

Article

Functional Outlier Detection by Means of h-Mode Depth and Dynamic Time Warping

Álvaro Rollón de Pinedo ^{1,2}, Mathieu Couplet ¹, Bertrand Iooss ^{1,3,*} , Nathalie Marie ⁴, Amandine Marrel ^{3,4} , Elsa Merle ⁵  and Roman Sueur ¹

¹ EDF R&D, 6 Quai Watier, 78400 Chatou, France; alvaro.rollon-de-pinedo@edf.fr (Á.R.d.P.); mathieu.couplet@edf.fr (M.C.); roman.sueur@edf.fr (R.S.)

² I-MEP2 Doctoral School, Université Grenoble-Alpes, 621 Avenue Centrale, 38400 Saint-Martin-d'Hères, France

³ Institut de Mathématiques de Toulouse, 31062 Toulouse, France; amandine.marrel@cea.fr

⁴ CEA, DES, IRESNE, 13108 Saint-Paul-lez-Durance, France; nathalie.seiler@cea.fr

⁵ LPSC Grenoble, 53 Avenue des Martyrs, 38000 Grenoble, France; elsa.merle@phelma.grenoble-inp.fr

* Correspondence: bertrand.iooss@edf.fr

Abstract: Finding outliers in functional infinite-dimensional vector spaces is widely present in the industry for data that may originate from physical measurements or numerical simulations. An automatic and unsupervised process of outlier identification can help ensure the quality of a dataset (trimming), validate the results of industrial simulation codes, or detect specific phenomena or anomalies. This paper focuses on data originating from expensive simulation codes to take into account the realistic case where only a limited quantity of information about the studied process is available. A detection methodology based on different features, such as h-mode depth or the dynamic time warping, is proposed to evaluate the outlyingness both in the magnitude and shape senses. Theoretical examples are used to identify pertinent feature combinations and showcase the quality of the detection method with respect to state-of-the-art methodologies of detection. Finally, we show the practical interest of the method in an industrial context thanks to a nuclear thermal-hydraulic use case and how it can serve as a tool to perform sensitivity analysis on functional data.

Keywords: functional data analysis; functional outlier detection; probabilistic modeling; computer experiments; sensitivity analysis



check for
updates

Citation: Rollón de Pinedo, Á.; Couplet, M.; Iooss, B.; Marie, N.; Marrel, A.; Merle, E.; Sueur, R. Functional Outlier Detection by Means of h-Mode Depth and Dynamic Time Warping. *Appl. Sci.* **2021**, *11*, 11475. <https://doi.org/10.3390/app112311475>

Academic Editor: Markus Goldstein

Received: 8 October 2021

Accepted: 23 November 2021

Published: 3 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Presently, the ever-increasing capabilities of measurement, generation, and storing of data have increased the scientific and industrial communities' interest in what are known as functional data. In its most general form, the domain of functional data analysis deals with any object with the form of a function, regardless of its dimension. For example, the most basic form these objects can adopt is one-dimensional (i.e., univariate) real functions, which can represent the evolution of a physical parameter of interest over time (or another variable). These data normally originate through an empirical measurement system or a simulation code.

The great interest of the functional data analysis is that it allows consideration of the intrinsic nature of the data, i.e., the underlying process that generates them is supposed to have certain restrictions of domain, regularity, continuity and so on. As the functional data are infinite-dimensional by nature, functional data analysis methods always rely on a dimension reduction technique, whether implicitly or explicitly. More precisely, it is the case in the context of classification [1], clustering [2], landmark research and registration [3] of functional data. By providing more synthetic descriptors of functional observations, functional data analysis methods allow a more practical treatment of data thanks to the available multivariate tools.

The domain of functional data analysis can be traced back to the works of Grenander [4], but the term itself was coined by Ramsay, [5]. Some of the most widely regarded references for the domain include the general works of [6] or [7], focused on non-parametric methods for functional data. Functional data can be found in a wide variety of contexts, among which we can mention: environmental sciences [8,9], medical sciences [10], economy [11] and others. In the context of this paper, our work is motivated using expensive simulation codes [12,13] that are commonly used in the nuclear industry to complement the nuclear safety assessment reports [14], and which require specific statistical tools. These thermal-hydraulic simulators, such as the system code CATHARE2 [15], can simulate the evolution of the essential physical parameters during a nuclear transient, and they are expensive to evaluate (several hours per run) and analyze due to the complexity of the simulated phenomena (highly nonlinear interactions between its parameters).

In general, in the context of nuclear safety, the main analyzed parameters are the scalar values known as *Safety Parameters*, which are representative of the severity of an accidental transient. However, the analysis of the whole evolution of the aforementioned physical parameters is a complex domain that has been subject to considerable research efforts in recent years, such as in the works of [16] or [17]. The detection and analysis of outlying transients in these sets of simulations are useful to detect unexpected physical phenomena, validate the simulations providing valuable descriptors of the functional outputs of interest, or quantify the extreme behavior or penalizing nature of specific subsets of those transients.

Despite the diversity of domains, there are several common points of interest between all of them. For example, in the most usual case of univariate functions, a typical difficulty that arises in the pre-treatment phase of the data are the subject of the discretization of the grid on which the data are observed. Univariate curves are not normally observed in their entirety (since they are intrinsically infinite-dimensional). Instead, they are indexed by some variable, usually the time in the univariate case, in a grid that can be homogeneous or not. Indeed, points are not necessarily equally spaced, as with biological measurements such as blood pressures, or with temporal data originated by a simulation code whose time step evolves accordingly with the convergence impositions of the numerical solver. Naturally, the representation of the underlying functional process is still a matter of discussion. It does not only depend on the spacing of each point in the time grid but also other notions such as the *density* of data. This quantity can differ by several orders of magnitude in the case of numerical simulators depending on the restrictions imposed to time steps to guarantee the convergence, the presence or not of random noise, etc. For a more profound analysis on the importance of these subjects, the reader can refer to [18].

This paper focuses on a functional outlier detection technique, which is an unsupervised problem, showing its sensitivity to both magnitude and shape outliers and its ranking capabilities. The importance of this work is closely related to the data quality domain, since the presence of anomalous behaviors that might have been originated by measurement errors, non-convergence of algorithms or the existence of non-physical values of numerical simulators may produce spurious results when treating a dataset. As explained in [19], when the underlying process that creates the data *behaves unusually*, it results in the creation of outliers. Therefore, it may contain useful information about *abnormal characteristics of the system*. Finally, the particularities of functional data treatment with respect to multivariate ones require specific tools to have acceptable detection capabilities [20].

The functional outlier detection domain has gained relevance in recent years, and the methods for detecting these outliers do not cease to improve. However, the techniques can differ significantly between them. Many of them rely on the notion of statistical depth [9,21,22] or other measures (e.g., entropy [23]). In contrast, others rely on other dimensionality reduction methods, clustering, or hypothesis testing on the coefficients of some functional decomposition [24], as well as graphical tools [25]. A simulation study comparing depth measurements to detect functional outliers can be found in [26]. Naturally, all these techniques showcase different detection capabilities, and they can be more adapted to the detection of a particular type of outliers.

In our study, we consider that the data correspond to independent and identically distributed (i.i.d.) realizations of a random process Z , taking its values in a functional space \mathcal{F} . In practice, any realization z_i of the random variable Z can only be observed in a limited number of points in the domain, i.e., it is observed as the random vector $(z_i(t_1), \dots, z_i(t_p))$, with $\{t_1, \dots, t_p\} \in \mathcal{T} \subset \mathbb{R}$. This data representation is not itself appropriate to the detection of outlying functions. Section 2 thus presents how to define and compute a modeling of data dedicated to the characterization of outliers, using Gaussian Mixture Models (GMM). The introduction of the measures and methods will be found in the Supplementary Material of this paper. Based on this data representation, Section 3 presents our functional outlier detection methodology. Sections 4 and 5 provide some analytical and industrial application examples. The Supplementary Material of this paper also contains complementary numerical tests. The properties, capabilities and limitations of the methodology are discussed in Section 6.

2. Adapting Gaussian Mixture Models for Outlier Detection

As mentioned in the introduction, functional data are elements of a functional space, typically functions defined on a continuous interval of \mathbb{R} . Measuring, storing and analyzing these data is however realized using numerical devices and computers, and hence impose a digital representation of data. For this reason, the value of functions is available only on a finite-dimensional sub-domain of their theoretical definition domain, i.e., in a discretized version. When dealing with time-dependent physical quantities, this discretization basically consists of a projection on a time grid which can be regular or not. To achieve an efficient detection of outlying functions, a transformation of the dataset is required, which brings out the features of the data that discriminate outlying ones from the rest of the set.

The use of probabilistic models in lower-dimensional feature spaces is useful to detect anomalous points assuming that they provide a good fit for the data and are able of capturing the central trend of the data. In this context, the use of Gaussian Mixture Models (GMM) is a common approach [20], since they can provide a good fit to the data if the number of components is not fixed, while providing an estimate of the generative probability of each point according to the model. However, their basic use suffers from well-known spurious effects in the outlier detection domain [19]. The main problems are:

- If the probabilistic model is adjusted including the potential outliers, they may bias the estimation of the underlying model. This is especially problematic if the outliers are assumed to be generated by a different distribution than the other data and are not only considered to be extreme realizations of the same underlying process as the others. On top of that, if the sample presents a high degree of contamination or the sample is small, this bias can greatly influence the detection.
- If the multivariate sample can be classified in several different clusters but they number of components is not well-chosen, the possibility of overfitting the probabilistic model to the data becomes a real problem. In this case, some small-sized clusters may appear overly adjusted to the outliers, which will not be identified as such.

These reasons motivate the modifications of the Expectation Maximization (EM) algorithm used in order to fit the GMM to the sample of data in the considered feature space, $\mathcal{S} \subset \mathbb{R}^R$, with $\mathcal{U} = \{u_r\}_{r=1}^R$ being the set of features such that $\forall u_r \in \mathcal{U}, u_r : \mathcal{F} \mapsto \mathbb{R}$. The objective is to generate a probabilistic model appropriate for outlier detection.

Let $\{z_1, \dots, z_i, \dots, z_n\}$ be an i.i.d. sample of functional data, with $\{\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_n\}$ being the set of associated features for each curve (for clarity, $\mathbf{u}_i = (u_1(z_i), \dots, u_r(z_i), \dots, u_R(z_i))$), the objective is the estimation of the set of parameters of the GMM $\{\omega, \mu_k, \Sigma_k\}_{k=1}^K$. This optimization problem is usually solved by maximum likelihood estimation via the EM algorithm [27]. The form of the problem is:

$$\max_{\omega \in \mathbb{R}^K, \{\mu_k, \Sigma_k\}_{k=1}^K} \sum_{i=1}^n \log \left(\sum_{k=1}^K f_k(\mathbf{u}_i; \mu_k, \Sigma_k) \right). \quad (1)$$

This well-known algorithm consists of maximizing the log-likelihood function of the Gaussian mixture when a certain amount of binary latent variables g such that $g_k \in \{0, 1\}$ and $\sum_{k=1}^K g_k = 1$, and which allow the representation of the corresponding k th gaussian component that is considered each time. This way, the conditional distribution of $p(\mathbf{u}|g_k = 1) = f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ and $p(\mathbf{u}) = \sum_{k=1}^K \omega_k f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. The last element required for the estimation of the GMM parameters in each step are the conditional probabilities of g given \mathbf{u} (usually called *responsibility* that component k takes in explaining the observation). Their values are:

$$\gamma(g_k) = p(g_k = 1|\mathbf{u}) = \frac{p(g_k = 1)p(\mathbf{u}|g_k = 1)}{\sum_{j=1}^K p(g_j = 1)p(\mathbf{u}|g_j = 1)} = \frac{\omega_k f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \omega_j f_j(\mathbf{u}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{2}$$

The estimation of the parameters is then performed in the following steps:

1. Initialize the values of the desired parameters $\{\omega_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$,
2. E-step: Evaluate the current responsibilities with the previous parameter values,
3. M-step: Re-estimate the parameters using the responsibilities,
4. Evaluate the log-likelihood function presented in Equation (1).

The modifications of the algorithm consist of iteratively check and reinitialize the estimated parameters during the procedure to detect the undesired and spurious effects by adding two steps after the estimation of the likelihood function:

1. Check if $\mathbf{u}_i = \boldsymbol{\mu}_k$ for any $k \in \{1, \dots, K\}$. The appearance of these singularities may maximize the log-likelihood function to infinity, since it is unbounded, and cause an overfitting of the data to isolated points. This phenomenon is well described in [28]. If this kind of anomaly is detected, the point is subtracted from the sample and the EM algorithm continues without it. Naturally, these points will be subject to close analysis since they are good candidates for potential outliers in the sample.
2. For each iteration step, ω_k can be viewed as the prior probability of $g_k = 1$, and $\gamma(g_k)$ can be seen as the posterior probability given the observed sample. If this posterior probability is considered too low (in our applications we shall take a value of 0.1 as the minimum weight of the mixing coefficients), we will consider that the corresponding component is either overfitting the data, or that it has detected a small subset of points which is not representative of the central trend of the data. In this case, the other calculated parameters of the components are kept and the values of means and covariances of the small cluster are reinitialized to a random value in the space.

These modifications allow for the GMM that is fitted to the sample in the feature space to stay general, and provide good candidates of outliers during the construction of the model, all of that while avoiding overfitting.

3. Outlier Detection

3.1. Test for Outlyingness

Thanks to these elements, we can construct a hypothesis test by making use of the components of the GMM adjusted model. This way, for any $\mathbf{u}_i \in \mathbb{R}^R$:

$$\begin{aligned} H_0 : & \mathbf{u}_i \text{ has been generated by } f_k \text{ with probability at least } p_{\alpha k} \\ H_1 : & \mathbf{u}_i \text{ is an outlier.} \end{aligned} \tag{3}$$

Under H_0 , $p(\mathbf{u}_i|g_k = 1) > p_{\alpha k}$, where $p_{\alpha k} = f_k(\mathbf{u}_\alpha|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ such that $\mathbb{P}(\mathbf{u}_\alpha|g_k = 1) \geq \alpha$. The set of data points less likely than \mathbf{u}_α is composed by points $\mathbf{u} \in \mathcal{S} \subset \mathbb{R}^R$ which verify:

$$(\mathbf{u} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u} - \boldsymbol{\mu}_k) \geq (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k). \tag{4}$$

And therefore, $\mathbb{P}[(\mathbf{u} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u} - \boldsymbol{\mu}_k) \geq (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)] = 1 - \mathbb{P}[L \leq (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)]$, where L follows a Chi-squared distribution, $L \sim \chi^2(k)$. By

performing this test over all the points considered in the feature space, we obtain a unique criterion for outlier detection, such that the outlying points will be the ones presenting p -values under a certain threshold $\alpha, \forall f_k$.

3.2. Ordering Score in the Feature Space

Once all the parameters are fixed at their estimated values, it is possible to quantify the extremal behavior of any point in the considered multivariate space through the notion of *Level Set* [29]. Formally, given an absolutely continuous probability measure ν with density p , the Minimum Level Set D^θ is the set of minimal volume (according to the Lebesgue measure) whose measure is θ :

$$D^\theta = \underset{D \subset \mathbb{R}^R, \nu(D) \geq \theta}{\operatorname{argmin}} \lambda(D) \tag{5}$$

where λ is the Lebesgue measure in \mathbb{R}^R . Due to the concave downward nature of the GMM [30], the set D is unique and $\nu(D^\theta) = \theta$. This way, a probabilistic score of outlyingness for functional data can be obtained via the probability mass retained in the associated Level Set of each functional datum z_i :

$$\hat{\theta}_i = \int_{\mathbb{R}^R} \hat{p}(\mathbf{u}) \mathbb{1}_{\{\hat{p}(\mathbf{u}) \geq \hat{p}(u_i)\}} d^R \mathbf{u} \tag{6}$$

where $\mathbb{1}$ is an indicator function, and \hat{p} is the adjusted probability density function of the GMM with the estimated parameters:

$$\hat{p}(\mathbf{u}; \omega, \hat{\mu}_k, \hat{\Sigma}_k) = \sum_{k=1}^K \hat{\omega}_k \frac{1}{2\pi \sqrt{|\hat{\Sigma}_k|}} \exp\left(-\frac{1}{2}(\mathbf{u} - \hat{\mu}_k)^T \hat{\Sigma}_k^{-1}(\mathbf{u} - \hat{\mu}_k)\right). \tag{7}$$

This integral must be solved by numerical methods. For $R = 1, 2$, there exists efficient software able to solve it, whereas several algorithms based on the Cholesky decomposition have been developed for the higher-dimensional cases [31]. In the applications presented here we will limit ourselves to two components, orienting each one of them to magnitude or shape outlier detection.

The above presented level set notion gives an unambiguous way of defining to what extend an observation is unlikely to be observed when assuming it was generated in accordance with the same probability law as the rest of the dataset. The GMM provides in turn the required underlying probabilistic model allowing this definition. With the resulting outlyingness scores θ_i , we have now available a properly constructed quantification tool for the degree of outlyingness of an observation. We can thus use common statistical approaches to implement an outlier detection procedure, such as considering the occurrence probability of a data point is too small when $\theta_i < \alpha$, where α is a significance level.

Finally, let us consider the more realistic case where the availability of data is actually limited, and the sample of functional data is too small to generate the GMM with an acceptable level of reliability (the global convergence of the EM algorithm is not necessarily reached). This is for instance the case for expensive industrial simulation codes, such as mechanical or thermal-hydraulic simulators. In this case, a natural extension of this idea for outlier detection can be implemented via bootstrap resampling, [32]. B groups are formed by successively drawn with replacement in the original sample. This way, the absence of data can be mitigated through the re-estimation of the GMM for each bootstrap group. If for B bootstrap groups $p_b(\mathbf{u})$ represents the GMM of the b th group, the form of the (bootstrap) estimator of outlyingness would then be:

$$\hat{\theta}_i = \int_{\mathbb{R}^R} \frac{1}{B} \sum_{b=1}^B \hat{p}_b(\mathbf{u}) \mathbb{1}_{\{\hat{p}_b(\mathbf{u}) \geq \hat{p}_b(u_i)\}} d^R \mathbf{u} \tag{8}$$

Throughout this reasoning, the hyperparameter K (the number of components of the mixture model) has been supposed to be fixed, but in practice, this is yet another input parameter of the GMM that must be provided *a priori* to the EM algorithm. Indeed, the actual form of the model is significantly different depending on the number of components that are considered. If that is the case, the use of an oversimplified mixture when modeling complex multivariate distributions can induce incorrect conclusions about the distribution of data, whereas an unnecessary increase in the number of components may lead to overfitting problems, unacceptable computational costs or imprecise conclusions.

This issue can be treated as a model-selection problem, and several metrics are available to estimate an appropriate number of components depending on the sample. Some examples are the Bayesian Information Criterion (BIC) [33] or the Integrated Completed Likelihood [34]. In this paper, the selection of the number of components is performed by means of the Bayesian Information Criterion:

$$BIC = 2 \log(\hat{l}) - k \log(n) \quad (9)$$

where \hat{l} represents the log-likelihood function for the GMM, k is the chosen number of components and n is the sample size used for the estimation. The second term introduces a penalty which depends on the number of components to mitigate overfitting effects.

3.3. Proposed Detection Algorithm

The practical implementation of the detection algorithm is presented below, and several clarifications will be made. First, the approach does not include a projection of the considered functional data onto a functional expansion to homogenize the dataset. This preliminary step may be important in applications with heterogeneous time grids, if the data are observed with noise etc. Here, it will be supposed that the domain of the data is already uniform. Furthermore, this kind of projections are not necessary most times when estimating non-parametric features, such as L^2 norms.

Secondly, it is important to note that in the ideal case, the estimation of the GMM should be made based on a non-contaminated dataset, i.e., without outliers. These existent outliers should be separated before the estimation if the objective is to model the joint distribution of the desired features to avoid the introduction of a non-controlled bias. However, this knowledge is not available *a priori*, which justifies the trimming procedure of Algorithm 1. It is worth mentioning the iterative step. Even though it is not explicitly noted, the K components that are retained depend on the chosen sample for each bootstrap group. The task of reestimating the parameters for each GMM to optimize the number of components is not computationally expensive for small values of K , which is appropriate for datasets that do not present a high degree of multimodality and to avoid overfitting the data. In our simulations, $K_{max} = 10$.

Another important remark on the algorithm is the *extraction* step, where the most outlying functions on the set according to the $\hat{\theta}_i$ are separated from the original dataset, and the procedure is re-applied to the remaining functions until a desired homogeneity level is reached. Naturally, in the case where many samples are available, some extreme data realizations are bound to appear. Therefore, regardless of the chosen value for α , these significant data might be separated from the set even though they do not actually contaminate the set since, although unlikely, these data appear due to the nature of the underlying generative process. Separating (trimming) all these data may introduce a bias in the estimation of the successive GMM. This is a known problem in the outlier detection domain. An example for the functional data case can be found in [9].

Algorithm 1: FOD algorithm

Result: Set of outlyingness scores $\{\theta_i\}_{i=1}^n$, and set of outliers $o \subset \{z_i\}_{i=1}^n$

1. Check if the sample of functional data is uniformly sampled. Otherwise, Project the data onto a basis of \mathcal{F} until a desired level of error $\|\hat{z}_i - z_i\|$ (with \hat{z}_i being the projected samples on the basis) is reached. Uniformly resample in this grid;
2. Choose a family of R feature functions and project $\{z_i\}_{i=1}^n$ onto the feature space $\mathcal{S} \subset \mathbb{R}^R$;

3. **for** $K \in \{1, \dots, K_{max}\}$ **do**

1. Choose (manually or randomly) a starting set of parameters $\nu^{old} = \{\mu_k^{old}, \Sigma_k^{old}, w_k^{old}\}_{k=1}^K$;
2. **E-step.** Evaluate the corresponding responsibilities $\gamma(g_k)$ for each given point $u_i \in \{u_i\}_{i=1}^n$;
3. **M-step.** Re-estimate the parameters through:

$$\mu_k^{new} = \frac{1}{n_k} \sum_{i=1}^n \gamma(g_k) u_i$$

$$\Sigma_k^{new} = \frac{1}{n_k} \sum_{i=1}^n \gamma(g_k) (u_i - \mu_k^{new})(u_i - \mu_k^{new})$$

$$w_k^{new} = \frac{n_k}{n}$$

where n_k represents the number of observations assigned to the k th Gaussian in the mixture, i.e., $n_k = \sum_{k=1}^K \gamma(g_k)$;

4. Check $\|u_i - \mu_k\|_{L^2} < \epsilon, \epsilon = \inf_S \|u_i - u_j\|_{L^2}, \forall i, j \in \{1, \dots, n\} \times \{1, \dots, n\}$;
5. For the given level α , check $w_k < \alpha, \forall k \in \{1, \dots, K\}$
6. Check for convergence through the likelihood function;
7. $BIC_K = BIC(\hat{p}_K)$. With \hat{p}_K being the estimated GMM model with K components;

end

4. Set $K = \min(\{BIC_K\}_{K=1}^{K_{max}})$;
5. For a level α , test $H_0, H_1 \forall u_i \in \{u_i\}_{i=1}^n$. Obtain o ;
6. Estimate

$$\hat{\theta}_i = \int_{\mathbb{R}^R} \frac{1}{B} \sum_{b=1}^B \hat{p}_b(u) \mathbb{1}_{\{\hat{p}_b(u) \geq \hat{p}_b(u_i)\}} d^R u, \theta_i \forall i \in \{1, \dots, n\}$$

4. Analytical Test Cases

Following [35–37] the detection capabilities of the algorithm can be assessed via controlled simulation examples. Such simulations make it possible to check whether a method succeeds in recognizing different kinds of outliers, when these are purposefully inserted in the dataset. Thus, inserted outliers are constructed to mimic typical outlying data engineers are facing in their daily practice. Some common notations for the analytical models are summarized in Table 1.

Table 1. Description of the common parameters of the models.

Notation	Description
$G(t)$	Centered Gaussian process of covariance function $\Sigma(t_1, t_2) = 0.3 \exp \frac{- t_1 - t_2 }{0.3}$
$Z(t)$	Functional random variable generating the main model
Z_0	Functional random variable generating the outliers
T_l	Random point uniformly generated in the definition domain of the function

4.1. Numerical Test Protocol

In all the simulation experiments, there will be a total number of $n = 50$ curves, 49 of which are generated by a reference model, and one is the contaminating outlier. The functions are defined in the interval $[0, 1]$, with a grid of 30 equally spaced points and $B = 10$ bootstrap groups. Using the previous notation for the DTW algorithm, $V = W = 30$ points. The description of the models for these control simulation tests is as follows:

- Model 1. $Z(t) = 4t + G(t)$ is the function generator for the reference set of curves. In this case, the outliers follow the distribution $Z_o(t) = 4t + G(t) + 2\mathbb{1}_{\{(t_I < t)\}}$.
- Model 2. The reference model for the curve generation remains $Z(t) = 4t + G(t)$, whereas the outliers are now generated from the distribution $Z_o(t) = 4t + G(t) + 2\mathbb{1}_{\{(t_I < t < t_I + 3)\}}$.
- Model 3. Here the reference model becomes $Z(t) = 30t(1 - t)^{3/2} + G(t)$. The outliers are generated from $Z_o(t) = 30(1 - t)t^{3/2} + G(t)$.
- Model 4. For this last case, we keep the reference model as it is for Model 1 and Model 2, but the outliers simply consist of the sole deterministic part $Z_o(t) = 4t$ (the Gaussian component is removed).

Similar models to 1 and 2 can be found in [35–37], although the multiplicative factor of the indicator function has been reduced to make the outliers less apparent. Model 3 was also considered in [25,35]. The fourth model was developed here to test the detection capabilities of pure shape outliers. These models (see Figure 1) or similar ones have already been tested with existing methodologies, and they were therefore included here to facilitate the comparability with other detection methods.

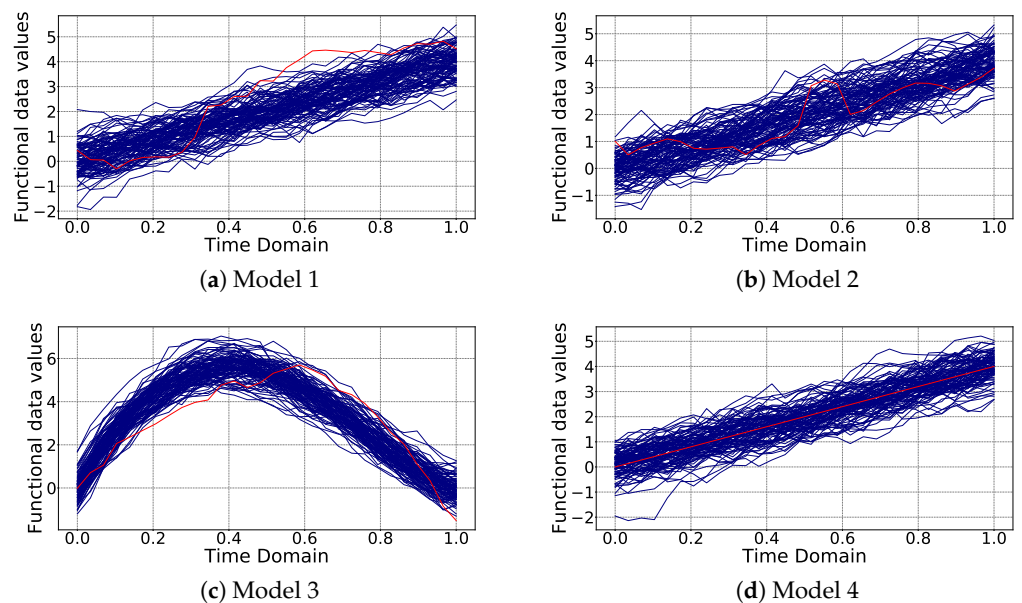


Figure 1. Examples of the four analytical test cases. The blue curves correspond to the 49 ones that are generated from the main model, whereas the red one corresponds to the outlier.

In all cases, a bivariate Gaussian mixture model is adjusted to a pair of selected features and the outlier detection procedure is applied thereafter. Four commonly used features in the functional data analysis framework will be considered:

- The modified band depth (BD) (see Equation (3) of the Supplementary Material),
- The h-mode depth (see Equation (4) of the Supplementary Material),
- The dynamic time warping (see Equation (7) of the Supplementary Material),
- The L^2 norm, which is one of the most intuitive and widely used metrics that can be applied to functional data. It takes the form: $\|z(t)\|_2 = \left(\int_{\mathbb{R}} |z(t)|^2 dt \right)^{1/2}$.

The detection procedure is applied to $N = 100$ replications of each model. We shall use two scores to evaluate the quality of the detection procedure. The first one will naturally be the estimated θ values of the outlier in each model and replication. This parameter is directly linked to the probability of being more anomalous than the outliers if the model is correct. Therefore, the distribution of values of $\theta_i, \forall i \in \{1, \dots, n\}$ constitutes an indicator of the detected outlying nature of the function.

The second score is the average ranking of the outlier with respect to the total population of data. Since the θ_i score provides an ordering of the anomalous nature of each element in the set of curves, so it is possible to rank the data accordingly to said metric. In industrial applications, this ranking can be followed by the engineer to analyze particular data (e.g., numerical simulations) from the most suspicious (potentially interesting) datum to less suspicious ones.

Several tests evaluating the performance of the different couples of measures can be found in the supplementary material of the paper. In particular, it can be seen that the use of the h-mode depth and the DTW seem appropriate for the detection of the considered types of outliers.

The use of the h-mode depth allows the detection of magnitude outliers taking into account the areas of the $(t, f(t))$ plane where the density of curves is less significant. This feature is useful since it allows the avoidance of problems derived from the consideration of multimodal distributions that may present several regions of high density and mask this type of outliers, a case which is not usually considered in the literature. Furthermore, the use of the normalized averaged version of the DTW that is considered in the paper allows the specific targeting of shape outliers. Its use allows the comparison of the functional data based on a pairwise comparison that only takes into account differences in shape, while being sensitive to shifts in the t-axis. In any case, their use has been tested against other similarity and depth measures, showcasing how the combination of these two *measures* provides the best identification grounds for outlier detection, both regarding the detection rates and the robustness.

Finally, we note that there exist similarities between the method presented here and in [23], but with significant differences. First, in the construction of the model, our paper proposes modifications to the EM algorithm in order to solve some of the main difficulties in the use of probabilistic models for outlier detection. Secondly, the measures selected in the algorithm presented in our paper allow for a certain explainability of why specific samples have been identified as outliers of shape, magnitude, or a mixture of both, which was also one of the main objectives. Lastly, the algorithm presented in our paper also aims at providing an order of outlyingness of the functional data, without necessarily labeling them as outliers. This outlyingness score aims at being used in domains where a quantitative score is more useful, such as in the sensitivity analysis one. An industrial example is provided in Section 5.

4.2. Comparison with State-of-the-Art Methodologies

The detection algorithm for the selected combination of features is performed to compare it to other detection methods. The selected methodologies are succinctly presented below.

4.2.1. Functional Boxplots

Given a sample of functional data z defined in a time domain \mathcal{T} indexed by the variable t , the 50% central region can be defined as:

$$C_{0.5} = \left\{ (t, z(t)) : \min_{r=1, \dots, n/2} z_r(t) \leq z(t) \leq \max_{r=1, \dots, n/2} z_r(t) \right\}. \quad (10)$$

This region can be interpreted as an analogous of the inter-quartile range for functional data, and it effectively corresponds to it pointwise. The whiskers of the functional boxplot can be computed by extending 1.5 times the pointwise extremes of the central region, such

that the outliers are detected if they surpass the frontiers defined by these whiskers. The in-depth analysis of this method can be found in [38].

4.2.2. High-Density Regions

Introduced by [39], the method consists of regrouping the values of the functional data in the considered time steps in a matrix and performing the Karhunen-Loève decomposition, obtaining the corresponding coefficients in a lower-dimensional feature space where the density of the components is estimated via kernel density estimation. This way, the High-Density Regions (HDR) can be defined as the regions such that: $\{u : f(u) \geq f_\alpha\}$, and will correspond to the region with highest probability density function with a cumulative probability of $1 - \alpha$, which can impose a detection criterion.

4.2.3. Directional Detector

As described in [40], let X be a stochastic process, $X : \mathcal{T} \rightarrow \mathbb{R}^p$, the Functional Directional Outlyingness is defined as:

$$FO(X, F_X) \int_{\mathcal{T}} \|O(X(t), F_{X(t)})\|^2 w(t) dt \quad (11)$$

where $w(t)$ is a weight function. This magnitude can be decomposed into two components, the Mean Directional Outlyingness (MO) and the Variation of Directional Outlyingness (VO). The detection algorithm is based on these quantities and the selection of cutoff values for inferred Mahalanobis distances based on standard boxplots.

4.2.4. Sequential Transformations

This algorithm from [35] relies on the transformation of a wide diversity of shape outliers into magnitude outliers, much easier to detect through standard procedures. Given a sequence of operators defined in \mathcal{F} (the functional space that generates the considered data) $\{\mathcal{G}_k\}, k = 0, 1, 2, \dots$, the method consists of sorting the raw and transformed data into vectors of ranks for each observation. The vectors of ranks are sorted according to a one-side depth notion, such as the *extreme rank depth* for instance, and a global envelope is constructed, which allows the outlier identification.

4.2.5. Results

The results of the application of the algorithm are given for the previously used 4 models and different degrees of contamination. The experiments were simulated 500 times for a sample of curves of $N = 100$, and three different degrees of contamination of outlying curves: 1%, 5% and 10% of outliers in the sample. The detection rates are summarized in Table 2.

First, we must note that the identification capabilities and rates are clearly reduced when the size of the outlying sample is increased. This reduction of the performance of any detection algorithm is logical, since higher degrees of contamination naturally pollute the functional sample, which increases the bias of the score that is used for outlier detection. In the same line, if the size of the outlying sample is considerable (10% of outliers for instance), an argument can be made to defend that this sample might not be outlying, and that it simply corresponds to another mode in a hypothetical multimodal functional sample. This kind of phenomenon, as well as masking effects, are described in detail in [19].

Looking at the results, we can appreciate that the performance of the proposed algorithm is indeed competitive and on par with existent methods, even for complex sets of functional data, such as Model 4. In this case, we can clearly appreciate how the inclusion of a measure specifically dedicated to the detection of shape differences allows the consistent detection of the outlier. This capability is especially significant when we compare it with the other methods, which prove to be unable to detect this kind of shape outlier. In the case of the widely used Functional Boxplots, this is to be expected since they are intended to detect magnitude outliers. Regarding the HDR method, its low detection capabilities

in this case are due to the fact that the low-dimensional representation through robust Functional Principal Component Analysis is not sufficiently precise to capture the outlying nature of the straight line. It is indeed possible that retaining a higher number of modes in this case could allow better detection capabilities, but this procedure greatly increases the curse of dimensionality problem (even if this subject is not treated in the paper by [39]), and it does not allow visualization purposes.

Table 2. Performance of the different algorithms on the test models. The results are expressed as a percentage (detection rates). DO: Directional Detector; FB: Functional Boxplots; HDR: High-Density Regions. The values given in bold correspond to the best results for each model.

N = 100, p = 1%	Model 1	Model 2	Model 3	Model 4
Algorithm	100.00	96.94	100.00	100.00
DO	59.26	39.51	100.00	0.00
FB	2.33	0.00	100.00	0.00
HDR	89.47	69.64	100.00	0.00
N = 100, p = 5%				
Algorithm	91.14	96.79	99.17	97.50
DO	58.23	54.40	100.00	0.00
FB	2.53	4.18	11.95	0.00
HDR	48.35	44.8	49.48	0.00
N = 100, p = 10%				
Algorithm	81.50	75.49	86.67	92.37
DO	47.25	45.97	99.63	0.00
FB	0.75	1.71	7.41	0.00
HDR	22.25	23.41	14.07	0.00

It is clear that Model 3 (being the only pure magnitude outlier among the considered models) is the most simple and easy to detect and virtually any method can consistently detect this kind of outlier when the sample is not overly polluted. Methods which rely the most on the density of curves in the functional space and their trends is more vulnerable to the bias induced in the sample by the curves, as they tend to identify the proportion of curves that *behave unusually* as belonging to a different mode of curves instead of genuine outliers. In the case of the functional boxplots, this is to be expected since by construction they are dedicated to the detection of magnitude outliers, which is useful if the contamination of the sample is made by a wide variety of magnitude outliers, but not so much if those outliers have all been generated by a homogeneous family of curves. In the case of the HDR plots, the existence of a homogeneous sample of outliers generates a set of points in their two-dimensional feature space of principal component scores with a high density of data.

In Models 1 and 2 the conclusions are similar (both models present a combination of slight magnitude and shape outliers). Most methods do not showcase any robustness for such slight magnitude outliers, contrary to the presented algorithm. The main conclusion that can be extracted from these tests is that most methods struggle to find outliers when they are not apparent, as is the case of the models presented here.

Finally, it must be mentioned that the Directional Detector is the most robust method when it comes to detecting the pure magnitude outlier presented in Model 3, as is the least sensitive method to more contaminated samples. The main advantage of this methodology is its capability of finding outliers in multivariate functional data sets.

4.3. Ranking Results

Finally, another advantage of the methodology presented in this paper is the ability to provide a scalar ranking criterion ($\theta_i \in [0, 1], \forall i \in \{1, \dots, n\}$) for a sample of functional data. This is not only useful from a clustering or outlying detection perspective, but also from an exploratory analysis one. In particular, this kind of score can be used to perform sensitivity analysis on the functional data (see an example in Section 5).

Depth measures are widely used in this setting, but some advanced techniques that have developed in recent years, such as the method presented in Section 4.2.4, which provides an efficient ordering of the data. Naturally, as it happened in the outlier detection setting, a good order measure should be capable of identifying a potential multimodality in the set of data, as well as handling the existence of magnitude, shape, or mixed outliers.

The ranking experiments are the same ones as the ones performed for outlier detection testing, with one outlier in the sample of 100 curves. The results are presented in Table 3.

Table 3. Average ranking of the outlier curve across the 100 replications of the experiments for the selected models. The Sequential transformations procedure is presented in [35]. The Modified Band Depth is presented in [36], and the standard Integrated Depth appears in [41]. In this case, the closer the value of a method is to 100, the more outlying it will be according to the corresponding ranking measure.

N = 100, p = 1%	Model 1	Model 2	Model 3	Model 4
Algorithm	98.36	97.80	99.57	93.06
Sequential Transformations	98.14	97.34	99.97	99.88
Modified Band Depth	84.00	61.39	98.49	1
Integrated Depth	83.15	59.81	98.42	1

In this case, we can appreciate that more advanced ranking methods such as the one presented here and the one presented in [35] provide a consistent ordering of the functional data. The results provided by these up-to-date methods show that for the sample of 100 curves, the introduced outlier is always found to belong to the 5% more outlying curves of the sample, and is frequently found to be the most outlying in simpler cases such as Model 3. Usual ranking techniques for functional data such as depth definitions fail to clearly identify the more outlying nature of the outlier in the sample, and cannot be reliably used as order measures in such homogeneous packages of univariate curves.

Finally, we can mention that the better ranking results provided by the Sequential Transformations algorithm for Model 4 with respect to the method presented in this paper can be explained due to the nature of the chosen transformations. In particular, the use of what they call \mathcal{D}_1 transformation in their paper, which consists of taking the first-order derivative of the functional data (in their notation, $\mathcal{D}_1[X(t)] = \frac{d^{(1)}X(t)}{dt}$), is obviously appropriate in the case of Model 4, where the pure shape outlier differs mainly by the values of derivatives. This means that its superior ranking capabilities for this specific model cannot be generalized for any type of shape outlier, and both methods provide comparable ranking results.

5. Industrial Test-Case Study

5.1. Presentation

In this section, the outlier detection methodology is applied to a real industrial dataset of time-dependent numerical simulations. We consider an Intermediate Break Loss of Coolant Accident (commonly called *IBLOCA* or simply *LOCA*) in a nuclear power plant, simulated with the CATHARE2 code. CATHARE2 is a *best estimate* computer code capable of recreating the main physical phenomena that may occur in the different systems involved in nuclear reactors, in particular in the 900 MW French Pressurized Water Reactors. It embeds two-phase modeling to calculate the thermal-hydraulic behavior of the coolant fluid in the reactor.

A LOCA accident is originated by a breach in the primary circuit, which is designed to evacuate the heat generated by the nuclear core. The sudden loss of large quantities of coolant implies a fast increase of the water temperature nearby the nuclear fuel rods, due to the residual power generated by the core during the accidental transient. This power discharge and the subsequent temperature elevation must be compensated by the injection of water through a dedicated safety system. This is supposed to ensure that fuel rods temperature would remain below the fusion point at all times. Hence, the main safety criterion in LOCA regarding the confinement of the fuel concerns its Peak Cladding Temperature (PCT), i.e., its maximum cladding temperature over the duration of the LOCA (here, the time-dependent cladding temperature is the maximum cladding temperature for all the fuel rods, whatever its localization in the core).

The particular statistical model under study involves many scalar input parameters ($p = 97$), which specify all sorts of physical phenomena whose relative influences in the PCT are difficult to assess a priori. These input variables can be classified into various categories, such as (i) initial conditions and limit states for the system (Primary pressure, starting thermal power, the primary pumps inertia...), (ii) some parameters of specific physical models and correlations that are used (thermal exchanges between the components, friction between fluid phases or some geometric parameters of the installation), as well as (iii) some scenario variables (existence of blockages in the heat exchangers, fuel use in its life-cycle or initial temperature of the safety water injection).

All of these scalar inputs of the simulation code are uncertain, and are hence represented by the random vector $X = (X_1, X_2, \dots, X_{97})$, whereas the output of interest would be Y , which is normally a critical safety parameter in this case, such as the aforementioned PCT.

The total number of simulations that can be performed is relatively limited for such a high-dimensional input vector, since each run of the code takes around one hour to finish. For this reason, the use of classical multivariate statistical techniques is not straightforward. This explains the widely spread use of space-filling design methods to maximize the coverage of the input space [42], as well as metamodeling techniques to better exploit the number of code runs available for the physical model. Briefly explained, space-filling designs try to optimally explore the space of input variables i.e., they establish criteria to better choose the analyzed points of this high-dimensional space (in the case of nuclear transients, there can easily exist more than a hundred input variables). In the case of metamodels, they are mathematical approximations of more complex physical models that, despite showing a higher precision in their calculations, take a much higher computation time. The use of metamodels helps to improve (increase) the total number of available simulations, so that the results that are finally obtained are more statistically relevant.

In this context, the consideration of the whole functional output (the evolution of the maximum cladding temperature, whatever its location) is expected to provide a better insight on the physical phenomena that govern in the transient than the scalar value of the PCT alone. In our case, 1000 Monte Carlo runs of the code were launched, generating the set of curves that is presented in Figure 2 for the evolution of the maximum cladding temperature during the transient.

5.2. Functional Outlier Detection

The previously presented outlier detection technique is applied to this set of curves. Both the h-mode depth and the DTW are the selected features to obtain the degree of outlyingness of each functional datum. The curves presenting a degree $\hat{\theta}$ of outlyingness over 95% are shown in Figure 3.

The first apparent result is that the main magnitude outlier is easily detected, since the curve that acts as the upper envelope of temperature in most points of the domain is the one presenting the highest value of θ ($\theta = 1.0$ actually in this case). This curve is not only anomalous in the magnitude sense, but also in the shape one, similar to Model 4 that was presented in the previous section (these are sometimes called *phase outliers*).

Two other magnitude outliers have been identified, and one of them is also a shape outlier, presenting an anomalous peak of temperature after about 120 s of simulation (physical time). The final main outlier is a pure shape one, remaining zones with high density of data during the whole domain of simulation, but presenting two peaks of temperature. Especially notorious is the first peak, which occurs around the 100 s of transient, in a time interval that does not match most curves.

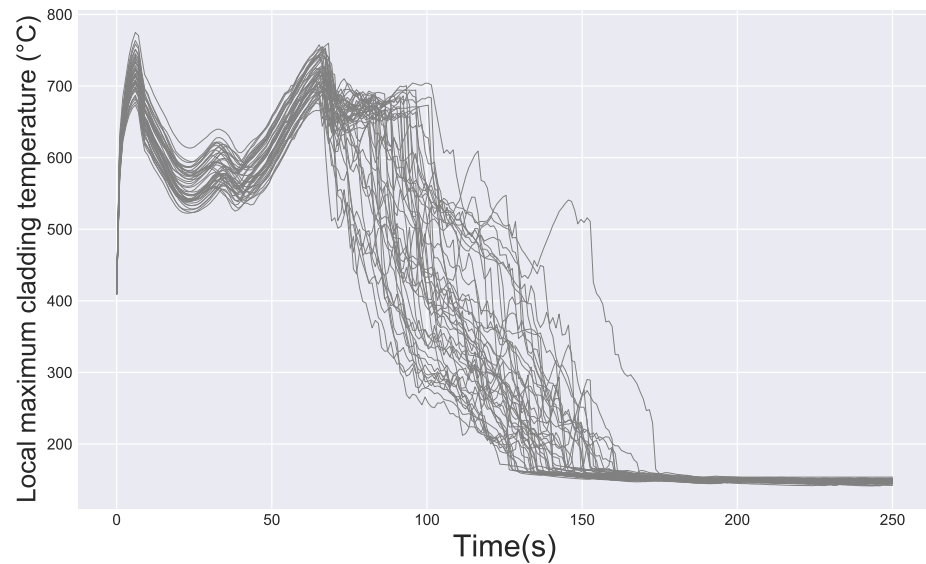


Figure 2. Examples of IBLOCA transients simulated with CATHARE2. Only 40 curves are displayed for clarity.

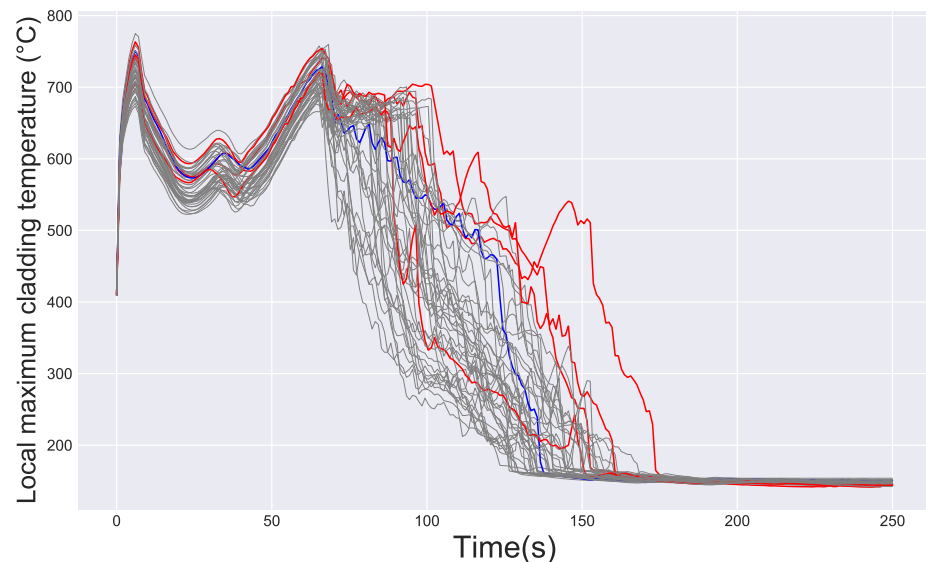


Figure 3. IBLOCA transient curves presenting the highest degree of outlyingness (red) and the least outlying curve (blue).

5.3. Sensitivity Analysis on Outliers

In this kind of numerical simulations, the detection of outputs that present a globally anomalous behavior is of critical importance, and characterizing what are the physical phenomena which have an actual influence on it can have the same importance, if not more. A way of performing this analysis is to establish some kind of dependence measure between the inputs of the simulation code and the outlying score θ . However, the high dimensionality of the problem and the possible correlations between the input variables

of the code can make this a difficult task. The Hilbert Schmidt Independence Criterion (HSIC) [43] can be a useful tool in this context to test the dependence between the scalar input variables of the code and the outlying score. This is a first step to understand which physical variables are the ones that actually influence the anomalous behavior of the outputs.

By performing statistical tests on the HSIC values of the couples (X_i, Y) in the design of experiments it is possible to quantify their dependence. Without going into the technical details of the procedure (see [44]), the HSIC represents a dependence measure between both variables, and can be used to build a statistical test with null hypothesis: \mathcal{H}_0 : the variable X_i and Y are independent. The hypothesis is rejected if the associated p -value of the test is inferior to a significance level threshold α . If \mathcal{H}_0 is rejected, then the existence of a dependence structure between the input variable X_i and the output exists.

In this case, we apply this measure to perform a Target Sensitivity Analysis, i.e., sensitivity analysis but quantifying the influence of the considered parameters in a restricted domain \mathcal{M} of the possible output values ($\mathcal{M} \subset \Theta$ such that $\mathcal{M} = \{\theta \in \Theta | \theta > 0.9\}$). This application to the set of input data and the obtained values of θ yields several influential variables that are shown in Table 4.

Table 4. Detected influential variables for $\theta \in \mathcal{M}$. The variables are not the actual values of the physical parameters, but **multiplicative coefficients** that increase or decrease their importance in a scenario.

Variable	Description
X_{16}	Friction between the water and the discharge line of the accumulators
X_{38}	Global heat transfer coefficient in reflood fuel/coolant
X_{45}	Pressure drop to model the constrained flow due to the deformation of the fuel
X_{62}	Friction coefficient between steam and water in the downcomer during the reflood phase
X_{64}	Friction coefficient between water and steam in core during the reflood phase
X_{68}	Heat transfer coefficient between steam and water in the downcomer

All these input variables can be considered to have an influence on the outlyingness of the resulting simulated curve, which represent the evolution of the maximum cladding temperature during the transient. Variable X_{16} represents the friction between the injected water in the primary circuit by the accumulators and the injection line. This parameter has been found to be influential in other similar studies since the compensation of the lost water during the transient is mainly guaranteed by this system (the accumulators), and therefore the line connecting it to the primary is of crucial importance. If this value of friction increases, the water flow will be reduced, with the consequent increase in the average temperatures of the fuel.

Regarding variables 38, 45 and 64, they are representative of physical phenomena occurring in the reactor pressure vessel during the transient. They model respectively: the heat transfer coefficient between the nuclear fuel and the surrounding coolant; the increase in pressure drop due to the deformation of the nuclear fuel due to the thermal-mechanical stress, prevents the coolant from ascending easily to the top of the reactor pressure vessel; and the friction between the steam and water in the core during the reflood phase of the transient. These elements are relevant since their evolution greatly influences the rewetting dynamic of the fuel, and the heat extraction in the short-term phase of the transient.

Finally, variables X_{62} and X_{68} are representative of phenomena which occur between the steam and coolant water in the downcomer (the annular part which links the injection line of the accumulators and the nuclear core). This element is critical during the reflood process, which is why increases in the friction between the ascending steam and the descending water in this element are penalizing from a safety point of view. This is due to the fact that if the friction coefficient increases, it means that the momentum of the injected water will be reduced, and the core will take longer to be filled. Similar conclusions can be

obtained for the heat exchange coefficient between both phases in the downcomer, since low values for this variable will imply lower heat extraction rates in the vessel.

The study of two-dimensional scatter plots between these input parameters and the outlying score θ could already prove to be useful to visualize how the inputs affect the outlying nature of the functional outputs. An example that illustrates the idea is presented in Figure 4.

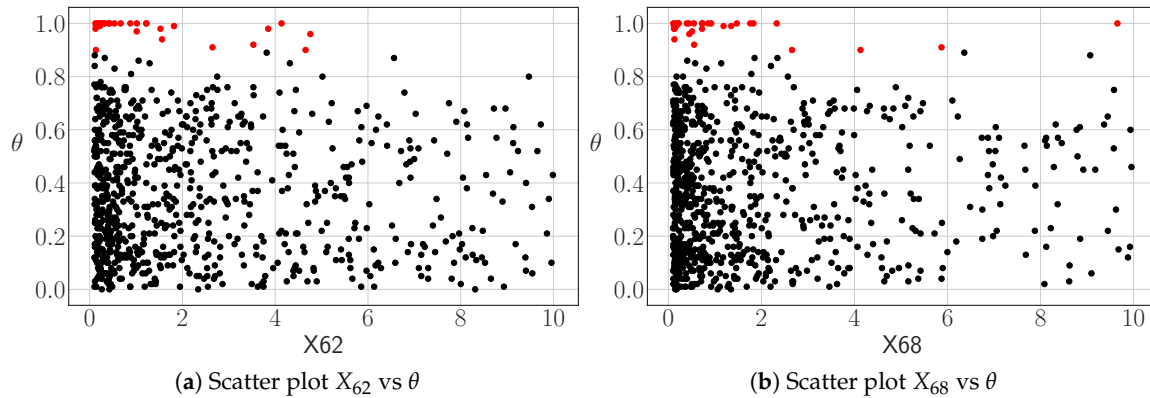


Figure 4. Scatter plots of two input variables of CATHARE2 and the outlyingness score θ . The points correspond to the bivariate plot of the values of the selected variable and the corresponding θ . The red dots correspond to the simulations that have been retained as outliers.

As it can be seen, outlying transient concentrate around specific subsets of the domain of the identified influential variables. These plots are useful to evaluate if the physical values that originate outlying simulations are physically coherent with their expected influence. In this particular case, for instance, lower values of friction should correspond to less penalizing and outlying configurations when the safety criterion is the Peak cladding temperature of the nuclear fuel. Therefore, the observed effect of this variable is actually not expected, and it corresponds to an anomalous effect in the coding of this particular transient that was later corrected by the engineers. In other words, the methodology and the analysis technique was capable of capturing not only extreme effects in the analyzed physical time-dependent variable (the Maximum Local Cladding temperature), but it was also capable of finding actual outliers, in the sense that those simulations showcase non-physical events in the particular modeling used in this study.

6. Discussion and Conclusions

This paper has dealt with a particular branch of unsupervised anomaly detection: the outlier detection problem in functional data. The main aspects to take into account when dealing with functional data or high-dimensional objects in general have also been developed, exposing its main challenges and advantages. A new time-dependent outlier detection methodology based on the use of non-parametric features has been proposed, assessed with synthetic data, and illustrated on thermal-hydraulic simulations, aiming at capturing the outlyingness of the elements of any considered functional dataset both in the magnitude and the shape senses. This is done via the joint use of the h-mode depth and the dynamic time warping, and by defining outliers as data which do not belong to the minimum volume set of a chosen probability. The underlying distribution in the feature space is modeled through a GMM whose estimation is adapted to the outlier detection framework by means of modifications of the EM algorithm. The maximal probability for which a datum is not regarded as an outlier anymore is used as a score θ of its outlyingness.

An original detection algorithm has been proposed, effectively allowing the trimming of functional data. This methodology, based on the use of two features, benefits from the notion of level set to treat real industrial problems based on time-dependent data even if the available data are scarce. Several features have been compared in this framework based

on some toy examples and two scores related to the outlyingness of functional data. Based on the results of these application cases, both the notions of dynamic time warping, and the h-mode depth have proven their efficiency when compared to other features such as the L^2 norm and the Modified Band Depth.

Finally, the analysis of simulations of the thermal-hydraulic behavior of a nuclear reactor during a Loss Of Coolant Accident has been carried out to illustrate the benefits of the method. This was achieved thanks to the use of sensitivity analysis tools (HSIC-based) capable of accounting for the dependence of the input variables of the numerical simulator.

Regarding the perspectives of this work, a primary objective would be the in-depth quantification of the causes for the detected anomalous characteristics of certain functions in real physical cases. In the case of numerical simulators, the identification of the inputs of the code that actually present an influence on the anomalous outputs can help engineers to detect possible defects of the code or finding physical phenomena of interest. This is also relevant to ensure the quality of the datasets that are used in the assessment reports of critical systems such as nuclear power plants.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/app112311475/s1>, file S1: Functional outlier detection by means of h-mode depth and dynamic time warping.

Author Contributions: Conceptualization, Á.R.d.P.; Methodology, Á.R.d.P.; Supervision, M.C., B.I., N.M., A.M., E.M. and R.S.; Writing—original draft, Á.R.d.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Research data are not shared due to confidentiality reasons.

Acknowledgments: Part of this work has been done in connection with the ACIDITHY project (French research program NEEDS, including CNRS, CEA, IRSN and EDF). We acknowledge two anonymous reviewers for their helpful remarks. Many thanks to Vincent Larget for his help with the industrial use case.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DTW	Dynamic Time Warping
i.i.d.	independent and identically distributed
GMM	Gaussian Mixture Models
HDR	High-Density Regions
LOCA	Loss of Coolant Accident
PCT	Peak Cladding Temperature
HSIC	Hilbert Schmidt Independence Criterion

References

1. Chamroukhi, F.; Nguyen, H.D. Model-based clustering and classification of functional data. *WIREs Data Min. Knowl. Discov.* **2019**, *9*, 1298. [[CrossRef](#)]
2. Slaets, L.; Claeskens, G.; Hubert, M. Phase and Amplitude-Based Clustering for Functional Data. *Comput. Stat. Data Anal.* **2012**, *56*, 2360–2374. [[CrossRef](#)]
3. Ieva, F.; Paganoni, A.; Pigoli, D.; Vitelli, V. ECG signal reconstruction, Landmark registration and functional classification. In Proceedings of the SCO 2011, Sharable Content Objects, 7th Conference about Electronic Support of Learning, Brno, Czech Republic, 22–23 June 2011.
4. Grenander, U. Stochastic processes and statistical inference. *Ark. Mat.* **1950**, *1*, 195–277. [[CrossRef](#)]
5. Ramsay, J.O. When the data are functions. *Psychometrika* **1982**, *47*, 379–396. [[CrossRef](#)]
6. Ramsay, J.; Silverman, B. *Functional Data Analysis*; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 1997.

7. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer: Berlin/Heidelberg, Germany, 2006; Volume 51.
8. Besse, P.; Cardot, H.; Faivre, R.; Goulard, M. Statistical modelling of functional data. *Appl. Stoch. Model. Bus. Ind.* **2005**, *21*, 165–173. [[CrossRef](#)]
9. Febrero-Bande, M.; Galeano, P.; González-Manteiga, W. Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics* **2008**, *19*, 331–345. [[CrossRef](#)]
10. Juang, W.C.; Huang, S.J.; Huang, F.D.; Cheng, P.W.; Wann, S.R. Application of time series analysis in modelling and forecasting emergency department visits in a medical centre in Southern Taiwan. *BMJ Open* **2017**, *7*, e018628. [[CrossRef](#)]
11. Sen, R.; Klüppelberg, C. Time series of functional data with application to yield curves. *Appl. Stoch. Model. Bus. Ind.* **2019**, *35*, 1028–1043. [[CrossRef](#)]
12. Santner, T.J.; Williams, B.J.; Notz, W. *The Design and Analysis of Computer Experiments*, 1st ed.; Springer Series in Statistics; Springer: Berlin/Heidelberg, Germany, 2003.
13. Roustant, O.; Joucla, J.; Probst, P. Kriging as an alternative for a more precise analysis of output parameters in nuclear safety—Large break LOCA calculation. *Appl. Stoch. Model. Bus. Ind.* **2010**, *26*, 565–576. [[CrossRef](#)]
14. IAEA. *Accident Analysis for Nuclear Power Plants with Pressurized Water Reactors*; Number 30 in Safety Reports Series; International Atomic Energy Agency: Vienna, Austria, 2003.
15. Geffraye, G.; Antoni, O.; Farvacque, M.; Kadri, D.; Laviolle, G.; Rameau, B.; Ruby, A. CATHARE 2 V2.5_2: A single version for various applications. *Nucl. Eng. Des.* **2011**, *241*, 4456–4463. [[CrossRef](#)]
16. Nanty, S. *Stochastic Methods for Uncertainty Treatment of Functional Variables in Computer Codes : Application to Safety Studies*. Ph.D. Thesis, Université Grenoble Alpes, Saint-Martin-d’Hères, France, 2015.
17. Auder, B. *Classification and Modelling of Computer Codes Functional Outputs: Application to Accidental Thermo-Hydraulic Computations in Pressurized Water Reactors (PWR)*. Ph.D. Thesis, Université Paris 6: Paris, France, 2011.
18. James, G.M.; Hastie, T.J.; Sugar, C.A. Principal Component Models for Sparse Functional Data. *Biometrika* **2000**, *87*, 587–602. [[CrossRef](#)]
19. Aggarwal, C. *Outlier Analysis*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2017.
20. Schmutz, A.; Jacques, J.; Bouveyron, C.; Chèze, L.; Martin, P. Clustering multivariate functional data in group-specific functional subspaces. *Comput. Stat.* **2020**, *35*, 1101–1131. [[CrossRef](#)]
21. López-Pintado, S.; Sun, Y.; Lin, J.K.; Genton, M.G. Simplicial band depth for multivariate functional data. *Adv. Data Anal. Classif.* **2014**, *8*, 321–338. [[CrossRef](#)]
22. Nagy, S.; Gijbels, I.; Hlubinka, D. Depth-Based Recognition of Shape Outlying Functions. *J. Comput. Graph. Stat.* **2017**, *26*, 883–893. [[CrossRef](#)]
23. Martos, G.; Hernández, N.; Muñoz, A.; Moguerza, J. Entropy measures for stochastic processes with applications in functional anomaly detection. *Entropy* **2018**, *20*, 33. [[CrossRef](#)]
24. Barreyre, C.; Laurent, B.; Loubes, J.M.; Boussouf, L.; Cabon, B. Multiple Testing for Outlier Detection in Space Telemetries. *IEEE Trans. Big Data* **2020**, *6*, 443–451. [[CrossRef](#)]
25. Arribas-Gil, A.; Romo, J. Shape outlier detection and visualization for functional data: The outliergram. *Biostatistics* **2014**, *15*, 603–619. [[CrossRef](#)]
26. Sguera, C.; Galeano, P.; Lillo, R. Functional outlier detection by a local depth with applications to NOx levels. *Stoch. Environ. Res. Risk Assess.* **2016**, *30*, 1115–1130. [[CrossRef](#)]
27. Moon, T. The expectation-maximization algorithm. *IEEE Signal Process. Mag.* **1996**, *13*, 47–60. [[CrossRef](#)]
28. Bishop, C.M. *Pattern Recognit. Mach. Learn.*; Springer: Berlin/Heidelberg, Germany, 2006.
29. Ahidar-Coutrix, A. *Surfaces quantile: Propriétés, convergences et applications*. Ph.D. Thesis, Université de Toulouse, Toulouse, France, 2015.
30. Hosseini, R.; Sra, S. Matrix Manifold Optimization for Gaussian Mixtures. In *Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2015; Volume 28.
31. Genz, A. Numerical Computation Of Multivariate Normal Probabilities. *J. Comput. Graph. Stat.* **2000**, *1*, 141–149.
32. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*; Macmillan: New York, NY, USA, 1994.
33. Schwarz, G. Estimating the Dimension of a Model. *Ann. Stat.* **1978**, *6*, 461–464. [[CrossRef](#)]
34. Biernacki, C.; Celeux, G.; Govaert, G. Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood. *Pattern Anal. Mach. Intell. IEEE Trans.* **2000**, *22*, 719–725. [[CrossRef](#)]
35. Dai, W.; Mrkvicka, T.; Sun, Y.; Genton, M. Functional outlier detection and taxonomy by sequential transformations. *Comput. Stat. Data Anal.* **2020**, *149*. [[CrossRef](#)]
36. López-Pintado, S.; Romo, J. On the Concept of Depth for Functional Data. *J. Am. Stat. Assoc.* **2009**, *104*, 718–734. [[CrossRef](#)]
37. Long, J.P.; Huang, J.Z. A Study of Functional Depths. *arXiv* **2015**, arXiv:1506.01332.
38. Sun, Y.; Genton, M.G. Functional Boxplots. *J. Comput. Graph. Stat.* **2011**, *20*, 316–334. [[CrossRef](#)]
39. Hyndman, R.J. Rainbow plots, bagplots and boxplots for functional data. *J. Comput. Graph. Stat.* **2009**, *19*, 29–45. [[CrossRef](#)]
40. Dai, W.; Genton, M. Directional outlyingness for multivariate functional data. *Comput. Stat. Data Anal.* **2019**, *131*, 50–66. [[CrossRef](#)]

41. Cuevas, A.; Fraiman, R. On depth measures and dual statistics. A methodology for dealing with general data. *J. Multivar. Anal.* **2009**, *100*, 753–766. [[CrossRef](#)]
42. Iooss, B.; Marrel, A. Advanced Methodology for Uncertainty Propagation in Computer Experiments with Large Number of Inputs. *Nucl. Technol.* **2019**, *205*, 1588–1606. [[CrossRef](#)]
43. Da Veiga, S. Global Sensitivity Analysis with Dependence Measures. *J. Stat. Comput. Simul.* **2015**, *85*, 1283–1305. [[CrossRef](#)]
44. De Lozzo, M.; Marrel, A. New improvements in the use of dependence measures for sensitivity analysis and screening. *J. Stat. Comput. Simul.* **2016**, *86*, 3038–3058. [[CrossRef](#)]