



**HAL**  
open science

# Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators

Edgar Jaber, Vincent Chabridon, Emmanuel Remy, Michael Baudin, Didier Lucor, Mathilde Mougeot, Bertrand Iooss

► **To cite this version:**

Edgar Jaber, Vincent Chabridon, Emmanuel Remy, Michael Baudin, Didier Lucor, et al.. Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators. 2024. hal-04381960v1

**HAL Id: hal-04381960**

**<https://hal.science/hal-04381960v1>**

Preprint submitted on 10 Jan 2024 (v1), last revised 13 Mar 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Sensitivity Analyses of a Multi-Physics Long-Term Clogging Model For Steam Generators

Edgar Jaber<sup>1,2,3\*</sup>, Vincent Chabridon<sup>1</sup>, Emmanuel Remy<sup>1</sup>,  
Michael Baudin<sup>1</sup>, Didier Lucor<sup>3</sup>, Mathilde Mougeot<sup>2</sup>,  
Bertrand Iooss<sup>1</sup>

<sup>1</sup>EDF R&D, 6 Quai Watier, 78401, Chatou, France.

<sup>2</sup>Paris-Saclay University, CNRS, Laboratoire Interdisciplinaire des  
Sciences du Numérique, 91405, Orsay, France.

<sup>3</sup>Paris-Saclay University, CNRS, ENS Paris-Saclay, Centre Borelli,  
91190, Gif-sur-Yvette, France.

\*Corresponding author(s). E-mail(s): [edgar.jaber@edf.fr](mailto:edgar.jaber@edf.fr);

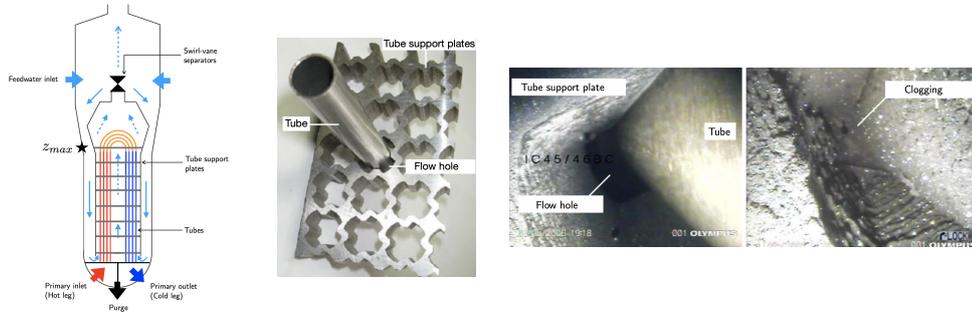
## Abstract

Long-term operation of nuclear steam generators can result in the occurrence of clogging, a deposition phenomenon that may increase the risk of mechanical and vibration loadings on tube bundles and internal structures as well as potentially affecting their response to hypothetical accidental transients. To manage and prevent this issue, a robust maintenance program that requires a fine understanding of the underlying physics is essential. This study focuses on the utilization of a clogging simulation code developed by EDF R&D. This numerical tool employs specific physical models to simulate the kinetics of clogging and generates time dependent clogging rate profiles for particular steam generators. However, certain parameters in this code are subject to uncertainties. To address these uncertainties, Monte Carlo simulations are conducted to assess the distribution of the clogging rate. Subsequently, polynomial chaos expansions are used in order to build a metamodel while time-dependent Sobol' indices are computed to understand the impact of the random input parameters throughout the whole operating time. Comparisons are made with a previous published study and additional Hilbert-Schmidt independence criterion sensitivity indices are computed. Key input-output dependencies are exhibited in the different chemical conditionings and new behavior patterns in high-pH regimes are uncovered by the sensitivity analysis. These findings contribute to a better understanding of the clogging phenomenon while opening future lines of modeling research and helping in robustifying maintenance planning.

**Keywords:** Global Sensitivity Analysis, Polynomial Chaos Expansions, Sobol' indices, Hilbert-Schmidt Independence Criterion, Clogging, Steam Generators

## 1 Introduction

Steam generators (SGs) are heat exchangers within pressurized water nuclear reactors (PWRs). PWRs are composed of two distinct water circuits responsible for the heat exchange. In this system, water from the primary circuit initially passes through the pressure vessel where it is heated due to nuclear reactions in the core. After this, it proceeds to the SG where it transfers its heat to the water in the secondary loops (mainly, three or four loops, hence three or four SGs). The resulting steam exits the SGs through their upper openings and then flows through turbines to generate electrical power. In this process, the primary fluid circulates within a bundle of U-shaped tubes, which are stabilized by tube-support plates (TSPs) (a general SG scheme and TSP is shown in Fig. 1). Adjacent to these tubes, the secondary fluid flows through the holes of the TSP and undergoes vaporization due to the heat from the primary fluid in the tubes.



**Fig. 1** SG scheme, TSP and example of video examination during an PWR outage (© EDF).

After prolonged periods of operation, a clogging phenomenon may appear in the flow-holes of the TSPs of some SGs. Clogging refers to a reduction in the flow area within the TSPs due to the accumulation of iron oxide deposits. These deposits primarily result from the corrosion of the secondary circuit, after which the oxides are carried by the secondary flow. This phenomenon has diverse impacts on the SGs, in particular it causes a localized redistribution of flow among the TSP flow-holes, elevating the risk of vibration, SG tube ruptures, and hampering the SGs' response to operational changes. To address this issue, the clogging rate (denoted as  $\tau_c$ ) is evaluated during PWR outages using non-destructive video examinations (an example is shown in Fig. 1). The clogging rate is an average quantity that gives a measure of how much the holes of a TSP are clogged. Subsequently, preventive maintenance consisting in chemical cleanings procedures are carried out to reduce the clogging rate and the secondary fluid pH can be supervised by chemical conditioning.

To achieve better control of this phenomenon and better plan the maintenance, EDF R&D has developed a physico-chemical clogging model and numerical simulation chain known as “THYC-Puffer-DEPOTHYC” (Feng et al., 2023; Prusek, 2012), that will be called “THYC-Puffer-DEPO” (TPD) in the rest of this paper for the sake of clarity (and to avoid any confusion between the THYC computer code and the DEPOTHYC one). This model is calibrated with respect to the real in-situ data by determining the optimal value of a scalar calibration parameter named  $a_v$  that will be detailed later (see equation (2)). This computational chain is a spatio-temporal simulator that takes a large number of inputs in order to fully model an entire SG and all the U-shaped tubes as shown in Fig. 1. However, among all these inputs, most of them are related to geometric quantities describing the SG and are thus purely deterministic, while only a few of them are tainted by uncertainties, either due to physical variability, or because of lack of knowledge about these physical quantities.

In accordance with the generic probabilistic framework for the treatment of uncertainties (De Rocquigny, Devictor, & Tarantola, 2008), input probability distributions have been elicited by expert judgment in order to quantify and propagate these uncertainties. Thus, a first study has been proposed by (Lefebvre et al., 2023). In this work, several techniques have been applied, from uncertainty propagation through Monte Carlo sampling, surrogate modeling based on an artificial neural network, global sensitivity analysis using Sobol’ indices (Sobol’, 1993, 2001) and Bayesian calibration of the  $a_v$  parameter to enhance and update the sensitivity analysis’ results. This first work enabled to open the path to some sound and relevant thoughts about the computational model and the input physical variables involved. However, this previous study (Lefebvre et al., 2023) was conducted on an earlier version of the computational chain (named “COLMATHYC”) that did not account for the long temporal evolution of clogging.

The primary objective of the present paper is to adopt a rather similar approach as the one proposed in (Lefebvre et al., 2023) (without going up to the Bayesian calibration step in the present study) while tackling several major differences with this previous work: firstly, a new version of the computational chain providing a long time-dependent evolution of the clogging rate; secondly, to try to incorporate the chemical conditioning (which is a key information arising from the operating strategy and that may drastically affect the clogging dynamics); and thirdly, to develop a fully “given-data” methodology in order to estimate complementary sensitivity measures that might be useful for a better understanding of the clogging phenomenon, its better monitoring and a more adapted predictive maintenance. All in all, these uncertainty studies, coupled with new future data, could help enhancing the whole computational chain (Cheng, 2023). The term “given-data” here means that the whole methodology relies on a single input-output dataset of Monte Carlo simulations. Such a framework is often encountered in industry as such complex simulation codes are expensive-to-evaluate and thus require to be deployed on high performance computing facilities (clusters) in order to make studies be tractable. Therefore, the main difficulty is to collect the maximum amount of relevant information (e.g., in order to build an efficient surrogate model or to perform sensitivity analysis) using this finite sample and dedicated estimators. The aim of global sensitivity analysis (GSA) is to study how

the output variability can be apportioned to the variability of the inputs (Da Veiga, Gamboa, Iooss, & Prieur, 2021; Saltelli et al., 2008). These two core topics, i.e surrogate modeling and GSA, play a central role in the UQ framework and have been widely studied in the literature. For instance, one can mention the works of (Heuveline & Schick, 2014) for the use of PCE applied to stochastic dynamical systems and (Schöbi, Sudret, & Wiart, 2015) for an efficient combination between PCE and Kriging in order to take advantage of both types of surrogate models. Regarding GSA, much effort has been put on providing robust estimators of Sobol’ indices (Janon, Nodet, & Prieur, 2014; Lamboni, 2016; Puy, Becker, Lo Piano, & Saltelli, 2022). The present work tries to exploit some rather standard tools based on polynomial chaos expansions (PCE) based surrogate model in order to estimate time-dependent Sobol’ indices, but also more recent kernel-based indices based on the so-called Hilbert-Schmidt independence criterion (HSIC) indices (Da Veiga, 2015; De Lozzo & Marrel, 2016; Gretton, Bousquet, Smola, & Schölkopf, 2005; Marrel & Chabridon, 2021) whose formulation is purely dedicated to the given-data framework.

The present paper is organized as follows. In Section 2, the physical and numerical modeling of the clogging phenomenon is briefly presented as well as the probabilistic modeling of the input uncertainties. Numerical results obtained from Monte Carlo simulations are presented and analyzed as a starting point of the proposed study. Then, Section 3 presents the PCE surrogate model and its ability, in a time-dependent context, to easily provide Sobol’ indices while putting an emphasis on the robust and accurate design of such a metamodel. Section 4 proposes to apply various types of HSIC-based indices (global, target and conditional) in order to capture some complementary information about time-dependent sensitivities. All the numerical results are discussed and analyzed regarding the current available knowledge in the literature. Finally, Section 5 draws some conclusions about the proposed tools and gives some perspectives for future work.

## 2 Physics and numerics of the clogging phenomenon in steam generators

### 2.1 Overview of the physical modeling

The clogging model proposed and described in (Feng et al., 2023; Prusek, 2012) is decomposed into three main steps. These steps are naturally associated with the different time scales in the clogging phenomenon. First, the two-phase flow of the secondary fluid inside the SG is described by a system of conservation laws for mass, momentum and energy. A stationary state of this system is then computed. Among these stationary thermal hydraulic quantities, there is  $\rho_m$ ,  $C_\ell$  and  $U_\ell$  which are respectively, the density of the liquid-gas mixture, the liquid quality and the velocity of the liquid phase. The liquid phase velocity field serves as a flux to a system of two transport equations governing the mass fractions of iron oxide  $\Gamma_p, \Gamma_s$  in its different chemical states (solid particles and soluble particles, respectively). To these transport equations, a solubility map of iron-oxide  $\Gamma_s^{\max}$  inside the SG is provided as well as initial conditions  $\Gamma_p(0), \Gamma_s(0)$ . Finally the third phenomenon is a local balance equation for the

time variation of the mass of iron oxide  $m_c$  with fluxes  $\Phi_p, \Phi_s$  linked respectively to the *vena contracta* and the *flashing* mechanisms (Prusek, 2012).

The clogging model boils down to the following system of mixed partial and ordinary differential equations (with the two initial conditions on the last line):

$$\begin{cases} \partial_t \tilde{\Gamma}_p + U_\ell \cdot \nabla \tilde{\Gamma}_p = f_p(\Phi_p, \Phi_s, \tilde{\Gamma}_s, \Gamma_s^{\max}) \\ \partial_t \tilde{\Gamma}_s + U_\ell \cdot \nabla \tilde{\Gamma}_s = f_s(\Phi_p, \Phi_s, \tilde{\Gamma}_s, \Gamma_s^{\max}) \\ \frac{dm_c}{dt} = \Phi_p + \Phi_s \\ \Gamma_s(0, \cdot) = \Gamma_s(0), \Gamma_p(0, \cdot) = \Gamma_p(0) \end{cases}, \quad (1)$$

where, for  $u = \{s, p\}$ , one has  $\tilde{\Gamma}_u = \rho_m C_\ell \Gamma_u$  and where  $f_u$  are essentially additive functions of their variables. As for the particle flux related to the *vena contracta* mechanism which is of interest here, it is given by (Prusek, 2012):

$$\Phi_p = a_v \frac{k_v (\rho_p - \rho_\ell) U_z^2 d_p^2}{\mu_\ell} \tilde{\Gamma}_p, \quad (2)$$

where  $a_v$  is the calibration non-dimensional parameter of the *vena contracta* mechanism,  $k_v$  is a geometric parameter of the flow-hole,  $\rho_p, \rho_\ell$  are the densities of particles and of the liquid phase respectively,  $U_z := U_m \cdot e_z$  is the vertical component of the homogeneous two-phase mixture velocity,  $\mu_\ell$  is the dynamic viscosity and  $d_p$  is the diameter of the particles. Once the system above is solved, the mass  $m_c$  is multiplied by the volumic mass of iron oxide to obtain the volume  $V_c$ . Finally the average clogging rate  $\tau_c$  is obtained through the following empirical correlation:

$$\tau_c = \alpha (1 - \exp\{-\beta V_c\}). \quad (3)$$

## 2.2 Overview of the numerical model

For the numerical model description, computer codes are denoted by functions  $g_* : \mathbf{x}_* \mapsto g_*(\mathbf{x}_*)$  where  $*$  stands for the name of the code and  $\mathbf{x}_*$  is its vector of *deterministic* input variables. In case the input vector is *random*, capital bold letters  $\mathbf{X}_*$  will be used. The output will sometimes be written  $\mathbf{Y}_* = g_*(\mathbf{X}_*)$ .

The stationary state of the conservation laws for the two-phase mixture is obtained using the THYC software developed by EDF R&D (David, 1999), denoted by  $g_{\text{THYC}}$  in the following. It is based on a numerical finite-volume scheme using a porous-medium approach to provide results at the component scale. Once the thermal hydraulic state  $\mathbf{Y}_{\text{THYC}}$  is computed, the  $g_{\text{DEPO}}$  code is used to solve the transport-balance equations. The output variable of interest is a *clogging rate increment* between two instants  $t < t'$ :

$$\Delta_t^{t'} \tau_c := \tau_c(t') - \tau_c(t). \quad (4)$$

Moreover, to better control the clogging rate, the pH of the secondary circuit may be modified by addition of basic organic solvents. This action is named *chemical conditioning* and the solubility maps of iron oxide have to be computed accordingly.

This is done with the code  $g_{\text{Puffer}}$  that provides the iron oxide solubility maps  $\Gamma_s^{\max}$  for a given SG geometry, thermal hydraulic state and chemical conditioning.

For long periods of simulation (e.g., simulating more than 10 years of the SG operating time), the stationary state for the thermal hydraulic quantities may be not physically valid anymore. Indeed the clogging deposits build additional head-losses that need to be taken into account. This is why a coupling has to be made between the three preceding codes which gives rise to the global architecture of the THYC-Puffer-DEPO simulation chain presented in Fig. 2.

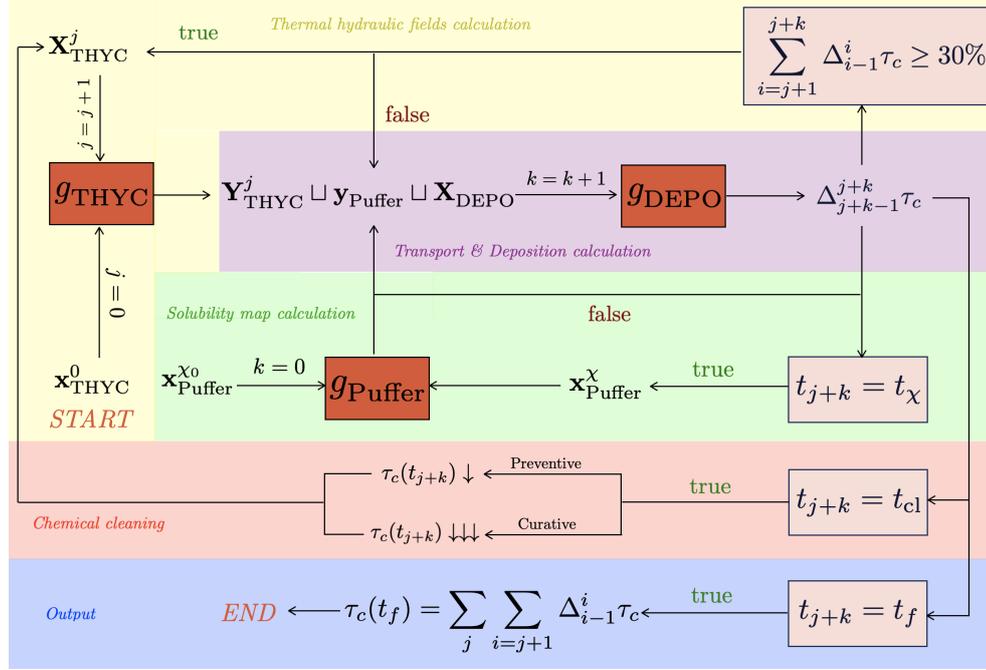


Fig. 2 THYC-Puffer-DEPO code architecture.

The code is initialized at iteration  $j = 0$  by computing the stationary thermal hydraulic state  $\mathbf{y}_{\text{THYC}}^0$ . The code  $g_{\text{THYC}}$  is supposed to have no uncertainties on its input variables. The vector  $\mathbf{x}_{\text{THYC}}^0$  contains SG geometry parameters and different physical correlations that are considered well known. After this, the code  $g_{\text{Puffer}}$  provides a solubility map  $\mathbf{y}_{\text{Puffer}}$  for a given chemical conditioning and SG geometry in  $\mathbf{x}_{\text{Puffer}}^0$ . Again,  $g_{\text{Puffer}}$  is supposed to have deterministic input variables. Once the solubility map  $\mathbf{y}_{\text{Puffer}}$  is built, the input vector for the deposit module  $g_{\text{DEPO}}$  is the concatenated outputs of the preceding codes as well as the uncertain input variables  $\mathbf{x}_{\text{DEPO}}$  that will be described further. The latter contains more than just uncertain variables but for the sake of clarity, it is assumed that it contains only the inputs related to the uncertainty analysis performed. The deposit module  $g_{\text{DEPO}}$  will calculate clogging increments  $\Delta_{j+k-1}^{j+k} \tau_c$  between times  $t_{j+k-1}$  and  $t_{j+k}$  after which

several conditions are checked. If the total amount of clogging between  $t_j$  and  $t_{j+k}$  is greater than 30%, then the head-losses have to be updated and the stationary thermal hydraulic states recalculated. The inputs  $\mathbf{X}_{\text{THYC}}^j$  are updated and some of its components become random given the uncertainty on the output of  $g_{\text{DEPO}}$ . If there is a modification of the chemical conditioning at time  $t_\chi$ , then there is a recalculation of the solubility maps. If  $t_{j+k}$  corresponds to a maintenance cleaning date of the SG, then the clogging rate is lowered according to the type of cleaning performed and  $g_{\text{THYC}}$  is relaunched. Finally, if  $t_{j+k}$  corresponds to the ending time of the simulation,  $t_f$ , the code provides the variation curve of the clogging rate  $\tau_c$ . For such a computational chain, a typical simulation horizon is 50 years (i.e., the typical operating duration of a SG without considering any other issue leading to a potential replacement). Finally, a unitary call of THYC-Puffer-DEPO takes about 5 hours on an high-performance computing infrastructure.

### 2.3 Uncertainty quantification and propagation

Among the numerous input variables of the code, the ones retained by experts for the uncertainty quantification treatment consists of  $d = 7$  independent input variables:

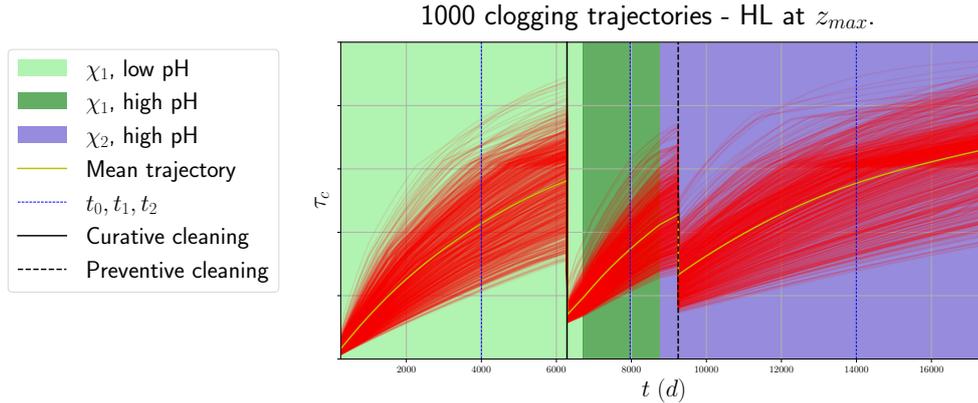
$$\mathbf{X}_{\text{DEPO}} = (\alpha, \beta, \epsilon_e, \epsilon_c, d_p, \Gamma_p(0), a_v). \quad (5)$$

As defined earlier, this vector does not contain all  $g_{\text{DEPO}}$  input variables but only gathers the uncertain ones. In comparison with the study proposed in (Lefebvre et al., 2023), the solubility map  $\Gamma_s^{\text{max}}$  is not considered here. The reason for this is that the previous version of the code - COLMATHYC - did not use  $g_{\text{Puffer}}$  for computing the solubility map and the clogging modeling has evolved since. The probabilistic modeling of these input variables is provided in Table 1. The input probability distributions as well as supports have been set by expert judgment and are in line with the previous study (Lefebvre et al., 2023). The probability laws elicited are either Gaussian ( $X_i \sim \mathcal{N}(\mu, \sigma^2)$  with mean  $\mu$  and variance  $\sigma^2$ ) or triangular ( $X_i \sim \mathcal{T}(a, b, c)$  with mean  $b$  on the interval  $[a, c]$ ). More information concerning the physics of the variables can be found in (Prusek, 2012).

Variable	Signification	Distribution
$\alpha$	First empirical correlation parameter in (3)	$\mathcal{N}(101.6, 4.0)$
$\beta$	Second empirical correlation parameter in (3)	$\mathcal{N}(0.0233, 0.0005)$
$\epsilon_e$	Porosities of the fouling deposits	$\mathcal{T}(0.2, 0.3, 0.5)$
$\epsilon_c$	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 <i>vena contracta</i> physical mechanism	$\mathcal{T}(0.1, 7.8, 12) \times 10^{-4}$

**Table 1** Probabilistic modeling of uncertain input variables.

In order to perform the uncertainty propagation through the THYC-Puffer-DEPO chain, a coupling was built with the OpenTURNS (Open source initiative for the Treatment of Uncertainties, Risks'N Statistics, see <https://openturns.github.io/www/>)

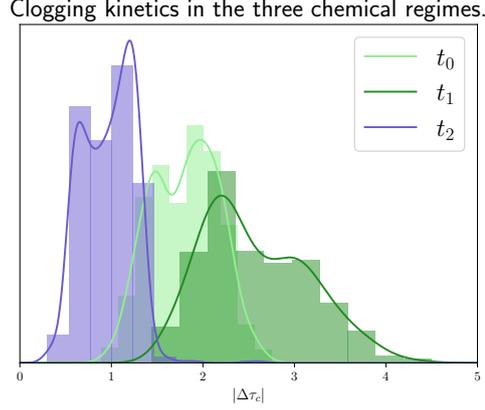


**Fig. 3** Trajectories obtained from Monte Carlo simulations (with curative and preventive chemical cleanings).

software, which is an open source C++/Python library developed by a consortium of industrial partners and public research institutions (Airbus Group, EDF, IMACS, ONERA and Phimeca) (Baudin, Dutfoy, Iooss, & Popelin, 2017).

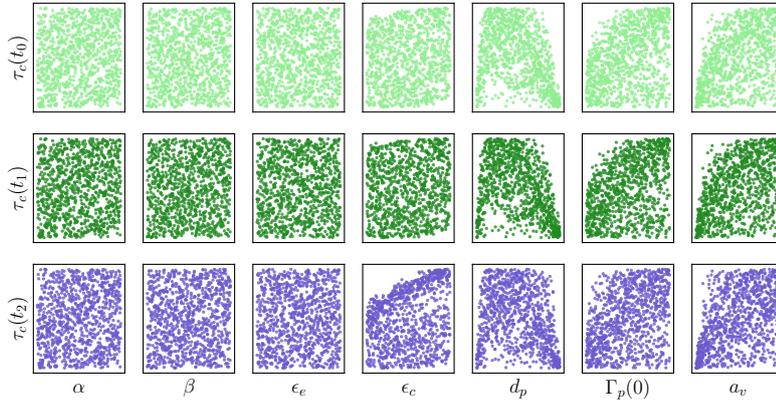
Crude Monte Carlo simulations were executed by drawing  $n = 10^3$  samples according to the previously mentioned joint probability distribution of  $\mathbf{X}_{DEPO}$  defined in Table 1. In order to leverage the computational cost, these simulations were launched in parallel on a cluster, in order to simulate a  $t_f = 50$  years period (approximately, one week on high-performance computing infrastructure for the whole computational process). The resulting trajectories are shown in Fig. 3. The remaining sections of this paper focus on the results obtained for the hot leg (HL) at the uppermost section of the SG in  $z_{max}$  (see the SG scheme in Fig. 1). This choice stems from the fact that the top portion of the HL is known to be the most clogged within the SG and is the primary area inspected by video examinations and therefore used for maintenance planning.

From the simulations in Fig. 3, several remarks can be drawn. Firstly, in the case of this particular SG, two distinct cleaning procedures have been executed all along the operating duration time of the SG: one is a curative treatment while the other one is a preventive cleaning. Typically, operators are aware that preventive cleaning is less effective in reducing clogging rates compared to curative cleaning. Secondly, one can see that the secondary fluid has undergone three distinct chemical conditionings at two different pH levels. These conditionings are denoted by  $\chi_i$ ,  $i \in \{1, 2\}$ , representing the two main chemical species, while the pH is represented as a binary variable with low or high as outcomes. A notable observation is that these various chemical conditionings yield differences in the clogging kinetics. This disparity becomes obvious in the data presented in Fig. 4, which illustrates the distributions of a clogging delta at different times in specific chemical conditionings. The findings reveal a certain deceleration of the simulated clogging kinetics when employing a more basic chemical conditioning. This fact is supported by operational observations.



**Fig. 4** Densities of clogging kinetics for various chemical regimes.

In addition to the time-dependent analysis, some scatter plots illustrating the relationships between inputs and output in the rank-space are provided in Fig. 5, at three distinct time points:  $t_0 < t_1 < t_2$ , which correspond to the three different chemical conditionings.



**Fig. 5** Scatter plots in the rank-space for the three different chemical conditionings.

The dependency structures observed in these scatter plots closely resemble those presented in (Lefebvre et al., 2023). Specifically, a nonlinear correlation is observed between the diameter of iron-oxide particles  $d_p$  and the clogging rate  $\tau_c$  at all the time points. In the physical space, this nonlinear correlation reveals that the particles responsible for high-clogging rates are the ones of intermediate diameter at all chemical conditionings and clogging levels. Additionally, a nonlinear correlation is better seen within at time point  $t_2$  (corresponding to the  $\chi_2$  with high-pH regime) between the clogging rate porosity  $\epsilon_c$  and  $\tau_c$ . It is important to note that this particular correlation is a novel finding compared to the previous study in (Lefebvre et al., 2023) where the

porosity of fouling and clogging were considered to be the same (i.e.,  $\epsilon_e = \epsilon_c = \epsilon$ ). Thus, the sensitivity analysis results found it to be non influential. It is worth pointing out that in this prior study, the pH did not play any important role since the solubility maps were constant on all the SG mesh and distributed according to the uncertainty law chosen. In the more advanced version of the clogging computational chain used in our work, the Puffer code provides a mesh-varying solubility map of iron oxide according to the different chemical conditionings.

Finally, linear correlations are observed between the clogging rate and the initial condition of particle mass-fraction  $\Gamma_p(0)$  as well as for the calibration parameter  $a_v$ . Consistent with the previous study, no significant correlation is identified with the empirical correlation parameters  $\alpha$  and  $\beta$ , as well as the fouling porosity rate  $\epsilon_e$ . These few preliminary visual observations on the input-output dependency-structures will be more accurately and quantitatively studied with the results of the given-data sensitivity analysis in the following sections.

### 3 Variance-based sensitivity analysis using polynomial chaos expansions

#### 3.1 Time-dependent surrogate modeling via polynomial chaos expansions

Given the time-costly nature of the computational chain and the main objective of performing global sensitivity analysis, a polynomial chaos expansions(PCE) meta-model is chosen as a suitable candidate regarding these purposes. Note that, if any other well-known family of metamodels would have been suitable too (e.g., Gaussian processes, artificial neural networks, support vector machines, ...), the use of a PCE metamodel here is motivated by the low-dimensional problem, the relatively low-size training sample ( $n = 10^3$ ) and the robust functional guarantees of these types of approximations (Le Maître & Knio, 2010). On top of that, the ability of this meta-model to provide Sobol' indices as a by-product is a non-negligible advantage when one wishes to perform given-data global sensitivity analysis. Similar methodology can be found in (El Garroussi, Ricci, De Lozzo, Goutal, & Lucor, 2022; Méndez Rojano, Zhussupbekov, Antaki, & Lucor, 2022).

The clogging rate output of the TPD code is now written as a generic function  $g$ :

$$g : (t, z, \mathbf{X}) \in [0, t_f] \times [0, z_{\max}] \times \mathcal{X} \mapsto g(t, z, \mathbf{X}) \in [0, 100] , \quad (6)$$

where  $\mathbf{X} := \mathbf{X}_{\text{DEP}} \in \mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_d \subset \mathbb{R}^d \sim \mathbb{P}_{\mathbf{X}}$  is the vector of inputs in (5),  $\mathcal{X}$  is the cartesian product of the support of the distributions in Table 1 and  $\mathbb{P}_{\mathbf{X}}$  the product measure of these laws (since the input variables are assumed to be independent). For a choice of inputs  $\mathbf{X}_0$ , the output is a function  $(t, z) \mapsto g(t, z, \mathbf{X}_0)$ . As already mentioned, all the simulation results will be considered at the specific altitude  $z_{\max}$  which corresponds to the top of the highest TSP in the HL of the SG. Therefore the output is a random vector of dimension equal to the time discretization  $N$ :

$$(g(t_1, z_{\max}, \mathbf{X}_0), \dots, g(t_N, z_{\max}, \mathbf{X}_0)) \in \mathbb{R}^N , \quad (7)$$

where in this case  $t_N = t_f$ . It is assumed the vector model lies in  $L^2_{\mathbb{P}_{\mathbf{X}}}(\mathcal{X}; \mathbb{R}^N)$ . A Hilbert basis  $\{\varphi_{\alpha}\}_{\alpha \in \mathbb{N}^d}$  made of tensorized orthonormal polynomials is then used (Xiu & Karniadakis, 2002). 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 is implemented in OpenTURNS). At each time step, the PCE of highest degree  $p \in \mathbb{N}$  is constructed as:

$$\tilde{g}(\mathbf{X}) = \sum_{\alpha \in \mathcal{J}} g_{\alpha} \varphi_{\alpha}(\mathbf{X}), \quad (8)$$

where  $g_{\alpha} = (g_{\alpha}^1, \dots, g_{\alpha}^N) \in \mathbb{R}^N$  for all  $\alpha \in \mathcal{J} \subseteq \{\alpha \in \mathbb{N}^d, \sum_{i=1}^d |\alpha_i| \leq p\}$ . The index set  $\mathcal{J} = \mathcal{J}_1 \cup \dots \cup \mathcal{J}_N$  is determined by performing, for each  $k \in \{1, \dots, N\}$ , the following two-step method (Blatman & Sudret, 2009), (Blatman & 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  $\mathcal{G}^k := \{g_{\alpha}^k\}_{\alpha \in \mathcal{J}_k}$  are calculated by solving the following least-square problem using the least-angle regression method (Blatman & Sudret, 2011):

$$\mathcal{G}^k = \arg \min_{g^k \in \mathbb{R}^{|\mathcal{J}_k|}} \mathbb{E} \left( \left| g^k(\mathbf{X}) - \sum_{\alpha \in \mathcal{J}_k} g_{\alpha}^k \varphi_{\alpha}(\mathbf{X}) \right|^2 \right). \quad (9)$$

Once the coefficients have been computed at each time-step, the index set is built by interweaving the basis terms.

In the metamodel-building procedure, the quality of the metamodel is measured through the use of the *predictivity coefficient*  $Q^2$  (Fekhari, Iooss, Muré, Pronzato, & Rendas, 2023). Since the problem is time-dependent, this coefficient is estimated at each time step, using a test dataset  $\mathcal{D}^{\text{test}}$  of size  $m < n$  at every time discretization  $(t_1, \dots, t_N)$  over  $[0, t_f]$ . Typically, the output trajectories presented in Fig. 3 are discretized in  $N = 75$  timesteps. Thus, one has:

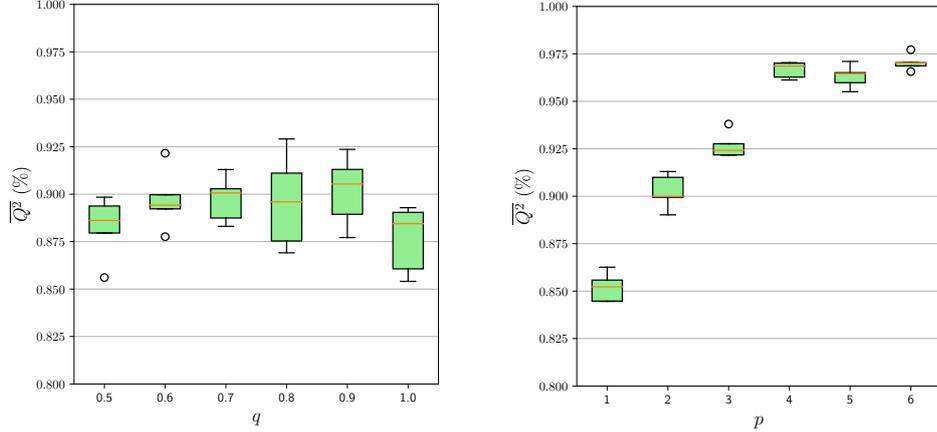
$$\forall k \in \{1, \dots, N\}, Q^2(t_k) = 1 - \sum_{j=1}^m \frac{|g(t_k, \mathbf{X}_j^{\text{test}}) - \tilde{g}(t_k, \mathbf{X}_j^{\text{test}})|^2}{\text{Var}(g(t_k, \mathbf{X}_j^{\text{test}}))}, \quad \overline{Q^2} = \frac{1}{N} \sum_{k=1}^N Q^2(t_k), \quad (10)$$

with  $\overline{Q^2}$  the time-averaged predictivity coefficient. Consequently, the hyperparameters  $q$  and  $p$  are optimized while maximizing the evolution of  $\overline{Q^2}$ .

To select the best hyperparameters, the variability of  $\overline{Q^2}$  with respect to the permutation of the training/test samples is studied. From the available data batch, 75% is used for training and 25% for testing while 5 permutation attempts are made. The  $q$ -norm is studied by fixing a maximal degree of 2. This  $q$ -norm being the number of basis elements retained, it is obvious that, for a full basis, the metamodel could overfit, hence the fall in  $\overline{Q^2}$  observed in Fig. 6. From these results, and in order to minimize

the computation time, a  $q$ -norm of 0.5 is chosen as a good compromise ( $\overline{Q^2} \approx 0.89$ ), since increasing it does not really improve  $\overline{Q^2}$ .

Then, the degree  $p$  is optimized and a satisfactory level of  $\overline{Q^2}$  is obtained for a degree 4 at a  $q$ -norm of 0.5 as shown in Fig. 7 ( $\overline{Q^2} \approx 0.97$ ). Finally, by considering the evolution of the sequence  $Q^2(t_k)$  for  $t_k \in \{t_1, \dots, t_N\}$ , one can clearly see a variation of the coefficient through time, with a decrease in the regions near the chemical cleaning discontinuities. However, globally, the PCE metamodel obtained can be confidently used for performing sensitivity analysis.



**Fig. 6** Boxplots for the  $q$ -norm validation with **Fig. 7** Boxplots for degree validation with  $q = p = 2$ .

## 3.2 Time-dependent Sobol' indices based on polynomial chaos expansion

### 3.2.1 Reminders on the theory

Assuming that, for every  $k \in \{1, \dots, N\}$ , the output component  $g^k \in L^2_{\mathbb{P}_{\mathbf{X}}}(\mathcal{X}, \mathbb{R})$ , one has access to the unique “functional analysis of variance” (FANOVA) decomposition (Hoeffding, 1948; Sobol', 1993, 2001) such that:

$$g^k(\mathbf{X}) = \sum_{\gamma \subseteq \{1, \dots, d\}} g^k_{\gamma}(X_{\gamma_1}, \dots, X_{\gamma_d}) = \sum_{\gamma \subseteq \{1, \dots, d\}} g^k_{\gamma}(\mathbf{X}_{\gamma}). \quad (11)$$

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

$$S_{\gamma}(t_k) = \frac{\text{Var}(g^k_{\gamma}(\mathbf{X}_{\gamma}))}{\text{Var}(g^k(\mathbf{X}))}. \quad (12)$$

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:

$$\mathcal{J}_\gamma = \{ \boldsymbol{\alpha} \in \mathbb{N}^d, |\boldsymbol{\alpha}| \leq p : \alpha_\ell > 0, \forall \ell \in \gamma \text{ \& } \alpha_\ell = 0, \text{ else} \}, \quad (13)$$

the terms in (13) can be regrouped and identified as the unique factors in the FANOVA decomposition such that:

$$g_\gamma^k = \sum_{\boldsymbol{\alpha} \in \mathcal{J}_\gamma} g_\alpha^k \varphi_\alpha(\mathbf{X}). \quad (14)$$

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

$$S_\gamma(t_k) = \frac{\sum_{\boldsymbol{\alpha} \in \mathcal{J}_\gamma} (g_\alpha^k)^2}{\sum_{|\boldsymbol{\alpha}| \leq p} (g_\alpha^k)^2}. \quad (15)$$

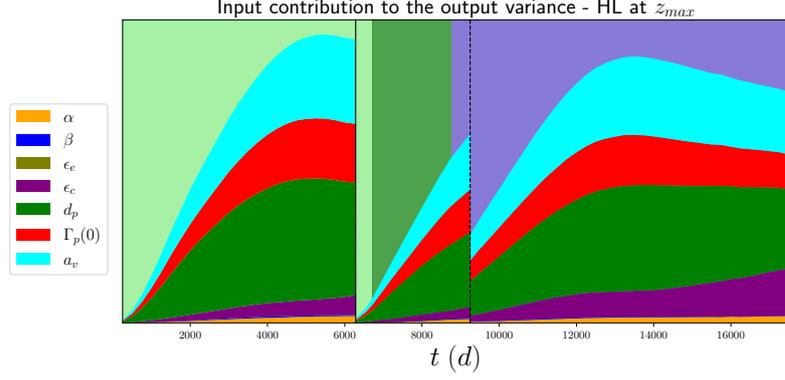
In what follows, special attention is given to the first-order index  $S_i$  and the total-order index  $S_i^T$  for  $i \in \{1, \dots, d\}$ , as well as the global interaction index  $S_*$ , defined by:

$$S_i(t_k) = \frac{\text{Var}(g_i^k(X_i))}{\text{Var}(g^k(\mathbf{X}))}, \quad S_i^T(t_k) = \sum_{\gamma|i \subset \gamma} S_\gamma(t_k) \text{ and } S_*(t_k) = 1 - \sum_{i=1}^d S_i(t_k). \quad (16)$$

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. Moreover, it will be of interest to plot the non-normalized first order indices  $t_k \mapsto \text{Var}(g_i^k(X_i))$ .

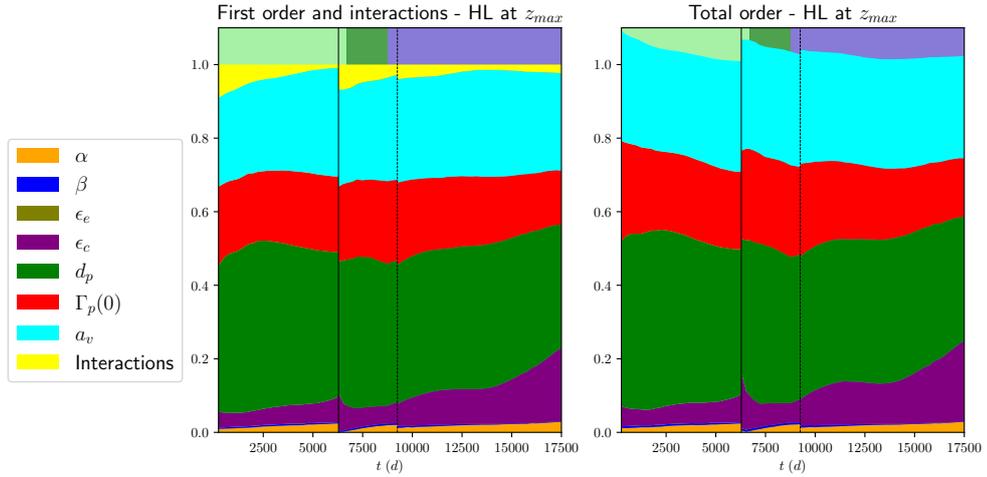
### 3.2.2 Analysis of numerical results on the THYC-Puffer-DEPO chain

In Fig. 8, the temporal evolution of the variance of every input is presented. Likewise, in 9 temporal evolutions of various Sobol' indices computed from the PCE meta-model are provided, regarding the three chemical conditionings which are recalled in the background of the plots. Before delving into the details, several initial observations merit consideration. Notably, alongside the instances of chemical cleaning, there are discernible spikes in the first and total order indices. It is imperative to emphasize that these spikes are a numerical artifact stemming from the construction of the metamodel and its susceptibility to discontinuities within the data. Another remark is that the proposed study corroborates the prior results obtained in (Lefebvre et al., 2023). The predominant portion of the output variance is primarily influenced by a subpart of the inputs, mainly  $d_p$ ,  $a_v$  and  $\Gamma_p(0)$ . It is worth noting that, in contrast to the previous study, the solubility maps  $\Gamma_s^{\max}$  are deterministic in this context due to their construction via the Puffer computer code. Moreover, one can see that all the



**Fig. 8** Evolution of the variance contribution of each uncertain input.

interactions terms have a negligible impact all along time evolution. This result is consistent with the total-order indices (see Fig. 9, on the right) where the indices seem to be almost equal to the first-order indices. As a consequence, one can say that the TPD code behaves here, conditionally to the input probabilistic model given in Table 1, almost as an additive function of the input variables. Moreover, this refined version of the code affords the opportunity to discern the impact of chemical conditioning on the hierarchy of influential variables. Specifically, in the  $\chi_2$  with high-pH regime, there is a notable transfer of variance contribution from  $d_p$  and  $\Gamma_p(0)$  to  $\epsilon_c$ , while the influence of  $a_v$  remains relatively constant across the three chemical conditionings.



**Fig. 9** Evolution of first-order, interactions (left) and total-order (right) time-dependent Sobol' indices.

The porosity rate of clogging deposits, denoted by  $\epsilon_c$ , quantifies the proportion of free space volume to the total volume occupied by the iron oxide deposit. It serves as a measure of how effectively the iron oxide particles are aggregated within the system. It is conventionally assumed to remain constant throughout the SG and is considered independent of the operational conditions. Our discovery presents an intriguing avenue for various physical interpretations. Iron oxides inherently possess charged particles that establish crystalline bonds, primarily Van der Waals interactions, based on their electronic structure. Altering the hydrogen ion charge of the solution can have a profound impact on this electronic structure. This phenomenon is notably exemplified in colloidal aggregation (Liu, Dai, & Hu, 2018). In light of this, it can be inferred that under high-pH conditions, the iron oxide deposits tend to exhibit a greater degree of aggregation. Consequently, this translates into an increased significance of uncertainty in clogging porosity on the system’s output. Furthermore, the stiffening or rigidification of clogging deposits in a high-pH regime could potentially account for the observed deceleration in clogging kinetics. These findings shed light on the intricate interplay between chemical conditions, particle aggregation, and clogging behavior within the system.

## 4 Kernel-based given-data sensitivity analysis

### 4.1 The Hilbert-Schmidt independence criterion for global sensitivity analysis

#### 4.1.1 Reminders on the theory

Another method for performing sensitivity analysis is through the use of kernel methods like the Hilbert-Schmidt independence criterion (HSIC). Introduced in the machine learning community by (Gretton et al., 2005), its use as a powerful importance measure in sensitivity analysis is due to (Da Veiga, 2015). It allows to uncover deep dependency structures between each input variable and the output quantity of interest. The underlying idea is to compare the joint probability distribution  $\mathbb{P}_{X_i, Y}$  with the product of distributions  $\mathbb{P}_{X_i} \mathbb{P}_Y$ . To do so, it makes use of generalized operators in Reproducing Kernel Hilbert Spaces (RKHSs) (Berlinet & Thomas-Agnan, 2004).

In (7), the output space is denoted by  $\mathcal{Y} = \mathcal{Y}_1 \times \dots \times \mathcal{Y}_N = g(t_1, \mathcal{X}) \times \dots \times g(t_N, \mathcal{X}) \subset \mathbb{R}^N$  where  $N$  is the time discretization. Let  $i \in \{1, \dots, d\}$  and  $k \in \{1, \dots, N\}$  be an input parameter index and a time-output index. Let  $\mathcal{F}_i, \mathcal{G}_k$  be two RKHSs of respective kernels  $\kappa_i, \kappa_k$  that are assumed to be characteristic (Sriperumbudur, Fukumizu, & Lanckriet, 2011). On the space  $\mathcal{F}_i \times \mathcal{G}_k$  consider the kernel  $\nu_{ik} = \kappa_i \otimes \kappa_k$ , defined for all  $(X, X') \in \mathcal{X}_i^2, (Y, Y') \in \mathcal{Y}_k^2$  by:

$$\nu_{ik}((X, Y), (X', Y')) = \kappa_i(X, X') \kappa_k(Y, Y'). \quad (17)$$

To alleviate the notations for  $Y_{ik} = g(t_k, X_i)$ , one will consider the pair  $(X_i, Y_{ik})$  which will be written as  $(X_i, Y_k)$ . Finally, the generalized covariance operator is defined as:

$$C_{X_i Y_k} = \mathbb{E}_{X_i, Y_k} [\nu_{ik}((X_i, Y_k), \cdot)] - \mathbb{E}_{X_i} \mathbb{E}_{Y_k} [\nu_{ik}((X_i, Y_k), \cdot)], \quad (18)$$

and the HSIC measure between the input  $X_i$  and the output  $Y_k$  is defined as the Hilbert-Schmidt norm of the generalized covariance operator:

$$\text{HSIC}(X_i, Y_k) = \|C_{X_i Y_k}\|_{\text{HS}}. \quad (19)$$

One theoretical result which strongly motivated the use of HSIC indices for global sensitivity analysis is that (Da Veiga, 2015; Gretton et al., 2005):

$$\text{HSIC}(X_i, Y_k) = 0 \iff X_i \perp Y_k. \quad (20)$$

Therefore, assessing independence between an input and an output can be achieved through the use of HSIC indices.

From a statistical viewpoint, one needs to estimate this quantity from a dataset. Various estimators have been derived, either based on U-statistics or V-statistics (Da Veiga et al., 2021, Ch. 6). For instance, the following trace-based estimator is often used in practice:

$$\widehat{\text{HSIC}}(X_i, Y_k) = \frac{1}{n^2} \text{Tr}(\mathbf{L}_i \mathbf{H} \mathbf{L}_k \mathbf{H}), \quad (21)$$

where  $\mathbf{L}_i, \mathbf{L}_k$  are the Gram-matrices on the samples  $(x_i^{(p)}, y_k^{(p)})_{1 \leq p \leq n}$ , thus given by  $(\mathbf{L}_i)_{pq} = \kappa_i(x_i^{(p)}, y_k^{(q)})$  and  $(\mathbf{L}_k)_{pq} = \kappa_k(x_i^{(p)}, y_k^{(q)})$  and  $\mathbf{H}$  is a shift matrix defined by  $(\mathbf{H})_{pq} = (\delta_{p,q} - \frac{1}{n})_{pq}$  for all  $1 \leq p, q \leq n$ . For a given input-output pair  $(X_i, Y_k)$ , a null-hypothesis can be derived:

$$\mathcal{H}_0 : \text{HSIC}(X_i, Y_k) = 0. \quad (22)$$

Let  $\widehat{S}_T = n \times \widehat{\text{HSIC}}(X, Y)$  be the test statistic and  $\widehat{S}_{T, \text{obs}}$  be the observed statistic. The corresponding  $p$ -value associated to  $\mathcal{H}_0$  is then defined by:

$$p_{\text{val}} = \mathbb{P}(\widehat{S}_T \geq \widehat{S}_{T, \text{obs}} | \mathcal{H}_0). \quad (23)$$

A large  $p$ -value is a sign of independence and in our applications a standard limit value  $p_{\text{val}} = 0.05$  is chosen above which independence achieves statistical significance. It is important to emphasize that it is not a measure of dependency but its non-nullity is a potential indicator of dependence and of influence on the output.

Compared to the Sobol' indices, there is no underlying FANOVA decomposition for the standard formulation of HSIC indices as used in the present paper (see (Da Veiga, 2021) for a deeper discussion). Usually a normalized index is used (Marrel & Chabridon, 2021) but the latter does not sum to one. It is defined by:

$$R_{\text{HSIC}}^2(X_i, Y_k) = \frac{\text{HSIC}(X_i, Y_k)}{\sqrt{(\text{HSIC}(X_i, X_i) \text{HSIC}(Y_k, Y_k))}}. \quad (24)$$

Deriving such FANOVA methods for HSIC indices is an ongoing topic of research (Da Veiga, 2021). The HSIC index only yields a global analysis and is not sensitive

to the interactions between the inputs (as with Sobol' indices). The use of HSIC here can be seen as complementary to the previous analysis in Section 3.

#### 4.1.2 Analysis of time-dependent global HSIC indices

Fig. 10 shows the corresponding  $R_{\text{HSIC}}^2$  indices and the associated  $p$ -values of the statistical tests, carried out on the TPD Monte Carlo sample used previously. 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  $\epsilon_c$  in the high-pH- $\chi_2$  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.

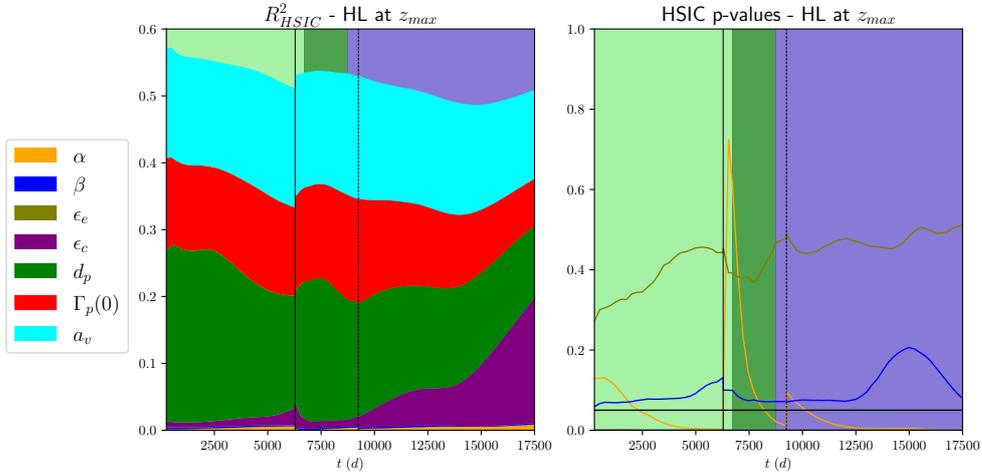


Fig. 10 Time variation of the estimated HSIC indices and  $p$ -values associated to the statistical test of independence.

## 4.2 Target HSIC indices

### 4.2.1 Reminders on the theory

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 & Chabridon, 2021). It makes use of a filter function  $f_{\mathcal{C}} : \mathcal{Y} \rightarrow \mathbb{R}_+$  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 & Chabridon, 2021). The filter function used is defined by:

$$\tilde{Y}_j = f_{\mathcal{C}}(Y_j) = \exp\left(-\frac{\inf_{y \in \mathcal{C}} \|Y_j - y\|}{s\hat{\sigma}_{Y_j}}\right) \in [0, 1], \quad (25)$$

where  $s = 1/5$  and  $\hat{\sigma}_{Y_j}$  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 & Chabridon, 2021). It can be optimized accordingly but this is out of the scope of this work. The V-statistic estimator (20) can still be used with the modified variable to yield the target HSIC indices:

$$\widehat{\text{T-HSIC}}(X_i, Y_j) = \widehat{\text{HSIC}}(X_i, \tilde{Y}_j), \quad (26)$$

and the corresponding  $p$ -value associated to the hypothesis test  $\mathcal{H}_0$  can still be used.

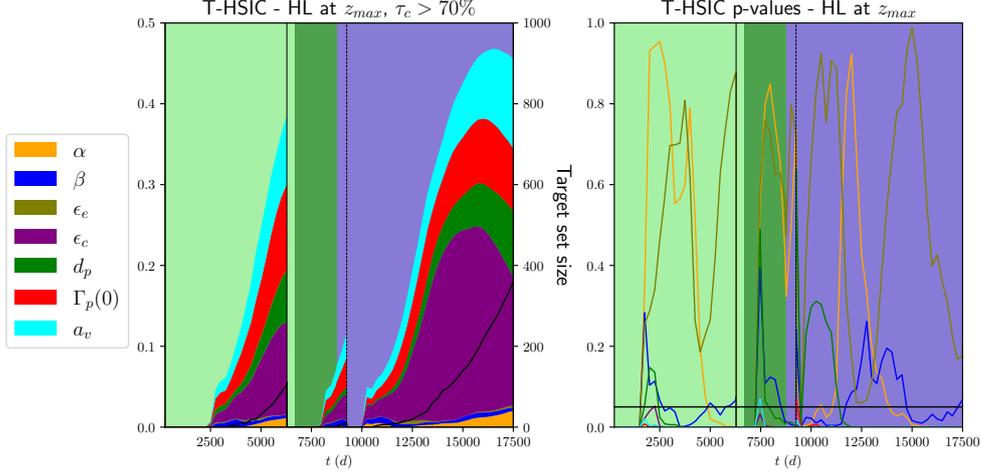
### 4.2.2 Analysis of time-dependent target HSIC indices

A threshold of 70% is chosen as a target region for the clogging simulation dataset. Indeed this highly-clogged regime is critical for the operation of the SG. The associated variation of T-HSIC indices amount is displayed in Fig. 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  $\epsilon_c$  becomes the most influential variable. It should be stressed that this was apparent in the previous analyses in Figs. 9 and 10 in the long-time high-pH  $\chi_2$  conditioning. Indeed it constitutes a new finding that is here quantitatively established with the help of the local target indices.

## 4.3 Conditional HSIC indices

### 4.3.1 Reminders on the theory

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 & Chabridon, 2021).



**Fig. 11** Time variation of the estimated, normalized T-HSIC indices, the associated target set size and the  $p$ -values associated to the statistical test of independence.

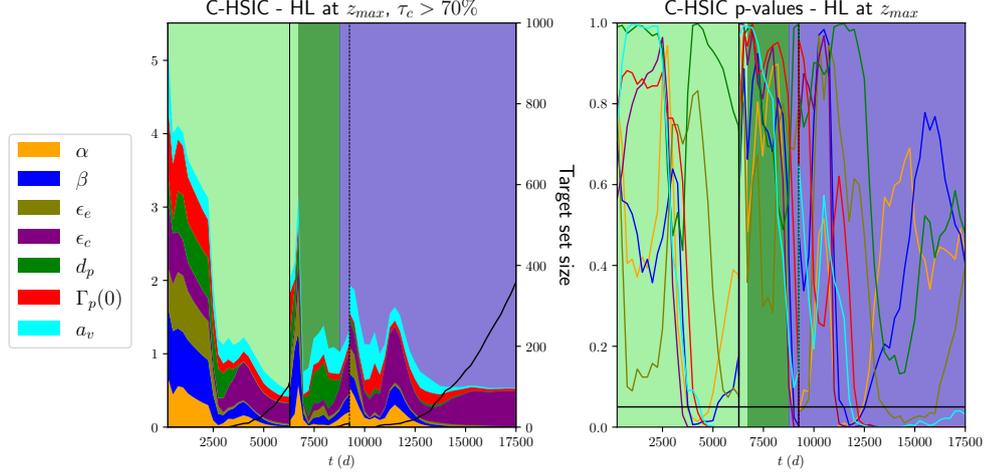
The conditioning formula gives:

$$Y_j | \{Y_j \in \mathcal{C}\} \sim \mathbb{P}_{Y_j}^f, \text{ s.t } \forall \text{ measurable } A \subset \mathcal{Y}, \mathbb{P}_{Y_j}^f(A) = \int_A \frac{f_{\mathcal{C}}(y)}{\mathbb{E}[f_{\mathcal{C}}(Y_j)]} d\mathbb{P}_{Y_j}(y). \quad (27)$$

By conditioning, the underlying input space is also modified; thus, for all  $i \in \{1, \dots, d\}$ ,  $X_i | \{Y_j \in \mathcal{C}\} \sim \mathbb{P}_{X_i}^f$  defined similarly. The joint distribution involved in the general covariance operator (18) is changed accordingly  $(X_i, Y_j) \sim \mathbb{P}_{X_i, Y_j}^f$ . Moreover a V-statistic estimator similar to (21) can be derived.

### 4.3.2 Analysis of time-dependent conditional HSIC indices

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 (25). 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. 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  $\chi_2$  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.



**Fig. 12** Time variation of the estimated, normalized C-HSIC indices and  $p$ -values associated to the statistical test of independence.

## 5 Conclusion and perspectives

In this study, the complex clogging rate phenomenon observed in the steam generators of pressurized water reactors is studied with the help of a panel of surrogate modeling and sensitivity analysis tools. The complex multiphysics THYC-Puffer-DEPOTHYC clogging code, developed by EDF R&D, is used and a complete uncertainty quantification and propagation is performed using Monte Carlo simulations. As a result, a  $n$ -size sample ( $n = 10^3$ ) of output trajectories is obtained for a specific steam generator while simulating the time evolution of the clogging rate over 50 operating years. The analysis of the results enable to recover qualitative results that are in accordance with a previous published study (Lefebvre et al., 2023) which tried to tackle the same phenomenon but with a different computational chain. However, the present study enabled to highlight new results in terms of inputs-output influence. These qualitative results are corroborated by quantitative results obtained via time-dependent Sobol' indices obtained as a by-product of a high-fidelity metamodel based on polynomial chaos expansions. Through these indices, one can understand that the interactions between inputs are relatively weak, leading to a quasi-additive structure of the model. The provided time-dependent Sobol' indices enable to identify and rank the most influential inputs. In addition, the use of another class of indices based on the Hilbert-Schmidt Independence Criterion (HSIC) enable to offer a refined (in terms of the nature of influence, i.e going beyond the output variance) and complementary insight about the input-output relationships. Such indices are versatile since they can offer three types of analyses (global, target and conditional). All in all, the results are still in coherence with the previous ones obtained with the variance-based importance measures, but they shed new light on the ranking in the different output critical domains, with respect to the evolution in time. As a perspective, some improvements could be

achieved through the use of other class of metamodels (see, e.g (Schöbi et al., 2015)) and other sensitivity indices could be investigated (see, e.g (Lüthen et al., 2023)).

## Acknowledgments

The authors would like to thank several people: Dr. Qingqing Feng, Dr. Stéphane Pujet and Morgane Garo Sail (EDF R&D / MFEE Department) for fruitful discussions about the clogging computational model, Jérôme Delplace (EDF Nuclear Division) for discussions about the industrial aspects as well as Dr. Régis Lebrun (Airbus) for fruitful discussions about OpenTURNS and the “Adaptive Stieltjes Algorithm” used in this work. This work is part of a PhD program funded by the French National Association for Technological Research (ANRT) under Grant n° 2022/1412.

## References

- Baudin, M., Dutfoy, A., Iooss, B., Popelin, A. (2017). Open TURNS: An industrial software for uncertainty quantification in simulation. R. Ghanem, D. Higdon, & H. Owhadi (Eds.), *Springer handbook on Uncertainty Quantification* (pp. 2001–2038). Springer.
- Berlinet, A., & Thomas-Agnan, C.M. (2004). *Reproducing Kernel Hilbert Space in Probability and Statistics*. Springer US.
- Blatman, G., & Sudret, B. (2009, 01). Anisotropic parcimonious polynomial chaos expansions based on the sparsity-of-effects principle. *Proc. 10th int. conf. struct. safety and reliability (icossar’2009)*. Osaka, Japan.
- Blatman, G., & Sudret, B. (2011). Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6), 2345-2367,
- Cheng, S.e.a. (2023). Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review. *IEEE/CAA Journal of Automatica Sinica*, 10, 1361-1387,
- 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., Prieur, C. (2021). *Basics and Trends in Sensitivity Analysis. Theory and Practice in R*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

- David, F. (1999, 3-8 October). Three Dimensional Thermal-Hydraulic Simulation In Steam Generators With THYC Exchangers Code - Application To The UTSG Model 73/19. San Francisco, CA.
- De Lozzo, M., & Marrel, A. (2016). Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators. *Stochastic Environmental Research and Risk Assessment*, *31*, 1437–1453,
- De Rocquigny, E., Devictor, N., & Tarantola, S. (Eds.). (2008). *Uncertainty in industrial practice - a guide to quantitative uncertainty managment*. Wiley and Sons.
- El Garroussi, S., Ricci, S., De Lozzo, M., Goutal, N., 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,
- Fekhari, E., Iooss, B., Muré, J., Pronzato, L., Rendas, J. (2023). Model predictivity assessment: incremental test-set selection and accuracy evaluation. N. Salvati, C. Perna, S. Marchetti, & R. Chambers (Eds.), *Studies in theoretical and applied statistics, sis 2021, pisa, italy, june 21-25* (p. 315-347). Springer.
- Feng, Q., Nebes, J., Bachet, M., Pujet, S., You, D., Deri, E. (2023, September). Tube support plates blockage of PWR steam generators: thermohydraulics and chemical modeling. Juan-les-Pins.
- Gautschi, W. (2004). *Orthogonal Polynomials: Computation and Approximation*. Oxford University Press.
- Gretton, A., Bousquet, O., Smola, A., Schölkopf, B. (2005). Measuring statistical dependence with Hilbert-Schmidt norms. *Proceedings algorithmic learning theory* (pp. 63–77). Springer-Verlag.
- Heuveline, V., & Schick, M. (2014). A hybrid generalized polynomial chaos method for stochastic dynamical systems. *International Journal for Uncertainty Quantification*, *4*, 37–61,
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, *19*(3), 293–325,
- Janon, A., Nodet, M., Prieur, C. (2014). Uncertainties assessment in global sensitivity indices estimation from metamodels. *International Journal for Uncertainty Quantification*, *4*, 21–36,

- Lamboni, M. (2016). Global sensitivity analysis: an efficient numerical method for approximating the total sensitivity index. *International Journal for Uncertainty Quantification*, 6, 1–17,
- Lefebvre, L., Segond, M., Spaggiari, R., Le Gratiet, L., Deri, E., Iooss, B., 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,
- Le Maître, O.P., & Knio, O.M. (2010). *Spectral Methods for Uncertainty Quantification: With Applications to Computational Fluid Dynamics*. Springer.
- Liu, J., Dai, C., 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,
- Lüthen, N., Roustant, O., Gamboa, F., Iooss, B., Marelli, S., Sudret, B. (2023). Global sensitivity analysis using derivative-based sparse Poincaré Chaos Expansions. *International Journal for Uncertainty Quantification*, 13, 57–82,
- Marrel, A., & 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,
- Méndez Rojano, R.M., Zhussupbekov, M., Antaki, J.A., 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,
- Prusek, T. (2012). *Modélisation et simulation numérique du colmatage à l'échelle du sous-canal dans les générateurs de vapeur*. Thèse de l'Université Aix-Marseille.
- Puy, A., Becker, W., Lo Piano, S., Saltelli, A. (2022). A comprehensive comparison of total-order estimators for global sensitivity analysis. *International Journal for Uncertainty Quantification*, 12, 1–18,
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer*. Wiley.

- Schöbi, R., Sudret, B., Wiart, J. (2015). Polynomial-Chaos-based Kriging. *International Journal for Uncertainty Quantification*, 5, 171–193,
- Sobol', I.M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modeling & Computational Experiment (Engl. Transl.)*, 1(4), 407–414,
- Sobol', I.M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1), 271-280,
- Sriperumbudur, B.K., Fukumizu, K., Lanckriet, G.R.G. (2011). Universality, Characteristic Kernels and RKHS Embedding of Measures. *Journal of Machine Learning Research*, 12, 2389–2410,
- Sudret, B. (2008). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering and System Safety*, 93(7), 964–979,
- Xiu, D., & Karniadakis, G.E. (2002). The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24, 271–280,