Bayesian Sparse Polynomial Chaos Expansion

Qian Shao[1,2], Thierry A. Mara[2], Anis Younès[3]

[1]Wuhan University (China),[2]University of La Reunion (France),[3]IRD UMR LISAH, Montpellier (France)

8th SAMO Conference, Dec. 1 2016, Le Tampon (La Reunion)

ANOVA-HDMR

Sobol' (MMCE, 1993) proves that if:

1. $y = f(\mathbf{x}) \in \mathcal{L}^2$
2. $\mathbf{x}$ uniformly distributed over unit hypercube $\mathbb{I}_n$

then, $f(\mathbf{x})$ can be *uniquely* cast as,

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{j>i}^{n} f_{ij}(x_i, x_j) + \cdots + f_{1\ldots n}(x_1, \ldots, x_n) \quad (1)$$

by imposing that

$$\int_0^1 f_{i_1\ldots i_s}(x_{i_1}, \ldots, x_{i_s})\mathrm{d}x_j = 0, \text{ if } j \in (i_1 \ldots i_s) \quad (2)$$

Eq.(1) is called ANOVA-HDMR of $f(\mathbf{x})$ and leads to Sobol' sensitivity indices

What is PCE?

$\mathcal{L}^2$ is a complete space $\Rightarrow \exists$ orthonormal bases $\{\psi_{\boldsymbol{\alpha}}(\mathbf{x}) : \boldsymbol{\alpha} \in \mathbb{N}\}$ such that $\forall f(\mathbf{x}) \in \mathcal{L}^2$, one can write:

$$f(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}) \tag{3}$$

Polynomial chaoses are such orthonormal (orthogonal) bases:

Hermite polynomials, Legendre polynomials, Jacobi polynomials etc...

What is PCE?

$\mathcal{L}^2$ is a complete space $\Rightarrow \exists$ orthonormal bases $\{\psi_{\boldsymbol{\alpha}}(\mathbf{x}) : \boldsymbol{\alpha} \in \mathbb{N}\}$ such that $\forall f(\mathbf{x}) \in \mathcal{L}^2$, one can write:

$$f(\mathbf{x}) = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}) \tag{3}$$

Polynomial chaoses are such orthonormal (orthogonal) bases:

Hermite polynomials, Legendre polynomials, Jacobi polynomials etc...

Denoting $\psi_\alpha(x)$ the α-th degree one-dimensional basis element, the p-th degree multi-dimensional basis elements write:

$$\psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \prod_{i=1}^{n} \psi_{\alpha_i}(x_i), \text{ with } p = \sum_{i=1}^{n} \alpha_i$$

N.B.: Computing Sobol' indices with PCE = Computing the $c_{\boldsymbol{\alpha}}$s

Bayesian sparse PCE

In practice, only truncated and sparse PCEs are investigated:

$$f(\mathbf{x}) \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \boldsymbol{\psi}_{\mathcal{A}}^T(\mathbf{x}) \mathbf{c}_{\mathcal{A}}$$

where $\mathcal{A}$ is a non-empty **finite** subset of $\mathbb{N}^n$

Bayesian sparse PCE

In practice, only truncated and sparse PCEs are investigated:

$$f(\mathbf{x}) \simeq \sum_{\boldsymbol{\alpha} \in \mathcal{A}} c_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \boldsymbol{\psi}_{\mathcal{A}}^{T}(\mathbf{x}) \mathbf{c}_{\mathcal{A}}$$

where $\mathcal{A}$ is a non-empty **finite** subset of $\mathbb{N}^n$

Bayesian sparse PCE was introduced in Shao et al. (submitted to CMAME, 2016) to address the following issues:

- How to find the optimal subset $\mathcal{A}$?
- What are the best polynomial degree p and level of interaction q, given the samples $\mathbf{X}$ and $\mathbf{y}$?
- How to assign uncertainty bounds to the Sobol' indices estimate?

Our Solution: PC coefficients estimated in a Bayesian framework + Model selection criterion

Model selection criterion: Kashyap information criterion (Kashyap, IEEE Trans., 1982)
KIC was defined in a Bayesian framework

$$KIC_{\mathcal{A}} = -2 \ln \mathcal{P}(y|\mathcal{M}_{\mathcal{A}}, \tilde{\mathbf{c}}_{\mathcal{A}}) - 2 \ln \mathcal{P}(\tilde{\mathbf{c}}_{\mathcal{A}}|\mathcal{M}_{\mathcal{A}}) - P_{\mathcal{A}} \ln(2\pi) - \ln |\tilde{\mathbf{C}}| \tag{4}$$

with
Likelihood: $y|\mathcal{M}_{\mathcal{A}}, \tilde{\mathbf{c}}_{\mathcal{A}}, \tilde{\sigma}^2_{\mathcal{A}} \sim \mathcal{N}(\mathbf{\Psi}^T_{\mathcal{A}} \tilde{\mathbf{c}}_{\mathcal{A}}, \tilde{\sigma}^2_{\mathcal{A}})$
Prior: $\mathbf{c}_{\mathcal{A}}|\mathcal{M}_{\mathcal{A}} \sim \mathcal{N}(0, \mathbf{C})$

The choice of **C** favours low-degree and low-interaction level terms
The lower $KIC_{\mathcal{A}}$, the better the current sparse PCE $\mathcal{M}_{\mathcal{A}}$.

Posterior: $\mathbf{c}_{\mathcal{A}}|\mathcal{M}_{\mathcal{A}}, \mathbf{y}, \tilde{\sigma}^2_{\mathcal{A}} \sim \mathcal{N}(\tilde{\mathbf{c}}_{\mathcal{A}}, \tilde{\mathbf{C}})$

with

$$\tilde{\mathbf{c}}_{\mathcal{A}} = \frac{\tilde{\mathbf{C}}\boldsymbol{\Psi}_{\mathcal{A}}^T\mathbf{y}}{\tilde{\sigma}^2_{\mathcal{A}}} \tag{5}$$

$$\tilde{\mathbf{C}} = \left(\frac{\boldsymbol{\Psi}_{\mathcal{A}}^T\boldsymbol{\Psi}_{\mathcal{A}}}{\tilde{\sigma}^2_{\mathcal{A}}} + \mathbf{C}^{-1}\right)^{-1} \tag{6}$$

$$\tilde{\sigma}^2_{\mathcal{A}} = \frac{\left(\mathbf{y} - \boldsymbol{\Psi}_{\mathcal{A}}\tilde{\mathbf{c}}_{\mathcal{A}}\right)^T\left(\mathbf{y} - \boldsymbol{\Psi}_{\mathcal{A}}\tilde{\mathbf{c}}_{\mathcal{A}}\right)}{N} \tag{7}$$

Algorithm

1. *Initialization*: Set initial polynomial degree p and interaction level q (e.g. $p = 2, 4,\ q = 1, 2$). Create the initial subset $\mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^n : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$

Algorithm

1. *Initialization*: Set initial polynomial degree p and interaction level q (e.g. $p = 2, 4$, $q = 1, 2$). Create the initial subset $\mathcal{A} = \{\boldsymbol{\alpha} \in \mathbb{N}^n : p_{\boldsymbol{\alpha}} \leq p, q_{\boldsymbol{\alpha}} \leq q\}$
2. *Ranking via correlation coefficient*: Set $P = \mathrm{Card}\,(\mathcal{A})$ and define the polynomials $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_P)$ associated to $\mathcal{A}$. Then, compute the Pearson correlation coefficients:

$$r_j = \frac{\mathbb{COV}\left[y, \psi_j(\mathbf{x})\right]}{\sqrt{\mathbb{V}[y]\mathbb{V}[\psi_j(\mathbf{x})]}}$$

   Re-order the polynomial basis vector $\hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \ldots, \hat{\psi}_j, \hat{\psi}_{j+1}, \ldots, \hat{\psi}_P)$ such that $r_j^2 \geqslant r_{j+1}^2$

3. *Ranking via partial correlation coefficient:* Compute the partial correlation coefficients:

$$r_{j|1,\ldots,j-1} = \frac{\mathbb{COV}\left[y, \hat{\psi}_j(\mathbf{x}) | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right]}{\sqrt{\mathbb{V}\left[y | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right] \mathbb{V}\left[\hat{\psi}_j(\mathbf{x}) | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right]}}$$

As previously re-order the PC basis elements $\tilde{\boldsymbol{\psi}} = (\tilde{\psi}_1, \ldots, \tilde{\psi}_j, \tilde{\psi}_{j+1}, \ldots, \tilde{\psi}_P)$ such that $r^2_{j|1,\ldots,j-1} \geqslant r^2_{j+1|1,\ldots,j}$. Set $\boldsymbol{\psi}_{\mathcal{A}} = \tilde{\psi}_0$ and $k = 0$.

3. *Ranking via partial correlation coefficient:* Compute the partial correlation coefficients:

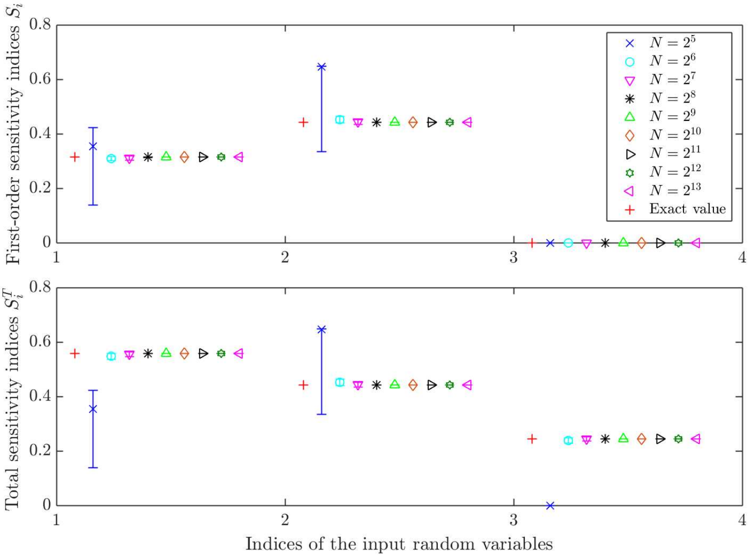
$$r_{j|1,\ldots,j-1} = \frac{\mathbb{COV}\left[y, \hat{\psi}_j(\mathbf{x}) | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right]}{\sqrt{\mathbb{V}\left[y | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right] \mathbb{V}\left[\hat{\psi}_j(\mathbf{x}) | \hat{\psi}_1(\mathbf{x}), \ldots, \hat{\psi}_{j-1}(\mathbf{x})\right]}}$$

   As previously re-order the PC basis elements $\tilde{\boldsymbol{\psi}} = (\tilde{\psi}_1, \ldots, \tilde{\psi}_j, \tilde{\psi}_{j+1}, \ldots, \tilde{\psi}_P)$ such that $r^2_{j|1,\ldots,j-1} \geqslant r^2_{j+1|1,\ldots,j}$. Set $\boldsymbol{\psi}_{\mathcal{A}} = \tilde{\psi}_0$ and $k = 0$.

4. *Model selection*: Set $k = k + 1$, $\boldsymbol{\psi}_{\mathcal{A}} = (\boldsymbol{\psi}_{\mathcal{A}}, \tilde{\psi}_k)$. Evaluate Eqs.(4-7). From the KIC estimate decide whether to keep or to remove $\tilde{\psi}_k$ from $\boldsymbol{\psi}_{\mathcal{A}}$. Resume until $k = P$.

5. *Enrichment of $\mathcal{A}$ or Stop*: If $\mathcal{A}$ contains elements of degree p or $(p-1)$ then set $p = p + 2$, if it contains elements of interaction level q then set $q = q + 1$. If p or q have been modified, then enrich $\mathcal{A}$ and resume from 2, otherwise stop.

The Ishigami function (3 inputs)



The Ishigami Function: $y = \sin x_1 + 0.1 \sin^2 x_2 + 7x_3^4 \sin x_1$,
$x_i \sim \mathcal{U}(-\pi, \pi)$

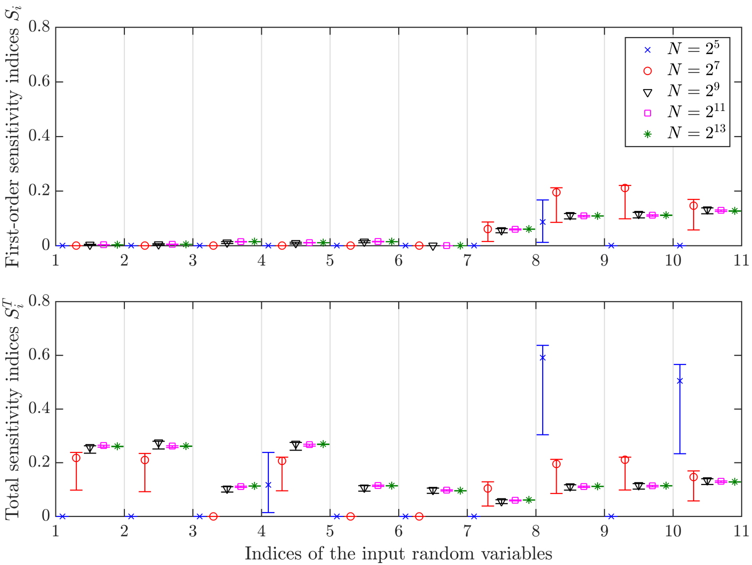
S_1^{ex}=0.3139	S_1^{PCE}=0.3086	
S_2^{ex}=0.4424	S_2^{PCE}=0.4424	
S_3^{ex}=0	S_3^{PCE}=0	
S_{12}^{ex}=0	S_{12}^{PCE}=0	$N = 64$ model runs
S_{13}^{ex}=0.2437	S_{13}^{PCE}=0.2364	
S_{23}^{ex}=0	S_{23}^{PCE}=0	
S_{123}^{ex}=0	S_{123}^{PCE}=0	

The Ishigami Function: $y = \sin x_1 + 0.1 \sin^2 x_2 + 7 x_3^4 \sin x_1$, $x_i \sim \mathcal{U}(-\pi, \pi)$

PCE structure for $N = 64$ model runs

0	0	0	$\tilde{c}_{000}$=3.4860
0	4	0	$\tilde{c}_{040}$=-1.9865
1	0	0	$\tilde{c}_{100}$=1.5015
3	0	0	$\tilde{c}_{300}$=-1.4191
0	6	0	$\tilde{c}_{060}$=1.3485
1	0	2	$\tilde{c}_{102}$=1.3233
3	0	2	$\tilde{c}_{302}$=-1.0833
0	2	0	$\tilde{c}_{020}$=-0.5966
1	0	4	$\tilde{c}_{104}$=0.5169
0	8	0	$\tilde{c}_{080}$=-0.3676
5	0	2	$\tilde{c}_{502}$=0.2295
5	0	4	$\tilde{c}_{504}$=0.1998
3	0	4	$\tilde{c}_{304}$=-0.1462

Morris function (20 inputs)



Conclusion

- The algorithm is virtually non-parametric
- Any statistic estimated with the BSPCE is a random variable
- Credible intervals can be assigned to estimated sensitivity indices
- Computational time depends on N and complexity of $f(\mathbf{x})$

Future works

- Account for model uncertainty, i.e.: Assess $(\mathbf{c}_{\mathcal{A}}, \mathcal{M}_{\mathcal{A}})\,|\mathbf{y}, \tilde{\sigma}^2_{\mathcal{A}}$ instead of $\mathbf{c}_{\mathcal{A}}|\mathcal{M}_{\mathcal{A}}, \mathbf{y}, \tilde{\sigma}^2_{\mathcal{A}}$
- Use BSPCE for optimal DOE