Enhanced Metamodeling and Sensitivity Analysis for Complex Models Using Tree-PCE

PhD Student: BEN SAID Faten[1,2]

Directors: REYGNER Julien[1], EL KADI ABDERREZZAK Kamal[2,3]

Supervisors: AURELIEN Alfonsi[1], GOEURY Cedric[2,3], ZAOUI Fabrice[2], DUTFOY Anne[2]

[1]CERMICS, [2]EDF R&D, [3]LHSV

14 June 2024

Table of contents

Context

- **Challenges in Hydro-Sediment Modeling:**
  - Accurate prediction of sediment transport and deposition.
  - Complex interactions between water flow and sediment movement.
  - Different sources of uncertainty (data, measures, empirical laws,..).
- **Impact of Uncertainty:**
  - Uncertainty affects model predictions and decision-making.
  - Understanding and quantifying uncertainty is crucial for model validation and improvement.
- **Goal:**
  - Perform sensitivity analysis to identify the most influential parameters.

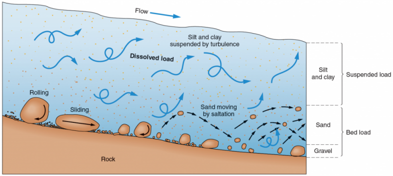


Figure: Sediment Transport

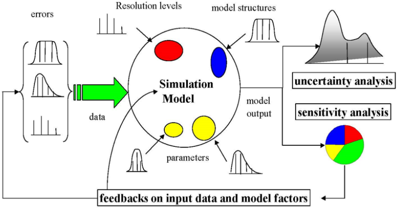


Figure: Uncertainty and Sensitivity Analysis

Sobol' indices [1]

Sobol' indices.

Let $X \in \mathbb{R}^d$ be a random vector, $G \in \mathbb{L}^2(P_X)$, $A \in \mathcal{P}_d$,

- The closed Sobol' index associated to A is defined as

$$S_A^{clos} = \frac{VarE[G(X)|X_A]}{VarG(X)}$$

  This index is also called the first-order Sobol' index associated to the input vector X_A.

- The total Sobol' index associated to X_A is defined as

$$S_A^T = 1 - S_{\overline{A}}^{clos}$$

Estimation of Sobol' indices by Polynomial Chaos method (PCE) [2]

PCE method.

Let $Y = G(X)$ with $G : X \in \mathcal{D}_X \subset \mathbb{R}^d \longmapsto Y = G(X) \in \mathbb{R}$.
Assume that $Y \in \mathbb{L}^2(P_X)$: Hilbert space, which allows writing :

$$Y = \sum_{|\alpha|=0}^{P} y_\alpha \Phi_\alpha(X) = \sum_{|\alpha|=0}^{P} y_\alpha \prod_{i=1}^{d} \phi_{\alpha_i}^{(i)}(X_i)$$

with

- $\{\Phi_\alpha\}_{|\alpha|=0}^{P}$ is a multivariate basis of the Hilbert space orthonormal with respect to the measure of X. such that $\Phi_\alpha = \prod_{i=1}^{d} \phi_{\alpha_i}^{(i)}(x_i)$ with $\{\phi_j^{(i)}\}_j$ is an univariate orthonormal basis with to the measure of X_i.
- $\{y_\alpha\}_{|\alpha|=0}^{P}$ are coefficients (the coordinates of Y in the basis $\{\Phi_\alpha\}_{|\alpha|=0}^{P}$).

Sobol' indices from PCE method

Let A $\subset \{1, \ldots, d\}$, We have

$$S_i = \frac{\sum_{\alpha \in J_i} y_\alpha^2}{\sum_{\alpha \in A, \alpha \neq 0} y_\alpha^2} \quad where \quad J_i = \{\alpha \in A, \alpha_i > 0, \alpha_{j \neq i} = 0\}$$

$$S_i^T = \frac{\sum_{\alpha \in J_i^T} y_\alpha^2}{\sum_{\alpha \in A, \alpha \neq 0} y_\alpha^2} \quad where \quad J_i^T = \{\alpha \in A, \alpha_i > 0\}$$

Limitation of PCE Method

- Dealing with discontinuous models, PCE method may encounter challenges in accurately capturing model's discontinuities.

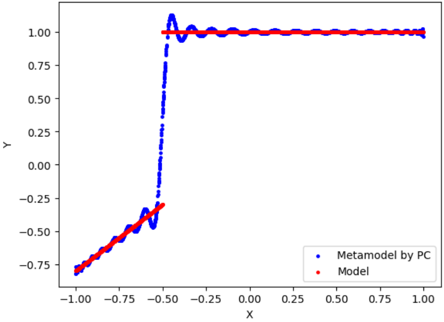


Figure: Approximation of a discontinuous function by PCE Method with total degree = 50

Aim of this work

Adaptation of Polynomial Chaos method for Models with Discontinuities :

1. Identify points of discontinuity and divide the domain into subdomains, each separated by a point of discontinuity.
2. Apply PCE within each subdomain.
3. Estimate the Sobol indices using the meta-models obtained.

Identify points of discontinuity.

Tree-PCE Algorithm

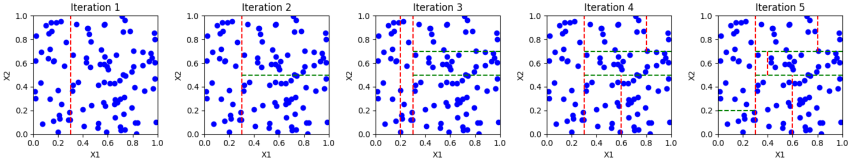


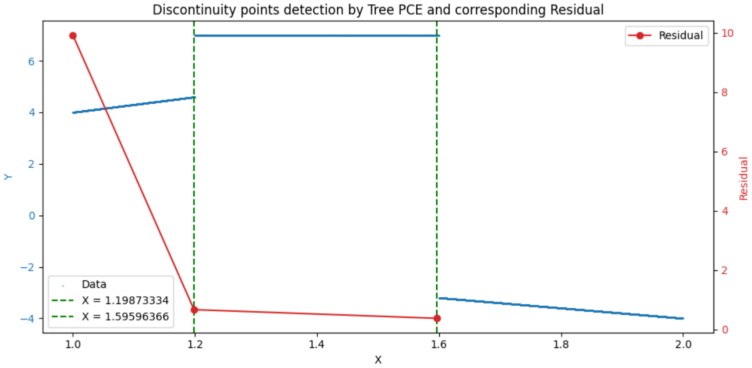
Figure: Tree-PCE iterations

- Select thresholds to classify data in a way that minimizes the Leave-One-Out (LOO) residual error of local PCEs.
- Refines iteratively the classification until the residual errors no longer improve.

Applications : Example 1D

For $X \in U([1,2])$,

$$Y = f(X) = \begin{cases} 3X+1 & \text{if } X < 1.2 \\ 7 & \text{if } 1.2 < X < 1.6 \\ -2X & \text{else} \end{cases}$$



Example 2D

For $X_1, X_2 \in U([1,2])$,

$$Y = \begin{cases} 3X_1 + 2X_2 & \text{if } X_1 < 1.7 \\ 15X_2 + 4X_1 & \text{if } X_1 < 1.5 \\ \begin{cases} -2X_2 & \text{if } X_2 < 1.8 \\ -8X_1 - 6X_2 & \text{if } X_2 < 1.3 \\ 8X_1 & \text{else} \end{cases} & \text{else} \end{cases}$$

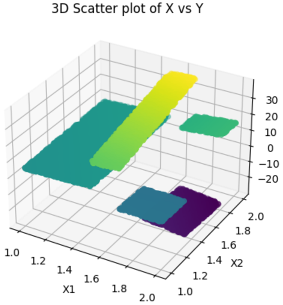


Figure: $Y = f(X)$

Detection of Points of Discontinuity by Tree-PCE

Detected points:

Step	X_1	X_2
1	1.696861	-
2	1.499787	-
2	-	1.797811
3	-	1.298169
4	1.794994	-

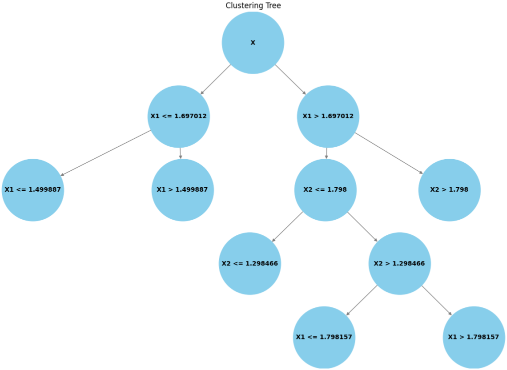


Figure: Clustering Tree

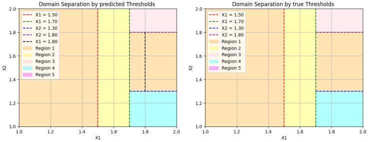


Figure: Comparison of Predicted Classes and Real Classes

Example with Triangular Domain

For $X_1, X_2 \in U([1,2])$,

$$Y = f(X) = \begin{cases} 0 & \text{if } X_1 < X_2 \\ 1 & \text{else} \end{cases}$$

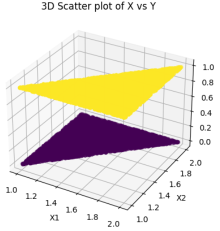


Figure: Y=f(X)

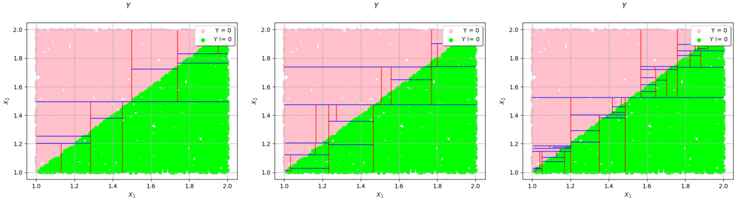


Figure: Application of Tree-PCE

Metamodel identified from Tree-PCE

The output of Tree-PCE is:

(i) a collection of d-dimensional rectangles $\mathcal{R}^r$, $r = 1, \dots, R$, indexed by some binary tree, which form a partition of the domain $\mathcal{D}_x$;

(ii) on each rectangle $\mathcal{R}^r$, which we denote by $\mathcal{R}^r = \prod_{i=1}^d \mathcal{I}^{(i),r}$, a local PCE model

$$\forall x \in \mathcal{R}^r, \qquad y \simeq \sum_{|\alpha| \leq p} y_\alpha^r \Psi_\alpha^r(x), \quad \Psi_\alpha^r(x) = \prod_{i=1}^d \Phi_{\alpha_i}^{(i),r}(x_i),$$

where for each $i \in \{1, \dots, d\}$, $(\Phi_j^{(i),r})_{j \geq 0}$ is a family of orthonormal polynomial in $L^2(f_{X_i|\mathcal{I}^{(i),r}})$, with in particular $\Phi_0^{(i),r}(x_i) = 1$ for any $x_i \in \mathcal{I}^{(i),r}$.

Metamodel identified from Tree- PCE.

Extending the definition of $\Phi_j^{(i),r}$ to $\mathcal{D}_{x_i}$ by setting $\Phi_j^{(i),r}(x_i) = 0$ if $x_i \notin \mathcal{I}^{(i),r}$, we thus obtain the global approximation

$$Y \simeq \tilde{Y} = \sum_{r=1}^R \sum_{|\alpha| \leq p} y_\alpha^r \Psi_\alpha^r(X).$$

Sobol indices from Tree- PCE

Sobol indices from Tree-PCE.

We have,

- $$\mathbb{E}[\tilde{Y}] = \sum_{r=1}^{R} \sum_{|\alpha|\leq p} y_\alpha^r \prod_{i=1}^{d} \mathbb{E}\left[\Phi_{\alpha_i}^{(i),r}(X_i)\right].$$

- $$\mathbb{E}[\tilde{Y}^2] = \sum_{r=1}^{R} \sum_{|\alpha|\leq p} (y_\alpha^r)^2 \mathbb{P}\left(X \in \mathcal{R}^r\right).$$

- $$\mathbb{E}[\tilde{Y}|X_I] = \sum_{r=1}^{R} \sum_{|\alpha|\leq p, \alpha_{I^c}=0} y_\alpha^r \left(\prod_{i\in I} \Phi_{\alpha_i}^{(i),r}(X_i)\right) \mathbb{P}(X_{I^c} \in \mathcal{R}_{I^c}^r),$$

- $$\mathbb{E}\left[\mathbb{E}[\tilde{Y}|X_I]^2\right] = \sum_{r,r'=1}^{R} \mathbb{P}(X_{I^c} \in \mathcal{R}_{I^c}^r)\mathbb{P}(X_{I^c} \in \mathcal{R}_{I^c}^{r'}) \sum_{\substack{|\alpha|\leq p, \alpha_{I^c}=0 \\ |\alpha'|\leq p, \alpha'_{I^c}=0}} y_\alpha^r y_{\alpha'}^{r'} C_{\alpha,\alpha'}^{r,r'}(I),$$

  with
  $$C_{\alpha,\alpha'}^{r,r'}(I) = \prod_{i\in I} \mathbb{E}\left[\Phi_{\alpha_i}^{(i),r}(X_i)\Phi_{\alpha'_i}^{(i),r'}(X_i)\right] = \prod_{i\in I} \int_{\mathcal{I}^{(i),r}\cap\mathcal{I}^{(i),r'}} \Phi_{\alpha_i}^{(i),r}(x_i)\Phi_{\alpha'_i}^{(i),r'}(x_i) f_{X_i}(x_i) dx_i, .$$

- if $\mathcal{I}^{(i),r} = \mathcal{I}^{(i),r'}$, then
  $$C_{\alpha,\alpha'}^{r,r'}(I) = \prod_{i\in I} 1_{\{\mathcal{I}^{(i),r}=\mathcal{I}^{(i),r'}, \alpha_i=\alpha'_i\}} \mathbb{P}(X_i \in I^{i,(r)})$$

Sobol indices :

$$S_A = \frac{\mathbb{E}\left[\mathbb{E}[Y|X_A]^2\right] - E[Y]^2}{E[Y^2] - E[Y]^2} \quad , \quad S_A^T = 1 - S_{A^c}$$

Example

For $X_1, X_2 \in U([1,2])$,

$$Y = f(X) = \begin{cases} 3X_1 + 5X_2 & \text{if } X_1 < 1.3 \text{ and } X_2 < 1.5 \\ 2 & \text{if } X_1 < 1.3 \text{ and } X_2 > 1.5 \\ -4X_2 & \text{if } X_1 < 1.3 \text{ and } X_2 < 1.5 \\ -3X_1 - 6X_2 & \text{else} \end{cases}$$

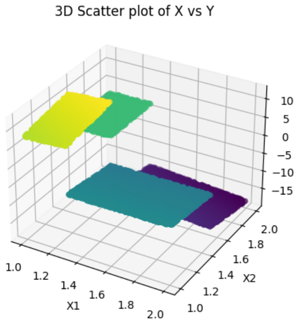


Figure: Y=f(X)

Detection of points of discontinuity by Tree-PCE

Detected points :

Step	X_1	X_2
1	1.292942	-
2	-	1.504695
3	-	1.494980

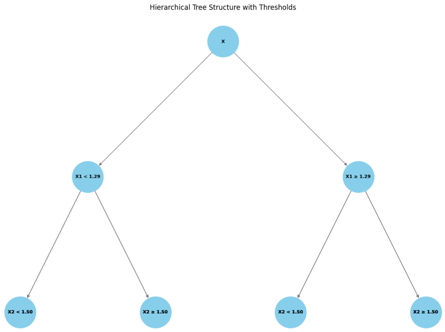


Figure: clustering Tree

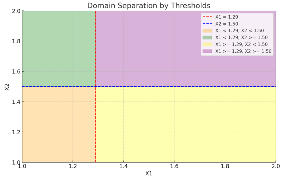


Figure: Domain Partition

Sobol indices from Tree-PCE

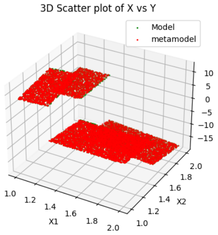


Figure: Metamodel vs Model

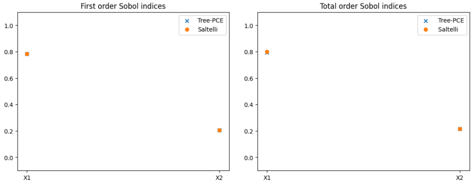
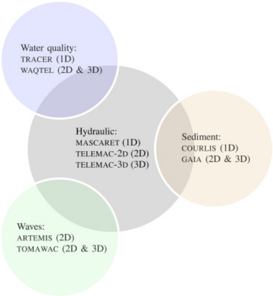


Figure: Comparison of Sobol' indices from Tree-PCE and Saltelli

Hydro-sediment test case

Simulation Model : TELEMAC-MASCARET [4].

- **TELEMAC-MASCARET :**
  - Used in the field of free-surface environmental flows.
  - Use high-capacity algorithms based on the finite-element or volume-element methods.
- **Telemac-2D :**
  - Solves the Saint-Venant equations in two dimensions in non-steady state.
- **Gaia:**
  - Simulate sediment and morphodynamic processes in coastal areas, rivers, lakes and estuaries,etc.



Hydro-sediment test case

Yen's [3] experience.

- Five tests conducted in a 180 canal bed, with the same sediment grain size distribution but varying input hydrographs.
- The canal bed comprised a layer of sand approximately 20 cm thick, with an initial median diameter $d_{50} = 1$ mm.
- The initial flow depth $h_0 = 5.44$ cm.
- Measurements were recorded at the peak and the end of each hydrograph.
- The hydro-sedimentary $2D$ model (TELEMAC-2D/GAIA) [4] is used to simulate the experience using the Meyer-Peter and Müller [5] formula for bedload

$$q_b^* = \begin{cases} \alpha_{MPM}(\theta - \theta_{cr})^{1.5} & \text{if} \quad \theta > \theta_{cr} \\ 0 & \text{else} \end{cases}$$

with q_b^* the non-dimensional bedload rate, α_{MPM} a factor (equal to 8 in the original formula), θ the Shields number, and θ_{cr} the critical Shields number for initiation of motion of sediments.

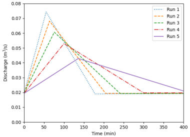


Figure: Hydrographs for experiments.

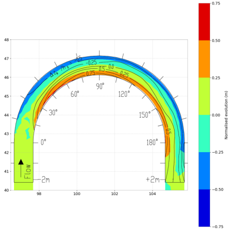
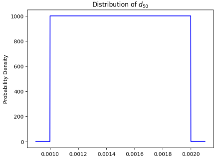


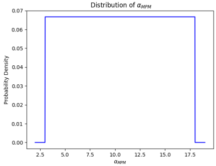
Figure: Contours of bed deformation for Run 4 at the end of the experiment. Negative values indicate erosion

Uncertainty Propagation

- Inputs :
  - Median diameter of sediment $d_{50} \sim U([0.001, 0.002])$
  - Coefficient of Meyer-Peter and Müller $\alpha_{MPM} \sim U([3, 18])$
  - d_{50} and α_{MPM} are independent.
- Output (QoI) :
  - Normalised bottom elevation variation $\frac{\Delta Z}{h_0} = \frac{Z_t - Z_{t_0}}{h_0}$

d_{50}

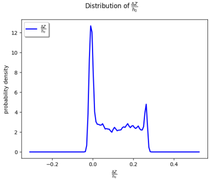




α_{MPM}

$\longrightarrow$ TELEMAC-2D + GAIA $\longrightarrow$

$\Delta Z / h_0$



Uncertainty Analysis

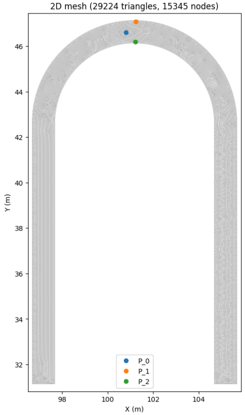


Figure: Points of interest

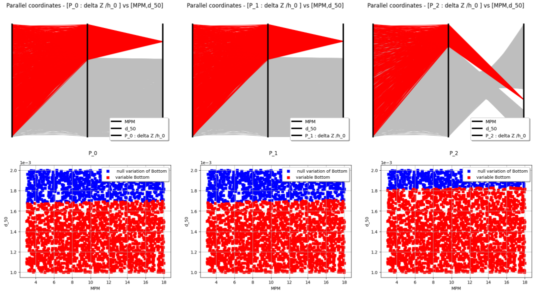


Figure: Bottom elevation variation

Application of Tree-PCE

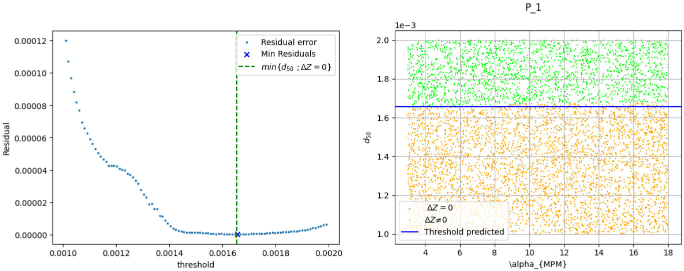


Figure: Detection of threshold

Comparison of Classic PCE and Tree-PCE

$N_{train} = 1000$

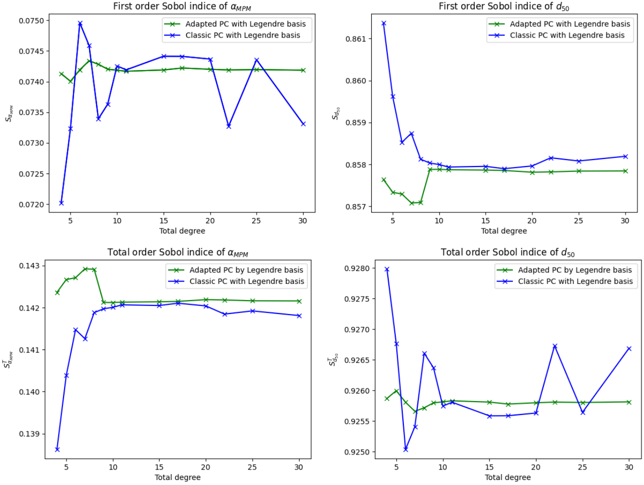


Figure: Convergence of Sobol' indices of d_{50} and α_{MPM}

Conclusion and perspectives

- Propagate uncertainty across 4 variables then all variables within the test case and implement Tree-PCE method in such scenarios.
- Model the dependencies between parameters and generalise the Tree-PCE method for dependent variables.

References

[1] A. Saltelli.
Making best use of model evaluations to compute sensitivity indices.
Computer Physics Communications, 2002, 145, 280–297.

[2] B. Sudret.
Global sensitivity analysis using polynomial chaos expansions.
Reliability Engineering & System Safety, 2008, 93(7), 964-979.

[3] Yen, C. L., & Lee, K. T. (1995). Bed topography and sediment sorting in channel bend with unsteady flow.
Journal of Hydraulic Engineering, 121(8), 591-599.

[4] J.M. Hervouet.
Free Surface Flows, 2007.

[5] E. Meyer-Peter, R. Müeller.
Formulas for bed-load transport.
Proc. IAHR World Congress, Stockholm, 1948.

Questions ?

Thank you for your attention !
Questions ?