

Métamodélisation de codes à entrées fonctionnelles

Pecnum 2015

S. Nanty^{1,3} C. Helbert² A. Marrel¹ N. Pérot¹ C. Prieur³

¹CEA, DEN, Saint-Paul-lez-Durance

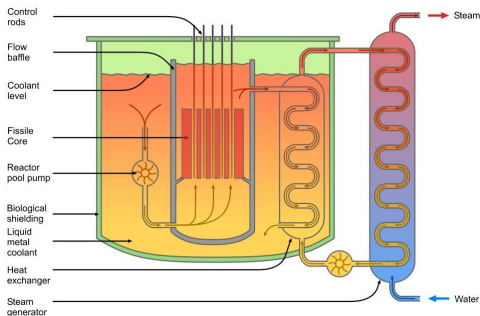
²Université de Lyon, UMR 5208, Ecole Centrale de Lyon, Institut Camille Jordan

³Université Joseph Fourier et INRIA, Grenoble

20 mai 2015

Contexte

- Étude de sûreté des réacteurs rapides refroidis au sodium
- Évaluation de la tenue des gaines des crayons de combustible lors d'un accident provoquant une forte augmentation de température dans le cœur
- Accident étudié : rupture de la liaison entre le sommier et une pompe primaire



Modélisation du phénomène physique

- Trio-U MCT utilisé pour modéliser les phénomènes thermo-hydrauliques dans le cœur du réacteur
- Coûteux en temps de calcul : une évaluation $\approx$ 2 heures
- Entrées scalaires et fonctionnelles
 - Sortie d'intérêt : température dans le cœur du réacteur
-
- Entrées fonctionnelles issues du code CATHARE, modélisant le réacteur à une échelle globale
- Coûteux en temps de calcul
- ⇒ Peu de réalisations des paramètres fonctionnels disponibles

Formalisation du problème (1)

- Entrées du code de calcul $\mathcal{M}$
 - $X_1, \dots, X_{m_X}$, variables aléatoires scalaires indépendantes de densité de probabilité μ_X
 - $Z_1, \dots, Z_{m_Z} : I \times \Omega \rightarrow \mathbb{R}$, variables aléatoires fonctionnelles dépendantes
- Sortie du code : Y , variable scalaire

$$Y = \mathcal{M}(Z_1, \dots, Z_{m_Z}, X_1, \dots, X_{m_X})$$

- Objectif de la métamodélisation : réaliser l'analyse de sensibilité du code $\mathcal{M}$

Formalisation du problème (2)

Problèmes

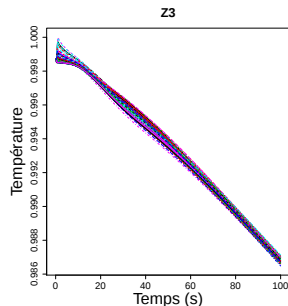
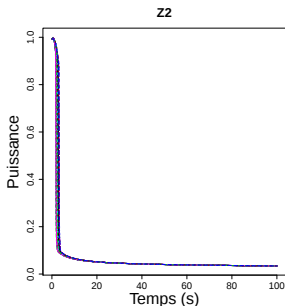
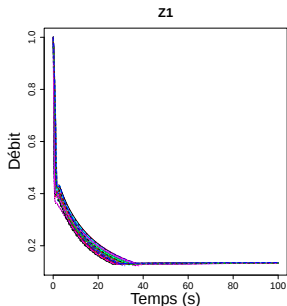
- Distribution de probabilité μ_Z de $(Z_1, \dots, Z_{m_Z})$ inconnue
 - Nécessité d'estimer cette loi
- Code $\mathcal{M}$ coûteux en temps de calcul
 - Utilisation d'un métamodèle
 - Nécessité de définir un plan d'expériences pour apprendre le métamodèle

Données

- Distribution de probabilité μ_X des variables scalaires
- Un échantillon probabilisé de n réalisations des variables fonctionnelles $\{z_i^{(j)}, 1 \leq i \leq m_Z, 1 \leq j \leq n\}$

Données de l'application CEA

- Echantillon de réalisations des trois variables aléatoires fonctionnelles normalisées :
 - Débit du caloporteur sodium en entrée du cœur
 - Puissance du cœur
 - Température du sodium en entrée du cœur



Méthodologie proposée

- 0. Quantification des incertitudes des entrées scalaires
- 1. Quantification des incertitudes des entrées fonctionnelles
 - 1.1 Décomposition sur une base fonctionnelle réduite
 - 1.2 Modélisation de la densité de probabilité des coefficients de la décomposition
- 2. Construction d'un métamodèle pour approximer $\mathcal{M}$
 - 2.1 Constitution d'un plan d'expériences
 - 2.2 Apprentissage du métamodèle

Méthodologie proposée

0. Quantification des incertitudes des entrées scalaires

1. Quantification des incertitudes des entrées fonctionnelles

1.1 Décomposition sur une base fonctionnelle réduite

1.2 Modélisation de la densité de probabilité des coefficients de la décomposition

2. Construction d'un métamodèle pour approximer $\mathcal{M}$

2.1 Constitution d'un plan d'expériences

2.2 Apprentissage du métamodèle sur l'échantillon

Méthodologie proposée

- 0. Quantification des incertitudes des entrées scalaires
- 1. Quantification des incertitudes des entrées fonctionnelles
 - 1.1 Décomposition sur une base fonctionnelle réduite
 - 1.2 Modélisation de la densité de probabilité des coefficients de la décomposition
- 2. Construction d'un métamodèle pour approximer $\mathcal{M}$
 - 2.1 Constitution d'un plan d'expériences
 - 2.2 Apprentissage du métamodèle sur l'échantillon

Etape 1.1 : Décomposition sur une base fonctionnelle

Définition

Soit $z : I \rightarrow \mathbb{R}$, $x \in I$

$$z(x) = \sum_{k=1}^{+\infty} \gamma_k \varphi_k(x) \longrightarrow \hat{z}(x) = \sum_{k=1}^d \gamma_k \varphi_k(x)$$

- γ_k coefficients
 - φ_k fonctions de base
 - d taille de la base
-
- Construction de la base par **Analyse en Composantes Principales Simultanée (ACPS)**
- **Réduction de dimension** : caractériser la loi de Z revient à caractériser la loi des paramètres $(\gamma_1, \dots, \gamma_d)$

Etape 1.1 : Analyse en Composantes Principales

- Cas multivarié
 - X variable aléatoire dans $\mathbb{R}^p$
 - Vecteurs de base de la décomposition : vecteurs propres de la covariance de X
- Cas fonctionnel [Ramsay et Silverman, 2005]
 - Z une variable aléatoire fonctionnelle
 - Fonctions de base φ : fonctions propres de

$$G : \xi \mapsto \left(t \mapsto \int_I \text{Cov}(Z(s), Z(t)) \xi(s) ds \right)$$

- Impossible à calculer en pratique
- ACP multivariée appliquée à la discrétisation de Z ou aux coefficients de Z sur une base fonctionnelle (splines, ondelettes...)

Etape 1.1 : ACP simultanée (ACPS)

- Objectif : étendre l'ACP au cas de plusieurs variables aléatoires fonctionnelles
- Soit $t_1 < \dots < t_p \in I$
- Soit $\mathbf{z}_i = (z_i(t_1), \dots, z_i(t_p))$ la version discrétisée de z_i , $i = 1, \dots, m_Z$
- L'ACP simultanée consiste à appliquer l'ACP multivariée à la concaténation des discrétisations des courbes normalisées :

$$(\mathbf{z}_1/N_1, \dots, \mathbf{z}_{m_Z}/N_{m_Z}) \in \mathbb{R}^{pm_Z},$$

avec les facteurs de normalisation $N_i = \sum_{j=1}^p \sqrt{\text{Var } Z_i(t_j)}$

Etape 1.2 : Modélisation de la distribution des coefficients de la décomposition

Objectif :

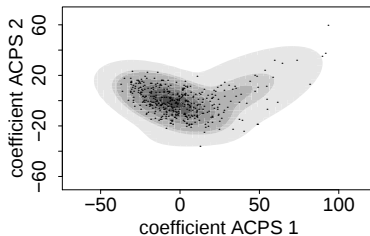
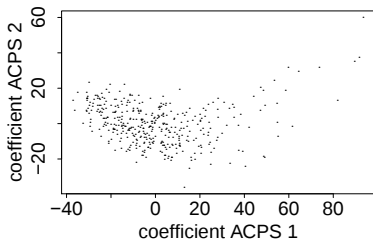
- Estimer la densité de probabilité des d coefficients de la décomposition ACPS

Contrainte :

- Grande dimension : $d \geq 10$
⇒ l'estimation à noyau n'est pas adaptée

Solution :

→ modélisation la densité par un **mélange de gaussiennes**



Etape 1.2 : Modélisation par un mélange de gaussiennes

- Densité de probabilité d'un mélange de gaussiennes :

$$f(\gamma|\mu_1, \Sigma_1, \dots, \mu_G, \Sigma_G) = \sum_{k=1}^G \tau_k \varphi(\gamma|\mu_k, \Sigma_k), \quad \forall \gamma \in \mathbb{R}^d$$

- G classes
- n points d'échantillonnage
- φ : densité de probabilité d'une loi normale
- τ_k, μ_k, Σ_k : proportion, moyenne et matrice de covariance de la classe k
- $\sum_{k=1}^G \tau_k = 1$
- Avantages / inconvénients
 - + Algorithme rapide d'estimation des paramètres
 - + Simulation très rapide de nouvelles réalisations
 - + Peut être utilisé en « grande » dimension
 - modèle paramétrique : hypothèses de modélisation
 - Le nombre de classes doit être choisi à l'avance

Etape 1.2 : Modélisation par un mélange de gaussiennes

- Densité de probabilité d'un mélange de gaussiennes :

$$f(\gamma|\mu_1, \Sigma_1, \dots, \mu_G, \Sigma_G) = \sum_{k=1}^G \tau_k \varphi(\gamma|\mu_k, \Sigma_k), \quad \forall \gamma \in \mathbb{R}^d$$

- G classes
- n points d'échantillonnage
- φ : densité de probabilité d'une loi normale
- τ_k, μ_k, Σ_k : proportion, moyenne et matrice de covariance de la classe k
- $\sum_{k=1}^G \tau_k = 1$
- Avantages / inconvénients
 - + Algorithme rapide d'estimation des paramètres
 - + Simulation très rapide de nouvelles réalisations
 - + Peut être utilisé en « grande » dimension
 - modèle paramétrique : hypothèses de modélisation
 - Le nombre de classes doit être choisi à l'avance

Etape 1.2 : Algorithme EM

- L'algorithme **Expectation-Maximization** (EM) [Dempster et al., 1977] estime les paramètres du mélange de gaussiennes.
- On introduit η_{ik} , la probabilité du i^e point dans la k^e classe.

Algorithme Expectation Maximization

1. Initialiser les paramètres $\tau_k^{(0)}$, $\mu_k^{(0)}$ et $\Sigma_k^{(0)}$
2. **Expectation** : Calculer $\eta_{ik}^{(j)}$
3. **Maximization** : Calculer $\tau_k^{(j+1)}$, $\mu_k^{(j+1)}$, $\Sigma_k^{(j+1)}$
4. Répéter les étapes 2 – 3 jusqu'à convergence

Etape 1.2 : Réduction du nombre de paramètres

- Nombre total de paramètres dans le modèle :

$$N_T = G - 1 + Gd + G \frac{d(d+1)}{2}$$

- G : nombre de classes

- N_T croît rapidement avec la dimension d

→ Solution : **estimation de matrices de covariance creuses**

Méthode sEM

- Une **pénalisation lasso** sur la matrice de covariance est ajoutée à la log-vraisemblance à l'étape 3 de maximisation :

$$\hat{\Sigma}_k = \operatorname{argmax}_{\Sigma_k} (\ell - \lambda \|\Sigma_k\|_1)$$

- Le paramètre de pénalisation λ est choisi par **validation croisée**.
- Le problème pénalisé est résolu par l'algorithme de [Wang, 2014].

Méthodologie proposée

- 0. Quantification des incertitudes des entrées scalaires
- 1. Quantification des incertitudes des entrées fonctionnelles
 - 1.1 Décomposition sur une base fonctionnelle réduite
 - 1.2 Modélisation de la densité de probabilité des coefficients de la décomposition
- 2. Construction d'un métamodèle pour approximer $\mathcal{M}$
 - 2.1 Constitution d'un plan d'expériences
 - 2.2 Apprentissage du métamodèle sur l'échantillon

Contexte et objectifs

Objectifs

- Approximation de $\mathcal{M}$ par un métamodèle
 - Nécessité d'avoir une base d'apprentissage de N simulations pour apprendre le métamodèle
- Métamodèle prédictif sur tout le domaine de variation des entrées
 - Un plan d'expériences probabilisé favoriserait certaines zones de l'espace des entrées
 - ⇒ Utilisation d'un plan d'expériences uniforme
 - ⇒ On n'utilise pas directement les courbes de l'échantillon disponible

Données disponibles

- Domaine de variation et distribution des variables $X_1, \dots, X_{m_X}$
- n réalisations probabilisées $z_1^{(j)}, \dots, z_{m_Z}^{(j)}$
- Loi estimée $\hat{\mu}_Z$ des variables fonctionnelles (cf. étape 1)

Méthodologie

1. Construction de l'échantillon D_X des variables scalaires
2. Construction de l'échantillon D_Z des variables fonctionnelles
3. Combinaison de D_X et D_Z

Echantillonnage des variables scalaires

- Si le domaine de variation de la variable X_i n'est pas borné :
 - Restriction de l'échantillonnage à l'intervalle des quantiles $\alpha/2$ et $(1 - \alpha/2)$ de la loi de X_i
 - Sinon, on garde l'intervalle de la distribution de X_i
- On note $\mathcal{D}_\alpha = I_1^\alpha \times \cdots \times I_{m_X}^\alpha$ le domaine restreint
- Échantillonnage choisi : plan hypercube latin (LHS) optimisé selon la discrédance

Echantillonnage des variables fonctionnelles (1)

- La loi des coefficients de la décomposition est non bornée (mélange de gaussiennes)
- Troncature de cette loi et donc de la loi des variables aléatoires fonctionnelles

Méthode

1. Tronquer la distribution des coefficients
2. Echantillonner uniformément sur le support de la loi tronquée

- Support de la troncature :

$$R_\alpha = \{\gamma \in \mathbb{R}^d : f(\gamma) \geq f_\alpha\},$$

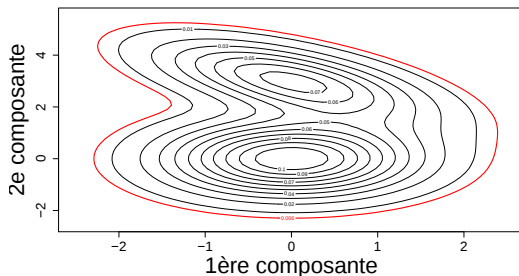
avec f_α tel que $\int_{R_\alpha} f(\gamma) d\gamma = 1 - \alpha$.

⇒ Union d'ellipsoïdes

- R_α contient les coefficients de plus hautes densités

Echantillonnage des variables fonctionnelles (2)

Densité de probabilité des coefficients et espace R_α (en rouge) :



- Choix de α
 - α élevé $\Rightarrow R_\alpha$ petit $\Rightarrow$ importante perte d'information pendant la troncature
 - α faible $\Rightarrow$ métamodèle moins bien estimé car trop faible ratio entre la taille de l'échantillon d'apprentissage et le volume de R_α
 - Choix de α discuté dans les applications

Échantillonnage uniforme sur un espace quelconque

- Problème plus difficile que l'échantillonnage dans un hypercube
- Méthode de rejet
 - Aucune assurance sur la qualité de l'échantillonnage
- Proposition de [Auffray et al., 2012] :
 - Algorithme d'échantillonnage d'un plan optimal pour le critère **maximin** sur un espace E quelconque
 - Algorithme de **recuit simulé**
 - Seule hypothèse : la fonction indicatrice de E , $\mathbb{I}(\cdot \in E)$, est connue

Algorithme [Auffray et al., 2012]

- Pour t de 0 à N_{iter}
 1. Choix de (x_i^t, x_j^t) dans X^t avec probabilités $\propto 1/\|x_i^t - x_j^t\|_2$.
 2. Choix d'un des deux points avec une probabilité $1/2$: x_k^t .
 3. Point candidat obtenu par marche aléatoire contrainte :

$$x_k^{prop} \sim \mathcal{N}(x_k^t, \tau_t \hat{\Sigma}) \mathbb{I}(\cdot \in E)$$

Plan d'expériences candidat :

$$X^{prop} = \{x_1^t, \dots, x_{k-1}^t, x_k^{prop}, x_{k+1}^t, \dots, x_N^t\}.$$

4. $X^{t+1} = X^{prop}$ avec probabilité

$$p_{prop}^t = \min \left(1, \exp \left(-\beta_t \left[\min_{x, y \in X^t} \|x - y\| - \min_{x, y \in X^{prop}} \|x - y\| \right] \right) \right)$$

et $X^{t+1} = X^t$ sinon.

Combinaison des plans d'expériences (1)

- [Muehlenstaedt et al., 2014] proposent de combiner les plans D^X et D^Z en optimisant un critère de remplissage de l'espace
- On définit une combinaison des deux plans :

$$D(\pi) = \left(\left(x^{(1)}, z^{(\pi(1))} \right), \dots, \left(x^{(n)}, z^{(\pi(n))} \right) \right),$$

- Le critère à maximiser est une variante régularisée du critère maximin :

$$\phi_q(D(\pi)) = \left[\sum_{1 \leq i < j \leq n} d \left(\left(x^{(i)}, z^{(\pi(i))} \right), \left(x^{(j)}, z^{(\pi(j))} \right) \right)^{-q} \right]^{1/q}.$$

- Minimiser ϕ_q est équivalent à maximiser le critère maximin pour q suffisamment grand (en pratique $q = 50$)
- Optimisation à l'aide d'un algorithme de recuit simulé [Morris et Mitchell, 1995]

Combinaison des plans d'expériences (2)

- Quelle distance choisir sur l'espace des entrées ?
- On cherche à ce que chaque variable ait une contribution équivalente dans la distance
- Distance proposée :

$$d_w \left(\left(x^{(i)}, z^{(i)} \right), \left(x^{(j)}, z^{(j)} \right) \right)^2 = \sum_{k=1}^{m_X} w_{x_k} \|x_k^{(i)} - x_k^{(j)}\|_2^2 + w_z \|\gamma^{(i)} - \gamma^{(j)}\|_2^2$$

- avec les pondérations suivantes :

$$w_{x_k} = 1 / \max_{x, x' \in I_k^\alpha} \|x - x'\|_2^2, \text{ pour } k = 1, \dots, m_X$$

$$w_z = 1 / \max_{\gamma, \gamma' \in R_\alpha} \|\gamma - \gamma'\|_2^2$$

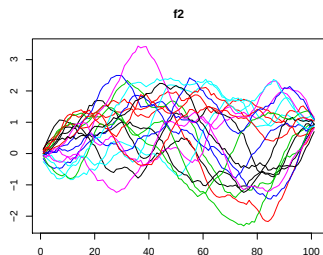
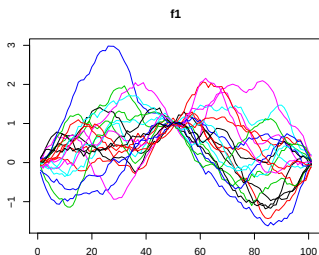
- Remarque : distance L^2 sur les fonctions $\Leftrightarrow$ distance L^2 sur leurs coefficients de l'ACPS

Construction du métamodèle

- Données :
 - Échantillon uniforme de N réalisations des variables scalaires $X_1, \dots, X_{m_X}$ et des coefficients $\gamma_1, \dots, \gamma_d$
- L'espace des entrées est de taille $m_X + d$
 - N sorties de $\mathcal{M}$ correspondantes
- Métamodèle processus gaussien [Rasmussen et Williams, 2006]
 - $Y_{GP}(X) = f(X) + Z(X)$
 - Z : processus gaussien centré de covariance Matérn 5/2
 - f : polynôme d'ordre 1
- Les hyper-paramètres sont estimés par maximum de vraisemblance à partir de l'échantillon d'apprentissage uniforme tiré sur $\mathcal{D}_\alpha \times R_\alpha$.
- Validation du métamodèle par le coefficient Q^2

Cas analytique : définition

- $m_Z = 2$ variables fonctionnelles Z_1, Z_2 sur $I = [0; 1]$
 - Z_1 : processus gaussien de noyau Matérn 3/2 avec $Z_1(-0, 01) = Z_1(1, 01) = 0$ et $Z_1(0, 5) = 1$
 - Z_2 : somme de Z_1 et d'un processus gaussien Z de noyau Matérn 3/2 avec $Z(-0, 01) = 0$ et $Z(1, 01) = 1$
- $m_X = 3$ variables scalaires i.i.d. $X_1, X_2, X_3 \sim \mathcal{U}(0, 1)$
- La sortie $Y = 2X_1^2 X_2 + 2 \int_0^1 (Z_1 + Z_2)(t)dt + X_3 \max_{t \in [0;1]} Z_2(t) + X_3$



Cas analytique : quantification des incertitudes (1)

- Données :
 - Base d'apprentissage : échantillon de 100 réalisations de (Z_1, Z_2)
- Méthode de caractérisation
 - Décomposition sur une base fonctionnelle : ACPS
 - Estimation de la densité : méthode sEM
- Critères proposés pour sélectionner la taille d et évaluer la qualité de la méthode :
 - **Critère C1** : Test d'adéquation entre la **densité de probabilité des coefficients** estimés et la densité réelle avec le test de [Fromont et al., 2012].
 - **Critère C2** : Erreur quadratique entre la **corrélation temps par temps** des variables estimées et des variables réelles.

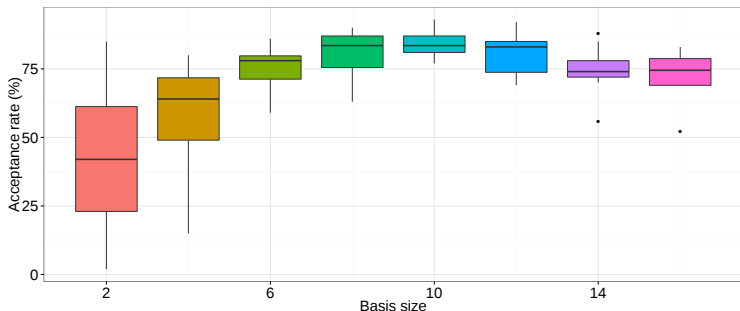
Cas analytique : quantification des incertitudes (1)

- Données :
 - Base d'apprentissage : échantillon de 100 réalisations de (Z_1, Z_2)
- Méthode de caractérisation
 - Décomposition sur une base fonctionnelle : ACPS
 - Estimation de la densité : méthode sEM
- Critères proposés pour sélectionner la taille d et évaluer la qualité de la méthode :
 - **Critère C1** : Test d'adéquation entre la **densité de probabilité des coefficients** estimés et la densité réelle avec le test de [Fromont et al., 2012].
 - **Critère C2** : Erreur quadratique entre **la corrélation temps par temps** des variables estimées et des variables réelles.

Cas analytique : quantification des incertitudes (2)

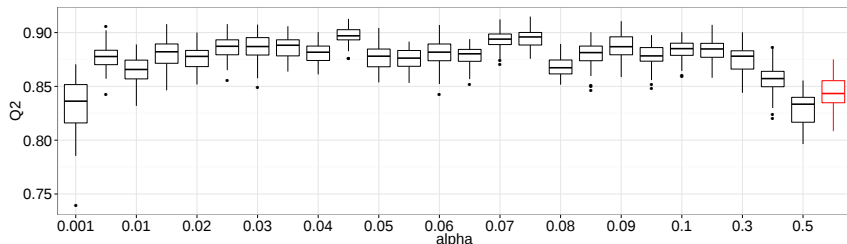
- **Critère C1** : Taux d'acceptation assez élevés pour $d \gtrsim 6$ (cf. figure)
- **Critère C2** : Bonne approximation des corrélations

→ Taille de la base ACPS : $d = 10$



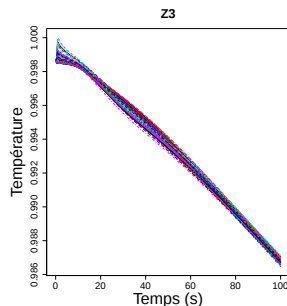
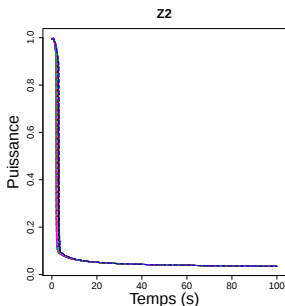
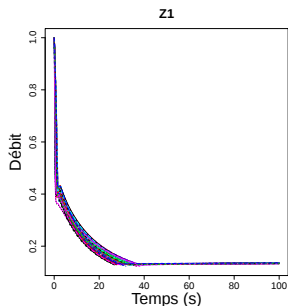
Cas analytique : construction du métamodèle

- Evolution du Q^2 du métamodèle appris sur R_α en fonction de α
 - Echantillons E_α de tailles 600
- Q^2 calculé sur une base de test tirée selon les distributions des variables d'entrée
 - Echantillon E de taille 600
- Q^2 calculé sur une base de test tirée selon la distribution des coefficients
- Les métamodèles appris sur les échantillons uniformes donnent de meilleurs Q^2 que celui appris sur l'échantillon non uniforme



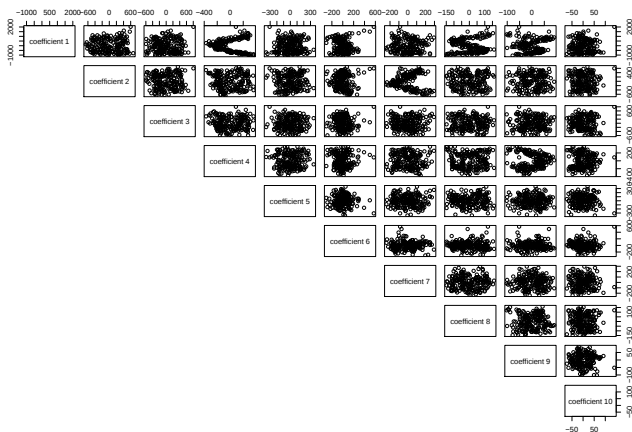
Application CEA : données

- En entrée de Trio-U MCT, 8 paramètres incertains
- $m_Z = 3$ variables fonctionnelles Z_1, Z_2, Z_3 sur $I = [0; 100]$
- Echantillon probabilisé de 200 réalisations de Z_1, Z_2, Z_3 disponible
- $m_X = 5$ variables scalaires (géométrie du cœur et propriétés des gaines des crayons)
- Sorties d'intérêt : températures du sodium et des gaines



Application CEA : quantification des incertitudes (1)

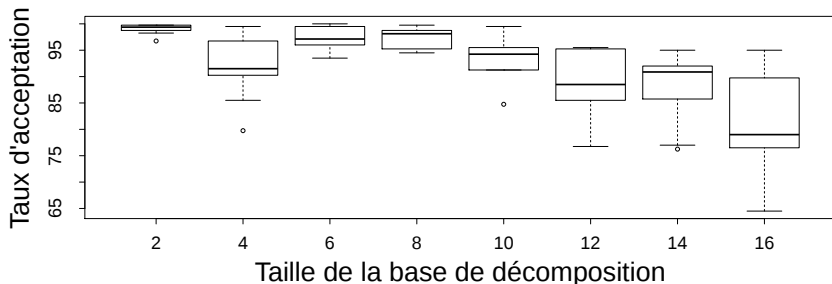
- Choix du nombre de gaussiennes par le critère BIC : $G = 3$
- 10 premiers coefficients de l'ACPS :



Application CEA : quantification des incertitudes (2)

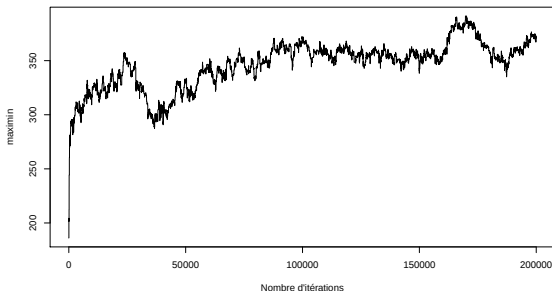
- Application de la méthode de quantification des incertitudes
- Critères calculés par validation croisée
- Critère C1 : Taux d'acceptation élevés pour $d \leq 10$ (cf. figure)
- Critère C2 : Bonne approximation des corrélations

→ Taille de la base ACPS : $d = 8$



Application CEA : échantillonnage

- On prend $\alpha = 0,05$ pour construire les domaines d'échantillonnage des variables scalaires et fonctionnelles
- Échantillon de taille 100
- D_X : plan LHS optimisé selon la discrédance centrée
- D_Z : plan optimisé selon le critère maximin
- Evolution du critère maximin en fonction du nombre d'itérations



Application CEA : construction du métamodèle

- Construction d'un métamodèle processus gaussien en dimension $m_X + d = 15$
 - Tendances : régression linéaire avec sélection de variables itérative
 - Calcul du Q^2 par leave-one-out :
 - $Q^2 \approx 0,98$ pour la température du sodium
 - $Q^2 \approx 0,99$ pour la température des gaines
- ⇒ Très bonne approximation du code sur le domaine tronqué

Conclusion et perspectives

- Développement d'une méthodologie de métamodélisation d'un code coûteux en temps de calcul et à entrées fonctionnelles :
 - Méthode de quantification des incertitudes de variables fonctionnelles
 - ACP simultanée
 - Estimation de la densité par un mélange de gaussiennes avec des matrices de covariance creuses
 - Proposition d'une méthode pour échantillonner uniformément sur un espace de variables scalaires et fonctionnelles
 - Construction d'un métamodèle processus gaussien
- Application à un cas analytique et à l'étude d'un cas de sûreté nucléaire
 - Bons résultats sur l'ensemble de la méthode

Perspectives :

- Comparer avec l'échantillonnage simultané des variables scalaires et fonctionnelles

References I



Auffray, Y., Barbillon, P. et Marin, J.-M. (2012).
Maximin design on non hypercube domains and kernel interpolation.
Statistics and Computing, 22(3):703–712.



Dempster, A. P., Laird, N. M. et Rubin, D. B. (1977).
Maximum likelihood from incomplete data via the EM algorithm.
Journal of the Royal Statistical Society. Series B (Methodological), 39:1–38.



Fromont, M., Laurent, B., Lerasle, M. et Reynaud-Bouret, P. (2012).
Kernels based tests with non-asymptotic bootstrap approaches for two-sample problem.
In *25th Annual Conference on Learning Theory*, volume 23, pages 1–22.



Morris, M. D. et Mitchell, T. J. (1995).
Exploratory designs for computational experiments.
Journal of statistical planning and inference, 43(3):381–402.



Muehlenstaedt, T., Fruth, J. et Roustant, O. (2014).
Computer experiments with functional inputs and scalar outputs by a norm-based approach.
arXiv preprint arXiv :1410.0403.

References II



Ramsay, J. O. et Silverman, B. W. (2005).
Functional Data Analysis.
Springer Series in Statistics. Springer.



Rasmussen, C. E. et Williams, C. K. I. (2006).
Gaussian Processes for Machine Learning.
MIT Press.



Wang, H. (2014).
Coordinate descent algorithm for covariance graphical lasso.
Statistics and Computing, 24(4):521–529.