]

Forecasts:

\[ \mu_m(t) \]

Sensitivity analysis is the same as before
Building regression trees with distances?

$$Q_{s,p} = \sum_{i \in R_m} d_{i,medoid} - \left\{ \sum_{i \in R_m} d_{i,\text{Lmedoid}} + \sum_{i \in R_m} d_{i,\text{Rmedoid}} \right\}$$

Similar to DGSA: Important parameter splits responses into discrete categories
Building regression trees with multiple functional outputs?

\[
\begin{bmatrix}
0 & d_{12} & d_{13} & \cdots & d_{1n} \\
d_{21} & 0 & d_{23} & \cdots & d_{2n} \\
d_{31} & d_{32} & 0 & \cdots & d_{3n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
d_{n1} & d_{n2} & d_{n3} & \cdots & 0
\end{bmatrix} = D_{oil}^* + D_{gas}^*
\]

\[
Q_{s,p} = \sum_{i \in R_m} d_{i,medoid} - \left\{ \sum_{i \in R_m} d_{i,Lmedoid} + \sum_{i \in R_m} d_{i,Rmedoid} \right\}
\]

Sensitivity analysis is the same as before, only now it describes parameters effect on both outputs.

Prediction: local mean functions of each output.
Applications to Shales

- 188 Horizontal wells
- Multi-Fracture completion
- 26 Covariates
  - 2 Geographical (X & Y)
  - 1 PVT: API Gravity
  - 14 Fracturing
  - 7 Petrophysical
  - 2 Other

Focus is on oil rates, for now
Characteristics

- Multi-variate functional output
- Spatial variation
- Input: continuous, categorical, functional
Dimension Reduction
Functional PCA

What do fPC’s describe?

PCA function 1 (Percentage of variability 89.7)
PCA function 2 (Percentage of variability 7.3)
Functional PCA

Principal component scores:

\[ \alpha_i^1 = \int \xi_1(t) \cdot Q_i(t) \, dt \]
\[ \alpha_i^2 = \int \xi_2(t) \cdot Q_i(t) \, dt \]
CDFs for all

Multivariate Functional Regression Tree
Multivariate Functional Regression Tree

Variable importance plot

- TotalPropellant
- TotalFluid.bbl
- StimLength
- StagesPumped
- AmtStickwater
- AmtAcid
- GeoX_Rel
- GeoY_Rel
- AmtResin
- AmtCrosslink
- AmtLinear
- GeoAPIGrav
- ProdVert300
- Geo3DSpacing
- ProdVert200
- CleanFluidTotal
- PetroSwt
- PetroCarb
- PetroVClay
- PetroVQtz
- PetroPyr
- PetroPor
- PetroTOC
- No_Screenouts
- Amt100mesh
## Summary

<table>
<thead>
<tr>
<th></th>
<th>OAT</th>
<th>Morris Method</th>
<th>Regression</th>
<th>Sobol’</th>
<th>DSGA</th>
<th>Tree-based SA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost</strong></td>
<td>Low ($Np+1$)</td>
<td>Low $L(Np+1)$</td>
<td>Low $L(Np+1)$</td>
<td>High Quasi-random, LHS $m(Np+2)$</td>
<td>Moderate LHS $m$</td>
<td>Moderate LHS $m$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Model Assumption</strong></td>
<td>Linear</td>
<td>Model free</td>
<td>Depends on regression model</td>
<td>Model free</td>
<td>Model free</td>
<td>Model free</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sensitivity measure</strong></td>
<td>Subjective</td>
<td>Subjective</td>
<td>Objective</td>
<td>Subjective</td>
<td>Objective</td>
<td>Subjective</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Interactions</strong></td>
<td>No</td>
<td>Yes, qualitative</td>
<td>Depends on regression model – symmetric</td>
<td>Yes</td>
<td>Yes, asymmetric</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Discrete parameter</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stochasticity</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes, With proxy only</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Input distribution</strong></td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>High-dimensional response</strong></td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Grujic, O., Silva, C.D. & Caers, J. Functional approach to data mining, forecasting, and uncertainty quantification in unconventional reservoirs, SPE-174849-MS, SPE-ATCE 2015


Code available on http://www.github.com/SCRFpublic
Video Material

Search: tinyURL.com/JefCaersYouTube

Decomposing The Problem

Use of a simplified version of the DNAPL test case for illustration

Data Science for shales, session 1 FPCA & Dice

Sensitivity Analysis