



Article

Risk Monitoring of Small Modular Reactors by Grey-Box Models: Feature Extraction and Global Sensitivity Analysis

Leonardo Miqueles ¹, Ibrahim Ahmed ¹, Francesco Di Maio ^{1,*} and Enrico Zio ^{1,2}

¹ Energy Department, Politecnico di Milano, 20133 Milan, Italy; leonardoandres.miqueles@polimi.it (L.M.); ibrahim.ahmed@polimi.it (I.A.); enrico.zio@polimi.it (E.Z.)

² Centre de Recherche sur les Risques et les Crises (CRC), Department of Economics, Management and Society, MINES Paris-PSL Université Paris, 06904 Sophia Antipolis, France

* Correspondence: francesco.dimaio@polimi.it

Abstract

Gray-Box (GB) models are being considered for risk monitoring of Small Modular Reactors (SMRs). Their effectiveness is linked to the proper selection of the model parameters. This paper proposes a systematic methodology for identifying the most influential parameters of a GB model for estimating safety-critical variables of an SMR during normal operation and accident scenarios. The GB integrates a reduced-order physics-based model (White-Box, WB) with a data-driven (Black-Box, BB) model that corrects the outputs of the WB using the condition-monitoring data collected by sensors positioned onto the SMR. The proposed method combines signal decomposition, specifically the Hilbert–Huang Transform (HHT), and global sensitivity analysis (SA), based on first-order Kucherenko indices, to quantify the contribution of non-stationary, correlated GB input parameters to the variability of the safety-critical output parameters of interest. The proposed approach is applied to the Small Modular Dual Fluid Reactor (SMDFR), and the obtained results demonstrate its effectiveness in identifying informative and physically interpretable features, reducing complexity and computational burden to enable real-time risk monitoring.

Keywords: Small Modular Reactors; risk monitoring; gray-box modeling; global sensitivity analysis; Hilbert–Huang transform; Kucherenko index

1. Introduction

Risk monitoring relies on timely and accurate estimations of the values of safety-critical parameters [1–3]. The compact core design of Small Modular Reactors (SMRs) challenges in-core sensor placement for directly measuring the values of safety-critical parameters for risk monitoring [4,5]. Virtual sensing [6,7] can be adopted for enabling the estimation (and prediction) of safety-critical parameters not directly measurable [8]. For the needed estimation accuracy and interpretability, Gray-Box (GB) models can be used, strategically combining physics-based White-Box (WB) models [9] with data-driven Black-Box (BB) models [7,10,11] to achieve transparency (WB) and computational efficiency (BB) [12,13].

GB models require a proper input parameter selection to obtain the expected benefits of output explainability and generalization capabilities with end-user trust, which are required for application in the nuclear power industry [14–16]. The systematic identification of physically meaningful and informative inputs for GB modeling is a critical issue, especially in the context of SMRs, because (i) input parameters (e.g., fuel and coolant temperature, mass flow rates) may show strong correlations, nonlinear behavior and



Academic Editor: Milica Ilic

Received: 17 February 2026

Revised: 21 April 2026

Accepted: 30 April 2026

Published: 7 May 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

non-stationarity [17–19] that may increase model complexity and reduce the GB model's interpretability; (ii) during transients, the characteristics of the inputs change since their evolution is non-stationary [1,2,20]. Proper feature extraction and sensitivity analysis (SA) techniques are then needed to determine the subset of informative input parameters that most contribute to the variability in the GB model output [20–25]. Various feature extraction techniques have been proposed for BB models, like embedded regularization techniques such as LASSO for sparse variable identification [26], dimensionality reduction approaches such as Principal Component Analysis (PCA) and Partial Least Squares (PLS) [27], and SHapley Additive eXplanatory (SHAP) values for input relevance ranking in regression tasks [28].

In this work, we propose a systematic methodology for identifying and selecting physically relevant, minimally redundant and dynamically informative input parameters to be fed to the BB model of the GB model used within a framework of SMR risk monitoring. The methodology integrates nonlinear time-series feature extraction with a global variance-based SA. The deployment of the approach consists of two steps. First, the Hilbert–Huang Transform (HHT) [29,30], a hybrid empirical model decomposition, is applied to the nonlinear and non-stationary condition-monitoring data to extract Intrinsic Mode Functions and instantaneous spectral features of the monitored parameters. These features serve as a refined representation of the reactor's dynamic behavior under transient and accident scenarios. Second, a global SA using first-order Kucherenko indices [31] is performed to quantify the contribution of each feature of the monitored parameters to the variability of the relevant safety-critical output parameters. Unlike classical Sobol indices [32], Kucherenko indices explicitly address correlated inputs, making them particularly suitable for SMR operational environments. This methodology, as a whole, enables the identification of a physically meaningful and interpretable set of measurable variables for GB model development, thereby improving model transparency and computational efficiency.

An application is shown with reference to the Small Modular Dual Fluid Reactor (SMDFR), a next-generation SMR concept characterized by compact design and, thus, with limited sensor placement feasibility. The obtained results demonstrate the capability of the methodology in identifying a subset of sensitive and physically meaningful features, thereby reducing the GB model's complexity and improving its computational efficiency, rendering feasible its use for risk monitoring purposes; the method is applicable to different designs of Nuclear Power Plants (NPPs), and is most useful for SMRs, whose design challenges in-core sensor placement.

The rest of the paper is structured as follows: in Section 2, the GB model for multi-step-ahead time-series forecasting and risk monitoring is presented; Section 3 introduces the proposed feature extraction and SA methods; Section 4 describes the SMR case study; Section 5 shows the numerical results; and Section 6 concludes the paper with some final remarks.

2. Problem Formulation

We consider the risk monitoring of an SMR in relation to a safety-critical parameter y (e.g., Peak Cladding Temperature (PCT)) which may not exceed a certain threshold value during operation. The dynamics of the non-measurable safety-critical parameter y can be modeled as:

$$y(t) = g(x(t), u(t), \theta) \quad (1)$$

where $x(t) = [x_1(t), x_2(t), \dots, x_m(t), \dots, x_M(t)]^T$ is the vector of M measurable input parameters (e.g., pressure, temperature, flow rate) at time t , $u(t)$ are variables affecting process evolution (e.g., coolant leakage rate during a loss-of-coolant accident) at time t and

θ is the vector of design parameters (e.g., heat transfer area, core geometry, fuel/coolant composition properties).

If considering time as a discrete independent variable, $t = t_0, t_1, \dots, t_n, \dots, t_N$, where $t_0 = 0, t_n - t_{n-1} = \Delta t$ and t_N is the end time of the transient considered, and treating each variable x_m as a time-series signal for which historical observations $x_m(t_{n-r+1} : t_n)$ over a window of length r are collected by sensors in $X(t_{n-r+1} : t_n) \in \mathbb{R}^{M \times r}$, a multi-step-ahead GB model (see Figure 1) [33] can be developed to predict the s -steps-ahead output $\hat{y}_{GB}(t_n : t_{n+s})$ of $y(t_n : t_{n+s}) \in \mathbb{R}^s$:

$$\hat{y}_{GB}(t_n : t_{n+s}) = \hat{y}_{WB}(t_n : t_{n+s}) + \Delta \hat{y}(t_n : t_{n+s}) \quad (2)$$

where $\hat{y}_{WB}$ is the output of a reduced-order physics-based WB model, which is fed by θ , $u(t_n)$ and the projection up to time t_{n+s} estimated by a BB model. The term $\Delta \hat{y}$ is the WB model discrepancy (error correction) between the WB model output and y . The WB model is an approximation of the SMR physics, as follows:

$$\hat{y}_{WB}(t_n : t_{n+s}) = f_{WB}(x(t_n), u(t_n : t_{n+s}), \theta) \quad (3)$$

where $f_{WB}(\cdot)$ retains the essential governing reactor physics, whereas the prediction discrepancy Δy is defined as:

$$\Delta y(t_n : t_{n+s}) = y(t_n : t_{n+s}) - y_{WB}(t_n : t_{n+s}) \quad (4)$$

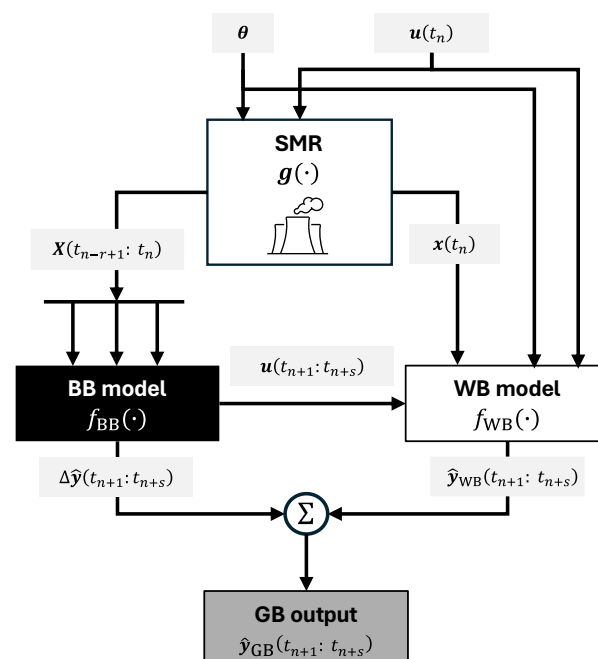


Figure 1. GB modeling for multi-step-ahead estimation of the safety-critical variable y .

In a GB-based framework for risk monitoring [33], $\Delta \hat{y}$ and u can be estimated by a data-driven BB model $f_{BB}(\cdot)$ as follows:

$$Y_{BB}(t_n : t_{n+s}) = f_{BB}(X(t_{n-r+1} : t_n)) \quad (5)$$

where $X(t_{n-r+1} : t_n) \in \mathbb{R}^{M \times r}$ corresponds to the memory matrix and $f_{BB} : \mathbb{R}^{M \times r} \rightarrow \mathbb{R}^{2 \times s}$. The BB model is developed to simultaneously approximate the discrepancy Δy and the variable u (i.e., $Y_{BB} = [\Delta \hat{y}, \hat{u}]^T$) required for the WB model estimation (Equation (3)), based

on r available observations taken from the monitored data. Given the above formulation, the objective of the present work is to identify and select an optimal subset of input variables $\mathbf{x}^* \subseteq \mathbf{x}$ that explain most of the variability in both Δy and u . Selecting such a subset is essential for the efficient implementation of $f_{\text{BB}}(\cdot)$ within the GB model, thereby ensuring reduced computational complexity and enhanced predictive performance for real-time risk monitoring of the SMR.

3. The Proposed Methodology

The proposed methodology is sketched in Figure 2. First, HHT is applied to the time series of the signal measurements in N_d available accident scenarios [29,30]. Without loss of generality, it is assumed that each accident scenario $d \in N_d$ is initiated at a fixed time $t_f \leq t_N$ and lasts until time t_N , with the measurements being taken at discrete time $t = t_0, t_0 + \Delta t, \dots, t_n, \dots, t_N$ with constant timestep Δt . For each scenario d , both measured variables $\mathbf{X}(t_{n-r+1} : t_n)$ and input signals $\mathbf{u}(t_{n+1} : t_{n+s})$ are collected, using fixed r (memory window length) and s (prediction window length) with time index $n = r, \dots, N - s$ (see Figure 3). At each timestep t_n , the SMR model $g(\cdot)$ and the WB model $f_{\text{WB}}(\cdot)$ are evaluated to compute the output vector $\mathbf{Y}_{\text{BB}}(t_n : t_{n+s})$ to be used as the target output of the BB model (see Section 2). Feature extraction is then performed on both the input $\mathbf{X}(t_{n-r+1} : t_n)$ and the target output $\mathbf{Y}_{\text{BB}}(t_n : t_{n+s})$ of the BB model using HTT [30] (see Section 3.1).

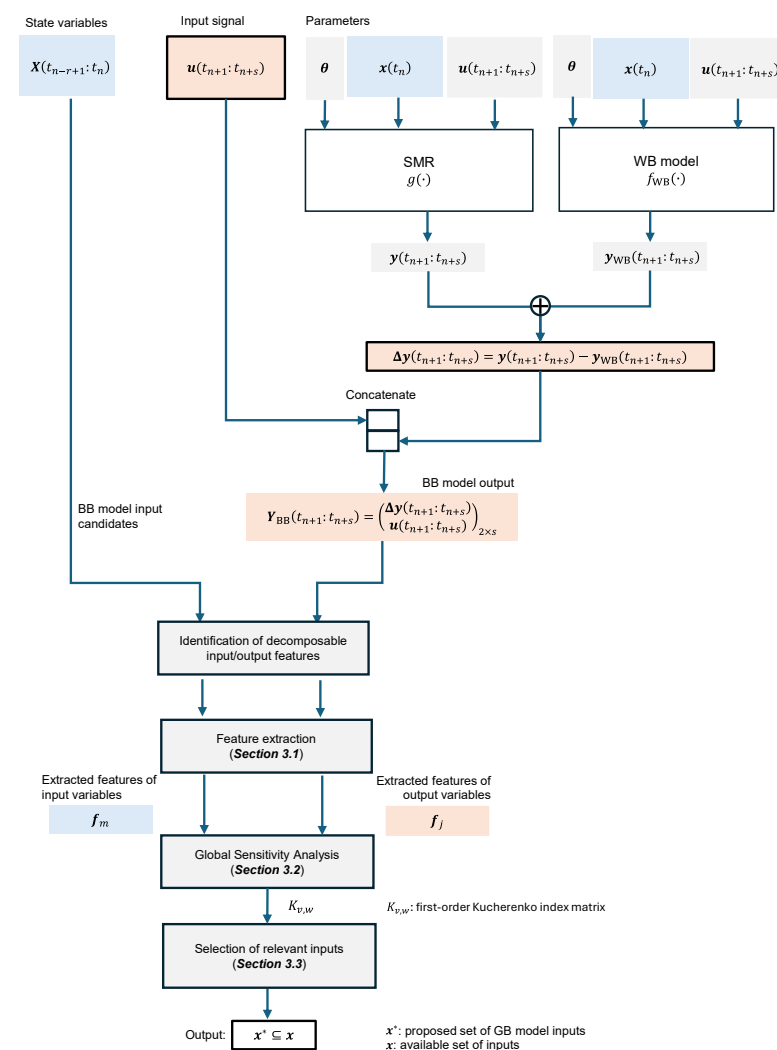


Figure 2. Proposed methodology.

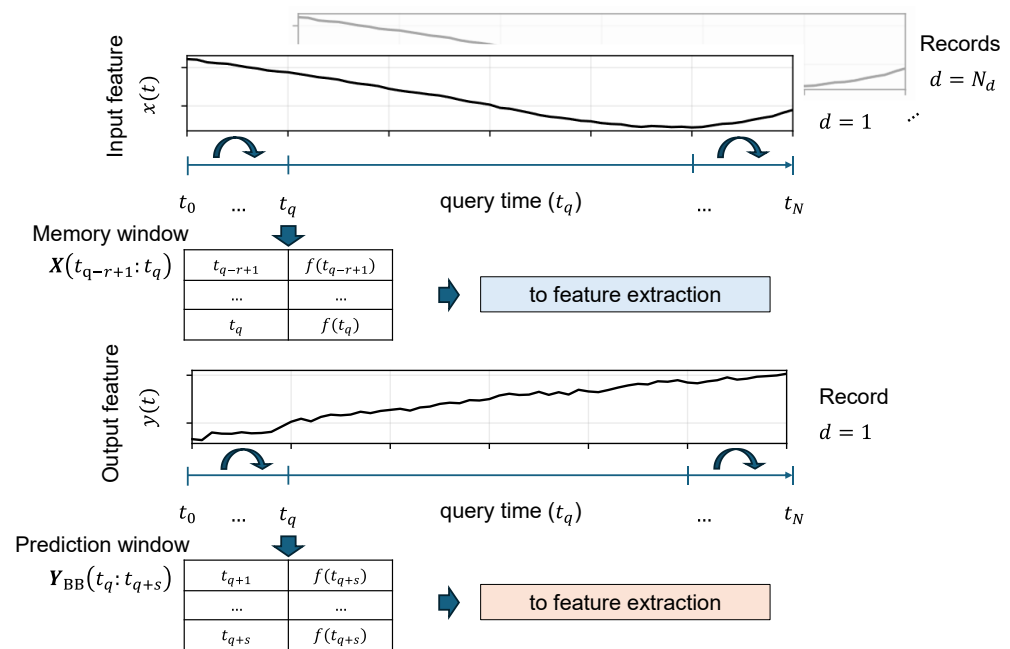


Figure 3. Data collection for feature extraction.

During a second step, a global SA based on Kucherenko indices $K_{p,q}$ [31] is performed on the extracted features $p \in V$ (input variables) and $q \in W$ (output variables), where V and W are the full sets of HHT-based features extracted from the inputs and the target outputs, respectively (see Section 3.2). The input selection is performed by aggregating the computed sensitivity indices $K_{p,q}$ from the feature space to the physical input/output parameter space. An aggregation strategy is employed in the framework of multi-output sensitivity analysis [34] to select a physically robust subset of BB inputs $\mathbf{x}^* \subseteq \mathbf{x}$ (see Section 3.3). The relationships among the extracted features, sensitivity indices, and the final input selection are schematically summarized in Figure 4.

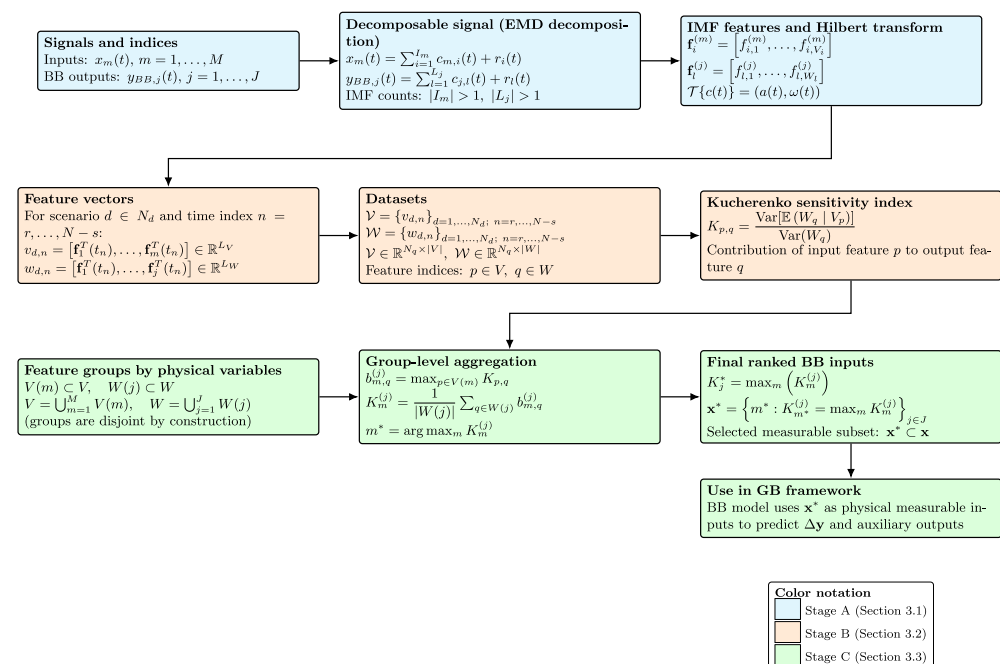


Figure 4. Schematic overview of the relationships among the variables and the symbols (Section 3).

3.1. Feature Extraction Using Hilbert–Huang Transform (HHT)

HHT is employed to extract representative time-frequency features prior to the SA of the nonlinear and non-stationary model of the input and output signals [29,30]. The HHT is an adaptive and data-driven feature extraction technique that decomposes signals through an Empirical Mode Decomposition (EMD) method, producing Intrinsic Mode Functions (IMFs) that inherently reflect true local oscillatory modes without relying on predefined basis functions [30]. This makes the HHT particularly suitable for feature extraction and denoising of nonlinear and non-stationary signals [29].

Without loss of generality, the vector of extracted features for each m -th input and j -th output are denoted as:

$$\mathbf{f}_m = [f_1^{(m)}, \dots, f_i^{(m)}, \dots, f_{I_m}^{(m)}]^T \quad (6)$$

$$\mathbf{f}_j = [f_1^{(j)}, \dots, f_l^{(j)}, \dots, f_{L_j}^{(j)}]^T \quad (7)$$

respectively, where $\mathbf{f}_m$ contains the vector of extracted features $f_i^{(m)}$ of the i -th IMF of the m -th input feature, $i = 1, \dots, I_m$. To obtain IMFs on a certain input feature m , the time-series signal must exhibit oscillatory content (i.e., a set of local extrema) so that the EMD can be applied [30,35]. For this reason, the signals are preprocessed to identify the non-decomposable input features. For each non-decomposable input signal (x_m), the set of extracted features includes the slope coefficient of linear regression ($\hat{\beta}_m$), mean ($\hat{x}_{m,1}$) and standard deviation ($\hat{x}_{m,2}$). In this case, then, the vector of input features renders:

$$\mathbf{f}_i^{(m)} = [\hat{\beta}_m, \hat{x}_{m,1}, \hat{x}_{m,2}] \quad (8)$$

with $I_m = 1$, thus $\mathbf{f}_m = \mathbf{f}_1^{(m)}$. For those decomposable input and output features, the EMD is conducted and a total of $|I_m|$ and $|J_l|$ IMFs are extracted, respectively, satisfying $|I_m| > 1$, $|J_l| > 1$. For each extracted i -th IMF of the m -th input, the vector of features $\mathbf{f}_i^{(m)}$ is expressed as:

$$\mathbf{f}_i^{(m)} = [f_{i,1}^{(m)}, \dots, f_{i,v}^{(m)}, \dots, f_{i,V_i}^{(m)}]^T \quad (9)$$

where $\mathbf{f}_i^{(m)} \in \mathbb{R}^{V_i}$, and $f_{i,v}^{(m)} \in \mathbb{R}$ correspond to the v -th extracted feature of the i -th IMF, where $v = 1, \dots, V_i$. Similarly, the extracted features of the j -th output can be defined as:

$$\mathbf{f}_j = [f_1^{(j)}, \dots, f_l^{(j)}, \dots, f_{L_j}^{(j)}]^T \quad (10)$$

$$\mathbf{f}_l^{(j)} = [f_{l,1}^{(j)}, \dots, f_{l,w}^{(j)}, \dots, f_{l,W_l}^{(j)}]^T \quad (11)$$

For those signals of input $x_m(t)$ and output $y_{BB,j}$ that are decomposed via EMD, the IMFs satisfy:

$$x_m(t) = \sum_{i=1}^{I_m} c_{m,i}(t) + r_i(t) \quad (12)$$

$$y_{BB,j}(t) = \sum_{l=1}^{L_j} c_{j,l}(t) + r_l(t) \quad (13)$$

where $c_{m,i}(t)$ and $c_{j,l}(t)$ are the i -th and l -th IMFs capturing oscillatory behaviors at a characteristic frequency, and $r_i(t)$ and $r_l(t)$ the residual trends. The Hilbert transform is applied to each extracted IMF $c(t)$ by introducing the operator $\mathcal{T}$ as follows:

$$\mathcal{T}\{c(t)\} = (a(t), \omega(t)) \quad (14)$$

where $a_{m,i}(t)$ ($a_{j,l}(t)$) corresponds to the Instantaneous Amplitude (IA) of the i -th IMF of the m -th signal at time t and $\omega_{m,i}$ ($\omega_{j,l}(t)$) corresponds to the instantaneous frequency (IF) of the i -th IMF of the m -th input (output) signal at time t . The operator $\mathcal{T}$ performs the Hilbert transform, the calculation of the analytic function and the extraction of the IA and IF (see [30]). Selected features, including mean IA ($\bar{a}_{m,i}$), IF ($\bar{\omega}_{m,i}$) and Energy ($\bar{e}_{m,i}$) are extracted for each IMF and stored in $f_i^{(m)}$ or $f_l^{(j)}$, correspondingly (see Section 4):

$$f_i^{(m)} = [\bar{a}_{m,i}, \bar{\omega}_{m,i}, \bar{e}_{m,i}] \quad (15)$$

$$f_l^{(j)} = [\bar{a}_{j,l}, \bar{\omega}_{j,l}, \bar{e}_{j,l}] \quad (16)$$

This procedure allows us to capture the nonlinear and transient dynamics of the reactor parameters, serving as explanatory input variables for the global SA, GSA (see Section 3.2). These features are not intended for direct physical interpretation by operators; rather, they are used internally for input selection (see Section 3.3). The final GB model operates on selected physical input variables, and interpretability is ensured through SHAP analysis in the input space (see Section 5.3).

3.2. Calculation of the Kucherenko Indices

Kucherenko indices are an extension of the standard Sobol indices for SA that can handle correlated inputs [31]. The dependent variables y_{BB} for the GSA contain the available residual error data Δy_{BB} and the safety-critical inputs u . The aim is to quantify the contribution of each feature f_m extracted from BB input vector x (and, by aggregation, of each physical/measurable input x_m) to the variance of y_{BB} extracted features f_j while accounting for correlations among features.

For each accident scenario $d \in N_d$ and query time index $n = r, \dots, N - s$, the extracted feature vectors from all BB input signals are concatenated:

$$v_{d,n} = [f_1^T(t_n), \dots, f_m^T(t_n)] \quad (17)$$

where $v_{d,n} \in \mathbb{R}^{L_V}$. Stacking all samples across all N_d scenarios produces the dataset $\mathcal{V}$:

$$\mathcal{V} = \{v_{d,n}\}_{d=1,\dots,N_d; n=r,\dots,N-s} \quad (18)$$

where $\mathcal{V} \in \mathbb{R}^{N_q \times V}$. The extracted features are stored in the dataset $\mathcal{V}$, collecting the samples for the query times indexed $n = r, \dots, N - s$, for each accident scenario $d \in N_d$. Likewise, the corresponding BB model output increments $\Delta y(t_{n+1} : t_{n+s})$ and control signal $u(t_{n+1} : t_{n+s})$ (thus forming Y_{BB}) are obtained to extract output features f_j , stored in matrix $\mathcal{W}$:

$$w_{d,n} = [f_1^T(t_n), \dots, f_j^T(t_n)] \quad (19)$$

$$\mathcal{W} = \{w_{d,n}\}_{d=1,\dots,N_d; n=r,\dots,N-s} \quad (20)$$

where $w_{d,n} \in \mathbb{R}^{L_W}$, and $\mathcal{W} \in \mathbb{R}^{N_q \times V}$. After processing all N_d scenarios and N query indexes for each d -th scenario, the first-order Kucherenko indices $K_{p,q}$ are computed in the feature space using matrices $\mathcal{V}$ and $\mathcal{W}$, as follows:

$$K_{p,q} = \frac{\text{Var}[\mathbb{E}[W_q | V_p]]}{\text{Var}[W_q]} \quad (21)$$

The sensitivity indices $K_{p,q}$ are obtained for each feature $p \in V$ and its contribution to features $q \in W$, where V and W correspond to the set of extracted features from the

BB input and output, respectively. As the analytical calculation of $K_{p,q}$ in Equation (21) is cumbersome [36], in this work we adopt the sample-based approach proposed by [36,37].

3.3. Selection of BB Model Inputs

The selection of the subset of measurable quantities ($\mathbf{x}^* \subset \mathbf{x}$) that will serve as inputs to the BB model is performed based on $K_{p,q}$, where $p \in V$ are the features extracted from the measurable inputs x_m and $q \in W$ are those extracted from the BB outputs. In practice, each measurable input x_m is represented by a group of extracted features ($V(m) \subset V$), where $V(m)$ are disjoint sets satisfying $V = \bigcup_{m=1,\dots,M} V(m)$, and, similarly, each j -th BB output is associated with a set of extracted features ($W(j) \subset W$), with $W = \bigcup_{j=1,\dots,J} W(j)$ and $j \in \{1, 2\}$. The BB input selection problem reduces to ranking the BB inputs by accounting for all features of a given BB output by a group-level sensitivity index [34,38] that aggregates feature-level sensitivity information $K_{p,q}$ into a single metric $b_{m,q}^{(j)}$ for an m -th input and j -th BB output by taking the maximum $K_{p,q}$ across the input features p belonging to the group $V(m)$.

To do that, for each measurable input (x_m) and for each output feature ($q \in W(j)$), we first evaluate the most influential contribution among all features extracted from x_m :

$$b_{m,q}^{(j)} = \max_{p \in V(m)} K_{p,q} \quad (22)$$

where $b_{m,q}^{(j)}$ measures the maximum sensitivity index of input x_m to the variance of the specific output feature q . Then, a group-level score is computed

$$K_m^{(j)} = \frac{1}{|W(j)|} \sum_{q \in W(j)} b_{m,q}^{(j)} \quad (23)$$

that quantifies the average explanatory power of the measurable input (x_m) of the j -th BB output (i.e., $m^* = \underset{m}{\operatorname{argmax}}(K_m^{(j)})$ is the input feature that explains j -th output variability better than any other input $m \neq m^*$).

To obtain a robust ranking of measurable inputs accounting for multiple BB outputs, the index $K_m^{(j)}$ is obtained for each BB output:

$$K_j^* = \max_m (K_m^{(j)}) \quad (24)$$

In this way, the proposed strategy for the selection of $\mathbf{x}^*$ rewards those signals that offer an overall advantage in terms of $K_{p,q}$ to explain the variability of each output feature. Therefore, $\mathbf{x}^*$ can be extracted by considering $\mathbf{x}^* = \left\{ m^* : K_{m^*}^{(j)} = \max_m (K_m^{(j)}) \right\}_{j \in J}$.

4. Case Study

The methodology is applied with reference to the SMDFR case study [39,40]. Given the lack of experimental datasets, a High-Fidelity (HF) model is used as a proxy of the state-space model ($g(\cdot)$) for generating data. The HF model is taken from [40]: it consists of a nodalization for each circuit of $L = 12$ nodes, $h = 1, \dots, N_h$. For each h -th node, the fuel temperature $T_f^h(t)$, the piping wall temperature $T_w^h(t)$ and the coolant temperature profile $T_k^h(t)$ are calculated along with the coolant mass flow rate ($\dot{m}_k^h(t)$). The safety critical parameter considered for risk monitoring is the maximum temperature within the tube piping system $T_p(t)$ at time t (i.e., a proxy of PCT):

$$y(t) = T_p(t) = \max_{h \in N_h} T_w^h(t) \quad (25)$$

We consider as the Initiating Event (IE) the occurrence of a loss-of-coolant accident (LOCA) in the primary coolant loop, in line with [40]. The considered variable $u(t)$ and system design parameters θ for the accidental scenario modeling are listed in Table 1.

Table 1. List of input variables and system design parameters.

Symbol	Notation	Description
Input variable ($u(t)$)	$\dot{m}_{\text{leak}}(t)$	Coolant mass flow leakage rate
Vector of system design parameters (θ)	$\dot{m}_{\kappa,\text{nom}}^h$	Nominal coolant mass flow rate in the h -th node
	$M_{\kappa,\text{nom}}^h$	Nominal mass contained in the h -th node
	$M_{\kappa,\text{nom}}$	Nominal total mass of coolant
	t_f	Failure time
	k_{leak}	Scale parameter
	A_{leak}	Shape parameter
	B_{leak}	Scale parameter
	ϵ	Signal disturbance

To mimic a realistic dataset, we simulate the LOCA as in [41], assuming at time t_f the onset in the primary coolant circuit with rate $\dot{m}_{\text{leak}}(t)$ and random white noise (ϵ):

$$\dot{m}_{\text{leak}}(t) = \begin{cases} B_{\text{leak}} \cdot k_{\text{leak}} \cdot (t - t_f)^{\frac{1}{A_{\text{leak}}}} - k_{\text{leak}} + \epsilon & \text{if } t \geq t_f \\ 0 & \text{if } t < t_f \end{cases} \quad (26)$$

where $k_{\text{leak}} \sim U(k_{\text{leak}}, \bar{k}_{\text{leak}})$, $\epsilon \sim N(0, \sigma_\epsilon)$, $A_{\text{leak}} \sim U(\underline{A}_{\text{leak}}, \bar{A}_{\text{leak}})$, $B_{\text{leak}} \sim U(\underline{B}_{\text{leak}}, \bar{B}_{\text{leak}})$.

Excessive overheating during a LOCA is prevented by the activation of the Protection System (PS) that consists of an Auxiliary Cooling System (ACS) that supplies coolant mass flow rate $\dot{m}_{\text{ACS}}(t)$. The resulting net coolant mass flow rate $\dot{m}_{\kappa,\text{net}}(t)$ is given by:

$$\dot{m}_{\kappa,\text{net}}(t) = \dot{m}_{\text{ACS}}(t) - \dot{m}_{\text{leak}}(t) \quad (27)$$

The variables $x = \{\dot{m}_{\text{leak}}, T_{f,12}^e, T_{\kappa,\text{in}}, T_{\kappa,12}^e\}$ that can be measured at each Δt by means of in-place sensors suitable for metal-cooled SMRs [42] are listed in Table 2.

Table 2. List of measurable variables (x) and system output (y).

Notation	Notation (HF Model)	Name
$u(t)$	$\dot{m}_{\text{leak}}$	Mass flow leakage rate
$x(t)$	$T_{\kappa,12}^e$	Outlet coolant temperature
	$T_{f,12}^e$	Outlet fuel temperature
	$T_{\kappa,\text{in}}$	Inlet coolant temperature
$y(t)$	$\max_h(T_w^h)$	Temperature of piping wall (PCT Proxy)

The GB model that is used to estimate the non-measurable parameters of interest such as the pipe wall temperature, $T_w^h(t)$, consists of a zero-dimensional WB model, from which the model discrepancy Δy is calculated (see Section 2), and a BB model $f_{\text{BB}}(\cdot)$ that corresponds to a Bi-directional Long-Short Term Memory (Bi-LSTM) network for multi-step-ahead estimation of u and Δy (see [33]). The GB model formulation for the case study is summarized in Table 3, whereas the GB model parameters are presented in Table 4. The BB and GB models use a memory window length (r) and prediction

window length (s) of $r = s = 200$ with sampling time $t_{\text{sam}} = 0.5\text{s}$. These settings allow us to capture the short-term evolution of the LOCA transient of interest and anticipate the enforcement of symptom-based operating procedures, while complying with computational time constraints [33].

Table 3. GB model configuration, SMDFR case study.

Notation	Variable	Description
y_{WB}	$T_{\text{p,WB}}$	Vector of WB-based estimates of $\max_h(T_w^h)$ (proxy of PCT)
Δy	ΔT_{p}	Vector of WB estimated errors of $\max_h(T_w^h)$ (proxy of PCT)
y_{GB}	$T_{\text{p,GB}}$	Vector of GB-based estimates of $\max_h(T_w^h)$ (proxy of PCT)

Table 4. GB and BB model parameters, SMDFR case study.

Notation	Description	Value
r	Memory data length	$r = 200$
s	Prediction window length	$s = 200$
M	Number of BB model inputs	$M = 4$
J	Number of BB model outputs	$J = 2$
Δy	BB error correction term	$\Delta y = \Delta T_{\text{p}}$
u	BB output signal	$u = \dot{m}_{\text{leak}}$

BB input/output features were extracted from time-series signals, obtained by simulating N_d accident scenarios via Monte Carlo (MC)-based random failure-injection [41]. Each scenario $d \in N_d$ includes a randomly sampled fault-injection time $t_f \leq t_N$. For each scenario d , the variable matrix $X(t_{n-r+1} : t_n)$ and the input signal $u(t_{n+1} : t_{n+s})$ are collected using fixed parameters (r, s) over the time index $n = r, \dots, N - s$. At each timestep t_n , the HF model $g(\cdot)$ and the WB model $f_{\text{WB}}(\cdot)$ are evaluated to compute the BB model targets $Y_{\text{BB}}(t_n : t_{n+s})$ (see Section 3, Figure 2). The parameters needed for MC failure-injection [41,43] are listed in Table 5.

Table 5. Parameters for accident scenario simulation.

Notation	Description	Value [Units]
N_d	Number of simulated accident scenarios d	$N_d = 150$
Δt	Fixed sampling time	$\Delta t = 0.5 \text{ [s]}$
t_n	Fixed observed time	$t_n \in \{\Delta t, 2\Delta t, \dots, t_N\} \text{ [s]}$
n	Fixed time index	$n \in \{1, \dots, N\} \text{ [-]}$
t_N	Fixed mission time, for each scenario d	$t_N = 1500 \text{ [s]}$
t_f	Failure time, sampled for each scenario d	$t_f \sim \text{Uniform}[500, t_N] \text{ [s]}$
k_{leak}	Scale parameter, sampled for each scenario d	$k_{\text{leak}} \sim \text{Uniform}[8, 10] \text{ [kg/s]}$
A_{leak}	Shape parameter, sampled for each scenario d	$A_{\text{leak}} \sim \text{Uniform}[9, 15] \text{ [-]}$
B_{leak}	Scale parameter, sampled for each scenario d	$B_{\text{leak}} \sim \text{Uniform}[0.2, 0.4] \text{ [s}^{-1}\text{]}$

5. Results

5.1. Feature Extraction with Hilbert–Huang Transform

Signals are firstly preprocessed to identify the non-decomposable input features, prior to feature extraction. As shown in Figure 5, input signals such as the inlet coolant temperature ($T_{\kappa,\text{in}}$) and outlet fuel/coolant temperature ($T_{f,12}^e$, $T_{\kappa,12}^e$) do not show oscillatory content, revealing monotonically increasing/decreasing values. These input features become non-decomposable signals and therefore the HHT is not applicable, whereas output feature signals ($\dot{m}_{\text{leak}}$, ΔT_p) reveal oscillatory content that is extracted via HHT (see Figure 6).

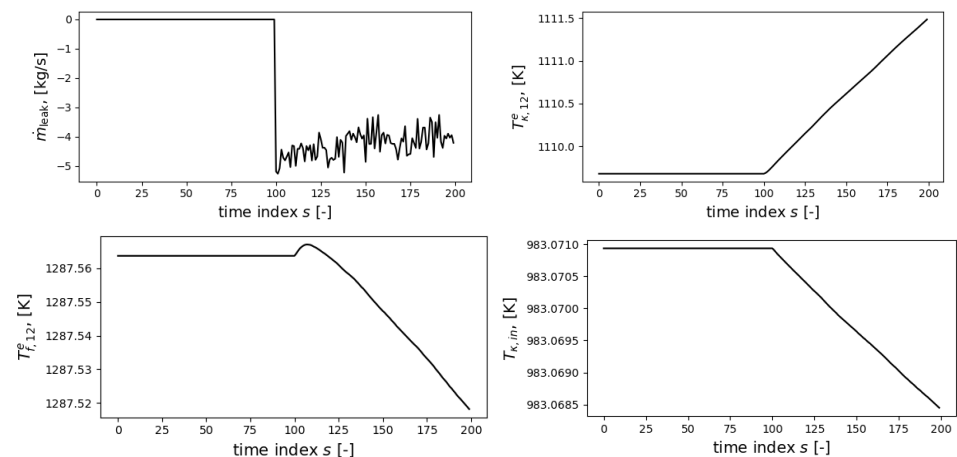


Figure 5. Examples of input signals.

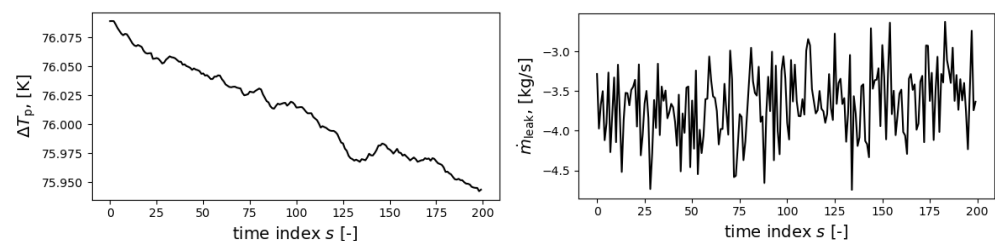


Figure 6. Example of output signals.

The generated dataset is processed by applying the Hilbert–Huang Transform (HHT) to extract the features (f_i) and (f_j) from input signals x and output signals y_{BB} , respectively. The list of extracted features for non-stationary signals is described in Table 6.

Table 6. Extracted features from non-stationary signals.

Feature	Expression ($x_m \in x$)	Expression ($y_{\text{BB},j} \in y_{\text{BB}}$)
Mean of Instantaneous Amplitude for each IMF i (input) or j (output)	$\bar{a}_{m,i} = \frac{1}{r} \sum_{k=n-r+1}^n a_{m,i}(t_k)$	$\bar{a}_{j,l} = \frac{1}{s} \sum_{k=n+1}^{n+s} a_{j,l}(t_k)$
Mean of instantaneous frequency for each IMF i (input) or j (output)	$\bar{\omega}_{m,i} = \frac{1}{r} \sum_{k=n-r+1}^n \omega_{m,i}(t_k)$	$\bar{\omega}_{j,l} = \frac{1}{s} \sum_{k=n+1}^{n+s} \omega_{j,l}(t_k)$
Mean Energy for each IMF i (input) or j (output)	$\bar{e}_{m,i} = \frac{1}{r} \sum_{k=n-r+1}^n a_{m,i}^2(t_k)$	$\bar{e}_{j,l} = \frac{1}{s} \sum_{k=n+1}^{n+s} a_{j,l}^2(t_k)$
Slope coefficient ($\hat{\beta}_{r,m}$, $\hat{\beta}_{r,j}$) of linear regression, mean ($\hat{r}_{m,1}$, $\hat{r}_{j,1}$) and standard deviation ($\hat{r}_{m,1}$, $\hat{r}_{j,2}$) of the residual ($\hat{r}_m$, $\hat{r}_j$).	$r_m(t) = \hat{\beta}_{r,m}t + \gamma, \quad t \in \{t_{n-r+1}, \dots, t_n\}$ $\hat{r}_{m,1} = \frac{1}{r} \sum_{k=n-r+1}^n r_m(t_k)$ $\hat{r}_{m,2} = \sqrt{\frac{1}{r} \sum_{k=n-r+1}^n (r_m(t_k) - \hat{r}_{m,1})^2}$	$r_j(t) = \hat{\beta}_{r,j}t + \gamma, \quad t \in \{t_{n+1}, \dots, t_{n+s}\}$ $\hat{r}_{j,1} = \frac{1}{s} \sum_{k=n+1}^{n+s} r_j(t_k)$ $\hat{r}_{j,2} = \sqrt{\frac{1}{s} \sum_{k=n+1}^{n+s} (r_j(t_k) - \hat{r}_{j,1})^2}$

For non-decomposable input signals (x_m), the features extracted are the slope coefficient of linear regression ($\hat{\beta}$), mean ($\hat{\mu}_1$) and standard deviation ($\hat{\sigma}_2$) (see Table 7).

Table 7. Extracted features from linear signals x_m .

Feature	Expression
Slope coefficient ($\hat{\beta}_m$) of linear regression m -th signal	$x_m(t) = \hat{\beta}_m t + \gamma, t \in \{t_{n-r+1}, \dots, t_n\}$
Mean ($\hat{x}_{m,1}$) of m -th input signal x_m	$\hat{x}_{m,1}(t_n) = \frac{1}{r} \sum_{k=n-r+1}^n x_m(t_k)$
Standard deviation ($\hat{x}_{m,2}$) of the m -th input signal x_m	$\hat{x}_{m,2}(t_n) = \sqrt{\frac{1}{r} \sum_{k=n-r+1}^n (x_m(t_k) - \hat{x}_{m,1})^2}$

For the BB input signals x_m , the input signal $m = 1$ (coolant leakage rate) was considered for feature extraction via HHT. A total of $|I_{m=1}| = 4$ IMFs were extracted. The selection of IMFs is based on their relative energy contribution (i.e., highest mean values in power distribution, see Figure 7b), and its adequacy is verified through the subsequent sensitivity analysis; low-frequency information is preserved via residual-based features, which are also included in the SA. Following this criterion, IMFs 1 and 2 were selected as they carry more information in terms of signal power for both short- and long-term trends (see Figure 7).

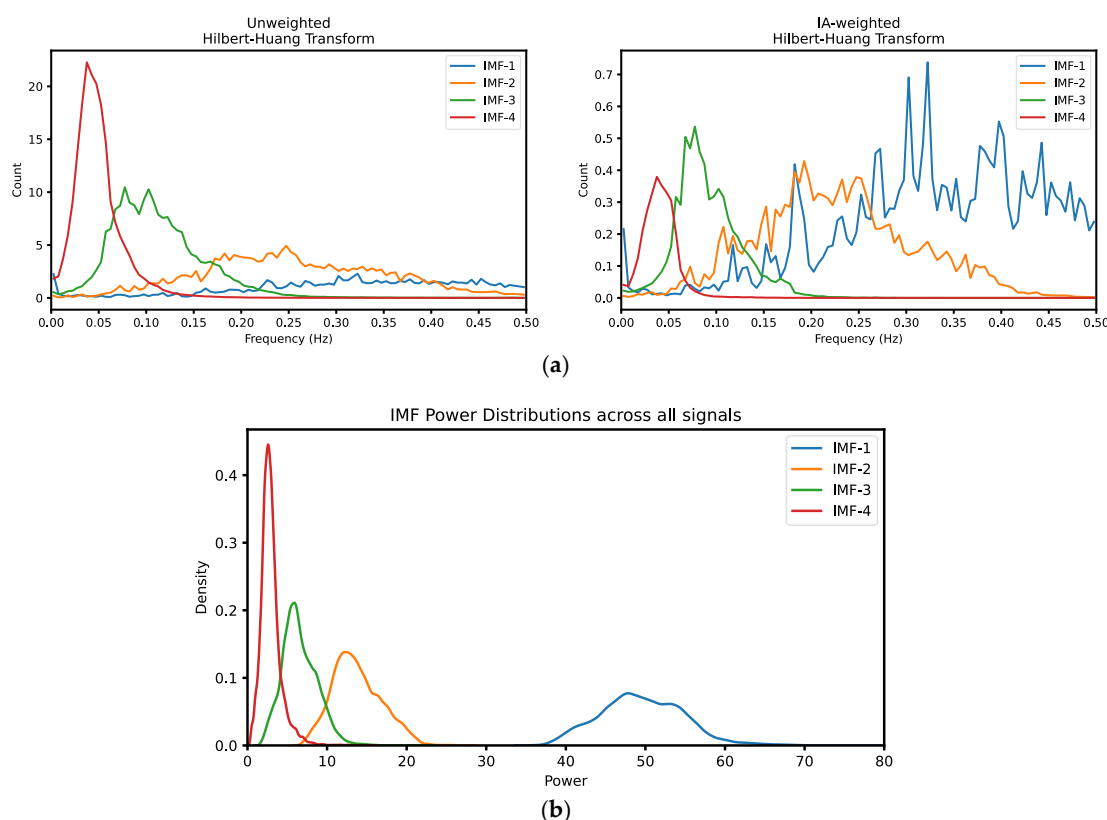


Figure 7. (a) Instantaneous frequency (left) and instantaneous frequency distribution weighted by IA (right), GB model input ($\dot{m}_{\text{leak}}$), for IMFs 1–4 unweighted and weighted by IA. (b) Power distribution of IMFs extracted for GB model input ($\dot{m}_{\text{leak}}$).

For the rest of the BB model inputs ($T_{f,12}^e$, $T_{\kappa,12}^e$, $T_{\kappa,\text{in}}$, where $m = 2, 3, 4$, correspondingly), features such as the slope coefficient of linear regression ($\hat{\beta}_m$), mean ($\hat{x}_{m,1}$) and

standard deviation ($\hat{x}_{m,2}$) of the time series were extracted (see Table 8). As stated, a total of $V = 18$ features were extracted.

Table 8. Extracted features of BB model inputs (f_m).

Notation	Type of Signal of BB Model	Variable	Extracted Features	Feature Code
u	Input, $m = 1$	Mass flow coolant leakage rate ($\dot{m}_{\text{leak}}$). Feature vector: f_1	IMFs 1 ($i = 1$)	
			• Mean of IA	leak_mean_amp_imf1
			IMFs 1 ($i = 1$)	
			• Mean Power	leak_mean_power_imf1
			IMFs 1 ($i = 1$)	
			• Mean of IF	leak_mean_freq_imf1
			IMFs 2 ($i = 2$)	
			• Mean of IA	leak_mean_amp_imf2
			IMFs 2 ($i = 2$)	
			• Mean Power	leak_mean_power_imf2
			IMFs 2 ($i = 2$)	
			• Mean of IF	leak_mean_freq_imf2
			Residual ($\hat{\beta}_{r,1}$)	
			• Beta coefficient	leak_slope_residual
			Residual ($\hat{r}_{1,1}$)	
			• Mean	leak_mean_residual
			Residual ($\hat{r}_{1,2}$)	
			• Standard deviation	leak_std_residual
x	Input, $m = 2$	Outlet coolant temperature ($T_{\kappa,12}^e$). Feature vector: f_2	Beta coefficient ($\hat{\beta}_m$)	slope_Te_f_12
			Mean ($\hat{x}_{m,1}$)	mean_Te_f_12
			Standard deviation ($\hat{x}_{m,2}$)	std_Te_f_12
	Input, $m = 3$	Outlet fuel temperature ($T_{f,12}^e$). Feature vector: f_3	Beta coefficient ($\hat{\beta}_m$)	slope_Te_c_12
			Mean ($\hat{x}_{m,1}$)	mean_Te_c_12
			Standard deviation ($\hat{x}_{m,2}$)	std_Te_c_12
	Input, $m = 4$	Inlet coolant temperature ($T_{\kappa,\text{in}}$). Feature vector: f_4	Beta coefficient ($\hat{\beta}_m$)	slope_Tin_c_12
			Mean ($\hat{x}_{m,1}$)	mean_Tin_c_12
			Standard deviation ($\hat{x}_{m,2}$)	std_Tin_c_12

For the BB output signals, the outputs $j \in \{1, 2\}$ were considered for feature extraction via HHT. For $j = 1$ ($\dot{m}_{\text{leak}}$), a total of $|L_{j=1}| = 4$ IMFs were extracted during the sifting process (see Figure 8), where IMFs 1 and 2 were selected as they contain more information in terms of signal power for both short- and long-term trends (see Figure 7). Following the same procedure for $j = 2$ (ΔT_p), a total of $|V_{j=2}| = 3$ IMFs were extracted, where IMFs 2 and 3 were selected during the feature extraction process (see Figure 9). The extracted features are summarized in Table 9. As stated, a total of $W = 18$ features were extracted.

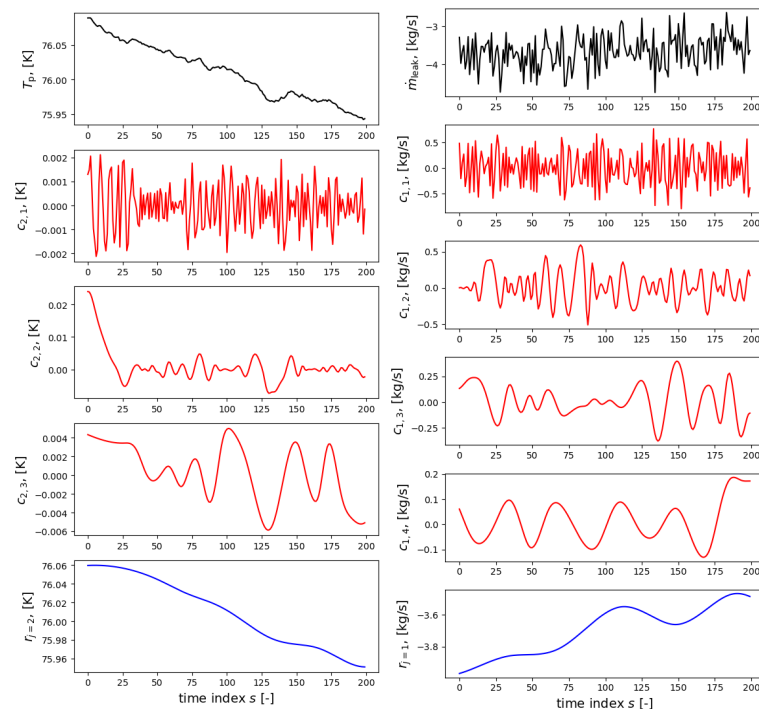


Figure 8. IMFs and residuals extracted for output signals, given a random sample. (Left): Coolant mass flow leakage rate ($\dot{m}_{\text{leak}}$, with $j = 1$). (Right): Error estimation of PCT (ΔT_p , with $j = 2$).

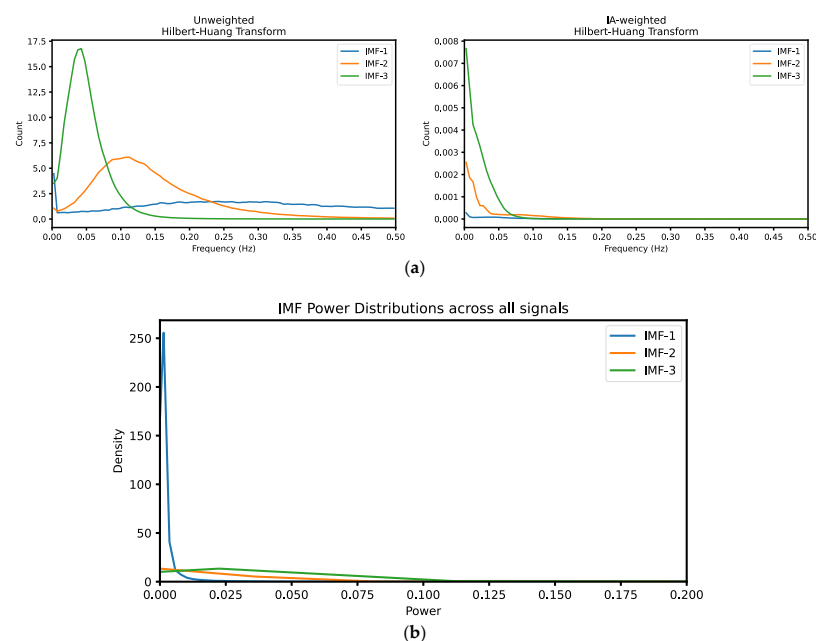


Figure 9. (a) Instantaneous frequency (left) and instantaneous frequency distribution weighted by IA (right), GB model output (ΔT_p), for IMFs 1–4 unweighted and weighted by IA. (b) Power distribution of IMFs extracted for GB model output (ΔT_p).

Table 9. Extracted features of BB model output (f_j).

Type of Signal of BB Model	Variable	Extracted Features	Feature Code
Output $j = 1$	Mass flow coolant leakage rate ($\dot{m}_{\text{leak}}$). Feature vector: f_1	IMFs 1 ($l = 1$)	
		• Mean of IA	leak_mean_amp_imf1
		IMFs 1 ($l = 1$)	
		• Mean Power	leak_mean_power_imf1
		IMFs 1 ($l = 1$)	
		• Mean of IF	leak_mean_freq_imf1
		IMFs 2 ($l = 2$)	
		• Mean of IA	leak_mean_amp_imf2
		IMFs 2 ($l = 2$)	
		• Mean Power	leak_mean_power_imf2
Output $j = 2$	Diff. of Peak Cladding Temperature (ΔT_p). Feature vector: f_2	IMFs 2 ($l = 2$)	
		• Mean of IF	leak_mean_freq_imf2
		Residual ($\hat{\beta}_{r,1}$)	
		• Beta coefficient	leak_slope_residual
		Residual ($\hat{r}_{1,1}$)	
		• Mean	leak_mean_residual
		Residual ($\hat{r}_{1,2}$)	
		• Standard deviation	leak_std_residual
		IMFs 1 ($l = 2$)	
		• Mean of IA	deltaTp_mean_amp_imf2
Output $j = 2$	Diff. of Peak Cladding Temperature (ΔT_p). Feature vector: f_2	IMFs 1 ($l = 2$)	
		• Mean Power	deltaTp_mean_power_imf2
		IMFs 1 ($l = 2$)	
		• Mean of IF	deltaTp_mean_freq_imf2
		IMFs 2 ($l = 3$)	
		• Mean of IA	deltaTp_mean_amp_imf3
		IMFs 2 ($l = 3$) Mean Power	deltaTp_mean_power_imf3
		IMFs 2 ($l = 3$)	
		• Mean of IF	deltaTp_mean_freq_imf3
		Residual ($\hat{\beta}_2$)	
Output $j = 2$	Diff. of Peak Cladding Temperature (ΔT_p). Feature vector: f_2	• Beta coefficient	deltaTp_slope_residual
		Residual ($\hat{r}_{2,1}$)	
		• Mean	deltaTp_mean_residual
		Residual ($\hat{r}_{2,2}$)	
Output $j = 2$	Diff. of Peak Cladding Temperature (ΔT_p). Feature vector: f_2	• Standard deviation	deltaTp_std_residual

5.2. Sensitivity Analysis with Kucherenko Indices

The dataset of extracted features ($\mathcal{V}$ and $\mathcal{W}$) is used to calculate the first-order Kucherenko indices. A matrix of first-order Kucherenko indices $K_{p,q}$ is shown in Figure 10. The obtained indices are postprocessed to select a subset of measured quantities x^* as BB model inputs (X_{BB}), given the computational time constraints for online risk monitoring. The SA based on the first-order Kucherenko indices highlights the importance of selecting a subset of input features for the BB model, given the following findings:

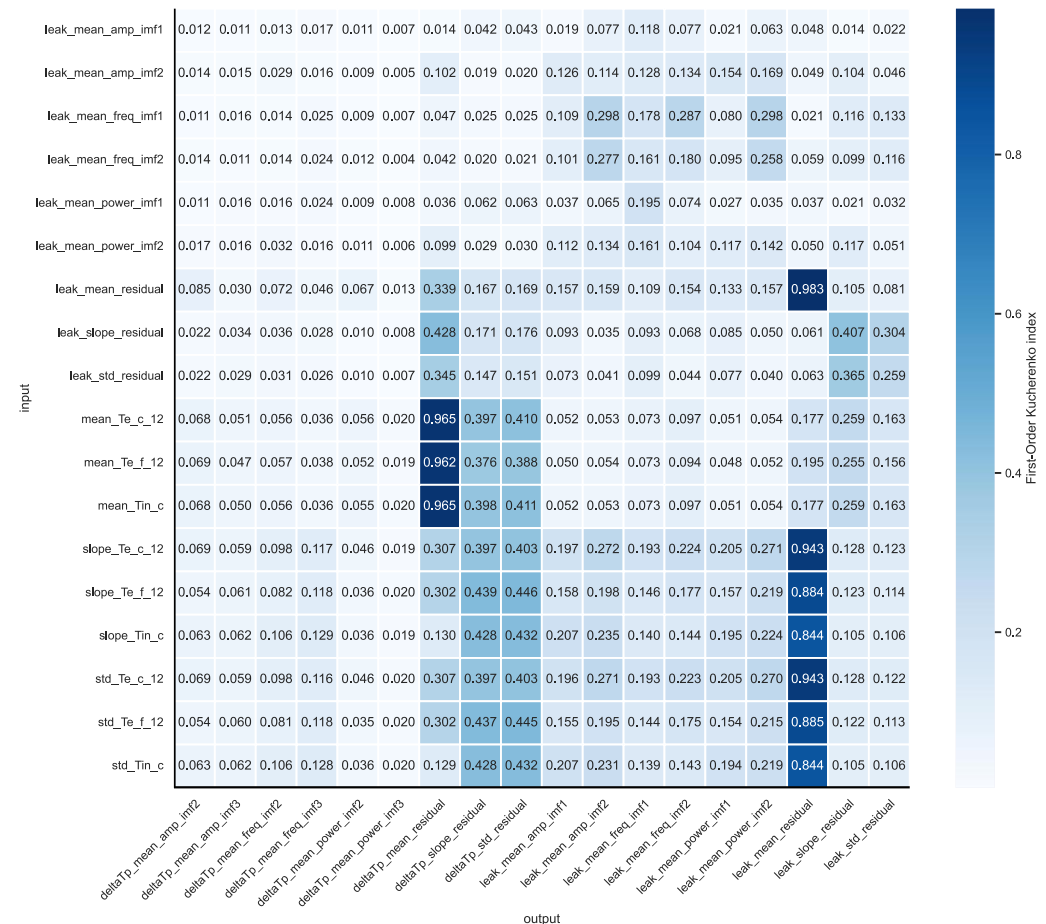


Figure 10. First-order Kucherenko index matrix.

- The features extracted from the residual of leakage rate $\dot{m}_{leak}$ achieve high sensitivity indices when predicting autoregressive long-term trends.
- Among the extracted features from the temperature signals, features from the outlet coolant temperature $T_{\kappa,12}^e$ and the inlet coolant temperature $T_{\kappa,in}$ present higher sensitivity levels when predicting the residuals of both BB outputs (i.e., denoised long-term trends of coolant leakage rate $\dot{m}_{leak}$ and the PCT estimation error ΔT_p during the LOCA scenario).
- The extracted features of the leakage rate do not substantially contribute to explaining the variability in the PCT estimation error ΔT_p , which can be explained due to the nonlinear effects of hydraulic phenomena and the thermohydraulic equations of the HF model.

5.3. Selection of BB Model Inputs

Each measurable input x_m corresponds to a group of features $V(m) \subset V$, whereas each BB output $j \in \{1, 2\}$ (leakage rate and ΔT_p) corresponds to an output-feature group $W(j) \subset W$. Following the strategy described in Section 3.3, the grouped sensitivity index

for each measurable input is computed as stated in Equations (22)–(24). Tables 10 and 11 show the maximum Kucherenko indices grouped by output j ($b_{m,q}^{(j)}$), from which the average score per input is computed by means of Equation (23) (see Table 12).

Table 10. Best scored Kucherenko indices $b_{m,q}^{(1)}$ for features grouped by output $j = 1$.

	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$	$q=7$	$q=8$	$q=9$
$b_{1,q}^{(1)}$	0.157	0.298	0.195	0.287	0.154	0.298	0.983	0.407	0.304
$b_{2,q}^{(1)}$	0.197	0.272	0.193	0.224	0.205	0.271	0.943	0.259	0.163
$b_{3,q}^{(1)}$	0.158	0.198	0.146	0.177	0.157	0.219	0.885	0.255	0.156
$b_{4,q}^{(1)}$	0.207	0.235	0.140	0.144	0.195	0.224	0.844	0.259	0.163

Table 11. Best scored Kucherenko indices $b_{m,q}^{(2)}$ for features grouped by output $j = 2$.

	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$	$q=7$	$q=8$	$q=9$
$b_{1,q}^{(2)}$	0.085	0.034	0.072	0.046	0.067	0.013	0.428	0.171	0.176
$b_{2,q}^{(2)}$	0.069	0.059	0.098	0.117	0.056	0.020	0.965	0.397	0.410
$b_{3,q}^{(2)}$	0.069	0.061	0.082	0.118	0.052	0.020	0.962	0.439	0.446
$b_{4,q}^{(2)}$	0.068	0.062	0.106	0.129	0.055	0.020	0.965	0.428	0.432

Table 12. Average scores $K_m^{(j)}$ per input group.

Output ($j=1$)		Output ($j=2$)	
$K_1^{(1)}$	0.3425	$K_1^{(2)}$	0.1215
$K_2^{(1)}$	0.3028	$K_2^{(2)}$	0.2435
$K_3^{(1)}$	0.2611	$K_3^{(2)}$	0.2498
$K_4^{(1)}$	0.2679	$K_4^{(2)}$	0.2516

As stated, input signal $m = 1$ (coolant mass flow leakage rate, $\dot{m}_{\text{leak}}$) is the best at explaining the overall variability of output $j = 1$ ($\dot{m}_{\text{leak}}$), whereas signal $m = 4$ (inlet coolant temperature, $T_{\kappa,\text{in}}$) is selected to explain the variability of $j = 2$ (PCT estimation error, ΔT_p).

The grouped Kucherenko indices K_j^* are then computed by means of Equation (23), obtaining $K_j^* = \{K_1^{(1)} = 0.3425, K_4^{(2)} = 0.2516\}$. Based on the grouped ranking, the selected measurable inputs are:

$$\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\} \quad (28)$$

As stated, these two signals form the selected optimal subset of BB inputs for the deployment of the GB model to be used for the monitoring of the SMDFR.

A BB model (Bi-directional Long-Short Term Memory, Bi-LSTM) [33] is trained for multi-step-ahead estimation, and embedded into the GB model. Figure 11 shows the performance (in terms of RMSE) of the GB model, when different sets of BB input variables ($\mathbf{x}^*$) are considered, including all possible pairs of input variables and all input variables $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e, T_{f,12}^e, T_{\kappa,\text{in}}\}$. It can be seen that the set of $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\}$ that has been identified by the proposed method allows the selection of a minimal yet informative set

of inputs, i.e., it provides similar results to the GB model that considers all the measurable variables as input features ($\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e, T_{f,12}^e, T_{\kappa,\text{in}}\}$).

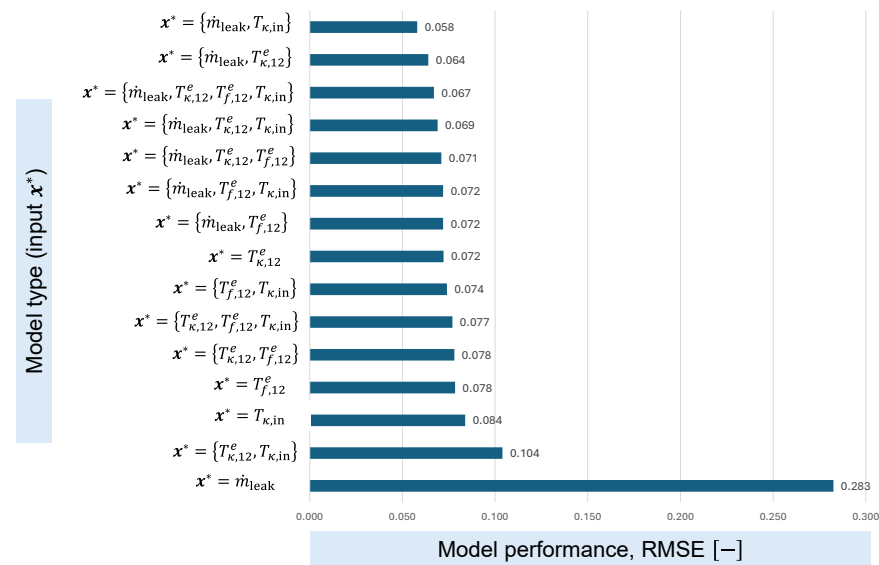


Figure 11. Comparison of instantiated GB models under different BB input features.

The GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\}$ is then tested against the GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e, T_{f,12}^e, T_{\kappa,\text{in}}\}$ and $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e\}$ in terms of computational time. The GB models are used to risk-monitor 20 random-trial LOCA scenarios (see [33] for further details): the results are summarized in Figure 12. The GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\}$ complies with the computational constraints imposed by regulatory requirements (i.e., a total computing time of 120 s, with no exceptions [1]), whereas the GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e, T_{f,12}^e, T_{\kappa,\text{in}}\}$ sometimes exceeds the constraint due to increased complexity (see the outliers *). The GB model candidate that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,12}^e\}$ satisfies the computational constraint; however, the proposed model with $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\}$ proves to be a better candidate in terms of accuracy and computational time.

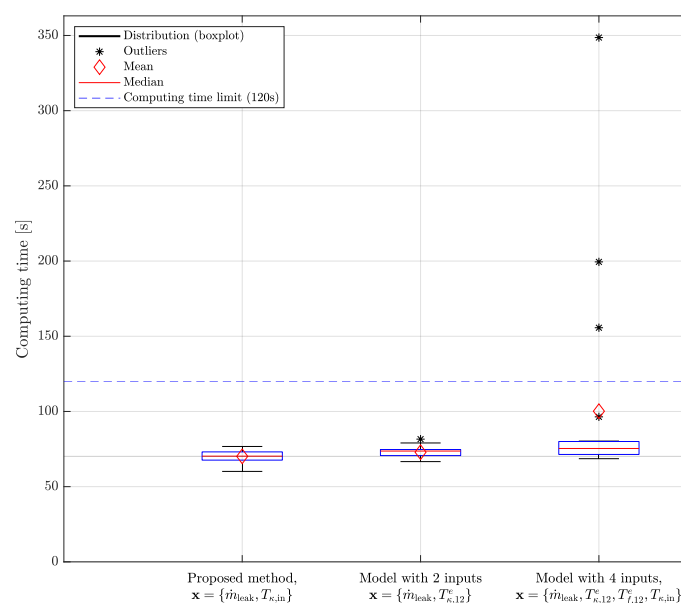


Figure 12. Boxplot (computing time) of proposed GB model against selected GB models.

The explainability of the GB models are then also analyzed with SHAP, as proposed in [33], to quantify the contribution of each model input feature to its output at each query time t_q [44,45] of a simulated accident scenario (see Table 13). Figure 13 shows the evolution of the SHAP values for the BB output features ($\dot{m}_{\text{leak}}$ and ΔT_p) predicted by the Bi-LSTM that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa,\text{in}}\}$ within the GB model at $t_q = 1100, \dots, 1300$ s. The results are summarized as follows:

Table 13. Accident scenario parameters.

Variable	Description	Value [Units]
r	Memory data length	$r = 200$ [—]
s	Prediction window length	$s = 200$ [—]
t_N	Mission time	$t_N = 1500$ [s]
t_q	Query time	$t_q \in \{1000, \dots, 1300\}$; $t_1 = 1000$ [s]: early stage of the LOCA scenario; $t_2 = 1100$ [s]: activation of Protection System (Auxiliary Cooling System); $t_3 = 1300$ [s]: late stage of the LOCA scenario, with enacted Protection System
t_f	Failure time	$t_f = 500$ [s]

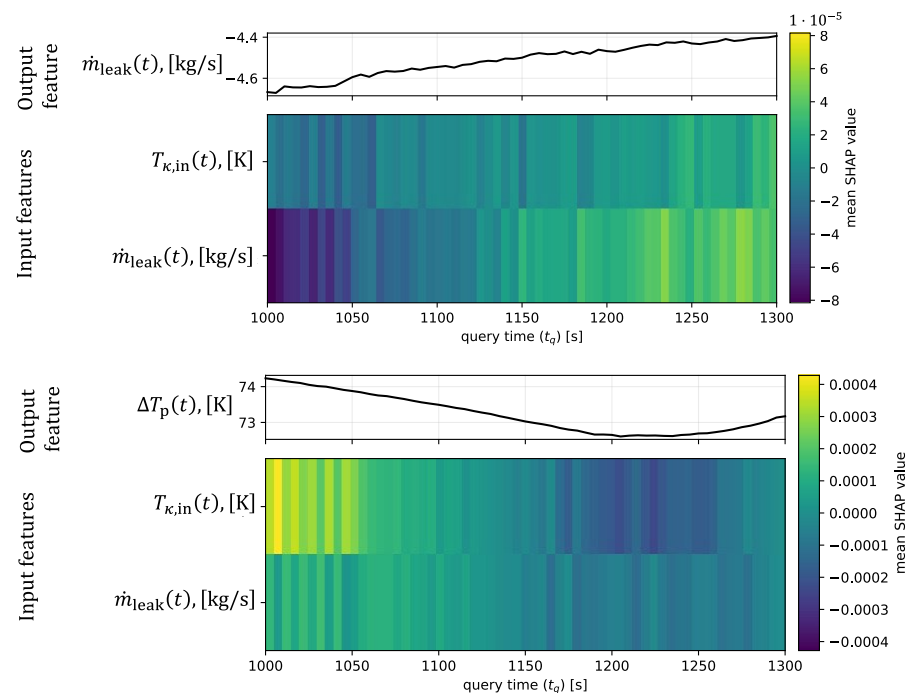


Figure 13. SHAP values during the accident scenario evolution, proposed GB model. (**Top**): predicted mass flow leakage rate ($\dot{m}_{\text{leak}}$). (**Bottom**): predicted PCT correction (ΔT_p).

- $\dot{m}_{\text{leak}}$ is shown in Figure 13 (top) to depend mostly on $\dot{m}_{\text{leak}}$ since $T_{\kappa,\text{in}}$ does not affect the severity of the leakage rate that is instead primarily determined by the rupture break size: initially, the rupture causes a leakage (e.g., -10 kg/s) (negative SHAP values); as the LOCA progresses and the system depressurizes, leakage reduces (e.g., -4 kg/s) (positive SHAP values). This confirms that hydraulic phenomena dominate this output, where the model correctly prioritizes historical leakage data to forecast future leakage rates.

- ΔT_p is shown in Figure 13 (bottom) to depend on $T_{\kappa, \text{in}}$: During the early LOCA stage, less coolant inventory means less heat removal, raising the PCT and lowering $T_{\kappa, \text{in}}$. The WB model overestimates this PCT rise, requiring a lower PCT correction (ΔT_p). The BB model uses $T_{\kappa, \text{in}}$ to estimate the coolant inventory state: when $T_{\kappa, \text{in}}$ is lower (less coolant inventory, e.g., at time $t = 1100$ s), the model decreases the PCT error correction (lower SHAP values) due to the overestimation by the WB model; when $T_{\kappa, \text{in}}$ is higher (more coolant inventory, e.g., at $t = 1000$ s), the model increases the PCT correction (higher SHAP values) to compensate for the error.

The SHAP-based analysis confirms that the BB model predictions are consistent with the accident dynamics and shows that the GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa, \text{in}}\}$ is explainable for risk monitoring applications. The GB model candidate $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa, 12}^e\}$ has also proven to be sufficiently explainable, as shown in [33].

For comparison purposes, Figure 14 shows the SHAP values of ($\dot{m}_{\text{leak}}$ and ΔT_p) where the BB of the GB model is fed with all measurable variables $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{\kappa, 12}^e, T_{f, 12}^e, T_{\kappa, \text{in}}\}$ as input features. Results are summarized as follows:

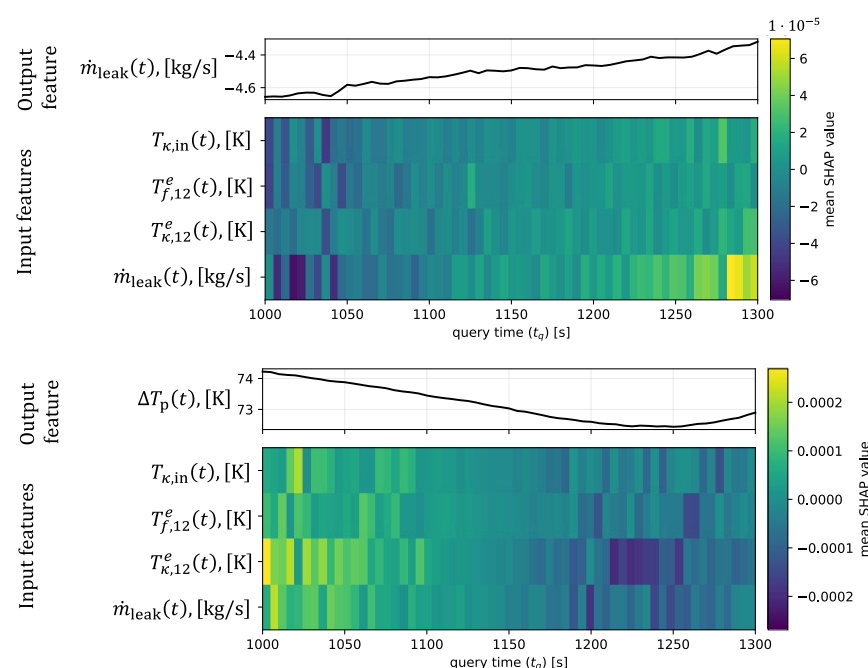


Figure 14. SHAP values during the accident scenario evolution, GB model with 4 input features. **(Top):** predicted mass flow leakage rate ($\dot{m}_{\text{leak}}$). **(Bottom):** predicted PCT correction (ΔT_p).

- $\dot{m}_{\text{leak}}$ is shown in Figure 14 (top) to depend mostly on $\dot{m}_{\text{leak}}$, confirming that hydraulic phenomena dominate this output, as previously shown in Figure 13. However, the inlet coolant temperature ($T_{\kappa, \text{in}}$) and the outlet fuel temperature ($T_{f, 12}^e$) show a contribution in terms of the SHAP values during early LOCA stages (e.g., $t_q = 1000$ s) that cannot be explained in terms of physical phenomena, as the coolant leakage rate ($\dot{m}_{\text{leak}}$) is primarily injected by means of MC-based random sampling.
- ΔT_p is shown in Figure 14 (bottom) to depend mostly on $T_{\kappa, 12}^e$, exhibiting a similar behavior as the inlet coolant temperature $T_{\kappa, \text{in}}$ in Figure 13. This is expected, as the outlet coolant temperature can be used as a proxy of the coolant inventory; however, the influence of the inlet coolant temperature ($T_{\kappa, \text{in}}$) and the outlet fuel temperature ($T_{f, 12}^e$) is not clear in terms of SHAP values during late LOCA stages (e.g., $t_q = 1200$ s, 1250 s).

For this reason, the explanatory power of the GB model with all measurable variables as input features lacks consistency, specifically when adding correlated variables such as the

outlet fuel temperature ($T_{f,12}^e$). Also, as stated, including more input parameters improves the overall accuracy of the GB model but at the expense of higher computing times.

Table 14 summarizes the insights of the comparison: we can claim that the GB model that uses $\mathbf{x}^* = \{\dot{m}_{\text{leak}}, T_{k,\text{in}}\}$ offers a suitable tradeoff between accuracy, explainability and computational time which makes it the perfect candidate for risk monitoring the PCT of the SMDFR. As stated previously, selecting the proposed GB inputs can be critical to satisfying regulatory requirements (e.g., see Figure 12). The physical consistency of the selected inputs with the underlying LOCA dynamics suggests that the results are not specific to the adopted WB model, but rather reflect the dominant thermohydraulic phenomena, supporting the robustness of the selection process.

Table 14. Overall performance of GB model configurations.

	Proposed Method	Full GB Model with 4 Inputs	Regulatory Requirements
Accuracy (RMSE) [K]	✓	✓	Accuracy below the tolerance error of current instrumentation (sensor: resistance temperature detector, Class A) [46,47]: $\text{RMSE} \leq 1.89 \text{ K}$
Mean computing time per query t_q [s]	✓	✗	Response time of two minutes for risk monitors [1]: $t_q \leq 120 \text{ [s]}$
Explainability (SHAP values)	✓	✗	Physical consistency of the model output, based on explainability techniques, to provide end-user trust [14]

6. Conclusions

This work presents a methodology for the selection of the inputs to a GB model to be embedded in a framework of risk monitoring for SMRs. The methodology integrates the HHT for adaptive feature extraction from nonlinear, transient input time series, and first-order Kucherenko sensitivity indices for global SA. Tested on the SMDFR, the approach supports the selection of a subset of measurable inputs from monitoring data for the estimation of the non-measurable safety-critical parameters of interest, the PCT. By feeding the selected inputs into the GB model, computationally efficient and accurate multi-step-ahead predictions are obtained both under nominal operating conditions and simulated accident scenarios. The proposed method and input selection scheme is agnostic to specific GB model architectures and accident types, thus making it useful for scalable deployment in risk monitoring applications, where physical interpretability and real-time computational demands are key constraints. For Digital Twin applications, the feature extraction and sensitivity analysis can be performed offline and updated only when significant changes in operating conditions are detected, ensuring a balance between computational cost and sustained model performance.

Future work will include the validation of the proposed methodology using real datasets from laboratory experiments, whose signals typically contain noise and systematic errors. Some pending challenges include the scalability and computational cost of Kucherenko index estimation and HHT feature extraction for high-dimensional, high-frequency sensor data in SMRs, and the extension to online input selection under different operating conditions. As stated, the identified input subset is scenario-dependent; however, the proposed methodology can be extended to enable adaptive input selection across different transient conditions. Finally, to apply the methodology to other SMR designs, the WB and BB models should be tailored on their specific features, for example measurable inputs for performing SA.

Author Contributions: L.M.: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft. I.A.: Conceptualization, Methodology, Supervision, Writing—review and editing. F.D.M.: Conceptualization, Methodology, Supervision, Writing—review and editing. E.Z.: Conceptualization, Methodology, Supervision, Writing—review and editing, Project administration. All authors have read and agreed to the published version of the manuscript.

Funding: Marie Skłodowska-Curie Actions. Horizon 2020-European Training Network on Grey-Box Models for Safe and Reliable Intelligent Mobility Systems: Grant agreement No. 955393.

Data Availability Statement: Data can be made available upon request.

Conflicts of Interest: The authors declare they have no known competing financial interests or personal relationships that could have influenced this work.

Nomenclature

Abbreviations

BB	Black-Box model
Bi-LSTM	Bidirectional Long-Short Term Memory
CDF	Cumulative Density Function
EMD	Empirical Mode Decomposition
GB	Gray-Box model
HF	High-fidelity model
HHT	Hilbert–Huang Transform
IA	Instantaneous Amplitude
IF	Instantaneous Frequency
IMF	Intrinsic Mode Function
LOCA	Loss-Of-Coolant Accident
LASSO	Least Absolute Shrinkage and Selection Operator
MC	Monte Carlo
MHD	Magnetohydrodynamic Pump
ML	Machine Learning
NPP	Nuclear Power Plant
PCT	Peak Cladding Temperature
PLS	Partial Least Squares
PK	Point Kinetic equations
PS	Protection System
SA	Sensitivity Analysis
SMDFR	Small Modular Dual Fluid Reactor
SMR	Small Modular Reactor
TH	Thermo-Hydraulic model
WB	White-Box model

Symbols

$g(\cdot)$	SMR state-space model
$x_m(t)$	m -th measurable input signal
$y(t)$	Model output of the SMR state-space model
$y_{WB}(t)$	White-box model output
$y_{GB}(t)$	Gray-box model output
$\Delta y(t)$	Modeling error estimation
y_{BB}	Vector of BB model outputs $[\Delta y, u]$
$u(t)$	Accident-related input signal (coolant leakage rate)
θ	Vector of system design parameters
t_n	Discrete time step
Δt	Sampling time
N	Total number of time samples
r	Memory window length

s	Prediction window length
t_N	Mission time
N_d	Number of simulated accident scenarios
t_f	Failure (fault-injection) time
$X(t_{n-r+1} : t_n)$	Memory matrix of input signals
$Y_{BB}(t_{n+1} : t_{n+s})$	BB multi-step predicted outputs
$c_{m,i}(t)$	i -th intrinsic mode function (IMF) of input x_m
$c_{j,l}(t)$	l -th IMF of output $y_{BB,j}$
$r_m(t), r_j(t)$	Residual of input/output signal after Empirical Mode Decomposition
I_m	Number of IMFs extracted from input m
L_j	Number of IMFs extracted from output j
$a_{m,i}(t)$	Instantaneous Amplitude (IA) of the i -th IMF of the m -th input
$a_{j,l}(t)$	Instantaneous Amplitude of the l -th IMF of the j -th output
$\omega_{m,i}(t)$	Instantaneous Frequency (IF) of the i -th IMF of m -th input
$\omega_{j,l}(t)$	Instantaneous Frequency of output l -th IMF of the j -th output
$\bar{a}_{m,i}, \bar{a}_{j,l}$	Mean Instantaneous Amplitude
$\bar{\omega}_{m,i}, \bar{\omega}_{j,l}$	Mean Instantaneous Frequency
$\bar{e}_{m,i}, \bar{e}_{j,l}$	Mean energy of each IMF
$\hat{\beta}_m, \hat{\beta}_j$	Slope coefficient of linear regression of the residual
$\hat{r}_{m,1}, \hat{r}_{i,2}$	Mean and standard deviation of input residual
$\hat{r}_{m,1}, \hat{r}_{j,2}$	Mean and standard deviation of output residual
f_m	Feature vector of the m -th input
f_j	Feature vector of j -th BB model output
$\mathcal{V}$	Matrix of input feature samples
$\mathcal{W}$	Matrix of output feature samples
V	Set of all extracted input features, $v \in V$
W	Set of all extracted output features, $w \in W$
$v_{d,n}$	Input feature sample at scenario d , time n
$w_{d,n}$	Output feature sample at scenario d , time n
$K_{p,q}$	First-order Kucherenko sensitivity index (features p and q)
$V(m)$	Subset of features corresponding to measurable input x_m
$W(j)$	Subset of features belonging to output j
$b_{m,q}^{(j)} = \max_{p \in V(m)} \{K_{p,q}\}$	Best feature of input m for output feature q
$K_m^{(j)}$	Grouped sensitivity index for input m
K_j^*	Maximum $K_m^{(j)}$ value for each output j
$\mathbf{x}^*$	Selected measurable input set for BB model
$\dot{m}_{\text{leak}}(t)$	Coolant leakage mass flow rate
k_{leak}	Leakage scale parameter
A_{leak}	Leakage shape parameter
B_{leak}	Leakage additional scale parameter
ϵ	Leakage disturbance term
$\dot{m}_{\text{ACS}}(t)$	Auxiliary Cooling System mass flow
$\dot{m}_{\kappa,\text{net}}(t)$	Net coolant mass flow (ACS–leakage)
$g(\cdot)$	High-fidelity model
$f_{\text{WB}}(\cdot)$	White-box model
$f_{\text{BB}}(\cdot)$	Black-box model
$\Delta y(t) = y(t) - y_{\text{WB}}(t)$	Modeling error
$Y_{\text{BB}} = [\Delta y, u]$	BB model output vector
$T_{f,h}(t)$	Fuel temperature at node h
$T_{w,h}(t)$	Wall temperature at node h
$T_{\kappa,h}(t)$	Coolant temperature at node h
$\dot{m}_{\kappa,h}(t)$	Coolant mass flow rate at node h
$T_p(t) = \max_h \{T_{w,h}(t)\}$	Peak Cladding Temperature (PCT) proxy

$T_{\kappa,\text{in}}$	Inlet coolant temperature
$T_{\kappa,\text{out}}$	Outlet coolant temperature
$T_{f,\text{in}}$	Fuel inlet temperature
$T_{f,\text{out}}$	Fuel outlet temperature

References

1. Nuclear Energy Agency. Risk Monitors: The State of The Art Report in their Development and Use at Nuclear Power Plants. OECD. 2004. Available online: https://www.oecd-nea.org/jcms/pl_18136/risk-monitors-the-state-of-the-art-report-soar-in-their-development-and-use-at-nuclear-power-plants-produced-on-behalf-of-the-iaea-and-the-nea-wgrisk?details=true (accessed on 6 June 2022).
2. Coble, J.B.; Coles, G.A.; Ramuhalli, P.; Meyer, R.M.; Berglin, E.J.; Wootan, D.W.; Mitchell, M.R. *Technical Needs for Enhancing Risk Monitors with Equipment Condition Assessment for Advanced Small Modular Reactors*; U.S. Department of Energy: Richland, WA, USA, 2013. [CrossRef]
3. Williams, Q.J.; Stewart, R.H.; Palmer, T.S.; Palmer, C.J.; Pope, C.; Shields, A.; Ritter, C. Selection of Sampling and Surrogate Modeling Methods for State-Point Evaluations of an AGN-201M Reactor. *Nucl. Sci. Eng.* **2025**, *200*, S391–S405. [CrossRef]
4. Nuclear Energy Agency. Small Modular Reactors: Challenges and Opportunities. 2021. Available online: https://www.oecd-nea.org/upload/docs/application/pdf/2021-03/7560_smr_report.pdf (accessed on 29 April 2026).
5. Fossum, K.L.; Bhowmik, P.K.; Sabharwall, P. Droplet Entrainment in Steam Supply System of Water-Cooled Small Modular Reactors: Experiment and Modeling Approaches. *J. Nucl. Eng.* **2024**, *5*, 563–583. [CrossRef]
6. Sarrañ, L.; Smith, K.M.; Hviid, C.A.; Rode, C. Grey-box modelling and virtual sensors enabling continuous commissioning of hydronic floor heating. *Energy* **2022**, *261*, 125282. [CrossRef]
7. Ahmad, I.; Ayub, A.; Kano, M.; Cheema, I.I. Gray-box soft sensors in process industry: Current practice, and future prospects in era of big data. *Processes* **2020**, *8*, 243. [CrossRef]
8. Hossain, R.; Ahmed, F.; Kobayashi, K.; Koric, S.; Abueidda, D.; Alam, S.B. Virtual sensing-enabled digital twin framework for real-time monitoring of nuclear systems leveraging deep neural operators. *npj Mater. Degrad.* **2025**, *9*, 21. [CrossRef]
9. Gong, L.; Peng, C.; Huang, Q. Deterministic Data Assimilation in Thermal-Hydraulic Analysis: Application to Natural Circulation Loops. *J. Nucl. Eng.* **2025**, *6*, 23. [CrossRef]
10. Tulleken, H.J.A.F. Grey-box modelling and identification using physical knowledge and bayesian techniques. *Automatica* **1993**, *29*, 285–308. [CrossRef]
11. Xue, Y.; Zhang, B.; Su, K.; Li, Y.; Zhu, H.; Pan, H. A preliminary study of digital twin for nuclear reactor dynamics: A synergy of machine learning and model predictive control. *Eng. Appl. Artif. Intell.* **2025**, *153*, 110940. [CrossRef]
12. Pintelas, E.; Livieris, I.E.; Pintelas, P. A Grey-Box Ensemble Model Exploiting Black-Box Accuracy and White-Box Intrinsic Interpretability. *Algorithms* **2020**, *13*, 17. [CrossRef]
13. Sahadath, M.H.; Cheng, Q.; Pan, S.; Ji, W. Characterization of DeepONet Performance for Neutron Transport Modeling. *Nucl. Sci. Eng.* **2026**, 1–21. [CrossRef]
14. IAEA. *Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry*; Technical Report No. NR-T-2.16; IAEA: Vienna, Austria, 2025. [CrossRef]
15. Gu, W.; He, Y.; Wang, D. Critical flow break source term near the outlet of a slit. *Prog. Nucl. Energy* **2025**, *188*, 105878. [CrossRef]
16. Yaseen, M.; Wu, X. Quantification of Deep Neural Network Prediction Uncertainties for VVUQ of Machine Learning Models. *Nucl. Sci. Eng.* **2023**, *197*, 947–966. [CrossRef]
17. Yadav, V.; Agarwal, V.; Jain, P.; Ramuhalli, P.; Zhao, X.; Ulmer, C.; Carlson, J.; Eskins, D.; Iyengar, R. *Technical Challenges and Gaps in Digital-Twin-Enabling Technologies for Nuclear Reactor Applications*; U.S. Nuclear Regulatory Commission: Washington, DC, USA. Available online: <https://www.nrc.gov/docs/ML2136/ML21361A261.pdf> (accessed on 2 March 2025).
18. Xiong, Q.; Du, P.; Deng, J.; Huang, D.; Song, G.; Qian, L.; Wu, Z.; Luo, Y. Global sensitivity analysis for nuclear reactor LBLOCA with time-dependent outputs. *Reliab. Eng. Syst. Saf.* **2022**, *221*, 108337. [CrossRef]
19. Alexanderian, A.; Gremaud, P.A.; Smith, R.C. Variance-based sensitivity analysis for time-dependent processes. *Reliab. Eng. Syst. Saf.* **2020**, *196*, 106722. [CrossRef]
20. Yu, H.; Chang, L.; Yang, M.; Chen, S.; Li, H.; Wang, J. Time series modeling and forecasting with feature decomposition and interaction for prognostics and health management in nuclear power plant. *Energy* **2025**, *324*, 135784. [CrossRef]
21. Nguyen, H.-P.; Baraldi, P.; Zio, E. Ensemble empirical mode decomposition and long short-term memory neural network for multi-step predictions of time series signals in nuclear power plants. *Appl. Energy* **2021**, *283*, 116346. [CrossRef]
22. Di Maio, F.; Pedroni, N.; Tóth, B.; Burgazzi, L.; Zio, E. Reliability Assessment of Passive Safety Systems for Nuclear Energy Applications: State-of-the-Art and Open Issues. *Energies* **2021**, *14*, 4688. [CrossRef]
23. Harter, J.R.; DeHart, M.D. Uncertainty quantification and sensitivity analysis of a nuclear thermal propulsion reactor startup sequence. *Front. Nucl. Eng.* **2025**, *4*, 1628866. [CrossRef]

24. Zio, E. Advancing nuclear safety. *Front. Nucl. Eng.* **2024**, *2*, 1346555. [\[CrossRef\]](#)
25. Kobayashi, K.; Kumar, D.; Bonney, M.; Chakraborty, S.; Paaren, K.; Usman, S.; Alam, S. Uncertainty Quantification and Sensitivity Analysis for Digital Twin Enabling Technology: Application for BISON Fuel Performance Code. In *Handbook of Smart Energy Systems*; Springer: Cham, Switzerland, 2023; pp. 1–13. [\[CrossRef\]](#)
26. Shi, W.; Machida, M.; Yamada, S.; Yoshida, T.; Hasegawa, Y.; Okamoto, K. Inverse estimation scheme of radioactive source distributions inside building rooms based on monitoring air dose rates using LASSO: Theory and demonstration. *Prog. Nucl. Energy* **2023**, *162*, 104792. [\[CrossRef\]](#)
27. Roche, A. Local optimization of black-box functions with high or infinite-dimensional inputs: Application to nuclear safety. *Comput. Stat.* **2018**, *33*, 467–485. [\[CrossRef\]](#)
28. Walker, C.; Ramuhalli, P.; Agarwal, V.; Lybeck, N.J.; Taylor, M. Development of Short-Term Forecasting Models Using Plant Asset Data and Feature Selection. *Int. J. Progn. Health Manag.* **2022**, *13*. [\[CrossRef\]](#)
29. Huang, N.E. Chapter 1: Introduction to the Hilbert-Huang Transform and its related mathematical problems. In *Hilbert–Huang Transform and Its Applications*, 2nd ed.; Huang, N., Shen, S., Eds.; World Scientific Publishing: Singapore, 2014; Volume 16, pp. 1–26. [\[CrossRef\]](#)
30. Huang, N.E.; Wu, Z. A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Rev. Geophys.* **2008**, *46*, RG2006. [\[CrossRef\]](#)
31. Kucherenko, S.; Tarantola, S.; Annoni, P. Estimation of global sensitivity indices for models with dependent variables. *Comput. Phys. Commun.* **2012**, *183*, 937–946. [\[CrossRef\]](#)
32. Sobol', I.M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Math. Comput. Simul.* **2001**, *55*, 271–280. [\[CrossRef\]](#)
33. Miqueles, L.; Ahmed, I.; Di Maio, F.; Zio, E. Virtual sensing by Grey-Box modelling within an Importance Sampling Monte Carlo Dynamic Event Tree framework for risk monitoring of Small Modular Reactors. *Reliab. Eng. Syst. Saf.* **2026**, *272*, 112629. [\[CrossRef\]](#)
34. Chen, L.; Huang, H. Global sensitivity analysis for multivariate outputs using generalized RBF-PCE metamodel enhanced by variance-based sequential sampling. *Appl. Math. Model.* **2024**, *126*, 381–404. [\[CrossRef\]](#)
35. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.-C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. Lond. Ser. A* **1998**, *454*, 903–995. [\[CrossRef\]](#)
36. Marelli, S.; Lamas, C.; Sudret, B. UQLab User Manual—Sensitivity Analysis. 2015. Available online: <https://www.uqlab.com/sensitivity-user-manual> (accessed on 20 February 2025).
37. Plischke, E.; Borgonovo, E. Fighting the Curse of Sparsity: Probabilistic Sensitivity Measures From Cumulative Distribution Functions. *Risk Anal.* **2020**, *40*, 2639–2660. [\[CrossRef\]](#)
38. Benoumechiara, N.; Elie-Dit-Cosaque, K. Shapley effects for sensitivity analysis with dependent inputs: Bootstrap and kriging-based algorithms. *ESAIM Proc. Surv.* **2019**, *65*, 266–293. [\[CrossRef\]](#)
39. Lewitz, J.; Huke, A.; Ruprecht, G.; Weißbach, D.; Gottlieb, S.; Hussein, A.; Czerski, K. The Dual Fluid Reactor—An Innovative Fast Nuclear-Reactor Concept with High Efficiency and Total Burnup. *Int. J. Nucl. Power* **2020**, *65*, 145–154.
40. Liu, C.; Luo, R.; Macián-Juan, R. A New Uncertainty-Based Control Scheme of the Small Modular Dual Fluid Reactor and Its Optimization. *Energies* **2021**, *14*, 6708. [\[CrossRef\]](#)
41. Miqueles, L.; Ahmed, I.; Di Maio, F.; Zio, E. Importance Sampling for Monte Carlo Dynamic Event Tree Analysis of Accident Scenarios in New-Generation Nuclear Power Plants. *Nucl. Sci. Eng.* **2025**, *200*, 1296–1322. [\[CrossRef\]](#)
42. Park, J.H.; An, Y.J.; Yoo, K.H.; Na, M.G. Leak flow prediction during loss of coolant accidents using deep fuzzy neural networks. *Nucl. Eng. Technol.* **2021**, *53*, 2547–2555. [\[CrossRef\]](#)
43. Wang, W.; Cammi, A.; Di Maio, F.; Lorenzi, S.; Zio, E. A Monte Carlo-based exploration framework for identifying components vulnerable to cyber threats in nuclear power plants. *Reliab. Eng. Syst. Saf.* **2018**, *175*, 24–37. [\[CrossRef\]](#)
44. Neubauer, A.; Brandt, S.; Kriegel, M. Explainable multi-step heating load forecasting: Using SHAP values and temporal attention mechanisms for enhanced interpretability. *Energy AI* **2025**, *20*, 100480. [\[CrossRef\]](#)
45. Lundberg, S.M.; Allen, P.G.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4768–4777. Available online: <https://arxiv.org/pdf/1705.07874> (accessed on 29 April 2026).

46. *BS IEC 60737:2010*; Nuclear Power Plants-Instrumentation Important to Safety-Temperature Sensors (In-Core and Primary Coolant Circuit)-Characteristics and Test Methods. International Electrotechnical Commission (IEC): Geneva, Switzerland, 2010.
47. *BS IEC 60751:2022*; Industrial Platinum Resistance Thermometers and Platinum Temperature Sensors. International Electrotechnical Commission (IEC): Geneva, Switzerland; British Standards Institution: London, UK, 2022.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.