

Calage bayésien pour du pronostic hybride – Application à l'étude du colmatage des générateurs de vapeur

Edgar Jaber

Séminaire des Doctorants ETICS 2024
27/09/2024

Encadrants : E. Remy, V. Chabridon (EDF R&D)

Co-directeurs : D. Lucor (LISN), M. Mougeot (Centre Borelli)

Collaborateurs : M. Keller, B. Iooss, M. Baudin, Q. Feng (EDF R&D)



Sommaire

Colmatage des GV et motivations

Outils à disposition

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Sommaire

Colmatage des GV et motivations

Outils à disposition

Modèle physique : THYC-Puffer-DEPOTHYC

Modèle statistique : ESTICOL

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Le générateur de vapeur (GV)

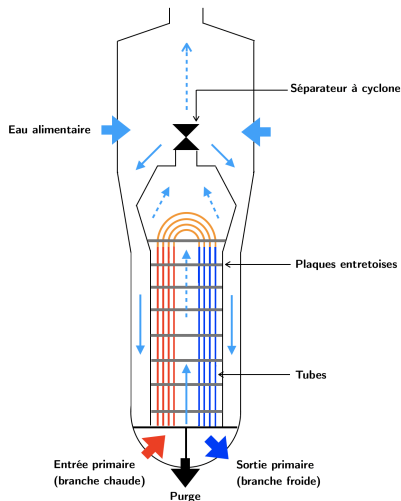


Figure – Schéma d'un générateur de vapeur (GV) typique (source : dessin propre).

Cas d'étude industriel

- ▶ Colmatage des plaques entretoises → phénomène complexe affectant les GVs du parc nucléaire français
- ▶ Apparaît sur de longues périodes d'exploitation → dû à la déposition progressive de produits de corrosion issus du circuit secondaire
- ▶ Augmente le risque de vibration des tubes et peut affecter la réponse du GV à des transitoires hypothétiques
- ▶ Taux de colmatage → mesuré par des *examens télévisuels* (ETV)
- ▶ ETV sont mesurés lors des arrêts de tranche → peu de mesures à disposition en tant que « vérité terrain »

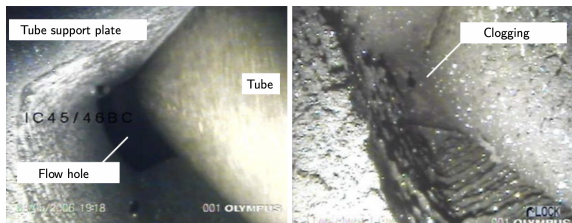


Figure – Exemple d'un ETV © EDF.

Cas d'étude industriel

- ▶ Réduire ponctuellement le colmatage → maintenances préventives par nettoyages chimiques (NC)
- ▶ Pour appuyer la planification des maintenances → travaux de R&D pour mieux comprendre, estimer et prédire le taux de colmatage τ_c
- ▶ Les différents outils issus de ces travaux sont aujourd'hui capitalisés dans la plateforme **Jumeau Numérique GV (JNGV)**
- ▶ Deux typologies d'outils :
 - ▶ *Modèle de simulation numérique* : THYC-Puffer-DEPOTHYC → basé sur une modélisation physico-chimique multi-échelle [Prusek, 2012; Feng et al., 2023]
 - ▶ *Modèle statistique* : ESTICOL → basé sur des données opérationnelles de la centrale et les ETV [Pincioli et al., 2021]
- ▶ But de la thèse : *hybrider* ces modèles pour un pronostic du colmatage plus robuste

Notions de pronostic

- ▶ Le taux de colmatage τ_c représente un indice de *dégradation*
- ▶ L'évènement d'intérêt pour la planification des maintenances est $\{\tau_c(t) > c\}$ (dépassement de seuil) → seuil conservatif
- ▶ Quantité d'intérêt est la RUL (Remaining Useful Life) [Bregon and Daigle, 2019], défini pour un certain GV à partir du temps présent t_P

$$RUL(GV, t_P, c) = \arg \min_{t > t_P} |\tau_c(t) > c|. \quad (1)$$

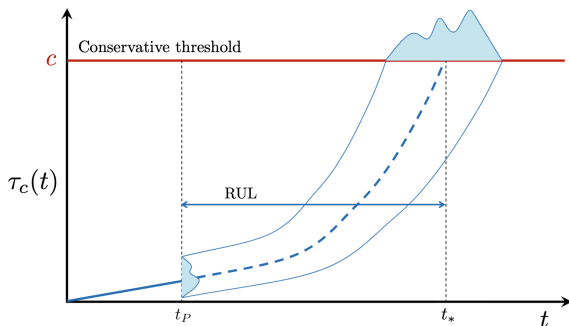


Figure – Illustration de la RUL, incertitudes sur un modèle de dégradation et impact sur le pronostic, adapté de [Bregon and Daigle, 2019].

Objectif de la thèse

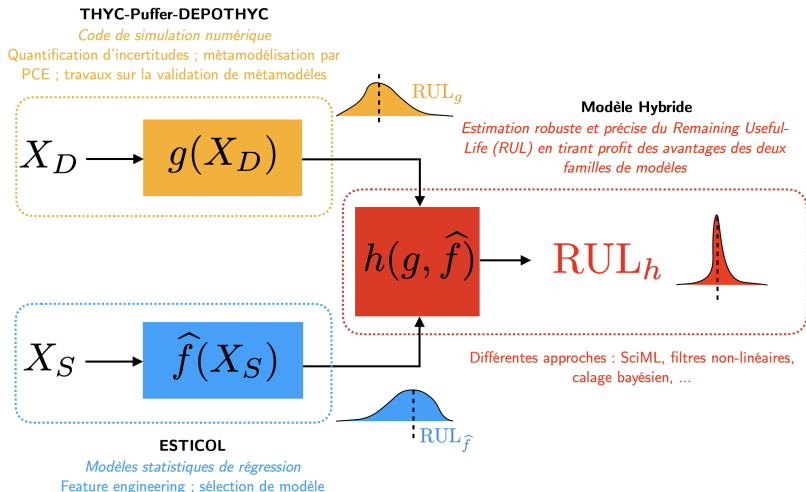


Figure – Schéma objectif de thèse (source : dessin propre)

Sommaire

Colmatage des GV et motivations

Outils à disposition

Modèle physique : THYC-Puffer-DEPOTHYC

Modèle statistique : ESTICOL

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Physique du colmatage

- ▶ Première brique : DEPOTHYC/COLMATHYC, développé par [Prusek, 2012] → système EDO-EDP mixte → permet de simuler la cinétique de colmatage à horizon de temps court
- ▶ Se base sur la validité physique des champs thermohydrauliques stationnaires → invalide sur des longues périodes d'exploitation
- ▶ Couplage multi-physique pour la simulation en temps long : THYC-Puffer-DEPOTHYC, développé par [Feng et al., 2023] → permet aussi de prendre en compte le conditionnement chimique (pH) du fluide grâce à un code de chimie (Puffer)
- ▶ Le modèle de colmatage correspond au système suivant (2 EDP de transport pour les fractions massiques + 1 EDO de croissance de matière pour la masse de magnétite) :

$$\begin{cases} \partial_t \Gamma_p + U \cdot \nabla \Gamma_p = f_p(\Phi_p, \Phi_s, \Gamma_p, \Gamma_s^{\max}) \\ \partial_t \Gamma_s + U \cdot \nabla \Gamma_s = f_s(\Phi_p, \Phi_s, \Gamma_s, \Gamma_s^{\max}) \\ dm_c/dt = \Phi_p + \Phi_s \\ \Gamma_s(0, \cdot) = \Gamma_s(0), \Gamma_p(0, \cdot) = \Gamma_p(0) \end{cases} \quad (2)$$

Le code THYC-Puffer-DEPOTHYC (TPD)

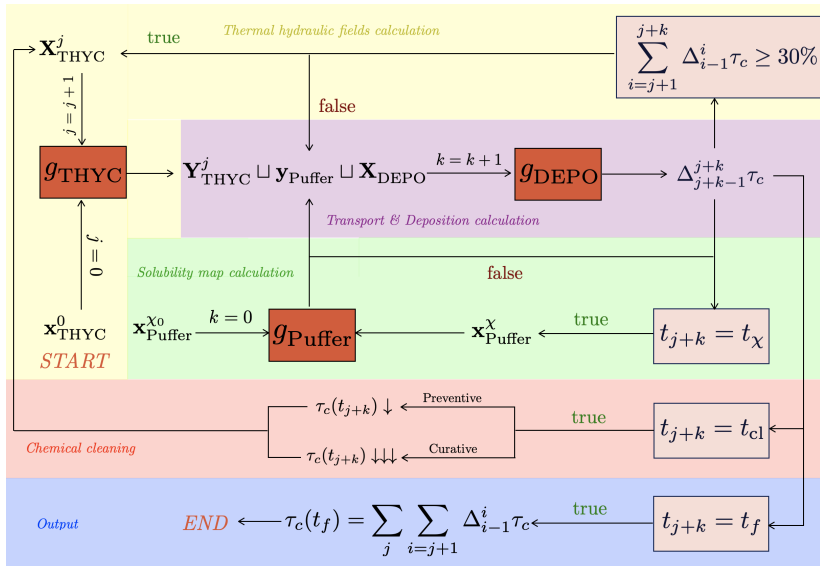


Figure – Architecture de TPD.

Jeu de données de TPD

- ▶ Par avis d'expert + travaux antérieurs [Lefebvre et al., 2023] : $d = 7$ variables d'entrée du module de dépôt g_{DEPO} entâchées d'incertitude (dont un paramètre de **calage**) → traitement probabiliste avec lois indépendantes données dans la table ci-dessous :

$$\mathbf{X}_{\text{DEPO}} = (\mathbf{X}, \theta) = (\alpha, \beta, \epsilon_e, \epsilon_c, d_p, \Gamma_p(0), a_v) \sim \otimes_{i=1}^d \mathbb{P}_{X_i} \quad (3)$$

- ▶ On dispose d'un n -échantillon en *given-data* de $n = 1000$ tirages Monte-Carlo (i.i.d.) tirés selon les lois données ci-dessous :

Variable	Signification	Distribution
α	Premier paramètre de corrélation empirique	$\mathcal{N}(101.6, 4.0)$
β	Second paramètre de corrélation empirique	$\mathcal{N}(0.0233, 0.0005)$
ϵ_e	Porosité des dépôts d'encrassement	$\mathcal{T}(0.2, 0.3, 0.5)$
ϵ_c	Porosité des dépôts colmatants	$\mathcal{T}(0.01, 0.05, 0.3)$
d_p	Diamètre des particules de magnétite (m)	$\mathcal{T}(0.5, 5.0, 10.0) \times 10^{-6}$
$\Gamma_p(0)$	Donnée initiale de la fraction massique solide	$\mathcal{T}(1.0, 4.5, 8.0) \times 10^{-9}$
a_v	Paramètre de calage	$\mathcal{U}[0, 15] \times 10^{-4}$

Table – Variables incertaines, signification physique et distribution.

Métamodèle de Chaos Polynomial vectoriel de TPD

- ▶ Code TPD coûteux en temps de calcul → Métamodélisation par chaos polynomial vectoriel (PCE) sur la sortie [Soize and Ghanem, 2004] en approche non-intrusive. Ici $\tilde{g}_{\text{TPD}} : \mathcal{X} \rightarrow \mathbb{R}^N$
- ▶ Avantages du PCE : garanties d'approximation fonctionnelle, permet d'utiliser la connaissance des distributions d'entrée, analyse de sensibilité facilitée, permet de gérer la discontinuité, métamodèle déjà validé dans des travaux précédents [Jaber et al., 2024]
- ▶ Travaux en cours : métamodélisation à base de processus gaussiens

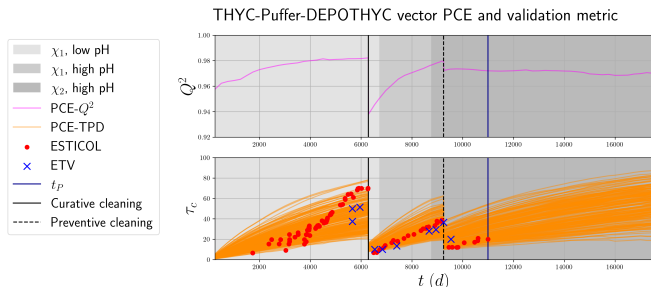


Figure – Trajectoires de PCE-TPD et variation du Q^2 .

ESTICOL

- ▶ ESTICOL \rightarrow modèle de régression utilisé pour estimer des taux de colmatage à un temps $t < t_P$ sur des features opérationnelles "agrégées" en une seule covariable $X^{(i)} \in \mathbb{R}$ (travaux basés sur [Pincioli et al., 2021])
- ▶ $\mathcal{D}_{\text{EST}} = \{(X^{(i)}, Y^{(i)})\}_{i=1}^P$ est le jeu de données construit avec Y_i un ETV au temps t_i et on résout le problème de minimisation :

$$(\hat{a}_{\text{EST}}, \hat{b}_{\text{EST}}) = \arg \min_{a,b} \sum_{i=1}^P w_i |aX^{(i)} + b - Y^{(i)}|^2, \quad \forall i \ w_i \propto \exp(t_i^2)$$

- ▶ Estimation à un temps $t_j < t_P$:

$$\hat{\tau}_c(t_j) = Y^{(j)} = \hat{a}_{\text{EST}} X^{(j)} + \hat{b}_{\text{EST}} \quad (4)$$

Sommaire

Colmatage des GV et motivations

Outils à disposition

Modèle physique : THYC-Puffer-DEPOTHYC

Modèle statistique : ESTICOL

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Méthodologie proposée

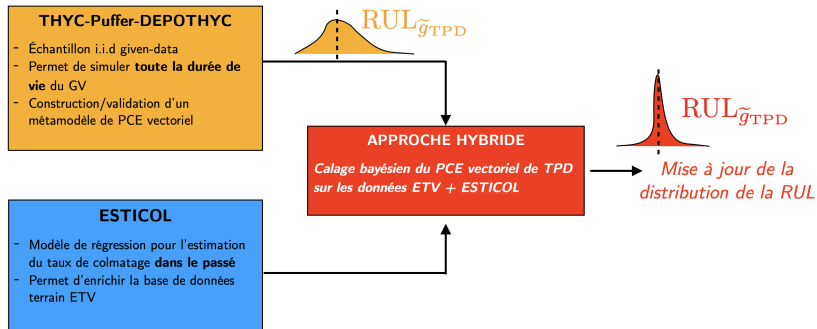


Figure – Schéma de la méthodologie hybride proposée.

a_v paramètre de calage du mécanisme de *vena contracta*

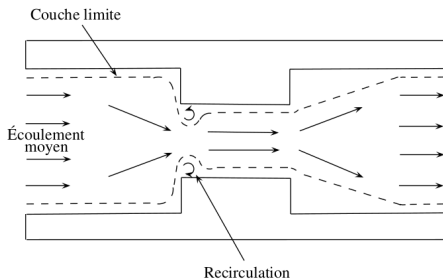


Figure – Mécanisme de vena contracta, source : [Prusek, 2012].

- ▶ Toutes les particules transportées en zone de recirculation sont déposées, d'où :

$$\Phi_p \propto K_p \Gamma_p$$

- ▶ La vitesse de déposition sur les parois K_p est donnée par :

$$K_p \propto a_v U^2 d_p^2 \rightarrow a_v \text{ à caler par rapport aux données terrain !}$$

Motivations méthodologie

- ▶ Travaux antérieurs d'analyse de sensibilité [Jaber et al., 2024], $a_v \rightarrow$ grande influence sur τ_c sur tout le temps de simulation
- ▶ Grande dispersion des trajectoires de colmatage autour de la date courante $t_p \rightarrow$ pose problème pour l'estimation d'une RUL
- ▶ Pratique actuelle \rightarrow méthodologie de calage déterministe COLMAREX [Prusek, 2012], vise à caler un a_v par GV *pour toute la durée de fonctionnement du GV*
- ▶ REX opérationnel \rightarrow cinétique de τ_c modifiée après un nettoyage chimique \rightarrow aspect non pris en compte par un a_v identique sur tout l'historique de fonctionnement + non prise en compte des incertitudes
- ▶ Question : que pourrait apporter le calage bayésien dans ce cas ?

Schéma général de la méthodologie de calage bayésien

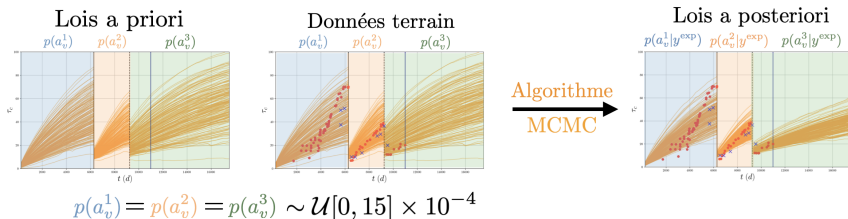


Figure – Principe général du calage bayésien.

- Littérature abondante de travaux en calage et inversion bayésienne à l'aide de métamodèle de PCE : [Marzouk et al., 2007; Marzouk and Najm, 2009; Yan and Zhou, 2019; Wagner et al., 2020]...

Calage bayésien du paramètre a_v

- ▶ On recale trois distributions de a_v sur chacune des 3 périodes → avant *nettoyage curatif* (CC), entre *nettoyage curatif et nettoyage préventif* (CC-PC) et après le *nettoyage préventif* (PC) → choix justifié par la modification de la cinétique observée après un NC
- ▶ Utilise le métamodèle de PCE \tilde{g}_{TPD} → procédure basée sur $m = n + 1$ calages où n est le nombre de NC réalisés sur un GV (dans notre exemple $m = 3$)
- ▶ Données terrain entre le k -ème et $k + 1$ -ème NC :

$$\{\mathbf{y}_k^* = (y_{k,1}^*, \dots, y_{k,n_k}^*)\} \subset \{\mathbf{y}_1^*, \dots, \mathbf{y}_m^*\} \quad (5)$$

$\mathcal{J}_{*,k}$ sont les indices de pas de temps respectif, $n_{*,k} := |\mathcal{J}_{*,k}|$ avec $* = \{\text{ETV}, \text{EST}\}$

- ▶ \mathbf{x}_0 correspond à la valeur *nominale* des paramètres incertains \mathbf{X} . Le paramètre de calage a_v^k est noté θ_k

Calage bayésien du paramètre a_v

- ▶ Sans discrédance de modèle on suppose [Carmassi et al., 2019] pour $k = 1, \dots, m$, avec $* = \{\text{ETV}, \text{EST}\}$:

$$\mathbf{y}_k^* = \mathcal{G}_k^*(\theta_k) + \boldsymbol{\eta}_k^*, \quad \boldsymbol{\eta}_k^* \sim \mathcal{N}(0, \sigma_*^2 I_{n_{*,k}}) \quad (6)$$

où $\mathcal{G}_k^*(\cdot)$ est la projection des sorties du métamodèle $\tilde{g}_{\text{TPD}}(\mathbf{x}_0, \cdot)$ sur les pas de temps de la $*$ -ème donnée entre le k et $k + 1$ -ème nettoyage

- ▶ Choix des priors pour tout k :

- ▶ $\theta_k \sim \mathcal{U}[0, 15] \times 10^{-4}$
- ▶ prior de Jeffreys pour $v := 1/\sigma_*^2$, $p(v) = 1/v$
- ▶ θ_k et $\boldsymbol{\eta}$ sont indépendants \rightarrow résidus donnent une vraisemblance gaussienne

- ▶ Si toutes les données terrain ont même écart-type alors on peut montrer que [Keller et al., 2022] :

$$p(\theta_k | \mathbf{y}_k^*) \propto \|\mathbf{y}_k^* - \mathcal{G}_k^*(\theta_k)\|^{-n_{*,k}} \quad (7)$$

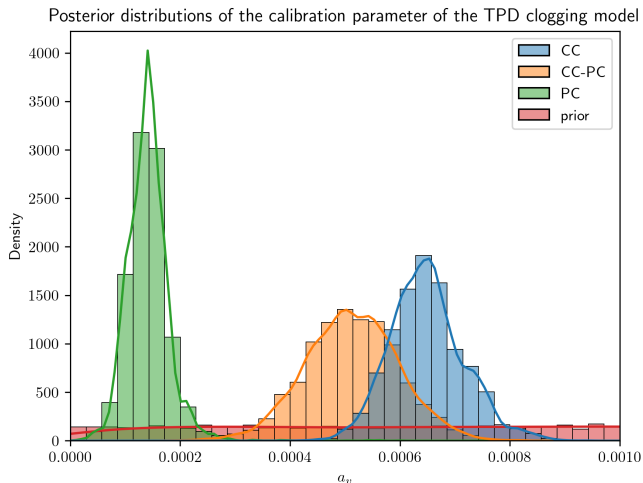
Calage bayésien du paramètre a_v

- ▶ On peut généraliser cette loi a posteriori pour q groupes de données terrain ($\mathbf{y}^{\text{exp},1}, \dots, \mathbf{y}^{\text{exp},q}$) de variances différents (dans notre cas $q = 2$ car ETV et ESTICOL n'ont pas même variance) pour $k = 1, \dots, m$:

$$p(\theta_k | \mathbf{y}_k^{\text{ETV}}, \mathbf{y}_k^{\text{EST}}) \propto \|\mathbf{y}_k^{\text{ETV}} - \mathcal{G}_k^{\text{ETV}}(\theta_k)\|^{-n_{\text{ETV},k}} \times \|\mathbf{y}_k^{\text{EST}} - \mathcal{G}_k^{\text{EST}}(\theta_k)\|^{-n_{\text{EST},k}}$$

- ▶ Le poids associé au k -ème type de donnée est lié au nombre de données disponibles $n_{\text{ETV},k}, n_{\text{EST},k}$
- ▶ Algorithme MCMC de type Random-Walk Metropolis-Hastings dans OpenTURNS [[Rubinstein and Kroese, 2011](#); [Baudin et al., 2017](#)] → échantillonnage dans des distributions $p(\theta_k | \mathbf{y}_k^{\text{ETV}}, \mathbf{y}_k^{\text{EST}})$ + test de convergence de Gelman-Rubin pour les chaînes de Markov (voir annexe)

Résultats numériques pour le calage de la chaîne TPD



- On trouve bien des modes différents en fonction des différentes cinétiques après NC.

Repropagation des incertitudes dans le PCE

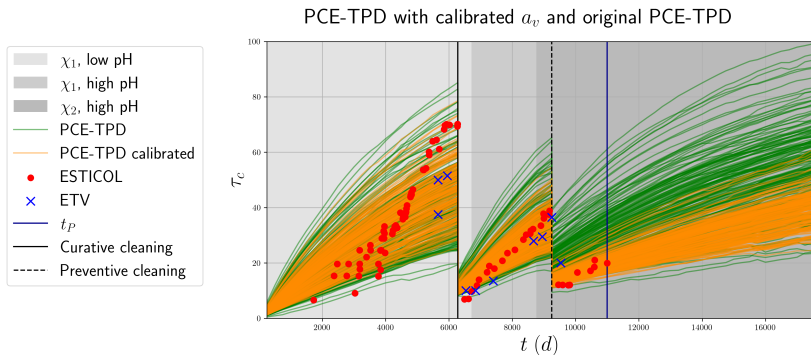


Figure – Simulations du métamodèle PCE avec a_v recalé

- ▶ Diminution de la dispersion du faisceau de trajectoires \rightarrow dispersion nettement réduite dans la dernière période \rightarrow peu de données mais période *utile* pour la prévision future
- ▶ Les incertitudes en prévision du PCE-TPD pour $t > t_P$ sont ainsi *réduites*

Sommaire

Colmatage des GV et motivations

Outils à disposition

Modèle physique : THYC-Puffer-DEPOTHYC

Modèle statistique : ESTICOL

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Conclusion et perspectives

- ▶ Le colmatage des GV est un phénomène complexe et de dynamique très lente → utilisation d'outils de modélisation physique et statistique ont pour objectif d'appuyer la planification future des maintenances
- ▶ Poursuite des travaux sur le calage avec un point de vue réduction d'incertitudes pour le pronostic
- ▶ Deux axes principaux de travail sur le calage bayésien :
 - ▶ Soit prendre en compte les incertitudes des paramètres d'entrée ;
 - ▶ Soit recalculer toutes les variables influentes (\mathbf{X}, θ) .

De plus : développer un métamodèle de GP et inclure un biais de métamodèle comme dans [Kennedy and O'Hagan, 2001] ;

Plus loin : exploration de pistes d'hybridation à travers l'utilisation de données de capteurs.

Travaux en cours : prise en compte des incertitudes dans \mathbf{X}

- ▶ En toute rigueur, la densité a posteriori simulée est $p(\theta_m | Y = y^{\text{exp}}, \mathbf{X} = \mathbf{x}_0)$, avec \mathbf{x}_0 correspondant à la valeur nominale des entrées de TPD hors le paramètre de calage
- ▶ L'hypothèse du plan d'expérience [Jaber et al., 2024] supposait l'indépendance des variables d'entrée $(\mathbf{X}, \theta) \sim p(\mathbf{X}) \times p(\theta)$, toutefois cette indépendance ne se propage pas sur les lois conditionnelles :

$$p(\theta | Y, \mathbf{X}) = \frac{p((\theta | Y), \mathbf{X})}{p(\mathbf{X})} \neq \frac{p(\theta | Y) \times p(\mathbf{X})}{p(\mathbf{X})}$$

- ▶ Une approche possible : intégration de la loi a posteriori [Perrin and Durantin, 2019] :

$$p(\theta | Y) = E_{\mathbf{X}}[p(\theta | Y, \mathbf{X})] = \int_{\mathcal{X}} p(\theta | Y, \mathbf{X} = \mathbf{x}) p(\mathbf{x}) d\mathbf{x} \quad (8)$$

- ▶ Objectif de travaux en cours → de rendre encore plus robuste le pronostic par une meilleure prise en compte des incertitudes

Sommaire

Colmatage des GV et motivations

Outils à disposition

Modèle physique : THYC-Puffer-DEPOTHYC

Modèle statistique : ESTICOL

Pronostic hybride par calage Bayésien

Conclusion et perspectives

Annexes

Graphes d'autocorrélation

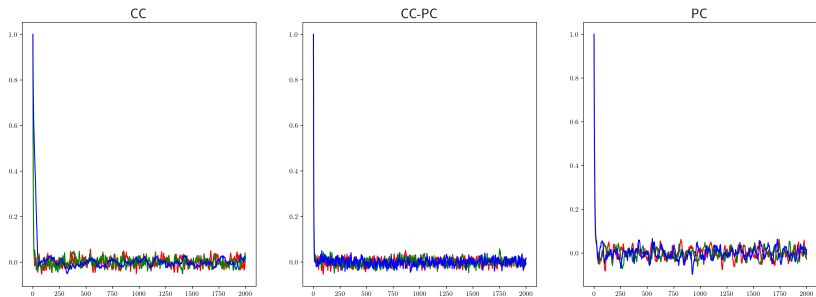


Figure – Graphes d'autocorrélation des trois chaînes de Markov sur les trois temps de simulation.

Statistiques de Gelman-Rubin

- ▶ On initialise J chaînes de Markov de manière indépendante (dans notre cas $J = 3$). Après la phase de *burn-in*, on a des échantillons $\theta_1^{(i)}, \dots, \theta_L^{(i)}$ pour $i = 1, \dots, J$ avec L le nombre d'itérations de la chaîne

- ▶ Calcule de la moyenne sur la chaîne, entre la chaîne

$$\bar{\theta}_i = \frac{1}{L} \sum_{k=1}^L \theta_k^{(i)}, \quad \bar{\theta}_* = \frac{1}{J} \sum_{i=1}^J \bar{\theta}_i$$

- ▶ Calcule de la variance des moyennes des chaînes et la moyenne des variances au sein d'une chaîne :

$$B = \frac{L}{J-1} \sum_{i=1}^J (\bar{\theta}_* - \bar{\theta}_i)^2, \quad W = \frac{1}{J} \sum_{i=1}^J \left(\frac{1}{L-1} \sum_{k=1}^L (\theta_k^{(i)} - \bar{\theta}_i)^2 \right)$$

- ▶ Enfin le test de Gelman-Rubin [Gelman and Rubin, 1992] se calcule selon :

$$R = \frac{(1 - 1/L)W + (1/L)B}{W} \rightarrow 1, \quad L \rightarrow \infty$$

Tests de GR et acceptation

- Paramètres MCMC : $J = 3$, marche aléatoire uniforme, initialisation aléatoire, taille des pas de la marche aléatoire 10^{-3}

	CC	CC-PC	PC
Taux d'acceptation moyen	0.195	0.273	0.272
Test de Gelman-Rubin	1.00	1.00	1.00

Table – Taux d'acceptation moyen et test de Gelman-Rubin.

References I

- Baudin, M., Dutfoy, A., Iooss, B., and Popelin, A. (2017). Open TURNS : An industrial software for uncertainty quantification in simulation. In Ghanem, R., Higdon, D., and Owhadi, H., editors, *Springer Handbook on Uncertainty Quantification*, pages 2001–2038. Springer.
- Bregon, A. and Daigle, M. J. (2019). *Fundamentals of Prognostics*, pages 409–432. Springer International Publishing.
- Carmassi, M., Barbillon, P., and Chiodetti, M. (2019). Bayesian calibration of a numerical code for prediction. *Journal de la société Française de statistique*, 160(1).
- Feng, Q., Nebes, J., Bachet, M., Pujet, S., You, D., and Deri, E. (2023). Tube support plates blockage of PWR steam generators : thermalhydraulics and chemical modeling.
- Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4) :457 – 472.

References II

- Jaber, E., Chabridon, V., Remy, E., Baudin, M., Lucor, D., Mougeot, M., and looss, B. (2024). Sensitivity analyses of a multi-physics long-term clogging model for steam generators. *International Journal for Uncertainty Quantification*.
- Keller, M., Damblin, G., Pasanisi, A., Schumann, M., Barbillon, P., Ruggeri, F., and Parent, E. (2022). Validation of a computer code for the energy consumption of a building, with application to optimal electric bill pricing. *Econometrics*, 10(4).
- Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63(3) :425–464.
- Lefebvre, L., Segond, M., Spaggiari, R., Le Gratiet, L., Deri, E., looss, B., and Damblin, G. (2023). Improving the Predictivity of a Steam Generator Clogging Numerical Model by Global Sensitivity Analysis and Bayesian Calibration Techniques. *Nuclear Science and Engineering*, 197(8) :2136–2149.

References III

- Marzouk, Y. M. and Najm, H. N. (2009). Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems. *Journal of Computational Physics*, 228(6) :1862–1902.
- Marzouk, Y. M., Najm, H. N., and Rahn, L. A. (2007). Stochastic spectral methods for efficient Bayesian solution of inverse problems. *Journal of Computational Physics*, 224(2) :560–586.
- Perrin, G. and Durantin, C. (2019). Taking into account input uncertainties in the Bayesian calibration of time-consuming simulators. *Journal de la Société Française de Statistique*, 160(2).
- Pincioli, L., Baraldi, P., Shokry, A., Zio, E., Seraoui, R., and Mai, C. (2021). A semi-supervised method for the characterization of degradation of nuclear power plants steam generators. *Progress in Nuclear Energy*, 131 :103580.
- Prusek, T. (2012). *Modélisation et simulation numérique du colmatage à l'échelle du sous-canal dans les générateurs de vapeur*. Thèse de l'Université Aix-Marseille.

References IV

- Rubinstein, R. and Kroese, D. (2011). *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Statistics. Wiley.
- Soize, C. and Ghanem, R. (2004). Physical systems with random uncertainties : Chaos representations with arbitrary probability measures. *SIAM Journal On Scientific Computing*, 26(2) :395–410.
- Wagner, P.-R., Fahrni, R., Klippel, M., Frangi, A., and Sudret, B. (2020). Bayesian calibration and sensitivity analysis of heat transfer models for fire insulation panels. *Engineering Structures*, 205 :110063.
- Yan, L. and Zhou, T. (2019). Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems. *Journal of Computational Physics*, 381 :110–128.