

## **Bayesian predictive inference of acidity in high-level liquid waste concentrators at La Hague**

Stéphane PUYDARRIEUX<sup>1</sup>, Ludovic MAKONG<sup>1</sup>, Frédéric LEPRÊTRE<sup>1</sup>, Maël CAYREFOURCQ<sup>2</sup>, Kévin TAOUDI<sup>2</sup>, Pierre GUENOUN<sup>2</sup>

1 : Orano Recyclage, La Hague, France

2 : Probayes, Montbonnot-Saint-Martin, France

### **Abstract**

In the reprocessing of spent nuclear fuel, concentrating fission product solutions, the primary source of radioactivity, is very important before vitrification. Evaporators heat these solutions while maintaining a consistent acidity level (between 2 N and 2.8 N) to effectively remove excess nitric acid. Given the complexity of the underlying phenomena, inherent industrial variability, and the limited number of laboratory acidity measurements, precise predictive modeling of acidity is essential for optimal evaporator operation.

Therefore, our work focuses on the quasi-continuous predictive modeling of acidity and its associated uncertainty. This approach integrates five years of historical operational data from the evaporators, along with laboratory acidity measurements, within a Bayesian framework. This methodology aims to ensure effective, reliable, and continuous control of the process throughout the evaporator's operation.

## **Inférence bayésienne prédictive de l'acidité des évaporateurs de produits de fission du site de La Hague**

Stéphane PUYDARRIEUX<sup>1</sup>, Ludovic MAKONG<sup>1</sup>, Frédéric LEPRÊTRE<sup>1</sup>, Maël CAYREFOURCQ<sup>2</sup>, Kévin TAOUDI<sup>2</sup>, Pierre GUENOUN<sup>2</sup>

1 : Orano Recyclage, La Hague, France

2 : Probayes, Montbonnot-Saint-Martin, France

### **Résumé**

Dans le retraitement du combustible nucléaire usé, concentrer les solutions de produits de fission, principale source de radioactivité, est une opération très importante avant leur vitrification. Les évaporateurs chauffent ces solutions tout en maintenant une acidité constante (entre 2 N et 2.8 N) pour éliminer efficacement l'excès d'acide nitrique. Étant donné la complexité des phénomènes en jeu, la variabilité industrielle inhérente et le nombre restreint de mesures d'acidité issues du laboratoire, une modélisation prédictive précise de l'acidité est souhaitée pour optimiser la conduite des évaporateurs.

Par conséquent, notre travail portera sur la modélisation prédictive quasi-continue de l'acidité et son incertitude associée. Cette approche intégrera un historique de cinq années de données chronologiques de fonctionnement des évaporateurs ainsi que des mesures d'analyse en laboratoire de l'acidité, le tout dans un cadre bayésien. Cette méthodologie vise à garantir un contrôle efficace, fiable et continu du procédé tout au long de l'exploitation de l'évaporateur.

## Introduction

Le site industriel d'ORANO La Hague joue un rôle crucial dans la gestion durable des combustibles nucléaires usés provenant des centrales françaises et étrangères, ainsi que de certains autres établissements. Ce site est spécialisé dans le retraitement de ces combustibles, via un processus complexe qui permet de séparer les matières valorisables, comme l'uranium et le plutonium, d'autres éléments chimiques appelés « produits de fission ».

Grâce à une série d'opérations mécaniques et chimiques complexes, l'uranium et le plutonium sont récupérés et envoyés vers d'autres installations industrielles pour être réutilisés dans la fabrication de nouveaux combustibles nucléaires, contribuant ainsi à la réduction des déchets et à l'optimisation des ressources.

Les produits de fission, en revanche, sont incorporés dans une matrice en verre par un procédé de vitrification, qui garantit leur confinement. Avant cette étape cruciale, les produits de fission sont d'abord concentrés dans des évaporateurs afin de limiter les volumes à gérer et de produire une solution pleinement compatible avec le procédé de vitrification.

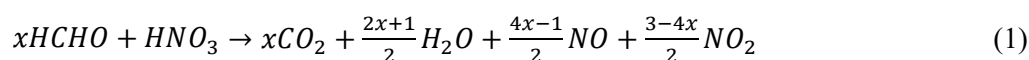
Le procédé de concentration des produits de fission dans les évaporateurs présente une dynamique de fonctionnement complexe. Cette complexité résulte des multiples interactions physico-chimiques influencées par la diversité des éléments réactifs en jeu et par une forte variabilité inhérente au procédé lui-même. Le pilotage du procédé repose principalement sur des mesures d'acidité issues d'analyses laboratoire réalisées toutes les huit heures. Ces résultats d'analyses comportent une part d'incertitude, parfois des biais, voire potentiellement des erreurs. Par ailleurs, le procédé est équipé de nombreux capteurs qui offrent une surveillance en quasi temps réel de la dynamique, mais ne sont pas exploités de manière optimale pour la maîtrise de l'acidité.

Afin de renforcer la fiabilité du pilotage de ce procédé, nos travaux se sont concentrés sur la prévision quasi-continue de l'acidité par des approches bayésiennes, qui offrent un cadre mathématique idéal pour intégrer toutes les connaissances historiques et a priori sur le procédé. Cet article présentera d'abord le procédé et les équations chimiques en jeu, puis leur modélisation pour quantifier l'acidité dans un cadre bayésien, avec une analyse comparative des différentes méthodes étudiées. Enfin, nous détaillerons la méthode retenue ainsi que les perspectives futures.

## Descriptif du procédé

Afin de concentrer ces produits de fission, qui se présentent en solution dans de l'acide nitrique, le procédé de dénitrification formique est mis en œuvre. Comme illustré dans la Figure 1, il s'agit d'un procédé en semi-batch dans lequel on alimente les produits de fission en continu dans un évaporateur tout en envoyant du formol dans l'équipement.

Ce réactif permet de décomposer l'acide nitrique en vapeur d'eau, vapeurs nitreuses et gaz carbonique comme suit :



où le facteur  $x$  est fonction de la température et de l'acidité (favorisant soit la formation de NO soit de NO<sub>2</sub> et se situe entre 0,25 et 0,75).

La maîtrise des débits d'alimentation en solutions de produits de fission et en formol permet de maintenir la concentration en acide dans l'évaporateur aux alentours des 2.3 N. Il n'y a pas d'accumulation d'acide dans l'évaporateur ce qui pourrait être très préjudiciable en termes de corrosion pour l'équipement. L'évaporateur est chauffé pendant l'alimentation en solutions pour vaporiser une partie du liquide. Les gaz produits (issus de la vaporisation du liquide et de la dénitrification formique) sont traités dans d'autres équipements puis recyclés

dans le procédé. Une fois la concentration voulue atteinte, l'évaporateur est refroidi et les solutions concentrées sont vidangées vers les unités du procédé de vitrification.

Lors de la concentration, les quantités de formol sont ajustées périodiquement en fonction des analyses chimiques des prélèvements réalisés dans l'évaporateur. Si l'acidité n'est pas à l'attendu, l'exploitant corrige le débit de formol pour se rapprocher au plus près de la valeur d'acidité souhaitée entre 2N et 2.8N.

Afin d'avoir une estimation quasi temps réel de la teneur en acide dans l'évaporateur, il a été proposé de développer un modèle bayésien prédictif. Celui-ci se nourrit des paramètres procédés relevés en continu (pressions, températures, débits, etc...) dans l'évaporateur et ses équipements connexes. L'objectif est de pouvoir réaliser une prévision fiable de l'acidité dans l'évaporateur, à partir de l'historique du cycle de concentration, afin de la comparer voire de la combiner aux résultats de l'analyse laboratoire.

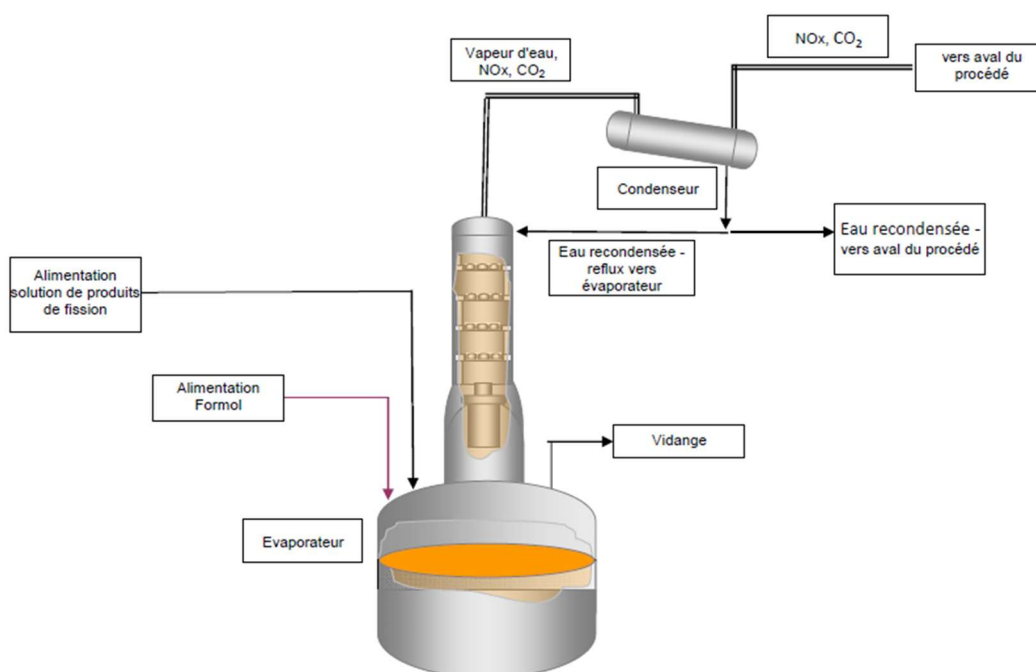


Figure 1. Schéma de fonctionnement du procédé de concentration des produits de fission

## Modélisation de l'acidité des évaporateurs

Nos travaux visent à faire une analyse des différentes mesures laboratoires pour réaliser une prédiction de cette mesure et ainsi diminuer l'incertitude de mesure associée. Nous travaillons sur la prédiction de l'acidité dans un évaporateur de la chaîne de traitement du combustible usé à partir (1) des données du procédé, (2) des valeurs historiques ainsi que (3) de paramètres suivis en temps réel.

### 1. Mise en équation du problème chimique

L'acidité dans l'évaporateur étudié répond tout d'abord à une équation chimique liant les produits de fission entrants (qui doivent être concentrés et dont la concentration présente une certaine variabilité) et du formol (ajouté indépendamment permettant de stabiliser cette acidité). Cette connaissance permet d'établir des principes de conduite de l'évaporateur afin de coller au mieux aux conditions favorables pour stabiliser l'acidité autour de la valeur cible (ces principes sont globalement la stabilisation du pH par le panachage des différentes origines de l'acide nitrique, abaques et autres connaissances pour l'introduction du formol dans le réacteur etc....)

La réaction de destruction de l'acide nitrique par le formol a été présentée en (1).

Le facteur  $x$  est fonction de la température et de l'acidité (favorisant soit la formation de  $NO$  soit de  $NO_2$  et se situe entre 0,25 et 0,75)

L'expertise apportée par les ingénieurs procédés d'Orano nous a permis de reformuler l'équation de réaction ci-dessus en explicitant le bilan matière. Concrètement, la transformation chimique est régie par la formule suivante :

$$Qté\ nitrate_t = Qté\ nitrate_{t-1} + Qté\ nitrate_{entrante} - reflux - FC * Qté\ formol \quad (2)$$

Où FC est le facteur de consommation de formol qui décrit la consommation de l'acide nitrique par le formol

Toutes les quantités sont mesurées par des capteurs en ligne et/ou des relevés laboratoires. L'inconnue principale est donc le facteur de consommation du formol. A partir d'un abaque issue d'une loi expérimentale, Orano estime approximativement ce facteur.

Dans l'optique de mesurer l'impact de l'apport de la science des données dans l'étude de cette réaction chimique, nous souhaitons évaluer la pertinence de la loi expérimentale. Nous avons donc calculé un FC dit empirique en utilisant des données réelles passées.

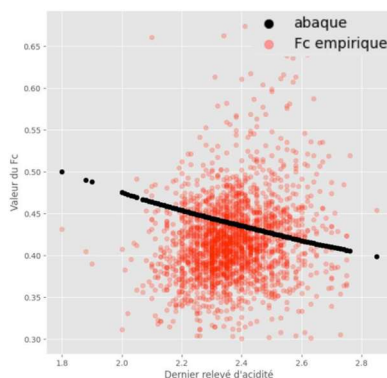


Figure 2. Lien entre le facteur de consommation du formol et l'acidité dans la cuve (abaque théorique vs résultats empiriques).

Il est clair que les résultats obtenus divergent substantiellement de la réalité observée (Figure 2). Cette observation a mené à l'hypothèse que certaines variables d'état du système pourraient expliquer les écarts par rapport aux prédictions de l'abaque théorique. Cela matérialise l'écart de comportement entre une expérience « in-vitro » de la situation réelle de l'exploitant.

## 2. Modèle

La difficulté provient du fait que :

L'ensemble de ces paramètres est sensible aux conditions d'utilisation : pression, température, conditions environnementales et chimiques etc...

Les acidités/concentrations/quantités intermédiaires ne sont pas connues assez précisément pour que la modélisation explicite physico-chimique soit aussi précise que souhaité (plus précisément, la conduite du procédé est déjà optimisée),

Chacune des mesures réalisées est elle-même sujette à caution car comporte une erreur ou incertitude intrinsèque,

Les mesures sont parfois difficiles à relier à l'expérimentation en cours du fait du passage par des cuves de stockage et circuits intermédiaires qui causent de légers effets de stagnation des réactifs avant introduction dans l'évaporateur (contexte de généalogie).

L'ambition est donc, à la lumière des teneurs connues, des paramètres utilisés et des points de mesure réalisés précédemment, de prédire cette acidité et de confronter ensuite cette prédiction à la mesure afin de réduire l'incertitude globale de la variable "acidité". Il est cependant difficile et coûteux de mesurer l'acidité dans la cuve ciblée et cela présente différents défauts : incertitudes (+/-5% d'erreur), instantanéité (calcul toutes les 8h), faisabilité (action de faire un prélèvement et de l'amener au laboratoire). Pour contourner cette difficulté, nous avons développé un modèle permettant d'estimer finement l'acidité dans l'évaporateur. Nous disposons de données issues de capteurs mais aussi de mesures laboratoires. Chaque information a une incertitude de mesure. D'autre part nous disposons d'une loi physique permettant de lier l'incertitude à un temps  $t$  avec d'autres informations dont nous disposons. Nous faisons le choix de modéliser les incertitudes par des gaussiennes. Un cadre théorique permettant de répondre à notre objectif tout en utilisant les spécificités du projet est une résolution bayésienne du problème. Le facteur de consommation du formol, désigné par le coefficient  $FC$ , constitue la seule variable inconnue dans notre modèle chimique. Afin d'approximer ce coefficient, nous nous appuyons sur une loi expérimentale déterminée en laboratoire. Cependant, l'importance de ce paramètre pour l'exactitude de nos prévisions a été mise en évidence par une étude de sensibilité approfondie, soulignant ainsi son rôle critique dans la prédiction précise de l'acidité dans l'évaporateur.

Dans le but de consolider les estimations du  $FC$ , nous avons fait le choix d'intégrer des informations pouvant impacter la qualité de la transformation chimique. C'est pourquoi nous souhaitons estimer le facteur de consommation du formol comme une fonction de plusieurs variables, définie en (3) :

$$\widehat{FC} = f(X, \alpha, \beta), \quad X \in \mathcal{E}, \alpha \in \mathbb{R}, \beta \in \mathbb{R}^p, p \text{ le nombre de paramètres procédé} \quad (3)$$

$\mathcal{E}$ , l'ensemble des mesures procédés utilisées

Comme évoqué précédemment, compte tenu des spécificités des données utilisées, une incertitude de mesure forte, nous allons optimiser notre problème dans un cadre bayésien. Nous cherchons donc à résoudre l'équation (4) :

$$\underset{\alpha, \beta \in (\mathbb{R}, \mathbb{R}^p)}{\operatorname{argmin}} \quad \|\text{Qté nitrate}_t - \text{Qté nitrate}_{t-1} - \text{Qté nitrate}_{\text{introduit}} + \text{reflux} - f(X, \beta) * \text{Qté formol}\| \quad (4)$$

Nous avons comparé deux modèles différents, un modèle additif et un modèle multiplicatif. Cette différence est caractérisée par la fonction qui estime le facteur de consommation du formol. Précédemment nous avons évoqué que certaines mesures d'état du système pourraient impacter le déroulé du procédé chimique, c'est pourquoi nous avons choisi d'intégrer ces informations  $X$  à la modélisation.

*Modèle multiplicatif :*

$$\forall X \in \mathcal{E}, \forall \alpha, \beta \in (\mathbb{R}, \mathbb{R}^p), \quad f(X, \alpha, \beta) = \exp(\alpha + X * \beta) \quad (5)$$

*Modèle additif :*

$$\forall X \in \mathcal{E}, \forall \alpha, \beta \in (\mathbb{R}, \mathbb{R}^p), \quad f(X, \alpha, \beta) = (\alpha + X * \beta) \quad (6)$$

Le modèle exponentiel se rapproche de l'abaque théorique, cependant nous avons fait face à des problèmes de stabilité numérique, c'est pourquoi nous avons souhaité le comparer à un modèle additif.

Maintenant que le problème a été mis en équation, il nous faut poser des hypothèses pour utiliser notre modèle bayésien. Le modèle élaboré a été confronté à diverses incertitudes, notamment :

- La légitimité des hypothèses de stationnarité et homogénéité des évaporateurs

- La complétion des variables explicatives contribuant aux écarts d'acidité subis, autrement dit : est-ce que la variabilité de l'acidité observée peut être expliquée par le bilan matière uniquement ?

Concernant le point 1, un test de stationnarité de Dickey Fuller a été effectué, cependant, au vu de l'écart conséquent entre deux mesures (~8h) alors que la réaction est supposée instantanée, nous n'avions pas de garanties quant à la bonne convergence du modèle.

### Résolution numérique

Pour résoudre notre problème bayésien, nous avons utilisé comme calculateur le progiciel pymc3 développé par Salvatier, J depuis 2016. Le solveur ADVI développé par Kucukelbir a été choisi en raison de ses avantages distincts pour modéliser des distributions postérieures complexes dans des contextes où l'inférence bayésienne classique devient inapplicable en raison des coûts computationnels élevés. Contrairement aux méthodes de Monte Carlo par chaîne de Markov (MCMC), qui peuvent être lentes et difficiles à calibrer, l'inférence variationnelle permet une approximation plus rapide et déterministe des distributions a posteriori en transformant le problème en une optimisation. Cependant, cette approche impose des contraintes importantes, notamment dans les hypothèses faites sur la forme de la distribution a posteriori, souvent en choisissant une famille paramétrique spécifique. Alors que MCMC construit un échantillon empirique au cours des itérations, permettant une plus grande flexibilité pour capturer la complexité de la distribution, l'inférence variationnelle peut introduire des biais si la forme de la distribution choisie n'est pas suffisamment flexible pour représenter fidèlement la distribution a posteriori réelle. Malgré ces limitations, l'inférence variationnelle reste une alternative robuste et pratique pour l'inférence statistique dans des systèmes complexes, offrant une scalabilité et une rapidité de calcul supérieures. Dans notre contexte, l'ADVI a notamment été choisie car nous disposons d'un bon a priori métier sur la famille de distribution a posteriori. En outre, les hypothèses sur les distributions a posteriori des estimateurs, et notamment l'hypothèse d'indépendance des estimateurs dans notre cas, seront discutées afin d'évaluer leur impact sur la précision et la fiabilité des résultats obtenus.

Nous avons donc fixé comme à priori que nos estimateurs du FC étaient tous indépendants, et étant donné qu'ils n'avaient pas de raison à être tous positifs, nous avons pris pour hypothèse que l'ensemble de notre vecteur d'estimateur suit une loi normale multivariée de matrice de covariance diagonale (i.e nos estimateurs sont indépendants).

$$\beta \sim N(\mu, \Sigma), \text{ avec } \mu \in \mathbb{R}^p, \Sigma \in M_{p,p}(\mathbb{R}) \text{ diagonal} \tag{7}$$

Nous avons favorisé ce type de modélisation pour intégrer une contrainte spécifique aux données. L'acidité en entrée est également mesurée toutes les 8h, et elle est donc fortement incertaine. La Figure 3 montre par exemple la sensibilité de la modélisation par simple estimation chimique en fonction de la valeur initiale paramétrée. L'idée d'utiliser soit une inférence bayésienne classique (Monte Carlo) soit une optimisation par ADVI, est de pouvoir intégrer et propager des incertitudes de mesures lors de l'optimisation. Ce point nous semblait crucial au vu de la sensibilité des estimations aux valeurs de l'acidité en entrée.

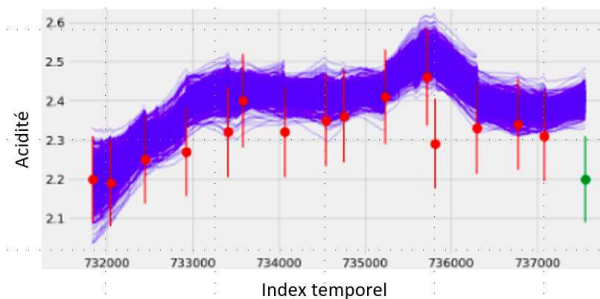


Figure 3. Etude de la sensibilité de la réaction chimique en fonction de l'acidité mesurée en entrée.

Nous avons donc intégré au modèle une variabilité à l'acidité en entrée. Nous l'avons caractérisée comme suivant :

$$\text{acid\_input} \sim N(\text{mesure}, \text{mesure} * 0.025) \quad (8)$$

Après avoir validé en menant des expériences avec un échantillonnage simple de Markov Chain Monte Carlo que les hypothèses (1) d'indépendance entre les estimateurs (matrice de Covariance diagonale) et (2) de lois normales pour les estimateurs étaient valides, nous avons exploité uniquement des algorithmes d'optimisation par inférence variationnelle : ADVI. L'avantage de l'ADVI sur le MCMC est la rapidité de convergence grâce à l'optimisation stochastique des paramètres d'une loi figée pour nos estimateurs.

Voici l'algorithme bayésien que nous avons mis en œuvre :

Loi à priori :

$$\alpha \sim N(\mu_\alpha, \Sigma_\alpha) \quad (9)$$

$$\beta \sim N(\mu_\beta, \Sigma_\beta) \quad (10)$$

$$\text{acid\_input} \sim N(\text{mesure}, \text{mesure} * 0.025) \quad (11)$$

Loi à posteriori :

$$\sigma_{\text{posterior}} \sim \text{IG}(\beta_\sigma, \alpha_\sigma) \quad (12)$$

$$\widehat{(\text{acid})} = \text{Qté nitrate}_{t-1} - \text{Qté nitrate}_{\text{introduit}} + \text{reflux} - (f(\widehat{X}, \alpha, \beta)) * \text{Qté formol} \quad (13)$$

$$\mu_{\text{post}} \sim N(\widehat{(\text{acid})}, \sigma_{\text{posterior}}) \quad (14)$$



## Résultats

ADVI vs MCMC :

En mêlant l'information chimique au modèle probabiliste, nous avons pris le parti d'évaluer l'intérêt d'utiliser un optimiseur tel que l'ADVI. Nous devons valider que les hypothèses prises étaient cohérentes avec la réalité métier.

Le coefficient de corrélation de Pearson calculé et présenté dans le Tableau 1 permet de valider l'hypothèses d'indépendance de nos estimateurs. Cette première hypothèse est donc validée.

Pour rappel ce coefficient se calcule comme suit (15)

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (15)$$

**Tableau 1.** Coefficient de corrélation de Pearson entre les différents estimateurs.

	<b>Perte de charge MCMC</b>	<b>Densité solution MCMC</b>	<b>Dernière mesure MCMC</b>
<b>Perte de charge ADVI</b>	1	0.03	-0.01
<b>Densité solution ADVI</b>	0.03	1	0.12
<b>Dernière mesure ADVI</b>	-0.01	0.12	1

En outre, il s'agit de vérifier que nos distributions empiriques issues du tirage de Monte Carlo suivent la loi à posteriori de nos estimateurs imposés lors de l'utilisation de l'inférence variationnelle.

Les itérations issues de l'échantillonnage par MCMC ont donné les résultats suivant pour nos estimateurs (Figure 4).

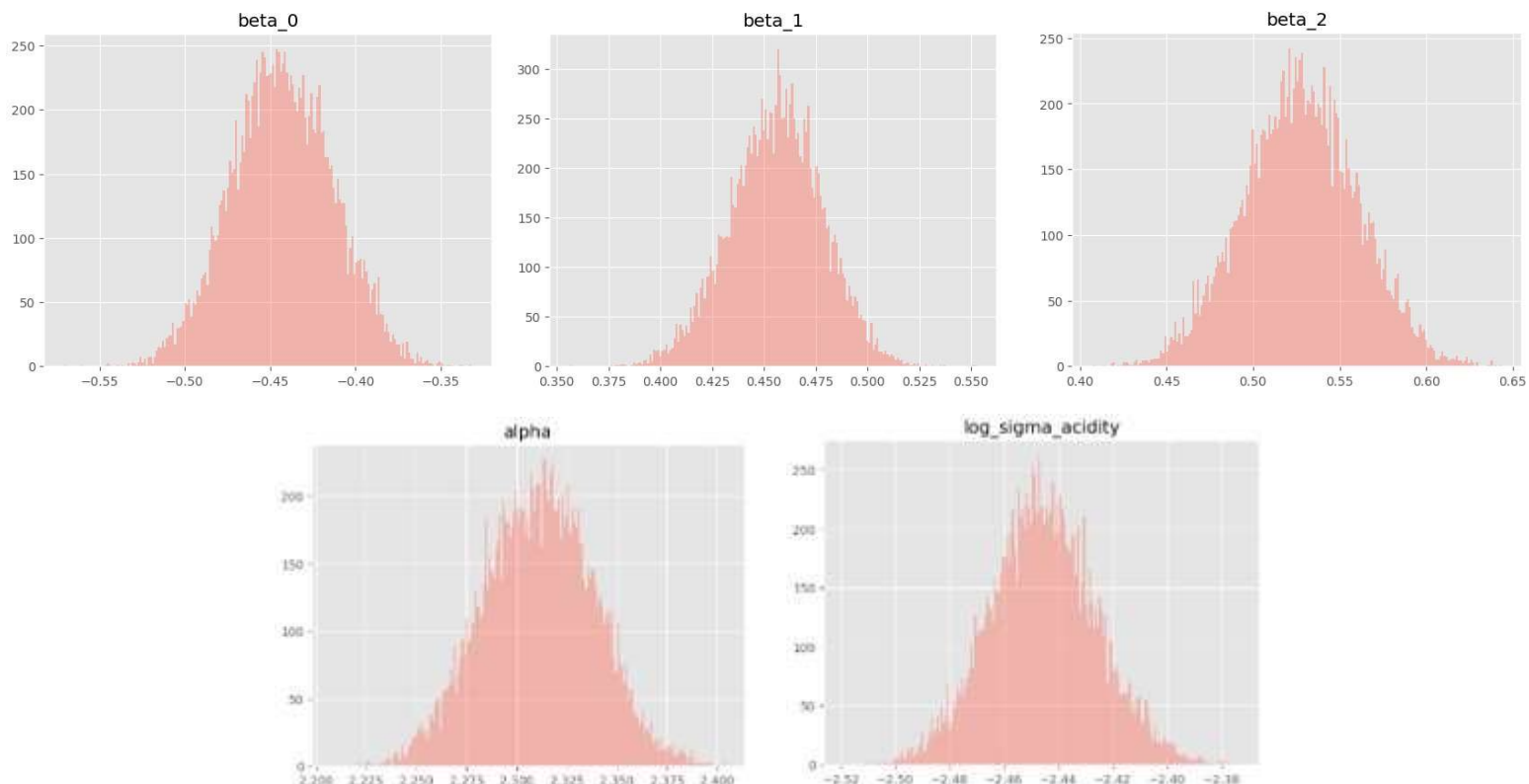


Figure 4. Distribution des estimateurs issues d'un échantillonnage par MCMC

Il s'avère que nos hypothèses, issues d'un apriori métier fort, sont validées. Les estimateurs sont bien indépendants (tab 1) et la loi empirique issue de l'échantillonnage suit bien une loi normale et suit la loi normale estimée par ADVI selon le test de Kolmogorov-Smirnov.

Dans le cadre d'un problème issu du monde du génie des procédés, et ayant de forts aprioris métiers sur le procédé, il peut être intéressant et pertinent d'utiliser un sampler ADVI.

Performances :

Nous avons comparé nos deux modèles à différentes méthodes d'estimations d'acidité dans la cuve cible :

- 1- Un modèle naïf qui consiste à prendre pour estimation de l'acidité à un instant  $t$  la dernière mesure connue

- $\forall t, \widehat{acid}_t = acid_{t-1}$

- 2- Un modèle issu de l'abaque théorique présenté en figure 1

- Pour cette méthode il s'agit d'appliquer la loi empirique issue de l'abaque. Rappelons que cette loi empirique met en lien l'acidité cible avec la valeur du facteur de consommation du formol.

- i.  $\widehat{FC} = f_{abaque}(\widehat{acidity\ cible})$

- ii.  $\widehat{acid}_t = Qté\ nitrate_{t-1} - Qté\ nitrate_{introduit} + reflux - \widehat{FC} * Qté\ formol$

À titre de comparaison, une pure démarche de type « data science » a été proposée pour évaluer les performances. L'implémentation d'un réseau de neurones MLP, optimisé sur la métrique d'erreur entre l'observation et la prédiction à l'instant suivant, a donné des résultats relativement mauvais. Ce constat s'explique aisément par les incertitudes présentes dans les observations, tant en entrée qu'en sortie, ainsi que par la difficulté à proposer un « feature engineering » pleinement informatif et descriptif du problème.

Dans la majeure partie de notre base de données, les variations de la valeur cible sont très faibles (la conduite de l'évaporateur est optimale), les changements sont donc souvent couverts par l'intervalle de confiance, rendant difficile l'évaluation précise de la performance du modèle.

Afin de palier à ce problème et évaluer l'apport de notre modèle pour la prédiction de l'acidité, nous avons testé nos prédictions d'abord sur le jeu de test entier, puis sur un sous-ensemble de notre base de données où les variations de la valeur cible sont supérieures à 0.05 entre deux valeurs successives.

Concrètement, nous avons évalué le modèle sur deux échantillons de données :

- $\chi_1 =$  Test data, intégralité des données de validation
- $\chi_2$ , sous échantillon du jeu de validation, tel que la variabilité absolue entre les deux relevés consécutifs soit supérieure à 0.05, i.e  $|PE_t - PE_{t-1}| > 0.05 \rightarrow$  Ce sont des mesures catégorisées comme difficiles à prédire.

Les résultats sont présentés sur la Figure 5 et les métriques sur le Tableau 2. Cette approche ciblée nous permet de concentrer notre évaluation là où les variations sont suffisamment marquées pour tester efficacement la précision de notre modèle.

*Mesure de la performance :*

Nous avons évalué les différents modèles à partir de deux métriques. Une RMSE (root mean squared error), et une métrique propre à notre problématique Pour évaluer la qualité des prédictions continues, nous introduisons une métrique, appelée, *Interval Tolerance Precision* (ITP), qui mesure la proportion de prédictions tombant dans un intervalle toléré autour des mesures laboratoires. Cette tolérance est contrôlée par un paramètre  $\tau$ , représentant le pourcentage d'écart admissible. Mathématiquement :

$$\forall \tau \in [0,0.5[ \text{ITP}(\tau) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(\hat{y}_i \in [y_i^*(1-\tau), y_i^*(1+\tau)])} \quad (16)$$

- $n$  est le nombre total d'observations de l'échantillon
- $y_i$  est la valeur réelle pour l'observation  $i$
- $\hat{y}_i$  est la prédiction associée
- $\mathbf{1}(\cdot)$  est une fonction indicatrice qui vaut 1 si la condition est vraie, 0 sinon

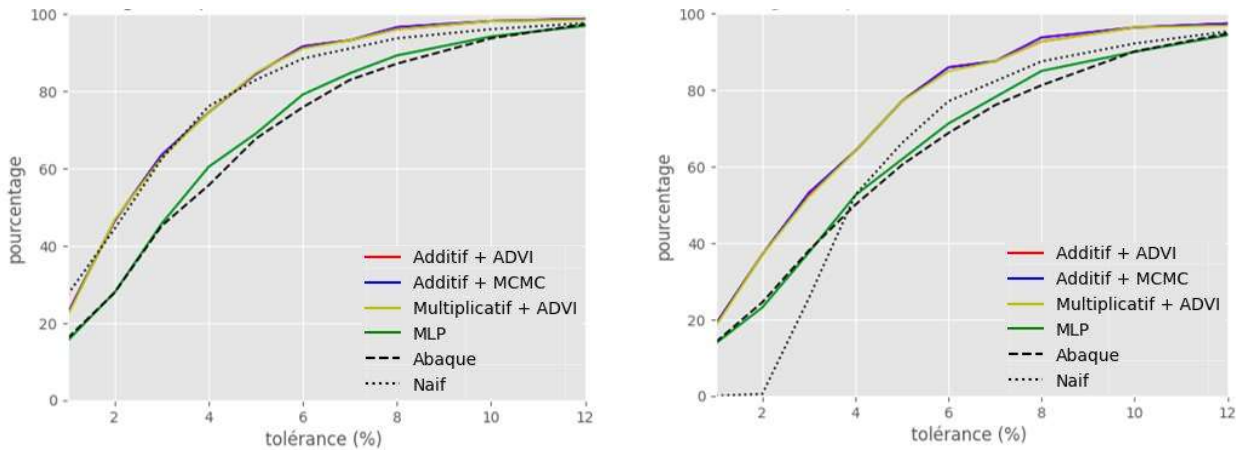


Figure 5. Performance des différents modèles sur l'échantillon de test complet  $\chi_1$  (à gauche) et sur le sous-échantillon  $\chi_2$  tel que  $|PE_t - PE_{t-1}| > 0.05$  (à droite)

Note de lecture : Pour le modèle Additif optimisé grâce au solveur ADVI, sur l'échantillon  $\chi_1$  84.5% des prédictions sont dans l'intervalle +/-5% autour de la mesure laboratoire (à gauche). Sur l'échantillon  $\chi_2$ , 77% des prédictions sont dans cet intervalle (à droite).

Ci-dessous ce tableau recense les performances de chaque méthode en considérant notre métrique pour une tolérance de  $\tau = 5\%$ .

	RMSE $\chi_1$	ITP $\chi_1$	RMSE $\chi_2$	ITP $\chi_2$
<b>Additif + ADVI</b>	0.088	84.5	0.10	77
<b>Additif + MCMC</b>	0.088	84.5	0.10	77
<b>Multiplicatif + ADVI</b>	0.088	84.8	0.10	77
<b>MLP</b>	0.12	69	0.14	63
<b>Abaque</b>	0.12	68	0.14	62
<b>Naïf</b>	0.096	82	0.13	66

Tableau 2. Performances de nos modèles.

Note de lecture : Pour le modèle Additif optimisé grâce au solveur ADVI, sur l'échantillon  $\chi_1$  84.5% des prédictions sont dans l'intervalle +/-5% autour de la mesure laboratoire. Sur l'échantillon  $\chi_2$ , 77% des prédictions sont dans cet intervalle.

L'apport de notre méthode de résolution et du modèle bayésien ici est clair. Elle permet plus de flexibilité dans les données en entrée, en outre, nos deux modèles bayésiens prennent en compte les incertitudes de mesure. Ce qui n'est pas le cas avec les autres modèles.

Nous pouvons tirer deux conclusions :

- 1- On a des performances similaires en utilisant un solveur par Monte Carlo ou un ADVI
- 2- Intégrer les incertitudes de mesures laboratoires est très pertinents dans ce cas d'étude. En effet, il surpasse les autres estimations, et notamment l'estimation machine learning à travers l'utilisation du MLP.

La Figure 6 présente par exemple différents échantillons de données où l'on peut voir toute l'information additionnelle apportée par ces travaux de modélisation.

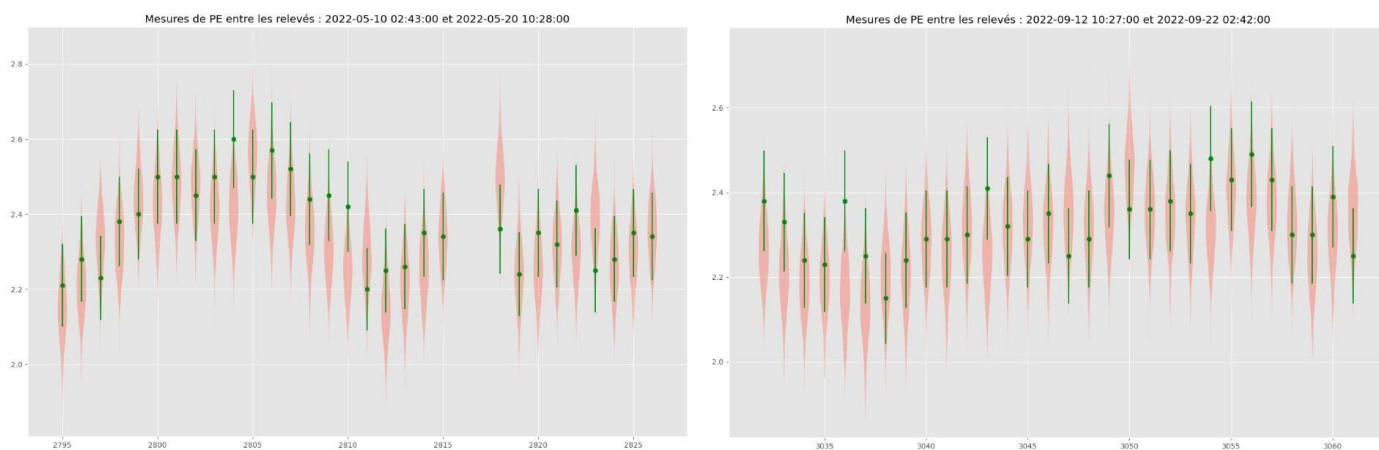


Figure 6. Evolution de la mesure laboratoire d'acidité dans le temps. En vert, la mesure et son incertitude, en rouge, les simulations issues de la modélisation bayésienne.

Un des avantages d'une modélisation qui combine des informations chimiques (bilan matière) et un modèle probabiliste est de rendre le modèle très interprétable. Nous avons veillé à ce que notre modèle respecte toujours la cohérence avec la réalité du procédé, et les graphiques ci-joints le confirment : la plupart du temps, nos simulations suivent la tendance d'évolution de l'acidité dans la cuve. Lorsque les simulations sont en accord avec les mesures, nous pouvons avoir confiance dans ces dernières. En revanche, lorsque nos simulations divergent des mesures, il est plus prudent de réaliser de nouvelles analyses en laboratoire, ce qui permet à Orano de mieux valider les résultats.

## Conclusion

Dans la première partie, nous avons présenté des généralités concernant le retraitement de combustibles nucléaires usés sur le site de la Hague avec un zoom sur le procédé de concentration des produits de fission par dénitrification formique dans des évaporateurs. L'enjeu consistait à souligner l'intérêt d'une conduite optimale des évaporateurs via une maîtrise précise et prédictive de l'acidité des évaporateurs informée par les données historiques de production et d'analyses laboratoire. Compte tenu des variabilités non-négligeables inhérentes au procédé, un cadre bayésien fut choisi afin de les intégrer dans l'estimation prédictive de l'acidité.

Dans la suite de l'article, des estimations quasi-continues de l'acidité des évaporateurs et leurs incertitudes furent précisément prédites à partir de deux approches bayésiennes interprétables. En effet, l'approche Monte Carlo par chaînes de Markov et celle par inférence variationnelle (ADVI) proposent des modèles prédictifs en cohérence avec le bilan chimique de la dénitrification formique tout en favorisant l'intégration de diverses incertitudes en jeu. Un comparatif de ces approches révèle aussi la rapidité d'inférence des prédictions issues de l'inférence variationnelle. Ces travaux ouvrent des perspectives sur l'application d'approches d'inférence bayésienne prédictive pour une conduite optimale des procédés industriels.

## Références bibliographiques

Salvatier, J., Wiecki, T., Fonnesbeck, C., *Probabilistic Programming in Python using PyMC*, <https://doi.org/10.48550/arXiv.1507.08050>

Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., and Blei, D. M. (2016). *Automatic Differentiation Variational Inference*. arXiv preprint arXiv:1603.00788.

George E.P Box University of Wisconsin-Madison (1980). *Sampling and Baye's Inference in Scientific Modelling and Robustness*.

Gelman, Carlin, Stern, Dunson, Vehtari and Rubin (1995 – 2003 – 2013). *Bayesian Data Analysis*.

Gelman (Departement of Statistics, Columbia University), et all (2017). *Expectation propagation as a way of life : A framework for Bayesian inference on partitioned Data*.

Nikhil Galagali, Youssef M. Marzouk (MIT, Cambridge) (2014). *Bayesian inference of chemical kinetic models from proposed reactions*.

Ryo Iwama, Hiromasa Kaneko, Meiji University, Japan (2021). *Design of ethylene oxide production process based on adaptive design of experiments and Bayesian optimization*.

## Nomenclature

### Éléments Chimiques :

$HCHO$	Fomaldéhyde
$HNO_3$	Acide nitrique
$CO_2$	Dioxyde de carbone
$H_2O$	Oxyde de dihydrogène (eau)
$NO$	Monoxyde d'azote
$NO_2$	Dioxyde d'azote
$FC$	Facteur de consommation du formol

### Notations mathématiques :

$A_t$  Pour tout capteur ou tout prélèvement A, il s'agit de la mesure de A à un instant t

$PE_t$  Il s'agit d'une mesure par prélèvement laboratoire à un instant t

$\widehat{FC} = f(X, \alpha, \beta)$ ,  $X \in \mathcal{E}$ ,  $\alpha \in \mathbb{R}$ ,  $\beta \in \mathbb{R}^p$ , p le nombre de paramètres procédé(3)

$\mathcal{E}$  L'ensemble des mesures procédées

$\mathbb{R}$  L'ensemble des nombres réels

$\hat{A}$  Pour toute mesure A, il s'agit de l'estimation de A. Par exemple :  $\widehat{FC}$ , il s'agit d'une estimation du facteur de consommation.

RMSE	Root Mean Squared Error : $\forall$ échantillon $Y$ $rmse(y_{pred}, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{pred_i} - y_i)^2}$
ITP	Métrique d'évaluation du modèle : $\forall \tau \in [0,0.5[$ $ITP(\tau) = \frac{1}{n} \sum_{i=1}^n (1 \text{ if } y_i \in [y_i * (1 - \tau); y_i * (1 + \tau)] \text{ else } 0)$
ADVI	Automatic Differentiation Variational Inference
MCMC	Markov Chain Monte Carlo
$N(\mu, \Sigma)$	Loi normale multivariée, de moyenne $\mu$ et de variance $\Sigma$ , de densité de probabilité : <ul style="list-style-type: none"> <li>• <math>\forall x \in \mathbb{R}^N, f(x, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{N}{2}}  \Sigma ^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu)\right)</math></li> </ul>
$IG(\alpha, \beta)$	Loi Inverse gamma, de densité de probabilité : <ul style="list-style-type: none"> <li>• <math>\forall (x, \alpha, \beta) \in (\mathbb{R}^{+*})^3, f(x, \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{x}\right)^{\alpha+1} \exp\left(-\frac{\beta}{x}\right),</math>                      avec <math>\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt</math></li> </ul>
MLP	Multi-Layer Perceptron : Réseau de neurones artificiels composé de plusieurs couches entièrement connectées. Il est souvent utilisé pour modéliser des relations complexes et non-linéaires.