



HAL
open science

Functional outlier detection by means of h-mode depth and dynamic time warping

Alvaro Rollon de Pinedo, Mathieu Couplet, Bertrand Iooss, Nathalie Marie,
Amandine Marrel, Elsa Merle, Roman Sueur

► **To cite this version:**

Alvaro Rollon de Pinedo, Mathieu Couplet, Bertrand Iooss, Nathalie Marie, Amandine Marrel, et al..
Functional outlier detection by means of h-mode depth and dynamic time warping. Applied Sciences,
2021, 11 (23), pp.11475. 10.3390/app112311475 . hal-02965504v2

HAL Id: hal-02965504

<https://hal.science/hal-02965504v2>

Submitted on 24 May 2021

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

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

FUNCTIONAL OUTLIER DETECTION BY MEANS OF H-MODE DEPTH AND DYNAMIC TIME WARPING

Álvaro Rollón de Pinedo^{1,2}, Mathieu Couplet¹, Bertrand Iooss¹, Nathalie Marie³, Amandine Marrel³, Elsa Merle⁴ and Roman Sueur¹

¹ EDF R&D, 6 quai Watier, Chatou, 78400, France

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

³ CEA, DES, IRESNE, F-13108, Saint-Paul-lez-Durance, France

⁴ LPSC Grenoble, 53 avenue des Martyrs, Grenoble, 38000, France

contact email - alvaro.rollon-de-pinedo@edf.fr

Abstract

This paper deals with the problem of finding outliers, i.e. data that differ distinctly from other elements of the considered dataset, when they belong to functional infinite-dimensional vector spaces. Functional data are widely present in the industry and may originate from physical measurements or numerical simulations. The automatic identification of outliers can help to ensure the quality of a dataset (trimming), to validate the results of industrial simulation codes, or to detect specific phenomena or anomalies. This paper focuses on data originated from expensive simulation codes, such as nuclear thermal-hydraulic simulators, in order 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, e.g. the h-mode depth or the Dynamic Time Warping, is proposed in order to evaluate the outlyingness both in the magnitude and shape senses. Theoretical examples are also used in order 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 methodology 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.

28 **Keywords:** Functional Data Analysis, Functional Outlier Detection, Probabilistic Model-
29 ing, Computer experiments, Sensitivity Analysis

30 1 Introduction

31 Nowadays, the ever increasing capabilities of measurement, generation and storing of data
32 have increased the interest of the scientific and industrial communities on what are known
33 as functional data. In its most general form, the domain of functional data analysis deals
34 with any object having the form of a function, regardless of its dimension. As an example,
35 the most basic form these objects can adopt are one-dimensional real functions, which might
36 represent the evolution of a physical parameter of interest over time. These data are normally
37 generated through an actual empirical measure, or a simulation code.

38 The great interest of the functional data analysis is that it allows to take into account the
39 intrinsic nature of the data, i.e. the underlying process that generates them is supposed
40 to have certain restrictions of domain, regularity, continuity and so on. As the functional
41 data are infinite-dimensional by nature, functional data analysis methods always rely on
42 a dimension reduction technique, whether implicitly or explicitly. More precisely, it is the
43 case in the context of classification (Chamroukhi and Nguyen, 2019), clustering (Slaets et al.,
44 2012), landmark research and registration (Ieva et al., 2011) of functional data. By providing
45 more synthetic descriptors of functional observations, functional data analysis methods allow
46 a more practical treatment of data thanks to the available multivariate tools.

47 The domain of functional data analysis can be traced back to the works of Grenander,
48 (Grenander, 1950), but the term itself was coined by Ramsay, (Ramsay, 1982). Some of
49 the most widely regarded references for the domain include the general works of Ramsay
50 and Silverman (1997) or Ferraty and Vieu (2006), focused on non-parametric methods for
51 functional data. Functional data can be found in a wide variety of contexts, among which
52 we can mention: environmental sciences (Besse et al., 2005; Febrero-Bande et al., 2008),

53 medical sciences (Juang et al., 2017), economy (Sen and Klüppelberg, 2019) and others. In
54 the context of this paper, our work is motivated by the use of expensive simulation codes
55 (Santner et al., 2003; Roustant et al., 2010) that are commonly used in the nuclear industry
56 in order to complement the nuclear safety assessment reports (IAEA, 2003), and which
57 require specific statistical tools. These thermal-hydraulic simulators, such as the system
58 code *CATHARE2* (Geffraye et al., 2011), are capable of simulating the evolution of the
59 essential physical parameters during a nuclear transient, and they are expensive to evaluate
60 (several hours per run) and analyze due to the complexity of the simulated phenomena
61 (highly non linear interactions between its parameters).

62 In general, in the context of nuclear safety, the main analyzed parameters are the scalar
63 values known as *Safety Parameters*, which are representative of the severity of an accidental
64 transient. However, the analysis of the whole evolution of the aforementioned physical pa-
65 rameters is a complex domain that has been subject to considerable research effort in recent
66 years, such as in the works of Nanty (2015) or Auder (2011). The analysis of outlying tran-
67 sients in these sets of simulations is useful in order to detect unexpected physical phenomena,
68 validate the simulations providing useful descriptors of the functional outputs of interest, or
69 quantify the extreme behavior or penalizing nature of specific subsets of those transients.

70 In spite of the diversity of domains, there are several common points of interests between
71 all of them. As an example, in the most usual case of one-dimensional functions, a typical
72 difficulty that arises in the pre-treatment phase of the data is the subject of the discretization
73 of the grid on which the data are observed. Without going into too much detail on the
74 subject, one dimensional curves are not normally observed in their entirety (since they are
75 intrinsically infinite-dimensional). Instead, they are indexed by some variable, normally the
76 time in the one-dimensional case, in a grid that can be homogeneous or not. Indeed, points
77 are not necessary equally spaced, as with biological measurements such as blood pressures,
78 or as with temporal data originated by a simulation code whose time step evolves accordingly
79 with the convergence impositions of the numerical solver. Naturally, the representation of the

80 underlying functional process is still a matter of discussion, and does not only depend on the
81 spacing of each point in the time grid, but also on other notions such as the *density* of data.
82 This quantity can differ by several orders of magnitude in the case of numerical simulators
83 depending on the restrictions imposed to time steps in order to guarantee the convergence,
84 as well as the presence or not of random noise etc. For a more profound analysis on the
85 importance of these subjects the reader can refer to James et al. (2000).

86 In this paper we focus on a functional outlier detection technique, showing its sensitivity
87 to both magnitude and shape outliers, and its ranking capabilities. The importance of this
88 work is closely related to the data quality domain, since the presence of anomalous behaviors
89 that might have been originated by measurement errors, non-convergence of algorithms or
90 the existence of non-physical values of numerical simulators may produce spurious results
91 when treating a dataset. As explained in Aggarwal (2017), when the underlying process
92 that creates the data *behaves unusually*, it results in the creation of outliers, and therefore
93 it may contain useful information about *abnormal characteristics of the system*. Finally, the
94 particularities of functional data treatment with respect to multivariate ones require specific
95 tools in order to have acceptable detection capabilities (Schmutz et al., 2020).

96 The functional outlier detection domain has gained relevance in recent years, and the methods
97 for detecting these outliers do not cease to improve. However, the techniques can differ
98 greatly between them. A large number of them rely on the notion of statistical depth
99 (Febrero-Bande et al., 2008; López-Pintado et al., 2014; Nagy et al., 2017), whereas others
100 rely on other dimensionality reduction methods, clustering or hypothesis testing on the
101 coefficients of some functional decomposition (Barreyre et al., 2020), as well as graphical
102 tools (Arribas-Gil and Romo, 2014). Naturally, all of these techniques showcase different
103 detection capabilities, and they can be more adapted to the detection of a particular type
104 of outliers.

105 In our study, we consider that the data correspond to independent and identically distributed
106 (i.i.d.) realizations of a random process Z , taking its values in a functional space \mathcal{F} . In

107 practice, any realization z_i of the random variable Z can only be observed in a limited
 108 number of points in the domain, i.e., it is observed as the random vector $(z_i(t_1), \dots, z_i(t_p))$,
 109 with $\{t_1, \dots, t_p\} \in \mathcal{T} \subset \mathbb{R}$. This data representation is not itself appropriate to the detection
 110 of outlying functions. Section 2 thus presents how to define and compute a modeling of data
 111 dedicated to the characterization of outliers. Based on this data representation, Section
 112 3 presents our functional outlier detection methodology. Sections 4 and 5 provide some
 113 analytical and industrial application examples. The properties, capabilities and limitations
 114 of the methodology are discussed in section 6.

115 2 Functional data representation and modeling

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

125 2.1 *Discussion and selected approach*

126 Functional outliers are commonly classified into magnitude and shape outliers. Magnitude
 127 outliers may be defined by functions that deviate from the bulk of curves at some point in
 128 the definition domain of the functions according to some distance metric defined in their
 129 functional space (Dai et al., 2020). There exist numerous methods to visualize and detect
 130 them, some of them being based on depth measures, as exposed by Sun and Genton (2011).

131 The other main kind of outliers are shape outliers. This type of functional outlier is sig-
 132 nificantly more difficult to detect, but several techniques able to deal with them have been
 133 developed in recent years. We can mention Slaets et al. (2012) or Arribas-Gil and Romo
 134 (2014). Some of the challenges in the use of depth measures in this case are exposed in Nagy
 135 et al. (2017). Most functional outlier detection methods can be classified into one of the
 136 three following categories (Jacques and Preda, 2014):

- 137 • Two-stage approaches: the functional data are firstly projected into the considered
 138 functional space, in what is usually called the *filtering* step, and then a classical mul-
 139 tivariate clustering procedure is applied on the coefficients of the expansion. In this
 140 case, if $\Phi = \{\phi_1, \dots, \phi_r\}$ is the set of functions that forms a complete orthonormal
 141 basis of \mathcal{F} , any function z_i of the space can be reconstructed from the sampled data
 142 through an expansion of the type $z_i = \sum_{j=1}^r a_j \phi_j$. In practice, we work with a finite
 143 number family of functions (a subset of \mathcal{F}), which induces a representation error, and
 144 that is commonly obtained by truncating an actual basis of \mathcal{F} . This allows to perform
 145 statistical hypothesis tests on the coefficients, which provides a detection criterion. An
 146 example of this approach can be found in (Barreyre et al., 2020).
- 147 • Non-parametric approaches: they are based on measures of proximity and dissimilarity
 148 between the functions. Multivariate clustering algorithms can usually be applied on
 149 these features (Dai et al., 2020).
- 150 • Probabilistic model-based approaches: they rely on the estimation of an underlying
 151 probability model, either on some non-parametric features applied to the curves, or
 152 on the coefficients of a basis expansion. An example of this approach applied to the
 153 coefficients of a functional Principal Components basis expansion can be found in
 154 the Ph.D. works of Nanty (2015), where the coefficients of the expansion are used to
 155 perform sensitivity analysis.

156 In our work, the detection procedure is based on the use of non-parametric measures and

157 the estimation of probabilistic models in order to reconstruct the joint probability density
 158 function of those features.

159 **2.2 Features definition of outlying functions**

160 Depth measures (Mozharovskyi, 2016) are a set of non-parametric features that have gained
 161 relevance in the functional outlier detection field in recent years (Cuevas and Fraiman, 2009).
 162 Generally speaking, let z_1, \dots, z_n be a set of objects observed in \mathbb{R}^p such that a random element
 163 Z describing the population is fixed, then a depth function is a mapping $D(\cdot, Z) : \mathbb{R}^p \rightarrow \mathbb{R}^+$
 164 which provides a center-outward ordering of the data. The same definition holds for the
 165 case where $p \rightarrow \infty$ for the functional framework. This functions are widely used for central
 166 tendency estimation, outlier detection and classification.
 167 Some of the most widely used definitions of depth measures in the functional framework are
 168 developed below.

- 169 • Band depths. Let z_1, \dots, z_n be a sample of functional data, then the basic definition of
 170 the Band Depth of a specific function z_i takes the form (López-Pintado et al., 2014):

$$S_{n,J}(z_i, Z) = \sum_{j=2}^J S_n^{(j)}(z_i|Z), J \geq 2 \tag{1}$$

171 such that:

$$S_n^{(j)}(z_i, Z) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq n} \mathbb{1}_{\{G(z_i) \subset B(z_{i_1}, z_{i_2}, \dots, z_{i_j})\}}, j \geq 2 \tag{2}$$

172 with $\mathbb{1}$ the indicator function. In this case, $G(z_i)$ is the graph of the function z_i , i.e.
 173 $G(z_i) = \{(t, z_i(t)) : t \in \mathcal{T}\}$, and B represents the band delimited by the j curves
 174 z_1, \dots, z_j . The parameter J restricts the maximum number of functions that delimit
 175 the bands. López-Pintado and Romo (2009) recommend the use of $J = 3$.

176 • A more flexible definition of this depth notion is the *Modified Band Depth*, which
 177 consists in replacing the indicator function $\mathbb{1}$ by a measure of the subset where the
 178 analyzed function is within the limits of the band. If $A_j(x) \equiv \{t \in \mathcal{T} : \min_{r=i_1, \dots, i_j} z_r(t) \leq$
 179 $z(t) \leq \max_{r=i_1, \dots, i_j} z_r(t)\}$ is the mentioned subset, then the Lebesgue measure λ of the
 180 subset, normalized by the measure of \mathcal{T} provides a measure of *how much time* the
 181 considered function remains within the bands. Taking this into account, the Modified
 182 Band Depth can be expressed as:

$$MBD_n^{(j)}(z_i, Z) = \sum_{j=2}^J \binom{n}{j}^{-1} \sum_{1 \leq i_1 < \dots < i_j \leq n} \lambda_r(A(z_i; z_{i_1}, z_{i_2}, \dots, z_{i_j})), 2 \leq j \leq n. \quad (3)$$

183 Contrary to the basic Band Depth, the MBD is sensitive to functions that deviate from
 184 the center of the functions even if it is only for small subsets within the domain, which
 185 is naturally essential in the outlier detection domain.

186 • Another widely spread definition of depth is the *h-modal* depth, proposed in Cuevas
 187 et al. (2007). It employs the notion of a kernel function in order to estimate the
 188 centrality of the curve by taking into account the degree of immersion of a certain
 189 curve with regard to the curves that lie closest to the analyzed one according to some
 190 distance notion defined in the considered functional space.

191 The *h-mode* depth of a realization $z_i \in \mathcal{F}$ with respect to the distribution of $Z \sim P \in$
 192 $\mathcal{P}(\mathcal{F})$ is defined as:

$$hM(z_i, Z) = \mathbb{E} \left(\frac{1}{h} K \left(\frac{\|z_i - Z\|}{h} \right) \right), \quad (4)$$

193 which can be substituted by its empirical version (with a sample of n functional data):

$$hM(z_i; Z_n) = \sum_{j=1}^K \left(\frac{1}{\hat{h}} K \left(\frac{\|z_i - z_j\|}{\hat{h}} \right) \right). \quad (5)$$

194 In this context, $\|\cdot\|$ is a norm defined on \mathcal{F} , with no *a priori* imposed limitations.
 195 K is a measurable kernel function $K : \mathbb{R} \rightarrow \mathbb{R}^+$ with h as the bandwidth parameter.
 196 The practical implementation of this depth notion consists in substituting the actual
 197 distribution P by its empirical version $P^* \in \mathcal{P}(\mathcal{F})$.

198 This definition strongly depends on the choice of the norm and the bandwidth param-
 199 eter. The authors give some orientations regarding this subject, proposing the L^2 and
 200 L^∞ norms, and taking h as the 15th percentile of the distribution of $\|z_i - z_j\|, \forall z_i, z_j \in$
 201 \mathcal{F} . For some results on the consistency of the *h-modal* depth, the reader can refer to
 202 Gijbels and Nagy (2015).

203 In addition to these depth notions, some other non-parametric features can be mentioned as
 204 they will help us characterizing functional data. The Time Series framework is significantly
 205 related to the functional data analysis domain, and also provides some useful metrics that
 206 can help to quantify the degree of similarity between ordered sequences. This is the case
 207 of the Dynamic Time Warping (DTW) algorithm, whose general form is presented below
 208 (Bellman and Kalaba, 1959).

209 Given two sequences $X := (x_1, x_2, \dots, x_V), V \in \mathbb{N}$ and $Y := (y_1, y_2, \dots, y_W), W \in \mathbb{N}$, as well as
 210 a feature space \mathcal{S} , and $x_v, y_w \in \mathcal{S}$ for $v \in [1 : V]$ and $w \in [1 : W]$, we can define a local cost
 211 measure (sometimes also called local distance measure), which is an application:

$$c : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+. \quad (6)$$

212 In this case, an (V, W) -warping path is a sequence $p = (p_1, \dots, p_L)$ with $p_l = (n_l, m_l) \in [1 :$
 213 $V] \times [1 : W], \forall l \in [1 : L]$ which also satisfies the following conditions:

- 214 • boundary condition: $p_1 = (1, 1)$ and $p_L = (V, W)$,
- 215 • monotonicity condition: $v_1 \leq v_2 \leq \dots \leq v_L$ and $w_1 \leq w_2 \leq \dots \leq w_L$,
- 216 • step size condition: $p_{l+1} - p_l \in \{(1, 0), (0, 1), (1, 1)\}$ for $l \in [1 : L - 1]$.

217 The total cost $c_p(X, Y)$ of a given warping path is

$$c_p(X, Y) := \sum_{l=1}^L c(x_{v_l}, y_{w_l}). \quad (7)$$

218 Finally, an *optimal warping path* between X and Y is a warping path p^* having minimal
 219 total cost among all possible warping paths. The *DTW distance* between X and Y is then
 220 simply defined as the total cost associated with the optimal warping path.

221 As the DTW might be expensive to evaluate, it is worth pointing out the existence of some
 222 accelerated versions of the algorithm reducing this cost when the number of sampling points is
 223 too high. Many of them are based on the restriction imposed to the set of acceptable (V, W) -
 224 warping paths (so that the whole cost matrix is not needed), by introducing weight functions
 225 that privilege certain specific paths, or by modifying the step-size condition (Tavenard and
 226 Amsaleg, 2013).

227 **2.3 Probabilistic modeling of features**

228 The main objective of this work is to develop a novel functional outlier detection technique
 229 which is as general as possible, and sensitive to the main types of outliers that are usually
 230 found in the industrial domain (i.e. shape and magnitude outliers). The first problem
 231 that we may encounter when setting such an objective is firstly the lack of a complete
 232 and indisputable definition of what constitutes an outlier in a set of data. Considering the
 233 definition provided in Dai et al. (2020) as data that behave *in an abnormal way* with respect
 234 to the other considered objects, this approach requires the definition of what an abnormality
 235 is, and it is usually quantified as the extremal values of a measure that is sensitive to the
 236 searched outliers.

237 A more general (but more difficult to apply) definition of what constitutes an outlier is a
238 subset of data that has been generated by a different process than the majority of data
239 present in the considered set (Hawkins, 1980). As an example, a set of measurements could
240 have a small amount of incorrect data points (measurement errors) that may not be obvious
241 at first (these data are not generated the same way as the others). This can also happen
242 in the simulation domain. Simply changing the compilers, computers or the version of
243 simulation codes can significantly change the outcome of any physical simulation. Finding
244 these abnormalities is fundamental in order to ensure the quality of any dataset.

245 Let us suppose that a certain number of features are available to describe our functional data
246 and are able to capture the specific characteristics of both central and abnormal observations.
247 If $\mathcal{U} = \{u_1, \dots, u_r, \dots, u_R\}$ represents this set of features, with no imposed a priori restrictions
248 on its size, such that $\forall u_r \in \mathcal{U}, u_r : \mathcal{F} \rightarrow \mathbb{R}$, then it would be possible to quantify the
249 anomalous behavior according to each measure through the extreme-value analysis theory.

250 The generalization of this theory is based on the use of probabilistic models that can be
251 adjusted to the data. Generally, these models are *generative*, i.e., they are based on the
252 estimation of the probability of occurrence of a data point (multivariate features in our case)
253 accordingly with an assumed underlying model. Once a parametric family has been chosen
254 for the generative model, its values must be estimated through an optimization algorithm.

255 The use of joint multivariate probabilistic models also has the advantage of providing a tool
256 able of taking into account the interaction between the different features used to evaluate the
257 dataset, in addition to providing a score of outlyingness related to a probability of occurrence.

258 When the underlying process that generates the data is unknown, the use of Gaussian Mix-
259 ture Models (GMM) (Reynolds, 2009) is practical due to the vast existent knowledge of
260 these models. Assuming that R descriptive features are available, the form of the associated
261 R -dimensional multivariate Gaussian mixture density function of the random vector $\mathbf{u} \in \mathbb{R}^R$

262 is

$$p(\mathbf{u}) = \sum_{k=1}^K \omega_k f_k(\mathbf{u}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (8)$$

263 where f_k represents each single Gaussian multivariate probability density function, $\boldsymbol{\mu}_k \in$
 264 \mathbb{R}^R represents its vector of means, and $\boldsymbol{\Sigma}_k \in \mathbb{R}^R \times \mathbb{R}^R$ is the corresponding covariance
 265 matrix. The weight of each individual density among the K components is represented by
 266 $\boldsymbol{\omega} \in \mathbb{R}^K$, $\sum_{k=1}^K \omega_k = 1$ and $\omega_k > 0 \forall k \in \{1, \dots, K\}$ and can be interpreted as the mixing
 267 probabilities of the components.

268 2.4 *Adapting GMM for outlier detection*

269 The use of probabilistic models in lower-dimensional feature spaces is useful in order to
 270 detect anomalous points assuming that they provide a good fit for the data and are able of
 271 capturing the central tendency of the data. However, their basic use suffers from well-known
 272 spurious effects in the outlier detection domain (Aggarwal (2017)). The main problems are:

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

283 These reasons motivate the modifications of the Expectation Maximization (EM) algorithm
 284 used in order to fit the GMM to the sample of data in the considered feature space, so that

285 these problems may be efficiently managed and the probabilistic model can be appropriate
 286 for outlier detection. Let $\{z_1, \dots, z_n\}$ be an i.i.d. sample of functional data, with $\{\mathbf{u}_1, \dots, \mathbf{u}_n\}$
 287 being the set of associated features for each curve, the objective is the estimation of the
 288 set of parameters of the GMM $\{\boldsymbol{\omega}, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$. This optimization problem is usually solved
 289 by maximum likelihood estimation via the EM algorithm (Moon, 1996). The form of the
 290 problem is:

$$\max_{\boldsymbol{\omega} \in \mathbb{R}^K, \{\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K} \sum_{i=1}^n \log \left(\sum_{k=1}^K f_k(\mathbf{u}_i; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right). \quad (9)$$

291 This well-known algorithm consists in maximizing the log-likelihood function of the Gaussian
 292 mixture when a certain amount of binary latent variables \mathbf{g} such that $g_k \in \{0, 1\}$ and
 293 $\sum_{k=1}^K g_k = 1$, and which allow to represent the corresponding k th gaussian component that is
 294 considered each time. This way, the conditional distribution of $p(\mathbf{u}|g_k = 1) = f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$
 295 and $p(\mathbf{u}) = \sum_{\mathbf{g}} p(\mathbf{g})p(\mathbf{u}|\mathbf{g}) = \sum_{k=1}^K \omega_k f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. The last element required for the
 296 estimation of the GMM parameters in each step are the conditional probabilities of \mathbf{g} given
 297 \mathbf{u} (usually called *responsibility* that component k takes in explaining the observation). Their
 298 values are:

$$\gamma(g_k) = p(g_k = 1|\mathbf{u}) = \frac{p(g_k = 1)p(\mathbf{u}|g_k = 1)}{\sum_{j=1}^K p(g_j = 1)p(\mathbf{u}|g_j = 1)} = \frac{\omega_k f_k(\mathbf{u}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^K \omega_j f_k(\mathbf{u}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \quad (10)$$

299 This way, the estimation of the parameters is performed in the following steps:

- 300 1. Initialize the values of the desired parameters $\{\boldsymbol{\omega}_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\}_{k=1}^K$,
- 301 2. **E-step**: Evaluate the current responsibilities with the previous parameter values,
- 302 3. **M-step** Re-estimate the parameters by using the responsibilities,
- 303 4. Evaluate the log-likelihood function presented in Eq. (9).

304 The modifications of the algorithm consist in iteratively check and reinitialize the estimated

305 parameters during the procedure in order to detect the undesired and spurious effects by
 306 adding two steps after the estimation of the likelihood function:

- 307 1. Check if $\mathbf{u}_i = \boldsymbol{\mu}_k$ for any $k \in \{1, \dots, K\}$. The apparition of these singularities may
 308 maximize the log-likelihood function to infinity, since it is unbounded, and cause an
 309 overfitting of the data to isolated points. This phenomenon is well described in Bishop
 310 (2006). If this kind of anomaly is detected, the point is subtracted from the sample
 311 and the EM algorithm continues without it. Naturally, these points will be subject to
 312 close analysis since they are good candidates for potential outliers in the sample.
- 313 2. For each iteration step, ω_k can be viewed as the prior probability of $g_j = 1$, and $\gamma(g_j)$
 314 can be seen as the posterior probability given the observed sample. If this posterior
 315 probability is considered too low (in our applications we shall take a value of 0.1 as the
 316 minimum weight of the mixing coefficients), we will consider that the corresponding
 317 component is either overfitting the data, or that it has detected a small subset of
 318 points which is not representative of the central trend of the data. In this case, the
 319 other calculated parameters of the components are kept and the values of means and
 320 covariances of the small cluster are reinitialized to a random value in the space.

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

324 **3** **Outlier detection**

325 **3.1** *Test for outlyingness*

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

$$H_0 : \mathbf{u}_j \text{ has been generated by } f_k \text{ with probability at least,} \quad (11)$$

$$H_1 : \mathbf{u}_j \text{ is an outlier.}$$

328 Under H_0 , $p(\mathbf{u}_j | g_k = 1) > p_{\alpha k}$, 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
 329 set of data points less likely than \mathbf{u}_α is composed by points which verify:

$$(\mathbf{u} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u} - \boldsymbol{\mu}_k) \geq (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k). \quad (12)$$

330 And therefore, $\mathbb{P}[(\mathbf{u} - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u} - \boldsymbol{\mu}_k) \geq (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)] = 1 - \mathbb{P}[L \leq (\mathbf{u}_\alpha -$
 331 $\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}^{-1} (\mathbf{u}_\alpha - \boldsymbol{\mu}_k)]$, where L follows a Chi-squared distribution, $L \sim \chi^2(k)$. By performing
 332 this test over all of the points considered in the feature space, we have a unique criterion for
 333 outlier detection, such that the outlying points will be the ones presenting p-values under a
 334 certain threshold $\alpha, \forall f_k$.

335 **3.2** *Ordering score in the Feature Space*

336 Once all of the parameters are fixed at their estimated values, it is possible to quantify
 337 the extremal behavior of any point in the considered multivariate space through the notion
 338 of *Level Set* (Ahidar-Coutrix, 2015). Formally, given an absolutely continuous probability
 339 measure ν with density p , the Minimum Level Set D^θ is the set of minimal volume (according
 340 to the Lebesgue measure) whose measure is θ :

$$D^\theta = \underset{D \subset \mathbb{R}^R, \nu(D) \geq \theta}{\operatorname{argmin}} \lambda(D) \quad (13)$$

341 where λ is the Lebesgue measure in \mathbb{R}^R . Due to the concave downward nature of the GMM
 342 (Hosseini and Sra, 2015), the set D is unique and $\nu(D^\theta) = \theta$. This way, a probabilistic score
 343 of outlyingness for functional data can be obtained via the probability mass retained in the
 344 associated Level Set of each functional datum z_i :

$$\hat{\theta}_i = \int_{\mathbb{R}^R} \hat{p}(\mathbf{u}) \mathbb{1}_{\{\hat{p}(\mathbf{u}) \geq \hat{p}(\mathbf{u}_i)\}} d^R \mathbf{u} \quad (14)$$

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

$$\hat{p}(\mathbf{u}; \boldsymbol{\omega}, \hat{\boldsymbol{\mu}}_k, \hat{\boldsymbol{\Sigma}}_k) = \sum_{k=1}^K \hat{\omega}_k \frac{1}{2\pi \sqrt{|\hat{\boldsymbol{\Sigma}}_k|}} \exp\left(-\frac{1}{2}(\mathbf{u} - \hat{\boldsymbol{\mu}}_k)^T \hat{\boldsymbol{\Sigma}}_k^{-1}(\mathbf{u} - \hat{\boldsymbol{\mu}}_k)\right). \quad (15)$$

347 This integral must be solved by numerical methods. For $R = 1, 2$, there exists efficient
 348 software able to solve it, whereas several algorithms based on the Cholesky decomposition
 349 have been developed for the higher-dimensional cases (Genz, 2000). In the applications
 350 presented here we will limit ourselves to two components, orienting each one of them to
 351 magnitude or shape outlier detection.

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

360 Finally, let us consider the more realistic case where the availability of data is actually limited,
 361 and the sample of functional data is too small to generate the GMM with an acceptable
 362 level of reliability (the convergence of the EM algorithm is not necessarily reached). This

363 is for instance the case for expensive industrial simulation codes, such as the mechanical
 364 or thermal-hydraulic simulators. In this case, a natural extension of this idea for outlier
 365 detection can be implemented via bootstrap resampling, Efron and Tibshirani (1994). B
 366 groups are formed by successively drawn with replacement in the original sample. This
 367 way, the absence of data can be mitigated through the re-estimation of the GMM for each
 368 bootstrap group. If for B bootstrap groups $p_b(\mathbf{u})$ represents the GMM of the b th group, the
 369 form of the (bootstrap) estimator of outlyingness would then be:

$$\hat{\theta}_i = \int_{\mathbb{R}^R} \frac{1}{B} \sum_{b=1}^B \hat{p}_b(\mathbf{u}) \mathbb{1}_{\{\hat{p}_b(\mathbf{u}) \geq \hat{p}_b(\mathbf{u}_i)\}} d^R \mathbf{u} \quad (16)$$

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

378 This question can be treated as a model-selection problem, and several metrics are available
 379 in order to estimate an appropriate number of components depending on the sample. Some
 380 examples are the Bayesian Information Criterion (BIC) (Schwarz, 1978) or the Integrated
 381 Completed Likelihood (Biernacki et al., 2000). In this paper, the selection of the number of
 382 components is performed by means of the Bayesian Information Criterion:

$$BIC = 2 \log(\hat{l}) - k \log(n) \quad (17)$$

383 where \hat{l} represents the log-likelihood function for the GMM, k is the chosen number of
 384 components and n is the sample size used for the estimation. The second term introduces a

385 penalty which depends on the number of components in order to mitigate overfitting effects.

386 **3.3** *Proposed detection algorithm*

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

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

405 Another important remark on the algorithm is the *extraction* step, where the most outlying
 406 functions on the set according to the $\hat{\theta}_i$ are separated from the original dataset, and the pro-
 407 cedure is re-applied to the remaining functions until a desired homogeneity level is reached.
 408 Naturally, in the case where a large number of samples is available, some extreme data re-
 409 alizations are bound to appear. So, regardless of the chosen value for α , these significant

Algorithm 1: Functional Outlier Detection

Extraction of $\mathcal{Z} = \{z_1, \dots, z_n\}$, $z_i \in \mathcal{F}$;
 Choose a base of features $\mathcal{U} = \{u_1, \dots, u_R\}$ such as those from section 2.2; Set the significance level α and the number of bootstrap groups B ; **while** $\exists \theta_i > \alpha$ **do**
 Estimate $u_j(z_i) | \forall z_i \in \mathcal{Z}, \forall u_j \in \mathcal{U}$; Estimate GMM parameters:
 $\{\omega^b, \mu_k^b, \Sigma_k^b\}_{k=1}^K; \forall b \in \{1, \dots, B\}, \forall k \in \{1, \dots, K_{max}\}$;
 Set $K = \underset{k=\{1, \dots, K_{max}\}}{\operatorname{argmax}} (BIC(k))$;
 Estimate $\theta_i, \forall z_i \in \mathcal{Z}$ defined by equation (16); Extract $z_i \in \mathcal{Z} | \theta_i < \alpha$ from the sample;
end

410 data might be separated from the set even though they do not actually contaminate the set
 411 since, although unlikely, these data appear due to the nature of the underlying generative
 412 process. Separating (trimming) all of these data may introduce a bias in the estimation of
 413 the successive GMM. This is a known problem in the outlier detection domain. An example
 414 for the functional data case can be found in Febrero-Bande et al. (2008).

415 4 Analytical test cases

416 Following Dai et al. (2020), López-Pintado and Romo (2009) and Long and Huang (2015) the
 417 detection capabilities of the algorithm can be assessed via controlled simulation examples.
 418 Such simulations make it possible to check whether a method succeeds in recognizing different
 419 kinds of outliers, when these are purposefully inserted in the dataset. Thus, inserted outliers
 420 are constructed so as to mimic typical outlying data engineers engineers are facing in their
 421 daily practice. Some common notations for the analytical models are summarized in Table
 422 1.

423 4.1 Numerical test protocol

424 In all of the simulation experiments, there will be a total number of $n = 50$ curves, 49 of which
 425 are generated by a reference model, and one is the contaminating outlier. The functions are

426 defined in the interval $[0, 1]$, with a grid of 30 equally spaced points and $B = 10$ bootstrap
 427 groups. Using the previous notation for the DTW algorithm, $V = W = 30$ points. The
 428 description of the models for these control simulation tests is as follows:

- 429 • Model 1. $Z(t) = 4t + G(t)$ is the function generator for the reference set of curves. In
 430 this case the outliers follow the distribution $Z_o(t) = 4t + G(t) + 2\mathbb{1}_{\{t_I < t\}}$.
- 431 • Model 2. The reference model for the curve generation remains $Z(t) = 4t + G(t)$,
 432 whereas the outliers are now generated from the distribution $Z_o(t) = 4t + G(t) +$
 433 $2\mathbb{1}_{\{t_I < t < t_I + 3\}}$.
- 434 • Model 3. Here the reference model becomes $Z(t) = 30t(1 - t)^{3/2} + G(t)$. The outliers
 435 are generated from $Z_o(t) = 30(1 - t)t^{3/2} + G(t)$.
- 436 • Model 4. For this last case, we keep the reference model as it is for Model 1 and
 437 Model 2, but the outliers simply consist of the sole deterministic part $Z_o(t) = 4t$ (the
 438 Gaussian component is removed).

439 Let us remark that, compared to the latter references, the multiplicative factor of the indi-
 440 cator functions has been reduced in order to make the outliers less apparent. All the outliers
 441 considered in this simulation study constitute shape outliers, and in some cases, such as in
 442 the third model, magnitude outliers as well.

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

- 446 • The h-mode depth (4).
- 447 • The Dynamic Time Warping (DTW) (7).
- 448 • The modified band depth (BD) (3)

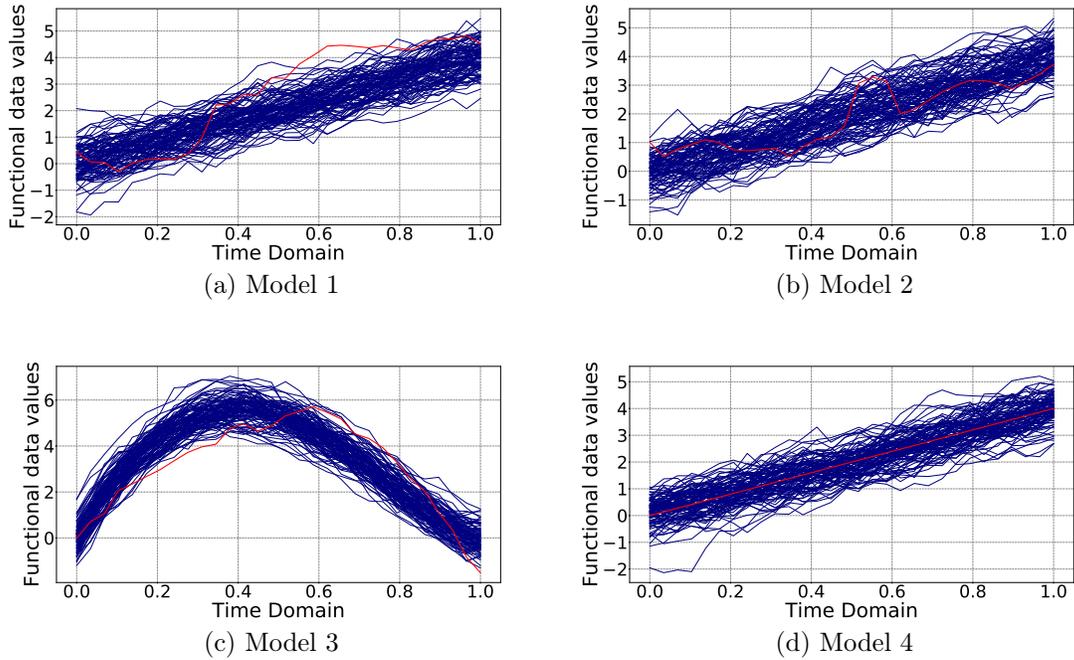


Figure 1: Examples of the four analytical test cases. The blue curves correspond to the 49 ones that are generated from the main model, whereas the red one corresponds to the outlier.

- 449 • The L^2 norm, which is one of the most intuitive and widely used metrics that can be
 450 applied to functional data. It takes the form: $\|z(t)\|_2 = \left(\int_{\mathbb{R}} |z(t)|^2 dt \right)^{1/2}$.

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

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

462 suspicious ones.

463 The Table 2 and Figures 2 and 3 summarize the results for all the replications of the ex-
 464 periments for every specific couple of features, i.e., the six possible combinations of h-mode
 465 depth (hM), modified band depth (BD), Dynamic Time Warping (DTW) and the L^2 metric.
 466 The average ranks of the outlier in each model accordingly to each chosen pair of features
 467 are shown in Table 2.

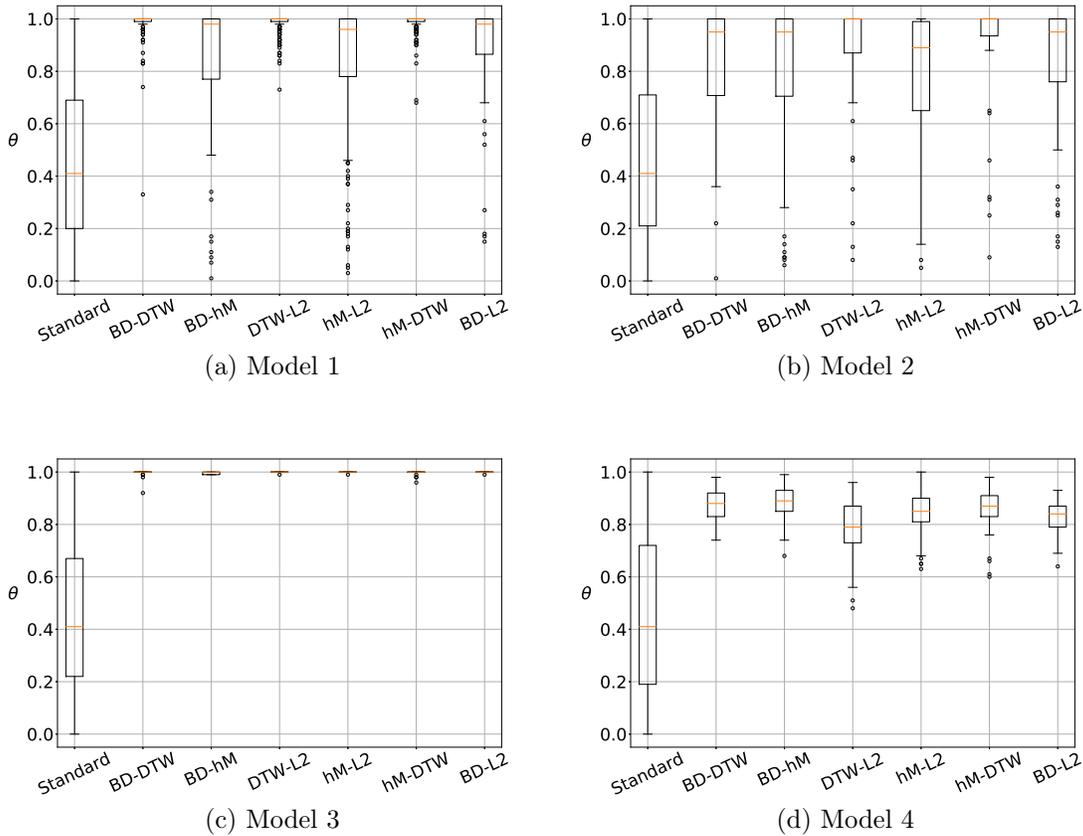


Figure 2: Boxplots of the outlyingness score for all combinations of features in each model in the $N = 100$ replications. The *Standard* boxplot takes into account the whole distribution of $\hat{\theta}_i$ for all the replications of each experiment.

468 As one can see from the Table 2 and Figure 3, the features that show the highest detection
 469 capabilities are the ones that include at least the h-Mode depth or the DTW as a component
 470 of the considered Gaussian mixture model. In the case of the first two models, it is the
 471 combination of both features that yields the best detection results, whereas it remains close
 472 to the best result for the third and fourth models.

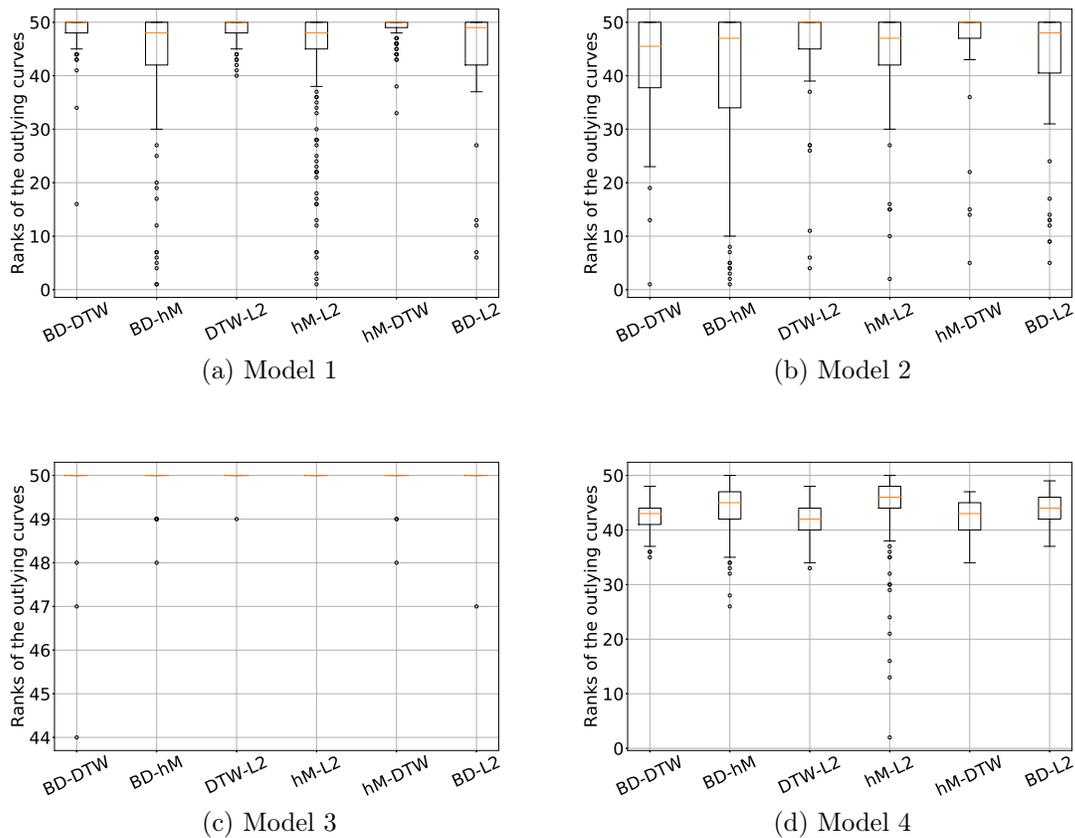


Figure 3: Boxplots of the ranking score of the outlier for all models over the $N = 100$ replications.

473 This result was expected, since the L^2 norm is a very general non-parametric measure which is
 474 probably not well suited for the direct application to the detection of anomalies in functional
 475 data, in spite of its usefulness for functional data characterization. The Modified Band Depth
 476 appears to be adapted for a quick detection of magnitude outliers, but not such a sensitive
 477 measure regarding shape outliers, which are far more complicated to define, identify and
 478 detect. That also explains why the scores for the third model are so high with respect to
 479 the others.

480 The presented scores can be used in order to compare different detection methods that could
 481 be based on identical features (multiple testing, use of level sets, functional boxplots...) as
 482 well as a tool to compare the usefulness of different features for a common detection on the

483 basis of a common detection algorithm.

484 In both cases (for the boxplots of the $\hat{\theta}_i$ and the rankings), it is possible to appreciate not
485 only the absolute detection capabilities that were mentioned before, but also the relative
486 dispersion of the data. This can also be interpreted as an indicator of robustness (which
487 depends on the choice of features). When looking at figures 2 and 3, several aspects can be
488 noted. The first obvious remark is that the detection capabilities for the third model are
489 far superior to those of the others. This is explained by the fact that this is the only one
490 that constitutes both a shape and magnitude outliers, which largely facilitates its detection,
491 even for less sensitive measures such as the L^2 distance. Another interesting point is that
492 for the first model, which is contaminated by a shape outlier, all of the best results are
493 obtained by the combinations that employ the DTW metric. This is also coherent, since it
494 is the feature that best takes into account the shape differences between the curves. Finally,
495 when analyzing the results of the experiments, it can be concluded that the use of a joint
496 model through the h-mode depth and the DTW provide not only the highest detection rates
497 in general, but also the smallest dispersion out of all the possible combinations. This is
498 mostly related to the fact that the DTW is the most sensitive feature when it comes to
499 analyzing shape outliers (it is specifically designed to provide a measure of correspondence
500 between sequences). These two features are the ones that will be retained for the industrial
501 application in the next section.

502 **4.2 *Comparison with state-of-the-art methodologies***

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

506 4.2.1 Functional Boxplots

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

$$C_{0.5} = \left\{ (t, z(t)) : \min_{r=1, \dots, n/2} z_r(t) \leq z(t) \leq \max_{r=1, \dots, n/2} z_r(t) \right\}. \quad (18)$$

509 This region can be interpreted as an analogous of the inter-quartile range for functional data,
 510 and it effectively corresponds to it pointwise. The whiskers of the functional boxplot can be
 511 computed by extending 1.5 times the pointwise extremes of the central region, such that the
 512 outliers are detected if they surpass the frontiers defined by these whiskers. The in-depth
 513 analysis of this method can be found in Sun and Genton (2011).

514 4.2.2 High-Density Regions

515 Introduced by Hyndman (2009), the method consists in regrouping the values of the func-
 516 tional data in the considered time steps in a matrix and performing the Karhunen-Loeve
 517 decomposition, obtaining the corresponding coefficients in a lower-dimensional feature space
 518 where the density of the components is estimated via Kernel Density Estimation (KDE). This
 519 way, the high density regions (HDR) can be defined as the regions such that: $\{\mathbf{u} : f(\mathbf{u}) \geq f_\alpha\}$,
 520 and will correspond to the region with highest probability density function with a cumulative
 521 probability of $1 - \alpha$, which can impose a detection criterion.

522 4.2.3 Directional detector

523 As described in Dai and Genton (2019), let X be a stochastic process, $X : \mathcal{T} \rightarrow \mathbb{R}^p$, the
 524 Functional Directional Outlyingness is defined as:

$$FO(X, F_X) \int_{\mathcal{T}} \|O(X(t), F_{X(t)})\|^2 w(t) dt \quad (19)$$

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

529 4.2.4 Sequential Transformations

530 This algorithm from Dai et al. (2020) relies on the transformation of a wide diversity of shape
 531 outliers into magnitude outliers, much easier to detect through standard procedures. Given a
 532 sequence of operators defined in \mathcal{F} (the functional space that generates the considered data)
 533 $\{\mathcal{G}_k\}, k = 0, 1, 2, \dots$, the method consists in sorting the raw and transformed data into vectors
 534 of ranks for each observation. The vectors of ranks are sorted according to a one-side depth
 535 notion, such as the *extreme rank depth* for instance, and a global envelope is constructed,
 536 which allows the outlier identification.

537 4.2.5 Results

538 The results of the application of the algorithm are given for the previously used 4 models and
 539 different degrees of contamination. The experiments were simulated 500 times for a sample
 540 of curves of $N = 100$, and three different degrees of contamination of outlying curves: 1%, 5%
 541 and 10% of outliers in the sample. The detection rates are summarized in the Table 3.
 542 Firstly, we must note that the identification capabilities and rates are clearly reduced when
 543 the size of the outlying sample is increased. This reduction of the performance of any
 544 detection algorithm is logical, since higher degrees of contamination naturally pollute the
 545 functional sample, which increases the bias of the score that is used for outlier detection. In
 546 the same line, if the size of the outlying sample is considerable (10% of outliers for instance),
 547 an argument can be made to defend that this sample might not be outlying, and that it
 548 simply corresponds to another mode in a hypothetical multimodal functional sample. This
 549 kind of phenomenon, as well as masking effects, are described in detail in Aggarwal (2017).

550 Looking at the results, we can appreciate that the performance of the proposed algorithm
551 is indeed competitive and on par with existent methods, even for complex sets of functional
552 data, such as Model 4. In this case, we can clearly appreciate how the inclusion of a measure
553 specifically dedicated to the detection of shape differences allows the consistent detection
554 of the outlier. This capability is especially significant when we compare it with the other
555 methods, which prove to be unable to detect this kind of shape outlier. In the case of the
556 widely used Functional Boxplots, this is to be expected since they are intended to detect
557 magnitude outliers. Regarding the HDR method, its low detection capabilities in this case are
558 due to the fact that the low-dimensional representation through robust Functional Principal
559 Component Analysis is not sufficiently precise to capture the outlying nature of the straight
560 line. It is indeed possible that retaining a higher number of modes in this case could allow
561 better detection capabilities, but this procedure greatly increases the curse of dimensionality
562 problem (even if this subject is not treated in the paper by Hyndman (2009)), and it does
563 not allow visualization purposes.

564 It is clear that Model 3 (being the only pure magnitude outlier amongst the considered
565 models) is the most simple and easy to detect and virtually any method can consistently
566 detect this kind of outlier when the sample is not overly polluted. Methods which rely the
567 most on the density of curves in the functional space and their trends is more vulnerable to
568 the bias induced in the sample by the curves, as they tend to identify the proportion of curves
569 that *behave unusually* as belonging to a different mode of curves instead of genuine outliers.
570 In the case of the functional boxplots, this is to be expected since by construction they
571 are dedicated to the detection of magnitude outliers, which is useful if the contamination
572 of the sample is made by a wide variety of magnitude outliers, but not so much if those
573 outliers have all been generated by a homogeneous family of curves. In the case of the HDR
574 plots, the existence of a homogeneous sample of outliers generates a set of points in their
575 two-dimensional feature space of principal component scores with a high density of data.
576 In Models 1 and 2 the conclusions are similar (both models present a combination of slight

577 magnitude and shape outliers). Most methods do not showcase any robustness for such slight
 578 magnitude outliers, contrary to the presented algorithm. The main conclusion that can be
 579 extracted from these tests is that most methods struggle to find outliers when they are not
 580 apparent, as it is the case of the models presented here.

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

585 **4.3 Ranking results**

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

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

596 The ranking experiments are the same ones as the ones performed for outlier detection
 597 testing, with one outlier in the sample of 100 curves. The results are presented in Table 4.

598 In this case, we can appreciate that more advanced ranking methods such as the one presented
 599 here and the one presented in Dai et al. (2020) provide a consistent ordering of the functional
 600 data. The results provided by these up to date methods show that for the sample of 100
 601 curves, the introduced outlier is always found to belong to the 5% more outlying curves of
 602 the sample, and is frequently found to be the most outlying in simpler cases like Model 3.

603 Usual ranking techniques for functional data such as depth definitions fail to clearly identify
 604 the more outlying nature of the outlier in the sample, and cannot be reliably used as order
 605 measures in such homogeneous packages of univariate curves.

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

614 5 Industrial test-case study

615 5.1 *Presentation*

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

624 A LOCA accident is originated by a breach in the primary circuit, which is designed to
 625 evacuate the heat generated by the nuclear core. The sudden loss of large quantities of coolant
 626 implies a fast increase of the water temperature nearby the nuclear fuel rods, due to the
 627 residual power generated by the core during the accidental transient. This power discharge

628 and the subsequent temperature elevation must be compensated by the injection of water
629 through a dedicated safety system. This is supposed to ensure that fuel rods temperature
630 would remain below the fusion point at all times. Hence, the main safety criterion in LOCA
631 with regard to the confinement of the fuel concerns its Peak Cladding Temperature (PCT),
632 that is its maximum cladding temperature over the duration of the LOCA (here, the time-
633 dependent cladding temperature is the maximum cladding temperature for all the fuel rods,
634 whatever its localization in the core).

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

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

647 The total number of simulations that can be performed is relatively limited for such a high-
648 dimensional input vector, since each run of the code takes around one hour to finish. For
649 this reason, the use of classical multivariate statistical techniques is not straightforward.
650 This explains the widely spread use of space-filling design methods in order to maximize the
651 coverage of the input space (Iooss and Marrel, 2019), as well as metamodeling techniques to
652 better exploit the number of code runs available for the physical model. Briefly explained,
653 space-filling designs try to optimally explore the space of input variables i.e., they establish
654 criteria in order to better choose the analyzed points of this high-dimensional space (in the

655 case of nuclear transients, there can easily exist more than a hundred input variables). In
656 the case of metamodels, they are mathematical approximations of more complex physical
657 models that, despite showing a higher precision in their calculations, take a much higher
658 computation time. The use of metamodels helps to improve (increase) the total number
659 of available simulations, so that the results that are finally obtained are more statistically
660 relevant.

661 In this context, the consideration of the whole functional output (the evolution of the max-
662 imum cladding temperature, whatever its location) is expected to provide a better insight
663 on the physical phenomena that govern in the transient than the scalar value of the PCT
664 alone. In our case, 1000 Monte Carlo runs of the code were launched, generating the set of
665 curves that is presented in Figure 4 for the evolution of the maximum cladding temperature
666 during the transient.

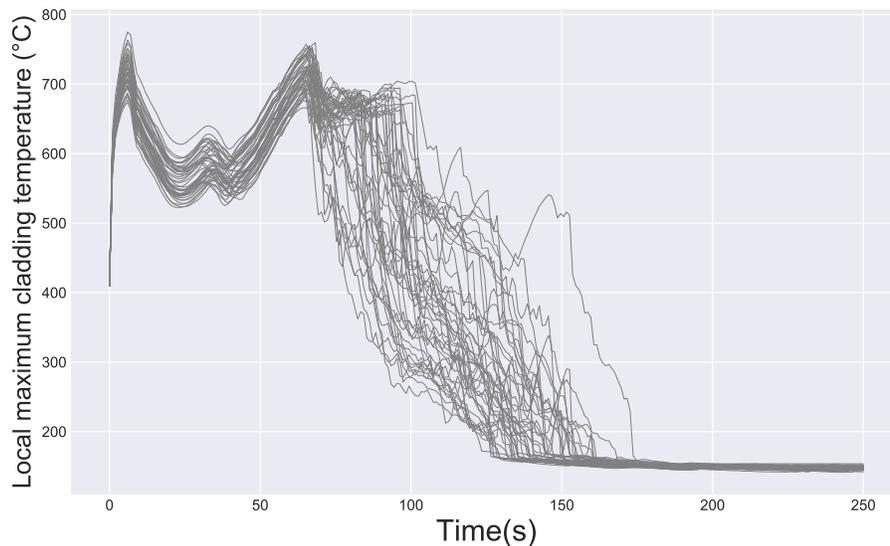


Figure 4: Examples of IBLOCA transients simulated with CATHARE2. Only 40 curves are displayed for clarity.

667 **5.2** *Functional outlier detection*

668 The previously presented outlier detection technique is applied to this set of curves. Both
 669 the h-mode depth and the DTW are the selected features in order to obtain the degree of
 670 outlyingness of each functional datum. The curves presenting a degree $\hat{\theta}$ of outlyingness over
 671 95% are shown in Figure 5.

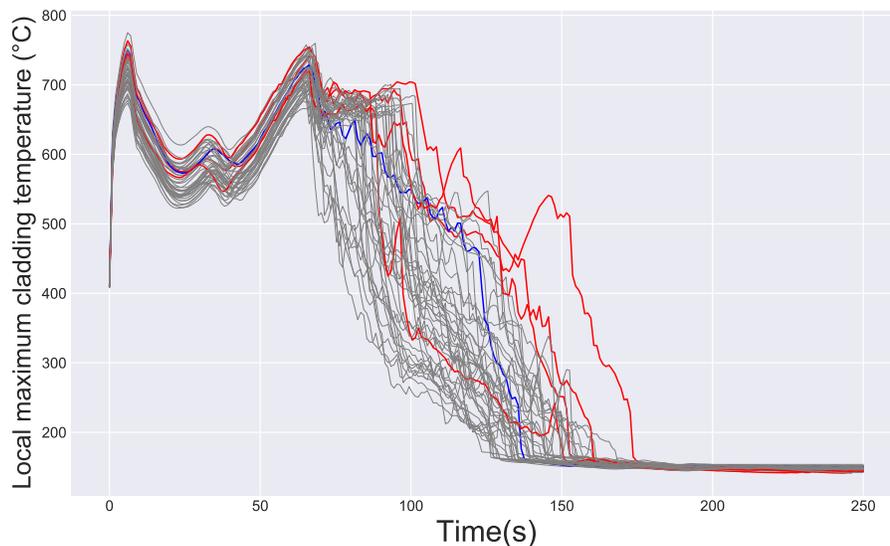


Figure 5: IBLOCA transient curves presenting the highest degree of outlyingness (red) and the least outlying curve (blue).

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

677 Two other magnitude outliers have been identified, and one of them is also a shape outlier,
 678 presenting an anomalous peak of temperature after about 120 seconds of simulation (physical
 679 time). The final main outlier is a pure shape one, remaining zones with high density of data
 680 during the whole domain of simulation, but presenting two peaks of temperature. Specially
 681 notorious is the first peak, which occurs around the 100 seconds of transient, in a time

682 interval that does not match the vast majority of curves.

683 5.3 *Sensitivity analysis on outliers*

684 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
685 which have an actual influence on it can have the same importance, if not more. A way
686 of performing this analysis is to establish some kind of dependence measure between the
687 inputs of the simulation code and the outlying score θ . However, the high dimensionality of
688 the problem and the possible correlations between the input variables of the code can make
689 this a difficult task. The Hilbert Schmidt Independence Criterion (Da Veiga, 2015) can be a
690 useful tool in this context in order to test the dependence between the scalar input variables
691 of the code and the outlying score. This is a first step in order to understand which physical
692 variables are the ones that actually influence the anomalous behavior of the outputs.

694 By performing statistical tests on the HSIC values of the couples (X_i, Y) in the design of
695 experiments it is possible to quantify their dependence. Without going into the technical
696 details of the procedure (see De Lozzo and Marrel (2016)), the HSIC represents a dependence
697 measure between both variables, and can be used in order to build a statistical test with null
698 hypothesis: \mathcal{H}_0 : the variable X_i and Y are independent. The hypothesis is rejected if the
699 associated p-value of the test is inferior to a significance level threshold α . If \mathcal{H}_0 is rejected,
700 then the existence of a dependence structure between the input variable X_i and the output
701 exists.

702 In this case, we apply this measure in order to perform a Target Sensitivity Analysis, i.e.,
703 sensitivity analysis but quantifying the influence of the considered parameters in a restricted
704 domain \mathcal{M} of the possible output values ($\mathcal{M} \subset \Theta$ such that $\mathcal{M} = \{\theta \in \Theta | \theta > 0.9\}$). This
705 application to the set of input data and the obtained values of θ yields a number of influential
706 variables that are shown in Table 5.

707 All of these input variables can be considered to have an influence on the outlyingness

708 of the resulting simulated curve, which represent the evolution of the maximum cladding
709 temperature during the transient. Variable X_{16} represents the friction between the injected
710 water in the primary circuit by the accumulators and the injection line. This parameters has
711 been found to be influential in other similar studies since the compensation of the lost water
712 during the transient is mainly guaranteed by this system (the accumulators), and therefore
713 the line connecting it to the primary is of crucial importance. If this value of friction increases,
714 the water flow will be reduced, with the consequent increase in the average temperatures of
715 the fuel.

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

724 Finally, variables X_{62} and X_{68} are representative of phenomena which occur between the
725 steam and coolant water in the downcomer (the annular part which links the injection line
726 of the accumulators and the nuclear core). This element is critical during the reflood process,
727 which is why increases in the friction between the ascending steam and the descending water
728 in this element are penalizing from a safety point of view. This is due to the fact that if
729 the friction coefficient increases, it means that the momentum of the injected water will be
730 reduced, and the core will take longer to be filled. Similar conclusions can be obtained for
731 the heat exchange coefficient between both phases in the downcomer, since low values for
732 this variable will imply lower heat extraction rates in the vessel.

733 The study of two-dimensional scatter plots between these input parameters and the outlying
734 score θ could already prove to be useful in order to visualize how the inputs affect the outlying

735 nature of the functional outputs. An example that illustrates the idea is presented in Figure
 736 6

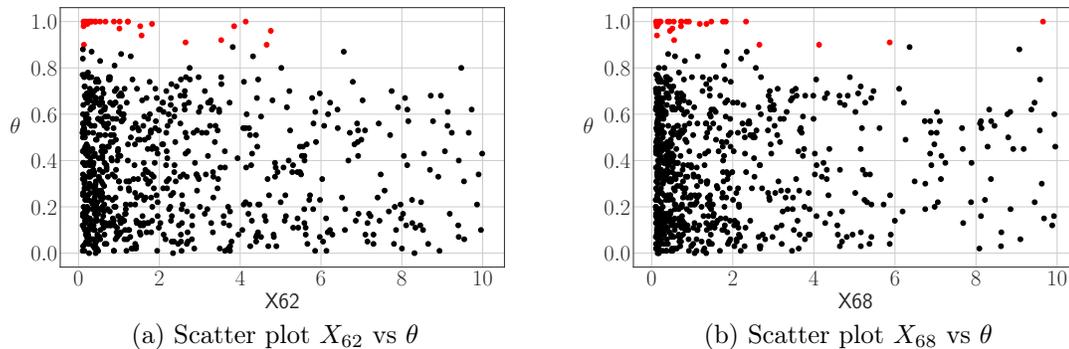


Figure 6: Scatter plots of two input variables of CATHARE2 and the outlyingness score θ . The points correspond to the bivariate plot of the values of the selected variable and the corresponding θ . The red dots correspond to the simulations that have been retained as outliers.

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

6 Discussion and conclusions

This paper has dealt with a fundamental branch of functional data analysis: the outlier detection problem. 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 utilization 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 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 in order 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 on the basis of 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 capable of accounting for the dependence of the input variables of the numerical simulator, Hilbert Schmidt Independence Criterion (HSIC).

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

776 real physical cases. In the case of numerical simulators, the identification of the inputs of
777 the code that actually present an influence on the anomalous outputs can help engineers to
778 detect possible defects of the code or finding physical phenomena of interest. This is also
779 relevant in order to ensure the quality of the datasets that are used in the assessment reports
780 of critical systems such as nuclear power plants.

781 References

- 782 Aggarwal, C. (2017). *Outlier Analysis*. Springer International Publishing, 2 edition.
- 783 Ahidar-Coutrix, A. (2015). *Surfaces quantile : propriétés, convergences et applications*. PhD
784 thesis, Université de Toulouse.
- 785 Arribas-Gil, A. and Romo, J. (2014). Shape outlier detection and visualization for functional
786 data: the outliergram. *Biostatistics*, 15(4):603–619.
- 787 Auder, B. (2011). *Classification and modelling of computer codes functional outputs: appli-*
788 *cation to accidental thermo-hydraulic computations in pressurized water reactors (PWR)*.
789 PhD thesis, Paris 6.
- 790 Barreyre, C., Laurent, B., Loubes, J.-M., Boussouf, L., and Cabon, B. (2020). Multiple
791 testing for outlier detection in space telemetries. *IEEE Transactions on Big Data*, 6:443–
792 451.
- 793 Bellman, R. and Kalaba, R. (1959). On adaptive control processes. *IRE Transactions on*
794 *Automatic Control*, 4(2):1–9.
- 795 Besse, P., Cardot, H., Faivre, R., and Goulard, M. (2005). Statistical modelling of functional
796 data. *Applied Stochastic Models in Business and Industry*, 21:165–173.
- 797 Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering

- 798 with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence*,
 799 *IEEE Transactions on*, 22:719 – 725.
- 800 Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- 801 Chamroukhi, F. and Nguyen, H. D. (2019). Model-based clustering and classification of
 802 functional data. *WIREs Data Mining and Knowledge Discovery*, 9(4):1298.
- 803 Cuevas, A., Febrero-Bande, M., and Fraiman, R. (2007). Robust estimation and classification
 804 for functional data via projection-based depth notions. *Computational Statistics*, 22:481–
 805 496.
- 806 Cuevas, A. and Fraiman, R. (2009). On depth measures and dual statistics. a methodology
 807 for dealing with general data. *Journal of Multivariate Analysis*, 100:753–766.
- 808 Da Veiga, S. (2015). Global sensitivity analysis with dependence measures. *Journal of*
 809 *Statistical Computation and Simulation*, 85(7):1283–1305.
- 810 Dai, W. and Genton, M. (2019). Directional outlyingness for multivariate functional data.
 811 *Computational Statistics & Data Analysis*, pages 50–66.
- 812 Dai, W., Mrkvicka, T., Sun, Y., and Genton, M. (2020). Functional outlier detection and
 813 taxonomy by sequential transformations. *Computational Statistics and Data Analysis*,
 814 149.
- 815 De Lozzo, M. and Marrel, A. (2016). New improvements in the use of dependence measures
 816 for sensitivity analysis and screening. *Journal of Statistical Computation and Simulation*,
 817 86(15):3038–3058.
- 818 Efron, B. and Tibshirani, R. (1994). *An Introduction to the Bootstrap*. Macmillan Publishers
 819 Limited. All rights reserved.

- 820 Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in
 821 functional data by depth measures, with application to identify abnormal nox levels. *En-*
 822 *vironmetrics*, 19:331 – 345.
- 823 Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and*
 824 *Practice*, volume 51. Springer.
- 825 Geffraye, G., Antoni, O., Farvacque, M., Kadri, D., Lavialle, G., Rameau, B., and Ruby,
 826 A. (2011). CATHARE 2 V2.5_2: A single version for various applications. *Nuclear*
 827 *Engineering and Design*, 241:4456–4463.
- 828 Genz, A. (2000). Numerical computation of multivariate normal probabilities. *Journal of*
 829 *Computational and Graphical Statistics*, 1:141–149.
- 830 Gijbels, I. and Nagy, S. (2015). Consistency of non-integrated depths for functional data.
 831 *Journal of Multivariate Analysis*, 140:259–282.
- 832 Grenander, U. (1950). Stochastic processes and statistical inference. *Ark. Mat.*, 1(3):195–277.
- 833 Hawkins, D. (1980). *Identification of Outliers*. Monographs on applied probability and
 834 statistics. Chapman and Hall.
- 835 Hosseini, R. and Sra, S. (2015). Matrix manifold optimization for Gaussian mixtures. In
 836 Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., and Garnett, R., editors, *Advances in*
 837 *Neural Information Processing Systems*, volume 28. Curran Associates, Inc.
- 838 Hyndman, R. J. (2009). Rainbow plots , bagplots and boxplots for functional data. *Journal*
 839 *of Computational and Graphical Statistics*, 19:29–45.
- 840 IAEA (2003). *Accident Analysis for Nuclear Power Plants with Pressurized Water Reactors*.
 841 Number 30 in Safety Reports Series. INTERNATIONAL ATOMIC ENERGY AGENCY,
 842 Vienna.

- 843 Ieva, F., Paganoni, A., Pigoli, D., and Vitelli, V. (2011). ECG signal reconstruction, land-
 844 mark registration and functional classification. In *SCo 2011, 7th conference*.
- 845 Iooss, B. and Marrel, A. (2019). Advanced methodology for uncertainty propagation in
 846 computer experiments with large number of inputs. *Nuclear Technology*, 205(12):1588–
 847 1606.
- 848 Jacques, J. and Preda, C. (2014). Model-based clustering for multivariate functional data.
 849 *Computational Statistics and Data Analysis*, 71:92–106.
- 850 James, G. M., Hastie, T. J., and Sugar, C. A. (2000). Principal component models for sparse
 851 functional data. *Biometrika*, pages 587–602.
- 852 Juang, W.-C., Huang, S.-J., Huang, F.-D., Cheng, P.-W., and Wann, S.-R. (2017). Applica-
 853 tion of time series analysis in modelling and forecasting emergency department visits in a
 854 medical centre in Southern Taiwan. *BMJ Open*, 7(11):e018628.
- 855 Long, J. P. and Huang, J. Z. (2015). A study of functional depths. *Preprint*,
 856 *arXiv:1506.01332*.
- 857 López-Pintado, S., Sun, Y., Lin, J. K., and Genton, M. G. (2014). Simplicial band depth for
 858 multivariate functional data. *Advances in Data Analysis and Classification*, 8(3):321–338.
- 859 López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal*
 860 *of the American Statistical Association*, 104(486):718–734.
- 861 Moon, T. (1996). The expectation-maximization algorithm. *Signal Processing Magazine*,
 862 *IEEE*, 13:47 – 60.
- 863 Mozharovskyi, P. (2016). Tukey depth: linear programming and applications. *Preprint*,
 864 *arXiv:1603.00069*.
- 865 Nagy, S., Gijbels, I., and Hlubinka, D. (2017). Depth-based recognition of shape outlying
 866 functions. *Journal of Computational and Graphical Statistics*, 26(4):883–893.

- 867 Nanty, S. (2015). *Stochastic methods for uncertainty treatment of functional variables in*
868 *computer codes : application to safety studies*. PhD thesis, Université Grenoble Alpes.
- 869 Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer series in statistics.
870 Springer.
- 871 Ramsay, J. O. (1982). When the data are functions. *Psychometrika*, 47:379–396.
- 872 Reynolds, D. (2009). *Gaussian Mixture Models*, pages 659–663. Springer US, Boston, MA.
- 873 Roustant, O., Joucla, J., and Probst, P. (2010). Kriging as an alternative for a more precise
874 analysis of output parameters in nuclear safety—large break LOCA calculation. *Applied*
875 *Stochastic Models in Business and Industry*, 26:565–576.
- 876 Santner, T. J., Williams, B. J., and Notz, W. (2003). *The design and analysis of computer*
877 *experiments*. Springer Series in Statistics. Springer-Verlag, 1 edition.
- 878 Schmutz, A., Jacques, J., Bouveyron, C., Chèze, L., and Martin, P. (2020). Clustering mul-
879 tivariate functional data in group-specific functional subspaces. *Computational Statistics*,
880 35:1101–1131.
- 881 Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6:461–
882 464.
- 883 Sen, R. and Klüppelberg, C. (2019). Time series of functional data with application to yield
884 curves. *Applied Stochastic Models in Business and Industry*, 35:1028–1043.
- 885 Slaets, L., Claeskens, G., and Hubert, M. (2012). Phase and amplitude-based clustering for
886 functional data. *Computational Statistics and Data Analysis*, 56:2360–2374.
- 887 Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and*
888 *Graphical Statistics*, 20(2):316–334.

889 Tavenard, R. and Amsaleg, L. (2013). Improving the efficiency of traditional DTW acceler-
890 ators. *Knowledge and Information Systems*, 42:215–243.

891 **7 Appendix**

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_I	Random point uniformly generated in the definition domain of the function

Table 1: Description of the common parameters of the models

Pairs of features	Model 1	Model 2	Model 3	Model 4
BD-DTW	48.663	41.272	49.621	42.376
BD-hM	41.342	39.067	49.833	43.643
DTW-L2	44.551	42.660	50	43.842
hM-L2	48.937	44.133	49.968	41.929
hM-DTW	49.225	45.154	49.852	42.343
BD-L2	44.254	41.418	49.944	43.672

Table 2: Average rankings of the outlier for each analytical model and combination of features.

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

Table 3: Performances of the different algorithms on the test models. The results are expressed as a percentage (detection rates). DO: Directional Detector; FB: Functional Box-plots; HDR: High-Density Regions.

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

Table 4: Average ranking of the outlier curve across the 100 replications of the experiments for the selected models. The Sequential transformations procedure is presented in Dai et al. (2020). The Modified Band Depth is presented in López-Pintado and Romo (2009), and the standard Integrated Depth appears in Cuevas and Fraiman (2009). 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

Variable	Description
X_{16}	Friction between the water and the discharge line of the accumulators
X_{38}	Global Heat Transfer Coefficient (HTC) in reflow 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 reflow
X_{64}	Friction coefficient between water and steam in core during the reflow phase
X_{68}	HTC between steam and water in the Downcomer

Table 5: 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.