université PARIS-SACLAY

Pronostic hybride à l'aide de codes de simulation et de modèles statistiques - Application à l'étude du colmatage des générateurs de vapeur

Hybrid prognostics using simulation codes and statistical models - Application to the study of steam generators clogging

Thèse de doctorat de l'Université Paris-Saclay

Ecole Doctorale de Mathématique Hadamard (EDMH) n° 574
Spécialité de doctorat : Mathématiques appliquées
Unité de recherche : Centre Borelli (ENS Paris-Saclay), UMR 9010 CNRS
Référent : Ecole Normale Supérieure de Paris-Saclay

Thèse CIFRE préparée en partenariat entre EDF R&D et les unités de recherche Centre Borelli et LISN sous la direction de **Mathilde MOUGEOT**, Professeure des Universités, la co-direction de **Didier LUCOR**, Directeur de Recherche CNRS et le co-encadrement de **Emmanuel REMY**, Ingénieur-Chercheur Expert et **Vincent CHABRIDON**, Ingénieur-Chercheur.

Thèse présentée et soutenue à l'ENS Paris-Saclay le 9 février 2026, par

Edgar JABER

Composition du jury :

Eleni CHATZI Professeure des Universités, ETH Zürich	Rapportrice
Pietro Marco CONGEDO Directeur de recherche, INRIA	Examinateur
Josselin GARNIER Professeur des Universités, Ecole Polytechnique	Président
Béatrice LAURENT-BONNEAU Professeure des Universités, INSA Toulouse	Rapportrice
Claudia SCHILLINGS Professeure des Universités, Freie Universität Berlin	Examinatrice
Emmanuel VAZQUEZ Professeur des Universités, CentraleSupélec	Examinateur

Thèse de doctorat

NNT: 2026UPASM002

Abstract (FR)

Ces travaux de thèse portent sur le développement de méthodes hybrides pour le pronostic de dégradation dans les systèmes industriels. L'application principale concerne le colmatage des générateurs de vapeur (GV) des réacteurs à eau pressurisée opérés par Électricité de France (EDF).

Deux grandes familles de modèles sont mobilisées : les modèles physiques et les modèles statistiques. Les approches hybrides cherchent à combiner ces paradigmes pour tirer parti de leurs complémentarités, en particulier pour mieux maîtriser l'incertitude prédictive. Ces outils sont cruciaux pour la maintenance préventive ou le remplacement d'équipements dans des infrastructures critiques. Lorsque des données de dégradation sont disponibles régulièrement, les techniques de filtrage (Kalman, particulaire) sont efficaces pour corriger les prédictions issues des simulations. Toutefois, dans le cas des GV, les données sont rares et les phénomènes physiques complexes, rendant le pronostic fortement dépendant du contexte. Des lignes directrices générales peuvent cependant être définies. Les modèles physiques présentent un biais structurel et une incertitude paramétrique liée à la méconnaissance des entrées. Leur bon usage nécessite une analyse de sensibilité et une quantification d'incertitude rigoureuse. Lorsque les simulations sont coûteuses, des métamodèles deviennent nécessaires. La première partie de la thèse propose une méthode non intrusive de quantification d'incertitude appliquée à un code industriel EDF. Les résultats mettent en évidence une incertitude de pronostic significative, en cohérence avec l'expertise métier. Il est ensuite crucial d'évaluer la qualité prédictive des émulateurs. Les méthodes de prédiction conforme offrent une approche robuste, sans hypothèses fortes, pour produire des intervalles de prédiction. Nous développons des estimateurs adaptés aux faibles régimes de données, appliqués à des processus Gaussiens scalaires. Contrairement aux intervalles bayésiens, ceux-ci sont peu sensibles aux priors mal spécifiés. Dans le cas déterministe, leur taille reflète l'erreur d'approximation, ce qui en fait un outil diagnostic pertinent. La suite du travail consiste à ajuster les lois a priori à l'aide des données disponibles, afin d'adapter la prédiction probabiliste au contexte opérationnel réel. Nous proposons une méthode de fusion de données inspirée de l'assimilation, adaptée à des sources d'information hétérogènes et peu nombreuses (modèles statistiques, observations terrains). Appliquée à un cas jouet de fissuration et au colmatage des GV, cette approche améliore les performances prédictives. Des questions subsistent sur la prise en compte des variables latentes et la modélisation de la discrépance. Enfin, dans une perspective exploratoire, les simulations recalibrées peuvent générer des trajectoires utiles à l'apprentissage temporel. Les données capteurs disponibles sur les systèmes réels ne mesurent pas directement la dégradation. Une question centrale est de savoir si ces signaux permettent d'inférer les états futurs de dégradation. En cas de corrélation, l'information exogène peut être exploitée via des features extraits des signaux.

Ces travaux s'inscrivent dans le développement des jumeaux numériques : l'intégration de modèles hybrides, la quantification d'incertitude et la fusion de données permettent de construire des cadres prédictifs robustes et certifiables pour les composants industriels des centrales nucléaires.

Abstract (EN)

This PhD thesis focuses on the development of hybrid methods for degradation prognostics in industrial systems. The main application concerns the clogging of steam generators (SGs) in pressurized water reactors operated by Électricité de France (EDF).

Two main families of models are used in prognostics: physics-based models and statistical, data-driven models. Hybrid approaches aim to combine both to leverage their respective strengths and improve robustness, particularly regarding predictive uncertainty. These tools are essential for planning maintenance or deciding on component replacement in long-lifespan critical infrastructure. In scenarios where degradation data is regularly available over time, filtering techniques (e.g., Kalman, particle filters) are effective at correcting simulation-based predictions using new observations. However, for complex systems like SGs, sparse data and complex physics make the prediction task strongly context-dependent. Still, general methodological principles can be established. Physics-based models come with structural biases and parametric uncertainty due to incomplete knowledge of input variables. Their use requires sensitivity analysis and rigorous uncertainty quantification, assuming the physical process is well modeled. When simulations are computationally expensive, surrogate models (or emulators) become necessary. The first part of this thesis develops a non-intrusive UQ method applied to an industrial clogging prediction code developed by EDF. Results align with expert knowledge and reveal significant prognostic uncertainty. It is then crucial to evaluate the predictive quality of the emulators. Conformal prediction offers a robust distribution-free framework to construct prediction intervals with guaranteed coverage. We develop estimators suited for limited-data settings, producing intervals for scalar Gaussian processes. Unlike Bayesian credible intervals, our bounds are less sensitive to prior misspecification. For deterministic codes, interval width reflects surrogate approximation error, making them useful diagnostic tools. The next phase involves conditioning the prior distributions on available heterogeneous data to improve predictive robustness. Unlike standard Bayesian calibration, typically applied using lab data, the goal here is to adapt probabilistic predictions to operational field contexts. We propose a data fusion approach inspired by data assimilation, tailored to sparse and heterogeneous sources (e.g., operator measurements, statistical models). Applied to a synthetic crack propagation case and SG clogging, the method significantly improves predictive performance. Open questions remain regarding latent variable uncertainty and discrepancy modeling. Finally, in an exploratory direction, recalibrated simulations can generate degradation trajectories suitable for time series learning. Real-world systems often rely on sensor data that do not directly measure degradation. A key research question is whether such exogenous signals can predict future degradation states. If correlation exists, unobserved degradation may be inferred through features extracted from sensor signals.

This work contributes to the development of digital twins, where hybrid modeling, uncertainty quantification, and data integration enable the construction of robust and certifiable predictive frameworks for industrial components such as those in nuclear power plants.

In all und jeder Zeit
Verknüpft sich Lust und Leid:
Bleibt fromm in Lust und seid
Dem Leid mit Mut bereit
- Alter Spruch

Laudate eum in sono tubae; laudate eum in psalterio et cithara.
Laudate eum in tympano et choro; laudate eum in chordis et organo.
Laudate eum in cymbalis benesonantibus; laudate eum in cymbalis jubilationis.
Omnis spiritus laudet Dominum. Alleluia.
- Psalmus 150, Biblia Sacra Vulgata

Acknowledgements

In no particular order, I would like to thank:

- Dr. Manuela Mocanu-Jaber (mom) and Roland Jaber (dad)
- The community at Fondation des États-Unis, with a special mention to the 5th floor
- Vincent Blot, PhD brother and Calanques co-explorer
- Baalu Belay-Ketema, for sharing S211 and the ups and downs of our research
- My Parisian Lyfrabuc friends
- Prof. Dr. Pauline Lafitte, professor at CentraleSupélec, for encouraging me to pursue a career in mathematics
- Jean-Philippe Jacob, prépa mathematics professor at the Lycée Carnot de Dijon
- Dr. Ruxandra Cristea, mathematics teacher, for instilling me the (belated) love of mathematics
- My supervisors in Paris-Saclay: Mathilde Mougeot and Didier Lucor
- My supervisors at EDF R&D: Emmanuel Remy, Vincent Chabridon and all my colleagues at EDF Labs Chatou
- The members of the jury, especially the reviewers: Prof. Dr. Eleni Chatzi and Prof. Dr. Béatrice Laurent-Bonneau
- Dr. Morgane Garo-Sail, manager of the EDF R&D project "Jumeau Numérique GV 2025"

This thesis was sponsored by the ANRT-CIFRE Grant No. 2022/1412.

To my love B. & JL.

Contents

Mathematical Notations

Symbol	Meaning
$\mathbb{R}, \mathbb{R}^N$	Set of real numbers, set of real vectors of dimension N
$D, [0, T]$	Space and time domains
t	Time variable
t_P	Present time
$\boldsymbol{x}$	Deterministic vector variable
$\boldsymbol{X}$	Random vector variable
X_i	i-th component of random vector $\boldsymbol{X}$
$\boldsymbol{X}^{(i)}$	i-th sample of random vector $\boldsymbol{X}$
$\mathcal{X}, \mathcal{Y}$	Output spaces of variables $\boldsymbol{X}$ and $\boldsymbol{Y}$
$p(\boldsymbol{X}), p_{\boldsymbol{X}}$	Density of random variable $\boldsymbol{X}$
$f\#p_{\boldsymbol{X}}$	Push-forward measure of distribution $p_{\boldsymbol{X}}$ through the function f
$\mathcal{P}(\mathcal{X})$	Set of probability measures over the set $\mathcal{X}$
$\boldsymbol{Y}$	Output vector variable (in regression problems)
$\boldsymbol{y}$	Output datapoint
N	Dimension of the output field
n	Size of the training set
m	Size of the test set
$\mathcal{D}$	Dataset of points
$\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{test}}$	Training and testing datasets
DoE	Design of Experiments
Q^2	Regression predictivity coefficient
$g(\cdot)$	Simulation code (e.g., THYC, DEPO, Puffer)
QoI	Quantity of interest function
$\mathcal{H}$	User-defined deterministic parameters in a computer model
θ, Θ	Parameters and set of parameter
$\widehat{f}, f_{\widehat{\theta}}$	Optimized regression functional models
$\boldsymbol{\mathcal{G}}$	Interpolation operator
pr_i	Projection operator on i-th marginal
$\mathbf{1}_A$	Indicator function of set A
$\boldsymbol{\Delta}^{p-1}$	Unitary simplex
Δ	Generic degradation rate
Δ_*	Degradation threshold
τ_c	Clogging rate
δ_a	Dirac-delta distribution in a, potentially a vector
$\mathcal{U}[a, b]$	Uniform distribution over the interval $[a, b]$
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
$\mathcal{T}(a, b, c)$	Triangular distribution over the interval $[a, c]$ with mode in b
$\mathcal{C}^k(D), \mathcal{C}^\infty(D)$	Set of k-times differentiable functions over D, set of smooth functions over D
$L^2_{p_{\boldsymbol{X}}}(\mathcal{X})$	Space of square-integrable functions over $\mathcal{X}$ w.r.t measure $p_{\boldsymbol{X}}$
$\mathbb{E}_{\boldsymbol{X}}, \mathbb{E}_{p_{\boldsymbol{X}}}$	Expected value operator w.r.t the law of $\boldsymbol{X}$ or the distribution $p_{\boldsymbol{X}}$

Glossary of Acronyms

Symbol	Meaning
BMU	Bayesian Model Updating
CEA	Commissariat à l'Énergie Atomique et aux Énergies Alternatives
CP	Conformal Prediction
CRC	Conformal Risk Control
DT	Digital Twin
EDF	Éléctricité de France
GP	Gaussian Process
HSIC	Hilbert-Schmidt Independence Criterion
JNGV	Jumeau Numérique Générateur de Vapeur (french for Steam Generator Digital Twin)
KLE	Karhunen-Loève expansion
kWh	Kilowatt per hour
LOO	Leave-One-Out
MCMC	Markov-chain Monte-Carlo
MSE	Mean-squared error
MWe	Electric megawatts
NPP	Nuclear Power Plant
NPWR	Nuclear Pressurized Water Reactor
PHM	Prognostics and Health Management
PINN	Physics-informed neural network
RB	Reduced basis
RUL	Remaining Useful Life
RMSE	Root mean-squared error
RWMH	Random-walk Metropolis-Hastings algorithm
SciML	Scientific Machine Learning
SG	Steam generator
TPD	THYC-Puffer-DEPO
TSP	Tube Support Plates
TVE	Televised Video Examination
UQ	Uncertainty Quantification
VPCE	Vector Polynomial Chaos Expansion
VV&UQ	Verification, validation and uncertainty quantification

Chapter 1

Introduction

General context

Nuclear power plants (NPPs) are a type of power plant that transforms heat from nuclear fission reactions into electrical current. They appear as crucial technological assets in the upcoming energy transition, given their very low carbon emissions per kWh produced. France has up to 80% of its daily electricity mix[1] that is composed from nuclear energy and, together with its operator EDF, it is one of the world leaders in the field. Among the different technologies of NPPs, the main reactor design of the French nuclear fleet are pressurized water reactors (NPWRs), with the entire fleet comprising 57 reactors. These are different generations of technologies with different nominal power outputs (namely 900, 1300 and 1450 MWe). The latest technology deployed is the European Pressurized Reactor (EPR) which has recently started operations in France at the Flamanville site, with a future power output of up to 1650 MWe. A NPWR is the assembly of three hydraulic circuits with heat exchangers linking them, as illustrated in Figure 1.1 below. The primary circuit transports the heated water through tubes passing in the first heat exchanger, namely the steam generator (SG). Each nuclear reactor has 3 or 4 SGs, depending on the reactor design. Inside SGs, cool water from the secondary circuit is heated and evaporated by heat transfer from the primary tubes. The outgoing steam is transfered as mechanic energy to a turbine that produces electric current with the help of an alternator. Finally, the cooling circuit allows to condense the water vapour back to liquid form, and back into the secondary circuit.

The clogging of SGs in NPWRs is a degradation phenomenon affecting certain plants of the French nuclear fleet (primarily on the 900 and 1300 MWe units). This complex degradation process arises from various factors, among which the most important are the erosion-corrosion of the secondary circuit tubes and the accidental pollution from raw water entering the cooling circuit (Prusek, 2012). Clogging is characterized by a clogging rate τ_c at the holes of the tube support plates (TSPs). The upper support plate is usually the only plate accessible to camera inspections, and it is empirically known to have the highest clogging rate (Prusek, 2012). This degradation process is difficult to monitor and diagnose. Indeed, there is little feedback from the system and data acquisition only occurs during plant shutdowns. Indirect acquisition methods based on regression from non-intrusive measurements based on Foucault eddy currents or transient indicators can be deployed (McNab, 1988; Girard, 2014), but it remains an ongoing topic of research. Moreover, this phenomenon has a very slow initiation kinetics, and due to the specific operational conditions of each NPWR, labo-

[1] see RTE-eco2mix website

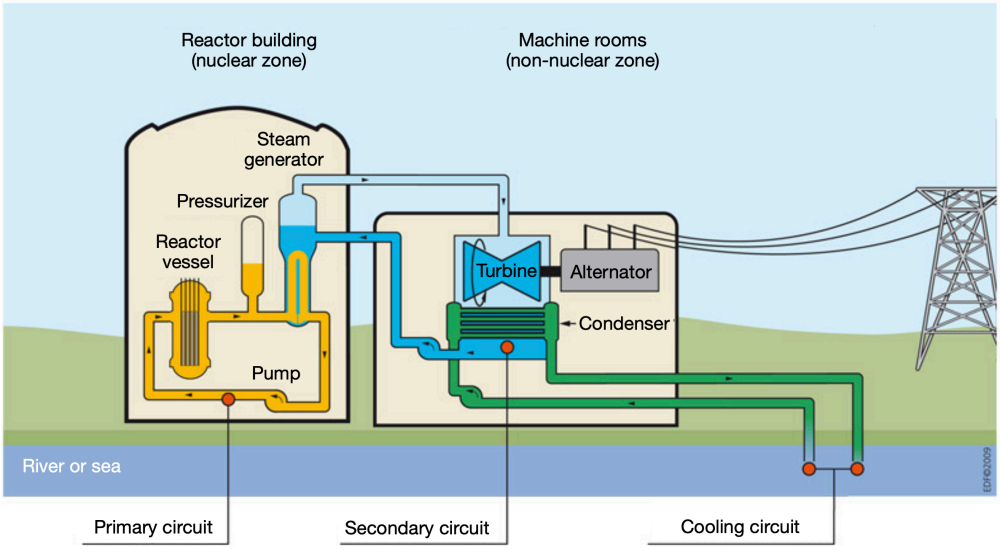

Figure 1.1: Diagram of a NPWR (© EDF, ASNR).

ratory experimental testing proves highly challenging. Nonetheless, studies have been conducted by EDF-CEA, with one notable example being the COLENTEC experimental loop (see Loubet et al. (2020)). Clogging proves to be hard to model due to the intertwined multiphysics interaction, notably involving fluid flows, heat transfer, chemistry of the secondary circuit, thermal-hydraulic evaporation conditions, complex component geometry and the variety of operating conditions. Although it does not pose a major safety threat, clogging has diverse impacts on the SGs – in particular, it can cause a localized redistribution of flow between the TSP flow holes, which elevates the risk of vibration, SG tube ruptures, and hampers the SG response to operational changes. To tackle these challenges, preventive chemical cleaning maintenances can be performed on the SG, and these maintenances have to be planned optimally. These consist in stopping the NPP and adding cleaning solvents inside the SG that disaggregate and remove part of the clogging deposits, and that are then evacuated through the purge of the SG (see Figure 2.1).

This problem lies at the core of the Prognostics and Health Management (PHM) framework (Vachtsevanos et al., 2006), whose purpose is to guide preventive maintenance actions in order to ensure reliable system performance. We denote by Δ a degradation index associated with a given system within a fleet of components. This index is generally represented as a function of time, $(t \mapsto \Delta(t))$, and may be scalar or vector-valued when degradation is evaluated at multiple locations within the component. For a prescribed degradation threshold Δ_*, the *remaining useful life* (RUL) (Vachtsevanos et al., 2006) is defined as

$$\mathrm{RUL}(t_P, \Delta_*) = \underset{t>t_P}{\arg\min}\{\Delta(t) \geq \Delta_*\}, \tag{1.1}$$

where t_P denotes the current observation time. In practice, the evolution of $\Delta(t)$ is influenced by a combination of deterministic phenomena such as physical laws, and stochastic phenomena stemming from inherent variability in material properties, loading, and environmental conditions. Consequently, in all its generality the degradation process must be treated as probabilistically, and the RUL becomes a *random variable* corresponding to the *first passage time* at which the stochastic

Figure 1.2: SG delivery and example of replacement of 900MW steam generators at the Cruas plant (© Mitsubishi Heavy Industries, Bouygues Travaux Publics)

process $\Delta(t)$ reaches threshold Δ_*. Prognostics therefore consists in estimating the probability distribution of this first passage time by means of a suitable degradation model in order to have a risk-informed decision strategy for maintenance planning. How are these uncertainties treated in practice will be discussed in the upcoming chapters and is dependent on the typology of prognostics model used. In this work, two classes of models are considered for clogging prognostics: physics-based models based on first principles and implemented through numerical simulation and statistical or data-driven models, such as supervised learning approaches. The goal of this thesis is to build so-called *hybrid models* from these two approaches, with the ultimate task to leverage the benefits of both model typologies in order to assist EDF engineers for a more robust and risk-informed maintenance planning.

General models used for prognostics

☛ Short history of prognostics

The study of system degradation prognostics can be traced back to the *industrial revolution* that started in the late 18th century until the 1850s. During that period, the spread of production mechanisation first exposed the economic and human costs of machine failure. Therefore, the need for predictable, durable, and safe operation of systems such as steam engines or locomotives motivated systematic experimentation on materials and mechanical components in order to avoid such accidents. Wöhler (1858, 1860) pioneering research first conducted experiments on railway axles fatigue that established the first empirical laws of mechanical degradation[a] thus laying the foundations of fatigue and strength testing. By the late nineteenth century, industrialisation had driven the creation of metrology laboratories and standards organisations, embedding empirical reliability considerations into engineering practice.

[a] The S-N curves (Stress vs Number of cycles) obtained in material fatigue testing are also called Wöhler curves.

The twentieth century extended these ideas into the probabilistic domain: Weibull (1951) developed statistical models of material life, and Barlow and Proschan (1965) formalised the reliability mathematical theory for complex systems. Later, the physics-of-failure paradigm introduced by Pecht (1992) and further expanded by Zio (2012) connected these statistical approaches with mechanistic degradation models, establishing the basis of modern prognostics and health management.

Physics-based models

👉 Short history on physics-based prognostics models

The transition from empirical engineering to physics-based prognostics began in the mid-twentieth century, driven by the desire to link component degradation directly to its underlying physical mechanisms. Among the early studies in fracture mechanics and materials science provided the mathematical framework for this shift: Griffith (1920) developed the energy-balance theory of fracture and the Paris and Erdogan (1960) introduced a crack-growth law based on mechanistic ordinary differential equations (see appendix 8.3), enabling prediction of RUL from measurable stress and crack-length data. These advances marked a conceptual turning point: failure was no longer a statistical event but the deterministic outcome of coupled physical processes such as fatigue, corrosion, wear, or diffusion. By the 1980s and 1990s, this mechanistic approach had evolved into the physics-based prognostics exemplified by Pecht (1992), which sought to model degradation using fundamental laws of mechanics, chemistry and materials science.

A useful taxonomy distinguishes physical degradation models by the mathematical form of their governing relation: empirical laws, ordinary differential equations (ODEs), and partial differential equations (PDEs). For example, the classic Archard's wear law (Delaney and Wang, 2025) expresses material loss as an empirical proportionality law and has been used for wear-based RUL prediction of gear and clutch systems (see for instance (Zhao et al., 2018)). Second, crack-growth models based on the Paris-Erdogan law (Paris and Erdogan, 1960) is based on an ODE and underpin many bearing and structural prognostics studies (see Cubillo et al. (2016)). Third, diffusion-limited ageing of solid electrolyte interphase growth in lithium-ion batteries is often modeled by a PDE-based continuum formulation e.g., solvent diffusion through the interphase plus reaction kinetics (see Sankarasubramanian and Krishnamurthy (2012)), and is crucial for better and reliable prognostics of battery state of health (see Li (2025) for more on battery prognostics). Each of these model types offers increasing mechanistic fidelity but also increasing complexity. Empirical laws have limited physics fidelity and may not extrapolate well; ODE models capture time-evolution but often assume simplified loads; PDE models can capture fine spatial distributions and transport phenomena but are typically computationally expensive and require detailed parameterisation and boundary-condition knowledge, thus their practical use in online prognostics has to be accompanied by some form of uncertainty quantification. Together, these three references illustrate the trade-offs in purely physics-based prognostics: interpretability and mechanism come at the cost of calibration burden, uncertainty handling, and computational tractability.

We assume that the degradation level Δ can be deduced as a quantity of interest $\Delta = \text{QoI}(u)$ where $u \in \mathbb{R}^N$ is solution to either a general PDE involving a parametric space-time differential operator $\mathcal{A}$ of the form:

$$(E_{\boldsymbol{X}}): \begin{cases} \mathcal{A}_{\boldsymbol{X}}(t, \boldsymbol{z}, \boldsymbol{X}) u_{\boldsymbol{X}}(t, \boldsymbol{z}) = F_{\boldsymbol{X}}(t, \boldsymbol{z}), \ (t, \boldsymbol{z}) \in [0, T] \times \boldsymbol{D} \\ u_{|\partial \boldsymbol{D}} = h_{\boldsymbol{X}}(t) \end{cases} \tag{1.2}$$

where $\boldsymbol{D}$ is a regular bounded domain in $\mathbb{R}^M$, $F_{\boldsymbol{X}}$ is a source term, $h_{\boldsymbol{X}}$ a regular boundary term and $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^d$ are input variables of the physical model. We now fix these variables at a given value $\boldsymbol{X} = \boldsymbol{x}_0$ and consider $(E_{\boldsymbol{x}_0})$. If there is no analytic solution to this system, numerical methods have to be used. Among the different methods possible, couplings between explicit/implicit Euler, or Runge-Kutta schemes in time and finite elements/differences in space (Quarteroni and Valli, 2008) can be used on well suited meshing of the space-time domain $[0, T] \times \boldsymbol{D}$. Moreover, these methods work if assumptions of ellipticity and coercivity of the differential operator in space are satisfied and subsequent stability and accuracy properties are satisfied as well. It should be noted that the final simulation time and the resolution of the domain mesh are crucial for the computation time of any numerical method used. For more complex physics problems involving couplings between transport, diffusion, and chemical reactions, multiphysics and multiscale approaches may be necessary. These approaches combine different numerical methods adapted to each physical phenomenon to solve the coupled equations consistently in time and space (Peksen, 2018). The various solutions are then used as parameters or initial conditions for other downstream equations. It should be noted that well-posedness of the resulting system becomes harder to guarantee. Such general deterministic simulation codes can be formalized as operators:

$$G : \mathcal{X} \times \mathcal{H} \to \mathbb{R}^N, \tag{1.3}$$

where the set $\mathcal{H}$ corresponds to user-defined numerical parameters whose variability comes from different code inputs (such as a choice of the discretization step or the closure laws). For fixed parameters $\boldsymbol{h} \in \mathcal{H}$, we denote $g(.) := G(., \boldsymbol{h}) \in \mathbb{R}^N$ the output of the simulation code taking only the model variables as input and therefore $g(\boldsymbol{X})$ is a numerical solution to the parametric PDE system $(E_{\boldsymbol{X}})$ with fixed numerical method parameters $\boldsymbol{h}$. In other words $g(\boldsymbol{X}) = \texttt{solve}(\boldsymbol{h}, E_{\boldsymbol{X}})$.

In the context of uncertainty quantification (UQ), $\boldsymbol{X}$ is treated as a random variable, and we assume that all the well known inputs were fixed in vector $\boldsymbol{h}$. One of the main objectives of UQ in this case is to make computer solvers more robust and predictive by assessing the image measure of a probabilistic modeling of $\boldsymbol{X}$. It can be approached in two ways: one is *non-intrusive* (useful for expensive industrial codes), and the other *intrusive*, allowing for more refined uncertainty evaluations. In the non-intrusive case, the process involves a few key steps (De Rocquigny et al., 2008) which are detailed in Figure 1.3 below.
After specifying the computer model and the uncertain parameters in step A, step B consists in defining a probabilistic model for the inputs, i.e., determining a distribution $p_{\boldsymbol{X}} \in \mathcal{P}(\mathcal{X})$, and then evaluating the pushforward measure $g\#p_{\boldsymbol{X}}$ via the code in step C. This is also referred as obtaining samples from the solution manifold:

$$\mathcal{M}_{\boldsymbol{X}} = \{g(\boldsymbol{X}), \ \boldsymbol{X} \sim p_{\boldsymbol{X}} \in \mathcal{P}(\mathcal{X})\} \subseteq \mathbb{R}^N. \tag{1.4}$$

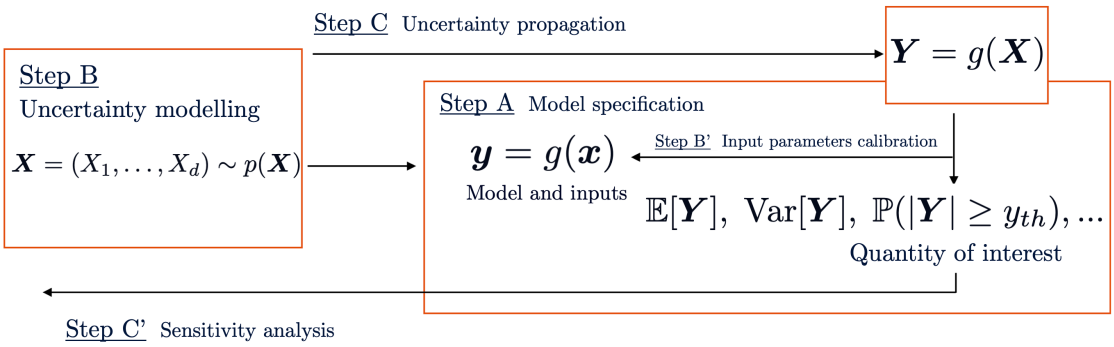

Figure 1.3: Generic methodology for non-intrusive uncertainty quantification of computer models for industrial applications, as based on (De Rocquigny et al., 2008).

Using typically Monte Carlo sampling (Rubinstein and Kroese, 2008), we obtain a dataset $\mathcal{D} = \{(\boldsymbol{X}^{(i)}, g(\boldsymbol{X}^{(i)}))\}_{i=1}^{n}$ from n code evaluations with which we can compute statistics or other quantities of interest. Often, due to mesh resolution and simulation duration, the computational time is significant. However, to study the output distribution and/or the sensitivity of input components, a much larger number of evaluations is often needed beyond the available budget. This is where supervised learning techniques can be employed to build surrogate models (also called "metamodels") $\widehat{g}$. Among the parametric families used are Gaussian processes (GP) (Rasmussen and Williams, 2006), polynomial chaos expansions (PCE) (Blatman and Sudret, 2009, 2011), or variants of neural networks (Lefebvre et al., 2023). These approaches are *black-box* in nature, meaning that the structure of the output is not assumed in advance; instead, learning procedures are implemented to approximate the output based on fixed-budget data. Depending on the knowledge of the governing equations and qualitative behavior of the solution, spatiotemporal interpolation methods can also be used to enhance the resolution. If the output dimension is too high and strongly correlated, then dimension reduction methods such as the Karhunen-Loève expansion (KLE) (Sudret and Der Kiureghian, 2002; Sullivan, 2015) can be applied. These methods are more widely known as reduced basis (RB) methods (Quarteroni et al., 2016), and can also be performed in an intrusive setting. These metamodels can be validated using standard regression validation metrics such as predictivity coefficients or more complex tools like conformal prediction which will be presented in this thesis.

Finally, there exist *intrusive* methods based on RB techniques (Le Maître and Knio, 2010; He et al., 2011; Quarteroni et al., 2016), or stochastic collocation methods (Xiu and Hesthaven, 2005; Xiu, 2010). These methods involve directly modifying the governing equations or numerical solvers to incorporate uncertainty quantification or reduced-order modeling (Luc, 2016). RB methods (which are similar, in a sense, to dimension reduction techniques in ML such as proper-orthogonal decomposition, see Hastie et al. (2009)) aim to construct a low-dimensional approximation of the solution space by identifying a set of representative basis functions, which are then used to solve the problem more efficiently. Stochastic collocation methods, on the other hand, discretize the stochastic space using collocation points and solve the deterministic problem at each point, enabling efficient propagation of uncertainties. These approaches are particularly useful for problems where the governing equations are well understood and computational resources are sufficient to handle the intrusive modifications.

👉 Physics-based model for clogging prognostics

For the SG clogging use-case, a simulation model has been developed, along with a numerical simulation tool THYC-Puffer-DEPO (initially proposed by Prusek et al. (2013) and enhanced and maintained by Feng et al. (2023)) enabling the simulation of clogging rate evolution over operational periods for NPWRs of up to 60 years (see chapter 2). The code is deterministic, producing identical outputs for identical inputs given the computational strategy of the numerical scheme used. In previous works (Lefebvre et al., 2023), certain input parameters of the model have been identified as having epistemic uncertainty, which may be reduced with improved knowledge, with the help of UQ methodologies. Thanks to its capability of extrapolating the clogging rate over the entire lifespan of the SG, it can also provide a first probabilistic RUL estimation.

Because THYC-Puffer-DEPO is a proprietary, closed-source solver with large runtimes and without direct access to the internal discretization or solver kernels, we adopt a parametric, non-intrusive UQ strategy: the code is treated as a deterministic black-box mapping from input parameters to full clogging trajectories, and uncertainty is introduced and quantified through probability models on those inputs (see chapter 3). This choice is further motivated by the intrinsic multiphysics and multiscale character of the model. Indeed, the solver couples thermal-hydraulics, chemistry and particle transport/adhesion processes across disparate spatial and temporal scales (see chapter 2). This yields complex response manifolds and intrusive modifications of the numerical scheme (e.g., RB methods) are impractical in a first approach; instead we perform parametric sensitivity analysis, Monte-Carlo sampling and design of experiments in the input space, and build non intrusive surrogates (GPs, PCEs, KLE-based field metamodels) to approximate the map g for fast code evaluations.

Statistical learning models

👉 Short history on statistical prognostics models

Statistical degradation models emerged from the need to generalise degradation behaviour across large populations of components with limited mechanistic knowledge. Building upon the probabilistic foundations laid by Weibull (1951) and Barlow and Proschan (1965), these approaches treated the RUL as a random variable drawn from empirically fitted lifetime distributions. With the advent of online data through digital sensors and monitoring systems, the field expanded toward stochastic process models such as hidden Markov models, gamma processes, and Wiener processes to represent uncertain degradation trajectories and estimate the RUL from noisy observations. These data-driven frameworks offered scalability and computational efficiency, but at the cost of physical interpretability and limited extrapolation beyond the training domain.

Statistical prognostics models may be broadly classified as *parametric* or *non-parametric* according to the assumptions imposed on the underlying degradation process. Parametric models postulate a fixed functional form governed by a finite set of parameters $\theta \in \Theta$, which are estimated from data; typical examples include Weibull or exponential lifetime models and stochastic process formulations

such as the Wiener or Gamma process. For instance Tayade et al. (2019) uses a linear regression model together with principal component analysis for RUL estimation of bearings, and Dhada et al. (2025) applies a hierarchical Weibull parametric model to multiple sub-fleets of similarly deteriorating systems for RUL prediction. In contrast, non-parametric models make fewer assumptions about the degradation law and instead try to infer it directly from data, as in kernel density estimators, GPs, random forests, or neural networks. Noteworthy applications of these non-parametric models for different prognostics tasks include the use of convolutional neural networks (Li et al., 2018) and long-short-term memory recurrent neural networks (LSTM) (Shi and Chehade, 2021) for the prediction of RUL of turbofan engines, support vector machines (García Nieto et al., 2015), and multi-layer perceptrons (Zhao et al., 2017) for the prediction of RUL of aircraft engines. Parametric methods offer interpretability and compactness but may be biased if the assumed structure is misspecified, whereas non-parametric approaches provide greater flexibility at the cost of higher data requirements and reduced physical transparency.

A supervised learning model (Hastie et al., 2009) for regression involves defining a sample $\mathcal{D} = \{(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})\}_{i=1}^{m}$ with $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^d$ and with independent and identically distributed (i.i.d.) targets $\boldsymbol{Y} \in \mathcal{Y} \subseteq \mathbb{R}^\ell$ such that $(\boldsymbol{X}, \boldsymbol{Y})$ follows an unknown generating distribution $p_{\boldsymbol{X},\boldsymbol{Y}}$, splitting it into a training and test sample $\mathcal{D} = \mathcal{D}_n \cup \mathcal{D}_m$ of respective sizes n and m, as well as choosing a (parametric) family of functional models $(f_\theta : \mathbb{R}^d \to \mathbb{R}^\ell)$ to determine a regression model. The goal is to minimize the risk on the training sample with respect to the empirical measure $\widehat{p}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})}$ - where $\delta_{(a,b)}$ here is the Dirac-function on $\mathbb{R}^{d+\ell}$-and a loss function $\mathcal{L} : \mathcal{Y}\times\mathcal{Y} \to \mathbb{R}_+$ to find the optimal parameters $\widehat{\theta} \in \Theta \subseteq \mathbb{R}^q$, i.e.,

$$\widehat{\theta} \in \arg\min_{\theta\in\Theta} \int_{\mathcal{X}\times\mathcal{Y}} \mathcal{L}(\boldsymbol{y}, f_\theta(\boldsymbol{x}))\widehat{p}_n(d\boldsymbol{x}, d\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\boldsymbol{Y}^{(i)}, f_\theta(\boldsymbol{X}^{(i)})). \tag{1.5}$$

Among the different parametric families, we can distinguish, among others, linear models, nonlinear models and kernel-based models (Hastie et al., 2009). Linear models, such as linear regression, assume a linear relationship between the covariates $\boldsymbol{X}$ and the target variable $\boldsymbol{Y}$, and in this case, $\widehat{\theta}$ has an explicit formula. For a linear predictor with output dimension $\ell = 1$, $f_\theta(\boldsymbol{x}) = \boldsymbol{x}^\top\theta$ with squared loss $\mathcal{L}(y, f) = (y - f)^2$, the OLS estimator is:

$$\widehat{\theta} = \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{Y}, \tag{1.6}$$

where $\mathbf{X} \in \mathbb{R}^{n\times d}$ has rows $(\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)})^\top$ and $\mathbf{Y} = (Y^{(1)}, \ldots, Y^{(n)})^\top$. Non-parametric models, like neural networks (Hastie et al., 2009), capture more complex relationships using nonlinear transformations of the data. Kernel-based models, such as support vector machines (SVM) consist in projecting the data into a higher-dimensional space through the kernel to make relationships linear in that space. These different families offer flexibility in model choice depending on the nature of the data and the study objectives. Once the optimization problem in Eq. (1.5) is solved, the resulting function is called an *estimator* $\widehat{f} := f_{\widehat{\theta}}$. In cases where the data exhibit a temporal structure, models, such as time series models, or memory-based models, like recurrent neural networks such as LSTM architectures can be used. These models capture temporal dependencies of potentially various ranges in the data. For example, an LSTM model can be defined to predict future degradation $\Delta(t)$ based on past observations (also called autoregressive steps) $(\Delta(t-1), \Delta(t-2), \ldots, \Delta(t-k))$,

possibly using exogenous covariates $\boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-k) \in \mathcal{X}$ jointly observed. The loss function $\mathcal{L}$ is then adapted to account for the sequential structure of the data, such as a cumulative loss over successive predictions. These approaches, however, require a significant volume of data and a rigorous validation phase to avoid overfitting. Classical regression methods can also be used effectively, ensuring that $\mathcal{D}_m$ contains only future values. Feature engineering approaches could be used to homogenize temporal dependencies (Kuhn and Johnson, 2019).

As in any modeling strategy, statistical models show some limitations. They often rely on the assumption of stationarity or independence, which may not hold in real-world scenarios. Additionally, they may struggle to generalize well when the data distribution changes over time (a phenomenon known as distribution shift) or when the available data is sparse or noisy. These models require careful tuning of hyperparameters, they are prone to prior misspecifications and are sensitive to the choice of the loss function and model architecture. Furthermore, statistical models may fail to capture complex physical interactions or domain-specific knowledge, which can be critical for applications in engineering systems such as prognostics. In order to enhance their robustness, a first path is to quantify the uncertainty of the prediction of $\widehat{f}$. This can be done using frequentist confidence intervals theory or Bayesian approaches.
Confidence interval theory provides a statistical framework for quantifying the uncertainty of predictions made by a model. Given a point estimate $\widehat{f}(\boldsymbol{X})$ for the scalar target variable Y, a confidence interval $[a(\boldsymbol{X}), b(\boldsymbol{X})]$ is constructed such that, under repeated sampling, the true value Y falls within this interval with a specified probability $1-\alpha$, where $\alpha \in [0,1]$ is the significance level. For example, in linear regression with Gaussian additive errors, the interval is typically based on the estimated standard error and quantiles of the normal distribution. Confidence intervals offer interpretable bounds on predictions and are widely used for uncertainty quantification in regression and forecasting tasks. Bayesian approaches treat the model parameters as random variables with prior distributions, then update these distributions based on observed data to obtain posterior distributions. For example, in a statistical learning framework, if a model parameter θ has a prior distribution $p(\theta)$ and the data likelihood is given by $p(\mathcal{D}|\theta)$, then the posterior distribution is given by Bayes' rule:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}, \tag{1.7}$$

where $p(\mathcal{D})$ is the marginal probability of the data, obtained by marginalization. This approach incorporates parameter uncertainties and propagates them into model predictions. Bayesian methods often require dedicated sampling techniques such as Markov Chain Monte Carlo (MCMC) sampling (Marin and Robert, 2007) or variational approximations to estimate posterior distributions. These methods are particularly useful when data is limited or when prior knowledge needs to be integrated into the model. Another approach to quantifying uncertainties in learning models is through conformal prediction methods (Vovk et al., 2005; Angelopoulos and Bates, 2023). In its generic approach, for any learning model $\widehat{f} : \mathcal{X} \to \mathcal{Y} \subseteq \mathbb{R}$ learned on a training set $\mathcal{D}_n$, a prediction interval $\widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)})$ for any new point $(\boldsymbol{X}^{(n+1)}, \boldsymbol{Y}^{(n+1)})$ is constructed for a coverage rate $\alpha \in (0,1)$ using the model $\widehat{f}$. This interval has the marginal coverage property:

$$\mathbb{P}\left(\boldsymbol{Y}^{(n+1)} \in \widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)})\right) \geq 1-\alpha, \tag{1.8}$$

where the probability is understood marginally, i.e., integrated over all possible permutations of

$\mathcal{D}_n \cup \{(\boldsymbol{X}^{(n+1)}, \boldsymbol{Y}^{(n+1)})\}$. The basic assumption for using conformal methods is *data exchangeability*, with i.i.d. data being a special case. Consequently, conformal methods do not naturally extend well to problems with correlated outputs, such as time series or spatial statistics. However, advances in this area, such as pseudo-exchangeability concepts (see Zaffran et al. (2022)), allow for overcoming this limitation.

☛ Statistical models for clogging prognostics

In the clogging context, degradation data is scarce and irregular: direct camera inspections occur only during rare shutdowns, measurements are often indirect or noisy, and operational conditions vary across NPWR units. This leads to challenges for purely data-driven methods: training sets are small, exchangeability and stationarity assumptions break down, temporal dependencies are poorly observed, and learned predictors tend to overfit or fail to transfer between SGs. As a result, many statistical estimators may produce unstable or overly confident RUL forecasts when used alone on clogging problems. To cope with these constraints, EDF developed dedicated indirect tools. ESTICOL is a proprietary indirect-estimation procedure designed to infer historical clogging rates from the BIL100 periodic tests (see chapter 2); it provides reconstructed clogging rates in past operation times. PREVICOL is an operational statistical tool that implements a linear regression model (see chapter 6) to extrapolate future clogging from available pointwise observations. Both methods reuse the same limited observational material: ESTICOL focuses on transforming sparse, indirect signals into usable estimates, while PREVICOL applies lightweight predictive models on specific feature engineering and both lack enough validation data.

Hybrid models

In the context of PHM, a *hybrid model* refers to a modeling framework that integrates physics-based models with data-driven statistical or machine learning models. The goal is to leverage the strengths of both paradigms: physical models provide mechanistic understanding and extrapolation capability, while data-driven models offer flexibility to capture unmodeled dynamics in operational scenarios, residual errors, or other unknown correlations (Liao and Kottig, 2014; Eker et al., 2019). Such hybrid approaches are particularly relevant when physical modeling is incomplete or too computationally intensive, and observational data is available but potentially noisy or sparse. A typical hybrid PHM framework may involve model correction strategies, Bayesian updating, physics-informed learning, or dual-filtering techniques. These models are instrumental in improving the robustness, accuracy, and explainability of predictive maintenance tools, especially for complex industrial systems with operational variability and limited degradation observability.

A hybrid surrogate model approach that is popular in the Scientific Machine learning (SciML) community today involves *physics-informed* learning, where one strategy is to regularize the loss functions with physical terms. This is done, for example, with physics informed neural networks (PINNs) (Raissi et al., 2019), or more recently polynomial chaos expansions (Novák et al., 2024). Moreover, in Bayesian GP approaches, prior kernels can be chosen to match fundamental solutions of equations, e.g., (Henderson et al., 2023) for the wave equation. These learning approaches regularize the loss function by adding a physical term that penalizes solutions violating physical laws. More

precisely, if we consider a parameter $\boldsymbol{X} = \boldsymbol{x}_0$ in Eq. (1.2) with a homogeneous Dirichlet boundary term, the regularization term is defined as:

$$\mathcal{L}_\varphi = \frac{1}{N_t N_{\boldsymbol{z}}} \sum_{i=1}^{N_t} \sum_{j=1}^{N_{\boldsymbol{z}}} \left\| \mathcal{A}_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) \widehat{u}_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) - F_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) \right\|^2, \tag{1.9}$$

where $\widehat{u}$ is the approximation given by the learning model (e.g., a neural network in the case of PINNs (Raissi et al., 2019)) and $\{(t_i, \boldsymbol{z}_j)\}_{i,j}$ are collocation points over $[0, T] \times \boldsymbol{D}$ sampled uniformly. The total loss function becomes:

$$\mathcal{L}_{\text{tot}} = \mathcal{L} + \lambda \mathcal{L}_\varphi, \tag{1.10}$$

where $\mathcal{L}$ is a supervised loss in Eq. (1.5) such as the quadratic loss, and $\lambda > 0$ is a hyperparameter controling the strength of physical regularization. Typically we might have some datapoints from experimental measurements $\{y(t_i, \boldsymbol{z}_j)\}_{i,j}$ then

$$\mathcal{L} = \frac{1}{N_t^{\text{exp}}} \frac{1}{N_{\boldsymbol{z}}^{\text{exp}}} \sum_{i=1}^{N_t^{\text{exp}}} \sum_{j=1}^{N_{\boldsymbol{z}}^{\text{exp}}} \left| y(t_i, \boldsymbol{z}_j) - \widehat{u}_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) \right|^2. \tag{1.11}$$

Such regularization promotes solutions that are consistent with physical principles, even under limited data. These methods are particularly useful when data is expensive, noisy or partially observed, as they combine the benefits of supervised learning and physical modeling. However, they struggle with multiphysics and multiscale coupled problems, as well as long-time integration and training is often expensive. Nevertheless, theoretical guarantees for approximation and generalization are known for specific operators (Mishra and Molinaro, 2021; Doumèche et al., 2023). These methods primarily act as spatiotemporal interpolators that refine solutions locally while preserving physical constraints. However, in their current form, they do not handle random parameterizations of $\boldsymbol{X}$, which is the subject of ongoing research (see for instance (Panahi et al., 2025)).

Surrogate models of the various types can be used for forward problems, meaning that we compute various statistics of the original pushforward measure $g\#p(\boldsymbol{X})$, but they can also be used for inverse problems (Stuart, 2010) such as Bayesian model updating which is a type of hybrid model. In this context one is interested in reducing the prior uncertainties in the input parameters, with the help of data $\boldsymbol{y} \in \mathbb{R}^N$. This data can come from a data-driven model or from actual sensor data. In loose terms, the following hypothesis is made:

$$\boldsymbol{y} = g(\boldsymbol{X}) + \boldsymbol{\eta}, \tag{1.12}$$

where $\boldsymbol{\eta}$ represents the noise, typically modeled as a random variable with a prior distribution (e.g., Gaussian). The goal is to estimate the posterior distribution of $\boldsymbol{X}$ given the observed data $\boldsymbol{y}$. This is achieved again using Bayes' theorem as in Eq. (1.7):

$$p(\boldsymbol{X}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{X}) p(\boldsymbol{X}). \tag{1.13}$$

Techniques such as MCMC or variational inference can be used to sample from or approximate the posterior distribution (Rubinstein and Kroese, 2011). In this context, the use of surrogate models is crucial if the computer model g is computationally intensive. These methods are particularly useful for calibrating simulation models or updating predictions in real-time as new data becomes

available. In the well-known Kennedy and O'Hagan (2001) framework, the model in Eq. (1.12) can be augmented by adding a model discrepancy, taking into account the misfit between the true value of the degradation and the predicted model.

Another typical Bayesian technique used for hybrid degradation prognostics (Liao and Kottig, 2014) is *data assimilation*, and usually the go-to choice is *filtering* (Jouin et al., 2016). It is particularly well-suited for time-dependent degradation problems with *continuous* inflow and monitoring of data. This method addresses the inverse problem of updating the degradation state probability at a given time step using all available data points preceding it. The most common filtering techniques include the Kalman filter and its variants (Evensen, 1994), as well as particle filters (Del Moral et al., 2006). The Kalman filter assumes linear dynamics and Gaussian noise, making it computationally efficient but limited in its applicability to nonlinear systems, even if there exist some extensions of the Kalman filter which leverage their limitations (see more in Sullivan (2015)). The particle filter, on the other hand, is a non-parametric approach that uses a set of particles to represent the posterior distribution, making it suitable for highly nonlinear and non-Gaussian problems.

For example, in the context of degradation, recall that $\Delta(t)$ represents the degradation state at time t, and let $\boldsymbol{y}_{1:t}$ denote the sequence of observations up to time t. The goal is to compute the posterior distribution $p(\Delta(t)|\boldsymbol{y}_{1:t})$. Using a particle filter, this is achieved by propagating a set of particles $\{\Delta^{(i)}(t)\}_{i=1}^{n}$ through the physical model and updating their weights based on the likelihood of the observations:

$$p(\Delta(t)|\boldsymbol{y}_{1:t}) \approx \sum_{i=1}^{n} w^{(i)}(t)\delta_{\Delta^{(i)}(t)}, \quad w^{(i)}(t) \propto w^{(i)}(t-1)p(\boldsymbol{y}(t)|\Delta^{(i)}(t)), \tag{1.14}$$

where $\delta_{\Delta^{(i)}(t)}$ is the Dirac delta function centered at the particle $\Delta^{(i)}(t)$. This approximation allows for efficient computation of the posterior distribution and facilitates the estimation of quantities of interest, such as the expected degradation state or the probability of exceeding a critical threshold, all while monitoring the non-degeneracy of the weights (see (Chopin and Papaspiliopoulos, 2020)). By propagating ℓ times ahead of the present time t_P, the resulting RUL distribution will be estimated through its cumulative distribution function:

$$\mathbb{P}(\text{RUL}(\Delta_*) \leq t_{P+\ell} \mid \boldsymbol{y}_{1:t_P}) \approx \sum_{i=1}^{n} w^{(i)}(t_P)\mathbf{1}\{\Delta^{(i)}(t_{P+\ell}) > \Delta_*\}. \tag{1.15}$$

Data assimilation methods (Geir Evensen, 2022) are particularly powerful for real-time, online updating, meaning that observations $\boldsymbol{y}$ are continuously available at all timesteps. Additionally, this framework supposes a Markovian structure for the degradation phenomenon, allowing to update the degradation probability only through its preceding timestep through a kernel $p(\Delta(t)|\Delta(t-1))$.

👉 Hybrid models for clogging prognostics

In this work we simplify the physical model by assuming no aleatoric noise on the true degradation and no model-form uncertainty of the simulator: this means we treat the proprietary THYC-Puffer-DEPO simulator as a non-intrusive, deterministic foundation for probabilistic RUL estimation. The remaining uncertainty is therefore parametric (coming from input parameters) and (in principle) reducible with additional data and improved priors. This working hypothesis allows us to focus the hybrid workflow on updating the clogging state simulated through assimilation of observations into the model in order to deliver risk-informed, operational RUL estimates. Because TPD is used as a black-box that produces whole degradation trajectories (rather than supporting stepwise updates), standard online filtering is not adapted; therefore, we treat the problem offline by using the available heterogeneous groups of data (sparse camera inspections and ESTICOL reconstructions) to reduce the simulator output uncertainty first by informing the distribution of the inputs via Bayesian updating and then through smoothing of the posterior trajectories (as we will discuss in chapter 5).

Problem statement and objectives

The main goal of this thesis work is to obtain robust clogging RUL predictions with respect to uncertainties, by fusing all the available information and numerical tools into a hybrid framework. This industrial challenge poses limitations to classical prognostics approaches, most notably because of the lack of substantial and consistent field data as well as the complex physics and different operational scenarios for each NPWR. Therefore one cannot hope to build purely data-driven generalizable algorithms nor be absolutely satisfied with a purely physical and numerical approach. Uncertainties are present in all the models and within this set of hypotheses, our goal will be to minimize the global uncertainty and to guarantee its control at every step in all its forms.

A more long-term objective of this work is to propose methodologies for creating *digital twins of nuclear power plant components* (NPPCs) (US-NRC, 2021). Digital twins are virtual representations of physical assets that integrate data, models, and simulations with associated uncertainties to enable real-time monitoring, diagnostics, and prognostics (NAS, 2024; Liang et al., 2024). In the context of nuclear power plants, digital twins could facilitate predictive maintenance, optimize operational strategies, and enhance crucial safety by providing a continuously updated view of components' state of health. Achieving this requires the integration of physical models, data-driven approaches, uncertainty quantification, and data assimilation techniques discussed above. However, to date, there is still much to be done for obtaining real-time monitoring of relevant data on existing NPPCs, starting with sensor qualification and optimal placement together with validation and verification of physical models and surrogates. The development of robust digital twins thus relies on hybrid modeling frameworks that can adapt to new information, incorporate domain knowledge, and provide reliable predictions under uncertainty for existing facilities. For new nuclear facilities, there is room for better control at the design stage to facilitate the deployment of digital twin methodologies.

Outline of the thesis

The rest of this manuscript is structured as follows. Each chapter corresponds to a key component in the general scheme illustrated in Figure 1.4:

- *Chapter 2: Clogging modeling and digital twins for clogging prognostics*
  - We outline the existing prediction tools for SG clogging at EDF R&D and explain the strategy for digital twinning.
  - We identify the main limitations in standard hybrid approaches from the literature and we formulate the main problem addressed in this thesis.
- *Chapter 3: Uncertainty Quantification of THYC-Puffer-DEPO*
  - We provide results on advanced sensitivity analysis techniques for the uncertainty quantification of the THYC-Puffer-DEPO clogging simulation code as based in the published work (Jaber et al., 2025b).
  - We present a certain number of different metamodeling strategies for the code THYC-Puffer-DEPO, namely: Gaussian processes, polynomial chaos expansions, Karhunèn-Loève based field metamodels as well as bias-minimizing surrogate aggregation strategy.
  - We exhibit new unknown features of the clogging model uncertainty in the clogging model and obtain a first view on the predictive RUL uncertainty from a purely model-driven approach on a specific SG.
- *Chapter 4: Robust predictive metamodeling with conformal prediction*
  - We present a background on conformal prediction methods and their potential for application in SciML tasks.
  - We showcase the cross-conformal estimator developed for diagnostics and qualification of scalar-valued Gaussian process surrogate models of deterministic computer codes, as linked to our published work (Jaber et al., 2025a).
  - We develop uncertainty bands for vector-valued Karhunen-Loève-GP surrogates using conformal risk control and show their potential use for prior model qualification.
  - Both strategies are applied to THYC-Puffer-DEPO and we discuss their implications for active learning and design.
- *Chapter 5: Offline Bayesian hybrid strategies*
  - This is the core chapter of the thesis in which we develop a generic methodology for offline data assimilation, allowing to update relevant input variables of the computer model as well as applying Kalman smoothing in order to obtain robust risk-informed RUL. It is associated with the submitted work (Jaber et al., 2026).
  - We discuss the potential integration of an additional discrepancy model within the Bayesian model updating step.
- *Chapter 6: Time-series based prognostics regression models*

– This is a prospective chapter on the potential use of available SG sensor monitoring datasets in view of quasi real-time monitoring of clogging prognostics.
– We present the statistical work for prognostics at EDF R&D called PREVICOL and show how it could be potentially enhanced.
– We explore their explanatory power for modeling the clogging rate with the help of different regression models, including linear and Ridge regression.

The thesis ends with a conclusion and perspectives on future work in 7 and the appendix with various additional results from the different chapters are found in 8.

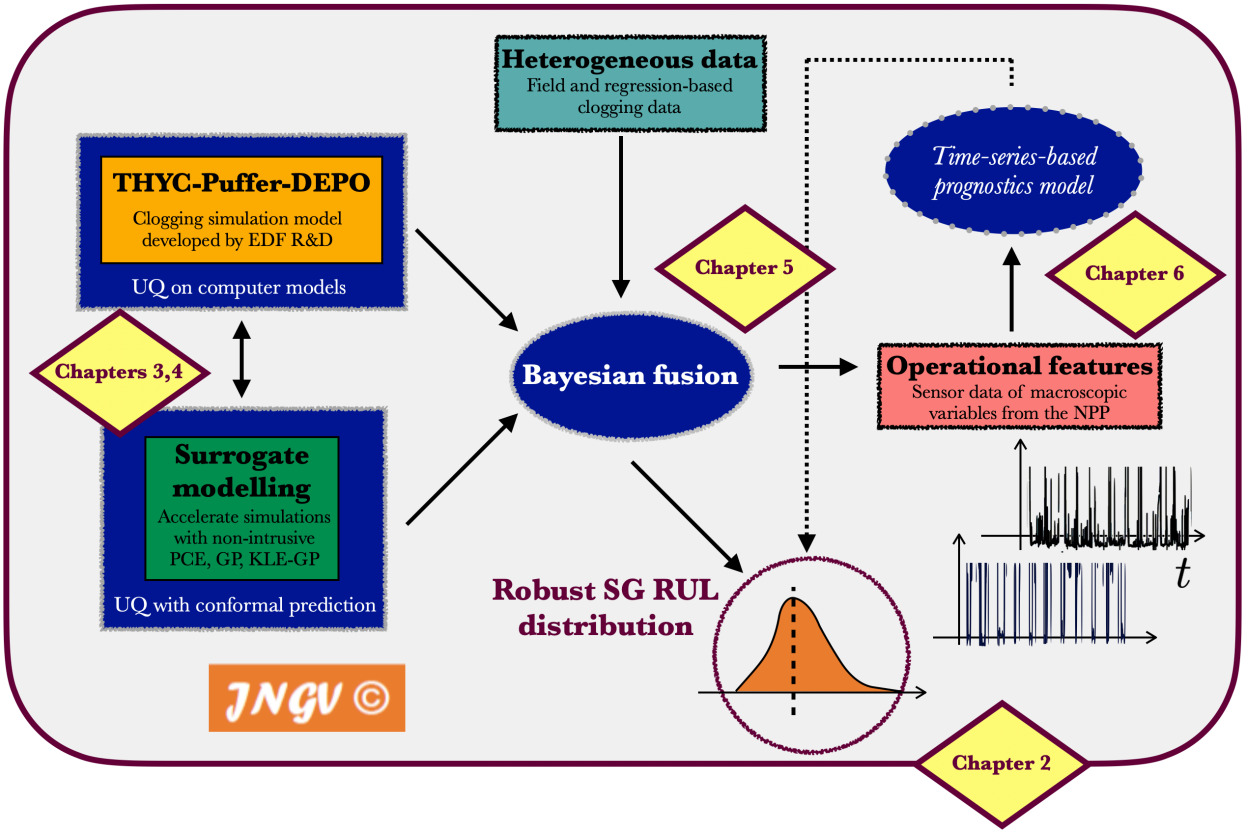

Figure 1.4: The digital twin framework developed and the articulation of the different thesis chapters.

Peer-reviewed journal papers

1. *Jaber, E., Chabridon, V., Remy, E., Baudin, M., Lucor, D., Mougeot, M., Iooss, B., Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators, Published in International Journal of Uncertainty Quantification, 2025, 10.1615/Int.J.UncertaintyQuantification.2024051489*

2. *Jaber, E., Blot, V., Brunel, N., Chabridon, V., Remy, E., Iooss, B., Lucor, D., Mougeot, M., Leite, A., Conformal Approach to Gaussian Process Surrogate Evaluation with Marginal Coverage Guarantees, Published in Journal of Machine Learning for Modeling and Computing, 2025, 10.1615/JMachLearnModelComput.2025054687*

3. *Jaber, E., Remy, E., Chabridon, V., Lucor, D., Mougeot, M., Fusion of heterogeneous data for robust degradation prognostics, Published in Journal of Reliability Engineering and System Safety, 2026, 10.1016/j.ress.2026.112435*

Summary of oral communications

- 04/2023, MASCOT-NUM 2023, Le Croisic, France - poster
- 09/2023, CJC-MA 2023, CentraleSupélec - poster
- 10/2023, ETICS 2023, Lège Cap-Ferret, France - slides
- 02/2024, SIAM Conference on Uncertainty Quantification 2024, Trieste, Italy - slides
- 09/2024, ETICS 2024, Saissac, France - slides (in French)
- 02/2025, DTE&AICOMAS 2025, Paris, France - slides
- 04/2025, SAMO 2025, Grenoble, France - poster
- 06/2025, INI Workshop: Calibrating prediction uncertainty : statistics and machine learning perspectives, Cambridge, U.K - slides, video
- 06/2025, UNCECOMP 2025, Rhodes, Greece - slides
- 08/2025, Mathematical and Computational Foundations of Digital Twins, CIRM Marseille, France - talk
- 09/2025, 9e rencontre PHM France, Grenoble, France
- 10/2025, UQSay#89, Online Seminar

Summary computational developments

All the results in this manuscript are reproducible and are tied chapter by chapter to the following `thesis GitHub repository`. This approach has the merit of ensuring reproducibility, allowing others to verify and build upon the work and can be readily used by the EDF engineers for deploying the methodologies:

1. `SA of a multi-physics long-term clogging model for steam generators`: implementations of sensitivity analysis and surrogate modeling techniques with OpenTURNS[2].

[2]see documentation here: `https://openturns.github.io/openturns/latest/index.html`

2. `Conformal approach for GP surrogates`: contributed to the repository with Vincent Blot, coupling between the MAPIE[3] conformal prediction library and OpenTURNS.

3. `Offline data fusion for robust degradation prognostics`: creating and simplifying classes for MCMC and metamodel assembly with OpenTURNS, implementation of the methodology on a toy-case and the main use case.

[3] see documentation here: `https://mapie.readthedocs.io/en/stable/`

Chapter 2

Steam generators digital twins for clogging prognostics

This chapter presents a comprehensive overview of the clogging phenomenon in steam generators within nuclear pressurized water nuclear power reactors, focusing on its physical modeling, statistical models, and integration into a digital twin framework for predictive maintenance. We detail the multi-physics mechanisms driving clogging, including corrosion, particle transport, and deposition, and describe the THYC-Puffer-DEPO simulation chain used to model long-term clogging kinetics. The ESTICOL statistical tool for historical clogging estimation based on operational data and televised video examinations is introduced, highlighting challenges due to sparse and heterogeneous measurements. The chapter further discusses the emerging role of digital twins in nuclear asset management, emphasizing the need for hybrid approaches that combine physics-based models, statistical methods, and uncertainty quantification. Finally, we formulate the hybrid prognostics problem for steam generator clogging, outlining the integration of simulation, data-driven models, and field data to enable robust, risk-informed estimates of remaining useful life.

Contents

2.1 Overview on the clogging phenomenon in steam generators

As explained in the introduction, SGs function as heat exchangers within NPWRs, which contain two separate water circuits facilitating heat transfer. In this configuration, water from the primary circuit goes initially through the pressure vessel, where it undergoes heating due to nuclear reactions in the core. It then enters the SG (illustrated in Figure 2.1 below) through the primary inlet, also called the *hot leg* at a temperature of about 320°C, transferring heat to the secondary loop and therefore cooling the fluid of the primary circuit, which exits through the *cold leg* at a temperature of 290°C - all of this maintained at a pressure of 155 bars. The secondary loop enters through the feedwater inlet at a temperature of 230°C and is heated to 285°C and the resultant steam exits the SG via its upper openings, flowing through turbines to produce electrical power at a pressure of about 70 bars. The primary fluid circulates through a bundle of U-shaped tubes (around 6000), supported by tube-support plates (TSPs). Adjacent to these tubes, the secondary fluid passes through the holes of the TSP, experiencing vaporization due to the heat from the primary fluid in the tubes (as represented in Fig. 2.1 below).

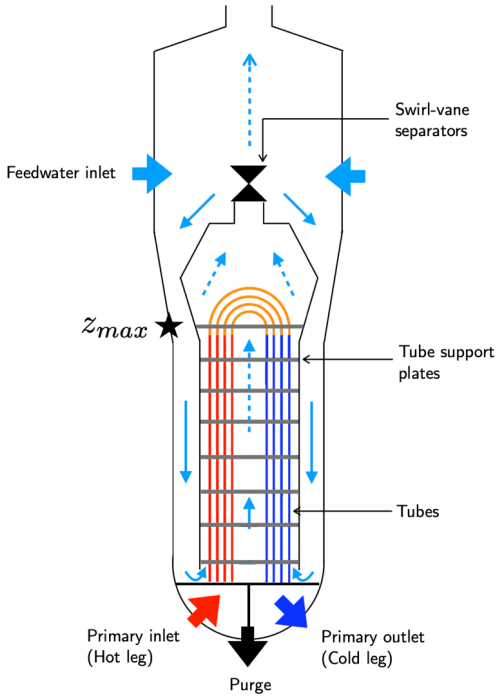

Figure 2.1: SG functioning scheme and principal sub-components. Source: personal drawing.

The French nuclear fleet is equipped with different models of SGs, and in this thesis we will be focusing on the 51B model (for more technical aspects of the system, see Girard (2014)) with additional results on the 6819 model in the appendix 8.1. Over extended periods of operation, certain SGs may experience a clogging process in the flow-holes of their TSPs. Clogging reduces the flow area within the TSPs due to the buildup of iron oxide deposits, primarily resulting from corrosion in the secondary circuit, with the oxides carried by the secondary flow. This process has various

repercussions on the SGs, causing localized changes in flow distribution among the TSP flow holes, thereby increasing the risk of vibration, SG tube ruptures, and affecting the SG responsiveness to operational adjustments. To tackle this issue, the clogging rate (referred to as τ_c) is assessed during PWR outages using non-destructive televised video examinations (TVEs), which are illustrated in Figure 2.2. These are performed at the $z_{\max}$ TSP altitude in the SG. The clogging rate provides an average indication of the extent to which the holes of a TSP are clogged. Chemical cleaning and pH monitoring are employed, among other reasons, to mitigate clogging, and to ensure a reliable and efficient operation of the NPP. The chemical cleaning process is the maintenance operation to be planned that is considered in this thesis. It comes in two different categories, so-called "curative cleaning" and "preventive cleaning", both having a varying degrees efficiency, but it manages to lower the clogging rate substantially. It is worth noting that this maintenance process can have damaging effects on the environment, therefore, its proper planning has multifold objectives and is of particular importance for EDF.

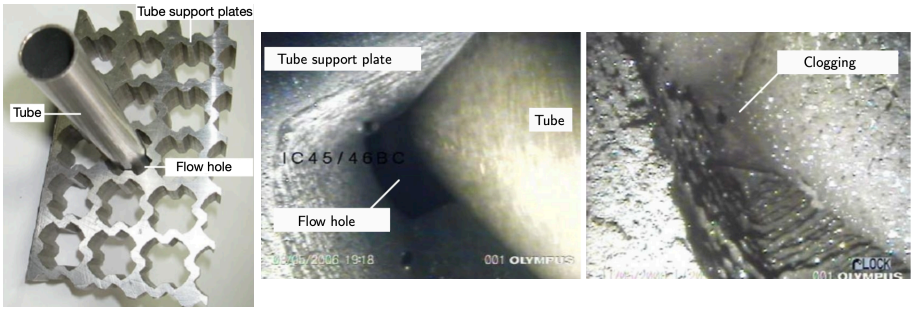

Figure 2.2: TSP and example of video examination during an PWR outage (© EDF).

As explained in the introduction, EDF R&D has developed two complementary types of tools to predict steam generator clogging: data-driven models and a physics-based simulation code. The data-driven approach relies on regression algorithms trained on operational indicators and TVE data. Depending on the objective, these models serve either to retrospectively estimate the historical health state of a specific generator or to forecast future clogging evolution. The tools developed for these two scopes are, respectively, ESTICOL (Pinciroli et al., 2021) (see section 2.3) for estimation, and PREVICOL (Mathieu et al., 2020) (see section 6.2) for forward prediction. On the physics side, no benchmark model currently provides a reliable fundamental understanding of the clogging rate τ_c, and the scientific literature remains sparse (Srikantiah and Chappidi, 2000; Prusek et al., 2013; Girard, 2014; Yang et al., 2017), making validation and benchmarking particularly difficult. The phenomenon itself evolves very slowly and is highly sensitive to operational conditions, complicating its reproduction under controlled experimental conditions since there is much variability between different NPPs. Although some indirect measurements and rare video field data are available, their quantity and quality remain insufficient to support fully reliable purely data-driven prediction. A physical clogging model has nonetheless been implemented and enhanced in the form of the THYC-Puffer-DEPO numerical code (Prusek, 2012; Prusek et al., 2013; Feng et al., 2023) (see section 2.2.4). However, due to the scarcity of trustworthy observational data, validating this model with high confidence remains a challenge. Consequently, decision-making regarding chemical cleaning planning must be performed under uncertainty, raising the key question of how to best leverage the available knowledge and models to achieve reliable predictions.

2.2 Physical model of clogging

2.2.1 Phenomenon description

In the flow holes of the SG, due in particular to the transport of impurities present in the secondary circuit, the liquid phase becomes enriched with particles that are likely to deposit during evaporation. Following Prusek et al. (2013) several sources of material contribute to clogging:

- *Generalized corrosion:* Iron from the secondary circuit piping oxidizes into magnetite. This phenomenon is largely driven by the water's pH, but also potentially by thermal-hydraulic factors;
- *Erosion-corrosion:* This mode of corrosion is driven by high flow velocity, leading to the thinning of secondary circuit components;
- *Ammoniacal corrosion:* This occurs due to degradation of copper alloy components in the presence of oxygen and ammonia and can be entirely eliminated by removing copper-based components (typically done for recent generations of NPPs);
- *Accidental pollution of the cooling circuit:* This results from the ingress of raw water containing various impurities.

According to Prusek (2012), corrosion and magnetite deposits account for about 90% of the clogging rate. The remaining components of the clogging deposits are difficult to assess, therefore in the remainder of the modeling, magnetite is considered to be solely responsible for all clogging deposits. Additionally, following Prusek et al. (2013), we assume these particles are spherical with diameter d_p.

2.2.2 Two-phase thermal-hydraulics

As previously mentioned, inside the SG, the secondary fluid is in a two-phase liquid-gas state. The thermal-hydraulic quantities of interest are denoted by (ρ_m, U_m, H_m) and correspond respectively to the mixture density (with the m subscript in the variables referencing to this solid-liquid mixture), the mass momentum of the mixture, and the mixture enthalpy. For these quantities, the following system of conservation laws is proposed in reference (Prusek et al., 2013):

- *Mass conservation:*

$$\partial_t \rho_m + \nabla \cdot (\rho_m U_m) = 0, \tag{2.1}$$

- *Momentum conservation:*

$$\partial_t(\rho_m U_m) + \nabla \cdot (\rho_m U_m \otimes U_m) + \nabla \cdot (\rho_m C_\ell C_g U_r \otimes U_r) = -\nabla P_m + \nabla \cdot (\tau_m) + \rho_m g + I_{fs}, \tag{2.2}$$

- *Energy conservation:*

$$\begin{aligned} &\partial_t(\rho_m H_m) + \nabla \cdot (\rho_m H_m U_m) + \nabla \cdot (\rho_m C_\ell C_g L U_r) = \\ &\partial_t P_m - \nabla \cdot \phi_m + E_{fs} + \nabla P_m \left[U_m + C_\ell C_g \rho_m \left(\frac{1}{\rho_g} - \frac{1}{\rho_\ell} \right) U_r \right]. \end{aligned} \tag{2.3}$$

It is worth mentioning that in order to solve this system of equations, additional closure laws (known as physical correlations), initial conditions, and geometric parameters of the domain must be specified and these are undisclosed for reasons of industrial privacy. The primary unknowns in this system are the mixture density ρ_m, the mixture velocity U_m, the liquid phase velocity U_ℓ, and the relative velocity of the gas phase with respect to the liquid phase U_r.

2.2.3 Deposition mechanisms

In the secondary circuit water, magnetite particles are present in two forms: *solid* and *soluble*. Throughout the remainder, quantities indexed by $h = s$ denote *soluble species*, while those with $h = p$ denote *solid particles*. Each form is associated with a mass fraction Γ_h. These mass fractions are governed by transport equations of the form:

$$\partial_t \rho_m C_\ell \Gamma_h + \underbrace{\nabla \cdot (\rho_m C_\ell \Gamma_h U_\ell)}_{\text{Convection}} + \underbrace{\nabla \cdot (\rho_m C_\ell D_h \nabla \Gamma_h)}_{\text{Diffusion}} = P_h + Q_h, \tag{2.4}$$

where C_ℓ is the liquid quality, P_h is the volumetric transfer between particle and soluble species, and Q_h is the volumetric transfer between soluble species and the TSP walls.
For soluble species, P_s represents the *precipitation* of soluble particles when their concentration exceeds the solubility limit. This is a "sink" term since it reduces soluble content. It is expressed integrally as:

$$\int_t^{t+\delta t} P_s dr = -\rho_m C_\ell \max(0, \Gamma_s - \Gamma_s^{\max}), \tag{2.5}$$

where $\Gamma_s^{\max}$ is the maximum space-dependent solubility of magnetite. Q_s is calculated from the fluxes of precipitated soluble species at the wall, contributing to clogging:

$$Q_s = -\Phi_{e,s} - \Phi_{c,s}, \tag{2.6}$$

where $\Phi_{e,s}$ and $\Phi_{c,s}$ denote, the flux of soluble species coming from fouling and clogging, respectively. For solid particles, P_p represents the creation of particles by *precipitation* of soluble species. This is a "source" term. Assuming instantaneous precipitation:

$$P_p = -P_s. \tag{2.7}$$

Finally, Q_p is calculated as a balance between deposited and removed particle fluxes:

$$Q_p = (\Phi_{e,p} - \Phi_r) + (\Phi_{c,p} - \Phi_r). \tag{2.8}$$

Here, the removal flux Φ_r is assumed negligible and omitted. We also assume that diffusion is negligible compared to convection in the transport equation.

The vena contracta mechanism

When the secondary flow reaches the flow hole, it encounters a constriction due to a narrowing of the flow section. As a result, the boundary layer near the walls thickens at the entrance of the flow hole, creating a recirculation zone. This zone promotes the deposition of solid particles in the flow. This mechanism is called the vena contracta.

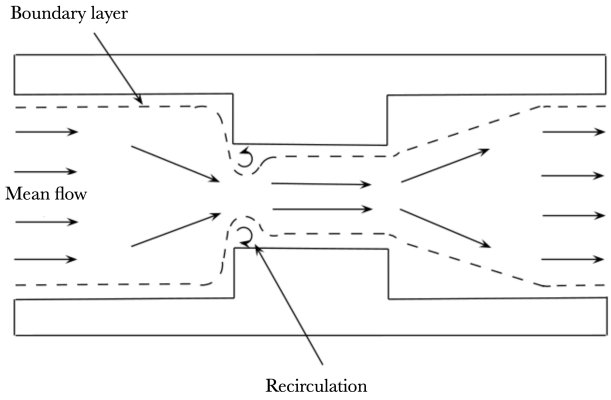

Figure 2.3: Vena contracta mechanism, adapted from Prusek et al. (2013).

For simplicity, we assume that all transported particles entering the recirculation zone are deposited. According to Prusek et al. (2013), the flux of solid species is given by:

$$\Phi_{c,p} = K_{c,p}\rho_m C_\ell \Gamma_p, \tag{2.9}$$

where $K_{c,p}$ is the wall deposition velocity given by:

$$K_{c,p} = a_v \frac{k_v(\rho_p - \rho_\ell)U_m^2 C_g d_p^2}{\mu_\ell}, \tag{2.10}$$

with μ_ℓ the dynamic viscosity of the fluid, ρ_p, ρ_ℓ the densities of particles and fluid respectively, C_g the vapor quality, d_p the magnetite particle diameter, a_v a model calibration parameter, and k_v a blockage parameter for the flow hole. This parameter is defined in terms of the clogging rate τ_c and the geometric parameters of the considered support plate as:

$$k_v = \frac{L_0 - R_0\,(1 - \tau_c)}{S}. \tag{2.11}$$

These geometric parameters of the TSP are described in Figure 2.4 below, for the so-called quatrefoil TSP design (more details on the designs are found in Girard (2014)).

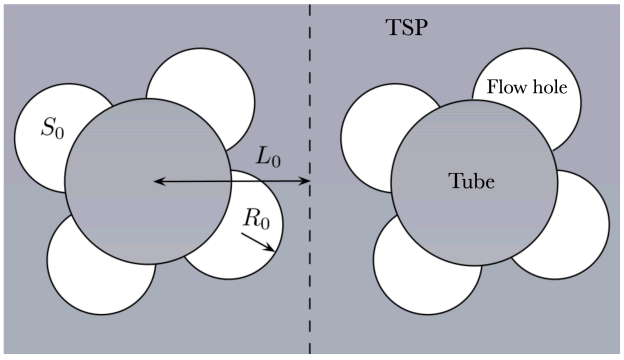

Figure 2.4: Geometric parameters of the TSP, adapted from Prusek et al. (2013).

In the flow holes, a local increase in the mass fraction Γ_s of soluble species occurs. This is partly due to the fluid contraction in the support plate passage (as described in the vena contracta mechanism). A sudden vaporization of the fluid follows, transferring soluble species into the liquid phase.

The flashing mechanism

When the solubility limit $\Gamma_s^{\max}$ is reached, the particles precipitate and consolidate the deposits in the recirculation zone: this is called flashing (Prusek et al., 2013). Again, it is assumed that all precipitated soluble species contribute to deposition:

$$\Phi_{c,s} = \Phi_g \Gamma_s^{\max}, \tag{2.12}$$

with Φ_g the vapor mass flux created by the flashing mechanism. This is determined from the liquid mass flux entering the flow hole:

$$\Phi_g = \Phi_\ell \frac{\Delta H_\ell}{H_{\ell g}}, \tag{2.13}$$

where $H_{\ell g}$ is the latent heat of vaporization (liquid to gas) and ΔH_ℓ is the enthalpy drop induced by the flow hole.

The clogging model

The mass of magnetite deposited by clogging m_c at each TSP level is modeled by the following ordinary differential equation:

$$\frac{dm_c}{dt} = \Phi_{c,p} + \Phi_{c,s}, \tag{2.14}$$

where $\Phi_{c,p}$ is the particle flux and $\Phi_{c,s}$ is the flux of precipitated soluble species on the wall due to clogging. This deposited magnetite mass can be converted into a clogging rate τ_c using an empirical correlation from Prusek et al. (2013), relating it to the clogging volume V_c:

$$\tau_c = \alpha \left(1 - \exp(-\beta V_c)\right), \tag{2.15}$$

with empirical fitting parameters (α and β) that will be detailed in the upcoming chapter.

To summarize, the deposition model corresponds to the following system:

$$\begin{cases} \partial_t \rho_m C_\ell \Gamma_s + \nabla \cdot (\rho_m C_\ell \Gamma_s U_\ell) = P_s + Q_s \\ \partial_t \rho_m C_\ell \Gamma_p + \nabla \cdot (\rho_m C_\ell \Gamma_p U_\ell) = P_p + Q_p \\ dm_c/dt = \Phi_{c,p} + \Phi_{c,s} \end{cases} \tag{2.16}$$

with initial conditions $\Gamma_s(0), \Gamma_p(0)$ and solubility $\Gamma_s^{\max}$, over the space-time domain $[0, T] \times D$.

The full long-term physical model of SG clogging consists of three nested temporal levels of models, allowing simulation of the clogging kinetics ($t \to \tau_c(t)$) over long periods of time (typically, from 40 to 60 years, and even more), as illustrated in Figure 2.5. First, the stationary thermal-hydraulic fields in the SG are computed. Using these fields, a transport mechanism carries the magnetite particles in both solid and soluble forms, which are likely to cause clogging following the vena contracta and flashing mechanisms. Once the mass fractions from both phenomena are determined, a growth equation for the deposited magnetite mass is solved, enabling us to track the variation in the clogging rate at the different TSPs of the SG.

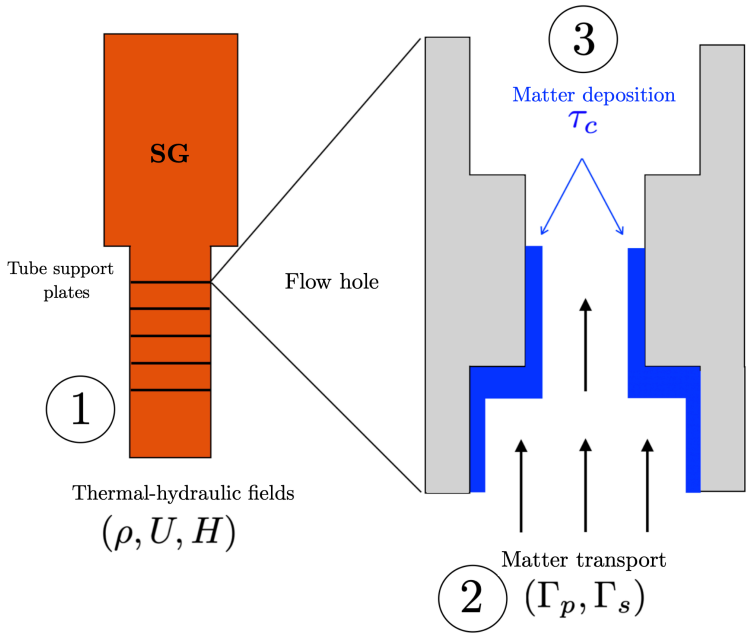

Figure 2.5: Nesting of the three physical chained temporal models of the SG clogging model.

Analysis of the solutions and their regularity

Among the parameters involved in Eq. (2.14), we can distinguish latent variables coming from the stationary quantities and other known geometric input variables that we denote as a generic vector $\boldsymbol{h}$, as well as input variables deemed to have epistemic uncertainty, denoted by $\boldsymbol{X}$. Finally the sum of fluxes gives rise to a nonlinear function f depending on clogging rate, latent variables, and uncertain parameters. The well-posedness of the two-phase thermal-hydraulic system is not clear at this time; this casts doubt on the complete theoretical existence of local solutions for the clogging process in a certain class of solutions with this modeling. However, for the coupling between transport and deposition, it is possible to establish local existence/uniqueness results, assuming the thermal-hydraulic quantities are well-defined with good regularity.

Suppose now that we consider a small interval $[s, s'] \subset [0, T]$ over which the deposition model is solved. According to the modeling of Prusek et al. (2013), ignoring the removal flux, we have a temporal ordinary differential equation for each TSP:

$$\frac{dm_c}{dt} = f(\tau_c(t), \boldsymbol{h}, \boldsymbol{X}), \text{ with } m_c(s) \text{ known.} \tag{2.17}$$

In the function f, we find the uncertain parameters $\boldsymbol{X}$, but also the solutions of Eq. (2.4), the stationary solutions of the thermal-hydraulic system in Eqs. (2.3), as well as other physical constants, all of which are considered latent variables in vector $\boldsymbol{h}$ depending only on space and not varying with time. Only the clogging term in the blocked flow parameter defined in Eq. (2.11) introduces temporal variation in f through the presence of $\tau_c(t)$. Moreover, the clogging rate τ_c is related to the clogging mass m_c by the phenomenological law in Eq. (2.15), rewritten as:

$$\tau_c(t) = \alpha\,(1 - \beta \exp[-m_c(t)/\rho_c]) = h(m_c(t)), \tag{2.18}$$

where $h \in \mathcal{C}^\infty(\mathbb{R})$. Furthermore, the term k_v in Eq. (2.11) is a linear term in f involved in the *solid* particle flux, which leads, up to constants, to:

$$f(\tau_c) \propto h(m_c).$$

Thus, by the global Cauchy-Lipschitz theorem, we have a solution $m_c \in \mathcal{C}^\infty([s, s'], \mathbb{R})$ and therefore by composition, $\tau_c = h(m_c) \in \mathcal{C}^\infty([s, s'], \mathbb{R})$. Thus, throughout the duration of *stationarity* of the latent variables (transport, thermal-hydraulics), assuming these are well-defined, we have a solution as regular as desired. Note also that the clogging rate is strictly increasing since:

$$\tau_c' = m_c' h'(m_c) = f(\tau_c, \boldsymbol{h}, \boldsymbol{X})\alpha\beta \exp(-m_c/\rho_c) > 0, \tag{2.19}$$

which is a useful property for performing time-interpolation in the subsequent computer model results, as we will explain later.
About the spatial regularity, we point out that the true quantity of interest studied is defined for each TSP height of the SG $z_i \in \{z_1, \ldots, z_{\max}\}$, denoting by $\boldsymbol{e}_i = (x, y, z_i)$ the spatial variable, the integrated and averaged clogging rate in the hot leg (HL) and cold leg (CL) of the SG is given by:

$$\tau_c(t, z_i) = \frac{1}{2} \iint (\tau_c(t, \mathrm{HL}, \boldsymbol{e}_i) + \tau_c(t, \mathrm{CL}, \boldsymbol{e}_i))\, dxdy. \tag{2.20}$$

In the rest of this thesis, we will only consider the specific height $z_i = z_{\max}$ and omit the spatial variable, writing simply $(t \mapsto \tau_c(t))$. All of this does not change the above results on temporal regularity under the appropriate assumptions.

2.2.4 The numerical simulation code: THYC-Puffer-DEPO

The numerical model THYC-Puffer-DEPO (TPD) is implemented through a simulation chain of computer codes, each represented as a function $g_* : \boldsymbol{x}_* \mapsto g_*(\boldsymbol{x}_*)$, where $*$ denotes the code name and $\boldsymbol{x}_*$ is its vector of deterministic input variables. When the input vector is random, it is denoted by capital bold letters $\boldsymbol{X}_*$, and the output may be written as $\boldsymbol{Y}_* = g_*(\boldsymbol{X}_*)$. The stationary solution of the conservation laws for the two-phase mixture is computed using the THYC software developed by EDF R&D (David, 1999), referred to as g_{THYC}. This code employs a finite-volume scheme with a porous-medium approach to deliver results at the component scale. After obtaining the thermal-hydraulic state $\boldsymbol{Y}_{\mathrm{THYC}}$, the g_{DEPO} code (Lefebvre et al., 2023) is used to solve the transport and balance equations. The main output of interest is the increment in clogging rate between two time points $t < t'$:

$$\Delta_t^{t'} \tau_c := \tau_c(t') - \tau_c(t)\,. \tag{2.21}$$

To manage the clogging rate, the pH of the secondary circuit can be adjusted by adding basic organic solvents, a process known as chemical conditioning. The corresponding chemical conditioning modifies the iron oxide solubility maps $\Gamma_s^{\max}$ and are generated by the g_{Puffer} code, which takes into account the SG geometry, thermal-hydraulic state, and chemical conditioning. In our applications we will consider two types of chemical conditionings: χ_1, χ_2 which have two different levels of pH: high and low (see Figure 3.1). For extended simulation periods (e.g., over 10 years of SG operation), the stationary thermal-hydraulic state may become invalid due to the accumulation

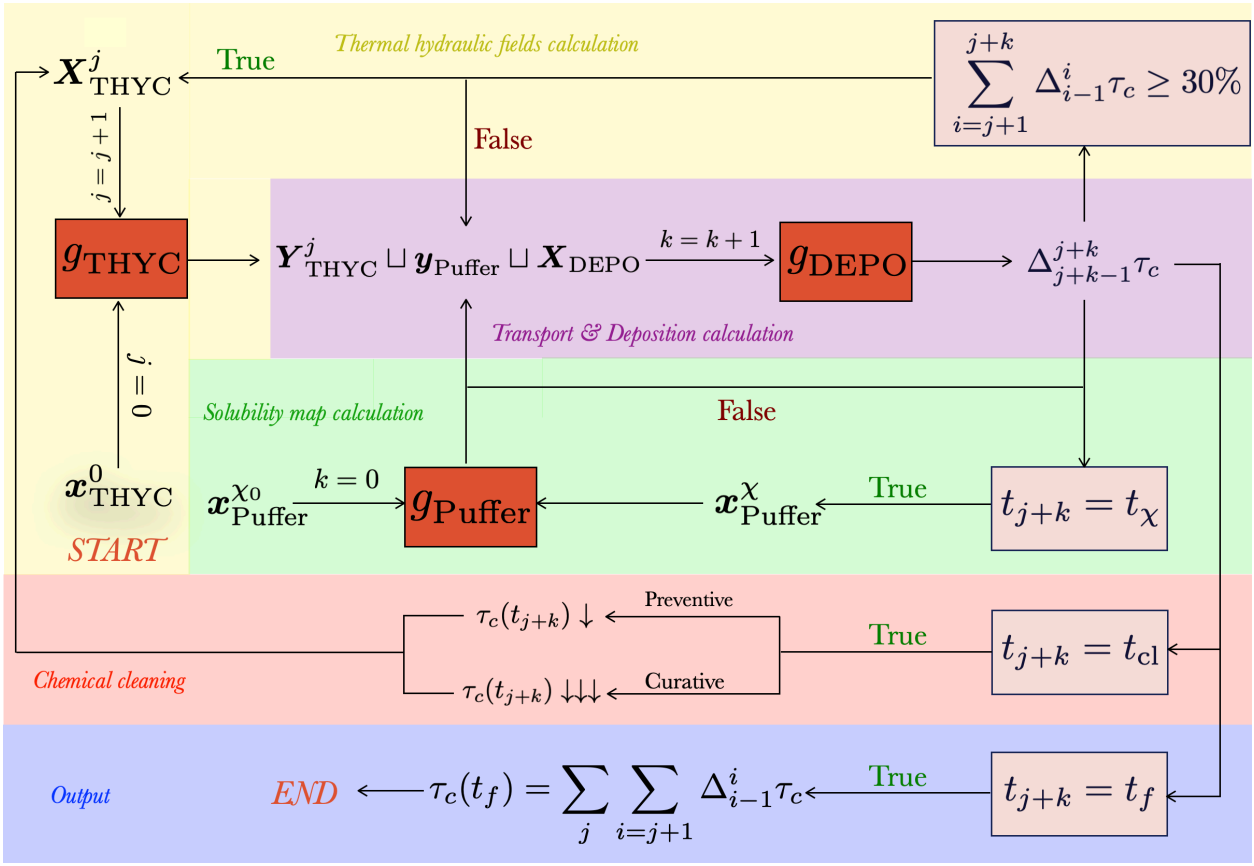

Figure 2.6: THYC-Puffer-DEPO simulation chain.

of clogging deposits, which increases head losses. Therefore, the three codes are coupled, forming the THYC-Puffer-DEPO simulation chain illustrated in Figure 2.6.
The simulation starts at iteration $j = 0$ (see "*START*" in Figure 2.6), by calculating the stationary thermal-hydraulic state $\boldsymbol{y}^{0}_{\text{THYC}}$, assuming no uncertainties in the input variables of g_{THYC}. The input vector $\boldsymbol{x}^{0}_{\text{THYC}}$ includes SG geometry as well as the physical correlations for closing the system of conservation laws in Eq. (2.3) (which are omitted for industrial privacy reasons). Our strategy is to propagate the uncertainties solely on the clogging module, and the UQ of the THYC code can be the topic of future work. Next, g_{Puffer} computes a solubility map $\boldsymbol{y}_{\text{Puffer}}$ for the specified chemical conditioning and geometry in $\boldsymbol{x}^{0}_{\text{Puffer}}$, also with deterministic inputs. The deposition module g_{DEPO} then receives the combined outputs of the previous codes and the uncertain input variables $\boldsymbol{X}_{\text{DEPO}}$, which are relevant for the uncertainty analysis. This module calculates clogging increments $\Delta^{j+k}_{j+k-1}\tau_c$ between times t_{j+k-1} and t_{j+k}, after which several checks are performed:

- if the accumulated clogging from t_j to t_{j+k} exceeds 30%, then the head losses are updated and the stationary thermal-hydraulic state is recalculated. The inputs $\boldsymbol{X}^{j}_{\text{THYC}}$ are updated, with some components becoming random due to the uncertainty in g_{DEPO}'s output,
- if chemical conditioning changes at time t_χ, solubility maps are recomputed using g_{Puffer},
- if t_{j+k} coincides with a maintenance cleaning, the clogging rate is reduced according to the cleaning type (i.e., either preventive or curative, as explained in 2.1) and g_{THYC} is rerun.

- When t_{j+k} reaches the simulation end time t_f, the code outputs the clogging rate evolution curve τ_c.

Typically, the simulation horizon is 60 years, corresponding to the expected SG operational lifetime, and a single THYC-Puffer-DEPO run takes about 5 hours on a high-performance computing platform, such as the GAIA computer at EDF[1].

2.3 Statistical model for clogging rates: ESTICOL

The tool ESTICOL (Pinciroli et al., 2021) is based on a regression algorithm using televised video examinations (TVEs) and operational data. One of the key features of the thesis use-case is that the observation data are scarce in time. Indeed, these televised measures can only be performed during NPP outages, which occur only at certain points in the asset's operational history. This means that for a specific SG, typically only a handful of at most 10 TVE measures are available. To enrich this database, alternative indirect measures (such as the *IComplexe* measures based on eddy currents, see Girard (2014)) could potentially be added to the set of outputs. However, such non-destructive testing for clogging techniques are still the subject of ongoing research, and have yet to be validated to use them confidently.

A complementary way to enrich the measurement database is to use statistical methods. The ESTICOL algorithm is built from a procedure that selects features from transient plant tests and correlates them with TVE observations; these transient tests are performed more frequently than TVEs. ESTICOL is intended to produce estimates of past and current clogging rates, i.e. the clogging level at the time of a test. The features are extracted from time series of the periodic BIL100 and EP-RGL4 tests used to monitor NPP safety. These tests are validated according to specific rules and documentation (Kerkar and Paulin, 2008), which depend on the plant power rating (900, 1300 MWe). The BIL100 measures, as accurately as possible, the thermal power generated in the reactor core; this measurement helps calibrate secondary-circuit components, including the SG. The BIL100 is performed at least monthly and is most reliable around 75% of nominal power. The thermal power is obtained via an enthalpy balance for each SG and requires properly qualified SG sensors.

The BIL100 is used as a reference for the EP-RGL4 test (Pinciroli et al., 2021; Kerkar and Paulin, 2008). The EP-RGL4 evaluates the spatial calibration of boron control rods, which regulate reactor power by absorbing neutrons via a boron solution of varying concentrations. This standardized transient typically lasta about two hours and appears as characteristic slope variations in power time series. It instantaneous employed to validate operating flexibility and monitoring capabilities. The underlying hypothesis is that features extracted from these transients can serve as proxies for changes in SG head losses potentially caused by clogging (as explained in Pinciroli et al. (2021)).
The output for ESTICOL is the clogging rate at time instant t_i for $i \in \{1, \dots, P\}$, denoted $Y^{(i)} = \tau_c(t_i)$. For the training dataset of ESTICOL on a specific SG, a feature is extracted from the closest EP-RGL4 signal within a year before the TVE, and the remaining EP-RGL4 signals are used for testing. An illustration of this principle is provided in Figure 2.7. From the training

[1] more information on the EDF R&D HPC infrastructure is available, in French, on the following EDF website

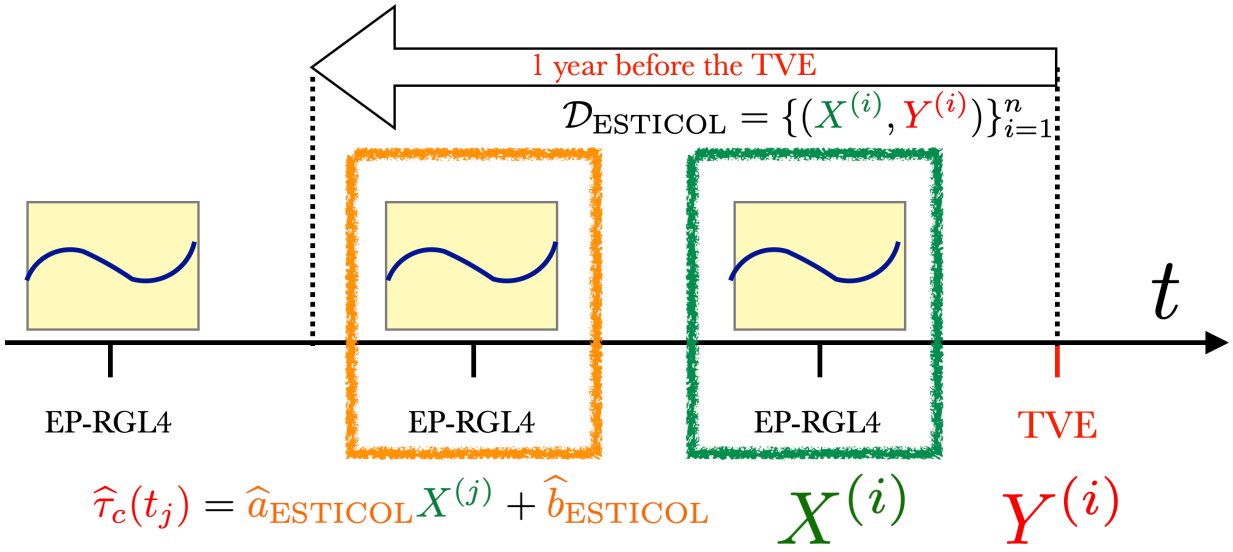

Figure 2.7: The ESTICOL algorithm.

database $\mathcal{D}_{\text{ESTICOL}} = \{(X^{(i)}, Y^{(i)})\}_{i=1}^{n}$ a weighted linear least-squares problem is solved:

$$(\widehat{a}_{\text{ESTICOL}}, \widehat{b}_{\text{ESTICOL}}) = \arg\min_{a,b} \sum_{i=1}^{P} w_i |aX^{(i)} + b - Y^{(i)}|^2, \quad \forall i, \quad w_i \propto \exp(t_i^2). \tag{2.22}$$

The weight allows to enhance the TVE values close to t_i. The estimation is then done by building the EP-RGL4 feature at a given time-instant $t \leq t_P$ and then using the affine relationship:

$$\widehat{\tau}_c(t_j) = \widehat{a}_{\text{ESTICOL}} X^{(j)} + \widehat{b}_{\text{ESTICOL}}. \tag{2.23}$$

Several practical limitations affect ESTICOL in this application. First, observations are extremely sparse in time: for a given SG one typically has at most on the order of $n = 10$ TVE measurements over its operational life. Training a model per SG with such small samples yields high estimator variance, limited ability to detect nonlinear effects, and instability of fitted coefficients (increasing the risk of overfitting). Second, TVE measurements themselves are noisy and irregularly spaced, and feature extraction from EP-RGL4 transients introduces additional timing and preprocessing uncertainties that impact predictive performance.

In spite of these issues, cross-validation strategies (for instance leave-one-out or time-aware folds) can be applied to provide a pragmatic assessment of predictive skill given the small data regime. Empirically, such validation has shown satisfactory predictive behavior for some SGs, especially those with relatively consistent operating histories and a handful of well-placed TVEs. Nonetheless, performance is heterogeneour, other SGs with few or noisy TVEs show poor generalization. These observations further motivate hybrid approaches (see section 2.2.4) that combine ESTICOL's inexpensive regression estimates with the physics-based THYC-Puffer-DEPO outputs to provide a robust RUL inference given the scarcity of the TVE data.

2.4 Digital twins for nuclear power plant components

In the context of extending the operational lifetime of the French nuclear fleet, accurate prediction of the clogging rate is essential for optimizing maintenance operations. This objective has been identified by the "Nuclear Plant of the Future" EDF program and is being pursued through the "Steam Generator Digital Twin" ("Jumeau Numérique Générateur de Vapeur" - JNGV) project (Deri et al., 2021). The JNGV initiative aims to develop a digital platform that integrates physical modeling, statistical methods, and data assimilation to estimate and forecast clogging in SGs as well as other degradation phenomena such as fouling (Prusek et al., 2013) or the mechanical rupture of TSPs.

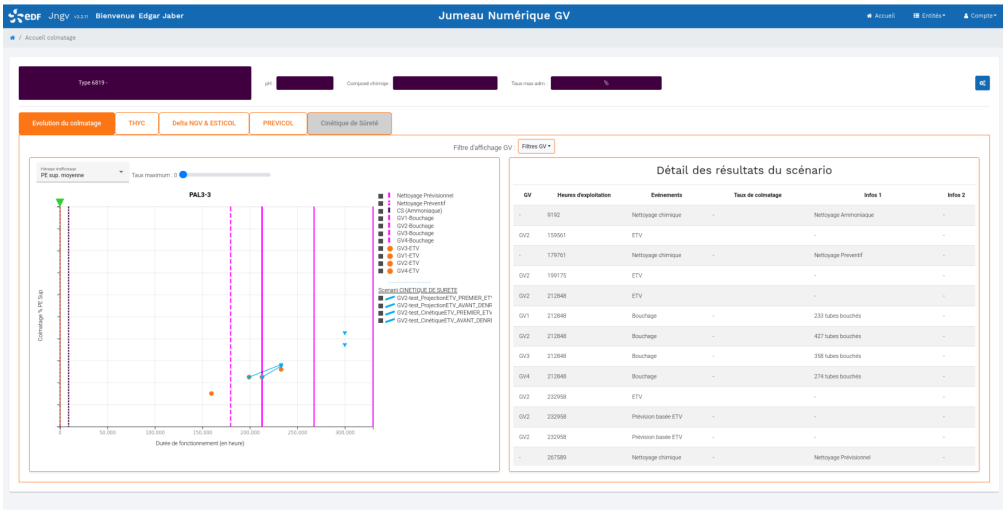

Figure 2.8: The JNGV computer interface (© EDF).

This objective falls into the broader scope of designing so-called Digital Twins (DTs) for engineering systems (in our case NPP components). DTs are an emerging field of technologies (NAS, 2024) that aims to bridge the gap between classical physical modeling and the growing amount of data in order to offer real-time guidance for informed decision-making on systems. It is therefore at the centre of our definition of hybrid methods. The standard definition used in the different communities was given by the American Institute of Aeronautics and Astronautics (AIAA, 2021) and is reproduced below:

A digital twin is a set of virtual information constructs that mimics the structure, context, and behavior of a natural, engineered, or social system (or system-of-systems), is dynamically updated with data from its physical twin, has a predictive capability, and informs decisions that realize value. The bidirectional interaction between the virtual and the physical is central to the digital twin.

For NPPs, this translates to designing systems maintaining continuous congruency with the actual state of the plant by adjusting in real-time to operational data. While still being at an early conceptual and design stage, guidelines and challenges for building such systems have been outlied in the following U.S Nuclear Regulation Committee report on digital twins (US-NRC, 2021). The main features that NPP-DTs must have is twofold. Firstly it must include modeling and simulation, including data analytics, machine learning/artificial intelligence methods, physics-based models, and

data-informed models (the latter meaning hybridizing physics-based models with real-time processing of plant data). Secondly it must have a robust data and information management system, covering accessibility, safe information and data storage as well as an operator interface with user-friendly visualization software. The report states that one of the primary purposes of NPP-DTs is to support diagnostics and prognostics of the health state and failure of structures, systems, and other NPPCs, enabling plant operators to anticipate failures and optimize preventive maintenance. The integration and use of such systems can save operating time and costs instead on relying on failure observation and subsequent action since it mitigates potential unwanted maintenance costs coupled with shutdowns of the NPP.

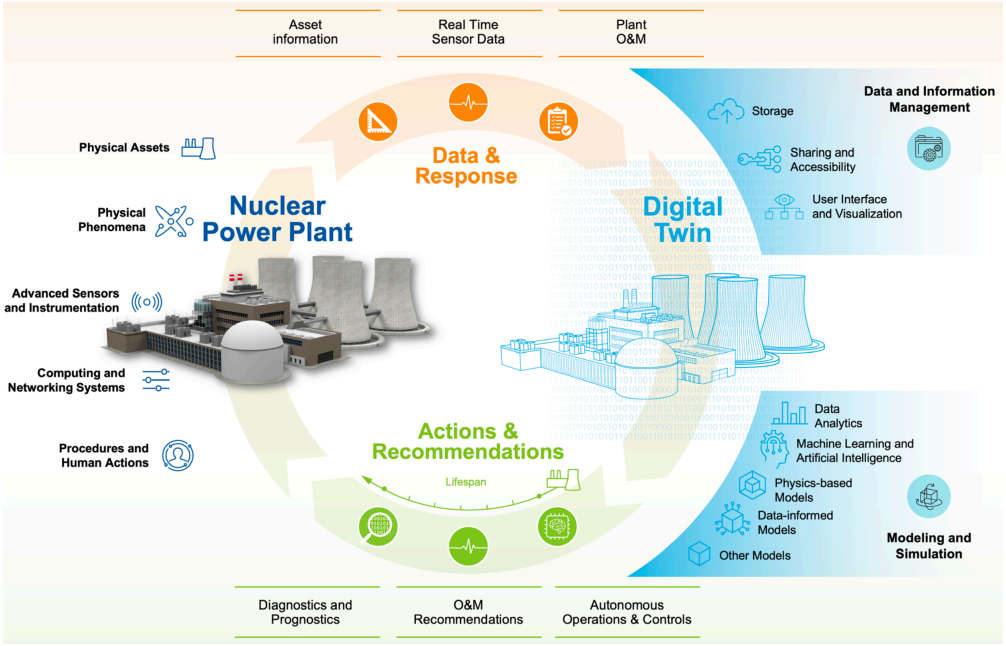

Figure 2.9: Summary of the workflow in a nuclear digital twin, (US-NRC, 2021).

Within DTs, large volumes of heterogeneous data could be processed rapidly to perform tasks such as anomaly detection, sensor malfunction identification, differentiation between true anomalies and sensor errors, failure prediction, and critical event forecasting. In the long term, DTs could enable more autonomous systems for plant operation and maintenance. However, several challenges must be addressed for effective implementation. Real-time data acquisition requires advanced sensors and instrumentation, and integrating these data streams into the DT is non-trivial, especially given the harsh plant environment. For modeling and simulation, verification, validation, and uncertainty quantification (VV&UQ) (Ghanem et al., 2017) are critical, including the propagation of aleatoric and epistemic uncertainties across coupled physical models. The lack of experimental validation data and the need to integrate heterogeneous models further complicates the process.

Data analytics faces challenges in integrating operator experience, potential experimental data, and synthetic simulation data, as well as in quantifying uncertainties and handling missing or inconsistent data. Machine learning algorithms can be valuable for processes too complex for first-principles modeling, particularly in operations, management, and design phases. Nevertheless, issues such as

the optimal amount and frequency of input data, scarcity of failure data, and quantification of output uncertainty and model performance must be considered. The selection of algorithms, input data typology, and training data complexity are also important factors. Explainability of AI models is especially significant in the nuclear domain, as regulatory acceptance depends on understanding the model's decision process. Among the different possible solutions proposed in (US-NRC, 2021), surveys can help assess the explainability of algorithms by evaluating statistics of operator satisfaction and their ability to predict model outcomes. This can also be done at an institutional level between different nuclear operators. Classical physics-based simulation models (e.g. for neutronics or thermal-hydraulics) face challenges in achieving real-time performance due to computational demands, and surrogate modeling (for instance using reduced-order models, see Kapteyn et al. (2020)) requires careful verification and validation. A current challenge (US-NRC, 2021) for data-informed models is their capability to update their internal states dynamically. This feature is key to synchronize with continuous streams of plant data rather than relying on static input variables.
Finally, a last desiderata for NPP-DTs are robust and secure data and information management systems, along with effective visualization interfaces. These are essential to ensure that nuclear operators can seamlessly interact with the different tools and can trustfully use them.

2.5 Formulation of the hybrid prognostics problem

Prognostics and Health Management (PHM) (see Eker et al. (2012); Biggio and Kastanis (2020)) is an interdisciplinary engineering field focused on monitoring, diagnosing, and forecasting the health status of industrial assets to optimize maintenance strategies. Prognostics aims to estimate the current health state (diagnostics) and predict its future evolution, enabling the calculation of the *remaining useful life* (RUL) (Escobet et al., 2019)—the time interval between the present and when the health indicator reaches a predefined threshold, signaling the need for maintenance before critical failure. Mathematically, for a time-dependent health indicator ($t \mapsto \Delta(t)$), a threshold $\Delta_* \in \mathbb{R}_+$, and current time t_P, the RUL is defined as:

$$\mathrm{RUL}(t_P, \Delta_*) = \underset{t>t_P}{\arg\min}\{\Delta(t) > \Delta_*\}. \tag{2.24}$$

Due to inherent uncertainties in both statistical and physics-based models, RUL is typically treated as a probabilistic quantity, as seen in Figure 2.10 below:
In the context of NPPs, the clogging of SGs represents a challenging PHM use-case, particularly for DT applications. While DT architectures for NPP components are emerging in the literature (Mengyan et al., 2024; Xue et al., 2025; Liu et al., 2025), EDF is currently the only major stakeholder developing DTs specifically for SGs (Deri et al., 2021) and its degradation processes. Other recent works have explored DT methodologies for predictive maintenance and operations management in NPPs, with some mostly at a design stage for new reactor technologies. These include:

- Oluwasegun and Jung (2020) develop a DT for anomaly detection and prognostics of control rod drive mechanisms using machine learning on plant instrumentation data.

- Kochunas and Huan (2021) and Zhao and Guan (2022) discuss the need for uncertainty quantification and multi-scale modeling in nuclear DTs.

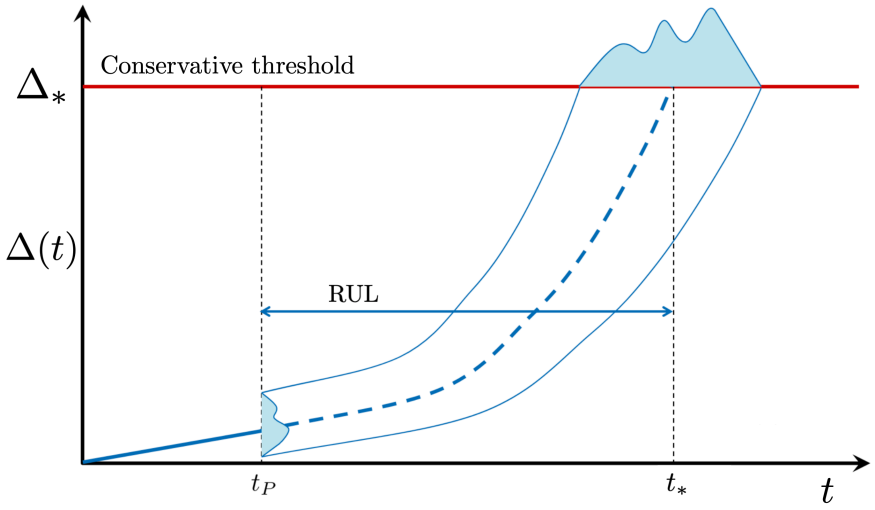

Figure 2.10: Prognostics process under uncertainty, adapted from: (Escobet et al., 2019).

- Zhang et al. (2022) present a DT for thermal performance monitoring, assimilating real plant measurements with simulation models for improved accuracy.
- Crespi et al. (2023) outline the potential of DTs across the nuclear lifecycle, emphasizing their transformative role for future reactor technologies.
- Amer et al. (2025) propose a modular probabilistic framework using dynamic Bayesian networks for maintenance scheduling in small modular reactors (SMR), integrating empirical, data-driven, and physics-based twins with real-time sensor data.
- General Electric's research project (Goldfarb et al., 2025) validates ML-based predictive maintenance DTs for SMR prototypes, focusing on remote monitoring and cost reduction.
- Rivas et al. (2025) introduce a dynamic optimization framework, combining system-level predictive maintenance via dynamic Bayesian networks and component-level assessment using deep neural networks, demonstrated on a pebble-bed reactor.

While significant progress has been made, the adoption of digital twins in the nuclear industry remains limited compared to sectors like civil engineering, primarily due to the specific challenges of nuclear prognostics. The field is still developing, but growing interest in the digitalization of future plant technologies is expected to accelerate research in this area. For the SG case study, which focuses on operational rather than conceptual aspects, the main challenges can be summarized as follows:

- *Nonlinear, multi-physics modeling:* SG clogging involves complex, coupled transport, chemical deposition, and flow dynamics over multi-year timescales, requiring computationally intensive simulations with limited benchmarking.
- *Uncertainty quantification:* The parametric uncertainty of the THYC-Puffer-DEPO simulation chain must be rigorously assessed, serving as a baseline for uncertainty reduction in hybrid prognostics, potentially leveraging surrogate modeling and conformal prediction.

- *Sparse and heterogeneous measurements:* Observations of clogging rates are not frequent over the operational period and originate from diverse sources, including field inspections (TVEs) and regression-based estimates (ESTICOL), complicating data integration.
- *Offline prognostics assessment:* Data acquisition is not continuous or on-demand, necessitating offline assimilation strategies distinct from online filtering or sequential updating methods common in other PHM domains.

In summary, the hybrid prognostics problem for SG clogging in NPPs requires integrating physics-based simulation (THYC-Puffer-DEPO) with statistical models (such as ESTICOL) and field data (such as the TVEs). The goal is to produce robust, risk-informed RUL estimates despite sparse measurements and complex system dynamics, leveraging the complementary strengths of physical modeling and data-driven approaches within a digital twin framework.

Chapter 3

Uncertainty quantification of THYC-Puffer-DEPO

This chapter presents a comprehensive uncertainty quantification and sensitivity analysis of the THYC-Puffer-DEPO simulation code, which models steam generator clogging phenomenon over asset lifespans. It stems mostly from the publication Faber et al. (2025b). The study details the probabilistic characterization of seven key input variables, the design of experiments, and the application of surrogate modeling techniques including Gaussian processes, vector Polynomial Chaos Expansions, and Karhunèn-Loève expansions. Sensitivity analysis using Sobol' indices and the Hilbert-Schmidt Independence Criterion (HSIC) identifies the iron oxide particle diameter, vena contracta calibration parameter, and initial particle mass fraction as dominant contributors to output uncertainty across chemical regimes. Notably, the porosity of clogging deposits emerges as a critical factor under high-pH conditions and long-term operation, revealing a previously unknown interplay between chemical environment and deposit microstructure.

Contents

3.1 Design of experiments and graphical analysis

For the uncertainty analysis of the THYC-Puffer-DEPO (TPD) code, a subset of $d = 7$ input variables in the DEPO mod was selected by domain experts (Lefebvre et al., 2023):

$$\boldsymbol{X}_{\text{DEPO}} = (\alpha, \beta, \epsilon_e, \epsilon_c, d_p, \Gamma_p(0), a_v)\,. \tag{3.1}$$

The vector components are assumed to be independent, and they are found in the equations of the deposition mechanisms in section 2.2.3. The empirical correlation parameters α, β from Eq. (2.15) are jointly estimated by a regression procedure. Therefore they could be subject to a correlation, but as we will see in the results, they have a negligible contribution to the clogging uncertainty. This vector includes only the uncertain parameters relevant for the study, rather than all possible inputs to g_{DEPO}, like for instance the initial value of soluble species $\Gamma_s(0)$ or geometric parameters found in Eq. (2.11). Notably, the solubility map $\Gamma_s^{\max}$ is excluded, as the current version of the code (which now incorporates g_{Puffer}) handles solubility differently compared to the earlier COLMATHYC-based approach described in Lefebvre et al. (2023). The probabilistic characterization of these variables, summarized in Table 3.1, is based on expert elicitation and remains consistent with previous work (Lefebvre et al., 2023). The input distributions are either Gaussian, $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ or triangular, $X_i \sim \mathcal{T}(a_i, b_i, c_i)$. These physical parameters are directly involved in the model equations, particularly in the expression for the clogging deposition velocity constant expressed in Eq. (2.10).

Variable	Signification	Distribution
α	First empirical correlation parameter in Eq. (2.15)	$\mathcal{N}(101.6, 4.0^2)$
β	Second empirical correlation parameter in Eq. (2.15)	$\mathcal{N}(0.0233, 0.0005^2)$
ϵ_e	Porosities of the fouling deposits	$\mathcal{T}(0.2, 0.3, 0.5)$
ϵ_c	Porosities of the clogging deposits	$\mathcal{T}(0.01, 0.05, 0.3)$
d_p	Iron oxide particle diameter (m)	$\mathcal{T}(0.5, 5.0, 10.0) \times 10^{-6}$
$\Gamma_p(0)$	Initial data for solid particles mass fraction transport equation	$\mathcal{T}(1.0, 4.5, 8.0) \times 10^{-9}$
a_v	Calibration parameter of the *vena contracta* physical mechanism	$\mathcal{T}(0.1, 7.8, 12) \times 10^{-4}$

Table 3.1: Probabilistic modeling of uncertain input variables.

The complete set of numerical results for uncertainty propagation and sensitivity analysis was produced by coupling the TPD computational chain with the OpenTURNS software package ("An Open source initiative for the Treatment of Uncertainties, Risks'N Statistics", an open-source C++/Python library developed by a consortium including Airbus Group, EDF, IMACS, ONERA, and Phimeca (Baudin et al., 2017). All scripts used in this section are available in the following GitHub repository.

Crude Monte Carlo simulations were performed by generating $n = 10^3$ samples from the joint probability distribution of $\boldsymbol{X}_{\text{DEPO}}$ as defined in Table 3.1 on a 51B type SG. To reduce computational time, simulations were executed in parallel on high performance computing facility infrastructure, covering a $t_f = 60$ year period (requiring approximately one week of computation on high-performance infrastructure). The resulting trajectories are presented in Figure 3.1. Analysis of the simulation results shown in Figure 3.1 leads to several observations. First, for this particular steam generator, two types of cleaning procedures have been implemented over its operational

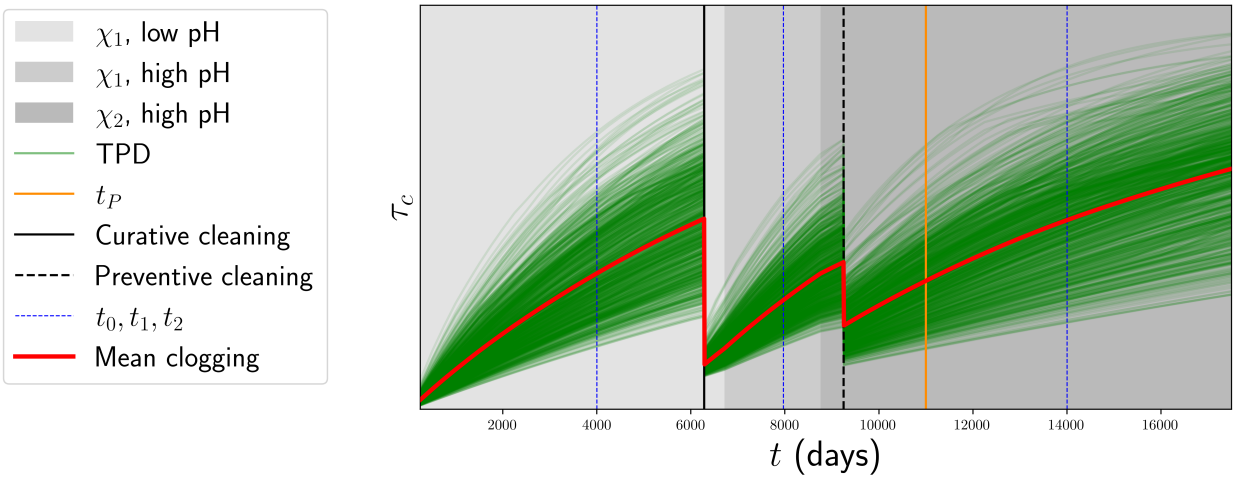

Figure 3.1: Clogging trajectories obtained from Monte Carlo simulations (with curative and preventive chemical cleanings).

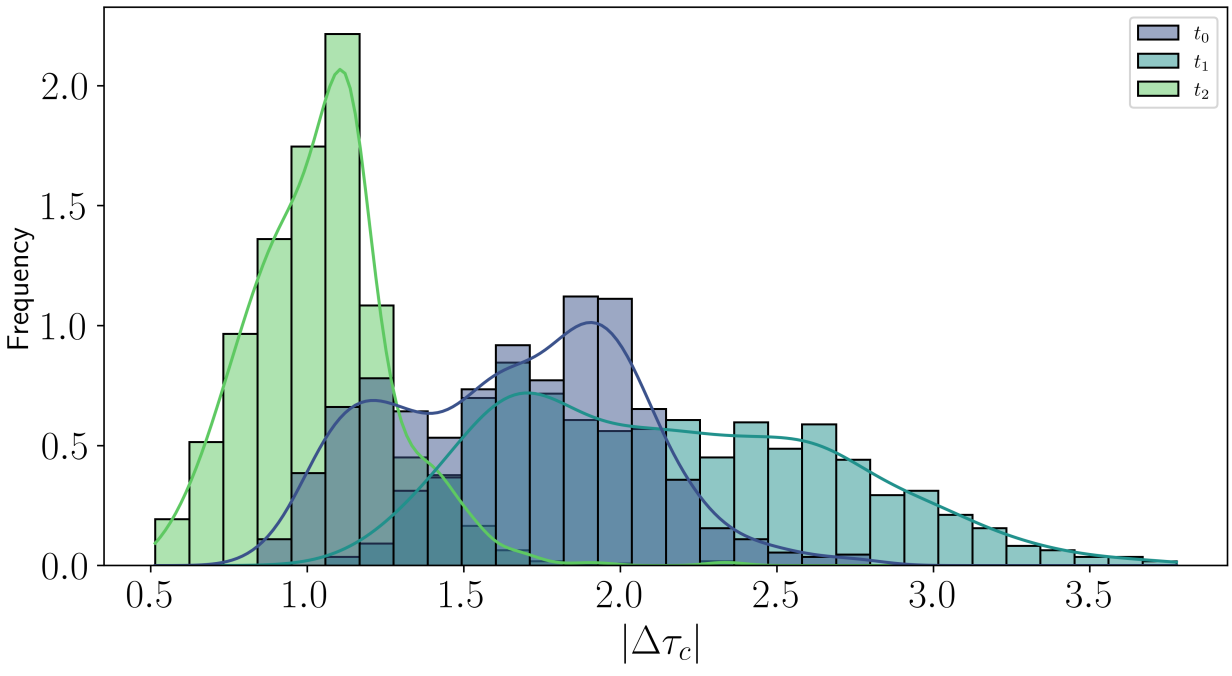

Figure 3.2: Histograms and kernel density estimation of clogging kinetics for various chemical regimes.

lifetime: curative and preventive. It is generally recognized by operators that preventive cleaning is less effective at reducing clogging rates than curative cleaning. Second, the secondary fluid has experienced three different chemical regimes at two pH levels, denoted as $\chi_i,\ i \in \{1, 2\}$ for the main chemical species, with pH categorized as either low or high. These different chemical regimes result in noticeable variations in clogging kinetics. This is clearly illustrated in Figure 3.2, which presents the distributions of clogging increments at various times under specific chemical conditions (χ_1-low pH, χ_1-high pH, χ_2-high pH). The results indicate a slowdown in simulated clogging kinetics when a more basic (higher pH) chemical regime is used, which is consistent with operational experience. Beyond the time-dependent analysis, Figure 3.3 provides scatter plots in the rank-space, showing the dependencies between inputs and output at three selected time points $t_0 < t_1 < t_2$, each corresponding to a different chemical regime. The dependency patterns ob-

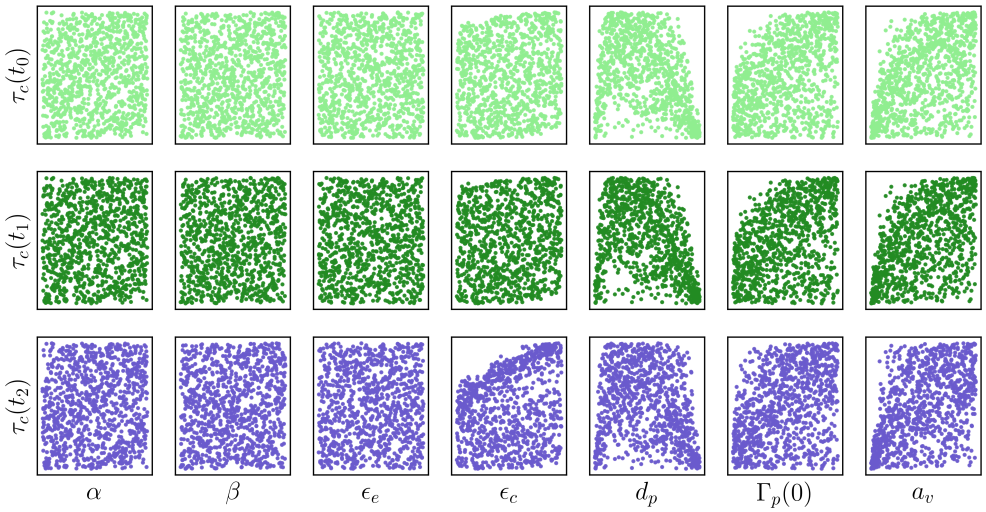

Figure 3.3: Scatter plots in the rank-space for the three different chemical conditionings.

served in these scatter plots are consistent with those reported in Lefebvre et al. (2023). Notably, a nonlinear dependence is evident between the iron-oxide particle diameter d_p and the clogging rate τ_c across all time points. Physically, this indicates that particles with intermediate diameters are primarily responsible for higher clogging rates, regardless of the chemical conditioning or clogging level. Furthermore, at time point t_2 (corresponding to the χ_2 high-pH regime), a pronounced nonlinear correlation emerges between the clogging deposit porosity ϵ_c and τ_c. This finding is new compared to the previous study (Lefebvre et al., 2023), where fouling and clogging porosities were assumed equal ($\epsilon_e = \epsilon_c = \epsilon$), leading to a lack of influence in sensitivity analysis. It is also important to note that, in the earlier work, pH had little impact since solubility maps were constant and distributed according to the prescribed uncertainty law. In contrast, the updated clogging computational chain TPD used here incorporates the Puffer code, which generates mesh-dependent iron oxide solubility maps reflecting different chemical conditionings. Additionally, linear correlations (with a sort of Gaussian copula pattern seen empirically) are observed between the clogging rate and both the initial particle mass-fraction $\Gamma_p(0)$ and the calibration parameter a_v. As in the previous study, no significant link is found with the empirical correlation parameters α and β, nor with the fouling porosity ϵ_e. These preliminary visual insights into input-output dependency structures will be further examined and quantified through the sensitivity analysis results presented in the following

sections.

3.2 Metamodeling techniques applied to TPD

In the rest of this section, we denote a generic computer model as g, and we do the same for THYC-Puffer-DEPO, thus whenever stated $g_{\text{TPD}} =: g$. Metamodels, also known as surrogate models, are essential in uncertainty quantification (De Rocquigny et al., 2008; Sullivan, 2015) because they approximate the behavior of complex computational codes with much lower computational cost. By replacing expensive simulations with fast, predictive models, metamodels enable efficient exploration of input uncertainties and facilitate downstream tasks such as sensitivity analysis and for inverse problems using Markov Chain Monte Carlo (MCMC) methods (discussed extensively in chapter 5). This acceleration is crucial when repeated evaluations of the original code would be prohibitively time-consuming.

In this section, we study several metamodeling approaches: Gaussian process regression (Kriging) in 3.2.1, vector-valued Polynomial Chaos Expansions (VPCE) in 3.2.2, and field metamodeling using Karhunèn-Loève expansions (KLE) in 3.2.3. Each method offers distinct advantages in terms of flexibility, interpretability, and computational efficiency. Using multiple surrogate models allows us to compare their predictive performance, assess robustness, and mitigate model-specific biases. Aggregating results from different surrogates can improve stability and generalizability, as we will explain in 3.2.4 ensuring that conclusions drawn from uncertainty and sensitivity analyses are not overly dependent on the assumptions of a single modeling approach. We emphasize that the same Design of Experiments (DoE) presented in 3.1 is used for building all the approximations.

3.2.1 Gaussian process regression or Kriging

Reminders

Assume a design of experiments $\mathcal{D}_n = \{(\boldsymbol{X}^{(i)}, g(\boldsymbol{X}^{(i)}))\}_{i=1}^n$ and define $\boldsymbol{X} := (\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(n)})$ the input sample matrix and $\boldsymbol{g}(\boldsymbol{X}) = (g(\boldsymbol{X}^{(1)}), \dots, g(\boldsymbol{X}^{(n)}))$ the output sample matrix. To construct a Gaussian process (GP) metamodel (Rasmussen and Williams, 2006) for g, one assumes that g is a realization of a GP $\mathcal{GP}(M, K)$, where $M : \mathcal{X} \to \mathcal{Y}$ is the mean function and $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is the covariance kernel. Conditioning this process on the available dataset $\mathcal{D}_n$ corresponds to Bayesian regression with a Gaussian prior on g, yielding a posterior distribution. When the outputs are noise-free, this regression reduces to GP interpolation, which is the standard approach for building GP surrogates of deterministic codes (Fang et al., 2006). Figure 3.4 illustrates this principle. For simplicity, set $M = 0$ (the so-called "ordinary Kriging") and use a Matérn-ν kernel, defined for $\nu = (2k+1)/2$, $k \in \mathbb{N}$, and $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$ as:

$$\begin{aligned} K(\boldsymbol{x}, \boldsymbol{x}') &= K_{(\nu,\theta,\sigma)}(\boldsymbol{x}, \boldsymbol{x}') \\ &= \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{|\boldsymbol{x} - \boldsymbol{x}'|}{\theta} \right)^{\nu} K_\nu \left(\sqrt{2\nu} \frac{|\boldsymbol{x} - \boldsymbol{x}'|}{\theta} \right), \end{aligned} \tag{3.2}$$

where K_ν is the modified Bessel function of the second kind and Γ is the gamma function. Hyperparameter ν controls the regularity of the process, as sample paths belong to $\mathcal{C}^{\lfloor \nu - 1 \rfloor}(\mathcal{X})$ (Gu et al.,

2018). The conditional process $\widetilde{\mathcal{G}} := \mathcal{G}|\mathcal{D}_n$ is a GP with posterior mean and covariance, for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$:

$$\widetilde{g}(\boldsymbol{x}) := k(\boldsymbol{x})^\top \boldsymbol{K}^{-1} \boldsymbol{g}(\boldsymbol{X}), \quad \widetilde{K}(\boldsymbol{x}, \boldsymbol{x}') := K(\boldsymbol{x}, \boldsymbol{x}') - k(\boldsymbol{x})^\top \boldsymbol{K}^{-1} k(\boldsymbol{x}'), \tag{3.3}$$

where

$$k(\boldsymbol{x}) := \left(K(\boldsymbol{x}, \boldsymbol{X}^{(1)}), \ldots, K(\boldsymbol{x}, \boldsymbol{X}^{(n)})\right)^\top \in \mathbb{R}^n, \quad \boldsymbol{K} := \left(K(\boldsymbol{X}^{(i)}, \boldsymbol{X}^{(j)})\right)_{1 \leq i,j \leq n} \in \mathcal{M}_n(\mathbb{R}). \tag{3.4}$$

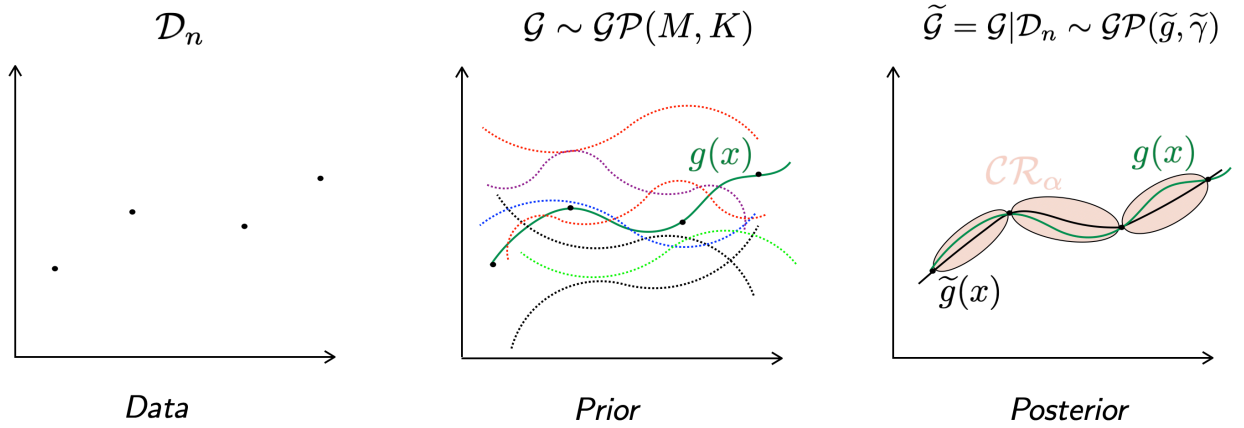

Figure 3.4: Illustration of GP interpolation metamodeling. The input-output DoE is computed with the deterministic code g. The data are assumed to be a realization of an underlying GP. In the absence of noise, the posterior process interpolates the data and provides nonzero credibility intervals outside $\mathcal{D}_n$.

The posterior mean (or mean predictor) $\widetilde{g}$ serves as a metamodel for g, i.e., $\widehat{g} = \widetilde{g}$. The kernel hyperparameters (σ^2, θ), for a given ν, can be estimated via maximum likelihood estimation (MLE) or cross-validation (CV) (Acharki et al., 2023). In ordinary Kriging, the MLE problem is:

$$(\sigma^2_{\text{MLE}}, \theta_{\text{MLE}}) \in \underset{(\sigma^2,\theta)}{\arg\min} \left\{ \boldsymbol{g}(\boldsymbol{X})^\top \boldsymbol{K}^{-1} \boldsymbol{g}(\boldsymbol{X}) + \log(\det \boldsymbol{K}) \right\}. \tag{3.5}$$

As shown in Acharki et al. (2023), MLE is preferable when the kernel is well-specified, while CV is more robust to misspecification, though this may not always hold for GP interpolation (Petit et al., 2023). The posterior standard deviation at $\boldsymbol{x}$ is given by:

$$\widetilde{\gamma}(\boldsymbol{x}) := \widetilde{K}^{1/2}(\boldsymbol{x}, \boldsymbol{x}). \tag{3.6}$$

For $\boldsymbol{x} \in \{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(n)}\}$, we have $\widetilde{\gamma}(\boldsymbol{x}) = 0$ and $\widetilde{g}(\boldsymbol{x}) = g(\boldsymbol{x})$, so the GP metamodel interpolates the data. If the outputs are noisy, the standard GP regression setting applies, and the covariance matrix includes a so-called "nugget effect" of the following form:

$$\boldsymbol{K}_\epsilon := \boldsymbol{K} + \sigma^2_\epsilon \boldsymbol{I}_n, \tag{3.7}$$

where $\boldsymbol{I}_n$ is the $n \times n$ identity matrix. The noise hyperparameter σ_ϵ is estimated by maximizing the full likelihood:

$$(\sigma^2_{\text{MLE}}, \theta_{\text{MLE}}, \sigma^2_{\epsilon,\text{MLE}}) \in \underset{(\sigma^2,\theta,\sigma^2_\epsilon)}{\arg\min} \left\{ \boldsymbol{g}(\boldsymbol{X})^\top \boldsymbol{K}_\epsilon^{-1} \boldsymbol{g}(\boldsymbol{X}) + \log(\det \boldsymbol{K}_\epsilon) \right\}. \tag{3.8}$$

Application to THYC-Puffer-DEPO

A naive GP constructed on all the output time instances independently does not take into account the temporal correlation. In this chapter, a stochastic method is proposed in order us to recover the right temporal evolution by using time-collocation points. In the probability modeling described in Table 3.1, the uncertain parameters are modeled using expert-informed probability densities which we denote by $p_{\boldsymbol{X}}$. We then assign a uniform probability density on the simulation timeline $[0, T]$ discretized at the hourly scale. Indeed, the output $g(\boldsymbol{X}) = (g(t_1, \boldsymbol{X}), \dots, g(t_N, \boldsymbol{X})) \in \mathbb{R}^N$ with N fixed such that the convergence of the simulation is reached. Typically $N \propto 100$ and the t_i correspond to the daily discretization of the simulation. Knowing from the equations that, locally, the clogging rate is smooth and strictly increasing (as seen in Eq. (2.19)), we can assume that the clogging rate is piecewise linear between the hourly discretization. We augment N to $D \gg N$ post-simulations with the help of linear interpolators $\boldsymbol{\mathcal{G}} : \mathbb{R}^N \to \mathbb{R}^D$ to get:

$$\boldsymbol{\mathcal{G}}(g(\boldsymbol{X})) = \left(\sum_{k=1}^{N} g(t_k, \boldsymbol{X}) \varphi_k(t_1), \dots, \sum_{k=1}^{N} g(t_k, \boldsymbol{X}) \varphi_k(t_D) \right) = Z(\boldsymbol{X}) \in \mathbb{R}^D, \tag{3.9}$$

where φ_k are the linear interpolators. The probability density on the timeline is then defined as $\mathcal{U}\{1, \dots, D\}$ and we draw independent samples $(t, \boldsymbol{X}) \sim \mathcal{U}\{1, \dots, D\} \otimes p_{\boldsymbol{X}}$. For each sample $(t^{(i)}, \boldsymbol{X}^{(i)})$, we first apply the interpolation operator to the output of the second marginal to get $\boldsymbol{\mathcal{G}} \circ g(\boldsymbol{X}^{(i)}) =: Z(\boldsymbol{X}^{(i)})$, and then we project the first marginal $t^{(i)}$ to get the clogging rate at time $t^{(i)}$. This is summed up in the Eq. (3.10) below:

$$(t^{(i)}, \boldsymbol{X}^{(i)}) \xrightarrow[\text{interpolation}]{\mathcal{L}} (\boldsymbol{X}^{(i)}, \boldsymbol{\mathcal{G}}(g(\boldsymbol{X}^{(i)}))) \xrightarrow[\text{projection}]{\text{pr}_i} ((t^{(i)}, \boldsymbol{X}^{(i)}), \text{pr}_i \circ Z(\boldsymbol{X}^{(i)})). \tag{3.10}$$

We thus create a DoE with a scalar output on which to build and optimize a regular scalar GP with an MLE method. An illustration of this DoE is given in the Fig. 3.5 below:

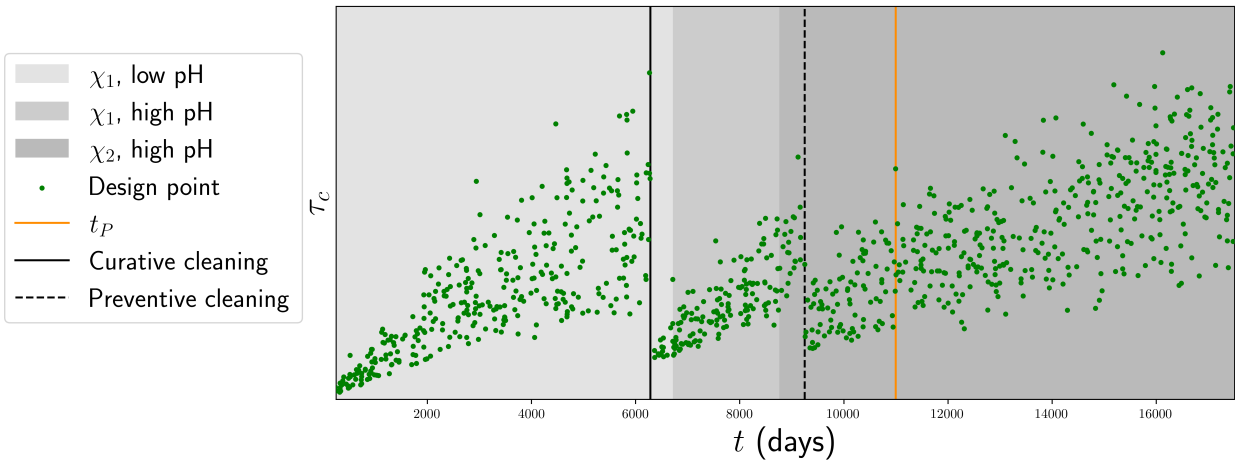

Figure 3.5: Stochastic time-collocation designs of experiments for THYC-Puffer-DEPO.

Afterwards, the GP metamodel of TPD is estimated by parametrizing the first marginal of the GP of Z, to the time indices of TPD, in other words:

$$\widehat{g}(\boldsymbol{X}) = (\widetilde{Z}(t_1, \boldsymbol{X}), \dots, \widetilde{Z}(t_N, \boldsymbol{X})). \tag{3.11}$$

The resulting trajectories as well as their average predictive coefficient are plotted in Figure 3.6. We consider three types of prior means: constant, linear and quadratic. The five covariance kernel priors considered are the absolute exponential (AbsExp), the squared exponential (SqExp) and the Matérn-ν for $\nu = 1/2, 3/2, 5/2$. The quality of the metamodel is measured through the use of the *predictivity coefficient* Q^2 (Fekhari et al., 2023). Since the problem is time-dependent, this coefficient is estimated at each time step, using a test dataset $\mathcal{D}_m = \{(\boldsymbol{X}^{(j)}, g(t_k, \boldsymbol{X}^{(j)}))\}_{k,j \in \{1,\dots,N\}\times\{1,\dots,m\}}$ of size $m < n$, typically coming from holdout validation train-test split on the original DoE. Thus, one has:

$$\forall k \in \{1, \dots, N\}, \quad Q^2(t_k) = 1 - \sum_{j=1}^{m} \frac{\left|g(t_k, \boldsymbol{X}^{(j)}) - \widehat{g}(t_k, \boldsymbol{X}^{(j)})\right|^2}{\widehat{\mathrm{Var}}(g(t_k, \boldsymbol{X}^{(j)}))}, \tag{3.12}$$

and we finally consider $\overline{Q^2}$, the time-averaged predictivity coefficient given by:

$$\overline{Q^2} = \frac{1}{N} \sum_{k=1}^{N} Q^2(t_k). \tag{3.13}$$

By looking at the time-averaged predictivity coefficient Q^2, we observe that our methodology mostly achieves good predictive performance on average. However, locally there are some problems due to the strong discontinuity created by the chemical cleanings. As is typical with GPs, our method is sensitive to the input dimension of the problem, and observe when looking at only one marginal $\theta = a_v$ (see appendix, Fig. 8.1), that the discontinuity is recovered for Matérn kernels as well as the absolute exponential. It is hard to say what would be the best GP surrogate in this case since they all achieve have a pretty good performance. It is worth nothing that discontinuity is also artificial and can vary between different SGs, so one is not necessarily interested in recovering it exactly. Moreover if we look locally around the discontinuities, there is a severe drop in the Q^2.

3.2.2 Vector-valued polynomial chaos expansions

Reminders

It is assumed the vector model g lies in $L^2_{p_{\boldsymbol{X}}}(\mathcal{X}; \mathbb{R}^N)$, which is the space of vector-valued measurable and square-integrable functions with respect to the probability measure $p_{\boldsymbol{X}}$. A Hilbert basis of tensorized orthonormal polynomials $\langle \psi_{\boldsymbol{\alpha}} \rangle_{\boldsymbol{\alpha} \in \mathbb{N}^d}$ is chosen (Xiu and Karniadakis, 2002) where for all $\alpha \in \mathbb{N}^d$, one has $\psi_{\boldsymbol{\alpha}} = \psi_{\alpha_1} \otimes \dots \psi_{\alpha_d}$. For Gaussian distributions, Hermite polynomials are used and for triangular distributions, since the orthonormal family is not analytically explicit, an adaptive Stieltjes algorithm (Gautschi, 2004, Ch .2, Sec. 2.2.3.1) is used for building it (to be more specific, an adaptive version of the initial algorithm based on a Gauss-Kronrod quadrature rule can be implemented, as done in OpenTURNS). For uniform distributions $\mathcal{U}[-1, 1]$, the orthonormal Hilbert basis is defined using the Legendre polynomials:

$$\psi_i = \sqrt{2i + 1/2}\, P_i, \quad \forall i \in \mathbb{N}, \tag{3.14}$$

where P_i are generated by the three-term recurrence relation:

$$\begin{aligned} &P_0(x) = 1 \\ &(i+1)P_{i+1}(x) = (2i+1)xP_i(x) - iP_{i-1}(x), \ \forall i \in \mathbb{N}. \end{aligned} \tag{3.15}$$

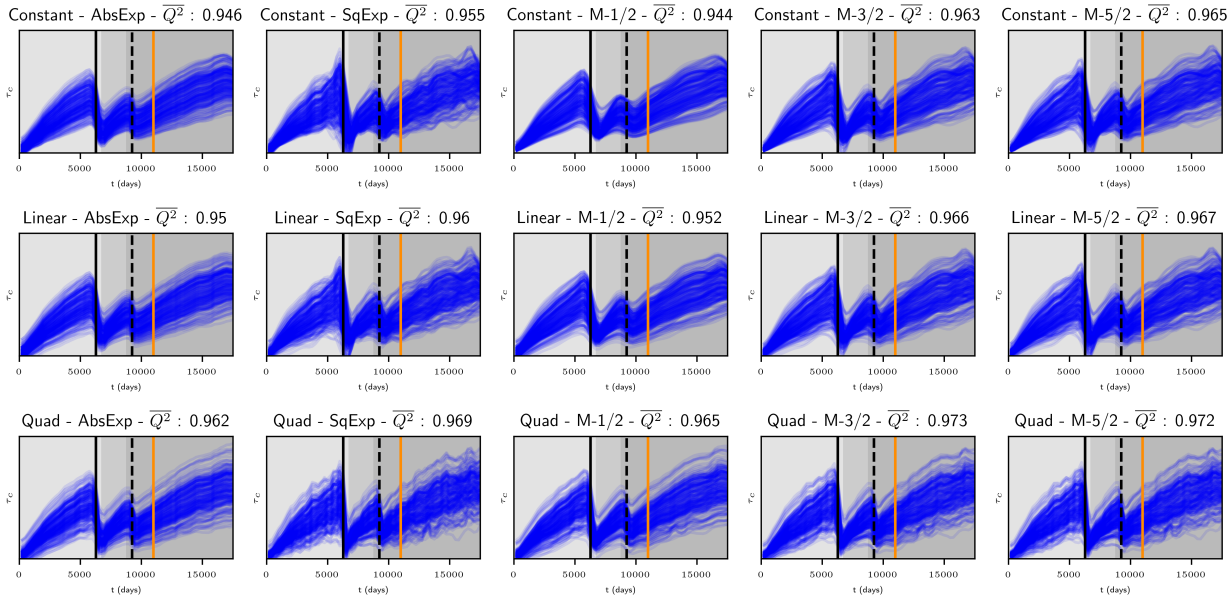

Figure 3.6: GP metamodels of TPD with different priors built with the stochastic-collocation methodology. The input dimension is $d = 1 + 7$, considering the time variable as well as the 7 variables in $\boldsymbol{X}_{\text{DEPO}}$.

At each time step, the vector polynomial chaos expansion (VPCE) of highest degree $p \in \mathbb{N}$ is constructed as the approximation:

$$\widetilde{g}(\boldsymbol{X}) = \sum_{\boldsymbol{\alpha} \in \mathcal{J}} \boldsymbol{g}_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \,, \tag{3.16}$$

where $\boldsymbol{g}_{\boldsymbol{\alpha}} = (g_{\boldsymbol{\alpha}}(t_1), \ldots, g_{\boldsymbol{\alpha}}(t_N)) \in \mathbb{R}^N$ for all $\boldsymbol{\alpha} \in \mathcal{J} \subseteq \{\boldsymbol{\alpha} \in \mathbb{N}^d, \sum_{i=1}^d |\alpha_i| \leq p\}$. The coefficient vector components will be denoted by $g_{\boldsymbol{\alpha}}^k := g_{\boldsymbol{\alpha}}(t_k)$. The index set $\mathcal{J} = \mathcal{J}_1 \cup \ldots \cup \mathcal{J}_N$ is determined by performing, for each time-step t_k with $k \in \{1, \ldots, N\}$, the following two-step method (see Blatman and Sudret (2009) and Blatman and Sudret (2011)):

- A hyperbolic enumeration rule of quasi-norm $0 < q \leq 1$ is chosen. This method allows to enhance the selection of marginal indices in the sparse basis for a choice of small quasi-norm;
- The coefficients $\mathcal{G}^k := \{g_{\boldsymbol{\alpha}}^k\}_{\boldsymbol{\alpha} \in \mathcal{J}_k}$ can be calculated in two ways:
  - Either by solving the following least-squares problem using the"least-angle regression" (LARS) method introduced by Blatman and Sudret (2011) (see Lüthen et al. (2021) for a comprehensive overview of these techniques in the context of sparse PCEs):

$$\mathcal{G}^k = \underset{g^k \in \mathbb{R}^{|\mathcal{J}_k|}}{\arg\min} \; \mathbb{E}\left(\left| g(t_k, \boldsymbol{X}) - \sum_{\boldsymbol{\alpha} \in \mathcal{J}_k} g_{\alpha}^k \psi_{\boldsymbol{\alpha}}(\boldsymbol{X}) \right|^2 \right) \,; \tag{3.17}$$

  - Or by performing quadrature, since following Hilbert space theory, one can show that:

$$g_{\boldsymbol{\alpha}}^k = \frac{\mathbb{E}_{\boldsymbol{X}}[g(t_k, \boldsymbol{X}) \psi_{\boldsymbol{\alpha}}(\boldsymbol{X})]}{\|\psi_{\boldsymbol{\alpha}}\|_{L^2_{\boldsymbol{X}}}^2}, \tag{3.18}$$

and if the polynomials are from a known family (such as Hermite or Legendre), the second-order moments have a closed-form expression. Moreover, the first term on the numerator can be expanded as follows:

$$\mathbb{E}_{\boldsymbol{X}}[g(t_k, \boldsymbol{X})\psi_{\boldsymbol{\alpha}}(\boldsymbol{X})] = \int_{\mathcal{X}} g(t_k, \boldsymbol{x})\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})dp_{\boldsymbol{X}}(\boldsymbol{x}) \simeq \sum_{i=1}^{n} g(t_k, \boldsymbol{X}^{(i)})\psi_{\boldsymbol{\alpha}}(\boldsymbol{X}^{(i)})w_i, \tag{3.19}$$

where the weights $\{w_i\}_{i=1}^n$ are estimated using a quadrature scheme such as Latin hypercube sampling, Gauss quadrature or even Monte Carlo sampling.

Once the coefficients have been computed at each time-step, the index set is built by interweaving the basis terms in order to to get the full expansion as in Eq. (3.16).

Application to THYC-Puffer-DEPO

The adoption of a vector-valued PCE metamodel for TPD is well-justified by several factors. First, the output dimension is moderate ($N \approx 100$), and the available training sample size ($n = 10^3$) is sufficient to support accurate polynomial expansions without overfitting. Second, PCEs offer strong theoretical guarantees for functional approximation in L^2 spaces, making them particularly suitable for problems with smooth, low-noise outputs (Le Maître and Knio, 2010). Crucially, PCE metamodels enable direct, analytic computation of Sobol' sensitivity indices as a natural by-product, streamlining global sensitivity analysis and avoiding the need for costly resampling or Monte Carlo procedures (Sudret, 2008; Da Veiga et al., 2021). This property is especially valuable for uncertainty quantification in computationally intensive codes. The methodology implemented here follows recent best practices in the literature (El Garroussi et al., 2022; Méndez Rojano et al., 2022), employing LARS regression for sparse basis selection and hyperparameter optimization via grid search over polynomial degree p and quasi-norm q. The optimal configuration identified is $p = 8$, $q = 0.8$, balancing model complexity and predictive accuracy. The trajectories on the test sample are displayed below:

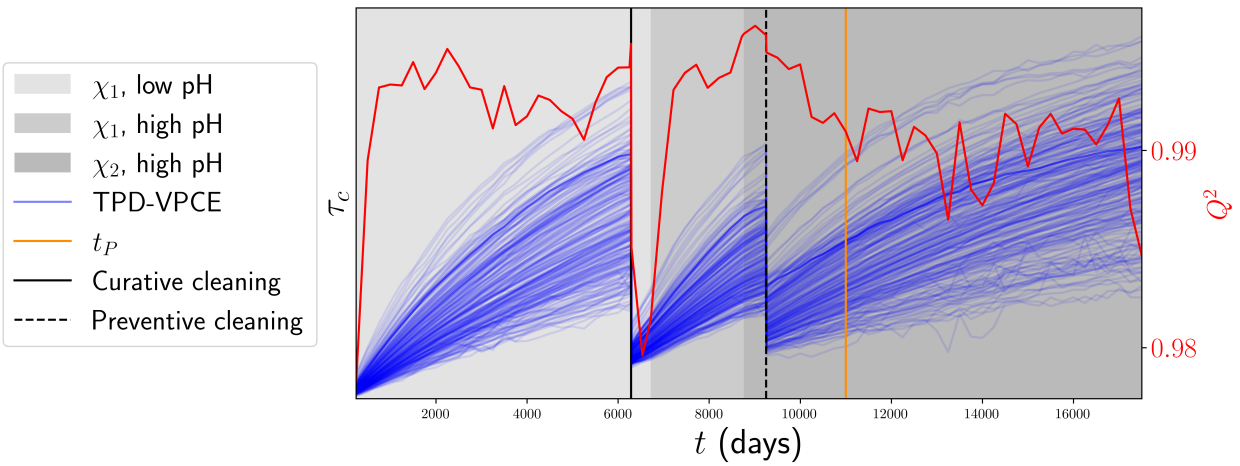

Figure 3.7: VPCE of TPD with best hyperparameters p, q by choice of $\overline{Q^2}$ optimized by a grid-search algorithm, and time-varying predictivity coefficient.

3.2.3 Field metamodeling using Karhunèn-Loève expansions

We consider a time-dependent simulation model $g : \mathcal{X} \to \mathbb{R}^N$, where, as before, each output $g(\boldsymbol{X}) = (g(t_1, \boldsymbol{X}), \ldots, g(t_N, \boldsymbol{X}))$ is a trajectory discretized on a time grid $\{t_1, \ldots, t_N\}$. The model is evaluated at n input samples $\{\boldsymbol{X}^{(i)}\}_{i=1}^n \sim p_{\boldsymbol{X}}$, resulting in an output data matrix:

$$\boldsymbol{g}(\boldsymbol{X}) = \boldsymbol{Y} = \left[g(\boldsymbol{X}^{(1)}), \ldots, g(\boldsymbol{X}^{(n)})\right] \in \mathbb{R}^{N \times n}. \tag{3.20}$$

To reduce the dimensionality of the output space, we apply a Karhunèn–Loève expansion (KLE) (Le Maître and Knio, 2010; Sullivan, 2015). To do so, one first needs to consider the empirical covariance matrix:

$$\widehat{C} = \frac{1}{n}\boldsymbol{Y}\boldsymbol{Y}^\top. \tag{3.21}$$

Then, perform a singular value decomposition (SVD) of $\boldsymbol{Y}$ such that:

$$\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{W}^\top, \tag{3.22}$$

where $\boldsymbol{V} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_m)^\top \in \mathbb{R}^{N \times m}$ contains the Karhunèn-Loève orthonormal eigenvectors (also known as "KLE modes"), $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ with σ_i^2 the empirical output variances,

$$\sigma_1 \geq \ldots \geq \sigma_m \geq 0,$$

and $m \leq \min(N, n)$ is the number of retained modes (or truncation level). The truncation level m is chosen to capture a prescribed proportion of total variance, explained by the KLE. Each trajectory $g(\boldsymbol{X}^{(i)})$ is then projected onto the retained KLE modes:

$$\xi_k(\boldsymbol{X}^{(i)}) = \langle g(\boldsymbol{X}^{(i)}), \boldsymbol{\Phi}_k \rangle = g(\boldsymbol{X}^{(i)})^\top \boldsymbol{\Phi}_k. \tag{3.23}$$

This yields m scalar-valued datasets denoted by:

$$\mathrm{DoE}_k = \left\{\left(\boldsymbol{X}^{(i)}, \xi_k(\boldsymbol{X}^{(i)})\right)\right\}_{i=1}^n, \quad k = 1, \ldots, m. \tag{3.24}$$

For each mode k, we construct a surrogate model $\widehat{\xi}_k(\boldsymbol{X})$, typically a Gaussian process or a polynomial chaos expansion. The full trajectory is then reconstructed as:

$$\widehat{g}(\boldsymbol{X}) = \sum_{k=1}^{m} \widehat{\xi}_k(\boldsymbol{X})\boldsymbol{\Phi}_k. \tag{3.25}$$

Application to THYC-Puffer-DEPO

A Karhunèn-Loève decomposition was applied to the output trajectories from the DoE. Analysis of the empirical output variances indicates that retaining two modes captures over 99% of the total variance. Accordingly, two Gaussian process metamodels were constructed to model the KLE modes using quadratic prior mean and absolute exponential covariance kernel. The resulting surrogate trajectories on the test sample are shown in Figure 3.8. While the predictive coefficient Q^2 drops noticeably for the trajectories between the two chemical cleanings, this is primarily attributed to the assumption of regular time discretization when performing the KLE decomposition, which does not fully match the TPD simulation data. In the other simulation regions, the Q^2 values remain high, indicating strong predictive performance.

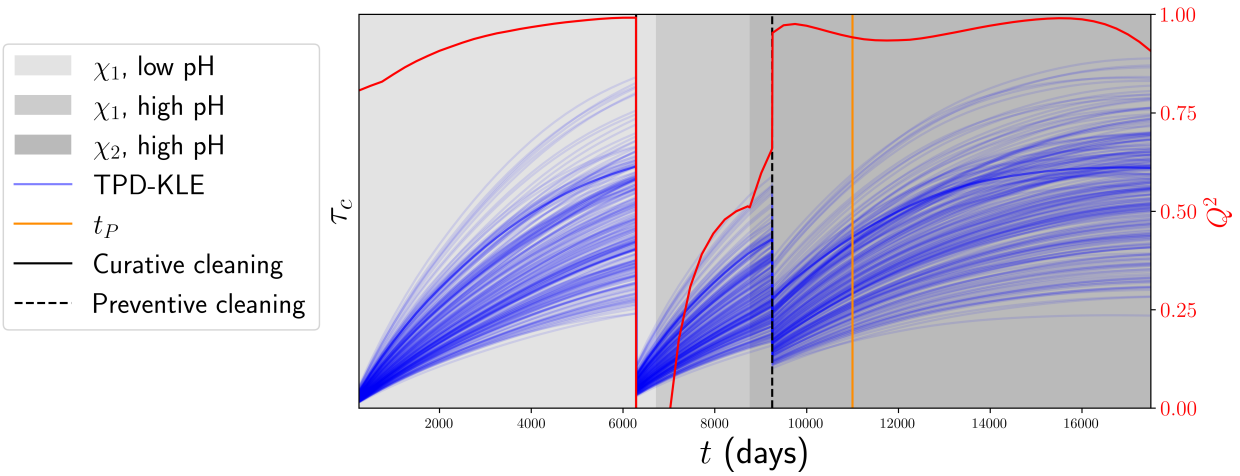

Figure 3.8: KLE-GP metamodel using a GP with prior quadratic mean and absolute exponential covariance kernel for learning the modes.

3.2.4 Metamodel aggregation strategy

Since in practice it is possible to use surrogates with different priors (i.e. various prior mean functions or covariance kernels on the GP (Rasmussen and Williams, 2006; Marrel and Chabridon, 2021) or prior degrees and q-norms for the VPCE) but that give equally good predictive means, we choose to average the p surrogates $(\widehat{g}^{(1)}, \ldots, \widehat{g}^{(p)})^\top$ by using a convex combination $(w_1, \ldots, w_p)^\top \in \boldsymbol{\Delta}^{p-1}$, similar to an expert aggregation technique (Cesa-Bianchi and Lugosi, 2006):

$$\widehat{g}^{\text{agg}}(\boldsymbol{X}) = \sum_{i=1}^{p} w_i \widehat{g}^{(i)}(\boldsymbol{X}). \tag{3.26}$$

This approach is useful for a couple of reasons. Firstly, since individual surrogates may have biases due to their specific assumptions or hyperparameters, it allows to reduce the total bias by averaging it out. Secondly, it increases the stability of the model since the combined prediction is less sensitive to the misspecification of any single surrogate, which leads to more stable and generalizable results. Finally, since there is no best surrogate in all generality, aggregation is a suitable way to incorporate all the available plausible models and therefore to minimize the potential bias. These surrogate aggregates will be further used in chapter 5 as part of the Bayesian model updating methodology (see section 5.2.3) in order to speed-up the computation of a computer model for Markov-chain Monte-Carlo (MCMC) sampling of data-informed posterior distributions. In that case, this aggregation technique allows to integrate part of the metamodeling bias in the posterior distributions. Another idea is to use the output of such surrogates for given-data sensitivity analysis, typically with an HSIC method (as will be detailed in section 3.3.2 below).

In summary, the metamodeling approaches presented, namely Gaussian process regression, vector-valued polynomial chaos expansions, and Karhunèn-Loève field decomposition, demonstrate that the THYC-Puffer-DEPO code can be efficiently and robustly approximated using a variety of surrogate models. Each technique offers complementary strengths in terms of predictive accuracy, interpretability, and computational efficiency, and the potential aggregation strategy further en-

hances model stability and reduces bias. These surrogates enable rapid uncertainty propagation for forward problems and can facilitate advanced sensitivity analyses, laying the groundwork for a thorough investigation of input-output relationships. Building on these metamodels, the next section leverages their analytic and computational advantages to perform detailed global sensitivity analysis, quantifying the influence of input parameters on clogging kinetics.

3.3 Sensitivity analysis

Sensitivity analysis is a fundamental tool (Da Veiga et al., 2021) in computational modeling and uncertainty quantification (Sullivan, 2015), aimed at understanding how variations in model inputs influence the uncertainty in the outputs. By systematically quantifying the impact of each input parameter, sensitivity analysis helps identify and rank the most influential variables. It can further guide model developers for simplification or enhancement, and supports robust decision-making under uncertainty. There are several approaches to sensitivity analysis, ranging from local methods such as derivative-based methods (Cacuci, 2003), which assess the effect of small perturbations around nominal values, to global methods, which consider the entire input space and capture nonlinearities and interactions. Popular global techniques (Da Veiga et al., 2021) include variance-based methods such as Sobol' indices (Sobol', 1993), dependency measures like the Hilbert-Schmidt Independence Criterion (HSIC) (Da Veiga, 2021), and screening methods such as Morris (Morris, 1991). Sensitivity analysis is widely used in engineering, environmental modeling, finance, and risk assessment, providing insights for model validation, calibration, and prioritization of data collection efforts (Saltelli et al., 2008).

3.3.1 Sobol' indices

Assuming that, for every $k \in \{1, \dots, N\}$, the output component $g^k \in L^2_{p_{\boldsymbol{X}}}(\mathcal{X}, \mathbb{R})$, one has access to the unique "functional analysis of variance" (FANOVA) decomposition (Hoeffding, 1948; Sobol', 1993) such that:

$$g^k(\boldsymbol{X}) = \sum_{\boldsymbol{\gamma} \subseteq \{1,\dots,d\}} g^k_{\boldsymbol{\gamma}}(X_{\gamma_1}, \dots, X_{\gamma_d}) = \sum_{\boldsymbol{\gamma} \subseteq \{1,\dots,d\}} g^k_{\boldsymbol{\gamma}}(\boldsymbol{X}_{\boldsymbol{\gamma}})\,. \tag{3.27}$$

For every $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)$, with $|\boldsymbol{\gamma}| \in \{1, \dots, 2^d - 1\}$, the Sobol' index of order $|\boldsymbol{\gamma}|$ at time t_k is defined by the ratio of the variance of $g^k_{\boldsymbol{\gamma}}$ over the total output variance:

$$S_{\boldsymbol{\gamma}}(t_k) = \frac{\mathrm{Var} g^k_{\boldsymbol{\gamma}}(\boldsymbol{X}_{\boldsymbol{\gamma}})}{\mathrm{Var} g^k(\boldsymbol{X})}. \tag{3.28}$$

While several sampling-based techniques are available in the literature to estimate these indices (see, e.g., (Da Veiga et al., 2021, Ch. 4) for a review), the PCE framework enables their direct computation as a simple by-product as proposed in (Sudret, 2008). By setting:

$$\mathcal{J}_{\boldsymbol{\gamma}} = \left\{ \boldsymbol{\alpha} \in \mathbb{N}^d, |\boldsymbol{\alpha}| \leq p : \alpha_\ell > 0,\ \forall \ell \in \boldsymbol{\gamma}\ \&\ \alpha_\ell = 0,\ \text{else} \right\}\,, \tag{3.29}$$

the terms in Eq. (3.29) can be regrouped and identified as the unique factors in the FANOVA decomposition such that:

$$g^k_{\boldsymbol{\gamma}} = \sum_{\boldsymbol{\alpha} \in \mathcal{J}_{\boldsymbol{\gamma}}} g^k_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{X})\,. \tag{3.30}$$

Sobol' indices are then directly assessed with simple calculations. Using the orthonormality of the basis terms, the following expression is obtained:

$$S_{\boldsymbol{\gamma}}(t_k) = \frac{\sum_{\boldsymbol{\alpha}\in\mathcal{J}_{\boldsymbol{\gamma}}}(g^k_{\boldsymbol{\alpha}})^2}{\sum_{1\leq|\boldsymbol{\alpha}|\leq p}(g^k_{\boldsymbol{\alpha}})^2}. \tag{3.31}$$

In what follows, special attention is given to the first-order index S_i and the total-order index S_i^T for $i \in \{1,\dots,d\}$, as well as the global interaction index S_*, defined by:

$$S_i(t_k) = \frac{\mathrm{Var}g^k_i(X_i)}{\mathrm{Var}g^k(\boldsymbol{X})}, \quad S_i^T(t_k) = \sum_{\boldsymbol{\gamma}|i\subset\boldsymbol{\gamma}} S_{\gamma}(t_k) \text{ and } S_*(t_k) = 1 - \sum_{i=1}^{d} S_i(t_k). \tag{3.32}$$

The first-order and total-order indices correspond respectively to the influence of a specific input X_i on the output variance and the total effect of input X_i taking into account its interactions with the other inputs.

Analysis of the Sobol' indices on THYC-Puffer-DEPO

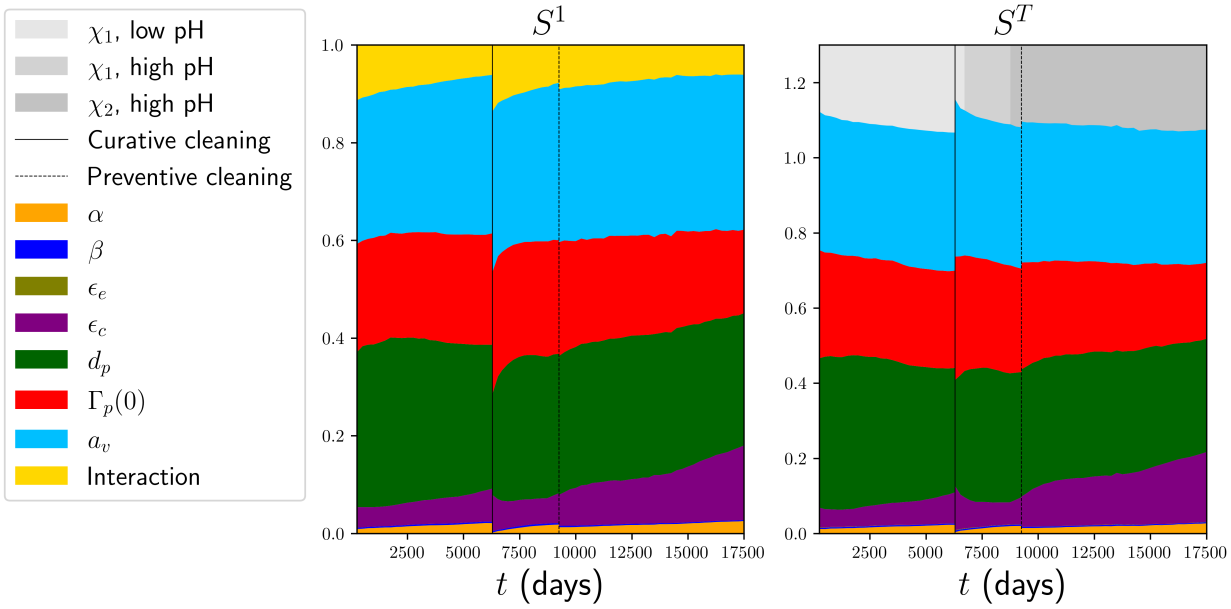

Figure 3.9: Evolution of first-order, interactions (left) and total-order (right) time-dependent Sobol' indices.

Figure 3.9 illustrates the time evolution of the Sobol' indices, as estimated from the PCE metamodel, across the three chemical regimes (the background shading indicates regime changes, a visual convention used throughout the sensitivity analysis). Several important insights emerge. First, abrupt peaks in both first-order and total-order indices align with chemical cleaning interventions; these are numerical artifacts reflecting the metamodel's sensitivity to discontinuities in the data. Second, the results corroborate earlier findings from (Lefebvre et al., 2023): the output variance is predominantly governed by a small subset of inputs, specifically d_p, a_v, and $\Gamma_p(0)$. Interaction effects remain

negligible throughout the simulation, as evidenced by the close agreement between first-order and total-order indices in Figure 3.9 (right), confirming that the TPD model exhibits nearly additive behavior with respect to its inputs under the adopted probabilistic framework. Notably, the updated code reveals a shift in the ranking of influential variables depending on the chemical regime: in the high-pH χ_2 regime, the variance contribution transitions from d_p and $\Gamma_p(0)$ to ϵ_c, while a_v consistently has a significant influence across all regimes.

The parameter ϵ_c, representing the porosity of clogging deposits, quantifies the fraction of void space within the iron oxide crystalline matrix and is often assumed to be constant and independent of operational conditions. However, the present results challenge this assumption. Iron oxide deposits are formed from charged particles whose aggregation and crystalline structure are sensitive to the solution's hydrogen ion concentration (pH), as observed in colloidal science (Liu et al., 2018). Under high-pH conditions, enhanced particle aggregation leads to denser, less porous deposits, thereby amplifying the impact of uncertainty in ϵ_c on the system's response. This increased aggregation (or stiffening) of deposits in basic environments may also account for the observed reduction in clogging rates. Overall, these findings underscore the intricate interplay between chemical conditioning, deposit microstructure, and clogging dynamics in the system.

Additionally, we have estimated the Sobol' indices using the Saltelli sampling-based approach as a benchmark to validate the metamodel-based results, more details are found in the appendix (see Figure 8.7). Both approaches turn out to be consistent, which ensures the robustness of the interpretations. This will be further confirmed with the help of HSIC indices in the following paragraph.

3.3.2 Hilbert-Schmidt independence criterion (HSIC)

Among various global sensitivity analysis measures, the Hilbert-Schmidt Independence Criterion (HSIC) is recognized as a powerful dependency measure for sensitivity analysis (Gretton et al., 2005; Da Veiga, 2015; Da Veiga et al., 2021). HSIC uncovers dependency structures between each input variable $X_j \in \boldsymbol{X} \in \mathbb{R}^d$ and the output of the physical code $Z_\ell = \mathrm{pr}_\ell \circ g(\boldsymbol{X}) \in \mathcal{Z}_\ell \subset \mathbb{R}$ at a certain time instance $t_\ell \in \mathcal{J}$. The fundamental idea behind HSIC is to compare the joint probability distribution $\mathbb{P}_{X_j,Z_\ell}$ of the couple (X_j, Z_ℓ) with the product of their marginal distributions $\mathbb{P}_{X_j}\mathbb{P}_{Z_\ell}$. This comparison is performed using generalized covariance operators in reproducing kernel Hilbert spaces (RKHSs) (Berlinet and Thomas-Agnan, 2004). A key property of HSIC is that a higher index value for a given variable indicates a stronger dependency with the output, this property in question is especially useful for variable selection (Da Veiga, 2015). Let $i \in \{1, \ldots, d\}$ be an input parameter index and $k \in \mathcal{J}$ be a data time instance. Consider two RKHSs, $\mathcal{F}_i$ and $\mathcal{G}_\ell$, with respective characteristic kernels κ_i and κ_ℓ (Sriperumbudur et al., 2011). On the product space $\mathcal{F}_i \times \mathcal{G}_\ell$, let's consider the product kernel $\nu_{j\ell} = \kappa_j \otimes \kappa_k$, for all $(X_j, X_j') \in \mathcal{X}_i^2, (Z_\ell, Z_\ell') \in \mathcal{Z}_k^2$ such as:

$$\nu_{jk}\left((X_j, Z_\ell), (X_j', Z_\ell')\right) = \kappa_i(X_j, X_j')\kappa_k(Z_\ell, Z_\ell'). \tag{3.33}$$

Then, the generalized covariance operator is defined as:

$$C_{X_j Z_\ell} = \mathbb{E}_{(X_j,Z_\ell)}\left[\nu_{jk}\left((X_j, Z_k), .\right)\right] - \mathbb{E}_{X_j}\mathbb{E}_{Z_k}\left[\nu_{jk}\left((X_j, Z_\ell), .\right)\right], \tag{3.34}$$

and the HSIC measure quantifying dependence is given by the Hilbert-Schmidt norm of the covariance operator:

$$\mathrm{HSIC}(X_j, Z_\ell) = \|C_{X_j Z_\ell}\|^2_{\mathrm{HS}} = \mathrm{Tr}(C^\top_{X_j Z_\ell} C_{X_j Z_\ell}). \tag{3.35}$$

A fundamental result, proven in (Gretton et al., 2005; Da Veiga, 2015), states that, for two characteristic kernels:

$$\mathrm{HSIC}(X_j, Z_\ell) = 0 \iff X_j \text{ and } Z_\ell \text{ are independent.} \tag{3.36}$$

As a counterpart, a nonzero HSIC value indicates a degree of dependence between the input X_j and the output Z_ℓ, justifying its use in global sensitivity analysis. Estimating HSIC from data $\{(X_i^{(j)}, Z_\ell^{(j)})\}_{j=1}^n$ requires statistical approximations. Usually, two kinds of estimators exist, notably the U-statistics and the V-statistics (Da Veiga et al., 2021, Ch. 6). A commonly used estimator formulates as:

$$\widehat{\mathrm{HSIC}}(X_j, Z_\ell) = \frac{1}{n^2}\mathrm{Tr}\left(\mathbf{L}_j \mathbf{H} \mathbf{L}_k \mathbf{H}\right), \tag{3.37}$$

where $\mathbf{L}_j, \mathbf{L}_k$ are Gram matrices with entries $(\mathbf{L}_j)_{pq} = \kappa_i(X_j^{(p)}, X_j^{(q)})$ and $(\mathbf{L}_k)_{pq} = \kappa_k(Y_k^{(p)}, Y_k^{(q)})$ for $p, q \in \{1, \ldots, n\}$. The matrix $\mathbf{H}$ is a centering matrix $(\mathbf{H})_{pq} = \delta_{p,q} - \frac{1}{n}$ (here $\delta_{p,q}$ is the Kronecker delta). In practical usage for real-valued random variables, we use the Gaussian kernel parameterized by the empirical standard-deviation $\widehat{\sigma}_{X_j}$ of the input sample:

$$\kappa_j(X_j^{(p)}, X_j^{(q)}) = \exp\left(|X_j^{(p)} - X_j^{(q)}|^2 / \widehat{\sigma}_{X_j}\right). \tag{3.38}$$

In order to better interpret the ranking results, a normalized version of the index (Marrel and Chabridon, 2021) is often considered:

$$R^2_{\mathrm{HSIC}}(X_j, Z_\ell) := \frac{\widehat{\mathrm{HSIC}}(X_j, Z_\ell)}{\sqrt{\widehat{\mathrm{HSIC}}(X_j, X_j)\widehat{\mathrm{HSIC}}(Z_\ell, Z_\ell)}} \in [0, 1]. \tag{3.39}$$

Analysis of the HSIC indices on TPD

Fig. 3.10 shows the corresponding R^2_{HSIC} indices and the associated p-values of the statistical tests, carried out on the TPD Monte Carlo sample presented in section 3.1. The sensitivities to the discontinuities are not present anymore. This is due to the given data estimation without relying on any approximation of the function. The main conclusions from the Sobol' analysis here persist, namely that the influential variables are a_v, d_p and $\Gamma_p(0)$ in all chemical conditionings and ϵ_c in the χ_2-high pH conditioning. Moreover the ranking of the influential variables for every fixed time is also the same as in the previous analysis. Here the use of the entire dataset allows to compute the p-value with an asymptotic estimator. All the influential variables are well under the threshold value. It can be concluded that with this complementary computation, the same results hold and therefore the robustness and relevance of the sensitivity analysis is achieved.

3.3.3 Target HSIC indices

With the help of HSIC indices, one can also easily access the target indices. In other words, this is about uncovering which indices contribute the most to the outputs belonging to a given critical region $\mathcal{C}$ (Marrel and Chabridon, 2021) corresponding potentially to a certain safety threshold. It

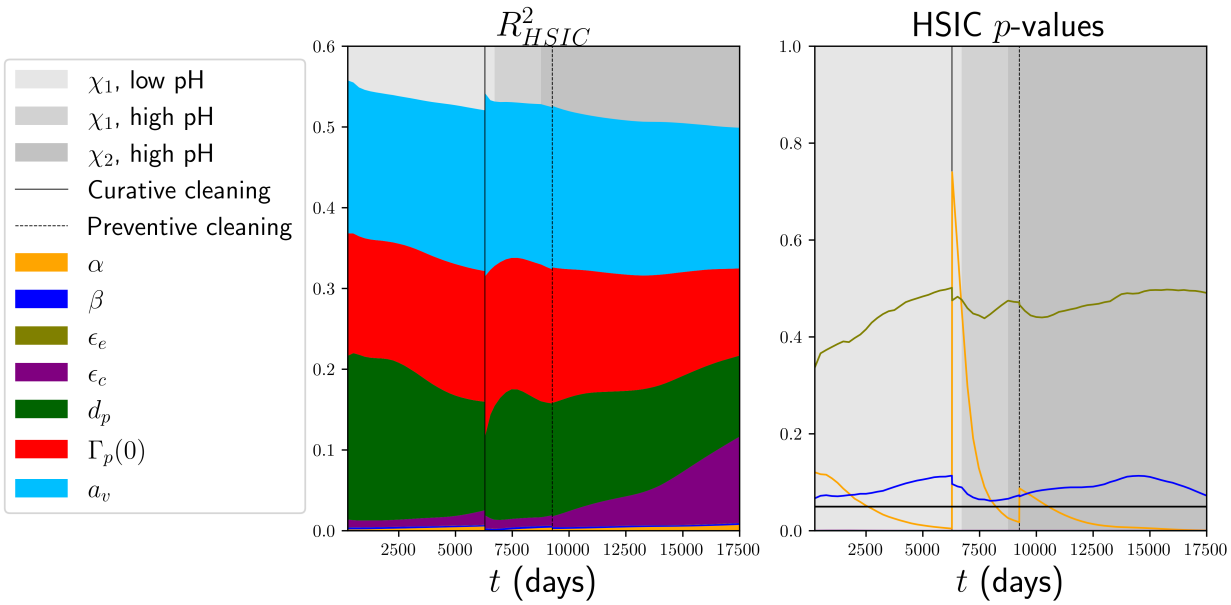

Figure 3.10: Time variation of the estimated HSIC indices and p-values associated to the statistical test of independence. The p-value horizontal line is set at 0.05 and the colors of the lines correspond to the different variables.

makes use of a filter function $f_{\mathcal{C}} : \mathcal{Y} \to \mathbb{R}_+$ that *selects* the simulations falling in $\mathcal{C}$ by transforming the output distribution. Usually a hard-thresholding with an indicator function $\mathbf{1}_{\mathcal{C}}$ is possible but this can forget outputs that are "close" to the region of interest and also leads to a notable decrease of the available data for estimation. This can be overcome using a modified covariance kernel as explained in (Marrel and Chabridon, 2021). The filter function used is defined by:

$$\widetilde{Y}_j = f_{\mathcal{C}}(Y_j) = \exp\left(-\frac{\inf_{y\in\mathcal{C}}\|Y_j - y\|}{s\widehat{\sigma}_{Y_j}}\right) \in [0,1]\,, \tag{3.40}$$

where $s = 1/5$ and $\widehat{\sigma}_{Y_j}$ is the estimated standard deviation of the output sample at t_j. The hyperparameter s may seem arbitrary but is the one retained for this type of analysis (Marrel and Chabridon, 2021). Its optimization can be the topic for future research, but remained outside of the scope of the current results. The V-statistic estimator in Eq. (3.37) can still be used with the modified variable to yield the target HSIC indices:

$$\widehat{\text{T-HSIC}}(X_i, Y_j) := \widehat{\text{HSIC}}(X_i, \widetilde{Y}_j), \tag{3.41}$$

and the corresponding p-value associated to the independence hypothesis test can still be used.

Analysis of the T-HSIC indices on TPD

An illustrative threshold of 70% is chosen as a target region for the clogging simulation dataset. This is not necessarily the prognostics threshold that is used in practice, indeed this highly-clogged regime is critical for the operation of the SG and is never reached in operation time. Nonetheless choosing such a threshold can test the limit cases of the physical clogging model presented in chapter 2. The

associated variation of T-HSIC indices amount is displayed in Fig. 3.11. For a certain amount of time, the T-HSIC index is negligible because there is no relevant data on the output set. Indeed all the outputs are weighted to zero because they are still far from the target region. Once the target set size increases, one can see the influential variables appear but the statistical test is not yet significant. Here a permutation-based estimator is used for computing the p-value, given the scarcity of the data. Indeed, this is linked to the number of critical simulations that appear progressively. From this local analysis of the highly-clogged simulations, it becomes manifest that the clogging porosity ϵ_c becomes the most influential variable. It should be stressed that this was apparent in the previous analyses in Figs. 3.9 and 3.10 in the long-time high-pH χ_2 conditioning. Indeed it constitutes a new finding that is here quantitatively established with the help of the local target indices.

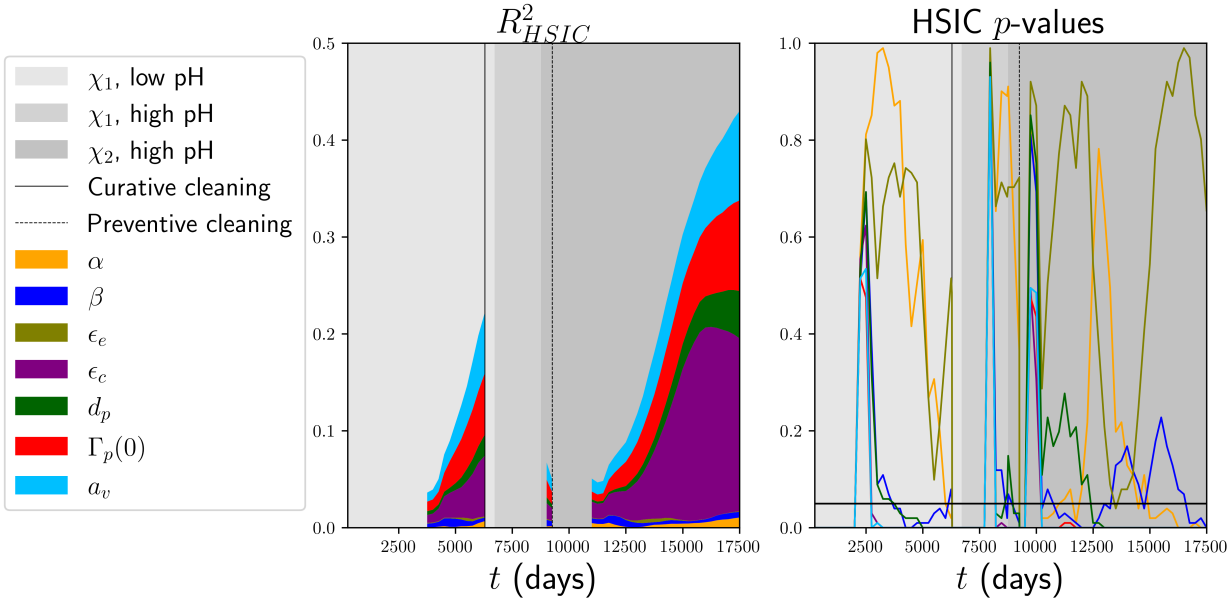

Figure 3.11: Time variation of the estimated, normalized T-HSIC indices, the p-values associated to the statistical test of independence.

3.3.4 Conditional HSIC indices

Conditional sensitivity analysis aims at identifying the influential input variables *knowing* that the output falls within a certain critical region $\mathcal{C}$. Reformulating this allows to uncover the variables driving the event $\{Y \in \mathcal{C}\}$. This can be achieved with the help of the filter function $f_{\mathcal{C}}$, by a standard procedure (Marrel and Chabridon, 2021). The conditioning formula gives:

$$Y_j|\{Y_j \in \mathcal{C}\} \sim p^f_{Y_j}, \text{ s.t for all measurable } A \subset \mathcal{Y},\ p^f_{Y_j}(A) = \int_A \frac{f_{\mathcal{C}}(y)}{\mathbb{E}\left[f_{\mathcal{C}}(Y_j)\right]} dp_{Y_j}(y). \tag{3.42}$$

By conditioning, the underlying input space is also modified; thus, for all $i \in \{1, \ldots, d\}$, $X_i|\{Y_j \in \mathcal{C}\} \sim p^f_{X_i}$ defined similarly. The joint distribution involved in the general covariance operator in Eq. (3.34) is changed accordingly $(X_i, Y_j) \sim p^f_{X_i,Y_j}$. Moreover a V-statistic estimator similar to Eq. (3.37) can be derived.

Analysis of the C-HSIC indices on TPD

The identical threshold value as in the target analysis is here maintained. The same progression in target set sizes persists due to the utilization of the identical weight function in Eq. (3.40). Notably, a significant distinction arises, as there is a noticeable absence of a null C-HSIC index throughout the entire simulation duration. This phenomenon can be attributed to the influence of conditioning on the input variables; however, the resultant effect lacks interpretability when considered in the context of sensitivity analysis. This phenomenon is evident in the uniform distribution of HSIC "shares" at the initial stages of the upper plots in Fig. 3.12. Once this conditioning effect stabilizes, the ranking of C-HSIC at the 70% threshold remains consistent, and the previous findings regarding the clogging porosity rate are reaffirmed. Following the implementation of preventive cleaning and under high-pH conditions in the χ_2 chemical conditioning, this variable emerges as the most influential, while the previously dominant variable, d_p, becomes negligible, as further corroborated by the p-value plot.

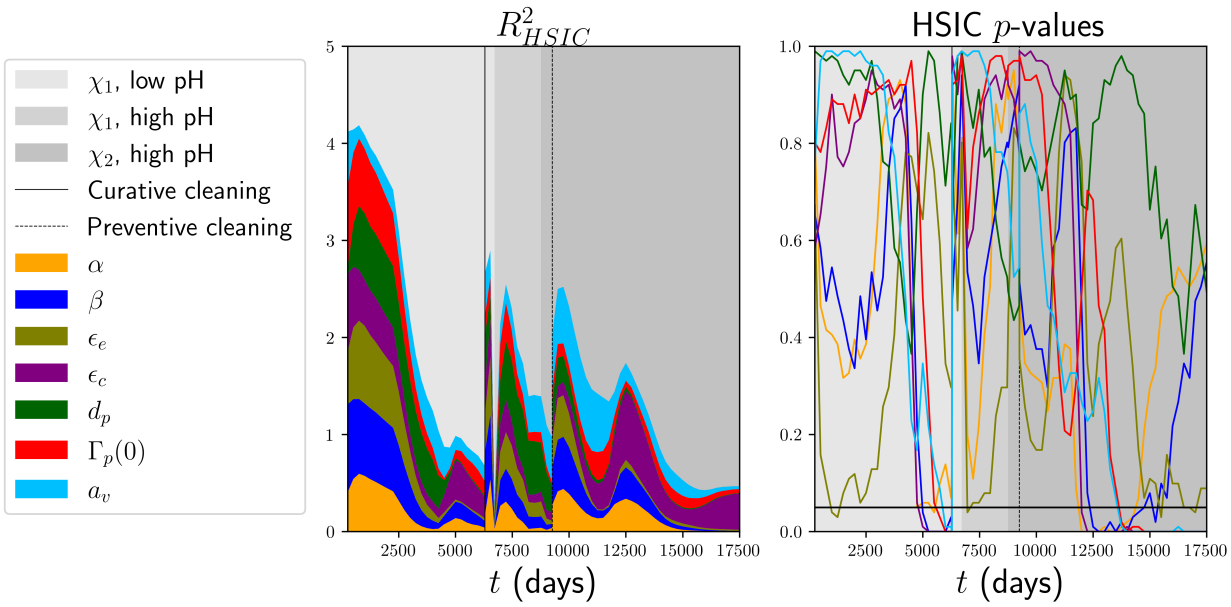

Figure 3.12: Time variation of the estimated, normalized C-HSIC indices, the p-values associated to the statistical test of independence.

3.4 Conclusion and perspectives

This chapter has presented a comprehensive uncertainty quantification and sensitivity analysis study for the THYC-Puffer-DEPO code, used as a simulation model of steam generator clogging phenomena over the lifespan of the asset. We have detailed the probabilistic characterization of the uncertain input parameters highlighted by physicists, the subsequent design of experiments, as well as the application of different surrogate modeling techniques, including Gaussian processes, vector Polynomial Chaos Expansions, and Karhunèn-Loève expansions.

Sensitivity analysis using both Sobol' indices and HSIC indices provided insights into the influence of the input variables on clogging kinetics. Across all chemical conditioning regimes, a subset of parameters, namely the iron oxide particle diameter, the calibration parameter for the vena contracta mechanism, and the initial particle mass fraction, were identified as dominant contributors to output uncertainty, corroborating with the previous findings in (Lefebvre et al., 2023). Notably, the porosity of clogging deposits emerged as a critical factor under high-pH conditions and long-term operation times, uncovering an unknown interplay between chemical environment and deposit microstructure. As mentioned before, this work has given rise to a publication as a journal article in (Jaber et al., 2025b) and the methodology applied to a different SG are found in the appendix 8.1.

At this stage, two remarks can be made. First, we have shown that the simulation code TPD can be approximated in a multitude of ways if we entrust merely the predictivity coefficient as the validation measure. This is valid for the VPCE and KLE models because their learning set has no intrinsic bias or noise, moreover the results show clearly that the approximating curves follow the right trend. However, for models such as the GP where there is more stochasticity involved, relying solely on the predictivity coefficient can be misleading, therefore one wish is to develop a more resilient validation and qualification technique by using, for instance, prediction intervals. Secondly, the methodologies developed and applied in this chapter provide a foundation to guide the planning of chemical cleanings for steam generators because the simulation code TPD can simulate the whole life-span of the asset. However, assuming this model as the most realistic, then the epistemic uncertainty observed by propagating the prior probability distributions (i.e the spread of the trajectories in 3.1) is significant. This uncertainty is, for the moment, too large in order to assist decision makers. Therefore it has to be reduced with the help of available tools and data presented in chapter 2. We will address these topics in the upcoming chapters.

👉 Perspectives of chapter 3

For the scientific and industrial perspectives of this chapter, the following items can be pursued:

- Analyse sensitivity of the results (both from metamodeling and sensitivity analysis) to the size of the DoE, to ensure robustness of the results.
- Propagate uncertainties in the other physical models involved, namely THYC and Puffer. For the latter, develop a rigorous probability modeling of the iron oxide water solubility field $\Gamma_s^{\max}$, and assess its influence on downstream outputs.
- Address model-form uncertainty and validation: design targeted lab experiments to reduce discrepancy between model and reality, develop and compare alternative clogging model variants.
- Clarify physical origins of the influence of the clogging porosity uncertainty: investigate experimentally and theoretically the dependence of clogging porosity ϵ_c on chemical conditions (the influence of pH on colloidal aggregation), and quantify its impact on clogging predictions.
- Understand the theoretical properties of $\widehat{g}^{\text{agg}}$, as well as the choice of its hyperparameters.
- Run the methodology on all the available SG models of the nuclear fleet and analyse the regularity of the results.
- Include sensitivity analysis tools and different surrogate models within the JNGV platform.

Chapter 4

Robust predictive metamodeling with conformal prediction

This chapter introduces conformal prediction (CP) methods for uncertainty quantification of Gaussian surrogate models, addressing the limitations of traditional Bayesian credibility intervals. We review the main CP approachesm)(full, split, and cross-conformal) and propose new cross-conformal estimators ($\mathcal{J}$+GP and $\mathcal{J}$-minmax-GP) tailored to GP metamodels, providing distribution-free, adaptive prediction intervals with finite-sample marginal coverage guarantees. Theoretical properties, including coverage and stability properties, are established, and practical evaluation metrics such as empirical coverage and interval adaptivity are discussed. Numerical experiments on THYC-Puffer-DEPO at a fixed time output and on the full vector outputs are addressed showcasing the robustness and informativeness of the proposed intervals. This work stems essentially from the published paper (Jaber et al., 2025a). Extensions to multidimensional outputs via conformal risk control strategies are also presented, highlighting the versatility of CP for reliable uncertainty quantification in surrogate modeling of computer simulation codes.

Contents

4.1 Overview on GP metamodel qualification

In the context of Gaussian process regression, a wide range of validation metrics have been developed to assess the predictive quality of fitted GP metamodels (Demay et al., 2022; Marrel and Iooss, 2024). Traditionally, the focus has been on evaluating the accuracy of the mean prediction, often quantified by the predictivity coefficient. However, recent research has emphasized the importance of also assessing the quality of the posterior predictive variance, as this reflects the model's uncertainty quantification capabilities. Techniques such as K-fold and leave-one-out cross-validation have been proposed to evaluate the robustness of these validation indicators (De Carvalho et al., 2022; Jaber et al., 2025b). Despite these advances, there is still no consensus on the most appropriate metrics for validating GP metamodels, and the question remains open, particularly when considering applications beyond mean prediction. A key requirement for an efficient surrogate model is its ability to adapt to local information in the training data. By conditioning on observed data, GP metamodels can capture underlying patterns and avoid overconfidence in regions with sparse or noisy data. The reliability of GP predictions is typically assessed using Bayesian credibility intervals, whose width reflects the model's confidence. These intervals are sensitive to the quality and quantity of training data, as well as to the choice of hyperparameters and covariance kernels. While careful tuning can improve the trustworthiness of GP uncertainty estimates, poor choices may result in intervals that are either too narrow (overconfident) or too wide (conservative). Recent approaches, such as that of Acharki et al. (2023), have sought to enhance the robustness of Bayesian credibility intervals by optimizing kernel hyperparameters to address model misspecification. However, these methods still fundamentally rely on the assumption of Gaussianity. To address these limitations, this chapter introduces the conformal prediction (CP) framework (Vovk et al., 2005) as a means to construct more reliable, distribution-free prediction intervals for GP metamodels. Unlike Bayesian approaches, CP does not require strong distributional assumptions or a well-specified prior for the covariance kernel. Instead, it leverages the flexibility of GP models while providing frequentist coverage guarantees for the resulting prediction intervals. This makes CP-based intervals particularly attractive for applications where the underlying model assumptions may be questionable or where robust uncertainty quantification is critical.

4.2 Conformal prediction UQ methods

Conformal prediction has become a widely used framework in machine learning for constructing prediction intervals with guaranteed coverage, regardless of the underlying predictive model (Vovk et al., 2005; Angelopoulos and Bates, 2023) and underlying data distribution. The only essential assumption is the exchangeability of the dataset, which is typically satisfied in computer experiments and standard ML settings where data points are independently and identically distributed. Given a training dataset $\mathcal{D}_n$ and a new test point $\boldsymbol{Z}^{(n+1)} = (\boldsymbol{X}^{(n+1)}, Y^{(n+1)})$, CP assumes that all $n+1$ points are *exchangeable* (Vovk et al., 2005). Formally, this means that for any permutation $\pi \in \mathfrak{S}(n+1)$:

$$(\boldsymbol{Z}^{(1)}, \ldots, \boldsymbol{Z}^{(n+1)}) \overset{\mathcal{L}}{=} (\boldsymbol{Z}^{(\pi(1))}, \ldots, \boldsymbol{Z}^{(\pi(n+1))}), \tag{4.1}$$

where $\overset{\mathcal{L}}{=}$ denotes equality in distribution. This means any point could serve as the test point, and i.i.d. data is a special case of exchangeability. CP thus provides a powerful, model-agnostic tool for uncertainty quantification in both regression and classification tasks. A central challenge in CP is

the construction of *adaptive* prediction intervals—intervals whose width varies according to the local prediction error of the metamodel. Adaptivity is closely linked to the expressivity of the surrogate model: ideally, intervals should be narrow where the model is accurate (e.g., near training points) and wider where the error is higher. For GP surrogate models such an adaptivity is a crucial property for meaningful uncertainty quantification and model validation.

For any confidence level $\alpha \in (0, 1)$, a conformal predictor with coverage $1 - \alpha$ is a measurable function (Vovk et al., 2005):

$$\begin{aligned} C_\alpha \colon \mathcal{Z}^n \times \mathcal{X} &\to 2^{\mathcal{Y}} \\ (\mathcal{D}_n, \boldsymbol{X}) &\mapsto C_\alpha(\mathcal{D}_n, \boldsymbol{X}) =: C_{n,\alpha}(\boldsymbol{X}), \end{aligned} \tag{4.2}$$

such that for a new test point $\boldsymbol{Z}^{(n+1)}$, the following *marginal* coverage holds:

$$\mathbb{P}\left(Y^{(n+1)} \in C_{n,\alpha}(\boldsymbol{X}^{(n+1)})\right) \geq 1 - \alpha, \tag{4.3}$$

where the probability is over the data $\widetilde{\mathcal{D}}_n = \mathcal{D}_n \cup \{\boldsymbol{Z}^{(n+1)}\}$. To construct such predictors, a *non-conformity score* is used, defined as a measurable function (Vovk et al., 2005):

$$\begin{aligned} R \colon \mathcal{Z}^n \times \mathcal{Z} &\to \mathbb{R} \\ (\mathcal{D}_n, \boldsymbol{Z}) &\mapsto R(\mathcal{D}_n, \boldsymbol{Z}), \end{aligned} \tag{4.4}$$

which quantifies how atypical $\boldsymbol{Z}$ is relative to $\mathcal{D}_n$. For example, if a metamodel $\widehat{g}$ is trained on $\mathcal{D}_n$, a simple non-conformity score is the residual:

$$R(\mathcal{D}_n, \boldsymbol{Z}^{(n+1)}) = |g(\boldsymbol{X}^{(n+1)}) - \widehat{g}(\boldsymbol{X}^{(n+1)})|. \tag{4.5}$$

The coverage in Eq. (4.3) is called *marginal* since it holds on average over all possible training sets and over all possible choice of $\boldsymbol{X}_{n+1}$. A stronger notion is *training-conditional* coverage (Angelopoulos and Bates, 2023), where for a conformal predictor $\widehat{C}_{n,\alpha}$:

$$\mathbb{P}\left(Y^{(n+1)} \in \widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)}) \mid \mathcal{D}_n\right) \geq 1 - \alpha. \tag{4.6}$$

Families of conformal prediction methods

There are three main approaches to constructing conformal predictors: *full-conformal* (transductive CP), *split-conformal* (inductive CP), and *cross-conformal* methods (Vovk et al., 2005; Papadopoulos et al., 2002; Vovk, 2015; Angelopoulos and Bates, 2023).

Full-conformal (Transductive CP):

This is the original conformal method, which computes prediction sets by retraining the model for each possible candidate output $y \in \mathcal{Y}$, thus making it costly. Given a training dataset $\mathcal{D}_n = \{(\boldsymbol{X}^{(i)}, Y^{(i)})\}_{i=1}^n$ and a new test input $\boldsymbol{X}^{(n+1)}$, the full-conformal (transductive) method constructs the prediction set for $Y^{(n+1)}$ as follows:

1. For each candidate value $y \in \mathcal{Y}$, form the augmented dataset $\widetilde{\mathcal{D}}_n = \mathcal{D}_n \cup \{(\boldsymbol{X}^{(n+1)}, y)\}$.

2. For each $i = 1, \ldots, n+1$, compute the non-conformity score $R(\widetilde{\mathcal{D}}_n, \boldsymbol{Z}^{(i)})$, where $\boldsymbol{Z}^{(i)}$ is the i-th point in the augmented dataset.

3. Compute the p-value for y:

$$p(y) = \frac{1}{n+1} \left| \left\{ i \in \{1, \ldots, n+1\} : R(\widetilde{\mathcal{D}}_n, \boldsymbol{Z}^{(i)}) \geq R(\widetilde{\mathcal{D}}_n, Z^{(n+1)}) \right\} \right| . \tag{4.7}$$

4. The prediction set at level $1 - \alpha$ is:

$$\widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)}) = \{ y \in \mathcal{Y} : p(y) > \alpha \} . \tag{4.8}$$

This procedure guarantees marginal coverage $\geq 1 - \alpha$ under exchangeability. The main computational cost comes from retraining the model for each candidate y, as the non-conformity scores must be recomputed for the augmented dataset for every y. However, it achieves marginal coverage $\geq 1 - \alpha$. Efficient implementations exist for specific models such as Ridge and Lasso regression (Vovk, 2001; Lei, 2019), k-nearest neighbors (Papadopoulos et al., 2008, 2011), and more recently, GP regression (Papadopoulos, 2024).

Split-conformal (Inductive CP):

Split-conformal methods (Vovk et al., 2005) partition the data $\mathcal{D}$ into a training set $\mathcal{D}_n$ of size n and a calibration set $\mathcal{D}_\ell$ of size ℓ. A metamodel $\widehat{g}$ is fitted on $\mathcal{D}_n$. For each calibration point $(\boldsymbol{X}^{(j)}, Y^{(j)}) \in \mathcal{D}_\ell$, compute the non-conformity score:

$$R_j^{\text{cal}} := |Y^{(j)} - \widehat{g}(\boldsymbol{X}^{(j)})|, \quad j = 1, \ldots \ell. \tag{4.9}$$

For a new test input $\boldsymbol{X}^{(n+1)}$, the split-conformal prediction interval at level $1 - \alpha$ is:

$$\widehat{C}_{n,\alpha}^{\text{split}}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{g}(\boldsymbol{X}^{(n+1)}) \pm \widehat{q}_{\ell,\alpha}^{\pm} \left\{ R_j^{\text{cal}} \right\} \right] , \tag{4.10}$$

where $\widehat{q}_{\ell,\alpha}^{\pm}$ denotes the $(1 - \alpha)$ empirical quantile over the calibration scores. For any finite subset $\{v_i\}_{i=1,\ldots,n}$ of an ordered set, the $(1 - \alpha)$-empirical quantile, with $\alpha \in (0, 1)$, is given by:

$$\widehat{q}_{n,\alpha}^{+} \{v_i\} := \text{the } \lceil (1 - \alpha)(n + 1) \rceil \text{-th smallest value of } v_1, \ldots, v_n , \tag{4.11}$$

with $\lceil \cdot \rceil$ denoting the ceil function. Similarly, the α-empirical quantile is given by:

$$\widehat{q}_{n,\alpha}^{-} \{v_i\} := \text{the } \lfloor \alpha(n + 1) \rfloor \text{-th smallest value of } v_1, \ldots, v_n , \tag{4.12}$$

where $\lfloor \cdot \rfloor$ denotes the floor function and such that:

$$\widehat{q}_{n,\alpha}^{-} \{v_i\} = -\widehat{q}_{n,\alpha}^{+} \{-v_i\} . \tag{4.13}$$

This method achieves marginal coverage $\geq 1 - \alpha$, but requires a dedicated calibration set, which may be suboptimal for small n.

Cross-conformal (Jackknife-based CP):

Cross-conformal methods, such as Jackknife and Jackknife+ (Barber et al., 2021), use the entire dataset for both training and calibration by fitting n leave-one-out (LOO) models. For Jackknife, the $(1-\alpha)$ empirical quantile of the LOO residuals is:

$$R_i^{\text{LOO}} := |g(\boldsymbol{X}^{(i)}) - \widehat{g}_{-i}(\boldsymbol{X}^{(i)})|. \tag{4.14}$$

The Jackknife prediction interval for a new $\boldsymbol{X}^{(n+1)}$ is:

$$\widehat{C}_{n,\alpha}^{\text{J}}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{g}(\boldsymbol{X}^{(n+1)}) \pm \widehat{q}_{n,\alpha}^{\pm}\left\{R_i^{\text{LOO}}\right\}\right]. \tag{4.15}$$

The standard Jackknife prediction interval does not have coverage guarantees without further hypothesis (Barber et al., 2021). However, the Jackknife+ modification has a marginal coverage guarantee by including LOO predictions in the quantile calculation :

$$\widehat{C}_{n,\alpha}^{\text{J+}}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{q}_{n,\alpha}^{\pm}\left\{\widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) \pm R_i^{\text{LOO}}\right\}\right], \tag{4.16}$$

and it achieves marginal coverage $\geq 1-2\alpha$. If the metamodel satisfies (δ,λ)-*out-sample stability* (for all i):

$$\mathbb{P}(|\widehat{g}(\boldsymbol{X}^{(n+1)}) - \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})| \leq \delta) \geq 1-\lambda, \tag{4.17}$$

then an δ-inflated Jackknife+ interval achieves marginal coverage $1-\alpha-2\sqrt{\lambda}$ (Barber et al., 2021):

$$\widehat{C}_{n,\alpha}^{\text{J+},\delta}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{q}_{n,\alpha}^{\pm}\left\{\widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) \pm R_i^{\text{LOO}}\right\} \pm \delta\right]. \tag{4.18}$$

More on the stability properties will be discussed in section 4.4 of this chapter. The Jackknife-minmax method is a more conservative variant, using the minimum and maximum LOO predictions for the interval bounds:

$$\begin{aligned}\widehat{C}_{n,\alpha}^{\text{J-mm}}(\boldsymbol{X}^{(n+1)}) = \Bigg[&\min_{i\in\{1,\dots,n\}}\left\{\widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})\right\} - \widehat{q}_{n,\alpha}^{-}\left\{R_i^{\text{LOO}}\right\},\\ &\max_{i\in\{1,\dots,n\}}\left\{\widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})\right\} + \widehat{q}_{n,\alpha}^{+}\left\{R_i^{\text{LOO}}\right\}\Bigg].\end{aligned} \tag{4.19}$$

and the latter achieves the $(1-\alpha)$ marginal coverage, as shown in Barber et al. (2021). Cross-conformal methods are less expensive than full-CP, but can still be costly for large datasets or expensive models. An illustration of the different cross-conformal methods is given in the figure below: A key limitation of these CP methods is the lack of theoretical adaptivity guarantees. Adaptivity (Romano et al., 2019) refers to the ability of conformal intervals to have variable widths depending on the test point, reflecting the expressivity of the surrogate model. This will be further discussed in the context of GPs.

4.3 J+GP cross-conformal estimator

We propose adapting the Jackknife+ method to GP metamodels to construct adaptive prediction intervals. Assume the GP is conditioned on a dataset $\mathcal{D}_n$, with hyperparameters $(\sigma^2_{\text{MLE}}, \theta_{\text{MLE}})$ of a

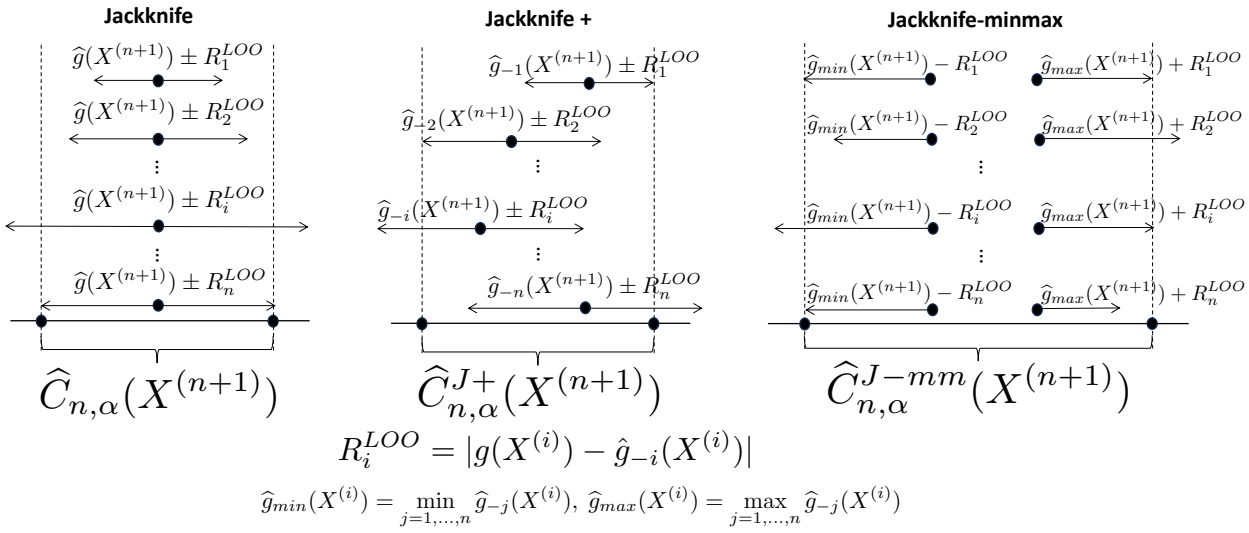

Figure 4.1: Illustration of the different cross-conformal estimators, adapted from (Barber et al., 2021).

Table 4.1: Marginal coverage and computational costs for CP methods (adapted from Barber et al. (2021)). n is the training size, $n_{\text{grid}} = |\mathcal{Y}|$ the output grid size, and m the test size.

Method	Marginal coverage	Training cost	Evaluation cost
Full	$\geq 1-\alpha$	$n \cdot n_{\text{grid}}$	$m \cdot n \cdot n_{\text{grid}}$
Split	$\geq 1-\alpha$	1	n
Jackknife+	$\geq 1-2\alpha$	n	$m \cdot n$
Jackknife-minmax	$\geq 1-\alpha$	n	$m \cdot n$

Matérn-ν kernel (for a fixed ν) estimated by maximum likelihood. This provides access to the posterior mean $\widetilde{g}$ and posterior standard deviation $\widetilde{\gamma}$. For each LOO model, denote the corresponding posterior mean and standard deviation by $\widetilde{g}_{-i}$ and $\widetilde{\gamma}_{-i}$, respectively, for $i = 1, \dots, n$. We define the LOO Gaussian non-conformity score, with a small constant $\varepsilon > 0$ to avoid division by zero in the interpolating case, as:

$$R_i^{\text{LOO}^\gamma} := \frac{|g(\boldsymbol{X}^{(i)}) - \widetilde{g}_{-i}(\boldsymbol{X}^{(i)})|}{\max\left(\varepsilon, \widetilde{\gamma}_{-i}(\boldsymbol{X}^{(i)})\right)}, \quad \forall i \in \{1, \dots, n\}. \tag{4.20}$$

For a new input $\boldsymbol{X}^{(n+1)} \in \mathcal{X}$ and a desired coverage level $1-\alpha \in (0,1)$, we define the "J+GP" conformal prediction interval, a variant of Jackknife+ tailored to GP models:

$$\widehat{C}^{\text{J+GP}}_{n,\alpha}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{q}^{\pm}_{n,\alpha}\left\{\widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)}) \pm R_i^{\text{LOO}^\gamma} \times \max\left(\varepsilon, \widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})\right)\right\}\right]. \tag{4.21}$$

This estimator provides adaptivity, as the interval width depends on $\boldsymbol{X}^{(n+1)}$. The marginal coverage guarantee matches that of Jackknife+, as stated in the following theorem.

Theorem 4.3.1. *Suppose $\mathcal{D}_n$ is exchangeable. For any $\boldsymbol{X}^{(n+1)} \in \mathcal{X}$ and coverage level $1-\alpha \in (0,1)$,*

$$\mathbb{P}\left(g(\boldsymbol{X}^{(n+1)}) \in \widehat{C}^{\text{J+GP}}_{n,\alpha}(\boldsymbol{X}^{(n+1)})\right) \geq 1-2\alpha. \tag{4.22}$$

The proof (see 8.2) follows the same reasoning as for Jackknife+ in Barber et al. (2021). Similarly, we introduce the "J-minmax-GP" estimator:

$$\begin{aligned}\widehat{C}_{n,\alpha}^{\text{J-mm-GP}}(\boldsymbol{X}^{(n+1)}) = \Big[&\min_i \left\{\widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)})\right\} - \widehat{q}_{n,\alpha}^{-}\left\{R_i^{\text{LOO}\gamma} \times \max\left(\varepsilon, \widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})\right)\right\}, \\ &\max_i \left\{\widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)})\right\} + \widehat{q}_{n,\alpha}^{+}\left\{R_i^{\text{LOO}\gamma} \times \max\left(\varepsilon, \widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})\right)\right\}\Big].\end{aligned} \tag{4.23}$$

This estimator inherits the same coverage guarantee as the standard min-max estimator, as formalized below.

Theorem 4.3.2. *Suppose $\mathcal{D}_n$ is exchangeable. For any $\boldsymbol{X}^{(n+1)} \in \mathcal{X}$ and marginal coverage level $\alpha \in (0,1)$,*

$$\mathbb{P}\left(g(\boldsymbol{X}^{(n+1)}) \in \widehat{C}_{n,\alpha}^{\text{J-mm-GP}}(\boldsymbol{X}^{(n+1)})\right) \geq 1-\alpha. \tag{4.24}$$

The proof (see 8.2) adapts the argument for Jackknife-minmax in Barber et al. (2021).

Table 4.2: Marginal coverage, training, and evaluation costs for the proposed intervals. n is the training sample size and m is the test sample size.

Method	Marginal coverage	Training cost	Evaluation cost
J+GP	$\geq 1-2\alpha$	n	$m \cdot n$
J-minmax-GP	$\geq 1-\alpha$	n	$m \cdot n$

4.4 About GP stability

In statistical learning theory, algorithmic stability refers to the sensitivity of predictions to perturbations in the training data (Bousquet, 2002). Two types of stability are usually defined: in-sample stability refers to how much the learned function changes when slightly modifying the training data, and out-of-sample stability refers to how well the algorithm generalizes to unseen data. We develop here on the latter because it helps to obtain a more accurate coverage lower bound for the Jackknife+ intervals, as shown in (Barber et al., 2021), and moreover provides training-conditional coverage guarantees (Liang and Barber, 2023). On a formal level, this amounts to finding $\lambda \in [0,1]$ for a fixed $\varepsilon > 0$ and $i \in \{1,\ldots,n\}$:

$$\mathbb{P}(|\widetilde{g}(\boldsymbol{X}^{(n+1)}) - \widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)})| \leq \varepsilon) \geq 1-\lambda. \tag{4.25}$$

The probability here is taken over the randomness of the training set $\mathcal{D}_n$, drawn from an unknown underlying distribution $\mathcal{D}_n \sim P$. Using standard GP leave-one-out formulae (Rasmussen and Williams, 2006):

$$Z := \widetilde{g}(\boldsymbol{X}^{(n+1)}) - \widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)}) \sim \mathcal{N}(0, \widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})), \tag{4.26}$$

where the LOO posterior variance is defined as:

$$\widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)}) = \widetilde{\gamma}(\boldsymbol{X}^{(n+1)})(1 - \boldsymbol{H}_{ii}), \tag{4.27}$$

and the matrix $\boldsymbol{H} \in \mathbb{R}^{n\times n}$ is the influence matrix:

$$\boldsymbol{H} = \boldsymbol{K}\boldsymbol{K}_{\epsilon}^{-1}, \tag{4.28}$$

where $\boldsymbol{K}_{\epsilon}$ is the covariance matrix with the additive nugget effect, defined in Eq. (3.7). Using standard Gaussian properties:

$$\mathbb{P}(|Z| \leq \delta) = 2\Phi\left(\frac{\delta}{\widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})}\right) - 1, \tag{4.29}$$

where Φ is the cumulative distribution function of the standard normal distribution. Thus rearranging, the following equality - and therefore lower bound - is obtained:

$$\mathbb{P}(|\widetilde{g}(\boldsymbol{X}^{(n+1)}) - \widetilde{g}_{-i}(\boldsymbol{X}^{(n+1)})| \leq \delta) = 1 - 2\left(1 - \Phi\left(\frac{\delta}{\widetilde{\gamma}_{-i}(\boldsymbol{X}^{(n+1)})}\right)\right). \tag{4.30}$$

A few conclusions can be drawn at this point. Firstly, in the absence of output noise, there is no out-of-sample stability guarantee. Indeed in this case the leave-one-out difference becomes deterministic, it is fixed once the training data is fixed because the leave-one-out variance becomes null. Therefore the prediction can change significantly when a single point is removed. Consequently, when approximating deterministic computer simulation codes with a GP (i.e performing GP interpolation), there will be no out-of-sample stability guarantees. However, whenever the noise on the outputs σ_{ϵ}^2 is non-zero, there is an improved stability and we get better generalization properties. This regularization effect can reduce overfitting if the noise is not too strong. Nonetheless, performing GP regression requires to tune the nugget effect, and if the noise is too strong, we risk losing the correlation between the error and the size of the CP-intervals, so a better interval adaptivity metric is required in this case. The different scenarios are summarized in Table 4.3 below.

Case	σ_{ϵ}^2	$\boldsymbol{H}_{ii}$	**Out-of-sample stability**	**Generalization property**
Interpolation	0	1	Low	Overfitting
Moderate noise	> 0	< 1	Good	Strong
High noise	$\gg 0$	$\to 0$	Very high	Risk of underfitting

Table 4.3: Different scenarios for GP out-of-sample stability.

4.5 Methodology evaluation

We use a two-step procedure to evaluate the performance of the proposed J+GP and J-minmax-GP estimators, comparing them to standard cross-CP methods and Bayesian credibility intervals.

Let $\widehat{C}^*_{n,\alpha}$ denote any prediction interval. On the test set $\mathcal{D}_m$, we first check empirical coverage for various target levels $1-\alpha \in [0,1]$:

$$\frac{1}{m}\sum_{i=n+1}^{n+m} \mathbf{1}\left\{g(\boldsymbol{X}^{(i)}) \in \widehat{C}^*_{n,\alpha}(\boldsymbol{X}^{(i)})\right\} \geq 1-\alpha. \tag{4.31}$$

Next, we assess the correlation between the prediction interval width and the model error. Ideally, intervals should be narrow when prediction errors are small and wider otherwise, so a significant correlation between interval width and residuals is expected. This metric quantifies the adaptivity of the intervals. Specifically, for a given coverage $1-\alpha \in (0,1)$, we compute the Spearman correlation on the test data and assess its significance via bootstrap:

$$0 \ll r_{\text{Spearman}}\left(\left\{\ell(\widehat{C}^*_{n,\alpha}(\boldsymbol{X}^{(i)})), |g(\boldsymbol{X}^{(i)}) - \widetilde{g}(\boldsymbol{X}^{(i)})|\right\}_{i\in\{n+1,\dots,n+m\}}\right), \tag{4.32}$$

where $\ell(\widehat{C}^*_{n,\alpha}(\boldsymbol{X}^{(i)}))$ is the interval length. The Spearman correlation is preferred for its robustness to outliers and ability to capture monotonic relationships. Bootstrap intervals are computed for statistical robustness. Metamodel surrogate quality is usually assessed using the predictivity coefficient on the test data (as mentioned in the previous section, see Eq. (3.12) and more details in (Iooss and Marrel, 2019)). This analysis can be complemented by empirical coverage and interval-residual correlation metrics, providing a comprehensive view of both predictive accuracy and uncertainty quantification. This framework also enables model selection, as it can be applied to different kernel choices and priors. In the following numerical experiments, we illustrate the methodology by comparing GP metamodels with various Matérn regularity parameters ν, demonstrating its ability to discriminate between models and select the most suitable one. While we focus on Matérn kernels, the approach is applicable to other kernel families as well.

To evaluate our methodology, we conduct a series of numerical experiments using both benchmark functions from the UQ and GP literature and the industrial THYC-Puffer-DEPO metamodeling use case. All examples are treated as deterministic, i.e., with no output noise, so GP interpolation is performed and theoretical out-of-sample stability is not guaranteed. The main objective is to assess the adaptivity of the J+GP and J-minmax-GP estimators by computing the Spearman correlation between the metamodel errors and the widths of the prediction intervals $\ell(\widehat{C}^*_{n,\alpha})$.

As a preliminary step, input data are standardized by subtracting the mean and dividing by the standard deviation, leveraging the known input distributions. We begin by a detailed study of a scalar-output GP metamodel for a fixed output of TPD as found in Jaber et al. (2025a). Afterwards, we present an application of the J+GP intervals on a stochastic collocation GP of TPD. An illustrative example of the method is provided in the appendix 8.2.

For each experiment, we summarize the characteristics of the computer experiments and report the performance of each GP metamodel using the predictivity coefficient (Eq. (3.12)) and the mean squared error (MSE):

$$\text{MSE} = \frac{1}{m}\sum_{i=n+1}^{n+m} |\widetilde{g}(\boldsymbol{X}^{(i)}) - g(\boldsymbol{X}^{(i)})|^2. \tag{4.33}$$

For each dataset, we compare the performance of the GP credibility intervals and the proposed J+GP and J-minmax-GP estimators. Three metrics are computed on the test dataset:

- Empirical coverage rate (Eq. (4.31));
- Empirical average width:

$$\overline{\ell}(\widehat{C}^*_{n,\alpha}) = \frac{1}{m}\sum_{i=n+1}^{n+m} \ell(\widehat{C}^*_{n,\alpha}(\boldsymbol{X}^{(i)})); \tag{4.34}$$

- Spearman correlation between interval width and error (Eq. (4.32)).

These metrics are computed for all coverage levels $1-\alpha$, and plots are shown for the Matérn kernel with the best performance. Summary tables are provided for three target coverage levels (90%, 95%, 99%) and three Matérn regularity parameters ($\nu \in \{1/2, 3/2, 5/2\}$). For each empirical coverage rate, we highlight the kernel with the smallest average width and the one with the highest Spearman correlation. Typically, the best-performing kernel differs across these metrics, so the choice depends on the desired balance between adaptivity and conservatism for the application.

Numerical results are obtained using Python, building on OpenTURNS for UQ and GP regression, and MAPIE (Cordier et al., 2023) for conformal prediction. A wrapper is implemented to make Scikit-learn (Pedregosa et al., 2011) GP constructors compatible with OpenTURNS and MAPIE. Only minor modifications are made to MAPIE to support our methodology, while retaining its standard conformal methods.

4.5.1 Fixed marginal output GP surrogate of THYC-Puffer-DEPO

Table 4.4 summarizes dataset sizes, training/test splits, and performance metrics (Q^2 and MSE) for three Matérn regularity parameters for the TPD DoE.

Table 4.4: Summary of GP metamodel performance metrics for TPD.

ν		TPD
	d	7
	N	1000
	%train	80
	%test	20
1/2	Q^2	0.990
	MSE	1.46
3/2	Q^2	0.996
	MSE	0.54
5/2	Q^2	0.997
	MSE	0.46

For the TPD use case, the GP metamodel achieves high predictive accuracy ($Q^2 \geq 0.99$) for all regularity parameters, making model selection challenging. The conformal predictors indicate that

Matérn-3/2 and Matérn-5/2 kernels offer advantages for robust metamodeling. Figure 4.2 shows empirical coverage, average interval width, and Spearman correlation for J+GP, J-minmax-GP, and GP credibility intervals as a function of target coverage.

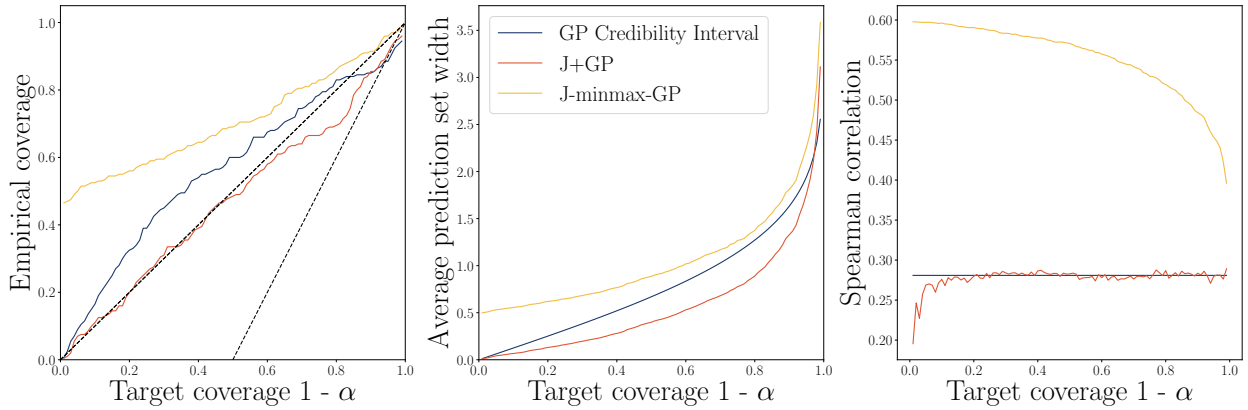

Figure 4.2: GP metamodel of TPD DoE (Matérn-3/2 kernel): empirical coverage, average interval width, and Spearman correlation between error and interval length for J+GP, J-minmax-GP, and GP credibility intervals, as a function of target coverage.

For the TPD case, GP credibility intervals show poor empirical coverage above the target threshold (~ 0.8), and J+GP intervals also underperform above $1 - \alpha = 0.5$, consistent with Theorem 4.3.1 (coverage guaranteed only above $1 - 2\alpha$). These results are conditioned on the training set; averaging over all train-test permutations would be needed for a full assessment. Low empirical coverage for credibility intervals may indicate model misspecification, making them unreliable for uncertainty quantification. The J+GP intervals have smaller average widths and similar correlation to GP credibility intervals, while J-minmax-GP intervals achieve higher correlations but larger widths. The J-minmax-GP correlation profile is regular and monotonic, likely due to the larger dataset and smoother sample. Overall, the Matérn-5/2 kernel with zero mean and MLE optimization is recommended for robust metamodeling of TPD in industrial uncertainty studies.

4.5.2 Stochastic collocation GP surrogate of THYC-Puffer-DEPO

For these experiments, we constructed a stochastic collocation GP for the multidimensional output by assigning a uniform distribution on the time domain and constraining the output trajectory to the original discretization grid (see Section 3.2.1). The cross-conformal approach is especially useful when the available dataset is small. The original TPD design of experiments has $n = 10^3$ points, which makes computing J+GP intervals computationally expensive (over 2 hours runtime for the previous table), though the resulting metamodel performs well. To make the approach more practical, we used a reduced dataset $\mathcal{D}_n = \{((t_i, \boldsymbol{X}^{(i)}), g(t_i, \boldsymbol{X}^{(i)}))\}_{1 \leq i \leq n}$ with $n = 200$, and applied an 80-20 train-test split. The J+GP intervals could then be computed much faster.

Figure 4.3 illustrates several test trajectories from a fitted stochastic collocation GP with a linear trend and a Matérn-1/2 kernel. Both J+GP and GP credibility intervals are shown. On average, both types of intervals perform similarly, with the Bayesian intervals being slightly more conservative. The empirical coverage at different time indices is generally respected. Overall, the J+GP

Table 4.5: THYC-Puffer-DEPO fixed marginal. Empirical coverage rate, average width, and Spearman correlation for different predictive intervals (standard Bayesian credibility, cross-conformal, and the proposed estimator) using various Matérn kernels and three confidence levels. In purple and underlined: the empirical coverage closest to the target coverage in absolute value. In red and bolded: lowest widths and highest Spearman correlations obtained under the target coverage condition.

Method	Matérn	Coverage			Average Width			Spearman corr.		
		90%	95%	99%	90%	95%	99%	90%	95%	99%
GP Credibility Intervals	1/2	0.960	0.975	0.975	4.717	5.621	7.387	0.463	0.463	0.463
	3/2	0.915	0.940	0.950	2.000	2.384	3.133	0.353	0.353	0.353
	5/2	0.850	0.885	0.945	1.632	1.944	2.555	0.281	0.281	0.281
J+	1/2	0.855	0.900	0.975	2.438	3.610	7.391	0.266	-0.223	0.132
	3/2	0.840	0.905	0.975	1.529	2.031	3.943	-0.355	0.043	0.202
	5/2	0.840	0.920	0.965	1.353	1.836	3.109	-0.052	0.301	0.273
J-minmax	1/2	0.860	0.920	0.975	2.763	3.943	7.711	0.666	0.666	0.666
	3/2	0.890	0.920	0.980	1.857	2.350	4.260	0.653	0.653	0.653
	5/2	0.905	0.950	0.980	**1.763**	**2.233**	3.505	**0.606**	**0.606**	0.606
J+GP	1/2	0.845	0.895	0.975	2.314	3.198	6.367	0.469	0.466	0.458
	3/2	0.840	0.925	0.955	1.523	2.058	3.215	0.351	0.345	0.352
	5/2	0.845	0.905	0.970	1.509	2.072	3.689	0.279	0.280	0.288
J-minmax-GP	1/2	0.870	0.920	0.975	2.638	3.523	6.700	0.617	0.592	0.546
	3/2	0.900	0.950	0.985	1.852	2.387	3.543	0.519	0.489	0.449
	5/2	0.895	0.960	0.995	1.918	2.477	**4.080**	0.424	0.390	**0.349**

intervals provide uncertainty quantification comparable to GP intervals, but are less conservative and do not require additional distributional assumptions. We also compared the distribution of interval widths for different prior kernels at a fixed coverage rate $1 - \alpha = 0.9$. As shown in Fig. 4.4, the Absolute exponential and Matérn-1/2 kernels yield the most concentrated interval widths, with a mean around 21% clogging rate. In contrast, the Squared exponential and Matérn-3/2, 5/2 kernels produce intervals with a smaller mean width but greater variability. The empirical coverage distributions across all kernels are similar, though some regions are undercovered and fall below the $1 - 2\alpha$ lower bound. This may be due to the marginal nature of the coverage guarantee, which holds over all permutations of train and test splits, or to the slightly different empirical coverage metric used here:

$$\frac{1}{m}\frac{1}{N}\sum_{k=1}^{N}\sum_{j=n+1}^{n+m}\mathbf{1}\{g(t_k, \boldsymbol{X}^{(j)}) \in \widehat{C}_{n,\alpha}^{\text{J+GP}}(t_k, \boldsymbol{X}^{(j)})\}. \tag{4.35}$$

This differs from Eq. (4.31) in the way the test set is used for estimation. Even though exchangeability in Eq. (4.1) holds for each sample point, the test dataset here has additional structure. In summary, the surrogate model designer has to decide here which GP model to opt for, since two distinct behaviors can be seen: one which favors higher variability (narrower, more adaptive intervals but potentially less accurate uncertainty quantification), and one which favors more conservative uncertainty quantification (wider intervals with higher coverage but less adaptivity). The choice depends on the application's requirements for adaptivity versus conservatism in uncertainty quantification.

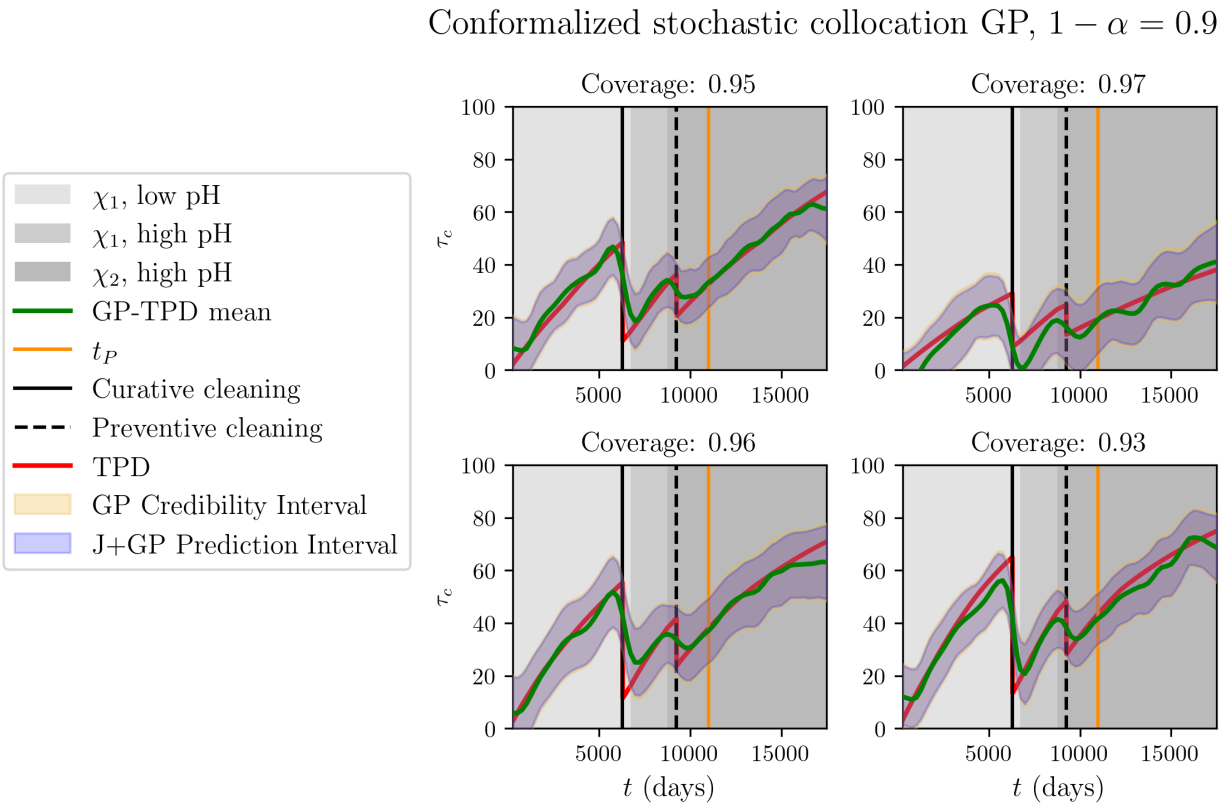

Figure 4.3: Stochastic GP metamodel of TPD dataset with a linear trend and a Matérn-1/2 kernel.

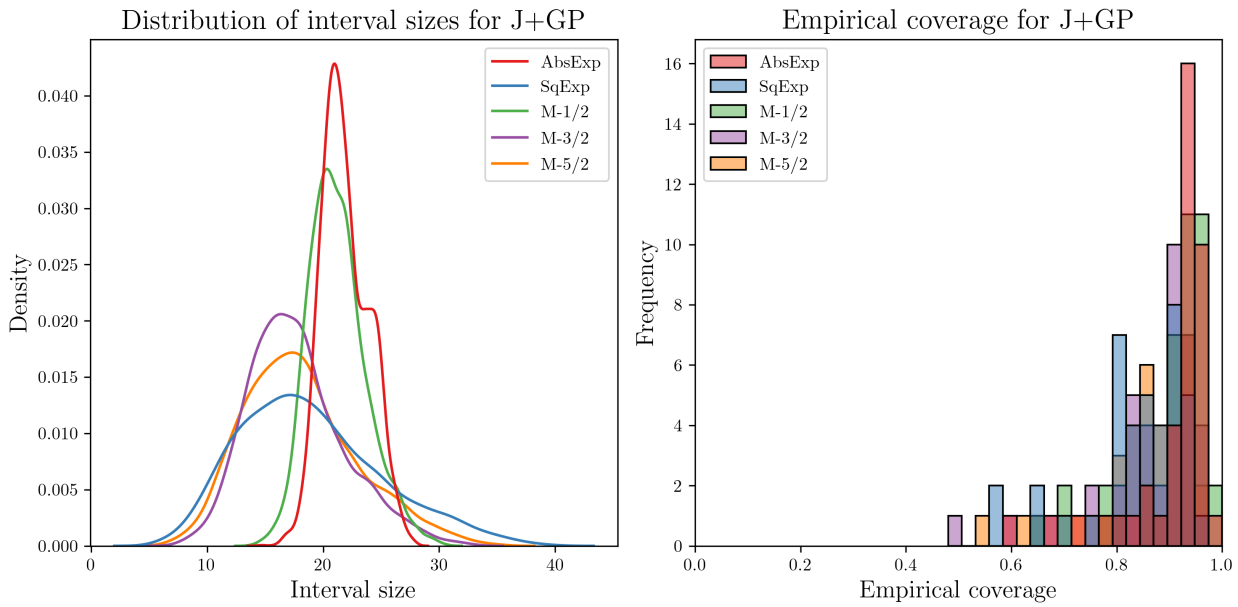

Figure 4.4: Distribution of interval sizes of the J+GP conformal predictor and empirical coverage distributions for a stochastic collocation GP with a linear trend and different kernels.

We conclude that the J+GP provides a good proxy for diagnosing the predictive performance of our stochastic collocation GP approach, and could be used as a model selection tool or for potentially for other active learning tasks.

4.6 Conformal risk control methods

In this chapter, we have so far focused on scalar outputs to construct conformal predictors, using approaches such as the stochastic collocation GP to handle multidimensional outputs indirectly. However, many simulation models (including TPD) produce inherently multidimensional outputs ($\mathcal{Y} \subseteq \mathbb{R}^N$), and it is preferable to construct conformal sets that account for this structure directly. A straightforward approach is to build conformal prediction intervals for each output dimension independently, but this quickly becomes inefficient as dimensionality increases.

4.6.1 Presentation of the methodology

A flexible way to handle multidimensional outputs is conformal risk control (CRC) (Angelopoulos et al., 2025; Blot et al., 2025), which directly enforces a user-specified expected loss for set-valued predictors. Rather than requiring the entire vector to lie inside a set, CRC controls a loss functional:

$$\mathcal{L} : \mathbb{R}^N \times 2^{\mathbb{R}^N} \to [0, 1], \tag{4.36}$$

and constructs sets $\Gamma(\boldsymbol{X})$ satisfying the following constraint on the expectation:

$$\mathbb{E}\big[\mathcal{L}(\boldsymbol{Y}, \Gamma(\boldsymbol{X}))\big] \leq \alpha, \tag{4.37}$$

for a chosen tolerance $\alpha \in [0,1]$. Under exchangeability the CRC procedure yields finite-sample, distribution-free control of this risk (marginally over data permutations). We employ CRC in the split-conformal setting. Partition the available data as $\mathcal{D} = \mathcal{D}_n \cup \mathcal{D}_\ell \cup \mathcal{D}_m$ with sizes n (training), ℓ (calibration) and m (test). Fit a surrogate on $\mathcal{D}_n$ and specify a parametric family of candidate bands $\{\Gamma_\lambda : \lambda \in \Lambda\} \subset 2^{\mathbb{R}^N}$. For each λ compute the empirical calibration risk on $\mathcal{D}_\ell$,

$$R_\lambda = \frac{1}{\ell} \sum_{j=1}^{\ell} \mathcal{L}\big(\boldsymbol{Y}^{(j)}, \Gamma_\lambda(\boldsymbol{X}^{(j)})\big), \tag{4.38}$$

and apply the standard conformal correction

$$R_\lambda^* = \frac{\ell}{\ell+1} R_\lambda + \frac{1}{\ell+1}. \tag{4.39}$$

Then, select the smallest $\lambda^* = \min\{\lambda \in \Lambda : R_\lambda^* \leq \alpha\}$. The band Γ_{λ^*}, satisfies the target risk guarantee marginally over the data under exchangeability.
For trajectory outputs (discretized at times $t_1, \ldots, t_N$) a simple and interpretable loss is the average coordinate miscoverage

$$\mathcal{L}(\boldsymbol{y}, \Gamma_\lambda) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}\{y_i \notin \mathrm{pr}_i(\Gamma_\lambda)\}. \tag{4.40}$$

A natural idea family is to scale λ by the pointwise standard deviation of the model (if available):

$$\Gamma_\lambda(\boldsymbol{X}) = \prod_{i=1}^{N} \big[\widehat{g}(t_i, \boldsymbol{X}) \pm \lambda\, \widehat{\sigma}(t_i, \boldsymbol{X})\big], \tag{4.41}$$

which yields bands whose width adapts locally through $\widehat{\sigma}(t_i, \boldsymbol{X})$. In practice Λ is taken as a monotone grid (so R_λ is nonincreasing in λ) and λ^* can be found by line search.

4.6.2 Application to KLE-GP surrogates

For a KLE-GP surrogate with p retained modes,

$$\widehat{g}(t, \boldsymbol{X}) = \sum_{k=1}^{p} \widetilde{\xi}_k(\boldsymbol{X}) \Phi_k(t), \qquad \widetilde{\gamma}(t, \boldsymbol{X}) = \sum_{k=1}^{p} \widetilde{\gamma}_k(\boldsymbol{X}) \Phi_k^2(t), \tag{4.42}$$

the CRC workflow above is applied with the loss in Eq. (4.41) to obtain Γ_{λ^*}. Calibration uses the KLE-reconstructed trajectories on $\mathcal{D}_\ell$ to estimate R_λ, and the corrected risk R_λ^* yields finite-sample control of the average fraction of out-of-band coordinates at level α.

These adaptive prediction bands for a multidimensional KLE-GP surrogate provide distribution-free uncertainty quantification. The relative size of these intervals is not changing since the value of λ is set for all the realizations of $\boldsymbol{X}$. The adaptiveness is given by the posterior standard deviation of the resulting GP. This still allows for efficient surrogate qualification, as we will see in the upcoming paragraph. Further enhancements of this method are to find two different values of λ for them not to be centered on the mean, or otherwise more recent work on having adaptive $\lambda(\boldsymbol{X})$ (see (Blot et al., 2025)). Nonetheless, the resulting intervals from our method are useful in a first approach for surrogate qualification and uncertainty analysis as we will see in the upcoming paragraph.

4.6.3 Application to KLE-GP surrogate of THYC-Puffer-DEPO

As an application, we construct a KLE-GP surrogates (see 3.2.3) for the TPD computer experiment. The dataset is split into training, calibration, and test sets of respective sizes 50, 30 and 20%, and the first $p = 2$ KLE modes are retained to capture the main variability of the output trajectories. After fitting the GP to the KLE coefficients, we compute the adaptive prediction bands Γ_{λ^*} calibrated to a target risk level α using the calibration set as explained above. Figure 4.5 illustrates several test trajectories along with the corresponding conformal prediction bands using a linear trend and a Matérn-1/2 kernel to learn the modes, demonstrating that the bands adapt to local uncertainty and provide valid coverage across the output domain. The distribution of interval widths and empirical coverage for different kernels are shown in Figure 4.6, highlighting the trade-off between adaptivity and conservatism. Indeed the use of a Squared-Exponential kernel, and similarly the Matérn-5/2 kernel to learn the modes produces small interval sizes but has a poor coverage. On the contrary the absolute exponential kernel and the Matérn-1/2 have a similar diagnostic, they are more adaptive and more conservative, thus have a better empirical coverage rate. Therefore the use of such kernels would be more suitable in this application. Our approach enables robust, distribution-free uncertainty quantification for high-dimensional outputs and can be leveraged for active learning by targeting regions with the largest prediction bands.

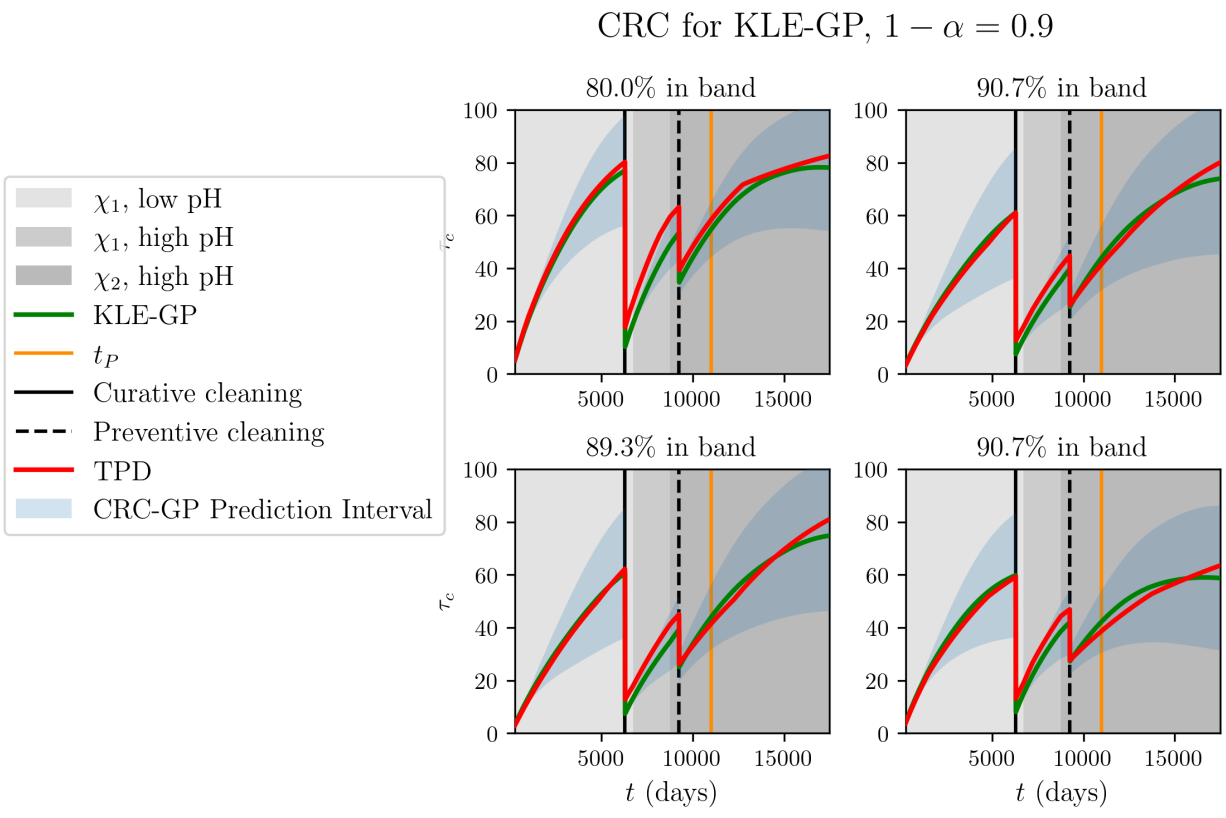

Figure 4.5: KLE-GP surrogate of TPD dataset with a linear trend and a Matérn-1/2 kernel.

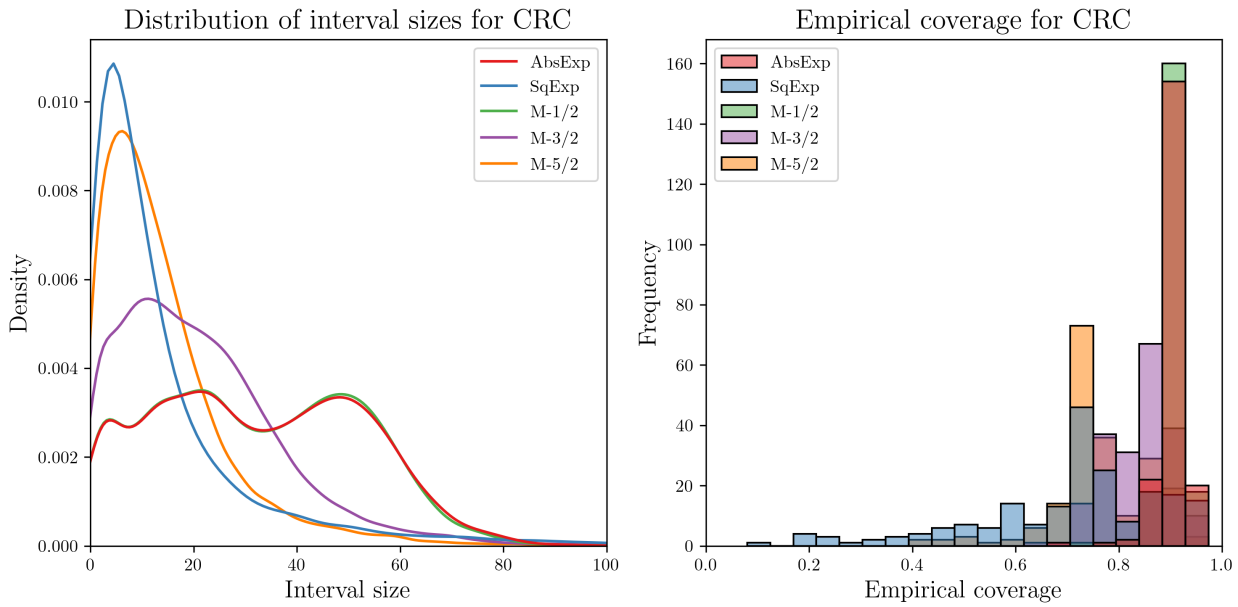

Figure 4.6: Distribution of interval sizes of the J+GP conformal predictor and empirical coverage distributions for a KLE-GP with a linear trend and different kernels.

4.7 Conclusion and perspectives

In summary, this chapter introduced new cross-conformal prediction methods, namely J+GP and J-minmax-GP, tailored for robust uncertainty quantification in Gaussian Process surrogate modeling. These estimators provide adaptive, distribution-free prediction intervals with finite-sample coverage guarantees, outperforming traditional Bayesian credibility intervals in terms of interval adaptivity and reliability, especially when model assumptions are questionable. The methodology enables practitioners to diagnose and select GP surrogate models more effectively, supporting informed kernel and hyperparameter choices.

The proposed new cross-conformal J+GP estimator yields smaller widths on average while keeping the correlation of widths with the metamodel error close to that of the Bayesian credibility intervals without need for further hypothesis, and the J-minmax-GP achieves better correlations than the standard Bayesian credibility intervals at the cost of larger intervals. We further show that inspection of our hypothesis-free CP intervals can help in choosing a more robust prior kernel for the GP metamodel of a computer code. After robust qualification and diagnostic of the predictive performance of surrogate models, the purpose now is to use them for some downstream task. The metamodel can be used to facilitate the evaluation of the pushforward measure and conformal prediction could also enables a prediction interval on the RUL estimation. Nonetheless, we could access the full distribution through Bayesian methods as we will detail in the following chapter. We have also proposed an adaptation of the conformal risk control paradigm for multi-output computer codes approximated by KLE with a GP. This method works well for computer DoEs that have at least 500 points, whereas the jackknife adaptation is better suited for computer experiments with a more reduced number of points.

👉 Perspectives of chapter 4

For the scientific perspectives of this chapter, the following items are of interest:

- Develop adaptive cross-conformal strategies for deterministic surrogates (e.g., PCE) that lack built-in stochasticity.
- Design and validate alternative adaptivity measures for prediction intervals beyond the interval-error Spearman correlation.
- Extend the conformal risk control adaptation to other reduced-basis surrogates, including intrusive reduced-order models and similar techniques.
- Analysing the sensitivity of the results to the different number of points used in each dataset.
- Perform conformal prediction on aggregate surrogate moels (see section 3.3.4).

Chapter 5

Offline Bayesian hybrid strategies

This chapter introduces a robust Bayesian framework for hybrid clogging prognostics, emphasizing uncertainty reduction in RUL prediction through the fusion of heterogeneous data sources. Rather than aiming for perfect calibration of the THYC-Puffer-DEPO physical model given the lack of experimental data, the methodology leverages Bayesian Model Updating and modular offline Data Assimilation to meaningfully reduce predictive uncertainty of the simulation outputs. Key contributions include the integration of heterogeneous groups of degradation data, iterative HSIC-based sensitivity analysis for prior updating, and the use of surrogate models and ensemble smoothing techniques for MCMC sampling and subsequent state updating. The approach is validated on the clogging simulation use-case and allows to inform specific priors and reducing the uncerainty of the predictive RUL. Full mathematical details and reproducible results are provided, building on the methodology presented in (Faber et al., 2026).

Contents

5.1 Bayesian techniques for reducing predictive uncertainty

We suppose a data assimilation problem for degradation prognostics under uncertainty where real-time degradation data is not available, only scarce and partial measurements, this is usually the case for many industrial applications with component degradation in complex systems. This makes standard filtering approaches (Jouin et al., 2016) impractical. Let's assume the computer simulation model $g : \boldsymbol{\mathcal{X}} \subset \mathbb{R}^d \to \mathbb{R}^N$, with uncertain inputs $\boldsymbol{X} = (X_1, \ldots, X_d) \sim p_{\boldsymbol{X}}$, produces a full trajectory for each input sample:

$$g(\boldsymbol{x}_0) = (g(t_1, \boldsymbol{x}_0), \ldots, g(t_N, \boldsymbol{x}_0)) \in \mathbb{R}^N, \tag{5.1}$$

where $\mathrm{pr}_\ell \circ g(\boldsymbol{X})$ or $g(t_\ell, \boldsymbol{X})$ represents the degradation index $\Delta(t_\ell)$. The model is treated as a grey-box: the underlying physics are known, but the code itself cannot be modified. We also consider q available heterogeneous degradation data groups, originating from different sensors or statistical models with varying fidelities. These are collected as $\mathcal{D} = (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q)$, with each $\boldsymbol{y}^i \in \mathbb{R}^{m_i}$ corresponding to distinct time indices $\mathcal{J}_i$, such that $\mathcal{J} = \cup_{i=1}^q \mathcal{J}_i$. For each group and time index we suppose:

$$y^i(t_\ell) = \Delta(t_\ell) + \eta_\ell^i, \tag{5.2}$$

where $\eta_\ell^i \sim \mathcal{N}(0, \sigma_i^2)$ and σ_i^2 is the known noise variance for group i. In our application $\Delta = \tau_c$, g will be THYC-Puffer-DEPO (see section 2.2.4) with its respective vector of input uncertainties $\boldsymbol{X} = \boldsymbol{X}_{\text{DEPO}}$ of dimension $d = 7$ (see section 3.1), and we will consider two groups of heterogeneous data, namely TVEs and ESTICOL (see section 2.3).

Sequential data assimilation is performed over predefined L time windows within the output time discretization. Figure 5.1 summarises the proposed methodology. For each window frame $\ell = 1, \ldots, L$, start with a design of experiments from the original computer model, apply the Bayesian Model Updating (BMU) methodology and sample from the pushforward of the updated inputs. On this ensemble, apply a smoothing technique to condition the trajectories on the heterogeneous data. This allows to integrate the available information in the computer model inputs, and to perform a diagnostics sanity check whereby confirming that data and simulation code are homogeneous. Once the prognostics window is reached, the same steps are applied, and then the RUL distribution is estimated empirically. This allows to guide decision with regards to replacement or maintenance planning of the system. If prior to acting the decision there is a new data point acquired, the steps can be reapplied to eventually correct the RUL mean and the subsequent planning.

Using data assimilation mathematical terminology, let $\boldsymbol{Z} = (\boldsymbol{X}, \boldsymbol{Y})$ be the state vector (Geir Evensen, 2022), where $\boldsymbol{Y} = g(\boldsymbol{X})$ are the model outputs over a given time window, and $\boldsymbol{X}$ are the uncertain, time-invariant parameters. The objective is to estimate the joint posterior $p(\boldsymbol{Z}|\mathcal{D})$. We do so by using a modular approach:

- Perform BMU to obtain the posterior distribution $p_{\boldsymbol{X}|\mathcal{D}}$ for the input parameters: $(\boldsymbol{X}|\mathcal{D}) \sim p_{\boldsymbol{X}|\mathcal{D}}$.
- Apply smoothing to estimate $p(\boldsymbol{Y}|\boldsymbol{X} \sim p_{\boldsymbol{X}|\mathcal{D}}, \mathcal{D})$, yielding the assimilated posterior approximation:

$$p(\boldsymbol{Z}|\mathcal{D}) \simeq p(\boldsymbol{Y}|\boldsymbol{X} \sim p_{\boldsymbol{X}|\mathcal{D}}, \mathcal{D}) p_{\boldsymbol{X}|\mathcal{D}}. \tag{5.3}$$

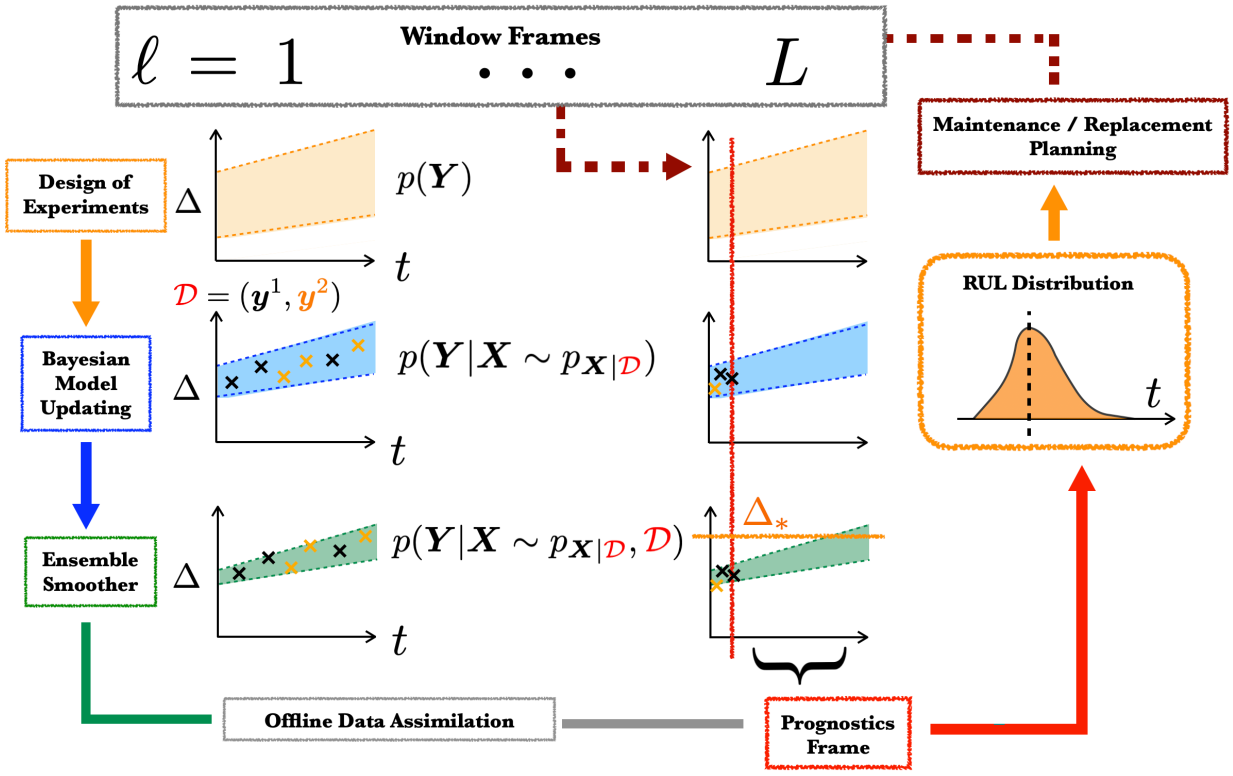

Figure 5.1: Proposed methodology for the offline data assimilation.

Once the prognostics window is reached, then the RUL distribution can be computed using the estimated density:

$$\mathbb{P}(\mathrm{RUL}(\Delta_*) \leq t_j \mid \mathcal{D}) = \int_{\mathbb{R}} \mathbf{1}\{\mathrm{pr}_{j+1}(\boldsymbol{y}) \geq \Delta_*\} p(\boldsymbol{y}|\mathcal{D}, \boldsymbol{X} \sim p_{\boldsymbol{X}|\mathcal{D}}) d\boldsymbol{y}, \tag{5.4}$$

where the integral is estimated using a Monte Carlo ensemble $\{(\boldsymbol{X}^{(i)}, g(\boldsymbol{X}^{(i)}))\}_{i=1}^n \sim p_{\boldsymbol{X}|\mathcal{D}} \otimes g\#p_{\boldsymbol{X}|\mathcal{D}}$:

$$\mathbb{P}(\mathrm{RUL}(\Delta_*) \leq t_j \mid \mathcal{D}) \approx \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\{\mathrm{pr}_{j+1} \circ g(\boldsymbol{X}^{(i)}) \geq \Delta_*\}. \tag{5.5}$$

5.2 Presentation of the different elementary tools

We consider the input space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_d \subset \mathbb{R}^d$ for the simulation model g. To minimize prior assumptions, we use a non-informative uniform prior with independent marginals: $p_{\boldsymbol{X}} = \mathcal{U}(\mathcal{X}_1) \otimes \ldots \otimes \mathcal{U}(\mathcal{X}_d)$. In usual industrial applications it is hard to know a prior dependency structure on the inputs if there are no prior physical information nor data to estimate it, therefore in the proposed methodology independency is assumed from the beginning. Sampling from this prior yields a design of experiments $\mathrm{DoE}_g^{p_{\boldsymbol{X}}} = \{(\boldsymbol{X}^{(i)}, g(\boldsymbol{X}^{(i)}))\}_{i=1}^n$, where n is the sample size. Since the simulation model may be computationally expensive, we construct p surrogate models $\widehat{\boldsymbol{g}} = (\widehat{g}^{\,(1)}, \ldots, \widehat{g}^{\,(p)})$ using the DoE, trained for instance with different input hyperparameters or

optimization procedures. These surrogates, built via supervised learning techniques such as GPs, PCEs (see sections 3.2.1 and 3.2.2), or even feed-forward neural networks (Tripathy and Bilionis, 2018; Lefebvre et al., 2023), approximate g as functions $\widehat{g} : \mathcal{X} \to \mathbb{R}^N$. Since no single surrogate is universally optimal, we aggregate p optimal surrogate models (see section 3.2.4) into a single robust predictor $\widehat{g}^{\,\text{agg}}$ using a weighted average on the simplex $\boldsymbol{\Delta}^{p-1}$, following expert aggregation strategies (Cesa-Bianchi and Lugosi, 2006):

$$\widehat{g}^{\text{agg}}(\boldsymbol{X}) = \sum_{i=1}^{p} w_i \widehat{g}^{\,(i)}(\boldsymbol{X}) = \boldsymbol{w}^\top \widehat{\boldsymbol{g}}(\boldsymbol{X}). \tag{5.6}$$

The weights in $\boldsymbol{w}$ are treated as hyperparameters. Assuming independent marginals in $\boldsymbol{X}$, we iteratively assess each marginal's contribution to output uncertainty to update the most influential input prior using heterogeneous data. The methodology proceeds for up to d iterations (input dimension), as summarized in Figure 5.2:

1. Initialize with $p_{\boldsymbol{X},0} = \mathcal{U}(\mathcal{X}_1) \otimes \ldots \otimes \mathcal{U}(\mathcal{X}_d)$ and generate the initial DoE.

2. Rank variable influence using a kernel-based sensitivity analysis HSIC (see section 3.3.2), averaged over all data time instances:

$$X_k = \underset{j \in \{1,\ldots,d\}}{\arg\max} \frac{1}{|\mathcal{J}|} \sum_{\ell \in \mathcal{J}} R^2_{\text{HSIC}}(X_j, g(t_\ell, \boldsymbol{X})). \tag{5.7}$$

   Select the most influential variable $\theta_k := X_k$ at iteration k, fixing the remaining variables $\boldsymbol{U}_k = \boldsymbol{u}_{0,k}$ at their nominal values.

3. Build and aggregate surrogate models as described above.

4. For q heterogeneous data groups $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q$, each with acquisition times $\mathcal{J}_i$, use a Bayesian approach to derive the posterior of θ_k given the data. The posterior integrates noise variance uncertainty and surrogate weights via Monte Carlo. Sampling from $p(\theta_k | \boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k})$ is performed using Random-Walk Metropolis-Hastings MCMC.

5. Compute the Kullback-Leibler divergence between prior and posterior to quantify information gain. If the data sufficiently informs θ_k, update its prior in $p_{\boldsymbol{X},k}$, regenerate the DoE, and retrain surrogates.

This process repeats until the information gain becomes negligible, at which point we pushforward the updated distributions.

5.2.1 Heterogeneous data groups

In some engineering prognostics applications, relevant degradation data can be collected through multiple, distinct methods. For example, direct measurements may be obtained from sensors that monitor the degradation indicator, while indirect methods may use correlated quantities and regression models—developed through feature engineering—to estimate the degradation state. Additionally, the available data often exhibit both temporal and structural heterogeneity: measurements may be acquired at irregular, non-periodic time points that do not necessarily align with the

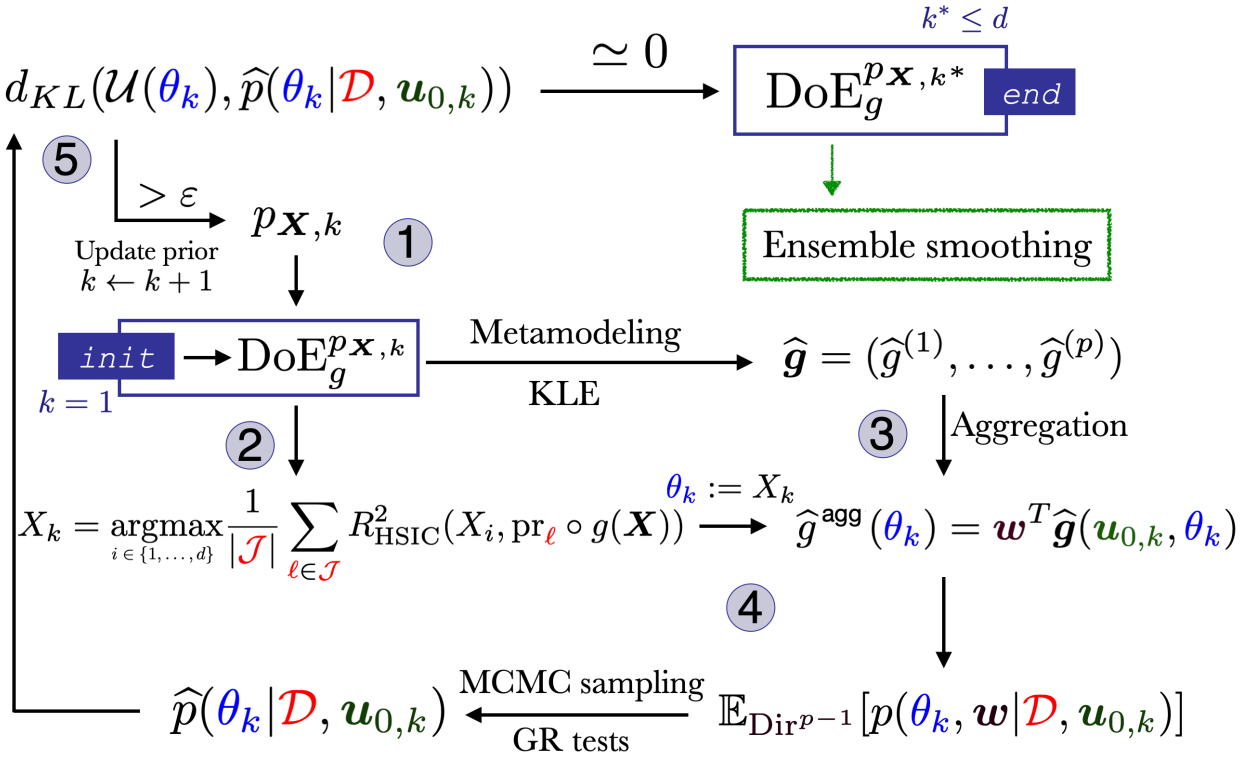

Figure 5.2: Proposed methodology for fusion of heterogeneous data in degradation prognostics.

simulation model's time grid, and the data may be sparse in time or space. To capture this, we consider q heterogeneous data groups, denoted as $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q$, where each group is associated with its own acquisition time grid $\mathcal{J}_i = \{t_1^i, \ldots, t_{m_i}^i\}$. The union of all acquisition times is $\mathcal{J} = \cup_{i=1}^q \mathcal{J}_i$, with total cardinality $|\mathcal{J}| = m_1 + \ldots + m_q$. For each group i and time index $j = 1, \ldots, m_i$, the observed data are modeled as:

$$y^i(t_j^i) = \Delta(t_j^i) + \eta_j^i, \tag{5.8}$$

where Δ is the true (latent) degradation process and $\eta_j^i \sim \mathcal{N}(0, \sigma_i^2)$ represents additive Gaussian noise. We assume the noise is homoskedastic within each group (constant variance σ_i^2 for group i), but heteroskedastic across groups (different variances between groups). This modeling approach allows us to flexibly incorporate multiple, heterogeneous sources of information, each with its own noise characteristics and sampling schedule, into the data fusion framework.

In the clogging use-case we treat ESTICOL and TVE as the heterogeneous data groups. For the SG-51B scenario we use approximately $m_{\text{ESTICOL}} \approx 100$ ESTICOL points and $m_{\text{TVE}} \approx 10$ TVE points. Their assumed variances are $\sigma^2_{\text{ESTICOL}} = 8/1.96\%$ and $\sigma^2_{\text{TVE}} = 5/1.96\%$, so that TVE measurements are considered more precise. Under the Gaussian noise assumption, these choices correspond to a 90% uncertainty band around the mean. The TPD DoE trajectories with the data points overlaid are shown in Figure 5.3.

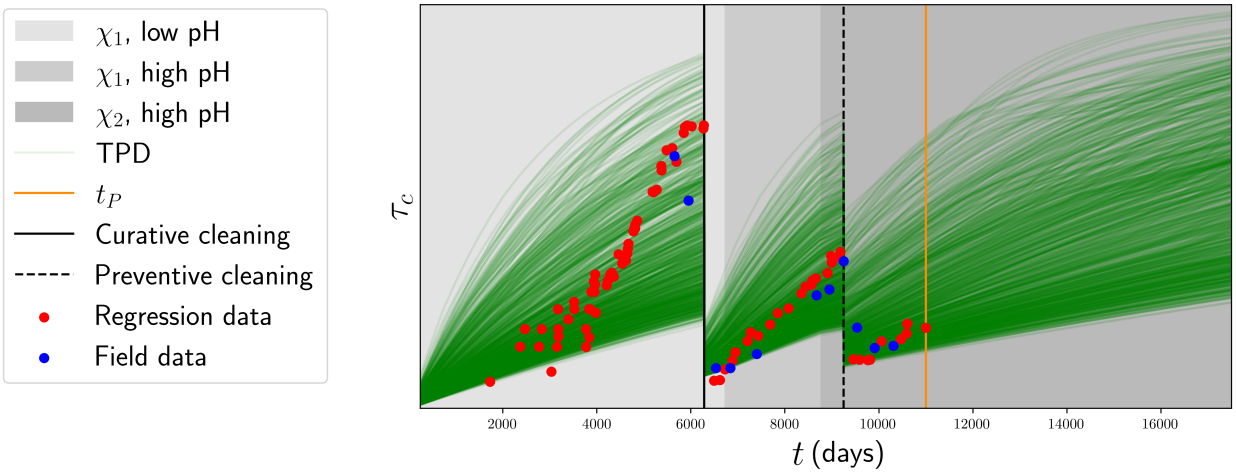

Figure 5.3: All the available data for Bayesian fusion on the SG type 51B use-case. In red the regression data points (ESTICOL) and field data points in blue (TVE).

5.2.2 Bayesian Model Updating

As explained in section 5.1, the state-space model variability is given all at once through a pushforward of the uncertain input variables independent probabilistic model. The problem is to estimate the posterior distribution after considering the heterogeneous data of the most influential input variable at iteration $k \leq d$. Assume that there is some value θ_k such that $\Delta(t_j^i) = g(t_j^i, \boldsymbol{u}_{0,k}, \theta_k)$, where θ_k is defined in Eq. (5.7) and $\boldsymbol{u}_{0,k}$ is the nominal value of the $d-1$ variables at step k. We therefore have:

$$y^i(t_j^i) = g(t_j^i, \boldsymbol{u}_{0,k}, \theta_k) + \eta_j^i, \quad \forall j \in \mathcal{J}_i. \tag{5.9}$$

Usually $\mathcal{J}_i$ is a coarser grid than the original time grid of the simulation model. Therefore, we need to interpolate the computer code output. As an interpolator operator, one can choose $\boldsymbol{\mathcal{G}}_i : \mathcal{X} \to \mathbb{R}^{|\mathcal{J}_i|}$ to interpolate the simulation model onto the time-grid of the data points. The choice of this interpolator is driven by prior regularity properties of the degradation trajectory such as monotonicity. Note that this hypothesis is verified in the SG clogging application, given the regularity of the clogging trajectory (see the corresponding paragraph in section 2.2.3). We drop the dependence on the latent variable in the rest of this paragraph, it is important to note however that the results are conditioned on these latent variables, as shown in Figure 5.2. In vector form, the problem is recast as:

$$\boldsymbol{y}^i = \boldsymbol{\mathcal{G}}_i(\theta_k) + \boldsymbol{\eta}^i, \ \forall i \in \{1, \ldots, q\}, \tag{5.10}$$

with $\boldsymbol{\eta}^i \sim \mathcal{N}(0, \sigma^2_{\boldsymbol{\eta}^i})$, where the covariance matrix of the noise is defined as $\sigma^2_{\boldsymbol{\eta}^i} := \sigma_i^2 I_{m_i}$. To simplify the presentation, it is assumed now that only one data group $\boldsymbol{y} \in \mathbb{R}^m$ is considered. The goal is to use the Bayes' theorem to estimate the posterior distribution of the influential input variable θ_k given the data:

$$p(\theta_k|\boldsymbol{y}) \propto p(\theta_k)p(\boldsymbol{y}|\theta_k). \tag{5.11}$$

This likelihood quantifies how well parameter θ_k explains the observed data. Assuming Gaussian residuals that are independent of θ_k, the likelihood is Gaussian with noise variance $\sigma^2_{\boldsymbol{\eta}}$ (homoskedas-

tic case). The posterior density is then:

$$p(\theta_k, \sigma_{\boldsymbol{\eta}}^2|\boldsymbol{y}) \propto p(\theta_k, \sigma_{\boldsymbol{\eta}}^2)\,(\sigma_{\boldsymbol{\eta}}^2)^{-m/2} \exp\left(-\frac{1}{2\sigma_{\boldsymbol{\eta}}^2}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta_k)\|^2\right), \tag{5.12}$$

where m is the number of data points. At this stage, it is possible to numerically integrate the standard deviation of the noise and modify the posterior distribution, but this often leads to numerical instabilities in MCMC chains. A carefully chosen prior on $\sigma_{\boldsymbol{\eta}}^2$ allows for analytical error integration using Bayes' theorem. We propose an extension of the proposition found in (Keller et al., 2022) to multiple heterogeneous (heteroskedastic) groups of data:

Proposition 5.2.1. *Assume* $\lambda := 1/\sigma_{\boldsymbol{\eta}}^2 \sim$ *Gamma*$(\frac{m}{2}, \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta_k)\|^2)$*, where* m *is the number of data points in* $\boldsymbol{y}$*;* $\theta \sim \mathcal{U}(\mathcal{X}_*)$*, and* $p(\theta, \lambda) \propto \lambda^{-1}$*. Then:*

$$p(\theta_k|\boldsymbol{y}) \propto \|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta_k)\|^{-m}. \tag{5.13}$$

Moreover, if multiple groups of data at different time-instances are considered for assimilation, then, one has $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q$*, with respective priors on the inverse of their standard deviations* $\lambda_i \sim$ *Gamma*$(\frac{m_i}{2}, \frac{1}{2}\|\boldsymbol{y}^i - \boldsymbol{\mathcal{G}}_i(\theta_k)\|^2)$*, and the generalization thus reduces to:*

$$p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^m) \propto \prod_{i=1}^{q} \|\boldsymbol{y}^i - \boldsymbol{\mathcal{G}}_i(\theta_k)\|^{-m_i}. \tag{5.14}$$

The proof is provided in appendix 8.3. To be exact, the posterior distribution obtained is also conditioned on the fixed latent variables, thus we have $p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k})$, and since we use an aggregation of p valid surrogates of the original degradation simulation model, we induce additional hyperparameters that we consider following a Dir^{p-1} distribution. Thus in Eq. (5.14), one can rewrite the posterior as:

$$p(\theta_k, \boldsymbol{w}|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k}) \propto \prod_{i=1}^{q} \|\boldsymbol{y}^i - \langle \boldsymbol{w}, \boldsymbol{\mathcal{G}}_i^{\text{agg}}(\theta_k)\rangle\|^{-m_i}, \tag{5.15}$$

where $\boldsymbol{\mathcal{G}}_i^{\text{agg}}(\theta_k) := (\boldsymbol{\mathcal{G}}_{i,1}(\theta_k), \ldots, \boldsymbol{\mathcal{G}}_{i,p}(\theta_k))$. At this stage we marginalize on the weights to obtain the posterior distribution of the influential input variable θ_k given the heterogeneous data as well as the aggregate surrogates:

$$p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k}) \propto \int_{\boldsymbol{\Delta}^{p-1}} \prod_{i=1}^{q} \|\boldsymbol{y}^i - \langle \boldsymbol{w}, \boldsymbol{\mathcal{G}}_i^{\text{agg}}(\theta_k)\rangle\|^{-m_i} d\boldsymbol{w}. \tag{5.16}$$

Then, a Monte Carlo approximation of the integral is performed, amounting to sampling $\boldsymbol{w}^{(1)}, \ldots, \boldsymbol{w}^{(M)}$ from the Dirichlet distribution Dir^{p-1} and computing the average of the likelihood functions:

$$p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k}) \propto \frac{1}{N} \sum_{i=1}^{M} \prod_{j=1}^{q} \|\boldsymbol{y}^j - \langle \boldsymbol{w}^{(i)}, \boldsymbol{\mathcal{G}}_j^{\text{agg}}(\theta_k)\rangle\|^{-m_j}. \tag{5.17}$$

Since for numerical purposes we will consider the log-likelihood function, $\log p(\theta_k|\boldsymbol{y}^1,\dots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k})$, then we will use the log-sum-exp trick in order to avoid numerical instabilities. The log-likelihood is rewritten as:

$$\log p(\theta_k|\boldsymbol{y}^1,\dots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k}) \propto \log \sum_{i=1}^{M} \exp\left(-\sum_{j=1}^{q} m_j \log(\|\boldsymbol{y}^j - \langle \boldsymbol{w}^{(i)}, \boldsymbol{\mathcal{G}}_i^{\text{agg}}(\theta_k)\rangle\|) - C\right) + C, \tag{5.18}$$

where $C := \max_{i=1,\dots,M} \sum_{j=1}^{q} m_j \left(\|\boldsymbol{y}^j - \langle \boldsymbol{w}^{(i)}, \boldsymbol{\mathcal{G}}_i^{\text{agg}}(\theta_k)\rangle\|\right)$. The log-likelihood is then used in the MCMC algorithm to sample from the posterior distribution of the influential input variable θ_k. This method cannot be used for integrating the latent variables uncertainties, since we have tried numerical experiments and have observed that this often averages the likelihood too much and the posterior distribution is not well estimated, especially when the dimension is important. The integration of the latent variables $\boldsymbol{u}_{0,k}$ remain an open-question with the current methodology and is left for future work.

5.2.3 MCMC with Random-Walk Metropolis-Hastings

The Random-Walk Metropolis-Hastings (RWMH) algorithm (Rubinstein and Kroese, 2011) is a MCMC method used to sample from the log-posterior distribution $\log p(\theta_k|\boldsymbol{y}^1,\dots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k})$. Starting from an initial guess $\theta_k^{(0)}$, the algorithm iteratively generates a candidate θ_k^* from a proposal distribution $q(\theta_k^*|\theta_k^{(i)})$. We choose uniform proposals centered on the mean value of the intervals $\mathcal{X}_*$ with different step-sizes. The candidate is then accepted with probability:

$$\alpha = \min\left\{1, \exp\Big[\log p(\theta_k^*|\boldsymbol{y}^1,\dots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k}) - \log p(\theta_k^{(i)}|\boldsymbol{y}^1,\dots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k}) + \log q(\theta_k^{(i)}|\theta_k^*) - \log q(\theta_k^*|\theta_k^{(i)})\Big]\right\}. \tag{5.19}$$

If accepted, the chain moves such that $\theta_k^{(i+1)} = \theta_k^*$; otherwise we stay at the same place such that $\theta_k^{(i+1)} = \theta_k^{(i)}$. This process is repeated for a large number of iterations to ensure convergence to the target posterior distribution. In order to test the convergence of the algorithm, the Gelman-Rubin diagnostic (Gelman and Rubin, 1992) is used. To do so, J Markov chains are initialized with different initial values and the RWMH algorithm is runned for a large number of iterations. After the burn-in phase, a chain $\theta_{k,1}^{(i)},\dots,\theta_{k,Q}^{(i)}$ for $i = 1,\dots,J$ is obtained where Q is the iteration number of the chain. The mean value of the chain and between chains are computed:

$$\overline{\theta}_{k,i} = \frac{1}{Q}\sum_{j=1}^{Q} \theta_{k,j}^{(i)}, \quad \overline{\theta}_{k,*} = \frac{1}{J}\sum_{i=1}^{J} \overline{\theta}_{k,i}, \tag{5.20}$$

as well as the variances of the means of the chains and the mean of the variances of one chain:

$$B = \frac{Q}{J-1}\sum_{i=1}^{J}(\overline{\theta}_{k,*} - \overline{\theta}_{k,i}^2), \quad W = \frac{1}{J}\sum_{i=1}^{J}\left(\frac{1}{Q-1}\sum_{j=1}^{Q}(\theta_{k,j}^{(i)} - \overline{\theta}_{k,j})^2\right). \tag{5.21}$$

The Gelman-Rubin diagnostic is defined as:

$$R = \frac{(1 - 1/Q)W + (1/Q)B}{W} \to 1,\ L \to \infty. \tag{5.22}$$

The algorithm is considered to have converged when R is close to 1 , i.e. for $Q \to +\infty$. The resulting sample is exponentiated and renormalized to obtain a draw from an estimate of the desired posterior, and also use kernel density estimation to obtain the functions.

5.2.4 Information gain and Kullback-Leibler divergence

The information gain (IG) quantifies the reduction in uncertainty about a random variable after observing data (Lindley, 1956). In the context of Bayesian inference, it measures how much the posterior distribution $p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k})$ differs from the prior uniform distribution $\mathcal{U}(\theta_k)$. This difference is formally captured by the Kullback-Leibler (KL) divergence, defined as:

$$d_{\mathrm{KL}}\left(\mathcal{U}(\theta_k, p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k}))\right) = \mathbb{E}_{(\theta_k|\boldsymbol{y}^1,\ldots,\boldsymbol{y}^q,\boldsymbol{u}_{0,k})} \log\left[\frac{p(\theta_k|\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q, \boldsymbol{u}_{0,k})}{\mathcal{U}(\theta_k)}\right]. \tag{5.23}$$

A higher KL divergence indicates that the data significantly updates the prior, leading to a more concentrated posterior distribution. This metric is crucial in the iterative updating of the methodology presented in figure 5.2 as it helps to determining whether the assimilation of new data justifies modifying the prior distribution of the input variables. In numerical implementations, a good trade-off value for the threshold of the Kullback-Leibler divergence was found to be $\epsilon = 0.1$. There are two possible outcomes of the algorithm: either the sensitivity assessment picks a variable whose prior has already been informed at a previous iteration, and therefore the Kullback-Leibler divergence is null, or it picks a variable that remains uninformed by the heterogeneous data, therefore posterior and prior are identical.

5.2.5 Ensemble Kalman Smoother

The Ensemble Kalman Smoother (EnKS) (Evensen and Van Leeuwen, 2000) is an extension of the Ensemble Kalman filter that assimilates all available observations to produce smoothed estimates of a system's state across all time steps. In the scalar, offline setting, we consider an ensemble of n trajectories $\{g(\boldsymbol{X}^{(i)})\}_{i=1}^n$, generated from some surrogate of the simulation model. Denote ensemble members as $g(\boldsymbol{X}^{(p)}) =: Y^{(p)} = (Y_{t_1}^{(p)}, \ldots, Y_{t_N}^{(p)})$. For the observation at time t_k^i let j be the index with $t_j = t_k^i$. Compute ensemble means and anomalies:

$$\overline{Y}_{t_\ell} = \frac{1}{n}\sum_{p=1}^{n} Y_{t_\ell}^{(p)}, \qquad A_{t_\ell}^{(p)} = Y_{t_\ell}^{(p)} - \overline{Y}_{t_\ell}. \tag{5.24}$$

Form cross-covariances between the observation index j and all state times:

$$C_{j\ell} = \frac{1}{n-1}\sum_{p=1}^{n} A_{t_j}^{(p)} A_{t_\ell}^{(p)}, \qquad C_{jj} = \frac{1}{n-1}\sum_{p=1}^{n} (A_{t_j}^{(p)})^2, \tag{5.25}$$

and compute the Kalman gain entries:

$$K_{j\ell} = \frac{C_{j\ell}}{C_{jj} + \sigma_i^2}. \tag{5.26}$$

For each ensemble member compute the innovation $d^{(p)} = y^i(t_k^i) - Y_{t_j}^{(p)}$ and update the ensemble across all times:

$$Y_{t_\ell}^{(p)} \leftarrow Y_{t_\ell}^{(p)} + K_{j\ell}\, d^{(p)}, \qquad \ell = 1, \ldots, N. \tag{5.27}$$

This update is repeated for each observation time t_k^i in a group and then for all groups, resulting in an ensemble of smoothed trajectories that incorporate the full observation sequence. The method assumes approximate linear-Gaussian behavior in the ensemble statistics and requires no model re-evaluation, making it suitable for post hoc assimilation. The procedure is described in Alg. 1 below.

Data: Ensemble $\{Y^{(p)}\}_{p=1}^n$ with $Y^{(p)} = (Y_{t_1}^{(p)}, \ldots, Y_{t_N}^{(p)}) \in \mathbb{R}^N$; observations $\{\boldsymbol{y}^i\}_{i=1}^q$ with times $\mathcal{J}_i = \{t_1^i, \ldots, t_{m_i}^i\}$; noise variances σ_i^2 (so $R_i = \sigma_i^2 I$).
Result: Smoothed ensemble $\{Y^{(p)}\}_{p=1}^n$
for $i = 1, \ldots, q$ **do**
  for *each observation time* $t_k^i \in \mathcal{J}_i$ **do**
    Let j be the index such that $t_j = t_k^i$ (projection index).
    Compute ensemble mean at t_j: $\overline{Y}_{t_j} = \frac{1}{n}\sum_{p=1}^n Y_{t_j}^{(p)}$.
    Form ensemble anomalies for all times: $A_{t_\ell}^{(p)} = Y_{t_\ell}^{(p)} - \overline{Y}_{t_\ell}$ for $\ell = 1, \ldots, N,\ p = 1, \ldots, n$.
    Compute cross-covariances (cf. (5.25)):
    $C_{j\ell} = \frac{1}{n-1}\sum_{p=1}^n A_{t_j}^{(p)} A_{t_\ell}^{(p)} \quad (\ell = 1, \ldots, N)$.
    Compute variance at observation time: $C_{jj} = \frac{1}{n-1}\sum_{p=1}^n (A_{t_j}^{(p)})^2$.
    Kalman gain for each state time t_ℓ (cf. (5.26)): $K_{j\ell} = \frac{C_{j\ell}}{C_{jj} + \sigma_i^2}$.
    for $p = 1, \ldots, n$ **do**
      Innovation (individual member): $d^{(p)} = y^i(t_k^i) - Y_{t_j}^{(p)}$.
      Update ensemble across all times (cf. (5.27)):
      $Y_{t_\ell}^{(p)} \leftarrow Y_{t_\ell}^{(p)} + K_{j\ell}\, d^{(p)}$ for $\ell = 1, \ldots, N$.
    end
  end
end

Algorithm 1: Ensemble Kalman Smoother (EnKS)

5.3 Numerical implementation

As numerical applications of the methodology, this chapter presents results on the THYC-Puffer-DEPO clogging simulation model for steam generators presented in section 2.2.4, for the SG-51B

model, and also on the SG-6819 (see appendix 8.3). We also study the BMU methodology on the Paris-Erdogan law for crack growth (Paris and Erdogan, 1960) (see appendix 8.3).

The numerical experiments described are fully reproducible. All code and application results are available in the following GitHub repository as discussed in our paper (Jaber et al., 2026). We use an optimized version of the RWMH code implemented in the OpenTURNS Python library, based on the Adaptive Metropolis algorithm to automatically tune the parameters of the Markov chain. It is also implemented in C++, allowing to have similar execution times to state of the art MCMC software such as STAN or JAGS.

Design of experiments and metamodels

Input variable	Distribution
α	$\mathcal{U}[100, 103]$
β	$\mathcal{U}[0.02, 0.025]$
ϵ_e	$\mathcal{U}[0.2, 0.5]$
ϵ_c	$\mathcal{U}[0.01, 0.3]$
d_p	$\mathcal{U}[0.5, 10.0] \times 10^{-6}$
$\Gamma_p(0)$	$\mathcal{U}[1.0, 8.0] \times 10^{-9}$
a_v	$\mathcal{U}[0, 15] \times 10^{-4}$

Table 5.1: Probabilistic modeling of uncertain input variables.

The input variables and their distributions are summarized in Table 5.1 above. These correspond to the same input variables as in section 3.1 with the difference that the prior distributions here are all uniform, as prescribed by the methodology presented in sections 5.1 and 5.2. A first full KLE metamodel is constructed on the output simulation time-grid using a constant mean and an absolute exponential kernel to predict the KL modes (see Eq. (3.23)). The resulting metamodel has a severe loss of predictive performance close to the discontinuities because of the chemical cleanings and otherwise a time-averaged Q^2 of about 0.9. Afterwards, we train $p = 12$ KLE metamodels on the linearly interpolated time instances of the data, with respective constant, linear and quadratic trends while considering both Matérn-ν with $\nu \in \{1/2, 3/2, 5/2\}$, squared exponential and absolute exponential covariance kernels. These metamodels are re-trained at each iteration step k of the methodology described in Figure 5.2 on a new DoE following a sampling from the updated distribution $\mu_{\boldsymbol{X},k+1}$ where the prior distribution of θ_k has been replaced with the posterior.

Numerical results

We perform $c = 3$ independent assimilation procedures based on scenarios before curative cleaning (CC), between curative and preventive cleaning (CC-PC), and after the last preventive cleaning (PC), corresponding to the current operational period of the SG. As can be seen from the results in Figure 5.4, the heterogeneous data fusion process has reduced the uncertainty in the input variables. In total, for each scenario, only 5 input variables are selected in the sensitivity analysis step, namely $\beta, \epsilon_c, d_p, \Gamma_p(0)$ and a_v. At each iteration, 5 MCMC chains are launched for computing the Gelman-Rubin statistic R and convergence is assessed by checking if $R \simeq 1$. The full methodology

on the three scenarios takes about 45 min on a regular computer. For all the scenarios, the most influential variable at the first iteration is a_v (the *vena contracta* calibration parameter, see section 2.2.3), thereby confirming prior results from chapter 3. The next iterations select parameters $\Gamma_p(0), d_p, \epsilon_c$ and the last considered is β. Each scenario ends with a re-selection and calibration of a_v which does not change from the first iteration.

For the sake of clarity, only the posterior distributions are displayed, as all the priors correspond to uniform distributions over the same support. The posterior distributions of parameters $\alpha, \beta, \epsilon_e$ remain nearly identical to their prior distributions, as shown in Figure 5.4, indicating that these parameters have minimal sensitivity to the output. In contrast, the distributions of input variables $\epsilon_c, d_p, \Gamma_p(0)$, and a_v are significantly updated. This aligns with the sensitivity analysis results obtained in section 3.3. Notably, parameter a_v exhibits three distinct modes corresponding to the different chemical cleanings. This behavior is linked to its role in the *vena contracta* flux, where it acts as a linear parameter directly proportional to the clogging kinetics (see Eq. (2.10)) The observed decrease in deposition speed after each chemical maintenance is accurately captured by the three decreasing modes in the posterior distribution. The updated distributions of these influential parameters lead to a more precise prediction of the clogging trajectories. Figure 5.5 illustrates the prior and posterior trajectories of the clogging rate. The posterior trajectories are better concentrated around the observed data, demonstrating the effectiveness of the proposed data fusion methodology in reducing the input uncertainty. However, it can be noted that the clogging data has a different variation profile in the first time window due to an inflexion point above a certain clogging level. This phenomenon is not simulated by the computer code, and it is most certainly due to an incompleteness of the physical model. Nonetheless, without further modifying the theory, since the RUL prediction is supposed to be grounded on the computer code, our methodology can enhance the TPD output by informing its output distribution realistic data. Although imperfect, the hybrid result allows to reduce the predictive RUL uncertainty.

Additionally, the methodology significantly improves the prediction of the RUL for a given threshold Δ_*. Figure 5.6 compares the prior and posterior RUL distributions after the last preventive maintenance. The posterior RUL distribution is narrower and more concentrated, reflecting the reduced uncertainty and enhanced reliability of the predictions. This improvement is crucial for maintenance planning and operational decision-making, as it provides a more accurate estimate of the time remaining before the next maintenance is required.

5.4 Bayesian model updating with model discrepancy

We have previously assumed that the following idealized hypothesis holds:

$$\Delta(t) = g(t, \boldsymbol{X}), \tag{5.28}$$

which postulates that the computer simulation model g is sufficiently accurate to describe the degradation index Δ without error. However, this assumption is often too strong. As pointed out in the seminal work of Kennedy and O'Hagan (2001), a more conservative and realistic approach introduces a model discrepancy term, leading to the formulation:

$$\Delta(t) = g(t, \boldsymbol{X}) + \xi(t), \tag{5.29}$$

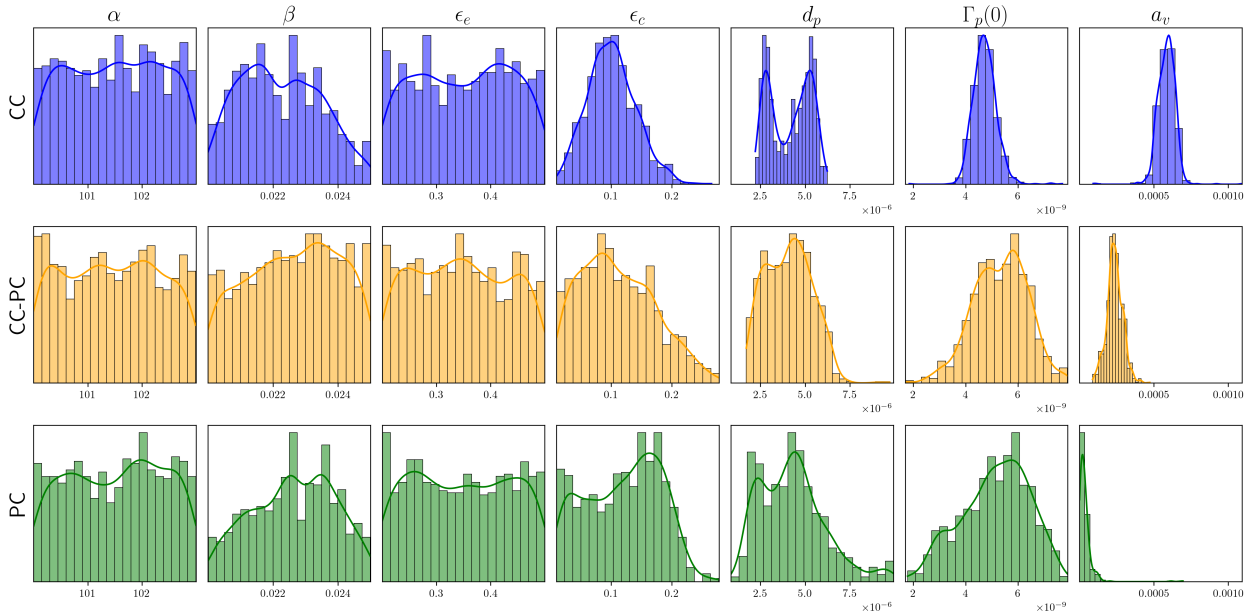

Figure 5.4: Posterior distributions of the input variables of TPD.

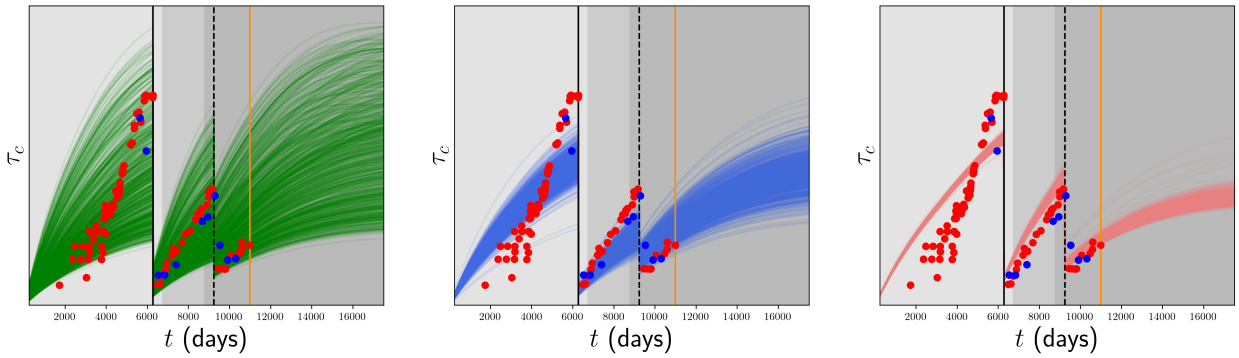

Figure 5.5: Prior and posterior trajectories of KLE-TPD emulator.

where $\xi(t)$ captures the structural bias between the computer model and the true physical system. This discrepancy term is modeled as a GP such that $\xi \sim \mathcal{GP}(0, \gamma_\xi)$, with the covariance function typically chosen as:

$$\gamma_\xi(i,j) = \sigma_\xi^2 \exp\left(-\nu_\xi |t_i - t_j|^2\right). \tag{5.30}$$

Consequently, the observation model in Eq. (5.9) becomes:

$$y(t_i^*) = g(t_i^*, \boldsymbol{X}) + \xi(t_i^*) + \eta_i, \quad \eta_i \sim \mathcal{N}(0, \sigma_\eta^2), \tag{5.31}$$

where the $\boldsymbol{y} = (y(t_1^*), \ldots, y(t_m^*))$ denotes the vector of data points. For the clarity of the presentation, we drop the details on the marginals from the BMU methodology presented in section 5.2.2, and we assume no need for an interpolating operator $\boldsymbol{\mathcal{G}}$ and consider, for the moment, a single group of data points $\boldsymbol{y}$. In vector form this is written as:

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{X}) + \boldsymbol{\xi} + \boldsymbol{\eta}, \quad \boldsymbol{\eta} \sim \mathcal{N}(0, \sigma_\eta^2 I_m). \tag{5.32}$$

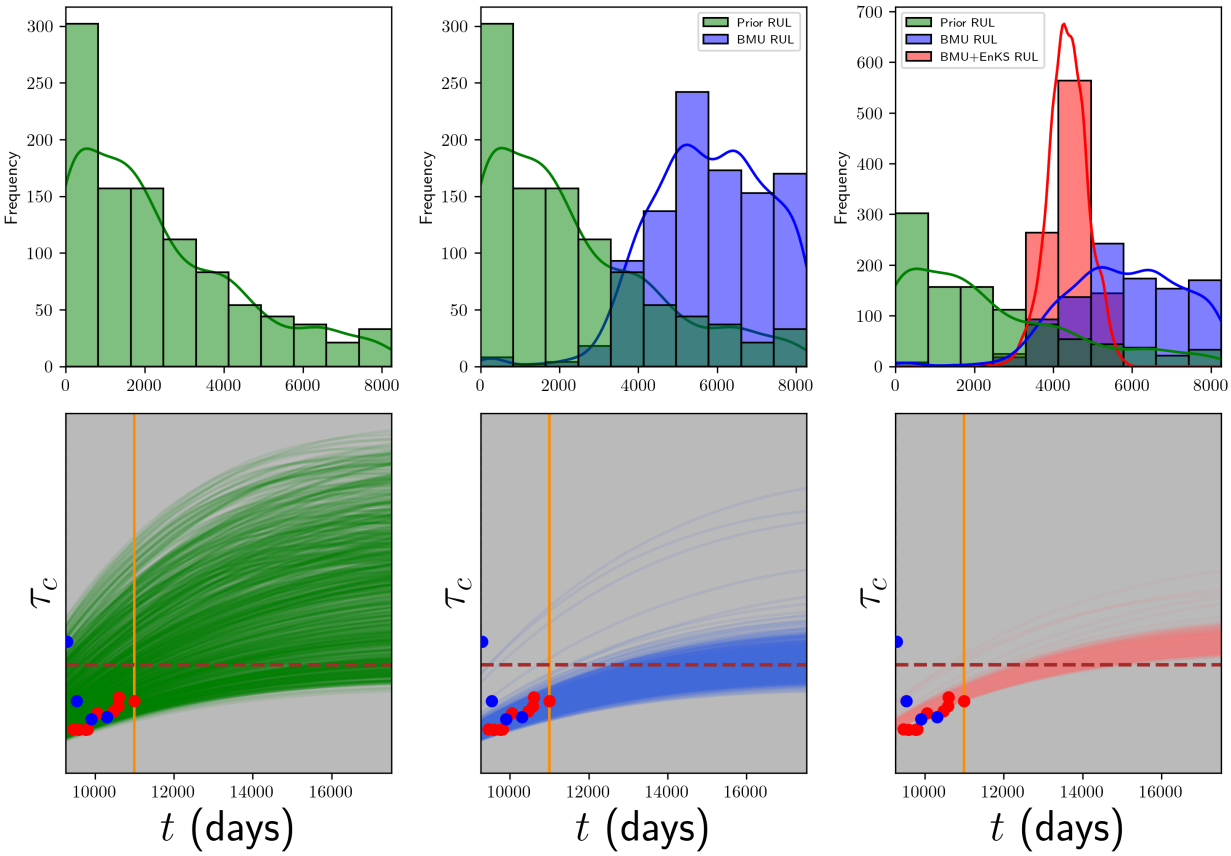

Figure 5.6: Prior and posterior RUL distributions before and after assimilation following the last preventive cleaning.

The computer model g, is supposed in Kennedy and O'Hagan (2001) to be a Gaussian process, and we build the corresponding surrogate on a DoE $\{((t_i, \boldsymbol{X}^{(i)}), g(t_i, \boldsymbol{X}^{(i)}))\}_{i=1}^n$ as in section 3.2.1 and we assume that the time-instances $\{t_1^*, \ldots, t_m^*\}$ are present in the first marginal of the inputs in the DoE. Denoting the optimized GP posterior mean function as $\widetilde{g}(t, \boldsymbol{X})$ and the posterior covariance function as γ_g, the output predictive vector at the data points is denoted $\widetilde{\boldsymbol{g}}(\boldsymbol{X}) = (\widetilde{g}(t_1^*, \boldsymbol{X}), \ldots, \widetilde{g}(t_m^*, \boldsymbol{X}))$. The predictive covariance matrix for $\boldsymbol{y}$ can be written as:

$$\Sigma_{\boldsymbol{y}}^* = \sigma_\eta^2 I_m + \boldsymbol{\gamma}_g^* + \boldsymbol{\gamma}_\xi^*, \tag{5.33}$$

where for all $1 \leq i, j \leq m$, $\boldsymbol{\gamma}_g^*(i,j) = \gamma_g((t_i^*, \boldsymbol{X}^{(i)}), (t_j^*, \boldsymbol{X}^{(j)}))$ is the GP kernel evaluated at predicted input points, and $\boldsymbol{\gamma}_\xi^*$ is the kernel of the GP modeling the discrepancy at the data time instances. The marginal likelihood for the data then becomes:

$$\mathcal{L}(\boldsymbol{X}, \sigma_\eta^2, \sigma_\xi, \nu_\xi \mid \boldsymbol{y}) \propto |\Sigma_{\boldsymbol{y}}^*|^{-1/2} \exp\left(-\frac{1}{2} \|\boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X})\|_{\Sigma_{\boldsymbol{y}}^*}^2\right), \tag{5.34}$$

where the Mahalanobis norm is defined as:

$$\|\boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X})\|_{\Sigma_{\boldsymbol{y}}^*}^2 = (\boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X}))^\top (\Sigma_{\boldsymbol{y}}^*)^{-1} (\boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X})). \tag{5.35}$$

Assuming independent priors $p(\cdot)$ over the parameters, the posterior distribution is given by:

$$\begin{aligned} p(\boldsymbol{X}, \sigma_\eta^2, \sigma_\xi, \nu_\xi \mid \boldsymbol{y}) &\propto p(\boldsymbol{X}) p(\sigma_\eta^2) p(\sigma_\xi) p(\nu_\xi) \times \mathcal{L}(\boldsymbol{X}, \sigma_\eta^2, \sigma_\xi, \nu_\xi \mid \boldsymbol{y}) \\ &\propto p(\sigma_\eta^2) p(\sigma_\xi) p(\nu_\xi) |\Sigma_{\boldsymbol{y}}^*|^{-1/2} \exp\left(-\frac{1}{2} \left\| \boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X}) \right\|_{\Sigma_{\boldsymbol{y}}^*}^2 \right). \end{aligned} \tag{5.36}$$

This equation can easily generalize to a group of heterogeneous data $\mathcal{D} = (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^q)$ defined as in section 5.2.1. In this case, the posterior distribution is given by:

$$p(\boldsymbol{X}, \sigma_\eta^2, \sigma_\xi, \nu_\xi \mid \mathcal{D}) \propto p(\sigma_\eta^2) p(\sigma_\xi) p(\nu_\xi) \prod_{i=1}^{q} \left| \Sigma_{\boldsymbol{y}^i}^* \right|^{-1/2} \exp\left(-\frac{1}{2} \| \boldsymbol{y}^i - \boldsymbol{\mathcal{G}}_i(\boldsymbol{X}) \|_{\Sigma_{\boldsymbol{y}^i}^*}^2 \right). \tag{5.37}$$

Identifiability issues in Bayesian calibration with discrepancy

A central difficulty in Bayesian calibration of computer models arises from the lack of identifiability between the calibration parameters and the model discrepancy function. This problem arises because, for any fixed value of $\boldsymbol{X}$, it is always possible to choose a discrepancy function $\xi(t)$ such that the sum $g(t, \boldsymbol{X}) + \xi(t)$ aligns with the observed data. As a result, the posterior distribution over $\boldsymbol{X}$ is not uniquely determined by the data, since deviations in the calibration parameters can be fully compensated by adjustments in the discrepancy function. This leads to confounding between $\boldsymbol{X}$ and $\xi(t)$, making it difficult to infer meaningful parameter values. The resulting posterior over $\boldsymbol{X}$ may thus be broad or multimodal, and predictive uncertainty can be significantly inflated. This fundamental issue of non-identifiability has been discussed extensively in the literature, notably in (Kennedy and O'Hagan, 2001; Brynjarsdóttir and O'Hagan, 2014; Tuo and Wu, 2015). In the following, we propose an adaptation inspired from Bayesian model averaging methods, in order to counter this problem. This is done through a specific choice of the hyperparameters σ_η^2 and σ_ξ.

A specific discrepancy modeling

We assume that we only have a specific data set $\boldsymbol{y}$, and that the scaling factor of the discrepancy can be written as $\sigma_\xi = \rho\tau^{-1}$, where $\tau \in [a, b] \subset \mathbb{R}$ is a precision factor, and $\rho \in [0, 1]$ is a weighting factor. Now define the standard deviation of the Gaussian noise also as $\sigma_\eta^2 = (1 - \rho)\tau^{-1}$, such that the predictive covariance in Eq. (5.33) can be rewritten as:

$$\Sigma_{\boldsymbol{y}}^* = \boldsymbol{\gamma}_g^* + \frac{\rho \boldsymbol{K}_\xi^* + (1 - \rho) I_m}{\tau}, \tag{5.38}$$

where we have denoted the exponential matrix on the time instances of the discrepancy covariance in Eq. (5.30) as $\boldsymbol{K}_\xi^*$. We see that with this specific choice of hyperparameters, the predictive covariance becomes a trade-off between the discrepancy and the noise matrices. We choose uniform priors for parameters $\boldsymbol{X} \sim \mathcal{U}(\mathcal{X})$, $\rho \sim \mathcal{U}[0, 1]$ and $\nu_\xi \sim \mathcal{U}[a_{\nu_\xi}, b_{\nu_\xi}]$. For the correlation length, it is assumed that the lower bound of the support is strictly positive, i.e. $a_{\nu_\xi} > 0$. Otherwise, ρ can become unidentifiable because the second term would boil down to $\tau^{-1} I_m$. For the precision parameter we choose a gamma prior on an interval $[a_\tau, b_\tau]$. The full likelihood has the following form:

$$\mathcal{L}(\boldsymbol{X}, \nu_\xi, \tau, \rho \mid \boldsymbol{y}) \propto \tau^{a-1} \times \mathbf{1}_{[a_{\nu_\xi}, b_{\nu_\xi}]}(\nu_\xi) \times \mathbf{1}_{[0,1]}(\rho) |\Sigma_{\boldsymbol{y}}^*|^{-1/2} \exp\left(-\frac{1}{2} \left\| \boldsymbol{y} - \widetilde{\boldsymbol{g}}(\boldsymbol{X}) \right\|_{\Sigma_{\boldsymbol{y}}^*}^2 \right). \tag{5.39}$$

We emphasize that the likelihood defined here can only be defined for a single data group $\boldsymbol{y}$.

Numerical example

As a test, since this model discrepancy framework can only treat a single type of data group, we test it on the THYC-Puffer-DEPO code, with only the field data corresponding to the TVEs. In order to have enough points, we do not subdivide the trajectories according to the time windows. Moreover, in TPD we only consider $\theta = a_v$, namely, the vena contracta calibration parameter defined in Eq. (2.10), and set the remaining input variables the parameters at their prior nominal value, as we have previously done in the BMU algorithm in Figure 5.2. The GP used here is trained on the full DoE presented in section 3.1, using a zero-mean and absolute exponential kernel prior. As we can see in the histograms of the posterior distributions in Figure 5.7, the weighting factor ρ has a mode closer to 1, meaning that that the discrepancy term may have a stronger impact than the noise term in Eq. (5.32). It is worth mentioning that these results are still at an early stage and more thorough investigation has to be done in order to understand these results.

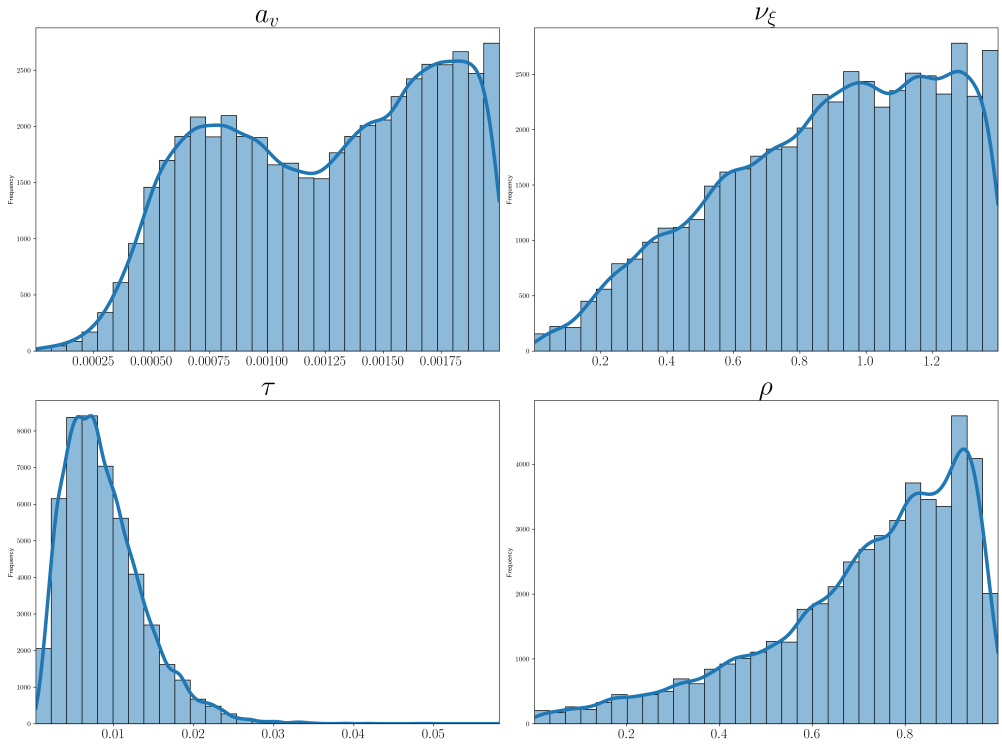

Figure 5.7: Posterior distributions for TPD with the discrepancy modeling using the entire TPD data.

5.5 Conclusion and perspectives

In conclusion, this chapter has presented a practical and modular Bayesian framework for hybrid clogging prognostics that leverages heterogeneous degradation data, surrogate modeling of a degradation computer simulation model and ensemble Kalman smoothing to substantially reduce predictive uncertainty in RUL estimation. The proposed modular method performs offline data assimilation starting with a BMU HSIC-driven prior updating combined with surrogate-aggregated

MCMC sampling and then by updating the full state vector by offline smoothing. The method demonstrates consistent RUL uncertainty distribution contraction on the THYC-Puffer-DEPO prognostics window. The extension to incorporate a model discrepancy highlights identifiability challenges and suggests some parametrizations that mitigate confounding between calibration parameters and structural error, but the results are still to be refined. Overall, the methodology offers a reproducible and computationally tractable pathway to improve prognostics when physical experiments are scarce and models are treated as grey boxes.

👉 Perspectives of chapter 5

For the scientific and industrial perspectives of this chapter, the following items can be pursued:

- Treatment of the latent-variable integration to avoid likelihood degeneration and to enable joint updating of correlated inputs (see appendix 8.3).
- Evaluate the sensitivity of the full methodology with respect to the number of datapoints used in each data group.
- Evaluate more advanced MCMC Bayesian samplers and variational approximations to scale inference to higher-dimensional input variable spaces.
- Extend the BMU methodology to field input variables, further enhance the discrepancy modeling to include different groups of data, and work on the interpretation of the results.

Chapter 6

Time-series based prognostics regression models

This final chapter adopts an exploratory perspective, focusing on the use of operational monitoring data for prognostics of steam generator clogging. We begin by presenting the available time-series sensor data. Then, building on previous work, namely the development of the static linear regression model PREVICOL, which involved handcrafted feature extraction, we assess whether these operational variables can support the construction of time-series forecasting models for the clogging rate. Due to the proprietary nature of the sensor data, much of the underlying physical context of the variables remains undisclosed. Nevertheless, from a broader standpoint on the design of real-time digital twins, investigating this modeling direction is both necessary and justified. We evaluate a series of classical regression methods, including linear regression, Ridge and Lasso regression, as well as ElasticNet. The chapter serves as a preliminary step towards assessing the viability of hybrid data-driven prognostics for long-term degradation modeling in nuclear systems.

Contents

6.1 Presentation of the available time-series data

Operational data acquisition constitutes a cornerstone of the operational regime in NPPs. Measurements are collected from a combination of analog and electronic sensors, with the latter being more deeply integrated into the plant's modern control systems and supervisory machines. A wide variety of physical quantities can be recorded, and the choice of which variables to monitor is determined by operational strategies and engineering requirements. These strategies are defined for the EDF-operated French nuclear fleet, by the UNIE (Unité d'ingénierie et d'opérations) engineering unit. The detailed strategy underlying the selection of physical quantities, sensor technologies and locations lies outside the scope of this presentation. Nonetheless, these topics are subject of emerging research, an intrusive method developed by EDF R&D involving RB methods for sensor placement in nuclear cores is found in (Argaud et al., 2018) as well as the recent paper of Karnik et al. (2024). Throughout this study, we simply assume that the engineering unit has provided the relevant sensor data for the specific SG under consideration.

The goal of this chapter is to implement a statistical regression strategy capable of providing real-time updates of clogging rate prognostics, as illustrated in the digital twin framework of Figure 1.4. One of the major challenges of clogging prognostics is that the phenomenon occurs in a real operational setting, where substantial variations of the stationary states arise and are not captured by the idealized physical model presented in Section 2.2.3. In practice, the thermal-hydraulic quantities seldom reach a perfect stationary state, making the assumptions of the physical model less realistic. Consequently, integrating operational monitoring data into our hybrid framework has the potential to significantly enhance the fidelity of the digital twin representation of the real system. Moreover, if such an online hybrid model can be constructed, feature importance analysis could reveal the main operational parameters driving the clogging kinetics and perhaps one could achieve a better real-time control of the degradation phenomenon.

Despite significant advances, several major challenges remain in the implementation of large-scale sensor monitoring within nuclear reactors. The U.S. Nuclear Regulatory Commission report on sensor instrumentation and communication technologies for digital twins (US-NRC, 2023) highlights multiple technical and operational difficulties. These include sensor qualification and reliability under harsh, radiative environments; ensuring the long-term survivability and applicability of sensors across diverse NPP configurations; developing autonomous power-harvesting solutions to maintain continuous operation; and obtaining statistically meaningful datasets for material reliability assessment and uncertainty quantification. Recent research efforts have sought to address some of these technological gaps through the use of artificial intelligence, notably via the development of machine learning-based virtual sensors such as those employing DeepONet architectures (Hossain et al., 2025). Alongside these hardware and modeling challenges, significant obstacles persist in the realm of data transmission and infrastructure, including the definition of standardized monitoring practices for key performance indicators and the design of robust, cost-effective communication networks suitable for long-term deployment in industrial environments.

Among the available operational monitoring variables, this work focuses on macroscopic physical quantities that have been identified by engineers as *potentially* correlated with the clogging phenomenon. This selection was made entirely on prior expert judgement based on empirical knowl-

edge. The dataset comprises 21 time series spanning approximately 10 years of one specific NPP operating history. As noted in Chapter 2, a given nuclear core includes three or four SG loops. For the remainder of this chapter, we restrict attention to the measurements done on the reference SG-51B studied through the THYC-Puffer-DEPO DoE in Chapter 3.

The time series contain numerous missing values and are only available over specific time frames rather than covering the entire simulation period of TPD up to the present time t_P. Overall, one can say the quality of this dataset is relatively poor and presents significant challenges for downstream learning and modeling tasks as we will see. In what follows, we present the results for the 51B SG, highlighting the superimposition[1] of the TPD-calibrated results with these operational data in Figure 6.1 below; for confidentiality reasons, the names and actual values of the variables have been altered and masked but the physical meanings and percentage of missing values are outlined in Table 6.1.

Name	**Variable**	**Domain**	%**Missing values**
X_1	`temperature_T_A`	operation	1.39
X_2	`chem_element_T_A`	chemistry	9.07
X_3	`power_T_A`	operation	2.95
X_4	`flow_rate_A`	operation	0.93
X_5	`pressure_T_A`	operation	1.60
X_6	`ph`	operation	6.05
X_7	`rotation_speed`	operation	20.52
X_8	`chem_element_T_B`	chemistry	8.16
X_9	`power_T_B`	operation	2.86
X_{10}	`flow_rate_B`	operation	8.00
X_{11}	`conductivity_T_A`	chemistry	11.04
X_{12}	`temperature_T_B`	operation	1.20
X_{13}	`pressure_T_B`	operation	94.81
X_{14}	`pressure_T_C`	operation	4.09
X_{15}	`conductivity_T_B`	chemistry	11.04
X_{16}	`chem_element_T_C`	chemistry	44.55
X_{17}	`temperature_T_C`	operation	0.76
X_{18}	`chem_element_T_D`	chemistry	7.16
X_{19}	`power_T_C`	operation	18.34
X_{20}	`conductivity_T_C`	operation	37.72
X_{21}	`flow_rate_C`	operation	17.71

Table 6.1: Selected time-series to be used in the model, grouped by physical domain and percentage of missing values.

[1] we have also selected a specific time-window in order to make the sensors start and end at the same time

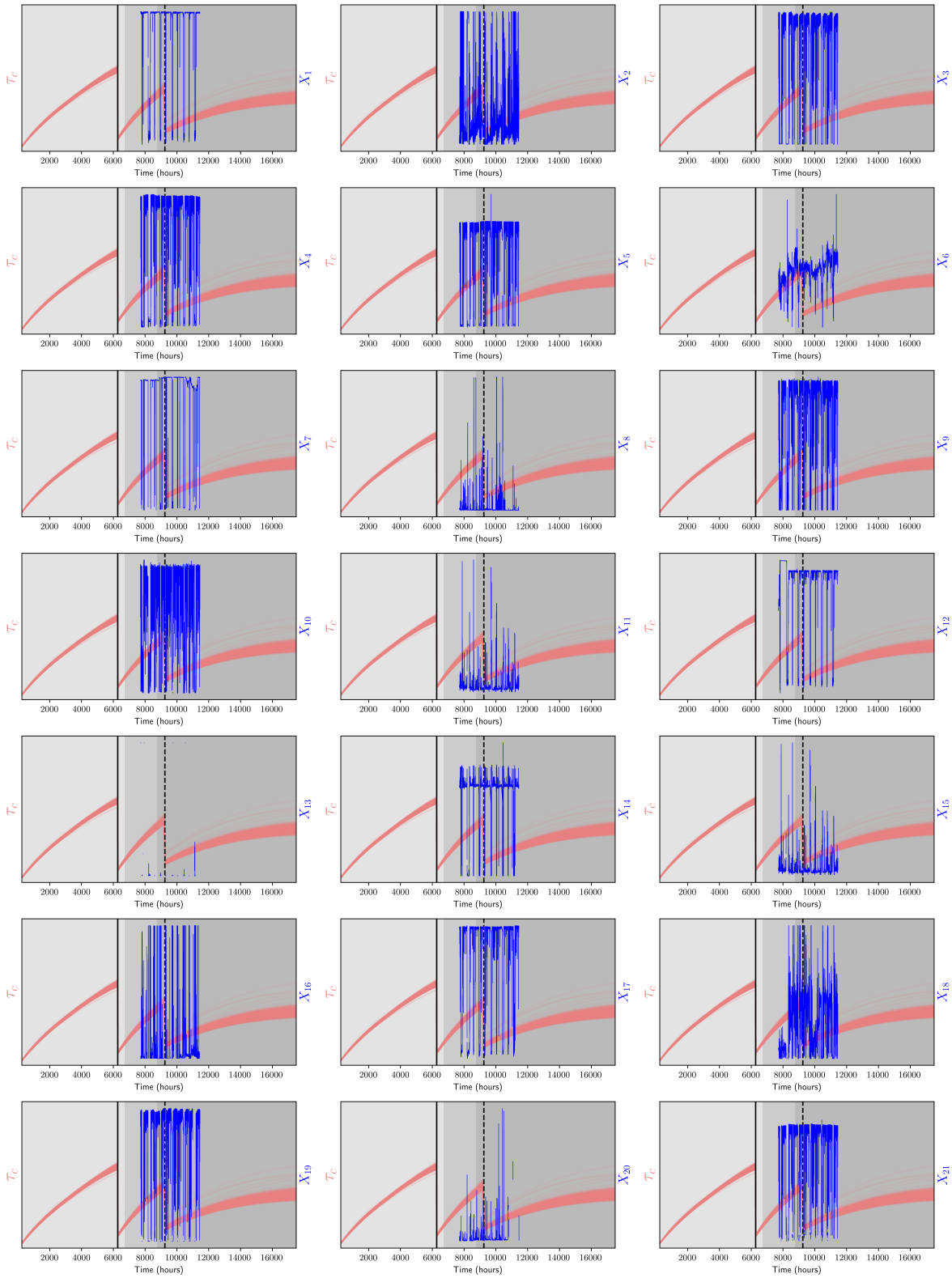

Figure 6.1: The 21 monitoring variables on the 51B steam generator, superimposed with an example of a clogging trajectory from the calibrated KLE-TPD trajectories from the results of section 5.3.

6.2 Statistical Model for Clogging Prognostics: PREVICOL

Building upon the principles of the ESTICOL framework (see Section 2.3), a regression-based strategy, known as PREVICOL, has been developed by EDF R&D to predict the future evolution of SG clogging. The central quantity of interest is the predicted increase in the clogging index:

$$\Delta\tau_c = \tau_c(t + \Delta t) - \tau_c(t), \tag{6.1}$$

where $\tau_c(t)$ denotes the clogging index at time t. The methodology is designed to simulate the clogging kinetics over a horizon of 10^4 operating hours. It extrapolates from the last available clogging measurement at t_P obtained via televised visual examination. The prediction provided by PREVICOL is defined as:

$$\tau_c^{\text{PREVICOL}}(t_P + 10^4) = \tau_c^{\text{TVE}}(t_P) + 10^4 \times \frac{\Delta\tau_c^{\text{PREVICOL}}}{\Delta t}(t_P), \tag{6.2}$$

where t_P is the last observation time. The clogging increment rate $\Delta\tau_c^{\text{PREVICOL}}/\Delta t$ is obtained from a statistical regression trained on historical field data.

The PREVICOL model was trained using TVE data collected from an entire fleet of SGs operating at a specific nominal power level. The dataset was preprocessed using a custom set of engineering rules and interpolation techniques to define meaningful differences between consecutive clogging measurements. A set of candidate features was extracted from multiple EDF operational databases, and a subset was selected following expert judgment, engineering heuristics, and empirical validation. Given the poor quality and irregularity of the time-series data, many of these features were engineered by aggregating (e.g., averaging) over appropriate temporal windows.

In total, $d = 28$ aggregated and engineered features were used as predictors in a linear regression model, trained by least-squares estimation. The resulting coefficient vector is denoted $\widehat{\theta}_{\text{PREVICOL}}$. Given a new input feature vector $\boldsymbol{Z}$, the predicted clogging increment is given by:

$$\Delta\tau_c^{\text{PREVICOL}}(t_P) = \widehat{\theta}_{\text{PREVICOL}}^T \boldsymbol{Z}. \tag{6.3}$$

Despite the thorough engineering effort behind PREVICOL, its predictive performance remains limited. The model exhibits poor generalization, especially due to the scarcity of labeled data and the heterogeneity across different SGs. In some instances, the number of available outputs is lower than the dimension of the input space, leading to overfitting or ill-posed estimation problems. In practice, model performance can vary drastically across different units, suggesting that a universal regression model applied uniformly to all SGs is unlikely to succeed. Given these limitations, the remainder of this work focuses on developing learning strategies that better exploit the time-dependent nature of the data.

6.3 Time-series-based prognostics models

This section investigates the feasibility of constructing data-driven models capable of forecasting the short-term evolution of the clogging index using multivariate operational time-series and the calibrated simulation outputs. Compared to the global PREVICOL model, this methodology is tailored for a specific SG. The overarching objective is to enhance the SG digital twin capabilities

by enabling near-real-time updates based on sensor data, independently of the physics-based model or its surrogates. To this end, we explore statistical learning techniques in a regression setting to map sequences of exogenous operational variables to simulated clogging values generated by the Bayesian-calibrated THYC-Puffer-DEPO model (Chapter 5). This modeling effort not only tests whether operational features contain predictive information about the clogging process but also aims to extend the PREVICOL methodology toward a more dynamic and responsive framework at a local level in the SG fleet.

6.3.1 Problem formulation

Let $\boldsymbol{X}(t) \in \mathbb{R}^d$ denote the d-dimensional vector of operational measurements at time t, and let $y_{\text{sim}}(t)$ be the corresponding clogging index obtained from the Bayesian-fused simulation described in chapter 5. The goal is to construct a predictive model f_θ such that:

$$y_{\text{sim}}(t+h) \approx f_{\theta^*}\big(y_{\text{sim}}(t), \boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-k)\big), \tag{6.4}$$

where k denotes the lookback window (memory), h the forecast horizon, and θ^* the optimized parameters of the model. We exclude additional autoregressive components $y_{\text{sim}}(t-j)$ with $j \geq 1$ in the model, to give weight to the exogenous variables relying solely on operational covariates. To formalize the learning setup, we define the stacked input vector:

$$\boldsymbol{Z}(t) = [y_{\text{sim}}(t), \boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-k)]^\top \in \mathbb{R}^{(k+2)\times d}, \tag{6.5}$$

and seek to approximate the map $f_{\theta^*}(\boldsymbol{Z}(t)) \approx y_{\text{sim}}(t+h)$.

6.3.2 Feature engineering

Given that $y_{\text{sim}}(t)$ is a deterministic simulation output subject to parametric uncertainties, we approximate its uncertainty by estimating the empirical standard deviation at each time step from the ensemble of Bayesian-fused TPD simulations. To improve regularization, we perturb the mean trajectory y_{sim} by adding a sample from a zero-mean Gaussian distribution scaled by this standard deviation. The operational dataset is temporally aligned with the occurrence of a maintenance event, specifically, the second preventive intervention in the 51B SG timeline. We split the data into pre- and post-maintenance segments and train two independent models $\widehat{f_1}, \widehat{f_2}$ on each segment. While it would be possible to include a categorical indicator to distinguish between regimes within a unified model, we opt for this separation as a first-stage design choice.

As illustrated in Figure 6.1, most sensor time-series are non-monotonic, with the exception of X_6 which shows weak accumulation. Since clogging is a cumulative degradation process, we augment the variables by computing cumulative sums over time. For each time step t_K, we define:

$$\widetilde{\boldsymbol{Z}}(t_K) = \sum_{i=1}^{K} \boldsymbol{Z}(t_{K-i}) \tag{6.6}$$

in order to cope with the missing data. To address missing data, we apply a linear interpolation operator $\mathcal{L}$ to each component of $\widetilde{\boldsymbol{Z}}(t)$ and the final engineered input becomes:

$$\boldsymbol{Z}^*(t_i) = \mathcal{L} \circ \widetilde{\boldsymbol{Z}}(t_i) + \epsilon_K, \tag{6.7}$$

where a small noise $\epsilon_K \sim \mathcal{N}(0, 0.01^2)$ is added to account for the sensor uncertainty. The resulting features for $\widehat{f_2}$ are shown in Figure 6.2 below. For our experiments, we consider lookback lengths $k \in \{10, 50, 100, 200\}$ and prediction horizons $h \in \{100, 500, 1000, 2000\}$, yielding datasets of the form:

$$\mathcal{D}_{k,h} = \{\boldsymbol{Z}^*(t_i), y_{\text{sim}}(t_i + h)\}_{i=1}^{n+m}, \tag{6.8}$$

which are split into training (past) and test (future) subsets using an 80/20% split. After training, model performance is evaluated on the testing set by computing the root-mean-square error (RMSE) and the Q^2 regression coefficient (see Eq. (3.12)). The RMSE is given by:

$$\text{RMSE} = \left(\frac{1}{m} \sum_{i=n+1}^{n+m} \left(y_{\text{sim}}(t_i + h) - \widehat{f}(\boldsymbol{Z}^*(t_i)) \right)^2 \right)^{1/2}. \tag{6.9}$$

6.3.3 Regression models

We restrict attention to a linear regression approach, following the PREVICOL philosophy: the engineered inputs are treated as independent observations and not as a sequential time-series. Concretely, we model each $\boldsymbol{Z}^*(t)$ as a flat regression input to a linear estimator and focus on training $\widehat{f_2}$, because it represents the forecasting window of interest. The dataset for the $\widehat{f_2}$ regression problem contains $\sim 50 \times 10^3$ points in total. In addition to ordinary least squares, we study penalized linear variants (Ridge and Lasso) whose regularization hyperparameters are optimized using blocked cross-validation appropriate for time series (see Figure 6.3 below). Concretely, we employ `sklearn.model_selection.TimeSeriesSplit` with 5 folds to avoid leakage from future to past. The selection criterion is the lowest cross-validated Q^2 and final performance is reported on the held-out test set using both RMSE and Q^2 as defined earlier.

Linear Regression

We begin with a simple linear regression model of the form:

$$f_\theta(\boldsymbol{Z}^*(t)) = \theta^\top \boldsymbol{Z}^*(t), \tag{6.10}$$

where $\theta \in \mathbb{R}^{(k+2)\times d} =: \Theta$ are model coefficients learned via ordinary least squares:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left(y_{\text{sim}}(t_i + h) - \theta^T \boldsymbol{Z}^*(t_i) \right)^2. \tag{6.11}$$

The baseline linear regression yields reasonable performance for most combinations of history length k and forecast horizon h as we can see in Figure 6.4. RMSE values are generally in the 1.05-1.10 range, with the smallest errors (around 1.05) observed for $(k = 10, h = 100)$ and $(k = 10, h = 1000)$. Larger errors (RMSE > 1.2) appear for $(k = 10, h = 500)$ and $(k = 200, h = 2000)$. The Q^2 scores are relatively stable, peaking near 0.23 at $(k = 10, h = 100)$ and remaining above about 0.21 for many configurations, though they drop to near zero or slightly negative for the worst cases. Overall, despite its simplicity and lack of regularization, the linear baseline explains a non-trivial portion of the variance for shorter horizons, but its predictive capability is limited (maximum $Q^2 \approx 0.23$).

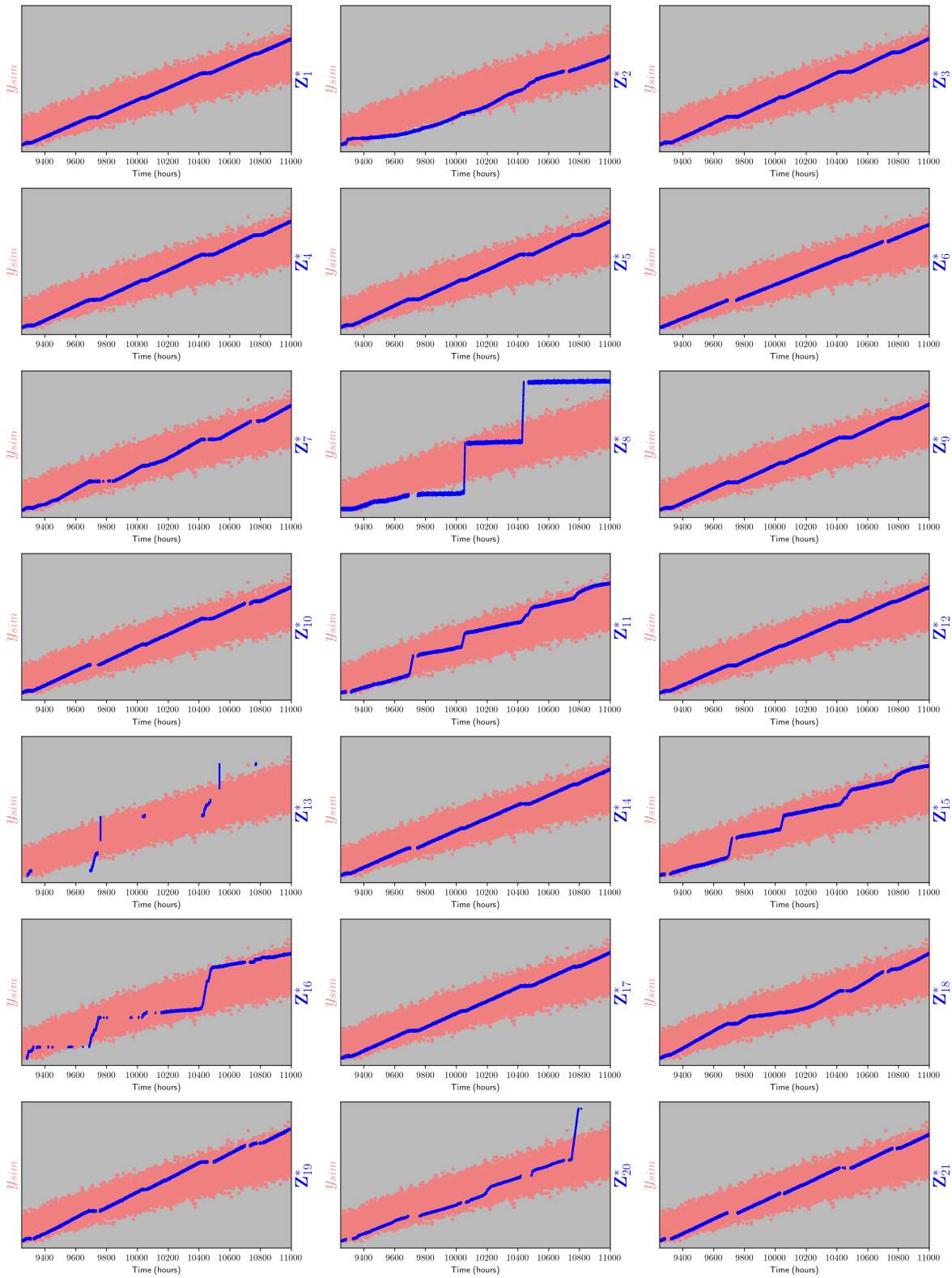

Figure 6.2: The 21 features after the preventive chemical cleaning until t_P used for training model $\widehat{f}_2$.

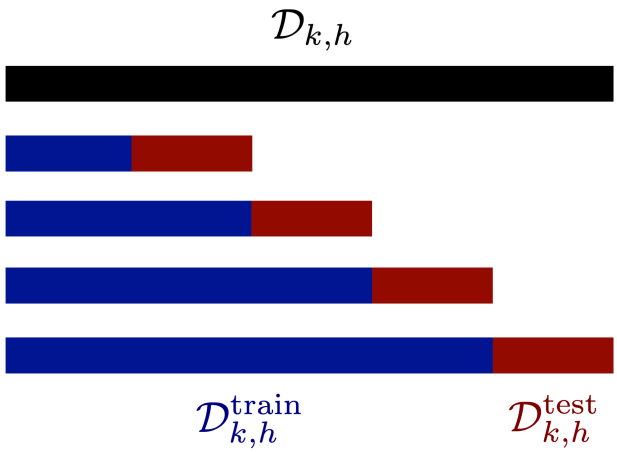

Figure 6.3: Illustration of the `sklearn.model_selection.TimeSeriesSplit` cross-validation dataset split method adapted for time-series, avoiding future time leakage. An example with 4 folds.

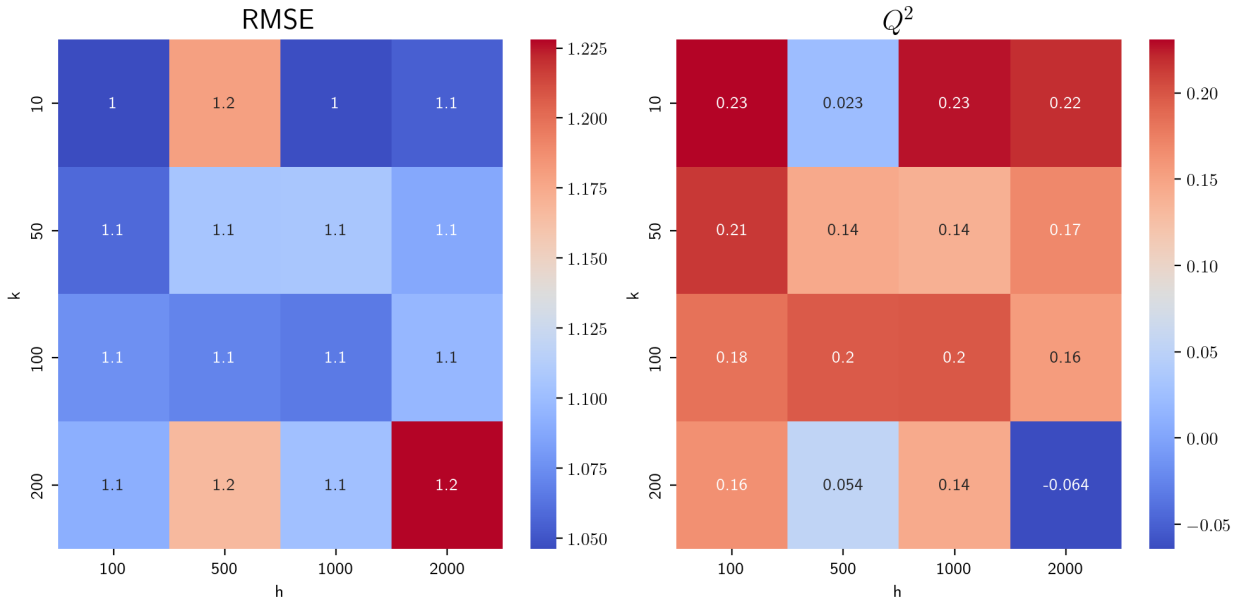

Figure 6.4: RMSE and Q^2 scores as a function of history length k and prediction horizon h for the linear regression model.

Ridge Regression

Ridge regression (Hoerl and Kennard, 1970) adds an ℓ^2 penalty to the least-squares objective:

$$\widehat{\theta}_{\text{Ridge}} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left(y_{\text{sim}}(t_i + h) - \theta^T \boldsymbol{Z}^*(t_i) \right)^2 + \alpha \|\theta\|_2^2, \tag{6.12}$$

where $\alpha > 0$ is the regularization strength tuned by `TimeSeriesSplit` CV. Ridge stabilizes the estimation when the number of features is large or collinear and prevents high-variance solutions. For each estimator we tune α over 41 logarithmically spaced values between 10^{-4} and 10^{4}. As shown in Figure 6.5, Ridge regression produces more consistent and generally better performance

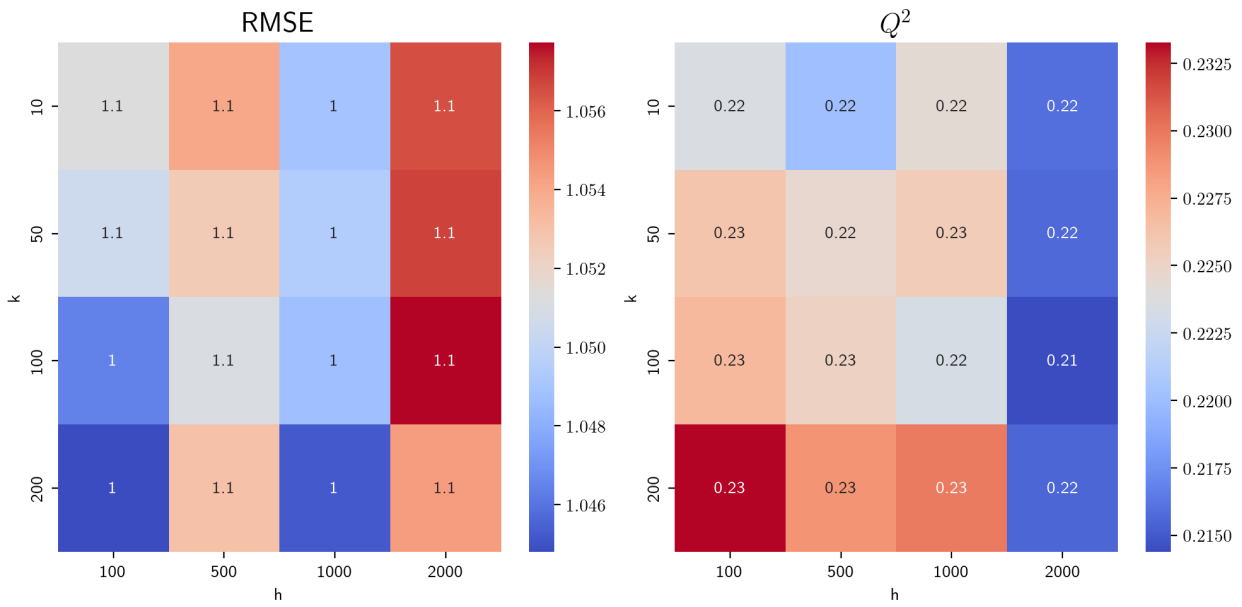

Figure 6.5: RMSE and Q^2 scores as a function of history length k and prediction horizon h for the Ridge regression model.

across the examined (k, h) pairs than the unregularized linear model. The best configuration is $(k = 200, h = 100)$ with RMSE ≈ 1.046 and a peak $Q^2 \approx 0.2325$, and Q^2 stays above 0.21 for most settings. It is worth noting that also on long horizons $h = 2000$ the models perform less well than the rest. This means that the statistical method based on these time-series is limited as far as the time-extrapolation horizon can go. The improvement in predictive accuracy is modest and the maximum Q^2 is comparable to the previous best linear model. Ridge's benefit appears to come from its ℓ_2 regularization, which mitigates overfitting and stabilizes estimates across different history lengths and horizons.

Lasso Regression

Lasso regression (Tibshirani, 1996) employs an ℓ^1 penalty that promotes sparsity and is useful to perform automatic feature selection in high-dimensional or correlated inputs. :

$$\widehat{\theta}_{\text{Lasso}} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left(y_{\text{sim}}(t_i + h) - \theta^T \boldsymbol{Z}^*(t_i) \right)^2 + \alpha \|\theta\|_1. \tag{6.13}$$

We tune α on the same logarithmic grid used for Ridge. Overall, Lasso achieves performance comparable to Ridge, with similar Q^2 scores and low average RMSE, as shown in Figure 6.6. The highest Q^2 is about 0.23, matching the linear and Ridge estimators, and is observed at $(k = 100, h = 100)$. Consistent with Ridge, predictive skill deteriorates for the longest horizon $(h = 2000)$, exhibiting lower accuracy and predictive power. RMSE values typically lie in the 1.05–1.08 range.

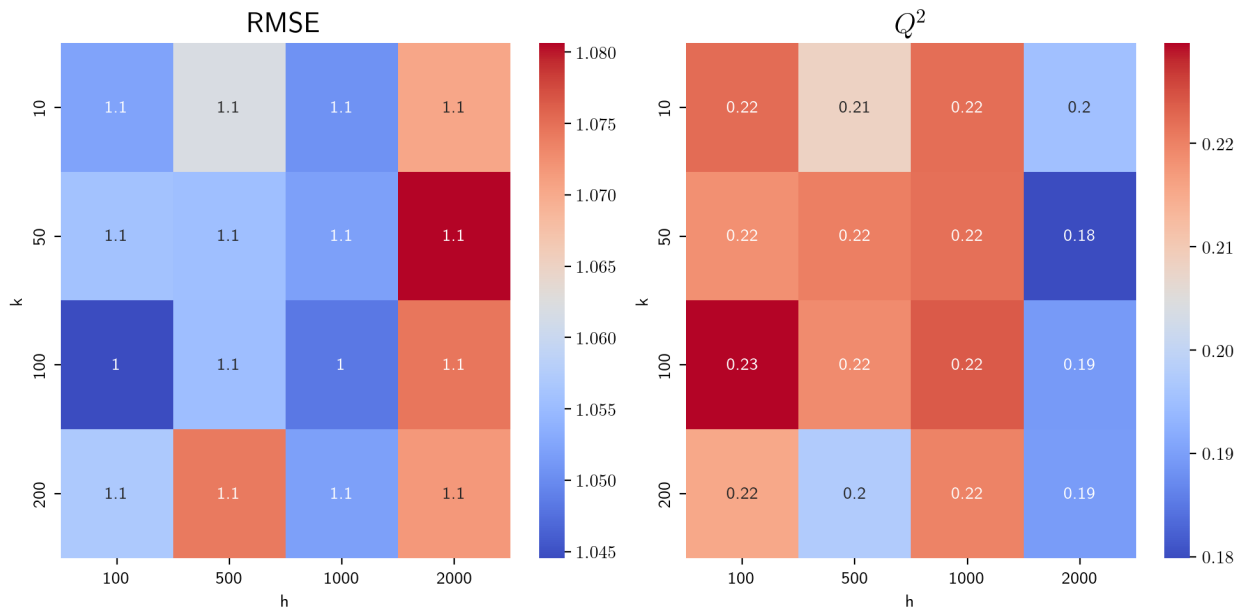

Figure 6.6: RMSE and Q^2 scores as a function of history length k and prediction horizon h for the Lasso regression model.

Elastic Net

Elastic Net combines ℓ^1 and ℓ^2 penalties:

$$\widehat{\theta}_{\text{EN}} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \left(y_{\text{sim}}(t_i + h) - \theta^T \boldsymbol{Z}^*(t_i) \right)^2 + \alpha \left(\rho \|\theta\|_1 + (1-\rho) \|\theta\|_2^2 \right), \tag{6.14}$$

where $\alpha > 0$ is the overall regularization strength and $\rho \in [0, 1]$ is the ℓ^1 ratio. We tune α on the same logarithmic grid and $\rho \in \{0.1, 0.5, 0.9\}$ with `TimeSeriesSplit` cross-validation. Elastic Net interpolates between Ridge and Lasso, offering both grouping behavior for correlated predictors and some sparsity. The results in Figure 6.7 indicate that the Elastic Net attains slightly improved average RMSE and Q^2 across the examined (k, h) combinations. As with Ridge and Lasso, the longest horizon ($h = 2000$) remains poorly predicted, reflecting the difficulty of extrapolation at that scale. Overall, for horizons below 1000 h the Elastic Net yields modest gains in predictive skill, with average Q^2 values around 0.23, comparable to the best performances observed with the simpler linear models.

Therefore, we conclude that reliable long-term prognostics are not attainable with the current dataset. Practically, horizons up to 1000 h (≈41 days) may be considered for maintenance planning, but the predictive performance remains modest and should be treated with caution. In our experiments the best results reach RMSE ≈ 1 and $Q^2 \approx 0.23$ using up to 200 lookback steps. It is unclear whether more complex models would uncover additional correlations. More promising gains are likely to come from explicitly modelling temporal dynamics of the time-series and, above all, from improving data quality at acquisition since the time series display numerous defects and the present feature engineering is largely ad hoc.

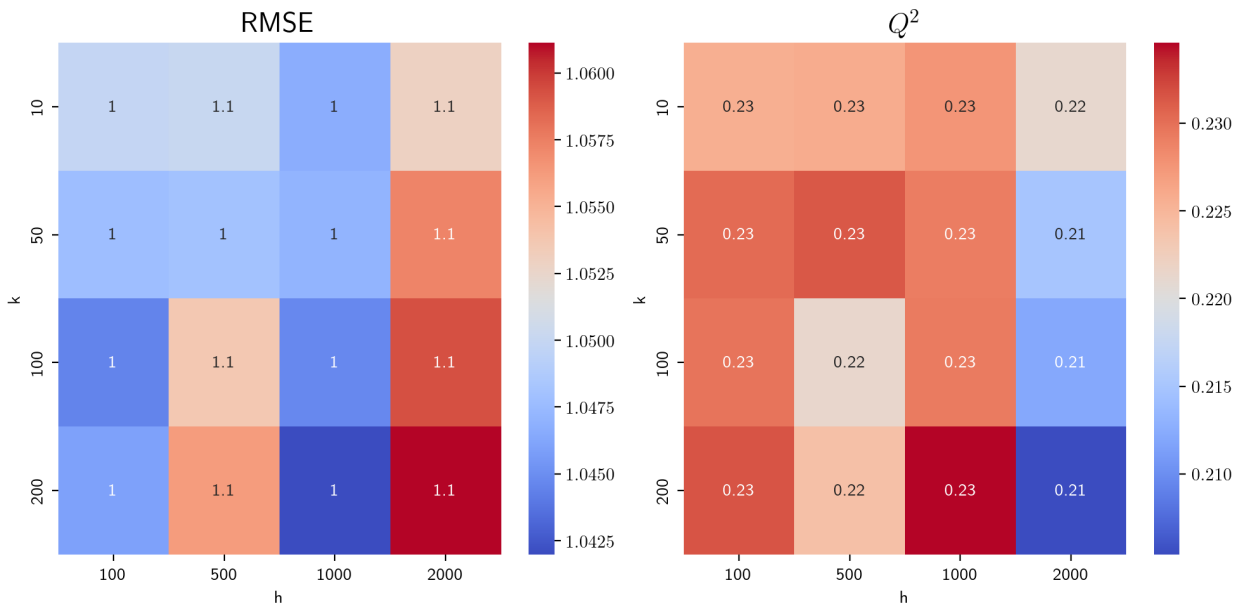

Figure 6.7: RMSE and Q^2 scores as a function of history length k and prediction horizon h for the Elastic Net regression model.

6.4 Conclusions and perspectives

The objective of this chapter was to assess whether multivariate operational monitoring data can support short-term prognostics of SG clogging within an online, hybrid framework where simulations provide training targets. Building on the Bayesian-fused TPD outputs from Chapter 5, we formulated forecasting tasks mapping recent operational histories to future simulated clogging indices, and evaluated a family of linear and penalized linear estimators across multiple history-horizon configurations.

Empirically, the experiments highlighted strong limitations driven primarily by the quality and structure of the available signals. The selected macroscopic variables suffer from substantial missingness, irregular time grids and unclear noise characteristics. Simple preprocessing (linear interpolation, cumulative augmentation and small additive noise) allowed model training, but only the linear-style estimators yielded modest predictive performance as measured by the Q^2 score (of about 0.23 for the regularized models on different (k, h) pairs). These results suggest that, in their current form, the operational features contain limited exploitable information for reliable short-term forecasting of clogging beyond what the Bayesian-fused simulator THYC-Puffer-DEPO already provides.

From a practical standpoint, the slow kinetics of clogging reduces the immediate operational need for high-frequency, short-horizon forecasts: the Bayesian-fused simulator remains the most informative tool for long-term RUL and maintenance planning. Nevertheless, a statistical layer could still add value if informative covariates or improved preprocessing strategies are identified, enabling timely detection of deviations from expected degradation trajectories. In summary, this chapter clarified that meaningful short-term clogging prognostics from the present operational dataset are not yet achievable with simple feature-based regressors. The hybrid paradigm that combines val-

idated physics-based simulation for outputs with principled online statistical inference remains a promising framework, but its success hinges on improved sensor data quality and developing more reliable feature engineering.

👉 Perspectives of Chapter 6

For the scientific perspectives of this chapter, the following items are of interest:

- Improve the interpolation and masking strategies used for handling missing data. Explore advanced imputation techniques or filtering methods tailored to the expected signal structure.
- Apply feature engineering methods aimed at learning only the stochastic fluctuations (spread) in the clogging trajectories, possibly by detrending the simulations.
- Investigate unsupervised learning methods to uncover latent representations or hidden correlations between sensor variables and clogging behavior.
- Explore conformal prediction techniques for uncertainty quantification in time series forecasting, following the approach of Zaffran et al. (2022).

Chapter 7

Conclusion and Perspectives

This thesis has addressed the problem of long-term prognostics of steam generator clogging in nuclear pressurized water reactors with the overarching objective of developing a hybrid methodology capable of producing robust, risk-informed predictions under uncertainty to assist maintenance planning of chemical cleanings. By combining physical modeling, statistical learning, and uncertainty quantification within a unified framework, this work contributes to the design of digital twins for critical components of nuclear power plants and in particular the steam generator (Deri et al., 2021).

After outlining the main physical mechanisms of clogging and the industrial context, the study began with a detailed analysis of the physical mechanisms involved in clogging, leading to the numerical formulation of the clogging simulation computational chain THYC-Puffer-DEPO in chapter 2 that allows to simulate the clogging kinetics over the entire lifespan of the SG. Treating this code as a deterministic solver, a full non-intrusive uncertainty quantification procedure was established in chapter 3, based on the probabilistic modeling of seven key input model parameters. The resulting uncertainty propagation and sensitivity analyses revealed the dominant influence of a subset of parameters in this input vector, namely the clogging deposit porosity ϵ_c, the iron oxide particle diameter d_p, the initial solid mass fraction $\Gamma_p(0)$, and the vena contracta calibration parameter a_v. These findings, validated by both Sobol' and HSIC indices, provided new physical insight into the interplay between chemical conditioning and deposit microstructure. In particular, the dependence of ϵ_c on the pH regime suggested a coupling between the colloidal aggregation kinetics and the thermal-hydraulic environment, which had not been explicitly documented before. Moreover, the uncertainty propagation on THYC-Puffer-DEPO provides a first measure of the uncertainty in the RUL prediction, which is central to the prognostics problem.

To reduce computational costs and facilitate global sensitivity analysis in chapter 3, several surrogate models were implemented and compared: Gaussian process regression, vector polynomial chaos expansions, and field decomposition through Karhunen-Loève expansions combined with different surrogates to learn the modes. Each of these techniques offered complementary advantages in terms of flexibility, interpretability, and numerical efficiency. Moreover, an aggregation strategy was subsequently proposed to mitigate metamodel-form bias and to ensure stable surrogate predictions.

A significant methodological contribution of this thesis lies in the introduction of conformal prediction techniques for the qualification of surrogate models of computer codes. The proposed J+GP and J-minmax-GP estimators extend classical cross-conformal schemes to Gaussian processes, yielding adaptive prediction intervals with finite-sample coverage guarantees. Unlike Bayesian credibility intervals, these conformal intervals require no assumption of Gaussianity and remain valid even under model misspecification. Their application to the clogging surrogate models demonstrated the ability of conformal prediction to deliver both reliability and adaptivity in uncertainty quantification, thus providing a mathematically rigorous complement to conventional surrogate model validation metrics such as the predictivity coefficient.

The central contribution of this thesis with respect to the prognostics problem was addressed in chapter 5 with the fusion of heterogeneous data sources through a modular data assimilation method. A general framework for Bayesian model updating with model discrepancy was established, allowing the posterior reduction of epistemic uncertainties using sparse field data such as televised video examinations and clogging estimates provided by the regression algorithm ESTICOL. The resulting hybrid strategy integrates physics-based and statistical elements, enabling the computation of reliable probabilistic RUL estimates. The approach also integrates an Ensemble Kalman Smoother, offering a refined uncertainty reduction estimation of degradation trajectories over the steam generator lifespan.

In chapter 6 a prospective study examined time-series regression approaches for clogging prognostics. Various architectures were evaluated across different lookback windows (inputs) and forecast horizons (output), including linear, Ridge, Lasso and ElasticNet regressors. These models were benchmarked to evaluate their ability to extrapolate the current clogging rate from past monitoring data. The results were inconclusive, with predictivity coefficients remaining unsatisfactory; nonetheless this chapter explored physics-informed learning strategies that fuse simulation outputs with monitoring inputs, forming a hybrid approach within a digital-twin framework.

Main contributions

The main scientific and technical contributions of this thesis work can be summarized as follows:

1. *A complete non-intrusive uncertainty quantification and sensitivity analysis of the industrial simulation chain THYC-Puffer-DEPO.*

2. *The construction and qualification of multiple typologies of surrogate models (GP, VPCE, KLE) and their convex aggregation, which can be useful for minimizing bias in downstream usage of (for instance in inverse problems).*

3. *The introduction of conformal prediction for surrogate model validation, with new J+GP and J-minmax-GP estimators providing marginal coverage guarantees, as well as extensions for field outputs using conformal risk control.*

4. *The development of an offline Bayesian hybrid data-fusion methodology integrating heterogeneous data-sources (simulation model and ESTICOL, TVE data).*

5. *A first exploration of time-series prognostics models for integrating monitoring data integration and computer simulation outputs, offering an additional perspective on hybrid learning approaches.*

This research fits within the strategic objectives of EDF's "Jumeau Numérique Générateur de Vapeur" research programme, aiming at the development and deployment of digital twin methodologies for the monitoring and predictive maintenance of steam generators. The proposed hybrid framework provides a reproducible and physically interpretable basis for risk-informed decision-making. From an industrial standpoint, the results of this thesis contribute to improving the planning of preventive chemical cleanings and therefore optimizing operational availability of the nuclear power plants.

From a scientific perspective, this thesis illustrates how uncertainty quantification and statistical learning can be coherently integrated within the broader paradigm of scientific machine learning for developing digital twins of complex operational systems. It builds a methodological bridge between deterministic computer simulation, probabilistic modeling, and data assimilation, all of which are key ingredients toward certified and trustworthy digital twins for engineering industrial systems.

Long-term scientific perspectives on SG clogging prognostics

- The current data-assimilation strategy is inherently offline; enabling an online (real-time) configuration requires more regular, non-destructive measurements (e.g., eddy-current inspections through the *IComplexe* methodology) to provide frequent clogging indicators and continuous assimilation into the THYC-Puffer-DEPO state.

- Build dedicated experimental validation datasets (e.g with the help of the COLENTEC experimental loop) and strengthen collaboration between experimentalists and modelers to calibrate and benchmark alternative physical formulations and closure laws in order to enhance the physical modeling.

- Benchmark clogging models with different levels of physical fidelity (especially taking into account the local nature of clogging at the TSP level), perform higher-order sensitivity analyses of phenomenological closure laws used in THYC, and systematically propagate parametric uncertainties across the different components of the computational chain (e.g. THYC, Puffer), possibly using stochastic field representations.

- Investigate intrusive reduced-order modeling techniques to accelerate parts of the computational chain, especially in order to develop iterative simulators required for the online updating in filtering.

- Move beyond global averages by characterizing the spatial structure and locality of clogging at the tube/TSP level (noting differences between cold and hot branches), and develop multiscale or macroscopic transient models that reconcile microscopic kinetics with plant-scale operating dynamics.

Long-term industrial perspectives

- *Integration into operational infrastructure:* In the medium term, embed the UQ and hybrid modules developed in this thesis within EDF's JNGV platform to promote adoption by the UNIE engineers. This requires developing intuitive visualization dashboards, secure data pipelines, and user-centered workflows so the modules become easy to deploy and operate seamlessly. It also requires applying the methodology to the entire SG fleet concerned by this degradation phenomenon.

- *Scalability and reuse across assets:* The proposed methodology is modular and broadly applicable: it can be perhaps adapted to other SG failure modes (e.g., tube rupture or fouling) and to digital twins in other fields such as aerospace, other power generation fields, and process industries where physics and scarce data must be reconciled under uncertainty.

- *Standards, governance and explainability:* Wider deployment depends on advances in standardization (analogous to VV&UQ protocols), regulatory and policy frameworks, and tools for explainable industrial digital twins. Establishing validation, verification, and uncertainty quantification workflows will be essential to gain both regulatory and operator confidence.

- *Societal and industrial impact:* Reliable digital twins can improve predictive maintenance, reduce unplanned outages, and support risk-informed decision making — contributing to increased operational availability and public confidence in nuclear technology, while helping meet the future energy needs of societies.

In conclusion, this thesis demonstrates that reliable prognostics and digital twins for complex industrial systems require the convergence of three methodological pillars: physics-based modeling, statistical learning, and uncertainty quantification. The hybrid approaches proposed herein contribute to this integration by providing methods that are not only predictive but also interpretable and verifiable. Beyond the specific case of steam generator clogging, they open promising avenues for the development of scientific machine learning tools with quantifiable credibility, forming the groundwork for the next generation of explainable and reliable digital twins in the nuclear industry and beyond.

Paris, ce 7 novembre 2025

Chapter 8

Appendix

Contents

8.1 Chapter 3

Stochastic GPs with one parameter

We consider the stochastic time-collocation GP methodology applied to a design of experiments including only the a_v uncertain variable. It can be seen in Fig 8.1 below that the discontinuity is well approximated in this low-dimensional setting.

Additional results on PCE optimization

To select the best hyperparameters of quasi-norm q and degree p, the variability of $\overline{Q^2}$ with respect to the permutation of the training/test samples is studied, as done in Jaber et al. (2025b). From the available data batch, 75% is used for training and 25% for testing while 5 permutation attempts are made. In Fig 8.2, we observe the box-plots of the time averaged predictivity coefficients for different degrees as a function of the q-norm. With the choice of the full-basis ($q = 1.0$), the metamodel overfits at all degrees, as seen through the notable drop in predictivity. Significant levels of $\overline{Q^2}$ are achieved for q-norms greater then 0.5 (with values $\overline{Q^2} \geq 0.98$) and the variability of the coefficient is reduced. By considering the evolution of the sequence $\{Q^2(t_k)\}_{k=1,\dots,N}$ in Fig 8.3 for PCE metamodels with degrees p and q-norms maximizing the time-averaged predictivity, we observe that there is a decrease of values at later times in the χ_2 high pH regime. Moreover, we see that for degrees $p = 7, 8, 9$, there is no major increase of Q^2 on average. For limiting the complexity of the metamodel, we therefore select the PCE with maximal degree $p = 7$ and hyperbolic enumeration rule $q = 0.7$. Globally, the resulting surrogate model can be confidently used for approximating the output and also to perform sensitivity analysis.

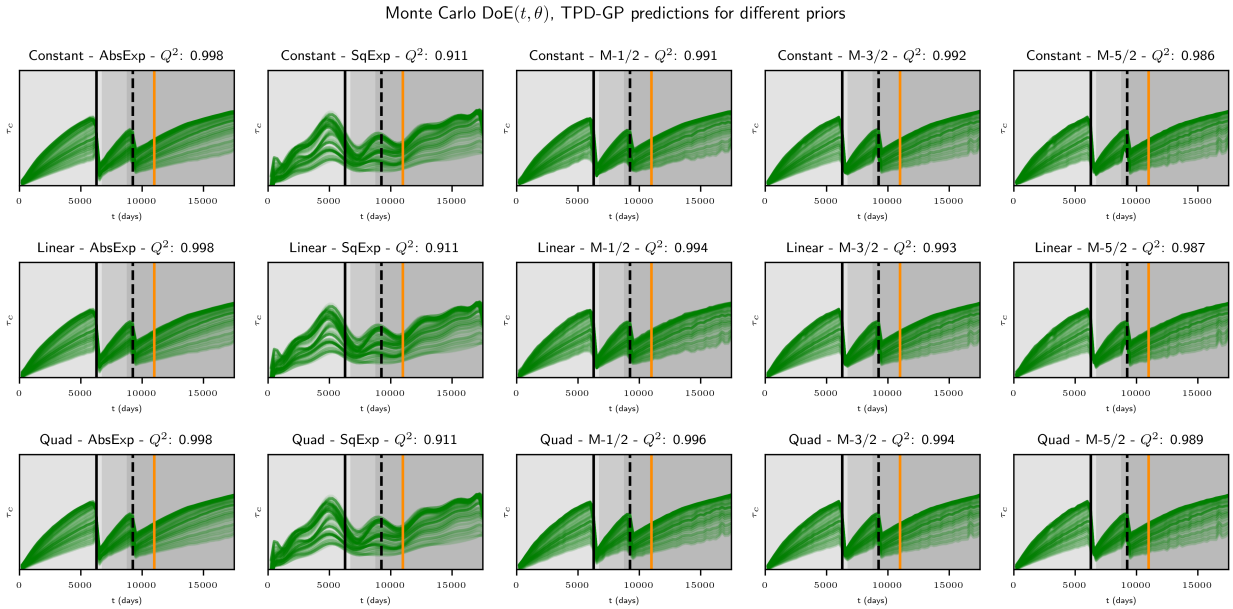

Figure 8.1: GP metamodels of TPD with different priors built with the stochastic-collocation methodology. Input dimension $d = 1$, considering only variable $\theta = a_v$.

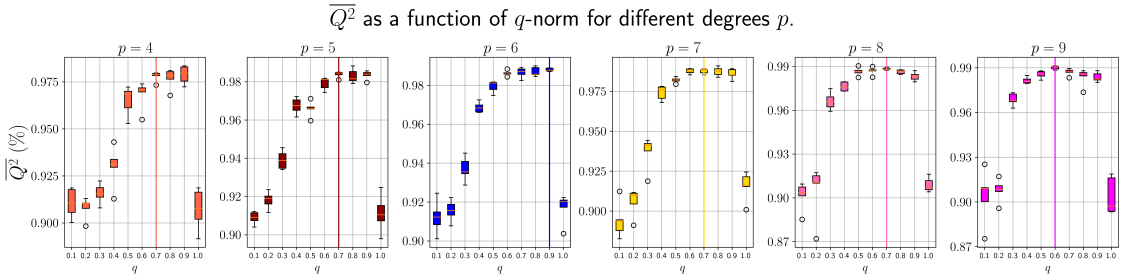

Figure 8.2: Time-averaged predictivity coefficient of PCE expansions of different degrees p and different choices of q-norm.

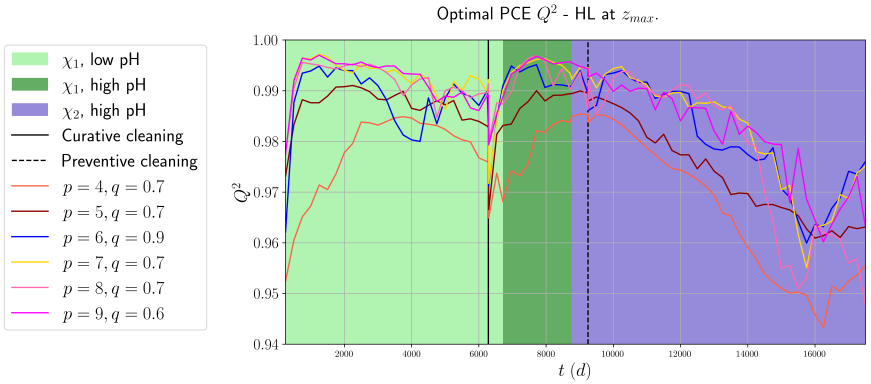

Figure 8.3: Time-variation of the predictivity coefficient for PCEs with (p, q)-hyperparameters maximizing the $\overline{Q^2}$.

Results on SG-6819

In order to assess the robustness of the findings, we perform the uncertainty analysis on a different type of SG, namely the SG-6819 Girard (2014). For this model, we propagate the same uncertainties as the ones defined in Table 3.1 and compute both the Sobol' and the normalized HSIC indices. The results obtained on the SG-6819 exhibit similar patterns and rankings to those obtained on the SG-51B model, indicating robustness of our findings across different SG models.

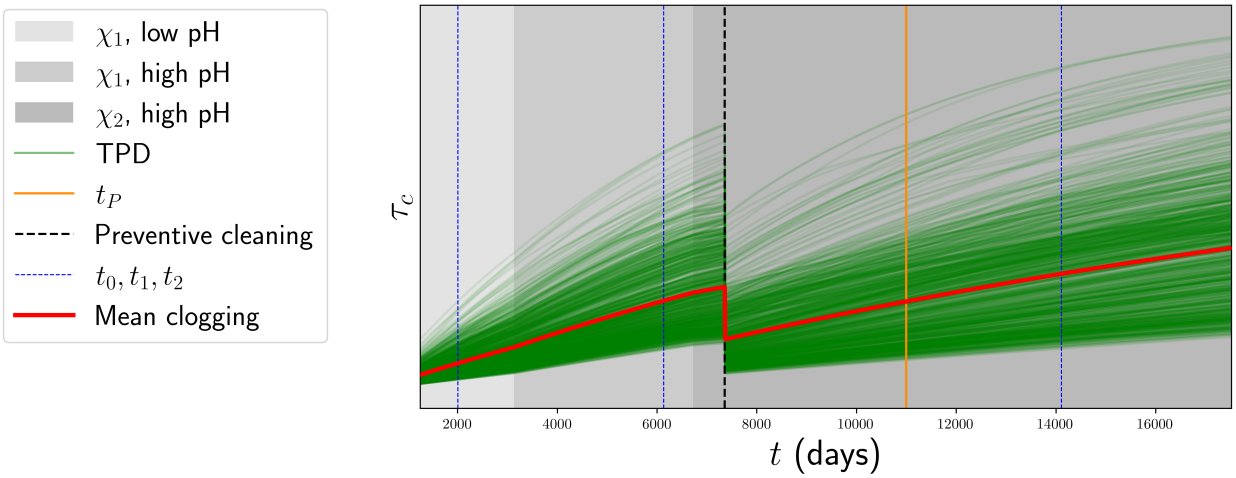

Figure 8.4: Clogging trajectories obtained from Monte Carlo simulations on SG-6819 model, with preventive chemical cleaning.

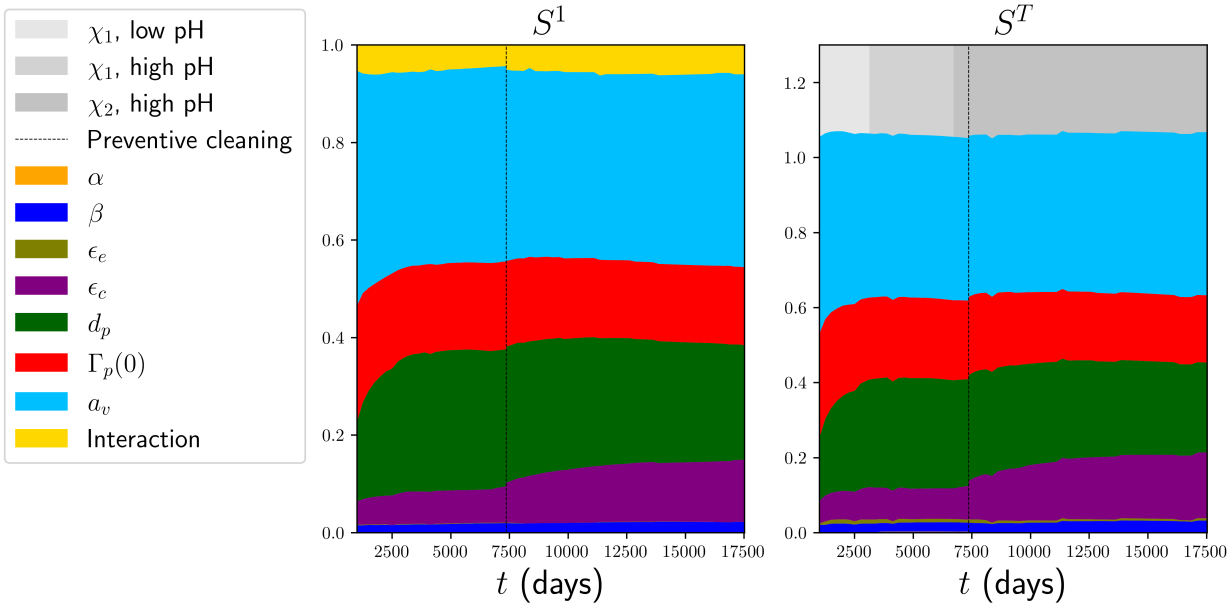

Figure 8.5: Evolution of first-order, interactions (left) and total-order (right) time-dependent Sobol' indices.

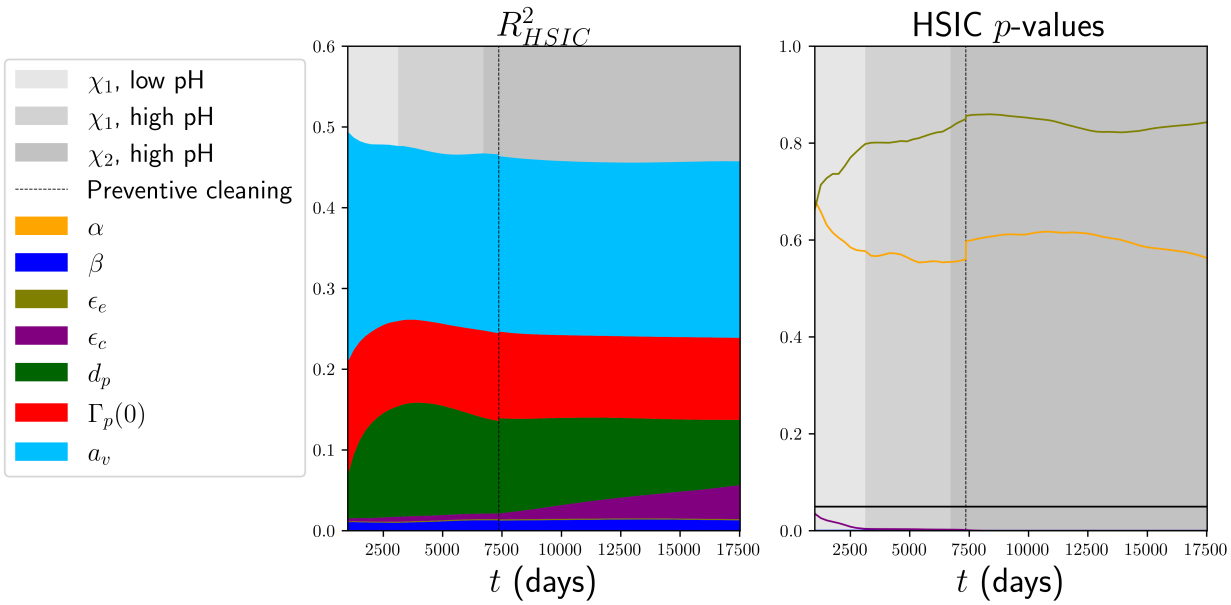

Figure 8.6: Time variation of the estimated normalized HSIC indices and p-values associated to the statistical test of independence. The p-value horizontal line is set at 0.05 and the colors of the lines correspond to the different variables.

Sobol' indices using the Saltelli method

In order to assess the robustness of the Sobol' index estimator, we perform a similar analysis by using the Saltelli method (Saltelli et al., 2008). We sample the points according to the specified design of experiments by using the KLE-GP surrogate model exposed in 3.2.3. The results obtained exhibit a very similar ranking to the one obtained with the VPCE-Sobol' estimation.

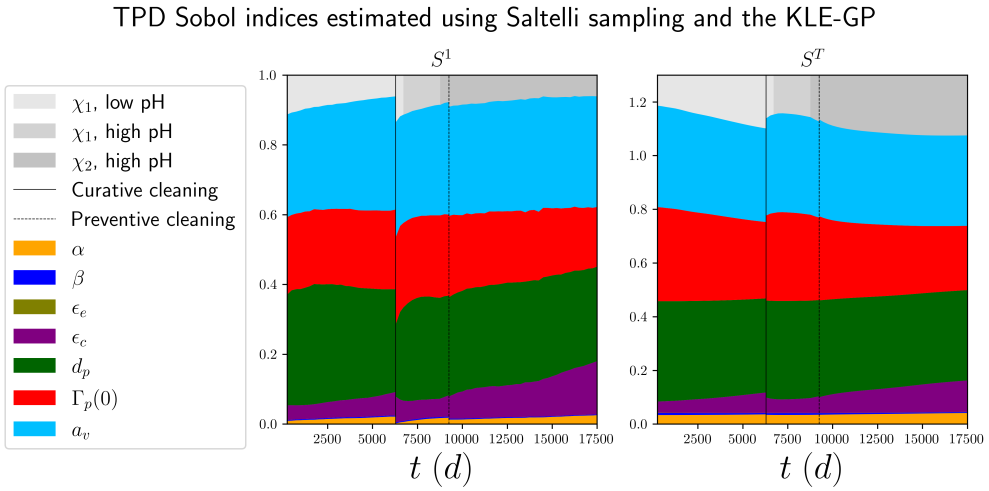

Figure 8.7: Time-varying Sobol' indices estimated using Saltelli sampling and generating the curves using the KLE-GP surrogate model on the SG-51B.

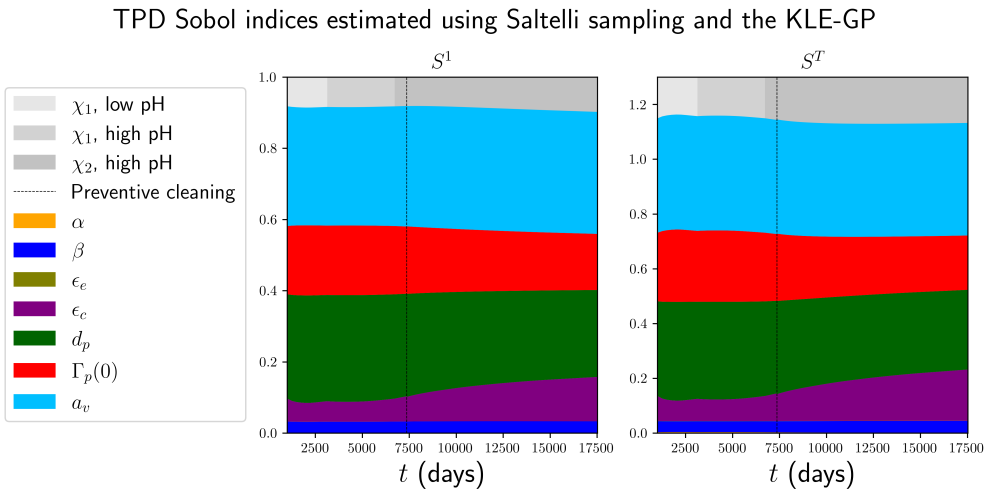

Figure 8.8: Time-varying Sobol' indices estimated using Saltelli sampling and generating the curves using the KLE-GP surrogate model on the SG-6819B.

8.2 Chapter 4

Example of a misspecified GP

To demonstrate the robustness of J+GP prediction intervals under model misspecification, we consider a one-dimensional function with both oscillatory and discontinuous behavior:

$$f(x) = \begin{cases} \sin(x), & \text{if } x > 1 \\ -x, & \text{if } x \leq 1 \end{cases}, \quad x \in [-10, 10]. \tag{8.1}$$

We use 10 training points and 90 test points on a uniform grid of 100 points. A zero-mean GP with a squared-exponential kernel is fitted, and kernel hyperparameters are optimized. Figure 8.9 shows the true function, training points, mean predictor, and prediction intervals for $1 - \alpha = 90\%$: Bayesian credibility intervals (yellow, left), J+ intervals (green, middle), and J+GP intervals (blue, right).

In this example, the GP is intentionally misspecified, resulting in Bayesian credibility intervals that fail to capture the true function for $x \leq 0$. In contrast, the J+GP intervals are more conservative and adaptive, successfully covering more of the true function values. The standard J+ intervals remain nearly constant across the input domain, lacking adaptivity. This illustrates that J+GP intervals provide more robust uncertainty quantification under model misspecification, with interval widths adapting to local data density.

Proof of theorem 4.3.1

We prove a more general version of the theorem. We assume that we are in a regression setting, and we use a model $\widehat{g}$ that has an estimator of its standard deviation $\widehat{\sigma}(X)$. Moreover, we show that a slight modification of the scaled non-conformity score by taking powers $\beta > 0$ of the standard deviation does not change the main results (such powers are used in Papadopoulos (2024) in the

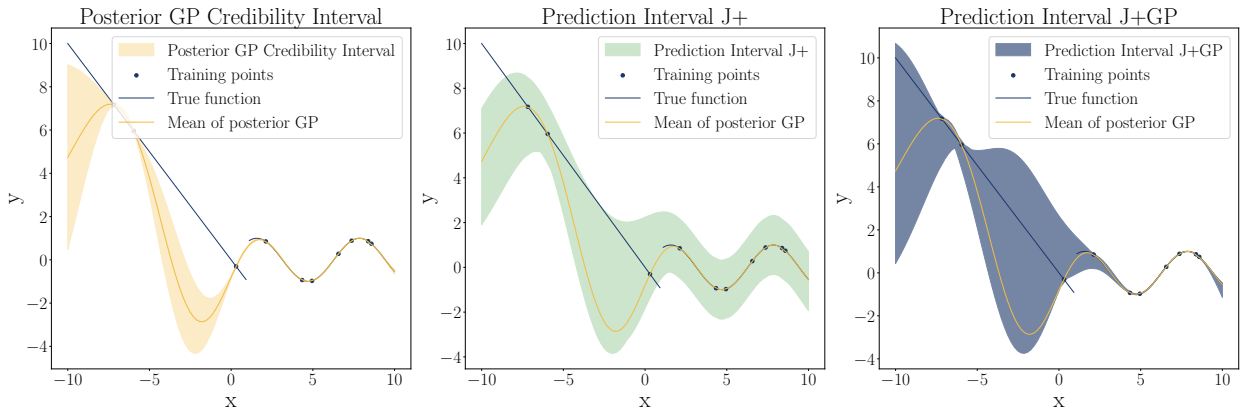

Figure 8.9: Misspecified GP metamodel (squared-exponential kernel) for the illustrative function f. Left: Bayesian credibility intervals ($1-\alpha = 90\%$). Middle: J+ intervals. Right: J+GP intervals.

full-conformal setting). For Gaussian processes, the predictor is the posterior mean $\widehat{g} = \widetilde{g}$ and the estimated standard deviation is the posterior covariance $\widehat{\sigma} = \widetilde{\gamma}$.

Proof. Assume that:

$$Y = g(X) + \epsilon, \tag{8.2}$$

with ϵ representing noise and that a statistical learning model $\widehat{g}$ is trained on the dataset $\mathcal{D}_n = \{(\boldsymbol{X}^{(i)}, Y^{(i)})\}_{i=1}^n$. Let $(\boldsymbol{X}^{(n+1)}, Y^{(n+1)}) \in \mathcal{X} \times \mathcal{Y}$ be a new point. We denote by $\mathcal{D}_{n+1} := \mathcal{D} \cup \{(\boldsymbol{X}^{(n+1)}, Y^{(n+1)})\}$. Let $\widehat{g}_{-(i,j)}$ for all $i \neq j \in \{1, ..., n+1\}$, be the statistical model learned on

$$\mathcal{D}_{n+1} \backslash \{(\boldsymbol{X}^{(i)}, Y^{(i)}), (\boldsymbol{X}^{(j)}, Y^{(j)})\}$$

By exchangeability we have that $\widehat{g}_{-(i,j)} = \widehat{g}_{-(j,i)}$ and $\widehat{g}_{-i} = \widehat{g}_{-(i,n+1)}$. Let us denote by $\widehat{\sigma}(\boldsymbol{X}^{(i)})$ an estimator of the standard deviation of $\widehat{g}$ and assume without loss of generality that $\widehat{\sigma} > 0$ and similarly for the corresponding LOO (we could take the max function with a small $\delta > 0$ otherwise). Similarly, as for the Gaussian non-conformity score, we define:

$$R_i^{\mathrm{LOO}\sigma} = \frac{|Y^{(i)} - \widehat{g}(\boldsymbol{X}^{(i)})|}{\widehat{\sigma}_{-i}^{\,\beta}(\boldsymbol{X}^{(i)})}, \tag{8.3}$$

we then proceed and define $R \in \mathcal{M}_{n+1}(\mathbb{R})$ as:

$$R_{ij} = \begin{cases} +\infty & \text{if } i = j, \\ |Y^{(i)} - \widehat{g}_{-(i,j)}(\boldsymbol{X}^{(i)})| \;/\; \widehat{\sigma}_{-(i,j)}^{\,\beta}(\boldsymbol{X}^{(i)}) & \text{if } i \neq j. \end{cases} \tag{8.4}$$

For simplifying the notations, we will now fix $\beta = 1$. We proceed in defining the matrix $A \in \mathcal{M}_{n+1}(\{0,1\})$:

$$A_{ij} = \mathbf{1}\{R_{ij} > R_{ji}\}. \tag{8.5}$$

It can be easily observed that $A_{ij} = 1 \Leftrightarrow A_{ji} = 0$ except when $j = i$. The strange set associated to A for $\alpha \in (0,1)$ is:

$$\mathcal{S}(A) := \left\{ i \in \{1, \ldots, n+1\} \;:\; \sum_{j=1, j\neq i}^{n+1} A_{ij} \geq (1-\alpha)(n+1) \right\}, \tag{8.6}$$

in other words, a point i is *strange* if the residual R_{ij} compared with R_{ji} for all $j \neq i$ is larger for a given fraction of comparisons.

We start by bounding the cardinal of $\mathcal{S}(A)$. Let i be a strange point. $A_{ij} = 0$ for *at most* $\alpha(n+1) - 1$ other strange points j since $A_{ij} = 1$ for at least $(1-\alpha)(n+1)$ and $i \neq j$. Let $s = |\mathcal{S}(A)|$, we now group pairs of strange points by $A_{ij} = 0$. For a chosen point i, there are at most s possibilities for the strange point j and for each one, $A_{ij} = 0$ at most $\alpha(n+1) - 1$ times. Thus there are at most $s \times (\alpha(n+1) - 1)$ *pairs* of strange points.

We can now bound the number of ways we can choose two points in $\mathcal{S}(A)$ and we obtain:

$$\frac{s(s-1)}{2} \leq s \times (\alpha(n+1) - 1), \tag{8.7}$$

and rearranging:

$$s \leq 2\alpha(n+1). \tag{8.8}$$

By assumption the dataset $\mathcal{D}_{n+1}$ is exchangeable. Thus, using permutation matrices Π that maps a $j \in \{1, \ldots, n+1\}$ to $n+1$ (such that $\Pi_{j,n+1} = 1$), we prove that:

$$\mathbb{P}(n+1 \in \mathcal{S}(A)) = \mathbb{P}(j \in \mathcal{S}(\Pi A \Pi^\top)) = \mathbb{P}(j \in \mathcal{S}(A)). \tag{8.9}$$

Therefore, any point is equally likely to be strange. We have then:

$$\mathbb{P}(n+1 \in \mathcal{S}(A)) = \frac{1}{n+1} \sum_{j=1}^{n+1} \mathbb{P}(j \in \mathcal{S}(A)) = \frac{\mathbb{E}[|\mathcal{S}(A)|]}{n+1} \leq 2\alpha. \tag{8.10}$$

We can now reconnect with the definition of prediction intervals. We denote the generic version of our proposed J+GP prediction interval as:

$$\widehat{C}^*_{n,\alpha}(\boldsymbol{X}^{(n+1)}) = \left[\widehat{q}^{\pm}_{n,\alpha} \left\{\widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) \pm R_i^{\mathrm{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})\right\}\right]. \tag{8.11}$$

Let us suppose that $Y^{(n+1)} \notin \widehat{C}^*_{n,\alpha}$. Then, for at least $(1-\alpha)(n+1)$ values i in $\{1, \ldots, n+1\}$, we have:

$$Y^{(n+1)} > \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) + R_i^{\mathrm{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)}), \tag{8.12}$$

or,

$$Y^{(n+1)} < \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) - R_i^{\mathrm{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)}). \tag{8.13}$$

Finally we can compute:

$$\begin{aligned}
(1-\alpha)(n+1) &\leq \sum_{i=1}^{n+1} \mathbf{1}\left\{Y^{(n+1)} \notin \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) \pm R_i^{\mathrm{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{R_i^{\mathrm{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)}) < |Y^{(n+1)} - \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})|\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{R_i^{\mathrm{LOO}\sigma} < \frac{|Y^{(n+1)} - \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})|}{\widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})}\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\frac{|Y^{(i)} - \widehat{g}_{-i}(\boldsymbol{X}^{(i)})|}{\widehat{\sigma}_{-i}(\boldsymbol{X}^{(i)})} < \frac{|Y^{(n+1)} - \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)})|}{\widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})}\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{R_{i,n+1} < R_{n+1,i}\right\} = \sum_{i=1}^{n+1} A_{n+1,i},
\end{aligned}$$

where the last equality above is obtained with the identities $\widehat{\sigma}_{-i}(\boldsymbol{X}^{(i)}) = \widehat{\sigma}_{-(i,n+1)}(\boldsymbol{X}^{(i)})$ and $\widehat{g}_{-i}(\boldsymbol{X}^{(i)}) = \widehat{g}_{-(i,n+1)}(\boldsymbol{X}^{(i)})$. Therefore $n+1 \in \mathcal{S}(A)$ and:

$$\mathbb{P}\left(g(\boldsymbol{X}^{(n+1)}) \notin \widehat{C}^{*}_{n,\alpha}(\boldsymbol{X}^{(n+1)})\right) \leq \mathbb{P}(n+1 \in \mathcal{S}(A)) \leq 2\alpha. \tag{8.14}$$

□

Proof of theorem 4.3.2

Proof. Assume the same hypothesis as in the previous theorem, and we make use of the same definitions and notations. We define the matrix $\widetilde{R} \in \mathcal{M}_{n+1}(\mathbb{R})$ as:

$$\widetilde{R}_{ij} = \begin{cases} +\infty & \text{if } i = j, \\ R_{ij} \times \widehat{\sigma}_{-(i,j)}(\boldsymbol{X}^{(n+1)}) & \text{if } i \neq j, \end{cases} \tag{8.15}$$

We redefine the matrix $A \in \mathcal{M}_{n+1}(\{0,1\})$:

$$A_{ij} = \mathbf{1}\{\min_{j'} \widetilde{R}_{ij'} \geq \widetilde{R}_{ji}\}, \tag{8.16}$$

where $\min_{j'} \widetilde{R}_{ij'}$ is the smallest residual for the point i when leaving out any point $j' \in \{1, \ldots, n\}$. We start by bounding the number of strange points, choose:

$$i_* \in \operatorname*{arg\,min}_{i \in \mathcal{S}(A)} \left\{ \min_{j'} \widetilde{R}_{ij'} \right\}. \tag{8.17}$$

We can observe that for all strange points $j \in \mathcal{S}(A)$, the matrix element $A_{i_* j}$ is null. Indeed, this comes since by definition:

$$\forall j \in \mathcal{S}(A),\ \widetilde{R}_{ji_*} \geq \min_{j'} \widetilde{R}_{jj'} \geq \min_{j'} \widetilde{R}_{i_* j'}. \tag{8.18}$$

We can then easily bound the number of strange points using that $i_* \in \mathcal{S}(A)$:

$$n + 1 - |\mathcal{S}(A)| \geq \sum_{j=1}^{n+1} A_{i_* j} \geq (1-\alpha)(n+1), \tag{8.19}$$

and a rearrangement gives:

$$|\mathcal{S}(A)| \leq \alpha(n+1). \tag{8.20}$$

Using the exchangeability property in the same fashion as the preceding proof, we have that:

$$\mathbb{P}\left(n+1 \in \mathcal{S}(A)\right) \leq \alpha. \tag{8.21}$$

Let us suppose now that $Y^{(n+1)} \notin \widehat{C}^{*-\text{minmax}}_{n,\alpha}$. Then, for at least $(1-\alpha)(n+1)$ values i in $\{1, \ldots, n+1\}$, we have:

$$Y^{(n+1)} > \max_{i=1,\ldots,n} \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) + R_i^{\text{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)}), \tag{8.22}$$

or,

$$Y^{(n+1)} < \min_{i=1,\dots,n} \widehat{g}_{-i}(\boldsymbol{X}^{(n+1)}) - R_i^{\text{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)}). \tag{8.23}$$

We denote $\widehat{g}_{\min}(\boldsymbol{X}^{(i)}) := \min_{j=1\dots,n} \widehat{g}_{-j}(\boldsymbol{X}^{(i)})$ and similarly for $\widehat{g}_{\max}$ and $\widetilde{R}_i(\boldsymbol{X}^{(n+1)}) := R_i^{\text{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})$. Finally, we can compute:

$$\begin{aligned}
(1-\alpha)(n+1) &\leq \sum_{i=1}^{n+1} \mathbf{1}\left\{Y^{(n+1)} \notin \left[\widehat{g}_{\min}(\boldsymbol{X}^{(n+1)}) - \widetilde{R}_i(\boldsymbol{X}^{(n+1)}), \widehat{g}_{\max}(\boldsymbol{X}^{(n+1)}) + \widetilde{R}_i(\boldsymbol{X}^{(n+1)})\right]\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\min_{j=1,\dots,n} |Y^{(n+1)} - \widehat{g}_{-j}(\boldsymbol{X}^{(n+1)})| \geq R_i^{\text{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\min_{j=1,\dots,n} \frac{|Y^{(n+1)} - \widehat{g}_{-j}(\boldsymbol{X}^{(n+1)})|}{\widehat{\sigma}_{-j}(\boldsymbol{X}^{(n+1)})} \times \widehat{\sigma}_{-j}(\boldsymbol{X}^{(n+1)}) \geq R_i^{\text{LOO}\sigma} \times \widehat{\sigma}_{-i}(\boldsymbol{X}^{(n+1)})\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\min_{j=1,\dots,n} \frac{|Y^{(n+1)} - \widehat{g}_{-(n+1,j)}(\boldsymbol{X}^{(n+1)})|}{\widehat{\sigma}_{-(n+1,j)}(\boldsymbol{X}^{(n+1)})} \times \widehat{\sigma}_{-(n+1,j)}(\boldsymbol{X}^{(n+1)}) \geq \right. \\
&\qquad \left. \frac{|Y^{(i)} - \widehat{g}_{-(i,n+1)}(\boldsymbol{X}^{(i)})|}{\widehat{\sigma}_{-(i,n+1)}(\boldsymbol{X}^{(i)})} \times \widehat{\sigma}_{-(i,n+1)}(\boldsymbol{X}^{(n+1)})\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\min_{j=1,\dots,n} R_{n+1,j} \times \widehat{\sigma}_{-(n+1,j)}(\boldsymbol{X}^{(n+1)}) \geq R_{i,n+1} \times \widehat{\sigma}_{-(i,n+1)}(\boldsymbol{X}^{(n+1)})\right\} \\
&= \sum_{i=1}^{n+1} \mathbf{1}\left\{\min_{j=1,\dots,n} \widetilde{R}_{n+1,j} \geq \widetilde{R}_{i,n+1}\right\} \\
&= \sum_{i=1}^{n+1} A_{n+1,i}.
\end{aligned}$$

Therefore $n+1 \in \mathcal{S}(A)$ and we conclude as in the preceding theorem. □

8.3 Chapter 5

Proof of proposition 5.2.1

Proof. By Bayes' theorem:

$$p(\sigma_{\boldsymbol{\eta}}^2|\theta, \boldsymbol{y}) = \frac{p(\theta, \sigma_{\boldsymbol{\eta}}^2|\boldsymbol{y})}{p(\theta|\boldsymbol{y})} \Leftrightarrow p(\theta|\boldsymbol{y}) = \frac{p(\theta, \sigma_{\boldsymbol{\eta}}^2|\boldsymbol{y})}{p(\sigma_{\boldsymbol{\eta}}^2|\theta, \boldsymbol{y})}. \tag{8.24}$$

Using equation (5.12) and applying the assumptions, we obtain:

$$p(\theta, \sigma_{\boldsymbol{\eta}}^2|\boldsymbol{y}) \propto p(\boldsymbol{y}|\theta)\lambda^{-1} = \lambda^{m/2-1} \exp\left(-\frac{\lambda}{2}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta)\|^2\right). \tag{8.25}$$

Consequently:

$$p(\theta|\boldsymbol{y}) \propto \frac{\lambda^{-1}p(\boldsymbol{y}|\theta)}{p(\lambda|\theta, \boldsymbol{y})} \propto \frac{\lambda^{m/2-1} \exp\left(-\frac{\lambda}{2}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta)\|^2\right)}{\lambda^{m/2-1}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta)\|^{2m/2} \exp\left(-\frac{\lambda}{2}\|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta)\|^2\right)} = \|\boldsymbol{y} - \boldsymbol{\mathcal{G}}(\theta)\|^{-m}. \tag{8.26}$$

Assuming the heterogeneous groups of data, we get by Bayes' theorem:

$$p(\theta|\boldsymbol{y}^1,\ldots,\boldsymbol{y}^q|\boldsymbol{y}^1,\ldots,\boldsymbol{y}^q) = \frac{p(\theta,\lambda_1,\ldots,\lambda_q)}{p(\lambda_1,\ldots,\lambda_q|\theta,\boldsymbol{y}^1,\ldots,\boldsymbol{y}^q)} \tag{8.27}$$

By writing the densities, we get

$$\begin{aligned} p(\theta|\boldsymbol{y}^1,\ldots,\boldsymbol{y}^q) &\propto \frac{\prod_{i=1}^q \lambda_i^{m_i/2-1}\exp\left(-\frac{\lambda_i}{2}\|\boldsymbol{y}^i-\boldsymbol{\mathcal{G}}_i(\theta)\|^2\right)}{\prod_{i=1}^q \lambda_i^{m_i/2-1}\|\boldsymbol{y}^i-\boldsymbol{\mathcal{G}}_i(\theta)\|^{m_i}\exp\left(-\frac{\lambda_i}{2}\|\boldsymbol{y}^i-\boldsymbol{\mathcal{G}}_i(\theta)\|^2\right)} \\ &\propto \prod_{i=1}^q \|\boldsymbol{y}^i-\boldsymbol{\mathcal{G}}_i(\theta)\|^{-m_i}. \end{aligned} \tag{8.28}$$

□

Crack propagation example: the Paris-Erdogan law

Fracture propagation models are essential in predicting the growth of cracks in materials under stress. One of the widely used models for this purpose is Paris-Erdogan's empirical law (Paris and Erdogan, 1960) which describes the rate of crack growth per cycle of loading in terms of the stress intensity factor range. The model is particularly useful in the field of fatigue analysis, where it helps in estimating the RUL of components subjected to cyclic loading. By integrating the crack growth rate equation over the number of loading cycles, the total number of cycles to failure can be estimated. This information is crucial for maintenance planning and ensuring the safety and reliability of structures subjected to cyclic loading. This model is widely used as a prognostics benchmark in the field of fracture mechanics to predict the fatigue life of components as well as for Bayesian model updating (Zárate et al., 2012; Peng et al., 2015).

Paris-Erdogan's law is an phenomenological relationship that relates the crack growth rate to the range of the stress intensity factor, ΔK. The law is expressed as:

$$\frac{da}{dN} = C(\Delta K)^m, \tag{8.29}$$

where a is the crack length, N is the number of loading cycles, C and m are material constants that need to be determined experimentally, ΔK is the stress intensity factor range, defined as $\Delta K = K_{\text{max}} - K_{\text{min}}$, where K_{max} and K_{min} are the maximum and minimum stress intensity factors during a loading cycle. Exponent m characterizes the material's resistance to crack growth, and coefficient C represents the crack growth rate under a given stress state (Bourinet, 2017).

Stress intensity factor

The stress intensity factor, K, is a measure of the stress state near the tip of a crack and is influenced by the applied load, crack size, and geometry of the component. For a mode I (opening mode) crack, the stress intensity factor is given by:

$$K = \sigma\sqrt{pa}Y, \tag{8.30}$$

where σ is the applied stress, a is the crack length and Y is a dimensionless geometry factor that depends on the shape and size of the component and the crack. Therefore, the stress intensity factor range, ΔK, can be expressed as:

$$\Delta K = (\sigma_M - \sigma_m)\sqrt{pa}Y. \tag{8.31}$$

Input probabilistic modeling

The uncertain input variables in the Paris-Erdogan's law fracture propagation model include the material constants C and m, which characterize the material's resistance to crack growth under cyclic loading. The initial crack length $a(0)$ is tracked to predict the component's remaining life. The applied stresses σ_M and σ_m influence the stress intensity factor and the crack growth rate. Lastly, the geometry factor Y accounts for the component's shape and its effect on the stress intensity factor. Uniform distributions with independent marginals are chosen here for the sake of illustration of the methodology, even though other distributions are chosen in practice for this model and work has established correlation between C and m (see (Bourinet, 2017)).

Input Variable	Nominal value	Distribution
C	$C_0 = 1 \times 10^{-10}$	$\mathcal{U}[0.9 \times C_0, 1.1 \times C_0]$
m	$m_0 = 3$	$\mathcal{U}[m_0 - 0.1m_0 + 0.1]$
σ_M	$\sigma_{M,0} = 100$ (MPa)	$\mathcal{U}[\sigma_{M,0} - 5.0, \sigma_{M,0} + 5.0]$
σ_m	$\sigma_{m,0} = 10$ (MPa)	$\mathcal{U}[\sigma_{m,0} - 1.0, \sigma_{m,0} + 1.0]$
Y	$Y_0 = 1.1$	$\mathcal{U}[Y_0 - 0.01, Y_0 + 0.01]$
$a(0)$	$a_0 = 1 \times 10^{-3}$ (m)	$\mathcal{U}[a_0 - 1 \times 10^{-4}, a_0 + 1 \times 10^{-4}]$

Table 8.1: Distributions of the Paris-Erdogan's law input variables.

Numerical results

In order to solve the crack size ordinary differential equation, an explicit Euler scheme of the Paris-Erdogan's law model is implemented, and set a RUL threshold at $D = 5$ cm. For illustration purposes, a series of $q = 4$ different data groups are generated by perturbing a nominal degradation curve with different values of Gaussian noise. The data is also generated at different time instances. As shown in Figure 8.10, the $n = 10^3$ Monte Carlo sample of trajectories from the Paris-Erdogan's law model show a wide range of possible outcomes of the RUL due to the uncertainty in the input variables. We start by interpolating linearly the resulting trajectories on the time instances of the data. Afterwards, we run the data fusion methodology presented to update the prior distributions of the input variables. Since the unitary call to the Euler scheme is not costly, there is no need for a metamodel here and we thus omit the integration on additional hyperparameters on $\mathbf{\Delta}^{p-1}$. The prior and posterior distributions of the Paris-Erdogan's law input variables are shown in Figure 8.11. The posterior distributions of parameters C, m, σ_M are modified and more centered around a mean value compared to the prior uniform distributions that are homogeneous on the whole support, indicating that the data assimilation process has reduced the uncertainty in the input variables. However, the data does not inform variables σ_m, Y and $a(0)$. We can see that the process has also reduced the uncertainty on the RUL, as the posterior distribution is better concentrated around a

mean value. Therefore, updated distributions can be used to make more accurate predictions of the RUL of the component based on the observed data.

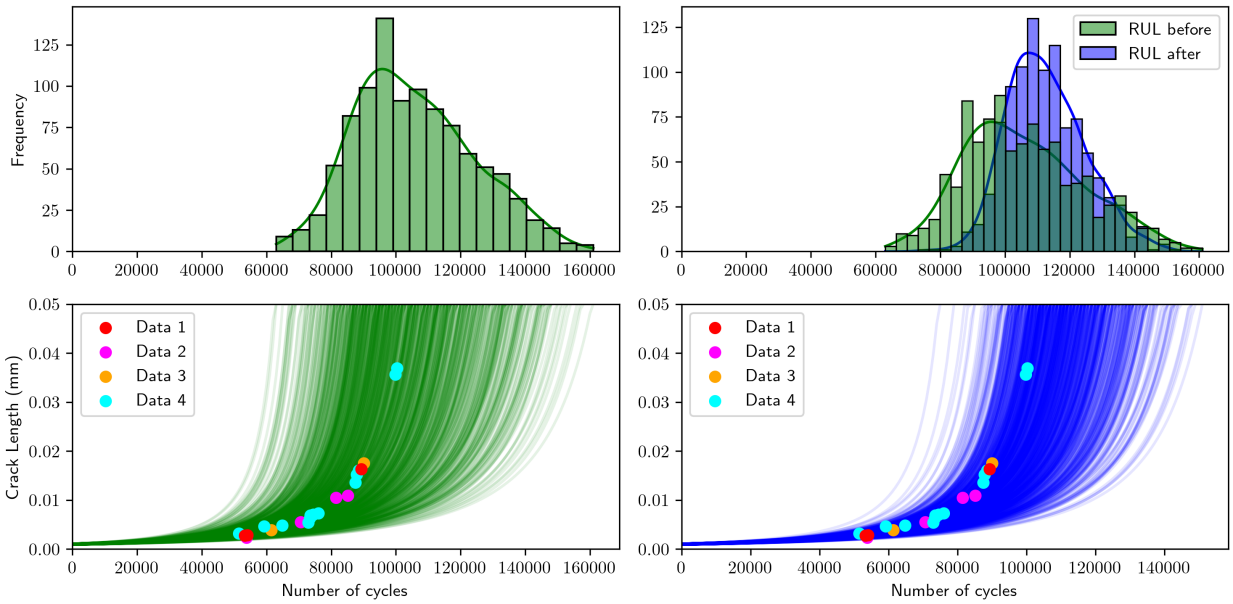

Figure 8.10: Monte Carlo sample of trajectories from the Paris-Erdogan's law before and after heterogeneous data fusion.

Results on SG-6819

We run the Bayesian fusion methodology on the SG-6819 to showcase the genericity of the proposed approach.

Uncertainty integration using VAEs

Since not all uncertain parameters are selected for calibration in the BMU framework in Figure.5.2, the proposed methodology allows for sampling from the conditional distribution $\log p(\theta|\boldsymbol{y}, \boldsymbol{U} = \boldsymbol{u}_0)$. In this setting, the uncertainty in the latent variables $\boldsymbol{U} = (U_1, \ldots, U_{d-1})$ must be accounted for to improve predictive robustness. A natural idea is to integrate over the latent variables, using:

$$\log \mathbb{E}_{\boldsymbol{U}}\, p(\theta|\boldsymbol{y}, \boldsymbol{U}) = \log p(\theta|\boldsymbol{y}). \tag{8.32}$$

However, the methodology does not allow us to compute this quantity directly, since we only have access to the analytical expression of $\log p(\theta|\boldsymbol{y}, \boldsymbol{U})$ in Eq. (5.18). As a result, we can only compute a lower bound using Jensen's inequality:

$$\mathbb{E}_{\boldsymbol{U}}\left[\log p(\theta|\boldsymbol{y}, \boldsymbol{U})\right] \leq \log p(\theta|\boldsymbol{y}). \tag{8.33}$$

To address this limitation, we propose a methodology inspired by variational autoencoders (VAEs).

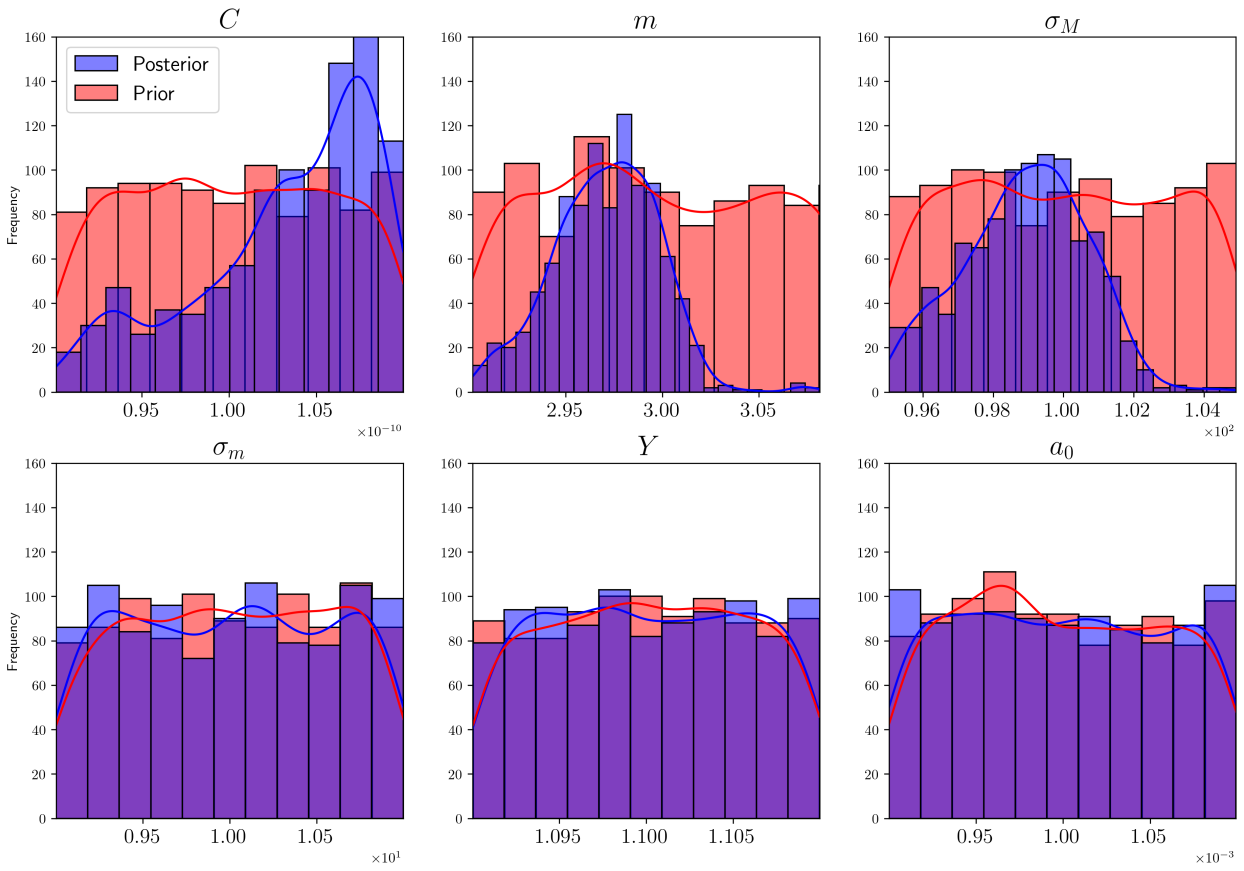

Figure 8.11: Prior and posterior distributions of the Paris-Erdogan's law input variables.

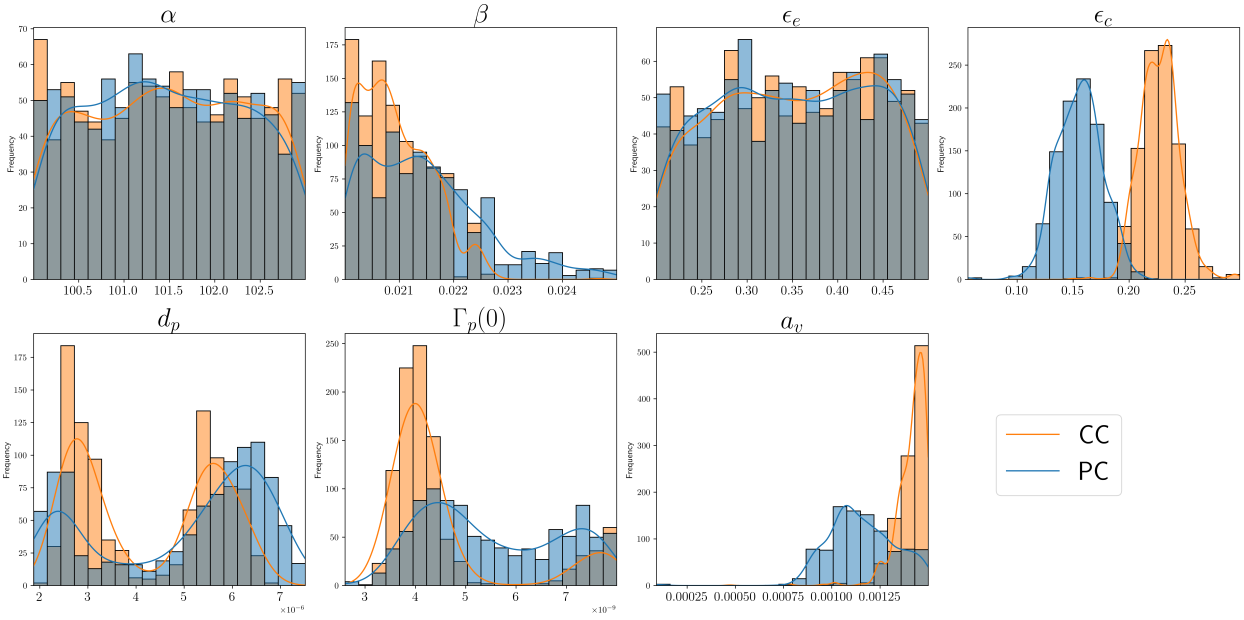

Figure 8.12: Posterior distributions of the input variables of TPD on SG 6819.

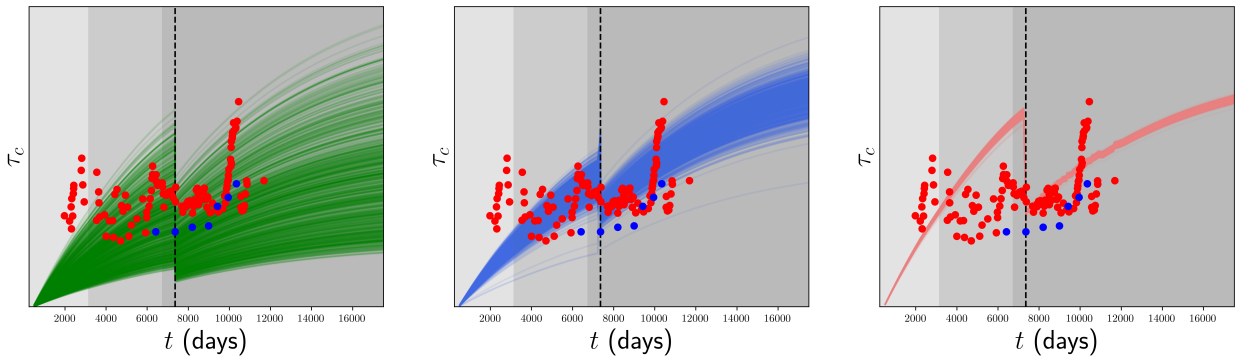

Figure 8.13: Prior and posterior trajectories of KLE-TPD emulator on SG 6818.

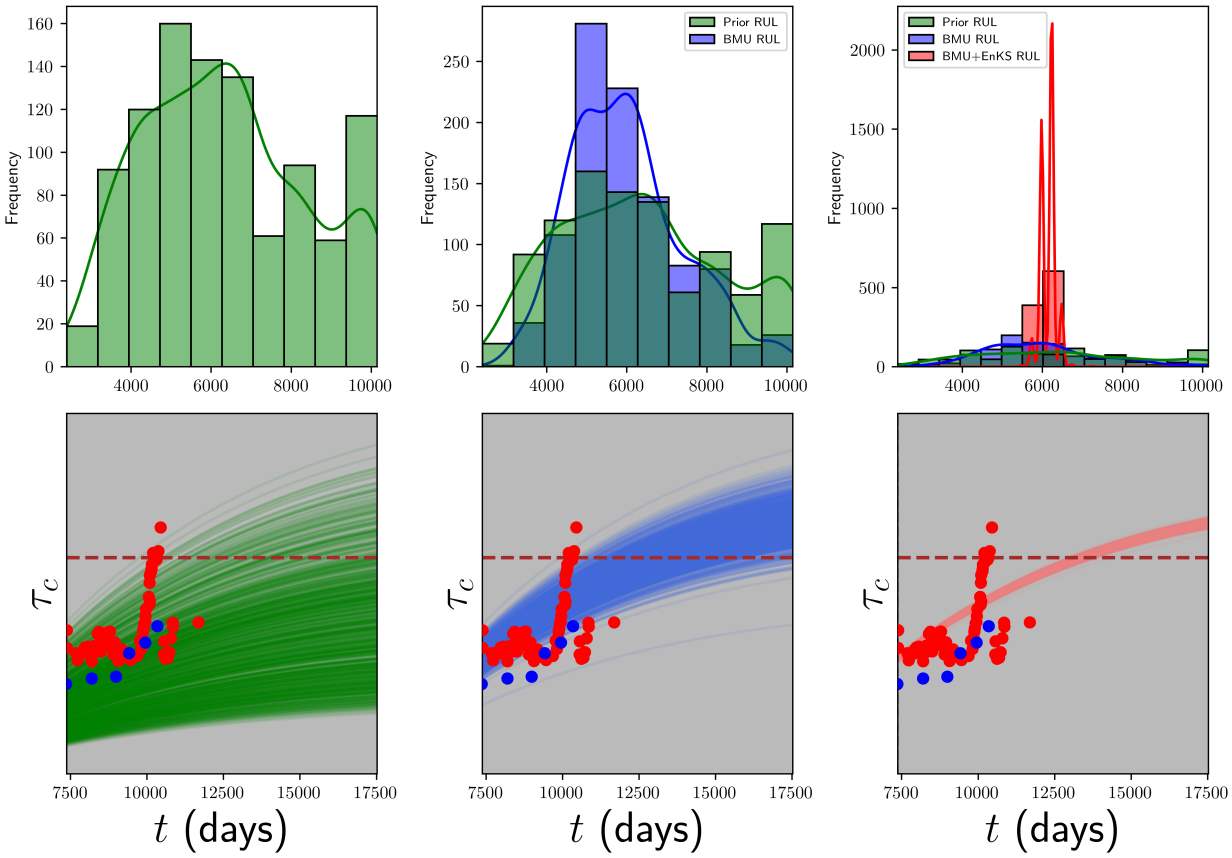

Figure 8.14: Prior and posterior RUL distributions before and after assimilation following the last preventive cleaning on SG 6819.

Evidence Lower Bound (ELBO)

In our context, the latent variables $\boldsymbol{U}$ are known and of fixed dimension $d-1$, and they follow a uniform prior distribution, typically imposed by the choice of the DoE. Let $p(\boldsymbol{U}|\theta, \boldsymbol{y})$ denote the posterior distribution of the latent variables, and let $q_\phi(\boldsymbol{U}|\theta, \boldsymbol{y})$ be a parametric approximation of

this posterior Then, we can write the following identity:

$$\log p(\theta|\boldsymbol{y}) = \mathbb{E}_{q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y})}\left[\log \frac{p(\theta,\boldsymbol{U}|\boldsymbol{y})}{q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y})}\right] + d_{\mathrm{KL}}\left(q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y}) \,\|\, p(\boldsymbol{U}|\theta,\boldsymbol{y})\right), \tag{8.34}$$

where d_{KL} denotes the Kullback-Leibler divergence. The conditioning identity $p((X|Y)|Z) = p(X|Y,Z)$ allows us to write:

$$p(\theta,\boldsymbol{U}|\boldsymbol{y}) = p(\theta|\boldsymbol{y},\boldsymbol{U})\, p(\boldsymbol{U}). \tag{8.35}$$

The Evidence Lower Bound (ELBO) is then defined as:

$$\mathrm{ELBO}_\phi(\theta|\boldsymbol{y}) = \mathbb{E}_{q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y})}\left[\log p(\theta|\boldsymbol{y},\boldsymbol{U})\right] - d_{\mathrm{KL}}\left(q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y}) \,\|\, p(\boldsymbol{U})\right). \tag{8.36}$$

Variational Autoencoders (VAEs)

By expanding the previous expressions, we recover the standard identity:

$$\log p(\theta|\boldsymbol{y}) = \mathrm{ELBO}_\phi(\theta|\boldsymbol{y}) + d_{\mathrm{KL}}\left(q_\phi(\boldsymbol{U}|\theta,\boldsymbol{y}) \,\|\, p(\boldsymbol{U}|\theta,\boldsymbol{y})\right). \tag{8.37}$$

Thus, maximizing the ELBO corresponds to minimizing the Kullback-Leibler divergence between the approximation and the true posterior. When we learn a sufficiently accurate approximation $q_{\phi^*}(\boldsymbol{U}|\theta,\boldsymbol{y})$ of the latent variable posterior, the KL divergence becomes negligible, and we obtain:

$$\log p(\theta|\boldsymbol{y}) \approx \mathrm{ELBO}_{\phi^*}(\theta|\boldsymbol{y}). \tag{8.38}$$

In traditional variational autoencoders, one learns both an *encoder* $q_{\phi^*}(\boldsymbol{U}|\theta,\boldsymbol{y})$ that maps inputs to the latent space, and a *decoder* $p_\psi(\theta|\boldsymbol{y},\boldsymbol{U})$ that reconstructs the input from the latent representation. In our case, however, the decoder is already given explicitly from the Bayesian model. This means we can sample from $p(\theta|\boldsymbol{y},\boldsymbol{U})$ directly using MCMC, together with a pre-trained surrogate model.

List of Figures

List of Tables

Contexte général

Les centrales nucléaires (CN) sont un type de centrale électrique qui transforme la chaleur issue des réactions de fission nucléaire en courant électrique. Elles apparaissent comme des atouts technologiques cruciaux dans la transition énergétique à venir, compte tenu de leurs très faibles émissions de carbone par kWh produit. La France tire jusqu'à 80% de son mélange électrique journalier[1] de l'énergie nucléaire et, avec son opérateur EDF, elle est l'un des leaders mondiaux dans le domaine. Parmi les différentes technologies de centrales, la conception principale du parc nucléaire français est celle des réacteurs à eau sous pression (REP), l'ensemble du parc comprenant 57 réacteurs. Il s'agit de générations technologiques différentes avec des puissances nominales distinctes (notamment 900, 1300 et 1450 MWe). La dernière technologie déployée est le Réacteur Européen à Eau Sous Pression (EPR) qui a récemment démarré ses opérations en France sur le site de Flamanville, avec une puissance future allant jusqu'à 1650 MWe. Un REP est l'assemblage de trois circuits hydrauliques avec des échangeurs de chaleur les reliant, comme illustré à la Figure 8.15 ci-dessous. Le circuit primaire transporte l'eau chauffée à travers des tubes passant dans le premier échangeur de chaleur, à savoir le générateur de vapeur (GV). Chaque réacteur nucléaire possède 3 ou 4 GV, selon la conception du réacteur. À l'intérieur des GV, l'eau froide du circuit secondaire est chauffée et évaporée par transfert de chaleur depuis les tubes primaires. La vapeur sortante est transformée en énergie mécanique pour une turbine qui produit du courant électrique à l'aide d'un alternateur. Enfin, le circuit de refroidissement permet de condenser la vapeur d'eau en phase liquide, qui retourne ensuite dans le circuit secondaire.

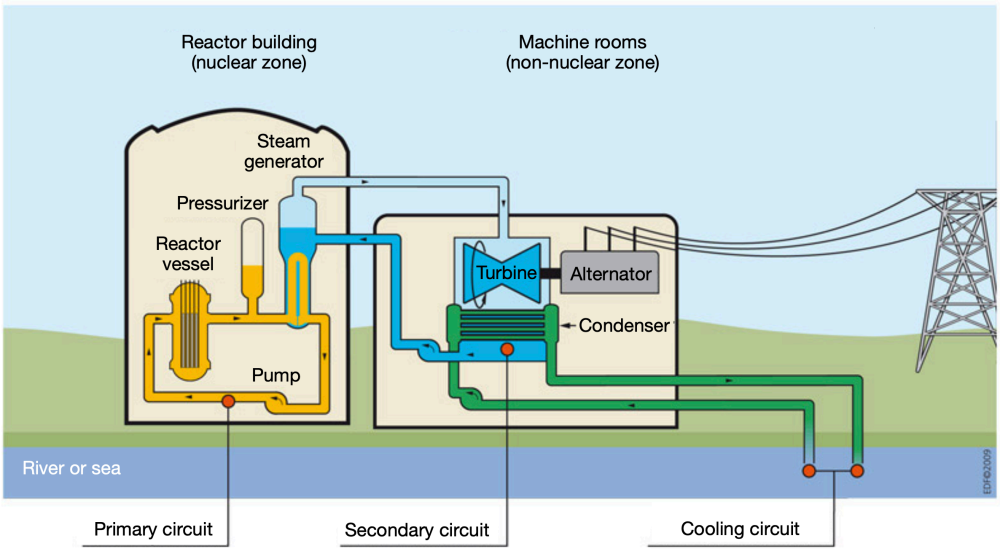

Figure 8.15: Schéma d'un REP (© EDF, ASNR).

Le colmatage des GV dans les REP est un phénomène de dégradation affectant certaines unités du parc nucléaire français (principalement les unités de 900 et 1300 MWe). Ce processus de dégradation complexe résulte de divers facteurs, parmi lesquels les plus importants sont l'érosion-corrosion des tubes du circuit secondaire et la pollution accidentelle par l'entrée d'eau brute dans le circuit de refroidissement (Prusek, 2012). Le colmatage se caractérise par un taux de colmatage τ_c au niveau

[1] voir le site RTE-eco2mix

Figure 8.16: Livraison de GV et exemple de remplacement de générateurs de vapeur 900MW sur la centrale de Cruas (© Mitsubishi Heavy Industries, Bouygues Travaux Publics)

des orifices des plaques entretoises (PE). La plaque de support supérieure est généralement la seule plaque accessible aux inspections par caméra, et il est empiriquement connu qu'elle présente le taux de colmatage le plus élevé (Prusek, 2012). Ce processus de dégradation est difficile à surveiller et à diagnostiquer. En effet, il y a peu de retours du système et l'acquisition de données n'a lieu que durant les arrêts d'exploitation. Des méthodes d'acquisition indirectes basées sur la régression à partir de mesures non-intrusives telles que les courants de Foucault ou des indicateurs transitoires sont déployées (McNab, 1988; Girard, 2014). De plus, ce phénomène présente une cinétique d'initiation très lente, et en raison des conditions opérationnelles spécifiques à chaque REP, des tests expérimentaux en laboratoire sont très difficiles. Néanmoins, des études ont été conduites par EDF-CEA, avec un exemple notable étant la boucle expérimentale COLENTEC (voir Loubet et al. (2020)). Le colmatage s'avère difficile à modéliser en raison des interactions multiphysiques imbriquées, impliquant notamment les écoulements de fluide, le transfert de chaleur, la chimie du circuit secondaire, les conditions d'évaporation thermo-hydrauliques et la géométrie complexe des composants. Bien qu'il ne constitue pas une menace majeure pour la sûreté, le colmatage a divers impacts sur les GV – en particulier, il peut provoquer une redistribution localisée des écoulements entre les orifices des plaques entretoises, ce qui augmente le risque de vibration, de rupture des tubes du GV, et entrave la réponse du GV aux changements opérationnels. Pour faire face à ces défis, des opérations de nettoyage chimique préventif peuvent être réalisées sur les GV, et ces interventions doivent être planifiées de manière optimale. Elles consistent à arrêter la CN et à introduire des solvants de nettoyage dans le GV qui désagrègent et éliminent une partie des dépôts de colmatage, puis sont évacués via la purge du GV (voir Figure 2.1).

Ce problème est au cœur du cadre de la Prognostics and Health Management (PHM) (Vachtsevanos et al., 2006), dont l'objectif est d'orienter les actions de maintenance préventive afin d'assurer des performances fiables du système. Nous notons par Δ un indice de dégradation associé à un système donné au sein d'une flotte de composants. Cet indice est généralement représenté comme une

fonction du temps, $(t \mapsto \Delta(t))$, et peut être scalaire ou vectoriel lorsque la dégradation est évaluée à plusieurs emplacements au sein du composant. Pour un seuil de dégradation prescrit Δ_*, la *durée de vie résiduelle* (RUL) (Vachtsevanos et al., 2006) est définie comme

$$\mathrm{RUL}(t_P, \Delta_*) = \arg\min_{t>t_P}\{\Delta(t) \geq \Delta_*\}, \tag{8.39}$$

où t_P désigne le temps d'observation courant. En pratique, l'évolution de $\Delta(t)$ est influencée par une combinaison de phénomènes déterministes tels que les lois physiques, et de phénomènes stochastiques provenant de la variabilité inhérente des propriétés des matériaux, des sollicitations et des conditions environnementales. Par conséquent, dans toute sa généralité, le processus de dégradation doit être traité de manière probabiliste, et la RUL devient une *variable aléatoire* correspondant au *temps d'arrêt* (first passage time) auquel le processus stochastique $\Delta(t)$ atteint le seuil Δ_*. Le pronostic consiste donc à estimer la distribution de probabilité de ce temps d'arrêt à l'aide d'un modèle de dégradation approprié afin d'avoir une stratégie décisionnelle informée par le risque pour la planification de la maintenance. La façon dont ces incertitudes sont traitées en pratique sera discutée dans les chapitres à vallenir et dépend de la typologie du modèle de pronostic utilisé. Dans ce travail, deux classes de modèles sont considérées pour le pronostic du colmatage : des modèles basés sur la physique fondés sur des principes premiers et implémentés par simulation numérique, et des modèles statistiques ou basés sur les données, tels que les approches d'apprentissage supervisé. L'objectif de cette thèse est de construire des *modèles hybrides* à partir de ces deux approches, avec la tâche ultime de tirer parti des avantages des deux typologies de modèles afin d'aider les ingénieurs d'EDF pour une planification de maintenance plus robuste et informée par le risque.

Modèles généraux utilisés pour le pronostic

Bref historique de le pronostic

L'étude de le pronostic de la dégradation des systèmes peut être retracée jusqu'à la révolution industrielle qui a débuté à la fin du XVIII^{e} siècle et s'est étendue jusqu'aux années 1850. Durant cette période, la généralisation de la mécanisation de la production a d'abord exposé les coûts économiques et humains des pannes machines. Ainsi, le besoin d'un fonctionnement prévisible, durable et sûr de systèmes tels que les machines à vapeur ou les locomotives a motivé des expérimentations systématiques sur les matériaux et composants mécaniques afin d'éviter de tels accidents. Les recherches pionnières de Wöhler (1858, 1860) ont d'abord mené des expériences sur la fatigue des essieux de chemin de fer qui ont établi les premières lois empiriques de la dégradation mécanique[a] posant ainsi les fondations des essais de fatigue et de résistance. À la fin du XIX^{e} siècle, l'industrialisation avait conduit à la création de laboratoires de métrologie et d'organismes de normalisation, intégrant des considérations empiriques de fiabilité dans la pratique de l'ingénierie. Le XX^{e} siècle a étendu ces idées au domaine probabiliste : Weibull (1951) a développé des modèles statistiques de durée de vie des matériaux, et Barlow and Proschan (1965) a formalisé la théorie mathématique de la fiabilité pour les systèmes complexes. Plus tard, le paradigme physics-of-failure introduit par Pecht (1992) et développé par Zio (2012) a relié ces approches statistiques aux modèles mécanistiques de dégradation,

établissant la base de le pronostic et de la gestion de la santé moderne.

[a]Les courbes S-N obtenues lors des essais de fatigue des matériaux sont aussi appelées courbes de Wöhler.

Modèles basés sur la physique

☛ Bref historique des modèles de pronostic basés sur la physique
La transition de l'ingénierie empirique vers le pronostic basée sur la physique a commencé au milieu du XX[e] siècle, motivée par le désir de lier la dégradation des composants directement à leurs mécanismes physiques sous-jacents. Parmi les études précoces, la mécanique de la rupture et la science des matériaux ont fourni le cadre mathématique pour ce changement : Griffith (1920) a développé la théorie énergétique de la rupture et Paris and Erdogan (1960) a introduit une loi d'évolution de fissure basée sur des équations différentielles ordinaires, permettant la prédiction de la RUL à partir de mesures de contrainte et de longueur de fissure (voir l'annexe 8.3). Ces avancées ont marqué un tournant conceptuel : la défaillance n'était plus un événement purement statistique mais l'issue déterministe de processus physiques couplés tels que la fatigue, la corrosion, l'usure ou la diffusion. Dans les années 1980 et 1990, cette approche mécaniste a évolué vers le pronostic basé sur la physique illustrée par Pecht (1992), qui visait à modéliser la dégradation à l'aide de lois fondamentales de la mécanique, de la chimie et de la science des matériaux.

Une taxonomie utile distingue les modèles physiques de dégradation selon la forme mathématique de leur relation gouvernante : lois empiriques, équations différentielles ordinaires (EDO) et équations aux dérivées partielles (EDP). Par exemple, la loi d'usure classique d'Archard (Delaney and Wang, 2025) exprime la perte de matériau comme une proportionnalité empirique et a été utilisée pour la prédiction de RUL d'engrenages et d'embrayages (voir par exemple (Zhao et al., 2018)). Ensuite, les modèles de croissance de fissures basés sur la loi de Paris-Erdogan (Paris and Erdogan, 1960) sont fondés sur une EDO et soutiennent de nombreuses études de pronostic des roulements et des structures (voir Cubillo et al. (2016)). Troisièmement, le vieillissement limité par diffusion de la couche SEI dans les batteries lithium-ion est souvent modélisé par une formulation continue basée sur des EDP, par exemple la diffusion du solvant à travers la couche et la cinétique de réaction (voir Sankarasubramanian and Krishnamurthy (2012)), et est cruciale pour des pronostics plus fiables de l'état de santé des batteries (voir Li (2025) pour plus sur le pronostic des batteries). Chacun de ces types de modèles offre une fidélité mécanistique croissante mais aussi une complexité croissante. Les lois empiriques ont une fidélité physique limitée et peuvent mal extrapoler ; les modèles EDO captent l'évolution temporelle mais supposent souvent des charges simplifiées ; les modèles EDP peuvent capturer des distributions spatiales fines et des phénomènes de transport mais sont typiquement coûteux en calcul et nécessitent une paramétrisation détaillée et des conditions aux limites, d'où leur usage pratique en pronostic en ligne doit être accompagné d'une quantification d'incertitude. Ensemble, ces références illustrent les compromis des approches purement basées sur la physique : interprétabilité et mécanismes au prix d'une charge de calibration, d'une gestion des incertitudes et d'une tractabilité computationnelle réduites.

Nous supposons que le niveau de dégradation Δ peut être déduit comme une quantité d'intérêt $\Delta = \mathrm{QoI}(u)$ où $u \in \mathbb{R}^N$ est la solution d'une EDP générale impliquant un opérateur différentiel

dépendant d'un paramètre $\mathcal{A}$ de la forme :

$$(E_{\boldsymbol{X}}): \begin{cases} \mathcal{A}_{\boldsymbol{X}}(t,\boldsymbol{z},\boldsymbol{X})u_{\boldsymbol{X}}(t,\boldsymbol{z}) = F_{\boldsymbol{X}}(t,\boldsymbol{z}),\ (t,\boldsymbol{z}) \in [0,T]\times \boldsymbol{D} \\ u_{|\partial \boldsymbol{D}} = h_{\boldsymbol{X}}(t) \end{cases} \tag{8.40}$$

où $\boldsymbol{D}$ est un domaine borné régulier de $\mathbb{R}^M$, $F_{\boldsymbol{X}}$ est un terme source, $h_{\boldsymbol{X}}$ un terme de bord régulier et $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^d$ sont des variables d'entrée du modèle physique. Nous fixons maintenant ces variables à une valeur donnée $\boldsymbol{X} = \boldsymbol{x}_0$ et considérons $(E_{\boldsymbol{x}_0})$. S'il n'existe pas de solution analytique à ce système, des méthodes numériques doivent être utilisées. Parmi les différentes méthodes possibles, des couplages entre schémas d'Euler explicite/implicite, ou des schémas de Runge-Kutta en temps et des éléments finis/différences en espace (Quarteroni and Valli, 2008) peuvent être utilisés sur un maillage adapté du domaine espace-temps $[0,T]\times \boldsymbol{D}$. De plus, ces méthodes fonctionnent si des hypothèses d'ellipticité et de coercivité de l'opérateur différentiel en espace sont satisfaites et si les propriétés de stabilité et de précision en résultent. Il est à noter que le temps final de simulation et la résolution du maillage spatial sont cruciaux pour le temps de calcul de toute méthode numérique utilisée. Pour des problèmes physiques plus complexes impliquant des couplages entre transport, diffusion et réactions chimiques, des approches multiphysiques et multiéchelles peuvent être nécessaires. Ces approches combinent différentes méthodes numériques adaptées à chaque phénomène physique pour résoudre les équations couplées de manière cohérente en temps et en espace (Peksen, 2018). Les différentes solutions sont ensuite utilisées comme paramètres ou conditions initiales pour d'autres équations aval. Il convient de noter que la positivité de la solution et la propriété de bien-posage du système résultant deviennent plus difficiles à garantir. De tels codes de simulation déterministes généraux peuvent être formalisés comme des opérateurs :

$$G : \mathcal{X} \times \mathcal{H} \to \mathbb{R}^N, \tag{8.41}$$

où l'ensemble $\mathcal{H}$ correspond aux paramètres numériques définis par l'utilisateur dont la variabilité provient des différentes entrées du code (comme le choix du pas de discrétisation ou des lois de fermeture). Pour des paramètres fixés $\boldsymbol{h} \in \mathcal{H}$, nous notons $g(.) := G(.,\boldsymbol{h}) \in \mathbb{R}^N$ la sortie du code de simulation ne prenant que les variables du modèle comme entrée et donc $g(\boldsymbol{X})$ est une solution numérique au système EDP paramétrique $(E_{\boldsymbol{X}})$ avec des paramètres numériques fixés $\boldsymbol{h}$. En d'autres termes $g(\boldsymbol{X}) = \texttt{solve}(\boldsymbol{h}, E_{\boldsymbol{X}})$.

Dans le contexte de la quantification d'incertitude (QI), $\boldsymbol{X}$ est traité comme une variable aléatoire, et nous supposons que tous les paramètres bien connus ont été fixés dans le vecteur $\boldsymbol{h}$. L'un des objectifs principaux de la QI dans ce cas est de rendre les solveurs numériques plus robustes et prédictifs en évaluant l'image de la mesure probabiliste modélisant $\boldsymbol{X}$. Cela peut être abordé de deux manières : l'une est *non-intrusive* (utile pour des codes industriels coûteux), et l'autre *intrusive*, permettant des évaluations d'incertitude plus fines. Dans le cas non-intrusif, la procédure implique quelques étapes clés (De Rocquigny et al., 2008) qui sont détaillées dans la Figure 8.17 ci-dessous. Après spécification du modèle numérique et des paramètres incertains à l'étape A, l'étape B consiste à définir un modèle probabiliste pour les entrées, c'est-à-dire déterminer une distribution $p_{\boldsymbol{X}} \in \mathcal{P}(\mathcal{X})$, puis à évaluer la mesure image $g\#p_{\boldsymbol{X}}$ via le code à l'étape C. Cela revient également à obtenir des échantillons de la variété de solutions :

$$\mathcal{M}_{\boldsymbol{X}} = \{g(\boldsymbol{X}),\ \boldsymbol{X} \sim p_{\boldsymbol{X}} \in \mathcal{P}(\mathcal{X})\} \subseteq \mathbb{R}^N. \tag{8.42}$$

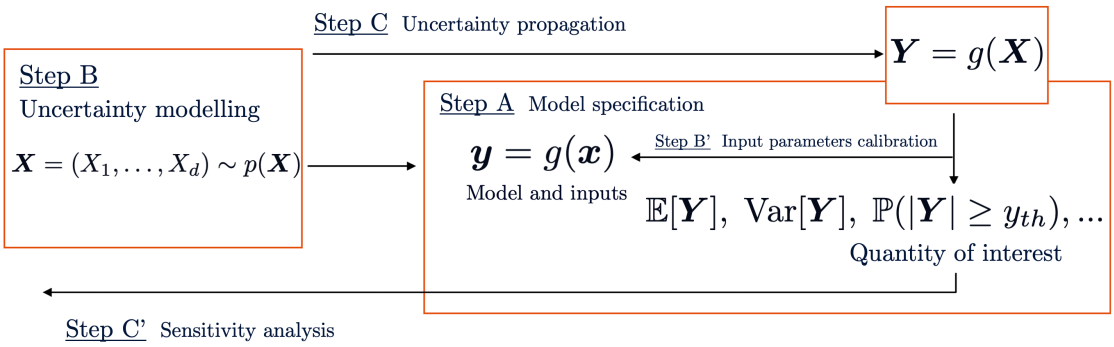

Figure 8.17: Méthodologie générique pour la quantification non-intrusive des incertitudes des modèles numériques pour des applications industrielles, d'après (De Rocquigny et al., 2008).

En utilisant typiquement l'échantillonnage Monte-Carlo (Rubinstein and Kroese, 2008), on obtient un jeu de données $\mathcal{D} = \{(\boldsymbol{X}^{(i)}, g(\boldsymbol{X}^{(i)}))\}_{i=1}^{n}$ à partir de n évaluations du code avec lequel on peut calculer des statistiques ou d'autres quantités d'intérêt. Souvent, en raison de la résolution du maillage et de la durée de simulation, le temps de calcul est significatif. Cependant, pour étudier la distribution de sortie et/ou la sensibilité des composantes d'entrée, un nombre beaucoup plus grand $N \gg n$ d'évaluations est souvent nécessaire au-delà du budget disponible. C'est là que les techniques d'apprentissage supervisé peuvent être employées pour construire des modèles de substitution (aussi appelés « métamodèles ») $\widehat{g}$. Parmi les familles paramétriques utilisées figurent les processus gaussiens (Rasmussen and Williams, 2006), les développements en chaos polynomial (Blatman and Sudret, 2009, 2011), ou des variantes de réseaux de neurones (Lefebvre et al., 2023). Ces approches sont de nature « boîte noire », ce qui signifie que la structure de la sortie n'est pas supposée a priori ; au lieu de cela, des procédures d'apprentissage sont mises en œuvre pour approximer la sortie à partir d'un jeu de données à budget fixe. Selon la connaissance des équations gouvernantes et le comportement qualitatif de la solution, des méthodes d'interpolation spatio-temporelle peuvent aussi être utilisées pour améliorer la résolution. Si la dimension de sortie est trop élevée et fortement corrélée, des méthodes de réduction de dimension telles que la décomposition de Karhunen-Loève (Sullivan, 2015; Sudret and Der Kiureghian, 2002) peuvent être appliquées. Ces méthodes sont plus largement connues sous le nom de méthodes de base réduite (RB) (Quarteroni et al., 2016), et peuvent aussi être réalisées en contexte intrusif. Ces métamodèles peuvent être validés en utilisant des métriques classiques de régression telles que les coefficients de prédictivité ou des outils plus complexes comme la prédiction conforme qui sera présentée dans cette thèse.

Enfin, il existe des méthodes *intrusives* basées sur des techniques RB (Le Maître and Knio, 2010; Quarteroni et al., 2016), ou des méthodes de collocation stochastique (Xiu and Hesthaven, 2005; Xiu, 2010). Ces méthodes consistent à modifier directement les équations gouvernantes ou les solveurs numériques pour incorporer la quantification d'incertitude ou la réduction d'ordre (Luc, 2016). Les méthodes de base réduite (similaires, en un sens, aux techniques de réduction de dimension en ML, voir Hastie et al. (2009)) visent à construire une approximation de faible dimension de l'espace des solutions en identifiant un jeu de fonctions de base représentatives, qui sont ensuite utilisées pour résoudre le problème plus efficacement. Les méthodes de collocation stochastique, quant à elles, discrétisent l'espace stochastique à l'aide de points de collocation et résolvent le problème

déterministe en chaque point, permettant une propagation efficace des incertitudes. Ces approches sont particulièrement utiles pour les problèmes où les équations gouvernantes sont bien comprises et les ressources computationnelles suffisantes pour gérer les modifications intrusives.

👉 Modèle physique pour le pronostic du colmatage

Pour le cas d'utilisation du colmatage des GV, un modèle de simulation a été développé, ainsi qu'un outil de simulation numérique THYC-Puffer-DEPO (initialement proposé par Prusek et al. (2013) et amélioré et maintenu par Feng et al. (2023)) permettant la simulation de l'évolution du taux de colmatage sur des périodes opérationnelles allant jusqu'à 60 ans (voir chapitre 1). Le code est déterministe, produisant des sorties identiques pour des entrées identiques compte tenu de la stratégie numérique utilisée. Dans des travaux antérieurs (Lefebvre et al., 2023), certains paramètres d'entrée du modèle ont été identifiés comme ayant une incertitude épistémique, qui peut être réduite avec une meilleure connaissance, à l'aide de méthodologies DE QI. Grâce à sa capacité d'extrapoler le taux de colmatage sur toute la durée de vie du GV, il peut également fournir une première estimation probabiliste de la RUL. Parce que THYC-Puffer-DEPO est un solveur propriétaire, fermé, avec des temps de calcul importants et sans accès direct aux noyaux de discrétisation ou de solveur internes, nous adoptons une stratégie de QI paramétrique non-intrusive : le code est traité comme une application déterministe en boîte noire faisant correspondre des paramètres d'entrée à des trajectoires complètes de colmatage, et l'incertitude est introduite et quantifiée via des modèles de probabilité sur ces entrées (voir chapitre 2). Ce choix est en outre motivé par le caractère intrinsèquement multiphysique et multiéchelle du modèle. En effet, le solveur couple la thermique-hydraulique, la chimie et le transport/adhésion des particules à travers des échelles spatiales et temporelles disparates (voir chapitre 1). Cela produit des variétés de réponse complexes et des modifications intrusives du schéma numérique (par exemple, les méthodes RB) sont impraticables dans une première approche ; à la place, nous réalisons une analyse de sensibilité paramétrique, un échantillonnage Monte-Carlo et un design d'expériences dans l'espace d'entrée, et construisons des substituts non intrusifs (PG, chaos polynomial, métamodèles de champ basés sur KL) pour approximer l'application g pour des évaluations rapides du code.

Modèles d'apprentissage statistique

👉 Bref historique des modèles statistiques de pronostic

Les modèles statistiques de dégradation ont émergé du besoin de généraliser le comportement de dégradation à travers de grandes populations de composants avec une connaissance mécanistique limitée. S'appuyant sur les fondations probabilistes posées par Weibull (1951) et Barlow and Proschan (1965), ces approches traitaient la RUL comme une variable aléatoire issue de lois de durée de vie ajustées empiriquement. Avec l'apparition des données en ligne via des capteurs numériques et des systèmes de surveillance, le domaine s'est étendu vers des modèles de processus stochastiques tels que les modèles de Markov cachés, les processus gamma et les processus de Wiener pour représenter des trajectoires de dégradation incertaines et estimer la RUL à partir d'observations bruitées. Ces cadres basés sur les données offraient scalabilité

et efficacité computationnelle, mais au prix d'une interprétabilité physique et d'une capacité d'extrapolation limitées hors du domaine d'entraînement.

Les modèles statistiques de pronostic peuvent être largement classés comme *paramétriques* ou *non-paramétriques* selon les hypothèses imposées au processus de dégradation sous-jacent. Les modèles paramétriques postulent une forme fonctionnelle fixe gouvernée par un nombre fini de paramètres $\theta \in \Theta$, qui sont estimés à partir des données ; des exemples typiques incluent les modèles de durée de vie de Weibull ou exponentiel et les formulations de processus stochastiques tels que le processus de Wiener ou le processus Gamma. Par exemple Tayade et al. (2019) utilise un modèle de régression linéaire associé à une analyse en composantes principales pour l'estimation de RUL de roulements, et Dhada et al. (2025) applique un modèle paramétrique hiérarchique de Weibull à plusieurs sous-flottes de systèmes se dégradant de manière similaire pour la prédiction de RUL. En revanche, les modèles non-paramétriques imposent moins d'hypothèses sur la loi de dégradation et cherchent à l'inférer directement à partir des données, comme dans les estimateurs de densité à noyau, les processus gaussiens, les forêts aléatoires ou les réseaux de neurones. Des applications notables de ces modèles non-paramétriques pour différentes tâches de pronostic incluent l'utilisation de réseaux de neurones convolutifs (CNN) (Li et al., 2018) et de réseaux de neurones récurrents à mémoire longue (LSTM) (Shi and Chehade, 2021) pour la prédiction de la RUL de turbomoteurs, des machines à vecteurs de support (García Nieto et al., 2015) et des réseaux de neurones artificiels (Zhao et al., 2017) pour la prédiction de RUL de moteurs d'avion. Les méthodes paramétriques offrent interprétabilité et compacité mais peuvent être biaisées si la structure supposée est mal spécifiée, tandis que les approches non-paramétriques offrent plus de flexibilité au prix d'exigences de données plus importantes et d'une transparence physique réduite.

Un modèle d'apprentissage supervisé (Hastie et al., 2009) pour la régression implique la définition d'un échantillon $\mathcal{D} = \{(\boldsymbol{X}^{(i)}, \boldsymbol{Y}^{(i)})\}_{i=1}^{m}$ avec $\boldsymbol{X} \in \mathcal{X} \subseteq \mathbb{R}^d$ et des cibles i.i.d. $\boldsymbol{Y} \in \mathcal{Y} \subseteq \mathbb{R}^\ell$ telles que $(\boldsymbol{X}, \boldsymbol{Y})$ suit une distribution génératrice inconnue $p_{\boldsymbol{X},\boldsymbol{Y}}$, la division en échantillon d'entraînement et de test $\mathcal{D} = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{test}}$ de tailles respectives n et ℓ, ainsi que le choix d'une famille (paramétrique) de modèles fonctionnels $(f_\theta : \mathbb{R}^d \to \mathbb{R}^\ell)$ pour déterminer un modèle de régression. L'objectif est de minimiser le risque sur l'échantillon d'entraînement par rapport à la mesure empirique $\widehat{p}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{(\boldsymbol{X}^{(i)},\boldsymbol{Y}^{(i)})}$ - où $\delta_{(a,b)}$ ici est la distribution de Dirac sur $\mathbb{R}^d \times \mathbb{R}^\ell$ - et une fonction de perte $\mathcal{L} : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_+$ pour trouver les paramètres optimaux $\widehat{\theta} \in \Theta \subseteq \mathbb{R}^q$, c.-à-d. :

$$\widehat{\theta} \in \arg\min_{\theta\in\Theta} \int_{\mathcal{X}\times\mathcal{Y}} \mathcal{L}(\boldsymbol{y}, f_\theta(\boldsymbol{x}))\widehat{p}_n(d\boldsymbol{x}, d\boldsymbol{y}) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\boldsymbol{Y}^{(i)}, f_\theta(\boldsymbol{X}^{(i)})). \tag{8.43}$$

Parmi les différentes familles paramétriques, on peut distinguer, entre autres, les modèles linéaires, non linéaires et basés sur des noyaux (Hastie et al., 2009). Les modèles linéaires, tels que la régression linéaire, supposent une relation linéaire entre les covariables $\boldsymbol{X}$ et la variable cible $\boldsymbol{Y}$, et dans ce cas, $\widehat{\theta}$ possède une formule explicite. Les modèles non-paramétriques, comme les réseaux de neurones (Hastie et al., 2009), capturent des relations plus complexes en utilisant des transformations non linéaires des données. Les modèles à noyau, tels que les SVM, consistent à projeter les données dans un espace de dimension supérieure via le noyau pour rendre les relations linéaires dans cet espace. Ces différentes familles offrent une flexibilité de choix selon la nature des données et les objectifs de l'étude. Une fois le problème d'optimisation en Eq. (8.43) résolu, la fonction

résultante est appelée un *estimateur* $\widehat{f} := f_{\widehat{\theta}}$. Dans les cas où les données présentent une structure temporelle, des modèles tels que les modèles de séries temporelles, ou des modèles à mémoire comme les réseaux récurrents (LSTM) peuvent être utilisés. Ces modèles captent des dépendances temporelles de portées potentiellement variées dans les données. Par exemple, un modèle LSTM peut être défini pour prédire la dégradation future $\Delta(t)$ à partir d'observations passées (également appelées étapes autorégressives) $(\Delta(t-1), \Delta(t-2), \ldots, \Delta(t-k))$, éventuellement en utilisant des covariables exogènes $\boldsymbol{X}(t), \boldsymbol{X}(t-1), \ldots, \boldsymbol{X}(t-k) \in \mathcal{X}$ observées conjointement. La fonction de perte $\mathcal{L}$ est alors adaptée pour tenir compte de la structure séquentielle des données, comme une perte cumulative sur des prédictions successives. Ces approches, toutefois, nécessitent un volume de données important et une validation rigoureuse pour éviter le surapprentissage. Des méthodes classiques de régression peuvent aussi être utilisées efficacement, en veillant à ce que $\mathcal{D}_{\text{test}}$ contienne uniquement des valeurs futures. Des approches d'ingénierie des caractéristiques peuvent être employées pour homogénéiser les dépendances temporelles (Kuhn and Johnson, 2019).

Comme pour toute stratégie de modélisation, les modèles statistiques présentent des limites. Ils reposent souvent sur l'hypothèse de stationnarité ou d'indépendance, qui peut ne pas tenir dans des scénarios réels. De plus, ils peuvent avoir du mal à généraliser lorsque la distribution des données change au cours du temps (phénomène de distribution shift) ou lorsque les données disponibles sont rares ou bruitées. Ces modèles nécessitent un réglage fin des hyperparamètres, sont sensibles à une mauvaise spécification a priori et au choix de la fonction de perte et de l'architecture du modèle. En outre, les modèles statistiques peuvent échouer à capturer des interactions physiques complexes ou des connaissances spécifiques au domaine, ce qui peut être critique pour des applications en ingénierie telles que le pronostic. Pour améliorer leur robustesse, une première voie consiste à quantifier l'incertitude de la prédiction de $\widehat{f}$. Cela peut être fait en utilisant la théorie des intervalles de confiance fréquentistes ou des approches bayésiennes. La théorie des intervalles de confiance fournit un cadre statistique pour quantifier l'incertitude des prédictions d'un modèle. Étant donné une estimation ponctuelle $\widehat{f}(\boldsymbol{X})$ pour la variable cible Y, un intervalle de confiance $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ est construit tel que, sous des échantillonnages répétés, la vraie valeur Y tombe dans cet intervalle avec une probabilité spécifiée $1-\alpha$, où α est le niveau de signification. Par exemple, en régression linéaire avec erreurs gaussiennes additives, l'intervalle est typiquement basé sur l'erreur standard estimée et les quantiles de la loi normale. Les intervalles de confiance offrent des bornes interprétables sur les prédictions et sont largement utilisés pour la quantification d'incertitude en régression et en prévision. Les approches bayésiennes traitent les paramètres du modèle comme des variables aléatoires avec des distributions a priori, puis mettent à jour ces distributions sur la base des données observées pour obtenir des distributions a posteriori. Par exemple, dans un cadre d'apprentissage statistique, si un paramètre du modèle θ a une distribution a priori $p(\theta)$ et la vraisemblance des données est donnée par $p(\mathcal{D}|\theta)$, alors la distribution a posteriori est donnée par la règle de Bayes :

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})}, \tag{8.44}$$

où $p(\mathcal{D})$ est la probabilité marginale des données, obtenue par marginalisation. Cette approche incorpore les incertitudes des paramètres et les propage dans les prédictions du modèle. Les méthodes bayésiennes nécessitent souvent des techniques d'échantillonnage dédiées telles que l'échantillonnage MCMC (Marin and Robert, 2007) ou des approximations variationnelles pour estimer les distributions a posteriori. Ces méthodes sont particulièrement utiles lorsque les données sont limitées ou

lorsque des connaissances a priori doivent être intégrées au modèle. Une autre approche pour quantifier les incertitudes dans les modèles d'apprentissage est la prédiction conforme (conformal prediction) (Vovk et al., 2005; Angelopoulos and Bates, 2023). Dans son approche générique, pour tout modèle d'apprentissage $\widehat{f} : \mathcal{X} \to \mathcal{Y} \subseteq \mathbb{R}$ appris sur un jeu d'entraînement $\mathcal{D}_{\text{train}}$, un intervalle de prédiction $\widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)})$ pour tout nouveau point $(\boldsymbol{X}^{(n+1)}, \boldsymbol{Y}^{(n+1)})$ est construit pour un taux de couverture $\alpha \in (0, 1)$ en utilisant le modèle $\widehat{f}$. Cet intervalle possède la propriété de couverture marginale :

$$\mathbb{P}\left(\boldsymbol{Y}^{(n+1)} \in \widehat{C}_{n,\alpha}(\boldsymbol{X}^{(n+1)})\right) \geq 1 - \alpha, \tag{8.45}$$

où la probabilité est comprise marginalement, c.-à-d. intégrée sur toutes les permutations possibles de $\mathcal{D}_{\text{train}} \cup \{(\boldsymbol{X}^{(n+1)}, \boldsymbol{Y}^{(n+1)})\}$. L'hypothèse de base pour utiliser les méthodes conformes est l'*échangeabilité* des données, les données i.i.d. étant un cas spécial. Par conséquent, les méthodes conformes ne s'étendent pas naturellement aux problèmes avec des sorties corrélées, comme les séries temporelles ou la statistique spatiale. Cependant, des avancées dans ce domaine, telles que des concepts de pseudo-échangeabilité (voir Zaffran et al. (2022)), permettent de surmonter cette limitation.

👉 Modèles statistiques pour le pronostic du colmatage

Dans le contexte du colmatage, les données de dégradation sont rares et irrégulières : les inspections directes par caméra n'ont lieu que lors d'arrêts rares, les mesures sont souvent indirectes ou bruitées, et les conditions opérationnelles varient entre les unités REP. Cela entraîne des défis pour les méthodes purement axées données : les jeux d'entraînement sont petits, les hypothèses d'échangeabilité et de stationnarité sont violées, les dépendances temporelles sont peu observées, et les prédicteurs appris ont tendance à sur-ajuster ou à échouer à transférer entre GV. En conséquence, de nombreux estimateurs statistiques peuvent produire des prévisions de RUL instables ou excessivement confiantes lorsqu'ils sont utilisés seuls pour des problèmes de colmatage. Pour faire face à ces contraintes, EDF a développé des outils indirects dédiés. ESTICOL est une procédure propriétaire d'estimation indirecte conçue pour inférer les taux historiques de colmatage à partir des essais périodiques BIL100 (voir chapitre 1) ; elle fournit des taux de colmatage reconstruits aux temps passés d'exploitation. PREVICOL est un outil opérationnel statistique qui implémente un modèle de régression linéaire (voir chapitre 5) pour extrapoler le colmatage futur à partir d'observations ponctuelles disponibles. Les deux méthodes réutilisent le même matériau d'observation limité : ESTICOL se concentre sur la transformation de signaux épars et indirects en estimations utilisables, tandis que PREVICOL applique des modèles prédictifs légers sur un travail d'ingénierie des caractéristiques spécifique et manquent toutes deux de données de validation suffisantes.

Modèles hybrides

Dans le contexte de la PHM, un *modèle hybride* se réfère à un cadre de modélisation qui intègre des modèles basés sur la physique avec des modèles statistiques ou d'apprentissage automatique basés sur les données. L'objectif est de tirer parti des forces des deux paradigmes : les modèles physiques fournissent une compréhension mécanistique et une capacité d'extrapolation, tandis que les modèles basés sur les données offrent une flexibilité pour capturer des dynamiques non modélisées en

conditions opérationnelles, des erreurs résiduelles ou d'autres corrélations inconnues (Liao and Kottig, 2014; Eker et al., 2019). De telles approches hybrides sont particulièrement pertinentes lorsque la modélisation physique est incomplète ou trop coûteuse en calcul, et que des données d'observation sont disponibles mais potentiellement bruitées ou éparses. Un cadre hybride typique en PHM peut impliquer des stratégies de correction de modèle, une mise à jour bayésienne, un apprentissage informé par la physique, ou des techniques de filtrage double. Ces modèles sont instrumentaux pour améliorer la robustesse, la précision et l'explicabilité des outils prédictifs de maintenance, en particulier pour des systèmes industriels complexes avec variabilité opérationnelle et observation de dégradation limitée.

Une approche de substitution hybride populaire dans la communauté Scientific Machine Learning (SciML) aujourd'hui implique l'apprentissage *physics-informed*, où une stratégie consiste à régulariser les fonctions de perte par des termes physiques. Ceci est réalisé, par exemple, avec les réseaux de neurones informés par la physique (PINNs) (Raissi et al., 2019), ou plus récemment les développements en chaos polynomial (Novák et al., 2024). De plus, dans les approches bayésiennes de processus gaussiens, des noyaux a priori peuvent être choisis pour correspondre aux solutions fondamentales d'équations, par exemple (Henderson et al., 2023) pour l'équation des ondes. Ces approches d'apprentissage régularisent la fonction de perte en ajoutant un terme physique qui pénalise les solutions violant les lois physiques. Plus précisément, si l'on considère un paramètre $\boldsymbol{X} = \boldsymbol{x}_0$ dans l'Éq. (8.40) avec une condition aux limites de Dirichlet homogène, le terme de régularisation est défini comme :

$$\mathcal{L}_\varphi = \frac{1}{N_t N_{\boldsymbol{z}}} \sum_{i=1}^{N_t} \sum_{j=1}^{N_{\boldsymbol{z}}} \left\| \mathcal{A}_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) \widehat{u}_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) - F_{\boldsymbol{x}_0}(t_i, \boldsymbol{z}_j) \right\|^2, \tag{8.46}$$

où $\widehat{u}$ est l'approximation donnée par le modèle d'apprentissage (par exemple, un réseau de neurones dans le cas des PINNs (Raissi et al., 2019)) et $\{(t_i, \boldsymbol{z}_j)\}_{i,j}$ sont des points de collocation sur $[0, T] \times \boldsymbol{D}$ échantillonnés uniformément. La perte totale devient :

$$\mathcal{L}_{\text{tot}} = \mathcal{L} + \lambda \mathcal{L}_\varphi, \tag{8.47}$$

où $\mathcal{L}$ est la perte supervisée standard en Eq. (8.43), et $\lambda > 0$ est un hyperparamètre contrôlant la force de la régularisation physique. Une telle régularisation favorise des solutions consistantes avec les principes physiques, même en présence de peu de données. Ces méthodes sont particulièrement utiles lorsque les données sont coûteuses, bruitées ou partiellement observées, car elles combinent les bénéfices de l'apprentissage supervisé et de la modélisation physique. Cependant, elles peinent avec des problèmes multiphysiques et multiéchelles couplés, ainsi que pour l'intégration sur de longues durées et l'entraînement qui est souvent coûteux. Néanmoins, des garanties théoriques d'approximation et de généralisation sont connues pour des opérateurs spécifiques (Mishra and Molinaro, 2021; Doumèche et al., 2023). Ces méthodes agissent principalement comme des interpolateurs spatio-temporels raffinant les solutions localement tout en préservant des contraintes physiques. Cependant, dans leur forme actuelle, elles ne traitent pas la paramétrisation aléatoire de $\boldsymbol{X}$, qui fait l'objet de recherches en cours (voir par exemple (Panahi et al., 2025)).
Les métamodèles des différents types peuvent être utilisés pour des problèmes avant (forward), signifiant que l'on calcule diverses statistiques de la mesure image originale $g\#p(\boldsymbol{X})$, mais ils peuvent aussi être employés pour des problèmes inverses (Stuart, 2010) tels que la mise à jour bayésienne du

modèle qui est un type de modèle hybride. Dans ce contexte, on cherche à réduire les incertitudes a priori sur les paramètres d'entrée, à l'aide de données $\boldsymbol{y} \in \mathbb{R}^N$. Ces données peuvent provenir d'un modèle basé sur les données ou de données capteurs réelles. En termes simples, l'hypothèse suivante est faite :

$$\boldsymbol{y} = g(\boldsymbol{X}) + \boldsymbol{\eta}, \tag{8.48}$$

où $\boldsymbol{\eta}$ représente le bruit, typiquement modélisé comme une variable aléatoire avec une loi a priori (par exemple gaussienne). L'objectif est d'estimer la distribution a posteriori de $\boldsymbol{X}$ étant donné les données observées $\boldsymbol{y}$. Cela se fait à nouveau en utilisant la règle de Bayes comme en Eq. (8.44) :

$$p(\boldsymbol{X}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{X})p(\boldsymbol{X}). \tag{8.49}$$

Des techniques telles que MCMC ou l'inférence variationnelle peuvent être utilisées pour échantillonner ou approximer la distribution a posteriori (Rubinstein and Kroese, 2011). Dans ce contexte, l'utilisation de métamodèles est cruciale si le modèle numérique g est coûteux en calcul. Ces méthodes sont particulièrement utiles pour calibrer des modèles de simulation ou mettre à jour des prédictions en temps réel à mesure que de nouvelles données deviennent disponibles. Dans le cadre bien connu de Kennedy and O'Hagan (2001), le modèle en Eq. (8.48) peut être augmenté en ajoutant un terme de biais (model discrepancy), prenant en compte le décalage entre la valeur vraie de la dégradation et la prédiction du modèle.

Une autre technique bayésienne typique utilisée pour le pronostic hybride (Liao and Kottig, 2014) est l'*assimilation de données*, et le choix courant est le *filtrage* (Jouin et al., 2016). Il est particulièrement adapté aux problèmes de dégradation dépendants du temps avec un flux continu et une surveillance des données. Cette méthode traite le problème inverse de la mise à jour de la probabilité de l'état de dégradation à un instant donné en utilisant toutes les données disponibles le précédant. Les techniques de filtrage les plus courantes incluent le filtre de Kalman et ses variantes (Evensen, 1994), et les filtres particulaires (Del Moral et al., 2006). Le filtre de Kalman suppose des dynamiques linéaires et un bruit gaussien, ce qui le rend efficace en calcul mais limité pour les systèmes non linéaires, même si des extensions existent pour atténuer ces limitations (voir Sullivan (2015)). Le filtre particulaire, quant à lui, est une approche non-paramétrique qui utilise un ensemble de particules pour représenter la distribution a posteriori, le rendant adapté aux problèmes fortement non linéaires et non gaussiens.

Par exemple, dans le contexte de la dégradation, rappelons que $\Delta(t)$ représente l'état de dégradation au temps t, et notons $\boldsymbol{y}_{1:t}$ la séquence des observations jusqu'au temps t. L'objectif est de calculer la distribution a posteriori $p(\Delta(t)|\boldsymbol{y}_{1:t})$. À l'aide d'un filtre particulaire, cela est réalisé en propageant un ensemble de particules $\{\Delta^{(i)}(t)\}_{i=1}^n$ à travers le modèle physique et en mettant à jour leurs poids en fonction de la vraisemblance des observations :

$$p(\Delta(t)|\boldsymbol{y}_{1:t}) \approx \sum_{i=1}^{n} w^{(i)}(t)\delta_{\Delta^{(i)}(t)}, \quad w^{(i)}(t) \propto w^{(i)}(t-1)p(\boldsymbol{y}(t)|\Delta^{(i)}(t)), \tag{8.50}$$

où $\delta_{\Delta^{(i)}(t)}$ est la distribution de Dirac centrée en la particule $\Delta^{(i)}(t)$. Cette approximation permet un calcul efficace de la distribution a posteriori et facilite l'estimation de quantités d'intérêt, telles que l'espérance de l'état de dégradation ou la probabilité de dépasser un seuil critique, tout en surveillant

la non-dégénérescence des poids (voir (Chopin and Papaspiliopoulos, 2020)). En propulsant ℓ pas en avant du temps présent t_P, la distribution de la RUL sera estimée par sa fonction de distribution cumulative :

$$\mathbb{P}(\text{RUL}(\Delta_*) \leq t_{P+\ell} \mid \boldsymbol{y}_{1:t_P}) \approx \sum_{i=1}^{n} w^{(i)}(t_P)\mathbf{1}\{\Delta^{(i)}(t_{P+\ell}) > \Delta_*\}. \tag{8.51}$$

Les méthodes d'assimilation de données (Geir Evensen, 2022) sont particulièrement puissantes pour la mise à jour en temps réel, en ligne, signifiant que des observations $\boldsymbol{y}$ sont continuellement disponibles à tous les pas de temps. De plus, ce cadre suppose une structure markovienne pour le phénomène de dégradation, permettant de mettre à jour la probabilité de dégradation uniquement à partir de son pas précédent via un noyau $p(\Delta(t)|\Delta(t-1))$.

👉 Modèles hybrides pour le pronostic du colmatage

Dans ce travail nous simplifions le modèle physique en supposant l'absence de stochasticité intrinsèque de la dégradation : cela signifie que nous traitons le simulateur propriétaire THYC-Puffer-DEPO comme une base déterministe non-intrusive pour l'estimation probabiliste de la RUL. L'incertitude restante est donc paramétrique (provenant des paramètres d'entrée) et en principe réductible avec des données supplémentaires et de meilleurs a priori. Cette hypothèse de travail nous permet de concentrer le flux de travail hybride sur la mise à jour de l'état de colmatage simulé via l'assimilation d'observations dans le modèle afin de délivrer des estimations RUL informées par le risque et robustes. Parce que TPD est utilisé comme une boîte noire produisant des trajectoires complètes de dégradation (plutôt que permettant des mises à jour pas à pas), le filtrage en ligne standard n'est pas adapté ; par conséquent, nous traitons le problème offline en utilisant les groupes hétérogènes de données disponibles (inspections caméra rares et reconstructions ESTICOL) pour réduire d'abord l'incertitude de sortie du simulateur en informant la distribution des entrées via une mise à jour bayésienne, puis via un lissage (smoothing) des trajectoires postérieures (comme nous le verrons au chapitre 4).

Énoncé du problème et objectifs

L'objectif principal de ce travail de thèse est d'obtenir des prédictions RUL robustes vis-à-vis des incertitudes, en fusionnant toutes les informations et outils numériques disponibles dans un cadre hybride. Ce défi industriel pose des limites aux approches classiques de pronostic, notamment en raison du manque de jeux de données de terrain substantiels et cohérents ainsi que de la complexité physique et des différents scénarios opérationnels pour chaque REP. Par conséquent, on ne peut espérer construire des algorithmes purement basés sur les données généralisables ni être totalement satisfait d'une approche purement physique et numérique. Les incertitudes sont présentes dans tous les modèles et, dans cet ensemble d'hypothèses, notre objectif sera de minimiser l'incertitude globale et de garantir son contrôle à chaque étape sous toutes ses formes.

Un objectif à plus long terme de ce travail est de proposer des méthodologies pour la création de *jumeaux numériques de composants de centrales nucléaires* (NPPCs) (US-NRC, 2021). Les jumeaux numériques sont des représentations virtuelles d'actifs physiques qui intègrent données, modèles

et simulations avec les incertitudes associées afin de permettre la surveillance en temps réel, le diagnostic et le pronostic (NAS, 2024; Liang et al., 2024). Dans le contexte des centrales nucléaires, les jumeaux numériques pourraient faciliter la maintenance prédictive, optimiser les stratégies opérationnelles et renforcer la sûreté grâce à une vision continuellement mise à jour de l'état de santé des composants. Pour y parvenir, il faut intégrer des modèles physiques, des approches basées sur les données, la quantification d'incertitude et des techniques d'assimilation de données discutées ci-dessus. Cependant, à ce jour, beaucoup reste à faire pour obtenir une surveillance en temps réel des données pertinentes sur les NPPC existants, en commençant par la qualification des capteurs et leur placement optimal ainsi que la validation et la vérification des modèles physiques et des substituts. Le développement de jumeaux numériques robustes repose donc sur des cadres de modélisation hybrides capables de s'adapter à de nouvelles informations, d'intégrer des connaissances de domaine et de fournir des prédictions fiables sous incertitude pour des installations existantes. Pour les nouvelles installations, il y a une marge d'amélioration lors de la phase de conception afin de faciliter le déploiement de méthodologies de jumeaux numériques.

Plan de la thèse

Le reste de ce manuscrit est structuré comme suit. Chaque chapitre correspond à un composant clé du schéma général illustré à la Figure 8.18 :

- *Chapitre 2 : Modélisation du colmatage et jumeaux numériques pour le pronostic du colmatage*
  - Nous exposons les outils de prédiction existants pour le colmatage des GV à EDF R&D et expliquons la stratégie de jumeau numérique.
  - Nous identifions les principales limitations des approches hybrides standard issues de la littérature et formulons le problème principal abordé dans cette thèse.
- *Chapitre 3 : Quantification des incertitudes de THYC-Puffer-DEPO*
  - Nous présentons des résultats sur des techniques avancées d'analyse de sensibilité pour la quantification des incertitudes du code de simulation de colmatage THYC-Puffer-DEPO basés sur le travail publié (Jaber et al., 2025b).
  - Nous présentons un certain nombre de stratégies de métamodélisation pour le code THYC-Puffer-DEPO, à savoir : processus gaussiens, développements en chaos polynomial, métamodèles de champ basés sur Karhunen-Loève ainsi qu'une stratégie d'agrégation de substituts minimisant le biais.
  - Nous mettons en évidence de nouvelles caractéristiques inconnues de l'incertitude du modèle de colmatage et obtenons une première vue sur l'incertitude prédictive de la RUL à partir d'une approche purement basée sur le modèle sur un GV spécifique.
- *Chapitre 4 : Métamodélisation prédictive robuste avec prédiction conforme*
  - Nous présentons un rappel sur les méthodes de prédiction conforme et leur potentiel d'application aux tâches SciML.

  - Nous présentons l'estimateur cross-conforme développé pour le diagnostic et la qualification de modèles de substitution à processus gaussien scalaire de codes déterministes, en lien avec notre travail publié (Jaber et al., 2025a).
  - Nous développons des bandes d'incertitude pour des substituts vectoriels basés sur KL et GPs en utilisant le contrôle de risque conforme et montrons leur potentiel pour la qualification a priori du metamodèle.
  - Les deux stratégies sont appliquées à THYC-Puffer-DEPO et nous discutons leurs implications pour l'apprentissage actif et le design.

- *Chapitre 5 : Stratégies hybrides bayésiennes offline*

  - Il s'agit du chapitre central de la thèse dans lequel nous développons une méthodologie générique d'assimilation de données offline, permettant de mettre à jour les variables d'entrée pertinentes du modèle numérique ainsi qu'appliquer un lissage de Kalman afin d'obtenir une RUL robuste et informée par le risque. Il est associé au travail soumis (Jaber et al., 2026).
  - Nous discutons de l'intégration potentielle d'un modèle de biais additionnel dans l'étape de mise à jour bayésienne du modèle.

- *Chapitre 6 : Modèles de régression pour le pronostic basés sur les séries temporelles*

  - Il s'agit d'un chapitre prospectif sur l'utilisation potentielle des jeux de données de surveillance capteurs disponibles des GV en vue d'une surveillance en temps réel du colmatage.
  - Nous présentons le travail statistique pour le pronostic à EDF R&D appelé PREVICOL et montrons comment il pourrait être potentiellement amélioré en utilisant des modèles d'apprentissage basés sur les séries temporelles.
  - Nous explorons leur pouvoir explicatif pour la modélisation du taux de colmatage à l'aide de différents modèles de régression, y compris la régression linéaire et la régression ridge.

La thèse se termine par une conclusion et des perspectives de travail futur en 7 et l'annexe contenant divers résultats additionnels des différents chapitres se trouve en 8.

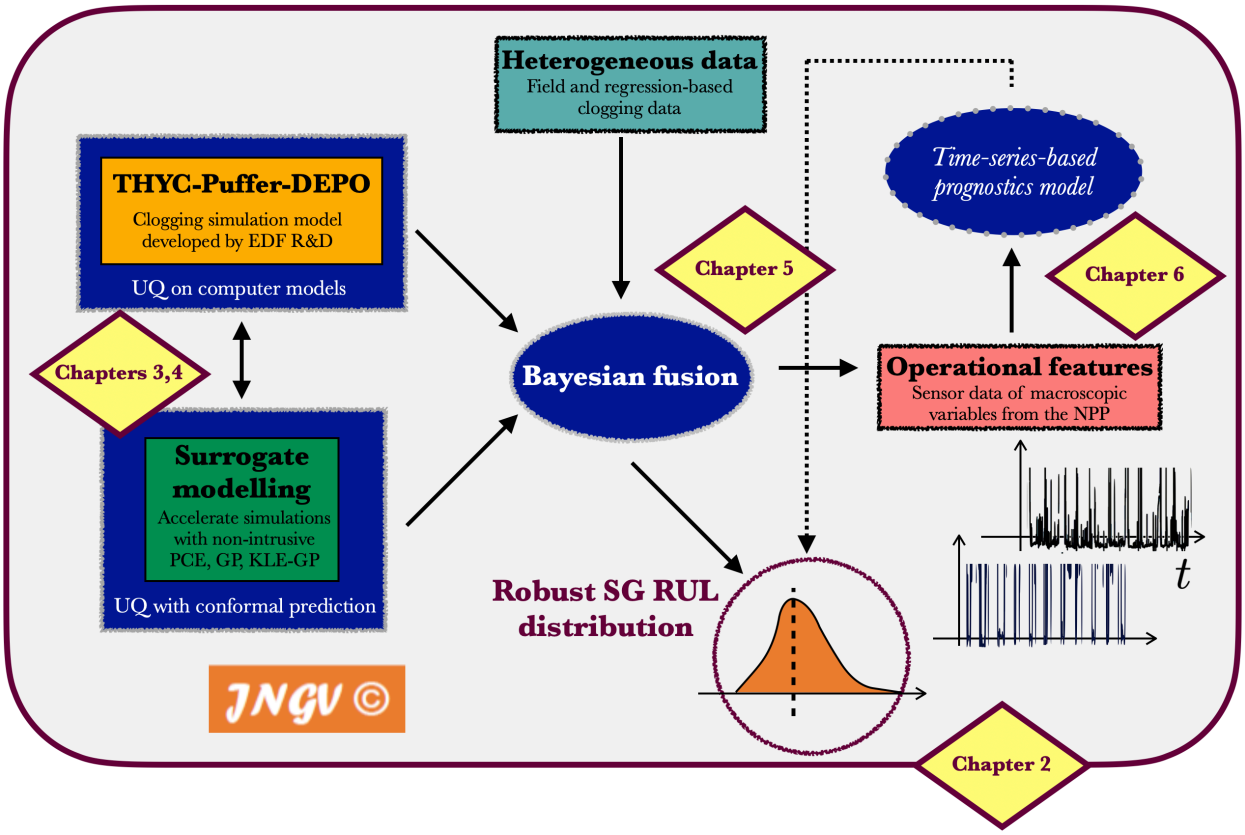

Figure 8.18: Le cadre du jumeau numérique développé et l'articulation des différents chapitres de la thèse.

Articles publiés dans des revues à comité de lecture

1. *Jaber, E., Chabridon, V., Remy, E., Baudin, M., Lucor, D., Mougeot, M., Iooss, B., Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators, Published in International Journal of Uncertainty Quantification, 2025, 10.1615/Int.J.UncertaintyQuantification.2024051489*

2. *Jaber, E., Blot, V., Brunel, N., Chabridon, V., Remy, E., Iooss, B., Lucor, D., Mougeot, M., Leite, A., Conformal Approach to Gaussian Process Surrogate Evaluation with Marginal Coverage Guarantees, Published in Journal of Machine Learning for Modeling and Computing, 2025, 10.1615/JMachLearnModelComput.2025054687*

3. *Jaber, E., Remy, E., Chabridon, V., Lucor, D., Mougeot, M., Fusion of heterogeneous data for robust degradation prognostics, Published in Journal of Reliability Engineering and System Safety, 2026, 10.1016/j.ress.2026.112435*

Résumé des communications orales

- 04/2023, MASCOT-NUM 2023, Le Croisic, France - poster
- 09/2023, CJC-MA 2023, CentraleSupélec - poster
- 10/2023, ETICS 2023, Lège Cap-Ferret, France - slides

- 02/2024, SIAM Conference on Uncertainty Quantification 2024, Trieste, Italie - slides
- 09/2024, ETICS 2024, Saissac, France - slides (en français)
- 02/2025, DTE&AICOMAS 2025, Paris, France - slides
- 04/2025, SAMO 2025, Grenoble, France - poster
- 06/2025, INI Workshop: Calibrating prediction uncertainty : statistics and machine learning perspectives, Cambridge, Royaume-Uni - slides, vidéo
- 06/2025, UNCECOMP 2025, Rhodes, Grèce - slides
- 08/2025, Mathematical and Computational Foundations of Digital Twins, CIRM Marseille, France - présentation

Résumé des développements computationnels

Tous les résultats de ce manuscrit sont reproductibles et sont liés chapitre par chapitre aux dépôts GitHub suivants `thesis GitHub repository`. Cette approche a le mérite d'assurer la reproductibilité, permettant à d'autres de vérifier et d'étendre le travail et pouvant être utilisée par les ingénieurs d'EDF pour déployer les méthodologies :

1. `SA of a multi-physics long-term clogging model for steam generators` : implémentations d'analyses de sensibilité et de techniques de métamodélisation avec OpenTURNS[2].
2. `Conformal approach for GP surrogates` : contribution au dépôt avec Vincent Blot, couplage entre la librairie de prédiction conforme MAPIE[3] et OpenTURNS.
3. `Offline data fusion for robust degradation prognostics` : création et simplification de classes pour MCMC et assemblage de métamodèles avec OpenTURNS, implémentation de la méthodologie sur un cas témoin et le cas d'utilisation principal.

[2]voir la documentation : `https://openturns.github.io/openturns/latest/index.html`
[3]voir la documentation : `https://mapie.readthedocs.io/en/stable/`

Bibliography

(2016). *Uncertainty Quantification in Computational Fluid Dynamics*. Lecture Notes in Computational Science and Engineering. Springer Cham.

Acharki, N., Bertoncello, A., and Garnier, J. (2023). Robust prediction interval estimation for gaussian processes by cross-validation method. *Computational Statistics & Data Analysis*, 178:107597.

AIAA (2021). Digital Twin: definition & value - An AIAA and AIA Position Paper.

Amer, A. W., Tsilifis, P., Saxena, A., Baglietto, E., and Khan, G. (2025). *Probabilistic Health Monitoring and Maintenance Scheduling for Nuclear Power Systems*.

Angelopoulos, A. N. and Bates, S. (2023). Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591.

Angelopoulos, A. N., Bates, S., Fisch, A., Lei, L., and Schuster, T. (2025). Conformal Risk Control.

Argaud, J.-P., Bouriquet, B., de Caso, F., Gong, H., Maday, Y., and Mula, O. (2018). Sensor placement in nuclear reactors based on the generalized empirical interpolation method. *Journal of Computational Physics*, 363:354–370.

Barber, R. F., Candès, E. J., Ramdas, A., and Tibshirani., R. J. (2021). Predictive inference with the jackknife+. *Annals of Statistics*, 49:486–507.

Barlow, R. E. and Proschan, F. (1965). *Mathematical Theory of Reliability*. John Wiley & Sons.

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, *Handbook on Uncertainty Quantification*, pages 2001–2038.

Berlinet, A. and Thomas-Agnan, C. M. (2004). *Reproducing Kernel Hilbert Space in Probability and Statistics*. Springer US.

Biggio, L. and Kastanis, B. (2020). Prognostics and health management of industrial assets: Current progress and road ahead. *Frontiers in Artificial Intelligence*, 3:578613.

Blatman, G. and Sudret, B. (2009). Anisotropic parcimonious polynomial chaos expansions based on the sparsity-of-effects principle. In *Proc. 10th Int. Conf. Struct. Safety and Reliability (ICOSSAR'2009)*, Osaka, Japan.

Blatman, G. and Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367.

Blot, V., Angelopoulos, A. N., Jordan, M. I., and Brunel, N. J.-B. (2025). Automatically Adaptive Conformal Risk Control.

Bourinet, J.-M. (2017). *FORM sensitivities to distribution parameters with the Nataf transformation*, pages 277–302. Springer International Publishing.

Bousquet, O. et Elisseeff, A. (2002). Stability and Generalization. *Journal of Machine Learning Research*, 2:499–526.

Brynjarsdóttir, J. and O'Hagan, A. (2014). Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30.

Cacuci, D. G. (2003). *Sensitivity & Uncertainty Analysis, Volume 1: Theory*. Chapman and Hall/CRC.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, Learning, and Games*. Cambridge University Press.

Chopin, N. and Papaspiliopoulos, O. (2020). *An introduction to Sequential Monte Carlo*. Springer Series in Statistics,. Springer Cham.

Cordier, T., Blot, V., Lacombe, L., Morzadec, T., Capitaine, A., and Brunel, N. (2023). Flexible and systematic uncertainty estimation with conformal prediction via the mapie library. In Papadopoulos, H., Nguyen, K. A., Boström, H., and Carlsson, L., editors, *12th Symposium on Conformal and Probabilistic Prediction with Applications*, volume 204, pages 549–581.

Crespi, N., Drobot, A. T., and Minerva, R., editors (2023). *Digital Twins for Nuclear Power Plants and Facilities*, pages 971–1022. Springer International Publishing, Cham.

Cubillo, A., Perinpanayagam, S., and Esperon Miguez, M. (2016). A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery. *Advances in Mechanical Engineering*, 8.

Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of Statistical Computation and Simulation*, 85:1283–1305.

Da Veiga, S. (2021). Kernel-based ANOVA decomposition and Shapley effects – Application to global sensitivity analysis. *arXiv 2101.05487*.

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

David, F. (1999). Three Dimensional Thermal-Hydraulic Simulation In Steam Generators With THYC Exchangers Code - Application To The UTSG Model 73/19. Proceedings of the Ninth International Topical Meeting on Nuclear Reactor Thermal Hydraulics (NURETH-9).

De Carvalho, T. M., Van Rosmalen, J., Wolff, H. B., Koffijberg, H., and Coupé, V. M. H. (2022). Choosing a metamodel of a simulation model for uncertainty quantification. *Medical Decision Making*, 42(1):28–42.

De Rocquigny, E., Devictor, N., and Tarantola, S., editors (2008). *Uncertainty in industrial practice - A guide to quantitative uncertainty managment*. Wiley and Sons.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(3):411–436.

Delaney, B. and Wang, Q. J. (2025). Archard's law: Foundations, extensions, and critiques. *Encyclopedia*, 5(3).

Demay, C., Iooss, B., Le Gratiet, L., and Marrel, A. (2022). Model selection based on validation criteria for gaussian process regression: An application with highlights on the predictive variance. *Quality and Reliability Engineering International*, 38(3):1482–1500.

Deri, E., Varé, C., and Wintergerst, M. (2021). *Development of Digital Twins of PWR Steam Generators: Description of Two Maintenance-Oriented Use Cases*, volume Volume 1: Operating Plant Challenges, Successes, and Lessons Learned; Nuclear Plant Engineering; Advanced Reactors and Fusion; Small Modular and Micro-Reactors Technologies and Applications of *International Conference on Nuclear Engineering*.

Dhada, M., Bull, L., Girolami, M., and Parlikad, A. (2025). Collaborative prognosis using a weibull statistical hierarchical model. *Reliability Engineering & System Safety*, 262:111110.

Doumèche, N., Biau, G., and Boyer, C. (2023). Convergence and error analysis of pinns.

Eker, O. F., Camci, F., and Jennions, I. K. (2012). Major Challenges in Prognostics: Study on Benchmarking Prognostics Datasets. *PHM Society European Conference*, 1(1).

Eker, O. F., Camci, F., and Jennions, I. K. (2019). A New Hybrid Prognostic Methodology. *International Journal of Prognostics and Health Management*, 10(2).

El Garroussi, S., Ricci, S., De Lozzo, M., Goutal, N., and Lucor, D. (2022). Tackling random fields non-linearities with unsupervised clustering of polynomial chaos expansion in latent space: application to global sensitivity analysis of river flooding. *Stochastic Environmental Research and Risk Assessment*, 36:693–718.

Escobet, T., Bregon, A., Pulido, B., and Puig, V. (2019). *Fault diagnosis of dynamic systems: quantitative and qualitative approaches*. Springer International Publishing.

Evensen, G. (1994). Sequential Data Assimilation With a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics. *Journal of Geophysical Research*, 99:10143–10162.

Evensen, G. and Van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, 128(6):1852–1867.

Fang, K.-T., Li, R., and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. CRC Press.

Fekhari, E., Iooss, B., Muré, J., Pronzato, L., and Rendas, J. (2023). Model predictivity assessment: incremental test-set selection and accuracy evaluation. In Salvati, N., Perna, C., Marchetti, S., and Chambers, R., editors, *Studies in Theoretical and Applied Statistics*, pages 315–347.

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.

García Nieto, P., García-Gonzalo, E., Sánchez Lasheras, F., and de Cos Juez, F. (2015). Hybrid pso-svm-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. *Reliability Engineering and System Safety*, 138:219–231.

Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press.

Geir Evensen, Femke C. Vossepoel, P. J. v. L. (2022). *Data assimilation fundamentals*. Springer Cham.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science*, 7(4):457 – 472.

Ghanem, R., Higdon, D., and Owhadi, H., editors (2017). *Handbook of Uncertainty Quantification*. Springer International Publishing, Cham.

Girard, S. (2014). *Clogging of Recirculating Nuclear Steam Generators*. Springer International Publishing.

Goldfarb, H., Huang, H., Iyer, N. S., Roychowdhury, S., Saxena, A., Subramaniam, A., Tang, L., Virani, N., Huning, A., Muhlheim, M., et al. (2025). Pmdt: Ai-enabled predictive maintenance digital twins for advanced nuclear reactors. Technical report, GE Global Research, Niskayuna, New York (United States).

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. In *Proceedings Algorithmic Learning Theory*, pages 63–77. Springer-Verlag.

Griffith, A. A. (1920). The phenomena of rupture and flow in solids. 221(582-593):163–198.

Gu, M., Wang, X., and Berger, J.-O. (2018). Robust gaussian stochastic process emulation. *Annals of Statistics*, 46(6A):3038–3066.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer, 2nd edition.

He, W., Williard, N., Osterman, M., and Pecht, M. (2011). Prognostics of lithium-ion batteries based on dempster–shafer theory and the bayesian monte carlo method. *Journal of Power Sources*, 196(23):10314–10321.

Henderson, I., Noble, P., and Roustant, O. (2023). Covariance models and gaussian process regression for the wave equation. application to related inverse problems. *Journal of Computational Physics*, 494:112519.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3):293–325.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Hossain, R., Ahmed, F., Kobayashi, K., et al. (2025). Virtual sensing-enabled digital twin framework for real-time monitoring of nuclear systems leveraging deep neural operators. *npj Materials Degradation*, 9:21.

Iooss, B. and Marrel, A. (2019). Advanced methodology for uncertainty propagation in computer experiments with large number of inputs. *Nuclear Technology*, 205:1588–1606.

Jaber, E., Blot, V., Brunel, N., Chabridon, V., Remy, E., Iooss, B., Lucor, D., Mougeot, M., and Leite, A. (2025a). Conformal approach to Gaussian process surrogate evaluation with marginal coverage guarantees. *Journal of Machine Learning for Modeling and Computing*.

Jaber, E., Chabridon, V., Remy, E., Baudin, M., Lucor, D., Mougeot, M., and Iooss, B. (2025b). Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators. *International Journal for Uncertainty Quantification*, 15:27–45.

Jaber, E., Remy, E., Chabridon, V., Mougeot, M., and Lucor, D. (2026). Fusion of heterogeneous data for robust degradation prognostics. *Reliability Engineering & System Safety*, page 112435.

Jouin, M., Gouriveau, R., Hissel, D., Péra, M.-C., and Zerhouni, N. (2016). Particle filter-based prognostics: Review, discussion and perspectives. *Mechanical Systems and Signal Processing*, 72-73:2–31.

Kapteyn, M. G., Knezevic, D. J., and Willcox, K. E. (2020). Toward predictive digital twins via component-based reduced-order models and interpretable machine learning.

Karnik, N., Abdo, M. G., Estrada-Perez, C. E., Yoo, J. S., Cogliati, J. J., Skifton, R. S., Calderoni, P., Brunton, S. L., and Manohar, K. (2024). Constrained optimization of sensor placement for nuclear digital twins. *IEEE Sensors Journal*, 24(9):15501–15516.

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.

Kerkar, N. and Paulin, P. (2008). *Exploitation des coeurs REP*. edp sciences, INSTN, Les Ulis.

Kochunas, B. and Huan, X. (2021). Digital twin concepts with uncertainty for nuclear power applications. *Energies*, 14(14).

Kuhn, M. and Johnson, K. V. (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models*. Chapman and Hall/CRC.

Le Maître, O. P. and Knio, O. M. (2010). *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Scientific Computation. Springer.

Lefebvre, L., Segond, M., Spaggiari, R., Le Gratiet, L., Deri, E., Iooss, 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.

Lei, J. (2019). Fast exact conformalization of the lasso using piecewise linear homotopy. *Biometrika*, 106(4):749–764.

Li, A. G. (2025). *Pulse Perturbation for Battery Management*. PhD thesis, Columbia University.

Li, X., Ding, Q., and Sun, J.-Q. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering and System Safety*, 172:1–11.

Liang, H., Moya, B., Seah, E., Weng, A. N. K., Baillargeat, D., Joerin, J., Zhang, X., Chinesta, F., and Chatzi, E. (2024). Harnessing hybrid digital twinning for decision-support in smart infrastructures.

Liang, R. and Barber, R. F. (2023). Algorithmic stability implies training-conditional coverage for distribution-free prediction methods. *arXiv:2311.04295*.

Liao, L. and Kottig, F. (2014). Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction. *IEEE Transactions on Reliability*, 63(1):191–207.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27(4):986–1005.

Liu, J., Dai, C., and Hu, Y. (2018). Aqueous aggregation behavior of citric acid coated magnetite nanoparticles: Effects of pH, cations, anions, and humic acid. *Environmental Research*, 161:49–60.

Liu, Y., Alsafadi, F., Mui, T., O'Grady, D., and Hu, R. (2025). Development of whole system digital twins for advanced reactors: Leveraging graph neural networks and sam simulations. *Nuclear Technology*, 211(9):2206–2223.

Loubet, L., Allinei, P.-G., Schindler, P., and Pointeau, V. (2020). Implementation of three gamma measuring stations on the COLENTEC loop in Cadarache for the on line observation of the clogging phenomena in Steam Generator. *EPJ Web of Conferences*, 225:03011.

Lüthen, N., Marelli, S., and Sudret, B. (2021). Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark. *SIAM/ASA Journal on Uncertainty Quantification*, 9:593–649.

Marin, J. and Robert, C. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer.

Marrel, A. and Chabridon, V. (2021). Statistical developments for target and conditional sensitivity analysis: application on safety studies for nuclear reactor. *Reliability Engineering and System Safety*, 214:107711.

Marrel, A. and Iooss, B. (2024). Probabilistic surrogate modeling by gaussian process: A review on recent insights in estimation and validation. *Reliability Engineering & System safety*, 247:110094.

Mathieu, F., De Caminel, A., Chabridon, V., Remy, E., Deri, E., Wintergerst, M., Trillon, D., and Molé, D. (2020). Apprentissage statistique pour la prévision de la dégradation d'un composant industriel. *Actes du 22ème Congrès de Maîtrise des Risques et Sûreté de Fonctionnement (Lambda-Mu22)*.

McNab, A. (1988). A review of eddy current system technology. *British Journal of Nondestructive Testing*, 30(7):249–255.

Méndez Rojano, R. M., Zhussupbekov, M., Antaki, J. A., and Lucor, D. (2022). Uncertainty quantification of a thrombosis model considering the clotting assay PFA-100®. *International Journal for Numerical Methods in Biomedical Engineering*, 38:e3595.

Mengyan, H., Xueyan, Z., Cuiting, P., Yixuan, Z., and Jun, Y. (2024). Current status of digital twin architecture and application in nuclear energy field. *Annals of Nuclear Energy*, 202:110491.

Mishra, S. and Molinaro, R. (2021). Estimates on the generalization error of physics-informed neural networks for approximating a class of inverse problems for pdes. *IMA Journal of Numerical Analysis*, 42(2):981–1022.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161–174.

NAS (2024). *Foundational Research Gaps and Future Directions for Digital Twins*. The National Academies Press.

Novák, L., Sharma, H., and Shields, M. D. (2024). Physics-informed polynomial chaos expansions. *Journal of Computational Physics*, 506:112926.

Oluwasegun, A. and Jung, J.-C. (2020). The application of machine learning for the prognostics and health management of control element drive system. *Nuclear Engineering and Technology*, 52(10):2262–2273.

Panahi, M., Porta, G. M., Riva, M., and Guadagnini, A. (2025). Modeling parametric uncertainty in pdes models via physics-informed neural networks. *Advances in Water Resources*, 195:104870.

Papadopoulos, H. (2024). Guaranteed coverage prediction intervals with gaussian process regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–12.

Papadopoulos, H., Gammerman, A., and Vovk, V. (2008). Normalized nonconformity measures for regression conformal prediction. In *26th International Conference on Artificial Intelligence and Applications*, pages 64–69.

Papadopoulos, H., Proedrou, K., Vovk, V., and Gammerman, A. (2002). Inductive confidence machines for regression. In Elomaa, T., Mannila, H., and Toivonen, H., editors, *ECML*, pages 345–356.

Papadopoulos, H., Vovk, V., and Gammerman, A. (2011). Regression conformal prediction with nearest neighbours. *J. Artif. Intell. Res.*, 40:815–840.

Paris, P. and Erdogan, F. (1960). A critical analysis of crack propagation laws. *Journal of Basic Engineering*, 85.

Pecht, M. (1992). *Product Reliability, Maintainability, and Supportability Handbook*. CRC Press.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.*, 12:2825–2830.

Peksen, M. (2018). *Multiphysics Modeling: Materials, Components, and Systems*. Academic Press.

Peng, T., He, J., Xiang, Y., Liu, Y., Saxena, A., Celaya, J., and Goebel, K. (2015). Probabilistic fatigue damage prognosis of lap joint using bayesian updating. *Journal of Intelligent Material Systems and Structures*, 26(8):965–979.

Petit, S. J., Bect, J., Feliot, P., and Vazquez, E. (2023). Parameter selection in gaussian process interpolation: An empirical study of selection criteria. *SIAM/ASA Journal on Uncertainty Quantification*, 11(4):1308–1328.

Pinciroli, 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*. PhD thesis, Université Aix-Marseille.

Prusek, T., Moleiro, E., Oukacine, F., Adobes, A., Jaeger, M., and Grandotto, M. (2013). Deposit models for tube support plate flow blockage in Steam Generators. *Nuclear Engineering and Design*, 262:418–428.

Quarteroni, A., Manzoni, A., and Negri, F. (2016). *Reduced Basis Methods for Partial Differential Equations*, volume 92. Springer International Publishing.

Quarteroni, A. M. and Valli, A. (2008). *Numerical Approximation of Partial Differential Equations*. Springer Publishing Company, Incorporated, 1st ed. 1994. 2nd printing edition.

Raissi, M., Perdikaris, P., and Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707.

Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning*. The MIT Press, Massachusetts.

Rivas, A., Delipei, G. K., and Hou, J. (2025). A system predictive maintenance framework for advanced reactors using a data-driven digital twin. *Nuclear Science and Engineering*, 199(3):358–387.

Romano, Y., Patterson, E., and Candes, E. (2019). Conformalized quantile regression. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32, pages 1–11.

Rubinstein, R. and Kroese, D. (2011). *Simulation and the Monte Carlo Method*. Wiley Series in Probability and Statistics. Wiley.

Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo Method*. John Wiley & Sons, New Jersey, 2nd edition.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Salsana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Wiley.

Sankarasubramanian, S. and Krishnamurthy, B. (2012). A capacity fade model for lithium-ion batteries including diffusion and kinetics. *Electrochimica Acta*, 70:248–254.

Shi, Z. and Chehade, A. (2021). A dual-lstm framework combining change point detection and remaining useful life prediction. *Reliability Engineering and System Safety*, 205:107257.

Sobol', I. M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling & Computational Experiment (Engl. Transl.)*, 1(4):407—414.

Srikantiah, G. and Chappidi, P. (2000). Particle deposition and fouling in PWR steam generators. *Nuclear Engineering and Design*, 200(1-2):285–294.

Sriperumbudur, B. K., Fukumizu, K., and Lanckriet, G. R. G. (2011). Universality, Characteristic Kernels and RKHS Embedding of Measures. *Journal of Machine Learning Research*, 12:2389–2410.

Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559.

Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7):964–979.

Sudret, B. and Der Kiureghian, A. (2002). Comparison of finite element reliability methods. *Probabilistic Engineering Mechanics*, 17(4):337–348.

Sullivan, T. J. (2015). *Introduction to uncertainty quantification*. Texts in Applied Mathematics, 63. Springer-Verlag.

Tayade, A., Patil, S., Phalle, V., Kazi, F., and Powar, S. (2019). Remaining useful life (rul) prediction of bearing by using regression model and principal component analysis (pca) technique. *Vibroengineering PROCEDIA*, 23:30–36.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Tripathy, R. K. and Bilionis, I. (2018). Deep uq: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *Journal of Computational Physics*, 375:565–588.

Tuo, R. and Wu, C. F. J. (2015). Efficient calibration for imperfect computer models. *The Annals of Statistics*, 43(6):2331 – 2352.

US-NRC (2021). Letter Report - TLR-RES/DE/REB-2021-17, Technical Challenges and Gaps in Digital Twin Enabling Technologies for Nuclear Reactor Applications.

US-NRC (2023). Letter Report - TLR-RES/DE/REB-2023-02,State-of-Technology and Technical Challenges in Advanced Sensors, Instrumentation, and Communication to Support Digital Twin for Nuclear Energy Application.

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., and Wu, B. (2006). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. John Wiley & Sons.

Vovk, V. (2001). Competitive On-Line Statistics. *International Statistical Review / Revue Internationale de Statistique*, 69(2):213.

Vovk, V. (2015). Cross-conformal predictors. *Annals of Mathematics and Artificial Intelligence*, 74(1–-2):9–28.

Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York.

Weibull, W. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18:293–297.

Wöhler, A. (1858). Bericht fiber die versuche, welche auf der königl niederschlesisch-miirkischen eisenbahn mit apparaten zum messen der biegung und verdehung von eisenbahnwagenachsen während der fahrt angestellt wurden. *Zeitschrift für Bauwesen*, 8:641–652.

Wöhler, A. (1860). Versuche zur ermittlung der auf die eisenbahnwagenachsen einwirkenden kräfte und die widerstandsfähigkeit der wagen-achsen. *Zeitschrift für Bauwesen*, 10:583–616.

Xiu, D. (2010). *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press.

Xiu, D. and Hesthaven, J. S. (2005). High-Order Collocation Methods for Differential Equations with Random Inputs. *SIAM Journal on Scientific Computing*, 27(3):1118–1139.

Xiu, D. and Karniadakis, G. E. (2002). The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24:271–280.

Xue, Y., Zhang, B., Su, K., Li, Y., Zhu, H., and Pan, H. (2025). A preliminary study of digital twin for nuclear reactor dynamics: a synergy of machine learning and model predictive control. *Engineering Applications of Artificial Intelligence*, 153:110940.

Yang, G., Pointeau, V., Tevissen, E., and Chagnes, A. (2017). A review on clogging of recirculating steam generators in pressurized-water reactors. *Progress in Nuclear Energy*, 97:182–196.

Zaffran, M., Feron, O., Goude, Y., Josse, J., and Dieuleveut, A. (2022). Adaptive conformal predictions for time series. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 25834–25866. PMLR.

Zhang, X., Hou, J., Zimmermann, H., and Xiang, Z. (2022). Preliminary Research on Digital Twin of Main Thermodynamic Systems in Nuclear Power Plant for Thermal Performance Monitoring. In *Innovative and Smart Nuclear Power Plant Design*, volume 12 of *International Conference on Nuclear Engineering*, page V012T12A005.

Zhao, D. and Guan, F. (2022). Research and application innovation of digital twin technologies for key equipment of nuclear power plants. In *Innovative and Smart Nuclear Power Plant Design*, volume 12 of *International Conference on Nuclear Engineering*, page V012T12A044.

Zhao, F., Tian, Z., et al. (2018). An integrated prognostics method for failure time prediction of gears subject to the surface wear failure mode. *IEEE Transactions on Reliability*, 67(1):316 – 327.

Zhao, Z., Liang, B., Wang, X., and Lu, W. (2017). Remaining useful life prediction of aircraft engine based on degradation pattern learning. *Reliability Engineering & System Safety*, 164:74–83.

Zio, E. (2012). *An Introduction to the Basics of Reliability and Risk Analysis*. World Scientific.

Zárate, B. A., Caicedo, J. M., Yu, J., and Ziehl, P. (2012). Bayesian model updating and prognosis of fatigue crack growth. *Engineering Structures*, 45:53–61.