

March 2022

“ICS for multivariate functional anomaly detection with applications to predictive maintenance and quality control”

Aurore Archimbaud, Fériel Boulfani, Xavier Gendre, Klaus Nordhausen,
Anne Ruiz-Gazen and Joni Virta

ICS FOR MULTIVARIATE FUNCTIONAL ANOMALY DETECTION WITH APPLICATIONS TO PREDICTIVE MAINTENANCE AND QUALITY CONTROL

Aurore Archimbaud¹, F riel Boulfani², Xavier Gendre³, Klaus Nordhausen⁴,
Anne Ruiz-Gazen² & Joni Virta⁵

¹ *Erasmus School of Economics, Erasmus University Rotterdam, Netherlands*

² *Toulouse School of Economics, University of Toulouse Capitole, France*

³ *ISAE-SUPAERO, Universit  de Toulouse, France*

⁴ *Department of Mathematics and Statistics, University of Jyv skyl , Finland*

⁵ *Department of Mathematics and Statistics, University of Turku, Finland*

Abstract

Multivariate functional anomaly detection has received a large amount of attention recently. Accounting both the time dimension and the correlations between variables is challenging due to the existence of different types of outliers and the dimension of the data. In the context of predictive maintenance and quality control, data sets often contain a large number of functional variables. However, most of the existing methods focus on a small number of functional variables. Moreover, in fields that have high reliability standards, detecting a small number of potential multivariate functional outliers with as few false positives as possible is crucial. In such a context, the adaptation of the Invariant Coordinate Selection (ICS) method from the multivariate to the multivariate functional case is of particular interest. Two extensions of ICS are proposed: point-wise and global. For both methods, the choice of the relevant components together with outlier identification and interpretation are discussed. A comparison is made on a predictive maintenance example from the avionics field and a quality control example from the microelectronics field. It appears that in such a context, point-wise and global ICS with a small number of selected components can be recommended.

Keywords: affine invariance, functional outlier map, global ICS, outliers, point-wise ICS, scatter matrices.

1 Introduction

Functional data analysis (FDA) is meanwhile a well-established field as demonstrated for example in text books like Ramsay and Silverman (2005); Kokoszka and Reimherr (2017). The main focus of FDA so far seems however to be on univariate functional data and time series as for example recent special issues on FDA like Kokoszka et al. (2017); Ferraty et al. (2022) show. Only recently research on multivariate functional methods started although multivariate functional data are frequently encountered in many different fields including meteorology (Suhaila et al., 2011), medicine (Erbas et al., 2007) and quality control (Mill n-Roures et al., 2018). More precisely, the observations that we consider here are functions of a univariate input variable (usually time) with multivariate output values. In what follows, the number of dimensions of the vector of output curves is denoted by p and the number of observed sets of these p curves is denoted by n with the assumption that $n > p$. Daily measurements of temperature, log precipitation and wind speed at some weather stations provide one example of such data. Other examples include fields with high reliability standards, such as automotive, avionics or aerospace engineering, where many parameters are measured over a certain period of time. In avionics, dozens of technical parameters, such as the airspeed, altitude, and so on, are recorded throughout flights by the aircraft. In microelectronics semiconductor fabrication, there is a large collection of process-control measurements, recorded by various

sensors during the processing of silicon wafers. In the usual non-functional multivariate framework, it is well-known that anomalies might not be outlying in any of the original variables but could exhibit a different correlation pattern compared to the main bulk of the data. While univariate outliers can be identified quite easily, specifically by visually searching for extreme values, multivariate outliers are more difficult to detect, especially in large dimension (Archimbaud et al., 2018). In the univariate functional context, outliers in the sense of their magnitude are usually distinguished from outliers in the sense of their shape. While magnitude outliers can be identified visually by tracking the extreme values of the curves, it can be difficult to detect observations that have neither especially high nor especially low values but instead exhibit anomalous patterns. When looking at multivariate functional data with a large number of dimensions p , both difficulties (multivariate and functional) are combined, and there are many possible types of outliers (see Hubert et al. (2015) for a taxonomy). Outliers are usually first divided into isolated and persistent outliers depending on whether their abnormal behaviour lasts for a short or long time. In the present contribution, focusing on high reliability applications, we assume that any extreme behaviours over a very short time is detected beforehand, and we focus on outliers that are quite persistent.

Detecting outliers in a multivariate functional framework is an issue that has received a large amount of attention very recently (see Rousseeuw et al. (2018), Staerman et al. (2019), Dai et al. (2020), Lejeune et al. (2020) and the references therein). Given that shape outliers are more difficult to identify than magnitude outliers, several recent papers tackle the problem of detecting shape anomalies in either a univariate functional framework (Nagy et al. (2017) and Harris et al. (2021)) or a multivariate functional framework (Dai et al. (2020) and Lejeune et al. (2020)).

Many papers in the univariate and multivariate functional data analysis literature consider the problem of outlier detection through a functional depth or pseudo-depth approach (e.g. Hubert et al. (2015), Kuhnt and Rehage (2016), Dai et al. (2020) and references therein). Each depth notion leads to a centrality index for the observed curves that allows for the identification of non-central curves as outliers. In a multivariate context with large dimension ($p \geq 5$), the depth-based methods are, however, computationally costly. Besides depths, there exist also other approaches such as the shape-based feature extraction method by Lejeune et al. (2020) and the isolation forest method by Staerman et al. (2019). Nevertheless, none of the proposed methods incorporate a dimension reduction step with regard to the dimension p . The lack of a dimension reduction step most likely explains why most examples discussed in the literature on multivariate functional outlier detection do not go beyond $p = 3$, and instead focus on the bivariate case (Kuhnt and Rehage (2016), Rousseeuw et al. (2018), Dai and Genton (2019), Dai et al. (2020), Staerman et al. (2019) and Lejeune et al. (2020)). To the best of our knowledge, robust PCA for univariate functional outlier detection (Sawant et al., 2012) has not been previously extended to the multivariate case.

In a non-functional framework, Archimbaud et al. (2018) consider the problem of outlier detection in large dimension (but still with $n > p$) and show that the Mahalanobis distance, which is a particular depth measure, works poorly. They propose to use instead the Invariant Coordinate Selection (ICS) method (see also Archimbaud et al. (2018b) and the R packages ICSOutlier by Archimbaud et al. (2018a) and ICSShiny by Archimbaud et al. (2018) for the implementation of the method in R), which is based on the joint diagonalization of two scatter matrices. The theoretical properties of the method are studied in Tyler et al. (2009) from the perspectives of mixtures of elliptical distributions and Independent Component Analysis. The method is similar to PCA in the sense that it allows a dimension reduction by calculating and selecting a small number of coordinates, or components. However, instead of relying on the eigendecomposition of one scatter matrix, it relies on the eigendecomposition of one scatter matrix relative to a second matrix (Nordhausen and Ruiz-Gazen, 2022). When observations are structured in groups as is the case in the presence of outliers (one large group accompanied by several small groups), ICS is able to retrieve the Fisher discriminant subspace without knowing the group memberships (see Theorem 3 in Tyler et al. (2009)). With respect to the outlier detection, this capability means that dimension reduction

through ICS is more likely to retain the outlyingness structure compared to PCA which has no guarantee of recovering the Fisher discriminant subspace. Moreover, ICS is affine invariant while PCA is only orthogonally invariant. Archimbaud et al. (2018) consider different pairs of scatter matrices where one scatter is more robust than the other. In fields with high reliability standards, such as automotive, avionics or aerospace industries, where only a small proportion of observations can be abnormal, the authors recommend the use of the regular covariance matrix and the so-called matrix of fourth moments as the scatter pair. They also exhibit the advantages of ICS over the use of Mahalanobis distances and robust PCA.

Recently, Li et al. (2016a) and Virta et al. (2020) proposed to generalize ICS to functional data in the context of Independent Component Analysis. While Li et al. (2016a) focus on the univariate case, Virta et al. (2020) consider multivariate functional data. To the best of our knowledge, there exists no extension of ICS to multivariate functional outlier detection. In the present paper we focus on the problem of detecting a small proportion of outliers (no more than 2%) as in Archimbaud et al. (2018), and we propose two functional ICS extensions. The first method, called “point-wise ICS”, is comparable with the point-wise approach used in Dai and Genton (2018) and Rousseeuw et al. (2018) but replaces, at each time step, the depth procedure with ICS. As detailed in Archimbaud et al. (2018), ICS consists of getting invariant coordinates, selecting the relevant components and calculating, for each observation, an outlyingness score using only the selected components. To select the relevant components, we propose to use the asymptotic test from Nordhausen et al. (2017, 2022) separately at each time point. Following Rousseeuw et al. (2018), the obtained scores are summarized by calculating certain average and dispersion measures. Using the Functional Outlier Map (FOM) from Rousseeuw et al. (2018), the amplitude and shape outliers can then be identified. The second ICS adaptation to functional data is called “global ICS” and consists in expanding the data on a basis, such as a Fourier or B-spline basis, and selecting a number D of basis functions. ICS is then applied to the resulting vectors of $p \times D$ coordinates. This procedure, which is very similar to the procedure described in Virta et al. (2020) in the case of Independent Component Analysis, is then supplemented by an outlier identification step. For global ICS, we follow the recommendations from Archimbaud et al. (2018) and use the scree plot as a simple tool to choose the relevant components and identify outliers using a Monte Carlo cutoff. To illustrate and compare “point-wise” and “global” ICS, we propose to examine in detail an example of daily weather measurements in small dimension. This example illustrates that point-wise and global ICS can identify amplitude and shape outliers with a multivariate perspective. Selecting a small number of components allows us to identify the most extreme multivariate amplitude and shape outliers. In the context of quality control, where univariate outliers have already been detected upstream, a small false positive rate is crucial. Thus, point-wise and global ICS with a small number of selected components are interesting approaches as illustrated on two real data sets.

This paper is organized as follows. Section 2 is divided into five subsections. First, in Subsection 2.1, we recall ordinary ICS for multivariate outlier detection, which is generalized in the following two subsections to the functional setting in two different ways: point-wise functional ICS and global ICS. Point-wise functional ICS in Subsection 2.2 consists in applying the ordinary ICS separately at each time point, whereas global ICS consists in implementing ICS only once on the coefficients of a functional basis expansion of each of the variables and is detailed in Subsection 2.3. Details concerning the criteria for the dimension selection are also discussed for both methods. Since each of the two approaches operates on data pre-processed in specific ways, guidelines to achieve these conditions in practice are discussed in Subsection 2.4. An illustrative and low-dimensional example of applying the two methods to the weather data set is given in Subsection 2.5. Section 3 is devoted to illustrate the two functional ICS approaches on two real data sets from the predictive maintenance and quality control fields. The performance of the methods is also evaluated in Section 3 through a small simulation study with $p = 2$ and $p = 20$. Section 4 concludes and gives research perspectives.

2 ICS for multivariate functional data

2.1 Ordinary ICS

For a p -variate data set $\mathbf{X}_n = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$, where $'$ denotes the transpose operator, a location vector $\mathbf{m}(\mathbf{X}_n)$ is an affine equivariant statistical functional taking values in \mathbb{R}^p and a scatter matrix $\mathbf{V}(\mathbf{X}_n)$ is an affine equivariant statistical functional taking values in the space of $p \times p$ symmetric positive definite matrices. By “affine equivariant” we mean that $\mathbf{m}(\mathbf{X}_n)$ and $\mathbf{V}(\mathbf{X}_n)$ are such that

$$\mathbf{m}(\mathbf{X}_n \mathbf{A} + \mathbf{1}_n \mathbf{b}') = \mathbf{A}' \mathbf{m}(\mathbf{X}_n) + \mathbf{b} \quad \text{and} \quad \mathbf{V}(\mathbf{X}_n \mathbf{A} + \mathbf{1}_n \mathbf{b}') = \mathbf{A}' \mathbf{V}(\mathbf{X}_n) \mathbf{A},$$

where \mathbf{A} is a full rank $p \times p$ matrix, \mathbf{b} a p -vector and $\mathbf{1}_n$ stands for the n -vector full of ones.

Scatter matrices are often computed with respect to a location vector and the most popular, and also most relevant, location vector for our purpose is the sample mean $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$.

One example of a scatter pair is the regular covariance matrix

$$\text{COV}(\mathbf{X}_n) = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})',$$

combined with the so-called scatter matrix of fourth moments

$$\text{COV}_4(\mathbf{X}_n) = \frac{1}{(p+2)n} \sum_{i=1}^n r_i^2 (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})',$$

where $r_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})' \text{COV}(\mathbf{X}_n)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$ is the classical squared Mahalanobis distance (see Nordhausen and Tyler (2015) and the references therein for other examples of scatter matrices). As illustrated on simulations by Archimbaud et al. (2018) in the context of a small proportion of outliers, this particular scatter pair is not only simple and fast to compute but also effective in detecting outliers when compared to other pairs that involve robust scatter estimators and a very popular combination in many other contexts (Nordhausen and Virta, 2019). The context in which there is a small proportion of outliers (less than 2%) is encountered in fields where the data quality standards are high, such as in avionics or aerospace, and it is precisely the context we are interested in.

ICS consists in the joint diagonalization of a scatter pair $(\mathbf{V}_1(\mathbf{X}_n), \mathbf{V}_2(\mathbf{X}_n))$. That is, we look for a $p \times p$ matrix $\mathbf{B}(\mathbf{X}_n)$ and a diagonal matrix $\mathbf{D}(\mathbf{X}_n)$ such that:

$$\mathbf{B}(\mathbf{X}_n) \mathbf{V}_1(\mathbf{X}_n) \mathbf{B}(\mathbf{X}_n)' = \mathbf{I}_p \quad \text{and} \quad \mathbf{B}(\mathbf{X}_n) \mathbf{V}_2(\mathbf{X}_n) \mathbf{B}(\mathbf{X}_n)' = \mathbf{D}(\mathbf{X}_n),$$

where \mathbf{I}_p denotes the $p \times p$ identity matrix. Diagonal of $\mathbf{D}(\mathbf{X}_n)$ contains the eigenvalues of $\mathbf{V}_1(\mathbf{X}_n)^{-1} \mathbf{V}_2(\mathbf{X}_n)$ in decreasing order, while the rows of the matrix $\mathbf{B}(\mathbf{X}_n) = (\mathbf{b}_1, \dots, \mathbf{b}_p)'$ contain the corresponding eigenvectors so that

$$\mathbf{V}_1(\mathbf{X}_n)^{-1} \mathbf{V}_2(\mathbf{X}_n) \mathbf{B}(\mathbf{X}_n)' = \mathbf{B}(\mathbf{X}_n)' \mathbf{D}(\mathbf{X}_n).$$

Using any location estimator $\mathbf{m}(\mathbf{X}_n)$, the corresponding scores,

$$\mathbf{Z}_n = (\mathbf{z}_1, \dots, \mathbf{z}_n)' = (\mathbf{X}_n - \mathbf{1}_n \mathbf{m}(\mathbf{X}_n)') \mathbf{B}(\mathbf{X}_n)',$$

are the affine invariant coordinates or components. As proved in Archimbaud et al. (2018), the Euclidian norm $\sqrt{\mathbf{z}_i' \mathbf{z}_i}$ of the i th observation, $i = 1, \dots, n$, is equal to the Mahalanobis distance of the i th observation from the location $\mathbf{m}(\mathbf{X}_n)$ in the sense of $\mathbf{V}_1(\mathbf{X}_n)$. Note that this property does not depend on the location estimators used in $\mathbf{V}_1(\mathbf{X}_n)$ and $\mathbf{V}_2(\mathbf{X}_n)$. The Mahalanobis distance does not offer the possibility of dimension reduction. However, this property of the invariant coordinates can be useful if the outliers belong to a space of reduced dimension and if we attempt to avoid false positives, as in high reliability

fields. Indeed, ICS offers the possibility to select the components that are helpful to detect rare and real anomalies, and consequently, it avoids false positives caused by noisy dimensions. In the case of a small proportion of outliers, the theoretical properties of ICS (see Archimbaud et al. (2018) for details) lead us to only focus on the invariant components associated to the largest eigenvalues. Archimbaud et al. (2018) propose various automatic selection procedures based on hypothesis testing, but they acknowledge the fact that these procedures tend to select too many components, and they propose the scree plot as an alternative. Once having selected k invariant components, the last step of the procedure is the identification of the outlying observations. For each observation $i = 1, \dots, n$, we calculate its squared “ICS distance”, which corresponds to the squared Euclidian norm in the invariant coordinate system restrained to the k first coordinates:

$$(\text{ICS distance})_{i,k}^2 = \sum_{j=1}^k \left(z_i^j\right)^2, \quad (1)$$

where z_i^j denotes the j th coordinate of the score \mathbf{z}_i . In Archimbaud et al. (2018), an observation is flagged as an outlier when its ICS distance using k components is larger than a cutoff based on Monte Carlo simulations from the standard Gaussian distribution. Being given a data dimension, a scatter pair and a number k of selected components, many Gaussian samples are generated and the ICS distances are computed. A cutoff is derived for a fixed level γ as the $1 - \gamma$ quantile of these distances.

2.2 Point-wise functional ICS

In this and the following subsections we generalize the ordinary ICS to multivariate functional data into two complementary ways. Both approaches assume that our sample consists of n observations of p -variate functions on $[0, 1]$. Furthermore, for point-wise functional ICS, we require that the np curves are aligned both across the n observations and the p features and that we have available the values of the curves on the equispaced grid $\{1/T, 2/T, \dots, 1\}$ for some $T \in \mathbb{N}$. See Subsection 2.4 on preprocessing the data in cases where these assumptions are not satisfied.

Thus, for each of the T time points on the grid, we have a sample of n observations of the point-wise values of the p curves. Point-wise functional ICS then consists of applying the ordinary ICS to each of these samples and calculating the resulting sets of n observations of p invariant coordinates for each of the T time points. For each time point t , we further select $k(t)$ components with a rule specified below and compute the set of n ICS distances, $(\text{ICS distance})_{i,k(t)}^2(t)$, $i = 1, \dots, n$. This process leads us to n curves of ICS distances (one per observational unit) on the grid.

If there is no dimension reduction (i.e., $k(t) = p$ for each time point t), the method is equivalent to the calculation of the Mahalanobis distance at each time point. However, the main advantage of ICS distances compared to Mahalanobis distances is that they can be based on a subset of components that form a more informative subspace. Basically all of those components that appear to be the most non-Gaussian should be selected. ICS can be considered in this case to be non-Gaussian component analysis (NGCA) (Nordhausen et al., 2017), where we assume that all of the outliers lie in a subspace that obviously is non-Gaussian and that this subspace is independent from the uncontaminated Gaussian subspace. Tyler et al. (2009); Nordhausen et al. (2017); Radojicic and Nordhausen (2020) then show that the eigenvalues d_1, \dots, d_p contained in $\mathbf{D}(\mathbf{X}_n)$ are the key to identifying the two subspaces because these can be seen as generalized measures of kurtosis. In the Gaussian subspace all eigenvalues must be equal and the exact values then only depend on the scatter matrices used in ICS. In the case of COV and COV₄, it can be shown that the “Gaussian” eigenvalues are equal to 1, and in our setting of only a few outliers, the “Non-Gaussian” eigenvalues are all larger than one. Thus, we have $d_1 \geq \dots \geq d_k > 1 = \dots = 1$, where the problem is now that k is unknown. One could, for example, use a scree plot or marginal tests as discussed in Archimbaud et al. (2018) or use successive applications of hypothesis tests of the form

$H_{0q} : k = q$ and test $H_{00}, H_{01}, H_{02}, \dots$ to find the value $\hat{k} = q$ where H_{0q} is the first test not rejected at a given significance level. The test statistic and its limiting distribution under the null hypothesis for a given q are presented in Nordhausen et al. (2017) where it is also shown that these statistics are fast to compute and lead to a consistent estimate of k for an appropriate sequence of significance levels.

In point-wise functional ICS, we follow the procedure detailed above, taking 1% as the initial significance level for testing H_{00} , and applying a type of Bonferroni adjustment by dividing the level by 2 for H_{01} , by 3 for H_{02} , and so on (see Archimbaud et al. (2018) for more details). Note that this sequence of statistical tests is performed separately for each of the T time points, implying that the number of selected components varies over time. The procedure leads to two natural graphical displays of the results. First, one can plot the ICS distances of the n observations according to the time, allowing the pin-pointing of both outlying observations and the times during which they exhibit their anomalous behavior. Second, we can plot the estimated dimension $k(t)$ with respect to the time, giving us a glimpse of the “complexity” of the data along the time. Additionally, the ICS distances can also be used to construct a 2-dimensional outlier map as in Rousseeuw et al. (2018). However, we have postponed the definition of this, along with some additional guidelines for interpreting the results, to Subsection 2.5 where the weather data is used to illustrate point-wise functional ICS.

Before moving to global functional ICS, we note that point-wise functional ICS could also be defined at the population level. Namely, let $L^2 := L^2([0, 1], dt)$ denote the Hilbert space of square integrable functions on $[0, 1]$ (with respect to the Lebesgue measure dt) endowed with the usual inner product $\langle f, g \rangle_{L^2} = \int_0^1 f(t) g(t) dt$. Then, let (Ω, \mathcal{F}, P) be a probability space and let $X(t) = (X_1(t), \dots, X_p(t))$ be a centered random element in $\mathcal{L} := L^2 \times \dots \times L^2$ that is measurable $\mathcal{F} \setminus \mathcal{B}$, where \mathcal{B} is the Borel sigma-algebra in the topology generated by the metric induced by the inner product $\langle f, g \rangle_{\mathcal{L}} := \langle f_1, g_1 \rangle_{L^2} + \dots + \langle f_p, g_p \rangle_{L^2}$. Then, point-wise functional ICS consists in applying the ordinary, population-level ICS to each of the random vectors $X(t) = (X_1(t), \dots, X_p(t))$ for $t \in [0, 1]$. In order to be well-defined, this process requires that $\text{Cov}\{X(t)\}$ is invertible and that $E\|X(t)\|^4 < \infty$ for each $t \in [0, 1]$, see, e.g., Ilmonen et al. (2010); Miettinen et al. (2015) for details.

2.3 Global functional ICS

For global functional ICS, we assume that our n observations of p -variate functions on $[0, 1]$ are aligned across the n observations but not necessarily across the p features. Additionally, we assume that, for each of the features, the curves have been projected into a D -dimensional functional space, see Subsection 2.4 for some guidelines to achieve this in practice. These projections are implemented separately for each feature which explains that there is no need to align the curves across the features. This gives us a $n \times pD$ data matrix which no longer consists of functions per se, but is rather made up of n observations of pD variables corresponding to the basis function coefficients of the curves. Assuming that $n > pD$, we define global functional ICS as the application of the ordinary ICS to this data set, producing n observations of pD invariant components.

We now follow Archimbaud et al. (2018), see Subsection 2.1, and use a test procedure to select the number of invariant coordinates using the test proposed by Nordhausen et al. (2017) as detailed previously. However, we instead recommend the scree plot, as advised by Archimbaud et al. (2018), and spot an elbow in the decrease of the eigenvalues or some large differences between successive eigenvalues. This graphical method is simple and tends to select fewer variables than the test procedure which selects all eigenvalues significantly larger than one. It cannot be used in the context of point-wise ICS, where the dimension selection must be performed at each time point and must be automated, but we recommend its use for global ICS. Having selected an appropriate dimension, we proceed by computing the ICS distances of the observations, the plot of which can again be used to identify outliers. We postpone the full illustration of this procedure to Subsection 2.5 where the two proposed functional ICS methods are applied to the weather data set.

Again, global functional ICS also admits a population-level formulation. An equivalent procedure is detailed in the case of independent component analysis in Virta et al. (2020) and, hence, we only give its rough outline in the following. Given the same setting as what was assumed in the population-level version of point-wise functional ICS earlier, assume that $E\|X\|_{\mathcal{L}}^4 < \infty$, where $\|\cdot\|_{\mathcal{L}}$ is the norm associated with the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{L}}$ defined in Subsection 2.2. Applying global functional ICS to X consists of two steps: dimension reduction through the Karhunen-Loeve expansion and subsequential projection onto the eigenfunctions of a specific fourth-order operator. In the first step, we let Σ_{ij} denote the cross-covariance operator between X_i and X_j and collect the p^2 cross-covariance operators into a “matrix of operators” $\Sigma = (\Sigma_{ij})_{i,j=1}^p$. Then, we consider the pD leading eigenfunctions of Σ , project X onto their span and standardize the obtained random element to have covariance operator equal to the identity operator (within the subspace). We denote the resulting finite-dimensional random element by X^* . In the second step, we compute the full set of eigenfunctions of the FOBI-operator $C(X^*)$, see Virta et al. (2020) for its precise formulation, and project X^* onto them. The resulting variables now correspond to the pD invariant coordinates.

2.4 Preprocessing multivariate functional data

The format of the data we observe in practice might not be as clean as we assumed in Subsections 2.2 and 2.3. For example, the numbers of time points can differ across both observations and features, and alignment or warping might be necessary. In this subsection, we briefly discuss some preprocessing steps one might need to carry out before applying the methods.

For point-wise ICS, the curves need to be aligned for both observations and features, while for global ICS, the alignment is necessary only across the observations, even if different features can have different alignments. In both data analyses of Section 3, the features are already aligned for each observation and as a result, the curves are only required to be aligned across observations. For the sake of simplicity, we use a linear interpolation. Other methods as the ones proposed by Tucker et al. (2013) and implemented in Tucker (2020) are also possible. In practice, we fix a value for the number of time points T , for example at the median value of the different curve time lengths. Then, we linearly interpolate each curve at T equally spaced points. For a curve with a length greater than T we reduce the number of points by averaging points and for the curves with a length smaller than T , we add points using linear interpolation. Finally, we rescale the time axis such that the observations occur at the times $\{1/T, 2/T, \dots, 1\}$.

For point-wise ICS, no other preprocessing is needed. For global ICS, a preliminary dimension reduction is necessary. We consider expansion of the functions associated with the aligned data on an orthonormal basis. The expansion of a function $f \in L^2$ according to the orthonormal basis $\{\xi_d\}_{d \in \mathbb{N}}$ is given by

$$\forall t \in [0, 1], f(t) = \sum_{d \in \mathbb{N}} c_d \xi_d(t), \text{ with } c_d = \langle f, \xi_d \rangle_{L^2}.$$

In our application, only discretizations of our functions are available and we estimate c_d on a regular grid of step $1/T$ in $[0, 1]$ by

$$\hat{c}_d = \frac{1}{T} \sum_{t=1}^T f\left(\frac{t}{T}\right) \xi_d\left(\frac{t}{T}\right).$$

To apply global ICS in practice, we only need these coefficients \hat{c}_d but we do not make use of the orthonormality of the basis. To illustrate this in the sequel, we focus on the Fourier basis which is orthonormal and on the B-splines basis which is not. Many other bases can also be used (see Ramsay and Silverman (2005)).

In order to reduce the dimension, a truncation is done and only the $D \in \mathbb{N}$ first coefficients are kept for each variable. The resulting truncation gives a data matrix of dimension $n \times pD$ that is then used

as the input for global functional ICS, as described in Subsection 2.3. We choose the same number of coefficients for each variable for the sake of simplicity but it is conceivable to let this number vary: if we choose a number D_j of coefficients for each of the variables $j = 1, \dots, p$, we get a data set of dimension $n(D_1 + \dots + D_k)$.

As detailed in Barreyre et al. (2019) in the context of satellite data, the choice of the basis to represent functional data in a reduced dimension could have an impact on the whole outlier detection procedure. In particular, the authors exhibit an artificial example in which isolated outliers are more likely to be identified when using a data-dependent basis such as PCA rather than a Fourier basis. However, as mentioned in the introduction, we do not focus on isolated outliers and we only consider Fourier and B-splines bases which lead to similar results in our experience.

2.5 Weather data example

In this subsection, we illustrate the point-wise and global functional ICS methods on a small Spanish weather dataset from the R package *fda.usc* (Febrero-Bande and de la Fuente, 2012). The data set consists of $p = 3$ variables that represent the daily average ($T = 365$) of temperature, wind speed and log precipitation records from 1980 to 2009 from $n = 73$ weather stations in Spain. As in Dai and Genton (2018) the curves are smoothed using a B-spline basis truncated at $D = 11$. No expert opinion on outlying weather stations is available but the example is small enough in terms of the number of observations and variables to be studied in detail by examining the curves. For the sake of readability, the curves are plotted in the appendix with different colours (see Figures 21 to 23 in the appendix). More details can be found in Dai and Genton (2019) and Dai et al. (2020). Each figure focuses on one of the 3 variables. In Figure 21, we coloured in red and identified by a number the curves that look different from the vast majority of curves in terms of their temperature behaviour. These are curves 45 and 56, which have very low temperatures, and the seven curves 34, 35, 36, 55, 57, 58 and 60, which have much flatter temperature curves than the other weather stations (see the left panel). These nine curves are also coloured in red on the wind speed (resp., log precipitation) plot in the middle (resp., right) panel of the same figure. In Figure 22, we have kept the red curves and coloured in blue the curves that were not already coloured in red and that look different from the others in terms of the wind speed. Curves 20 and 59 take large values with several large bumps. Curves 51 and 72 take small wind speed values. Finally, in Figure 23, we have kept the red and blue curves and coloured in green the curves that were not already coloured in red or blue and that look different from the others in terms of the log precipitation. These are curves 33, 39, 44 and 66, which have large values of log precipitation with a small dispersion. All together, this process gives 17 curves out of 73 (23%) that can be suspected as outlying. In the present paper, the objective is to detect only a small percentage (approximately 2%) of the observations as outliers, which corresponds to at most two or three observations in this small data set. Clusters of outliers, such as some curves among the red and green ones in Figure 22, are not relevant in our context, since we are rather looking for observations that differ as much as possible from other observations (while accounting for the interactions between the variables) and that are unique in their outlying behaviour. For both generalizations of ICS to functional data, a detailed outlier detection analysis of the weather data set is presented hereafter. Differences between point-wise and global ICS are also discussed.

Point-wise functional ICS

We applied point-wise functional ICS to the weather data and the resulting time courses of the estimated dimensions and squared ICS distances are given in Figure 1. When no component is selected, no outliers are detected and the ICS distances are thus equal to 0 for these time points. As mentioned previously, the number of selected components varies over time. However, on Figure 1 (left panel), the plot is quite structured with only one selected component except in the middle of the year where two components are selected. Note that selecting more components usually leads to more outliers. Looking at the ICS distance

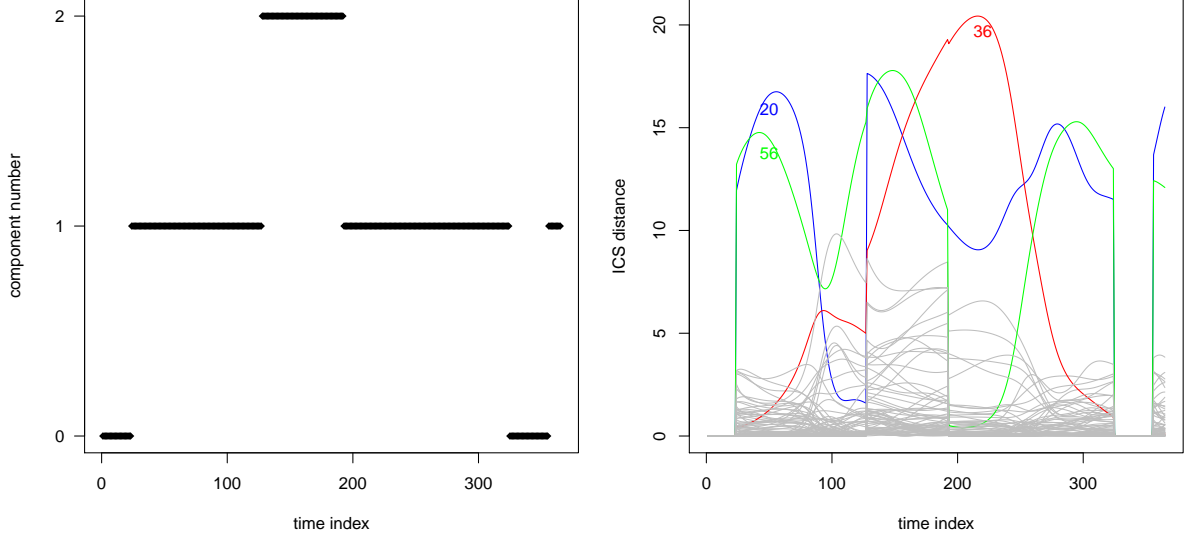


Figure 1: **Weather data set** - Point-wise ICS, Left: Number of components selected, Right: square ICS distance.

curves (the right panel of Figure 1), it is possible to identify the curves 20, 36 and 56 as outlying and detect at which periods of time they differ from the other curves. Observations 20 and 56 differ from the other curves at similar periods of time (at approximately the 50th, 150th and 300th days of the year), but their ICS distances are not the same. Curve 36 differs from the other curves essentially at approximately the 200th day. In the presence of many curves, it can be tedious to identify outliers by looking at such a plot. Moreover, it is costly and not recommended to flag the outliers by calculating a cutoff based on Monte Carlo simulations at each time point. We thus propose to summarize the information and to flag the outliers by using a functional outlier map (FOM), as defined by Rousseeuw et al. (2018). For each observation, we calculate and plot a weighted average (fICS) and a measure of variability (vICS) for the ICS squared distances:

$$\text{fICS}_i = \sum_{t=1}^T W(t) (\text{ICS distance})_{i,k(t)}^2(t) \quad (2)$$

and

$$\text{vICS}_i = \frac{\text{stdev}_i}{1 + \text{fICS}_i}, \quad (3)$$

where $W(\cdot)$ is a weight function such that $\sum_{t=1}^T W(t) = 1$ and stdev_i denotes the standard deviation of the $(\text{ICS distance})_{i,k(t)}^2(t)$ values over time. As mentioned in Rousseeuw et al. (2018), vICS is a relative metric that is preferable to the usual standard deviation. We could also use a weighted standard deviation as suggested in Rousseeuw et al. (2018). The number of components involved in the calculation of the ICS distances may vary from one time to another, and the larger the number of components selected is, the larger the ICS distance. This relationship means that if we limit ourselves to uniform weights $W(t) = 1/T$, time points with a large number of selected components can have a larger impact in the fICS calculation than time points with a small number of selected components. This relationship is especially true for time points at which no components are selected because the ICS distances are zero and such time points do not contribute to the fICS value. To give truly uniform weights over time, it should be

recommended to divide the squared ICS distances at each time by the number of selected dimensions as discussed in Section 3. This standardization is equivalent to take $W(t) = 1/(Tk(t))$ in (2). The FOM is a scatterplot of fICS and vICS. Large values of fICS correspond to curves that are outlying during a long period or for the entire period of time. Note that such curves are not necessarily shifted curves since a high ICS distance can correspond to a multivariate outlier (an observation outlying in the correlation structure but not necessarily extreme). Large vICS values correspond to curves whose behaviour differs from the other curves during some subperiods of time. Figure 2 gives the FOM for the weather data set. Note that the cutoff curve (red dashed curve on Figure 2) is calculated as in Rousseeuw et al. (2018) using the combined functional outlyingness (CFO) with a quantile order of 0.95. Its calculation is adapted from Rousseeuw et al. (2018). For each observation $i = 1, \dots, n$, we define

$$\text{LfICS}_i = \log \left[0.1 + \sqrt{\left(\frac{\text{fICS}_i}{\text{med}(\text{fICS})} \right)^2 + \left(\frac{\text{vICS}_i}{\text{med}(\text{vICS})} \right)^2} \right],$$

where $\text{med}(\cdot)$ stands for the median. An observation i is flagged as an outlier if

$$\frac{\text{LfICS}_i - \text{med}(\text{LfICS})}{\text{MAD}(\text{LfICS})} > \Phi^{-1}(\alpha), \quad (4)$$

where MAD denotes the median absolute deviation, Φ the standard normal cumulative distribution function and α a quantile order. For $\alpha = 0.95$, expression (4) yields the dashed red curve in Figure 2 which is part of an ellipse. The scatterplot clearly distinguishes the curves 20, 36 and 56 from the other curves both in terms of fICS and vICS (with slightly more variability for observation 36 than for observations 20 and 56). As already noticed in Figure 1, these three curves have a large average but also have widely dispersed ICS distances over time.

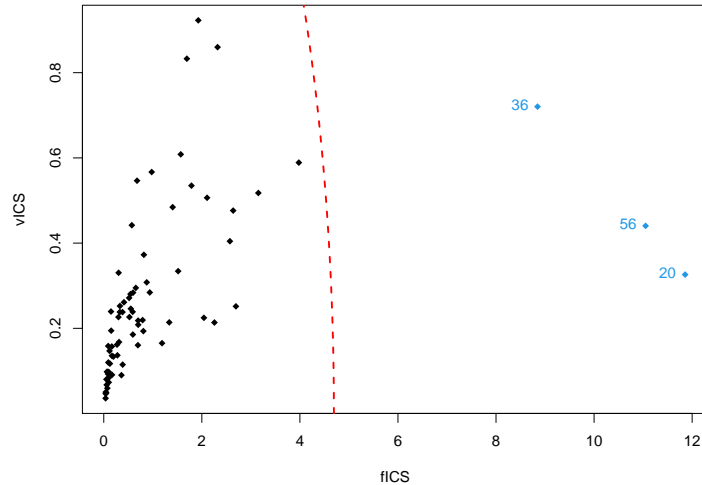


Figure 2: **Weather data set** - Point-wise ICS: FOM with the automatic dimension selection and the cutoff quantile of order 0.95.

The impact of the dimension reduction step can be analysed by looking at Figure 3, where we give the FOM when the number of selected components is set to $k = 1$ (resp., $k = 2$ and $k = 3$) during the whole time period, in the left (resp., middle and right) panel. As anticipated, the number of observations

flagged as outliers increases when the dimension increases. Note that the case $k = p = 3$ corresponds to the Mahalanobis distance and leads to the detection of most of the 17 curves detailed previously and suspected of atypical behaviour (in particular, the green and red curves in Figure 23 in the appendix). In contrast, selecting $k = 1$ gives a similar result to the result obtained with automatic selection in Figure 2. The result obtained with $k = 2$ is in between, with essentially the curves 33, 39, 44 and 66 (the green curves on Figure 23 in the appendix) detected in addition to the curves 20, 36 and 56. In the context that we are interested in, which has a small proportion of outliers to detect, we prefer the automated selection or the choice $k = 1$. The question that arises is how can we interpret the outlyingness of the three observations 20, 36 and 56.

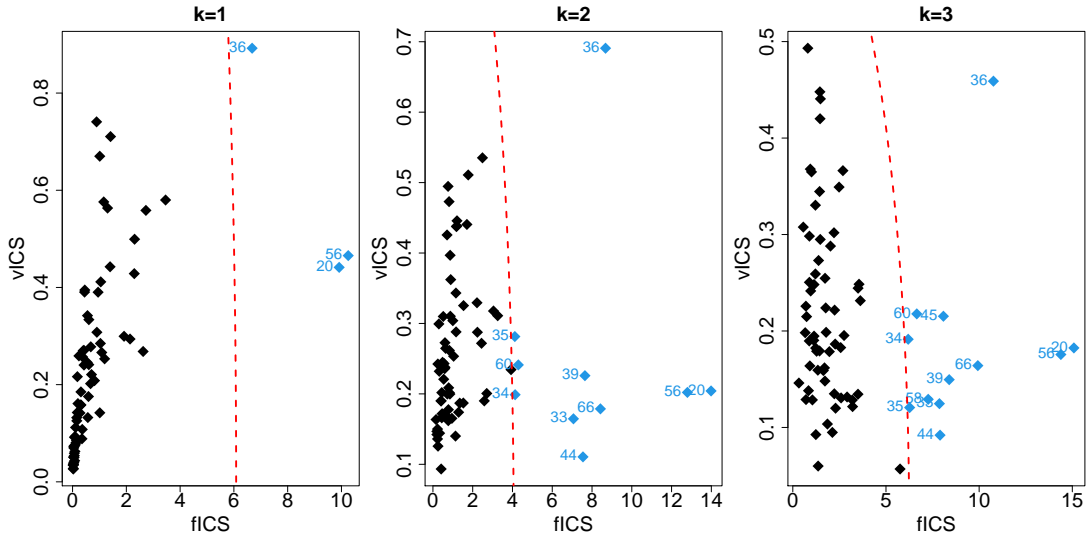


Figure 3: **Weather data set** - Point-wise ICS: FOM with $k = 1$ (resp., $k = 2$ and $k = 3$) on the left (resp., middle and right) panel and the cutoff quantile of order 0.95.

Interpreting outlying curves in the context of point-wise ICS is possible but not easy because of the temporal dimension combined with the possible presence of many variables. For the weather data, it is possible to interpret the three outlying curves 20, 36 and 56 by looking closely at Figures 21 and 22 in the appendix. Curve 36 is unique in the sense that it has a very particular wind speed curve with very high values and a unimodal and peaked shape around the 200th day. Observation 56 is also very special in the sense that it takes quite large values in terms of the wind speed, with some small bumps around the 50th and 300th days and a clear decrease around the 200th day. Other curves exhibit large values for the wind speed and, in particular, all of the red curves in Figure 21 have large and unusual behaviour in terms of temperature. However, observation 56 is the only curve among the red curves that has large wind speed values jointly with small temperature values. Observation 20 is even more atypical than 56 for large wind speed values, even though its temperature values are not so small. Note that curve 59 is also quite different from all of the other curves that have large values of wind speed and medium temperature values. However, as seen in Figure 4, curve 59 (in cyan) is not detected using point-wise ICS.

To make the interpretation easier, we propose to calculate at each time point the correlations between the selected invariant components and the initial variables. Examining such correlation curves is possible if the number of initial variables and the number of selected components are not too large, which is the case for the weather data set, where $p = 3$ and the number of selected components is not larger than 2. The plot on the left (resp., middle and right) panel of Figure 5 gives the correlation curve between the temperature (resp., wind speed and log precipitation) and the first ICS component. Values larger than

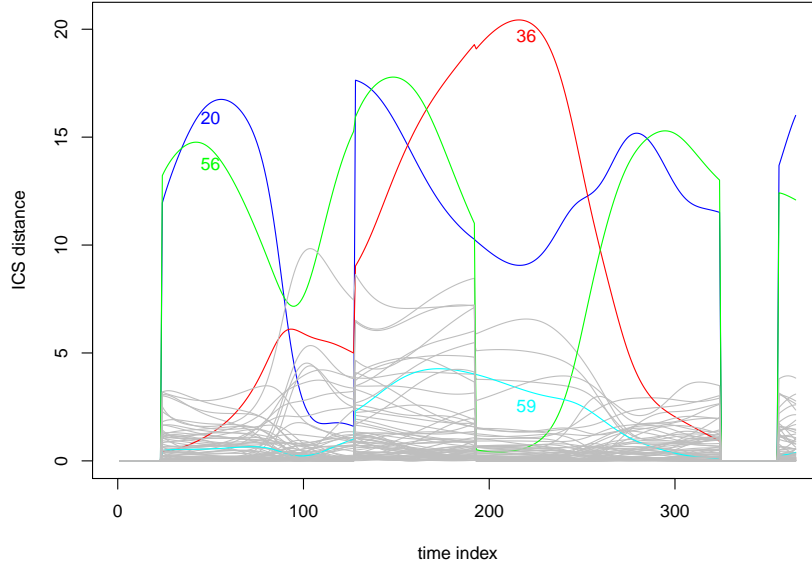


Figure 4: **Weather data set** - Point-wise ICS: square ICS distance.

0.20 in absolute value are plotted in red. Note that the range of correlation values differs from one plot to another to zoom in. It seems clear that the first invariant component is highly positively correlated with the wind speed with a smaller correlation between the 100th and 200th days. During the same period of time, the correlation of the first component becomes negative with the log precipitation and moderately negative with the temperature. These plots in conjunction with Figure 4 confirm our previous interpretation of the outlying curves 20, 36 and 56. At approximately days 50, 200 and 300, the first component is strongly positively correlated with the wind speed, and observations 20 and 56 (resp., 36) take very large wind speed values at days 50 and 300 (resp., 200). Curves 20 and 56 are also outlying at day 150, when the component is positively correlated with the wind speed but also negatively correlated with the temperature and the log precipitation, which corresponds to the fact that these curves have large wind speed together with a small temperature and log precipitation. We do not give the plot of the correlations for the second component, but this component is highly positively correlated with the wind speed only between day 100 and day 200. This finding explains the fact that curve 36 is closer in terms of the ICS distance from 20 and 56 in Figure 2, where the second component of ICS accounted for between day 120 and day 180 (see the left plot of Figure 1) compared with Figure 3, which has only one component (left plot). The second component is also very much positively correlated with the log precipitation during the first 100 days which explains that the green curves in Figure 23 are detected as outliers on Figure 3 when $k = 2$ at every time (middle plot). In the case of large p , the previous interpretations could become intractable, and global ICS, applied to the data next, appears as a good alternative.

Global functional ICS

We applied global ICS to the weather data and, as described in Subsection 2.3, we use now the scree plot as in Figure 6 to determine an appropriate number of invariant coordinates. There are $p \times D = 3 \times 11 = 33$ eigenvalues and we have several possible dimension choices k when looking at large differences between

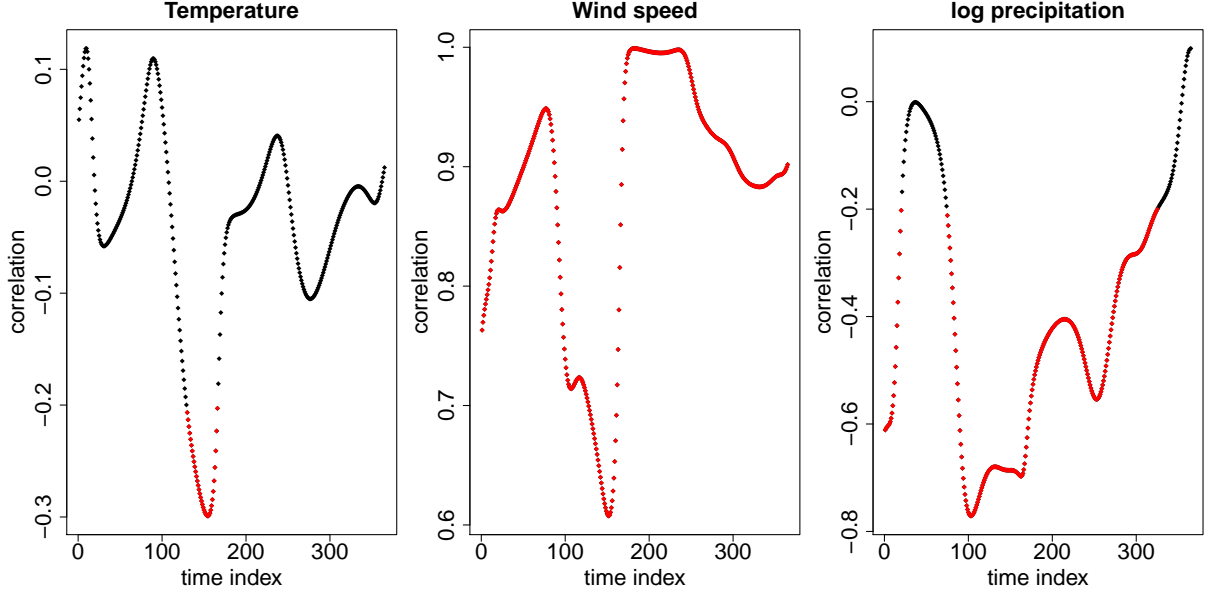


Figure 5: **Weather data set** - Point-wise ICS: correlations between the first component and the three initial variables by time. Red colour corresponds to correlation with absolute value larger than 0.20.

successive eigenvalues. We consider values for k that corresponds to a large jump between the k th and $(k + 1)$ th eigenvalues. This procedure leads us to choose $k = 2$ (squares on the plot), 3 (squares and triangle) or 4 (squares, triangle and cross) components. We then calculate and plot the ICS distance of each observation using the expression (1) for the given number of selected components. Note that we get only one ICS distance for each observation in global functional ICS. We also plot a cutoff line obtained by Monte Carlo simulations as proposed by Archimbaud et al. (2018) and previously detailed. Figure 7 illustrates the ICS distance plot for $k = 2$ (resp., $k = 3$ and $k = 4$) selected components on the left (resp., middle and right) panel. Selecting only two components leads to detecting curves 20 and 56 as outliers while selecting three (resp., four) components leads to also detecting curve 59 (resp., 59 and 36).

It is possible to interpret the outliers by looking at the correlations between the basis coefficients of each initial variable and the selected invariant coordinates. Figure 8 gives the correlations of the four components in different plots. On every plot the points correspond to the different B-spline coefficients (11 per variable) and they are grouped by variable in 3 columns, with the temperature on the left side of the first vertical green line, the wind speed in between the first and second green lines, and the log precipitation on the right of the second green line. Interestingly, we note that the correlations are quite similar for a given component among the different coefficients of the same variable. For the first three components leading to the detection of curves 20, 56 and 59, the correlation structure is globally similar and has negative correlations between the components and wind speed coefficients and low or medium positive correlations with the temperature and log precipitation coefficients. The first three plots differ when looking at the correlations with the different B-splines coefficients in detail. However it is not easy to interpret these coefficients. With regard to the correlations with the fourth component on the right bottom plot, they differ from the other plots and exhibit positive correlations with both the temperature and the wind speed, and negative correlations with the log precipitation. This finding explains that curve 36 which has a very high level of wind speed together with a large temperature is detected with $k = 4$. Note that Fig. 9 from Dai et al. (2020) plot, for the same weather data set, give joint outlying curves that are not detected using univariate methods, and among them are observations 56 (green on Fig. 9)

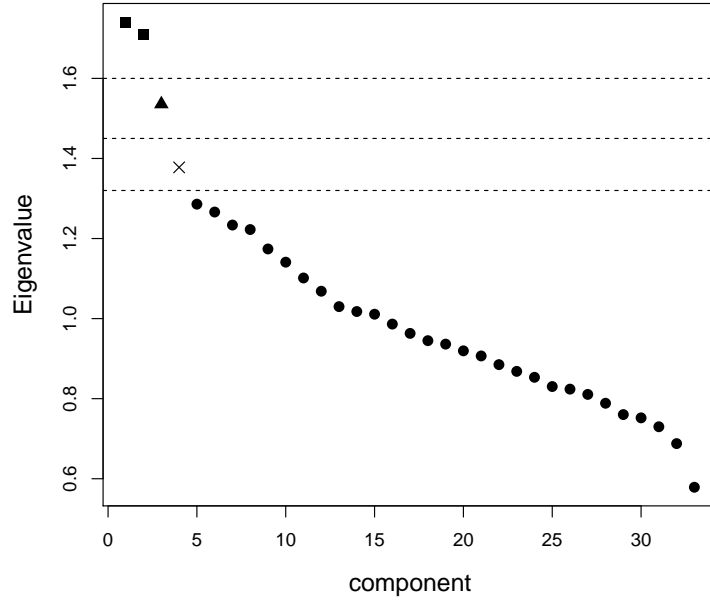


Figure 6: **Weather data set** - Global ICS: scree plot.

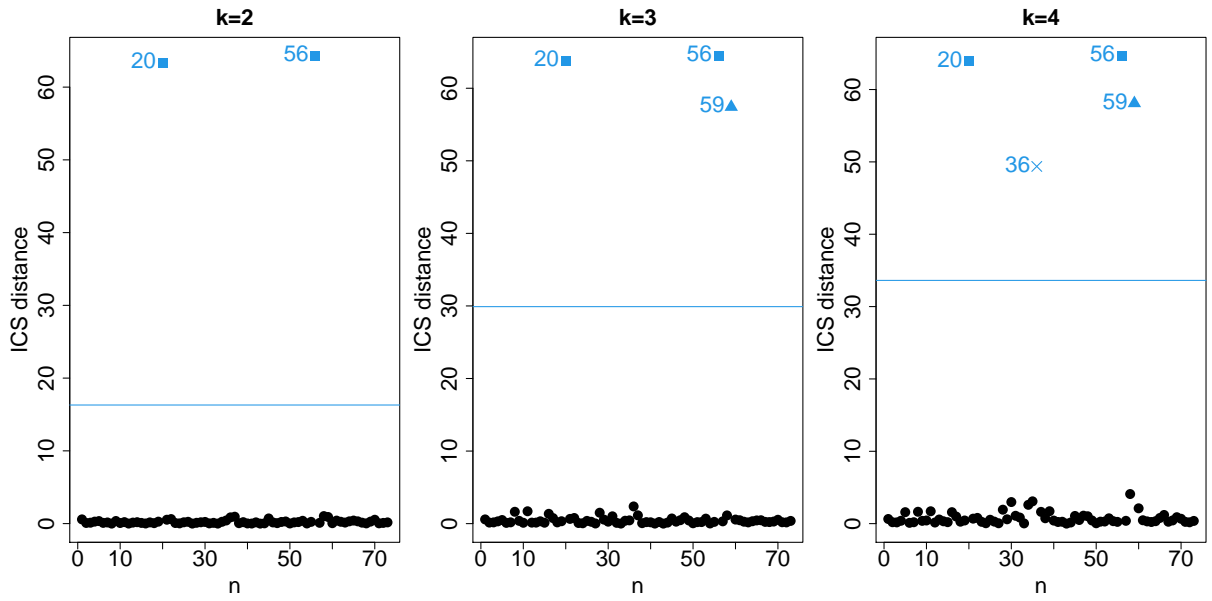


Figure 7: **Weather data set** - Global ICS: square ICS distance with $k = 2$ (resp., $k = 3$ and $k = 4$) selected components on the left (resp., center and right) panel and the cutoff quantile of order 0.975.

and 59 (purple on Fig. 9).

Each of our two functional ICS approaches has advantages and disadvantages. One disadvantage of the point-wise approach is that it does not account for the temporal dependence in the data. This limitation is not the case for the global approach which accounts for the temporal behaviour of the data through functional dimension reduction. However, global ICS depends on the choice of a functional basis, and interpreting the outlying curves using the coefficients in the basis is not easy to accomplish. An advantage of global ICS is that the domains for the functional variables do not need to be the same. Because it only involves one joint diagonalization, global ICS is much less expensive in terms of the calculation time than point-wise ICS. Additionally, global ICS uses one and only one dimension selection, which can be made using the scree plot. With regard to point-wise ICS, it allows for a graphical representation of the results in the form of curves, which is helpful to deepen the results.

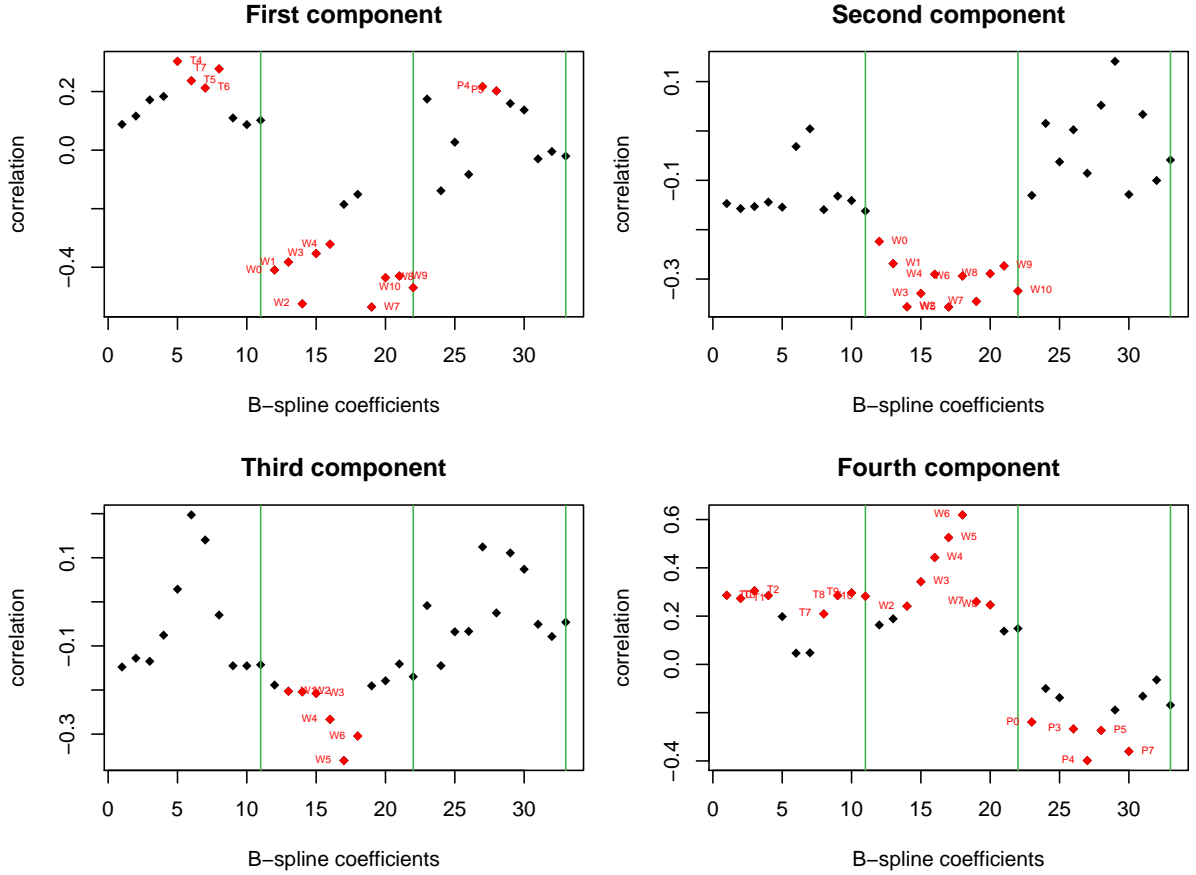


Figure 8: **Weather data set** - Global ICS: correlations between the first four components and the 11 B-spline coefficients of the three initial variables. Red colour corresponds to correlation with absolute value larger than 0.20. The green lines are separators between the features Temperature, Wind speed and log precipitation.

3 Data analysis and simulation

In the next two subsections, we consider two real data sets that come from the fields of aeronautics and microelectronics manufacturing. In the third subsection, we give results on a small simulation study and assess the performance of the ICS methods compared to other multivariate functional outlier detection methods. The first example concerns the predictive maintenance of an aeronautical electrical generator before its failure. The second example concerns a quality control process in a semiconductor microelectronics fabrication. The first data set is completely confidential, while the second one is available online. For both examples, we work under the assumption that only a small proportion of the observations is expected to be diagnosed as anomalies (no more than 2%), as is usually the case in areas where quality is of a high standard. Moreover, for each of these two data sets, we have prior information on some particular observations that we would like to detect.

The goal is to compare our two functional ICS approaches to detect outlying curves. We also include in the comparison the directional outlier detection method proposed by Rousseeuw et al. (2018) using a point-wise Directional Outlyingness (DO) measure. Both point-wise ICS and the method by Rousseeuw et al. (2018) make use of the functional outlier map (FOM) tool. For point-wise ICS, the weights take into account the number of selected dimensions at each time point while, for DO, the weights are uniform. We use the *fOutl* and the *fom* functions available in the R package *mrfDepth* (Segaert et al., 2020) to implement the method in Rousseeuw et al. (2018). The *fom* function is customized to be able to change the colours and the cutoff quantile order. From our experience on these two data sets, but also on dozens of other data files, the default quantile level 0.95 for the FOM leads to an excessive number of outliers and has to be adjusted. Depending on the field of interest one can, for example, use the simple 1-2% rule, meaning that we do not flag more than 1% to 2% of observations as outliers. On the other hand, it is usually worthwhile to choose the cutoff by taking into account also the plot characteristics. If the FOM exhibits some rare (less than 2%) observations clearly separated from the main bulk of the data, it is recommended to flag these observations and only these ones as outliers. In the two examples below the cutoffs are such that the observations far from the main bulk of the data are flagged as outliers while keeping the proportion of outliers lower than 2%.

As functional expansion used in global ICS, we consider the Fourier basis but note that the results obtained with a B-spline basis are similar. Concerning the choice of the dimension D , based on our experience of using global ICS on many real examples, the number of observations n and the number of features p are to be taken into account. The number of observations per dimension $p \times D$ should not be smaller than 10. This criterion gives the rule of thumb that D should be smaller than $n/(10p)$, implying that, regardless of the value of D , the number of observations n should always be larger than $10p$. In the main body of the text, we focus on particular D values in $\{5, 11, 15\}$ using the recommended rule of thumb. Comparisons with results obtained using the other values of D for the quality control example are given in the appendix.

3.1 Predictive maintenance of an aeronautical electrical generator

Our first application uses aircraft flight data. In the aeronautics literature, several recent papers tackle the problem of anomaly detection on flight data even if the literature is still rather sparse. Most articles consider a multivariate time series framework (see Li et al. (2015), Li et al. (2016b), Memarzadeh et al. (2020) and the references therein). One exception is the paper by Jarry et al. (2020) which considers PCA in a univariate functional framework together with a clustering method, to detect atypical approaches using landing radar records.

Our application concerns the monitoring of aircraft electrical generators during routine flights. Electrical generator failures lead to delays or cancellations of flights which can be extremely costly to the manufacturer and the airline. To reduce this cost, electrical design engineers are willing to detect abnormal generator behaviour before it turns into a failure. The goal is to raise a warning when successive

Table 1: Recorded features by flight to detect abnormal electrical generator behaviour.

Notation	Description	Unit
X^1	Generator oil temperature	C°
X^2	Engine speed	Knots (kts)
X^3	Generator load	KVA
X^4	Static air temperature	C°
X^5	Computed air speed	kts
X^6	Altitude	ft

flights are detected as outliers and to suggest performing a maintenance action on the generator. Such a process is called predictive maintenance. We consider $n = 590$ flights of varying duration from a given aircraft and a given generator. What makes this data set particularly interesting is that during the 591st flight, which followed the 590 flights for which the data are available, an electrical generator failed. The objective with this example is to figure out whether functional ICS could have detected outlying generator behavior in advance. For each flight, we observe $p = 6$ features with a sampling rate of one record per second. These features are identified as relevant by electrical engineers and are detailed in Table 1. To be able to apply the selected approaches, we must first align the flights to obtain equal numbers T of time points. To account for the different flight phases, which correspond to different electrical behaviours, we split each flight into takeoff, cruise and landing phases. Each flight phase is aligned separately, and the whole flights are rebuilt afterwards. In the present example, the flights are aligned on the duration of $T = 2900$ seconds. In Figure 24 of the appendix, we plot the aligned flights by the features for the engine speed, the static air temperature, the computed air speed and the altitude. For confidentiality reasons, we cannot plot the generator oil temperature and the generator load curves but we use the data in the analysis. The reason why some flights are coloured in red will be clarified later. Given that the size of the sample is $n = 590$, and considering that the maximum percentage of outliers is between 1 and 2%, the maximum number of flights detected as outliers cannot exceed 12. Let us now implement and compare the two functional ICS approaches on this real data set.

First, we apply global ICS on the aligned flights using $D = 5$ since the rule of thumb $D < n/(10p)$ leads to $D < 9$. With $D = 5$, we have a data set with dimension 590×30 . Global ICS consists of computing the squared distances using the first k invariant components. To select the number of components, we use the scree plot (see Figure 9) and consider the values of k that correspond to large jumps between the k th and $(k + 1)$ th eigenvalues. In this example, the large jumps correspond to 3 different possibilities for k , namely $k = 1$ (square symbol), 2 (triangle) and 4 (diamond). For each value of k , we compute the squared ICS distances by flight and the ICS cutoff using the Monte Carlo calculation with $m = 100$ replications and a level $\gamma = 0.025$ (function *dist.simu.test* from the R package *ICSOutlier* (Archimbaud et al., 2018b)). The method simulates squared ICS distances using the first k invariant components from a Gaussian population whose dimension is that of our data set. Then, the cut-off corresponds to the mean $(1 - \gamma)$ -quantile over these m replicates. The results are given in Figure 10, where the flights are ordered by date, and we plot a black dashed vertical line that represents the flight that precedes the generator loss (number 590). The symbols of the flights that are detected as abnormal are the same as the symbols used in the scree plot (square when detected on the first component, triangle on the second and diamond on the third and fourth components). With $k = 4$, we have already too many outliers compared to the 2% limit and thus we should consider $k = 1$ or $k = 2$. When $k = 1$, only the flights 587 to 590 are detected as outliers while for $k = 2$, more flights (11, 17, 53, 66, 260 and 456) are detected. However, even with $k = 2$, the only successive flights detected are the four that precede the generator loss. Looking at the correlations between the first invariant component and the 5 Fourier coefficients of

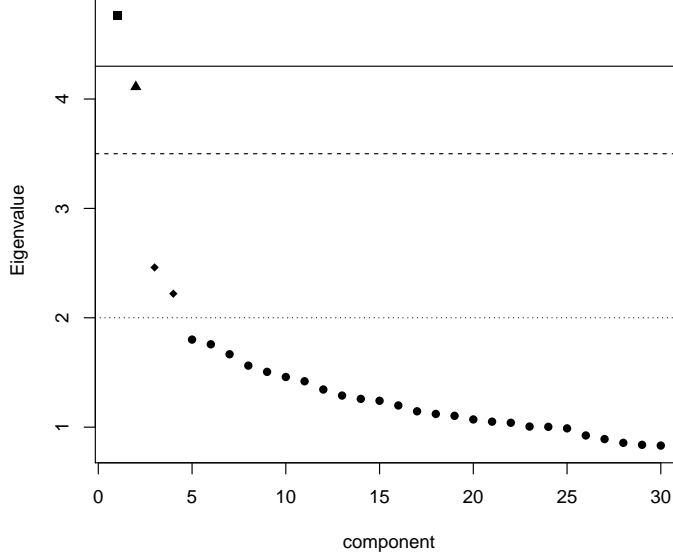


Figure 9: **Aeronautical data set** - Global ICS: scree plot.

the engine speed feature in Figure 25, we can see that the four flights differ from the others because of the engine speed variable. Figure 24 where flights 587 to 590 have been plotted in red helps to confirm our interpretation of the outlyingness of the flights. In practice, it is not possible to monitor each feature separately, and the procedure we propose gives an automatic way to detect abnormal behaviors preceding a failure of an electrical generator by taking into account many features.

Next, we apply point-wise functional ICS on the aligned curves and use the automatic component selection procedure described previously as in Nordhausen et al. (2017). The number of selected components by time (see Figure 11) varies from two to six and is higher during the cruise period than during take-off and landing. Using fICS and vICS, the FOM is plotted on the left panel of Figure 12. The cutoff curve has been adjusted to detect observations that clearly differ from the main bulk of the data in terms of fICS. It can be seen that flights 587 to 590 are detected together with six other flights (11, 17, 34, 53, 308 and 456). Of these, four flights (11, 17, 53 and 456) were also detected by global ICS with $k = 2$ (see the middle panel of Figure 10). The squared ICS distances are plotted in Figure 13 with the same colour code as in Figure 12 (red for the four successive outliers and blue for the additional outliers detected by point-wise ICS). It can be seen that the red curves differ from the other curves at the end of the flights while the blue curves are outlying during the cruise period. However, it is difficult to analyse the reasons for the outlyingness of the red and blue flights by looking at the correlation curves because of the high number of initial features and the large and highly variable numbers of components. To visualize the flights detected as anomalies by global and point-wise ICS, and to interpret their outlyingness, we plot the original curves with different colours in Figure 14 (see the caption for details on the colours). It can be seen that all of the flights that are detected as outliers either by global or point-wise ICS exhibit a different behaviour compared to the other flights. Global and point-wise ICS do not give exactly the same results but have in common many curves detected as outliers. We now compare the ICS results with the FOM obtained using the DO index as defined in Rousseeuw et al. (2018). Using the default options leads to very poor results (see Figure 26); as a result, we tried out other options (available but

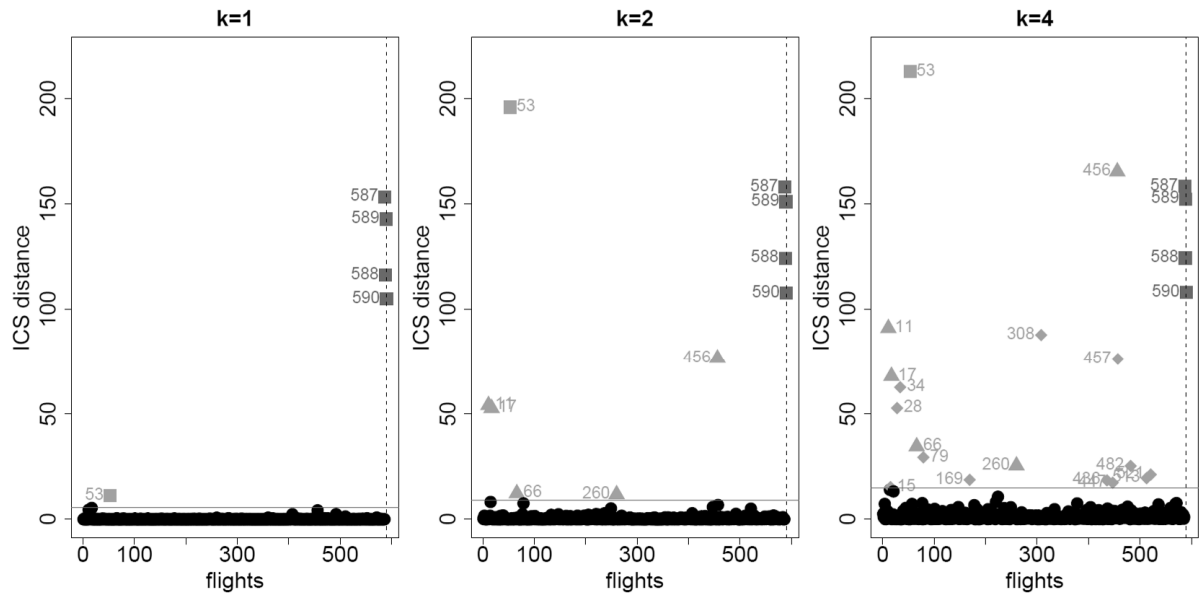


Figure 10: **Aeronautical data set** - Global ICS: square ICS distance with $k = 1$ (resp., $k = 2$ and $k = 4$) selected components on the left (resp., center and right) panel and the cutoff quantile of order 0.975.

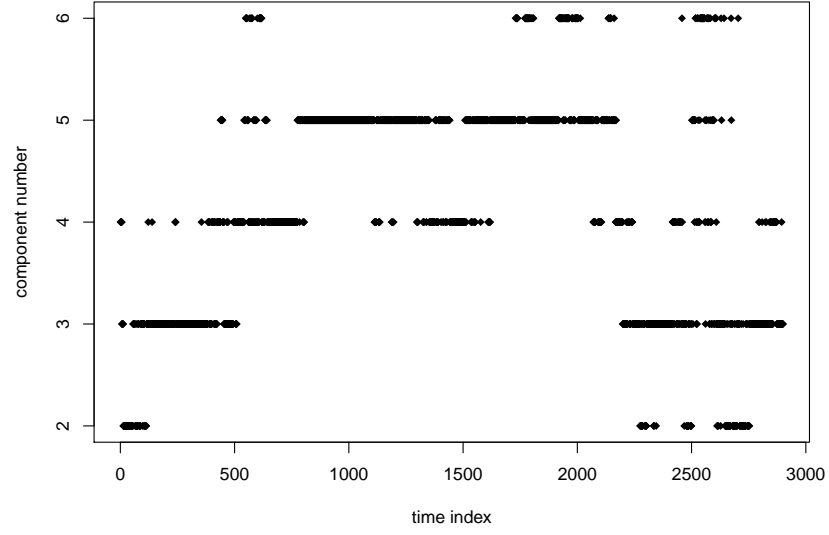


Figure 11: **Aeronautical data set** - Point-wise IC: number of selected components.

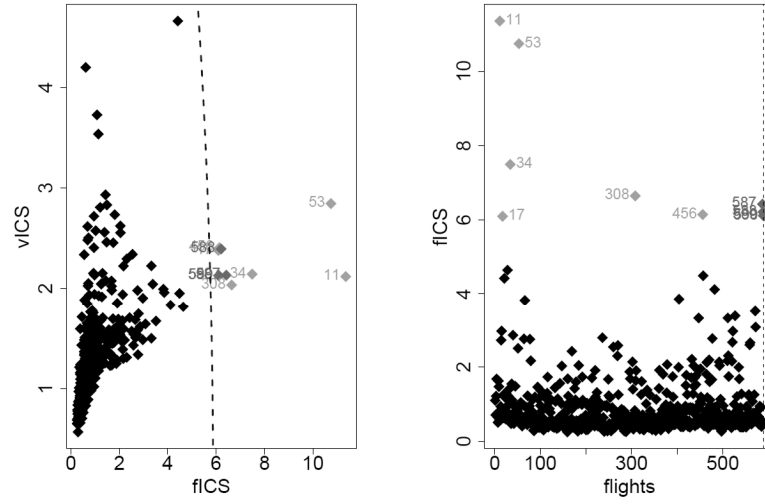


Figure 12: **Aeronautical data set** - Point-wise ICS, Left: FOM, Right: fICS.

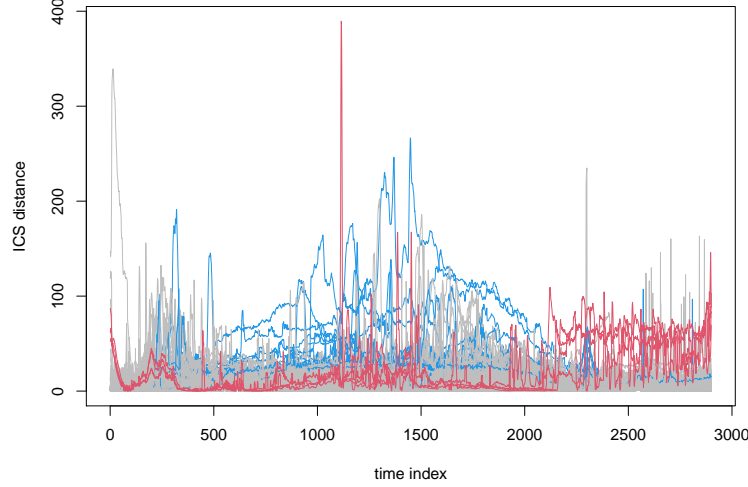


Figure 13: **Aeronautical data set** - Point-wise ICS: square ICS distances with an automatic components selection. Normal flights are coloured in grey whereas the four successive flights detected as outliers are coloured in red and the isolated flights detected as outliers are in blue.

not detailed in the documentation) that lead to Figure 15, which permits us to detect the flights 587 to 590 that precede the failure only. However, the method is quite unsatisfactory because for almost half of the time points (46%), an exact fit is detected and it is not possible to calculate the outlyingness index. Figure 27 in the appendix gives the time points at which it is not possible to calculate the DO index and they correspond to the cruise part of the flights. This finding explains why the method does not detect any of the curves 11, 17, 34, 53, 66, 260, 308 and 456 which are outlying precisely during the cruise period. Thus, we conclude that the results obtained by the DO method are only good by chance and it is worrying that the method cannot be applied for almost half of the time points.

3.2 Semiconductor quality control application

Anomaly detection is crucial in quality control and especially in semiconductor microelectronics, which has safety-critical applications, such as automotive electronics, medical devices, and aerospace systems. Archimbaud (2018) gives a review of the common unsupervised methods that are used in practice together with their implementations in R software. It appears that only a few multivariate methods, such as Mahalanobis distance or Principal Component Analysis, are used by manufacturers. Some recently published articles in the field of industrial process monitoring consider a multivariate functional framework (see for example Liu et al. (2020) and the references therein).

The database that we consider is available online: <https://www.cs.cmu.edu/~bobski/data/data.html>. It contains a collection of sequences of measurements (or runs) that are recorded by one vacuum-chamber sensor during the etch process applied to one silicon wafer during the manufacture of semiconductor microelectronics. An etch process is a complex process during which layers of various materials are applied to a silicon wafer and selectively removed to define circuit elements on the wafer. Each run, among the 1194 runs, has an assigned classification of normal or abnormal and 127 runs are flagged as abnormal. This number corresponds to 10.6% of the observations which is much larger than the 2%

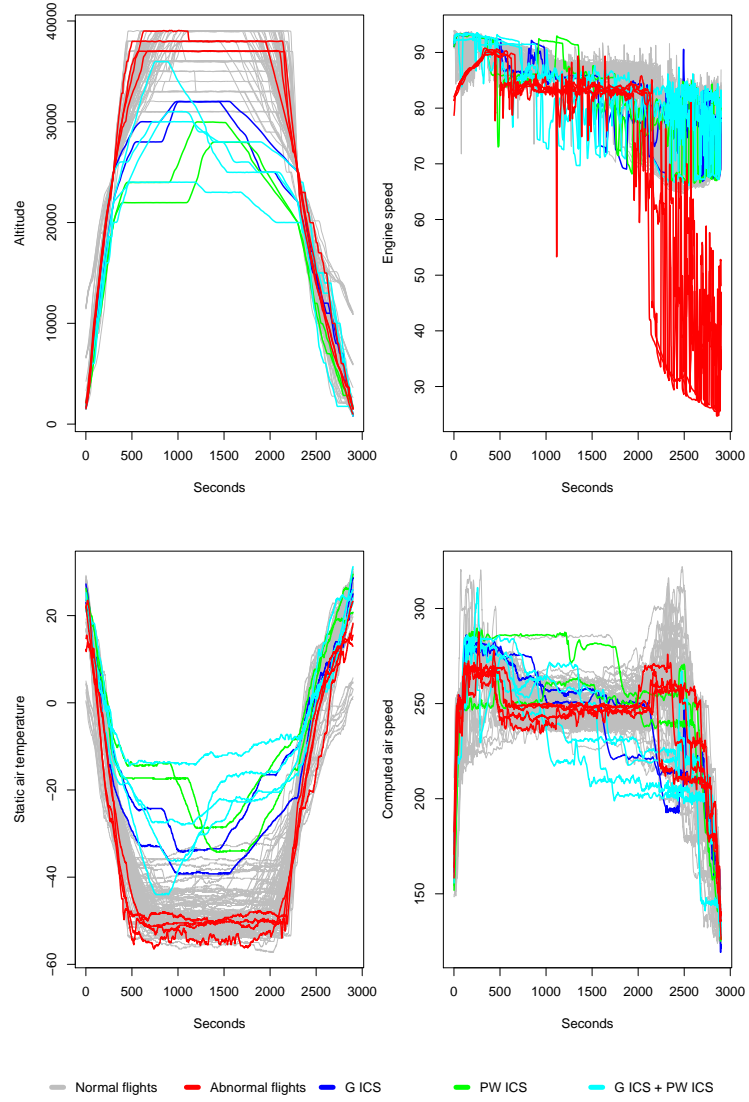


Figure 14: **Aeronautical data set** - Detected outliers. Normal flights are coloured in grey whereas the four successive flights detected as outliers are coloured in red. The curves 66 and 260 detected as outliers only by global ICS with $k = 2$ (G ICS) are in blue, the curves 34 and 308 detected as outliers only by point-wise ICS (PW ICS) with an automatic component selection are in green, while the curves 11, 17, 53 and 456 detected as outliers by both G ICS and PW ICS are in cyan.

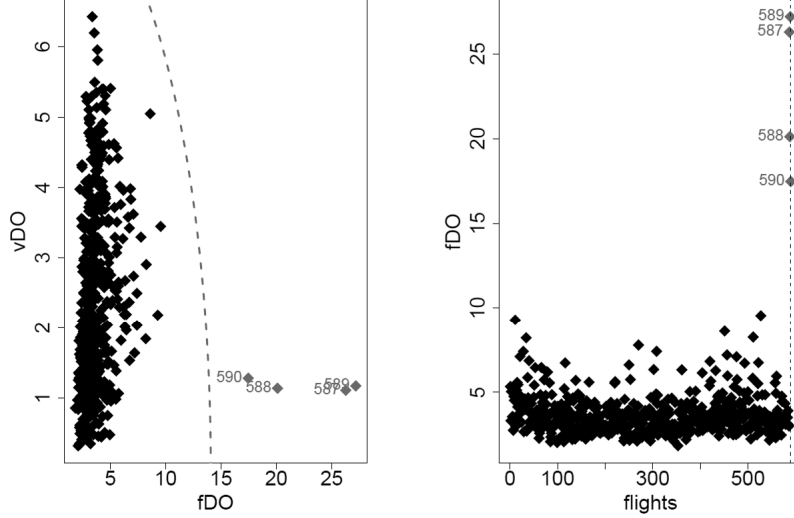


Figure 15: **Aeronautical data set** - DO method using the R package mrfDepth (with distOptions = list(rmZeroes = TRUE,maxRatio = 3) in the fOutl function). Left: FOM, Right: fDO by flight.

Table 2: Sensor descriptions for the semiconductor manufacturing data set.

Notation	Sensor description
Sensor 6	Radio frequency forward power sensor
Sensor 7	Radio frequency reflected power sensor
Sensor 8	Chamber pressure sensor
Sensor 11	405 nm emission sensor
Sensor 12	520 nm emission sensor
Sensor 15	Direct current bias sensor

we are interested in. Among the 127 abnormal runs, four runs are severe or very severe faults which correspond to 0.3% of the observations. The objective of our analysis is to check whether the functional ICS methods can detect these four runs as outliers. Six sensors have been identified by domain experts as being critical for monitoring purposes (see Olszewski (2001), p. 68-69, for more details on the sensors). Description of the $p = 6$ sensors is given in Table 2. We have $n = 1194$ runs with a duration that varies from 104 to 198 records. The runs are aligned by linear interpolation to a duration of $T = 150$ records. The aligned runs are plotted in Figure 28 for each sensor. The four runs with severe and very severe faults are 73, 107, 317 and 351 and are highlighted in red.

For global ICS, a dimension reduction is applied on the aligned runs for each sensor using the Fourier basis, with the number of basis functions equal to $D = 11$. The appendix gives the results for other values of D (see Figure 29), but applying the rule of thumb $D < n/(10p)$ leads to $D < 19$, and thus, we focus on $D = 11$. The global ICS results on this 1194×66 data set are given in Figures 16 and 17. The scree plot in Figure 16 shows that we can consider $k = 1, 3, 6$ or 7 , and we use different symbols to differentiate between the eigenvalues (square for the first two eigenvalues, triangle for the third, diamond

for the fourth to the sixth and cross for the seventh). Figure 17 gives the squared ICS distances and the ICS cutoff (red line) for each of the four values of k . To compute the ICS cutoffs we use again the function `dist.simu.test` with a quantile of order $1 - \gamma = 0.975$. The symbols correspond to the ones used in Figure 16. A value $k = 6$ or 7 leads to the detection of a dozen observations which corresponds to approximately 1% of the observations. Such a choice leads to the detection of the four runs with severe or very severe faults together with some other runs.

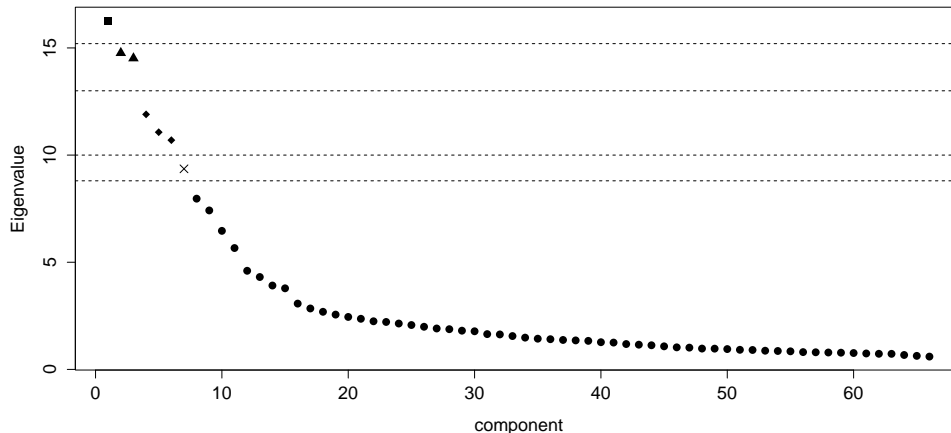


Figure 16: **Semiconductor data set** - Global ICS: scree plot.

For point-wise functional ICS, we apply $T = 150$ ordinary ICS with $n = 590$ and $p = 6$. For each ordinary ICS, we select the number of invariant components automatically as detailed in Section 2.2 and the corresponding plot is given in Figure 18. Figure 19 gives the FOM with fICS weighted by the number of components selected at each time point. If we tolerate a detection rate of 1%, we can use the weighted fICS and detect the four runs with severe and very severe outliers with six other runs. We note that the runs 122, 964, 311 and 326 are detected by both functional ICSs but that some additional observations are detected by only one of the methods. Note also that if we ignore the run 162 for global ICS and 489 for point-wise ICS which are normal runs, all the runs declared as outliers are abnormal runs (but with no severe or very severe faults). We also give the FOM for the Directional Outlyingness index proposed by Rousseeuw et al. (2018) in Figure 20. Once again, some exact fit problems do not allow us to calculate the DO index 50% of the time (see Figure 31 in the appendix for a plot of the time points, where the calculation is not possible). This finding could explain the fact that among the 2% of observations (24 runs) that have the largest DO index, only one run (number 317) has severe or very severe faults. The other 3 runs with severe or very severe faults (73, 107 and 351) are not detected among the 2% of outliers.

We conclude this section with a small simulation study to assess the performance of the ICS extensions and to compare the new proposals with other existing methods.

3.3 Simulation

We consider two different setups. We first follow the simulation framework of Subsection 5.2 from Dai and Genton (2019) with 100 curves generated at 50 equidistant time points on $[0, 1]$, conducting a total of 500 replications of this setting. We consider bivariate functional models ($p = 2$) as in Dai and Genton (2019). The uncontaminated model (Model 5 in Dai and Genton, 2019) is a bivariate Gaussian process with a cross-covariance function from the Matérn class, while the contaminated models (Model 6 to 9 in

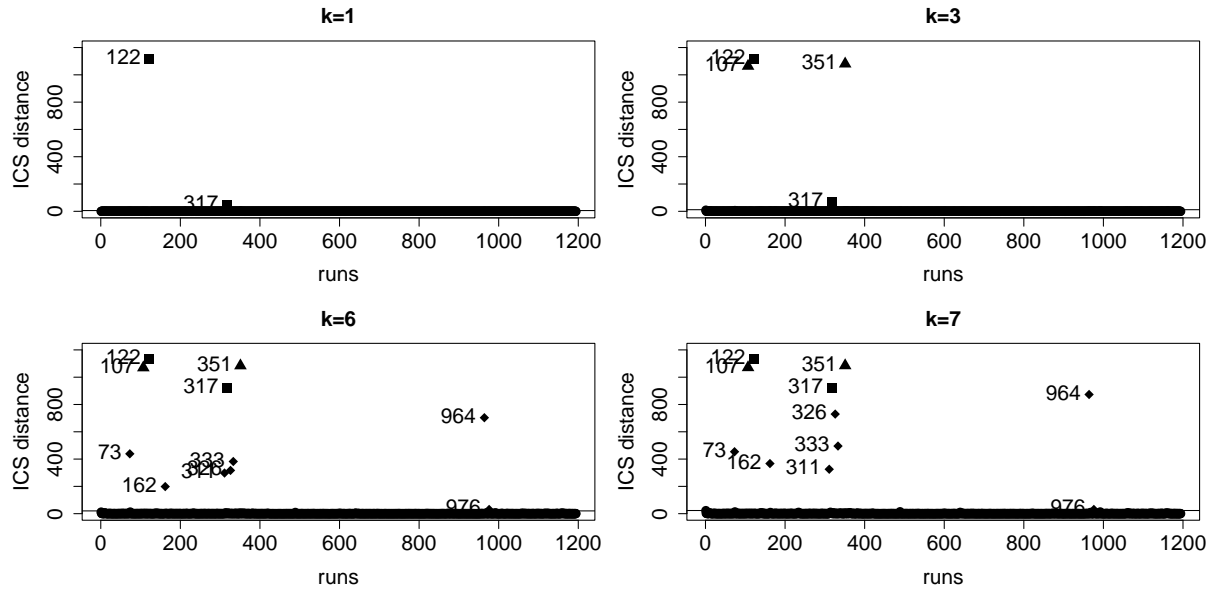


Figure 17: **Semiconductor data set** - Global ICS: square ICS distance for $k = 1, 3, 6$ and 7 , using the cutoff quantile of order 0.975.

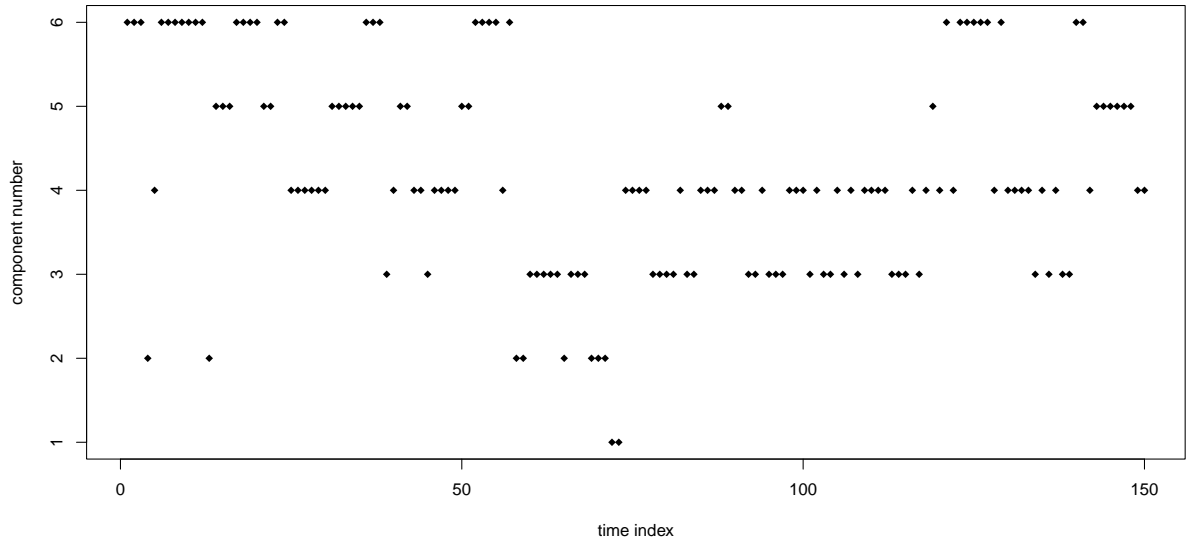


Figure 18: **Semiconductor data set** - Point-wise ICS: number of selected components.

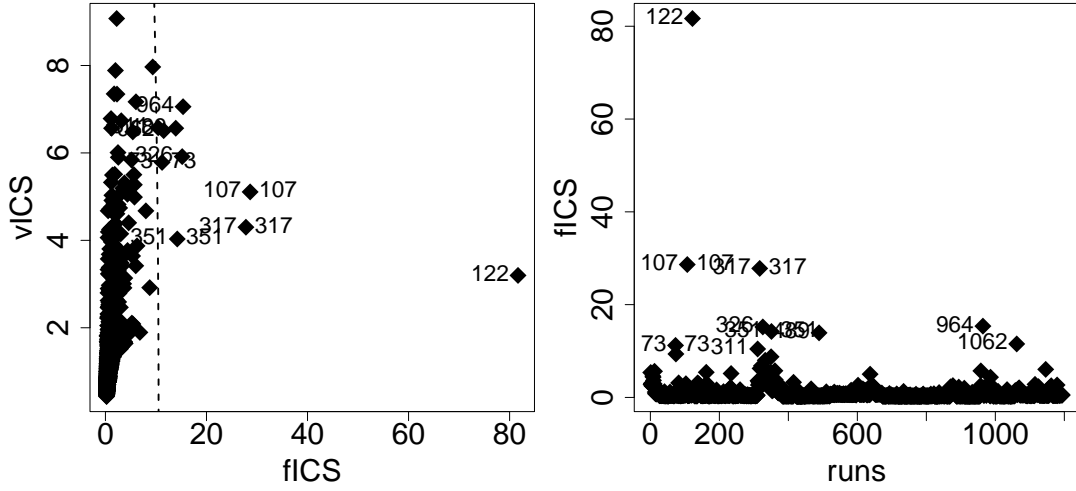


Figure 19: **Semiconductor data set** - Point-wise ICS, Left: FOM, Right: fICS.

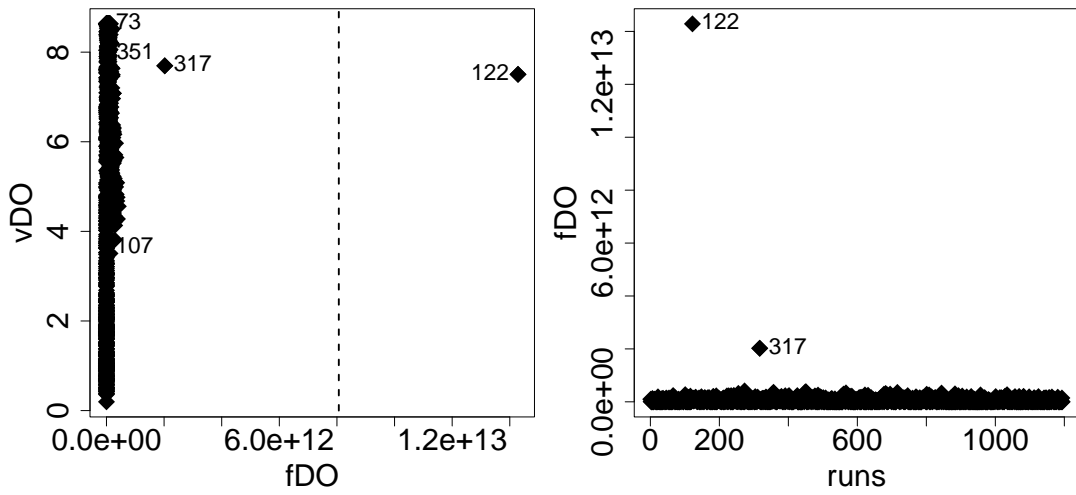


Figure 20: **Semiconductor data set** - DO method, Left: FOM, Right: fDO.

Table 3: Average correct outlier detection rate (p_c) and false outlier detection rate (p_f) for the four models under consideration together with their standard deviations based on 500 repetitions for different data dimensions p .

Model	p	Model 6		Model 7	
		p_c	p_f	p_c	p_f
G ICS	2	86.9 (3.2)	0.0 (0.0)	56.8 (4.1)	0.0 (0.1)
	20	99.0 (1.4)	0.0 (0.0)	71.5 (0.7)	0.0 (0.0)
PW ICS	2	100.0 (0.0)	0.6 (0.9)	100.0 (0.0)	0.8 (1.1)
	20	100.0 (0.0)	0.6 (0.2)	100.0 (0.0)	0.6 (0.1)
Tot.Dir.Out	2	100.0 (0.0)	0.0 (0.1)	100.0 (0.0)	0.0 (0.1)
	20	100.0 (0.0)	0.0 (0.0)	100.0 (0.0)	0.0 (0.0)
Model	p	Model 8		Model 9	
		p_c	p_f	p_c	p_f
G ICS	2	21.8 (16.1)	0.9 (1.0)	33.6 (13.8)	0.3 (0.7)
	20	31.6 (5.7)	2.2 (0.5)	31.1 (4.9)	1.6 (0.5)
PW ICS	2	68.1 (18.3)	0.9 (1.1)	98.7 (3.8)	0.8 (1.1)
	20	44.0 (24.0)	0.4 (0.2)	99.5 (0.7)	0.7 (0.6)
Tot.Dir.Out	2	82.3 (13.3)	0.0 (0.0)	91.9 (8.5)	0.0 (0.1)
	20	10.0 (2.8)	0.0 (0.0)	64.0 (5.7)	0.1 (0.0)

Dai and Genton, 2019) are obtained by replacing 10% of the observations of the uncontaminated model by some outliers. For Models 6 and 8, the outliers are obtained by multiplying the Gaussian process by a constant larger than one over the whole time period. The constant is equal to 4 for Model 6 and 1.7 for Model 8. For Models 7 and 9, the outliers are also obtained by multiplying the Gaussian process by a constant larger than one but only on a part of the time period (10% randomly chosen on $[0, 1]$). The constant is equal to 12 for Model 7 and 5 for Model 9. Following Dai and Genton (2019), we calculate the correct outlier detection rate and the false outlier detection rate over the 500 simulations to assess the properties of the outlier detection methods.

In Dai and Genton (2019), several multivariate functional outlier detection methods are compared, and the best results are clearly obtained for the directional outlyingness method (denoted by Tot.Dir.Out). Thus, we compare the global and point-wise ICS to Tot.Dir.Out only. To go beyond the bivariate case, we consider a second setup of data sets with $n = 1000$ curves and $p = 20$ dimensions. For this new setup, the first two dimensions are generated exactly as before using the different uncontaminated and contaminated models, while the 18 additional dimensions are generated using 9 times the uncontaminated bivariate model (Model 5 in Dai and Genton, 2019). For both dimensions ($p = 2$ and 20), we compare Tot.Dir.Out with global and point-wise ICS (denoted respectively by G ICS and PW ICS), and the results are reported in Table 3 for Models 6 to 9. The results for the uncontaminated model are not reported because the three methods give very similar results with less than 1% of false positives. For both ICS methods, the choice of the number of components is based on the test given by Nordhausen et al. (2017) with an initial level of 10%. For PW ICS, the observations are flagged as outliers using equation (4) and $\alpha = 95\%$. For G ICS, we use the Fourier basis with a number of basis functions equal to $D = 5$ which corresponds to $D = n/(10p)$ in both setups ($n = 100, p = 2$ and $n = 1000, p = 20$). The cutoff value for the ICS distances (see equation (1)) is obtained through Monte Carlo simulations with a 5% level.

For the bivariate setup, the results obtained by PW ICS in terms of true positives are comparable to Tot.Dir.Out but at the cost of a small increase in the number of false positives (less than 1% however). The results are not as good for G ICS in terms of true positives for Model 8 and 9 where outliers do

not differ a lot from the rest of the data, especially for Model 8. Looking at the results for $p = 20$, we see that PW ICS and Tot.Dir.Out are still comparable for Model 6 and 7 in terms of true positives. However, the results are much better for PW ICS than for Tot.Dir.Out when looking at Model 8 and 9. For Model 8, even G ICS gives better results than Tot.Dir.Out in terms of true detection with a small deterioration in terms of false detection. Moreover, note that Tot.Dir.Out uses a projection depth based on the Stahel-Donoho outlyingness measure (see Dai and Genton, 2019). This measure is based on a projection pursuit algorithm which is costly in large dimensions and only approximate.

Overall, the ICS extensions and especially the point-wise method prove to be attractive compared to other methods, especially in a large dimensional setup and with no need for an approximate algorithm.

4 Conclusions and perspectives

Given the complexity of outlier detection in the framework of multivariate functional data, it appears to be unrealistic to find a method that works in all situations. The methodology we propose is particularly suitable when there is a small number of observations likely to be real anomalies but that may not be identified when looking at univariate functional characteristics only. In this context, we propose two generalizations of ICS to multivariate functional data with many diagnostic plots to help the data analyst in the detection and interpretation of outlying curves. Both approaches have advantages and disadvantages. Compared to global ICS, point-wise ICS does not take into account the time dependence, is more costly in terms of computation and does not allow simple interpretation of the outliers. Nevertheless, point-wise ICS depends neither on the choice of the functional basis nor on the choice of the number of elements in this basis. Moreover, the constraint that the number of observations has to be larger than ten times the dimension is not required for point-wise ICS. Finally, the results from our small simulation study indicate that point-wise ICS can achieve better performance than global ICS. In the context of a small proportion of outliers, it is important to have detection methods that lead to the identification of only a few anomalies with the possibility of understanding their anomalous behaviour. Functional ICS, either global or point-wise, with the covariance matrix and the matrix of fourth moments as the scatter pair, is particularly suitable for such a context.

Regarding the future research perspectives for the global ICS method, it is worth investigating the impact of the choice of the functional basis and the number of basis coefficients, possibly separately for each feature. Moreover, even if global ICS can be applied with a non-orthonormal basis, it is relevant to study the consequences of such a choice on the procedure. In particular, the expressions of the $\text{COV} - \text{COV}_4$ scatter pair should be adapted as suggested by Virta et al. (2020). Let us also mention the case of curves measured on different domains (Happ and Greven, 2018) such as images (Rousseeuw et al., 2018) for which the global ICS could be extended. Further work on the automatic selection of invariant components using a multiple testing approach, instead of the scree plot, is also of interest.

5 Acknowledgements

We would like to thank an associate editor and two reviewers for their suggestions and comments that helped us a lot to improve the paper. This work was partly supported by the French *Agence Nationale de la Recherche* through CIFRE contract 2017/1354 and through the Investments for the Future (Investissements d’Avenir) program, grant ANR-17-EURE-0010, the Academy of Finland (Grant 335077) and by a grant of the Dutch Research Council (NWO, research program Vidi, project number VI.Vidi.195.141). The authors are grateful to Camille Girou for some preliminary work on the semiconductor data set and to Martina Salvignol for her help on the aeronautical data set.

References

- Archimbaud, A. (2018). Détection non-supervisée d’observations atypiques en contrôle de qualité: un survol. *Journal de la Société Française de Statistique*, 159(3):1–39.
- Archimbaud, A., May, J., Nordhausen, K., and Ruiz-Gazen, A. (2018). *ICSShiny: ICS via a Shiny Application*. R package version 0.5.
- Archimbaud, A., Nordhausen, K., and Ruiz-Gazen, A. (2018). ICS for multivariate outlier detection with application to quality control. *Computational Statistics & Data Analysis*, 128:184–199.
- Archimbaud, A., Nordhausen, K., and Ruiz-Gazen, A. (2018a). *ICSOutlier: Outlier Detection Using Invariant Coordinate Selection*. R package version 0.3-0.
- Archimbaud, A., Nordhausen, K., and Ruiz-Gazen, A. (2018b). Unsupervised outlier detection with ICSOutlier. *The R Journal*, 10(1):234–250.
- Barreyre, C., Boussouf, L., Cabon, B., Laurent, B., and Loubes, J.-M. (2019). Statistical methods for outlier detection in space telemetries. In *Space Operations: Inspiring Humankind’s Future*, pages 513–547. Springer.
- Dai, W. and Genton, M. G. (2018). Multivariate functional data visualization and outlier detection. *Journal of Computational and Graphical Statistics*, 27(4):923–934.
- Dai, W. and Genton, M. G. (2019). Directional outlyingness for multivariate functional data. *Computational Statistics & Data Analysis*, 131:50–65.
- Dai, W., Mrkvička, T., Sun, Y., and Genton, M. G. (2020). Functional outlier detection and taxonomy by sequential transformations. *Computational Statistics & Data Analysis*, 149:106960.
- Erbas, B., Hyndman, R. J., and Gertig, D. M. (2007). Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine*, 26(2):458–470.
- Febrero-Bande, M. and de la Fuente, M. O. (2012). Statistical computing in functional data analysis: The R package *fda.usc*. *Journal of statistical Software*, 51:1–28.
- Ferraty, F., Kneip, A., Kokoszka, P., and Petersen, A. (2022). 2nd special issue on functional data analysis. *Econometrics and Statistics*, 21:112–113.
- Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522):649–659.
- Harris, T., Tucker, J. D., Li, B., and Shand, L. (2021). Elastic depths for detecting shape anomalies in functional data. *Technometrics*, 63:466–476.
- Hubert, M., Rousseeuw, P. J., and Segaert, P. (2015). Multivariate functional outlier detection. *Statistical Methods & Applications*, 24(2):177–202.
- Ilmonen, P., Nevalainen, J., and Oja, H. (2010). Characteristics of multivariate distributions and the invariant coordinate system. *Statistics & Probability Letters*, 80(23):1844–1853.
- Jarry, G., Delahaye, D., Nicol, F., and Feron, E. (2020). Aircraft atypical approach detection using functional principal component analysis. *Journal of Air Transport Management*, 84:101787.
- Kokoszka, P., Oja, H., Park, B., and Sangalli, L. (2017). Special issue on functional data analysis. *Econometrics and Statistics*, 1:99–100.

- Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press, Boca Raton.
- Kuhnt, S. and Rehage, A. (2016). An angle-based multivariate functional pseudo-depth for shape outlier detection. *Journal of Multivariate Analysis*, 146:325–340.
- Lejeune, C., Mothe, J., Soubki, A., and Teste, O. (2020). Shape-based outlier detection in multivariate functional data. *Knowledge-Based Systems*, 198:105960.
- Li, B., Van Bever, G., Oja, H., Sabolová, R., and Critchley, F. (2016a). Functional independent component analysis: an extension of the fourth-order blind identification. Technical Report 16/02, The Open University, School of Mathematics and Statistics.
- Li, L., Das, S., John Hansman, R., Palacios, R., and Srivastava, A. N. (2015). Analysis of flight data using clustering techniques for detecting abnormal operations. *Journal of Aerospace Information Systems*, 12(9):587–598.
- Li, L., Hansman, R. J., Palacios, R., and Welsch, R. (2016b). Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring. *Transportation Research Part C: Emerging Technologies*, 64:45–57.
- Liu, J., Chen, J., and Wang, D. (2020). Wavelet functional principal component analysis for batch process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 196:103897.
- Memarzadeh, M., Matthews, B., and Avrekh, I. (2020). Unsupervised anomaly detection in flight data using convolutional variational auto-encoder. *Aerospace*, 7(8):115.
- Miettinen, J., Taskinen, S., Nordhausen, K., and Oja, H. (2015). Fourth moments and independent component analysis. *Statistical Science*, 30(3):372–390.
- Millán-Roures, L., Epifanio, I., and Martínez, V. (2018). Detection of anomalies in water networks by functional data analysis. *Mathematical Problems in Engineering*, 2018.
- Nagy, S., Gijbels, I., and Hlubinka, D. (2017). Depth-based recognition of shape outlying functions. *Journal of Computational and Graphical Statistics*, 26(4):883–893.
- Nordhausen, K., Oja, H., and Tyler, D. E. (2022). Asymptotic and bootstrap tests for subspace dimension. *Journal of Multivariate Analysis*, 188:104830.
- Nordhausen, K., Oja, H., Tyler, D. E., and Virta, J. (2017). Asymptotic and bootstrap tests for the dimension of the non-Gaussian subspace. *IEEE Signal Processing Letters*, 24(6):887–891.
- Nordhausen, K. and Ruiz-Gazen, A. (2022). On the usage of joint diagonalization in multivariate statistics. *Journal of Multivariate Analysis*, 188:104844.
- Nordhausen, K. and Tyler, D. E. (2015). A cautionary note on robust covariance plug-in methods. *Biometrika*, 102(3):573–588.
- Nordhausen, K. and Virta, J. (2019). An overview of properties and extensions of FOBI. *Knowledge-Based Systems*, 173:113–116.
- Olszewski, R. T. (2001). Generalized feature extraction for structural pattern recognition in time-series data. Technical report, Carnegie-Mellon Univ. Pittsburgh PA School of Computer Science.

- Radojicic, U. and Nordhausen, K. (2020). Non-Gaussian component analysis: Testing the dimension of the signal subspace. In Maciak, M., Pesta, M., and Schindler, M., editors, *Analytical Methods in Statistics. AMISTAT 2019*, pages 101–123. Springer, Cham.
- Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer, New York, 2nd edition.
- Rousseeuw, P. J., Raymaekers, J., and Hubert, M. (2018). A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27(2):345–359.
- Sawant, P., Billor, N., and Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27(1):83–102.
- Segaert, P., Hubert, M., Rousseeuw, P., and Raymaekers, J. (2020). *mrfDepth: Depth Measures in Multivariate, Regression and Functional Settings*. R package version 1.0.13.
- Staerman, G., Mozharovskiy, P., Cl  men  on, S., and d’Alch   Buc, F. (2019). Functional isolation forest. In Lee, W. S. and Suzuki, T., editors, *Proceedings of The Eleventh Asian Conference on Machine Learning*, volume 101, pages 332–347. PMLR.
- Suhaila, J., Jemain, A. A., Hamdan, M. F., and Zin, W. Z. W. (2011). Comparing rainfall patterns between regions in Peninsular Malaysia via a functional data analysis technique. *Journal of Hydrology*, 411(3-4):197–206.
- Tucker, J. D. (2020). *fdasrvf: Elastic Functional Data Analysis*. R package version 1.9.4.
- Tucker, J. D., Wu, W., and Srivastava, A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66.
- Tyler, D. E., Critchley, F., D  mbgen, L., and Oja, H. (2009). Invariant coordinate selection. *Journal of the Royal Statistical Society: Series B*, 71(3):549–592.
- Virta, J., Li, B., Nordhausen, K., and Oja, H. (2020). Independent component analysis for multivariate functional data. *Journal of Multivariate Analysis*, 176:104568.

Appendix

Supplementary figures for the Weather data set

Visualization of the outliers on the aligned curves In Figures 21 to 23 we give the curves for the three variables, and we highlight the abnormal curves using colors and add the curves number. In Figures 21 (resp., 22 and 23) we coloured in red (resp., blue and green) the abnormal curves observed in the temperature (resp., wind and log precipitation) on the 3 variables.

Supplementary figures for the Aeronautic data set

Visualization of the outliers on the aligned curves In Figure 24, we plot 100 normal flights in grey and highlight the successive flights detected as outliers in red. It can be seen that the four red curves differ from the others at the end of the flight with an engine speed that drops compared to the other curves. The failure occurred during the flight that followed, the data of which were not included in our statistical analysis.

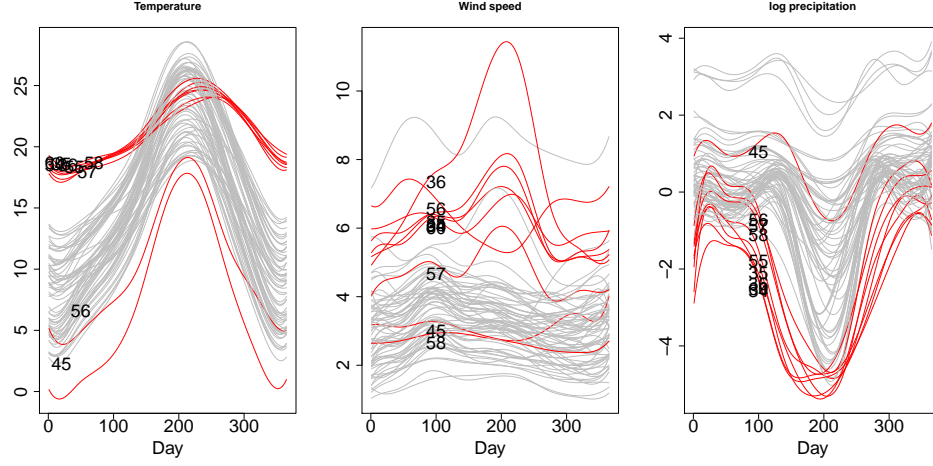


Figure 21: **Weather data set** - Suspected outlying curves (34,35,36,45,55,56,57,58,60) flagged in red based on the temperature.

Global ICS Figure 25 gives the correlations between the first invariant component and the 5 Fourier coefficients of the 6 initial variables. The first component is clearly only correlated with coefficients associated with the engine speed (X^3).

Directional outlyingness method The result of the DO method proposed by Rousseeuw et al. (2018), without changing the *distOptions* in the *fom* function is given in Figure 26. The four successive flights detected as outliers are coloured in red, the abnormal flights detected by the FOM cutoff on the vDO axis are coloured in green and those detected by the fDO axis are coloured in blue.

The abnormal flights detected are very different from the flights detected using global or point-wise ICS. Figure 27 gives the time points where the weight is equal to zero, meaning that it was not possible to calculate the outlyingness index at these time points. In fact, all of the cruise period is removed from the analysis. This observation is not only true for this example. Of the dozens of data files we have studied in the aeronautical domain, we have encountered this problem systematically at least at some time points and often on nearly half of the time points.

Supplementary figures for the Semiconductor data set

Visualization of the outliers on the aligned curves In Figure 28, we plot in grey the normal curves and in red the runs that were classified as abnormal in the data file. It is not easy to interpret the outlyingness of the red curves but it seems that there is a mix between shift and amplitude outliers.

Global ICS In Figures 29 and 30 respectively, we give the scree plot and the ICS distances by run using the first 6 ICS components for each $D = 5, 11$ and 15 . The results for $D = 11$ and 15 are similar, but for $D = 5$, when using the cutoff we recommend, we detect a number of outliers much larger than the 2% we can tolerate. Nevertheless, when looking at the three plots of Figure 30, we can note that the observations associated with large ICS distances compared to the majority group are very similar.

Directional outlyingness method We observe in Figure 31 that the *fom* function cannot run for 50% of the time points that correspond to weights equal to 0, essentially at the beginning and at the end of the time period.

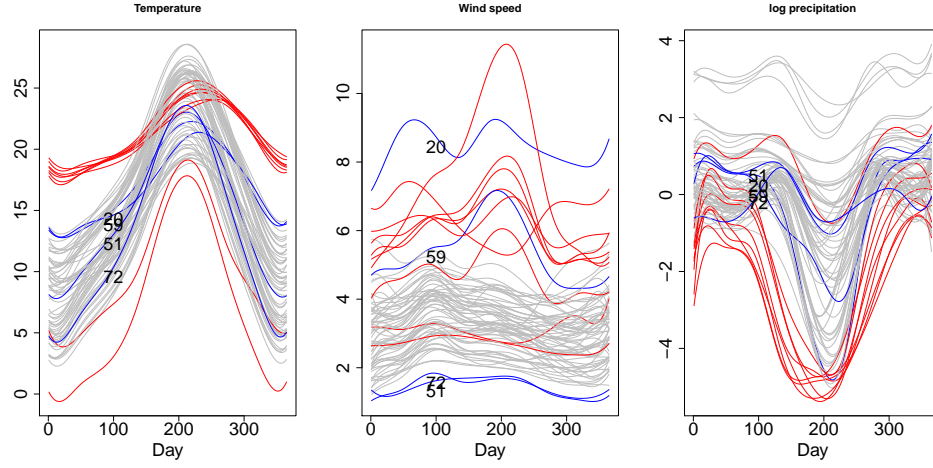


Figure 22: **Weather data set** - Additional suspected outlying curves (20,51,59,72) flagged in blue based on the wind speed.

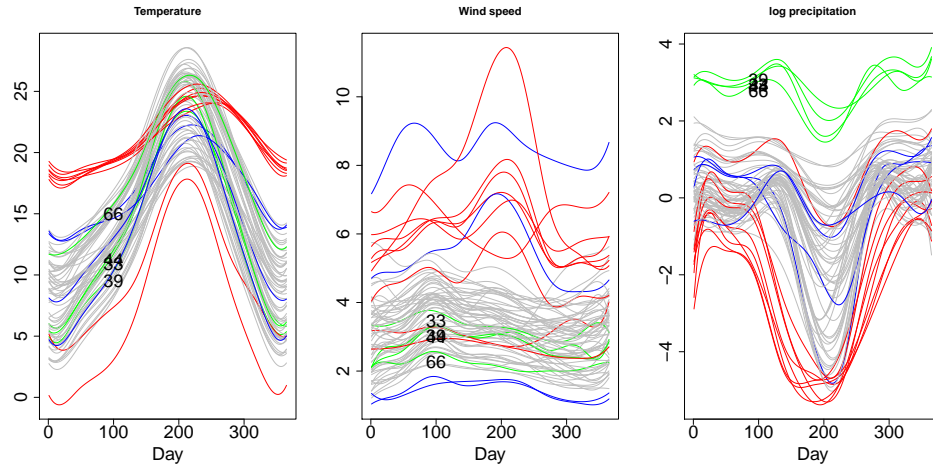


Figure 23: **Weather data set** - Additional suspected outlying curves (33,39,44,66) flagged in green based on the log precipitation.

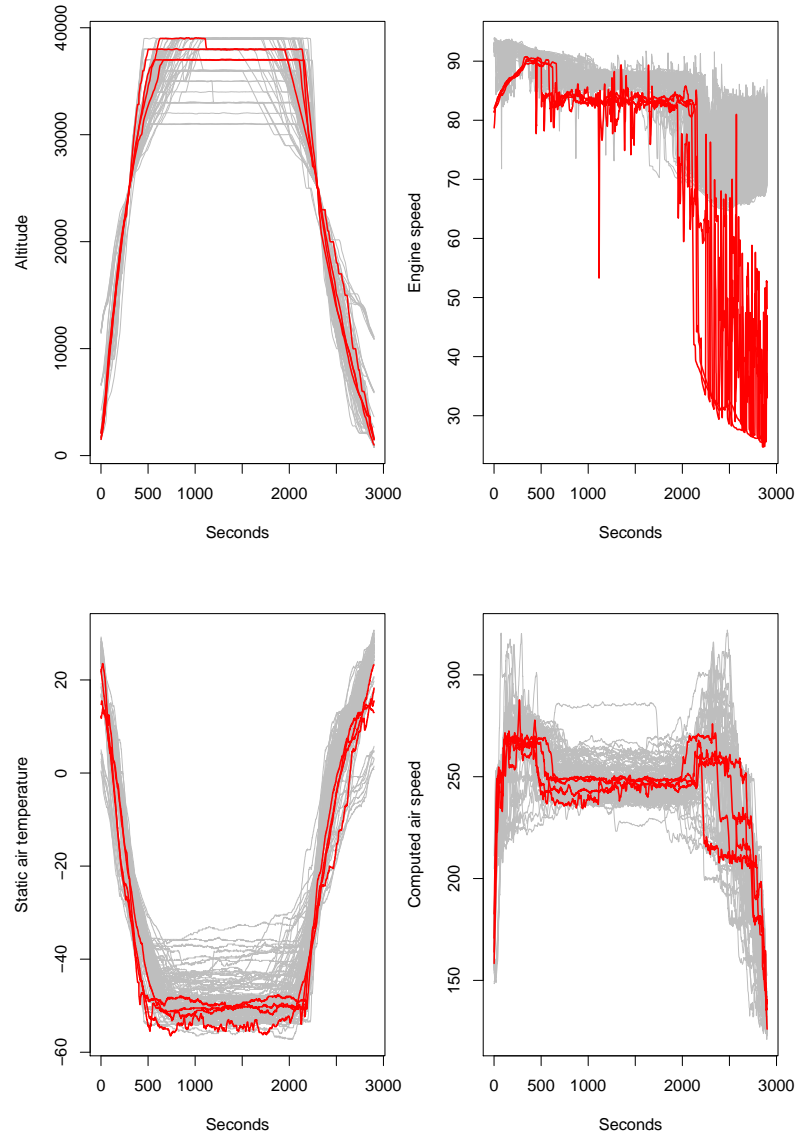


Figure 24: **Aeronautical data set** - Observed flights after alignment by feature with $T = 2900$ seconds. In red the flights that precede the generator loss.

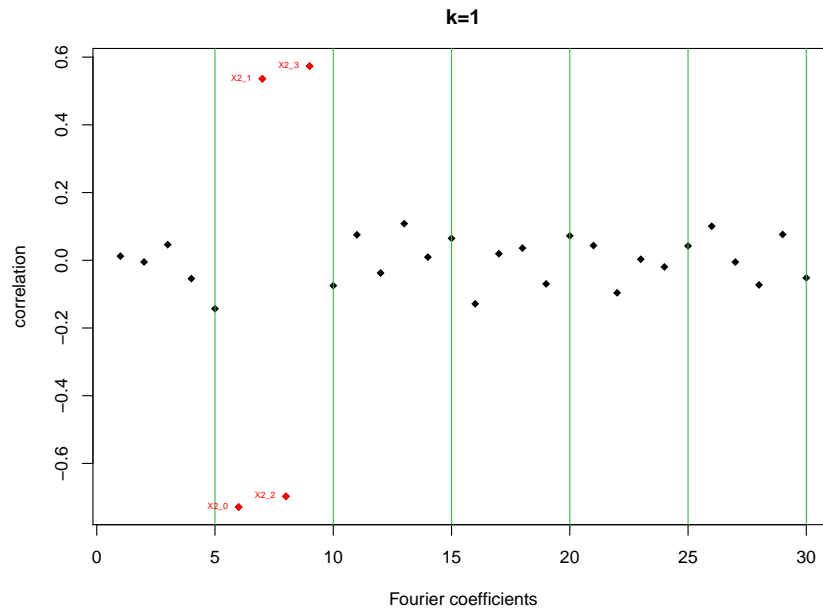


Figure 25: **Aeronautical data set** - Global ICS: correlations between the first invariant component and the 5 Fourier coefficients of the 6 initial variables. Red colour corresponds to correlation with absolute value larger than 0.20. The green lines are separators between the features which are ordered from X^1 to X^6 .

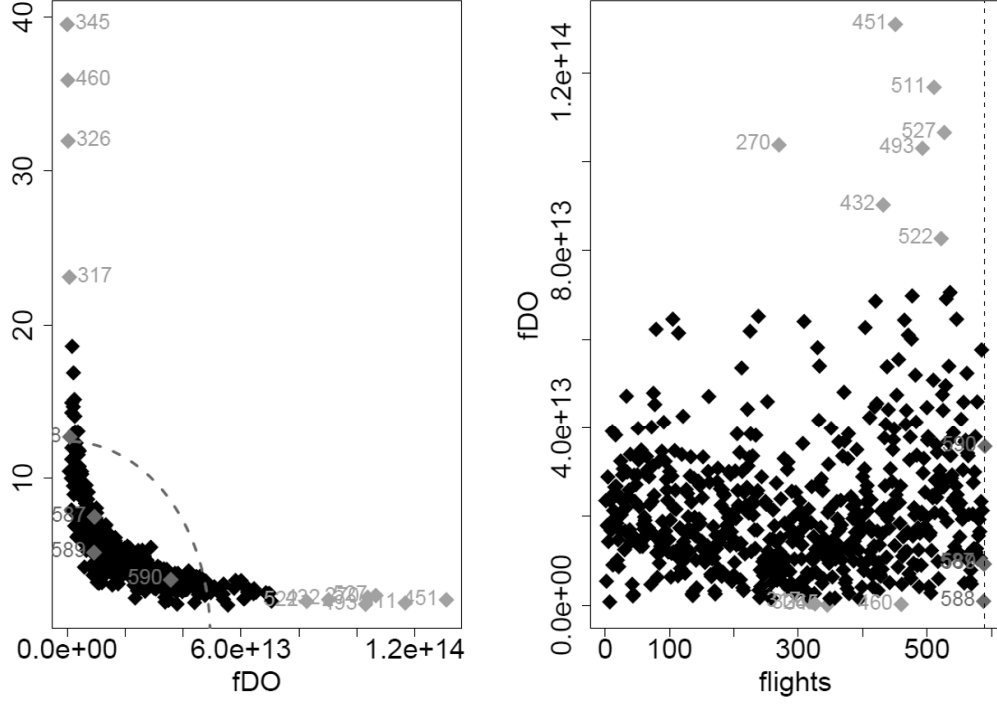


Figure 26: **Aeronautical data set** - DO method, Left: FOM, flights. Right: fDO.

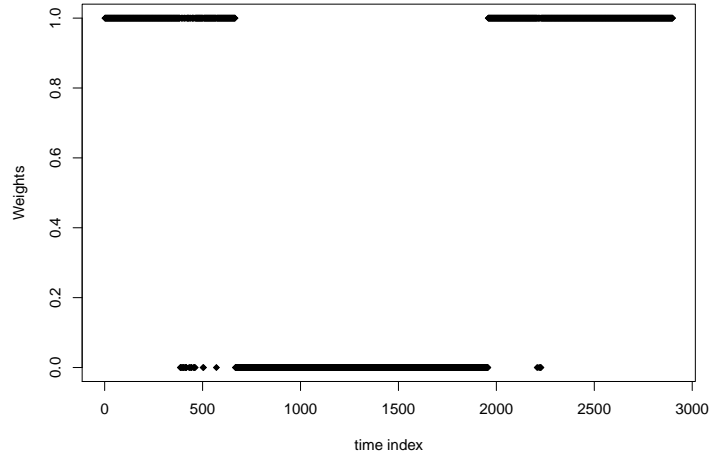


Figure 27: **Aeronautical data set** - DO method: 0/1 weights with a zero weight at time points where the method cannot run (46% of the points).

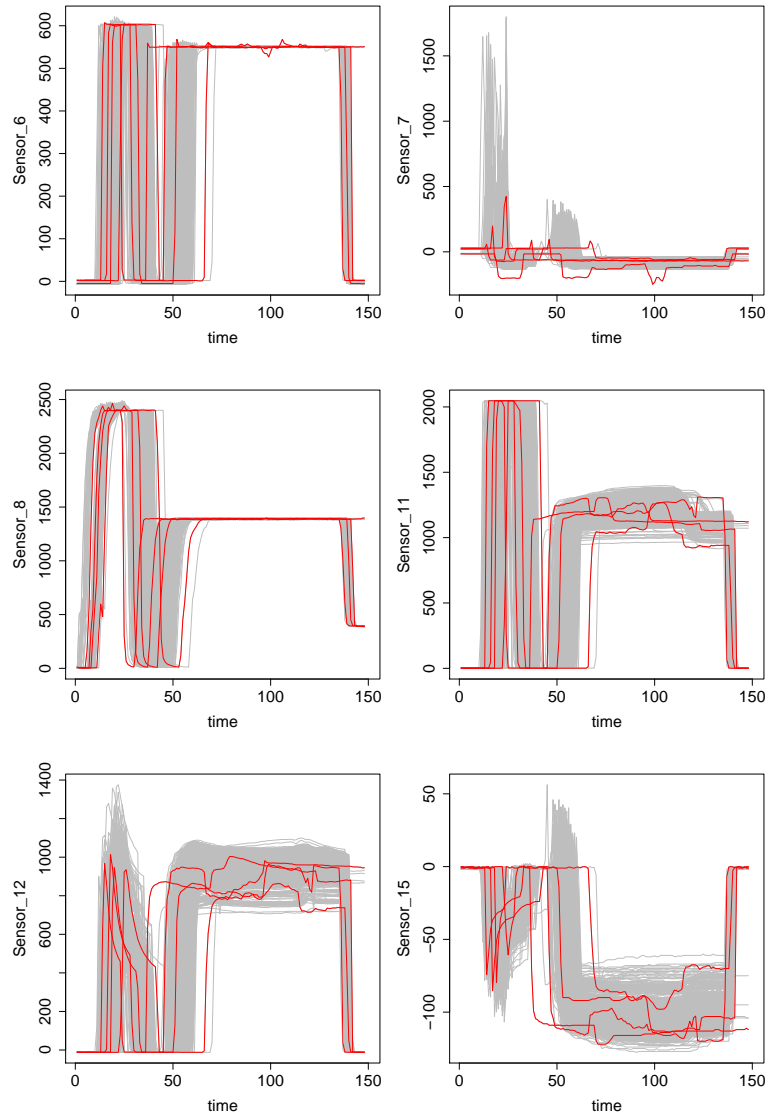


Figure 28: **Semiconductor data set** - Observed runs by sensor after alignment. The 4 abnormal (severe or very severe faults) runs are coloured in red.

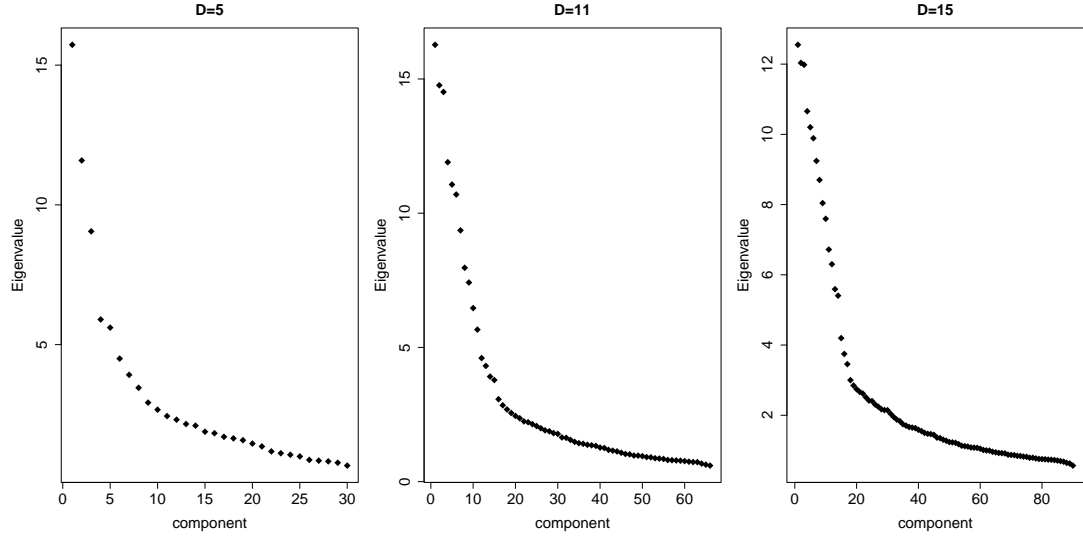


Figure 29: **Semiconductor data set** - Global ICS: scree plot with $D = 5$ (left), 11 (center) and 15 (right) coefficients for the Fourier basis.

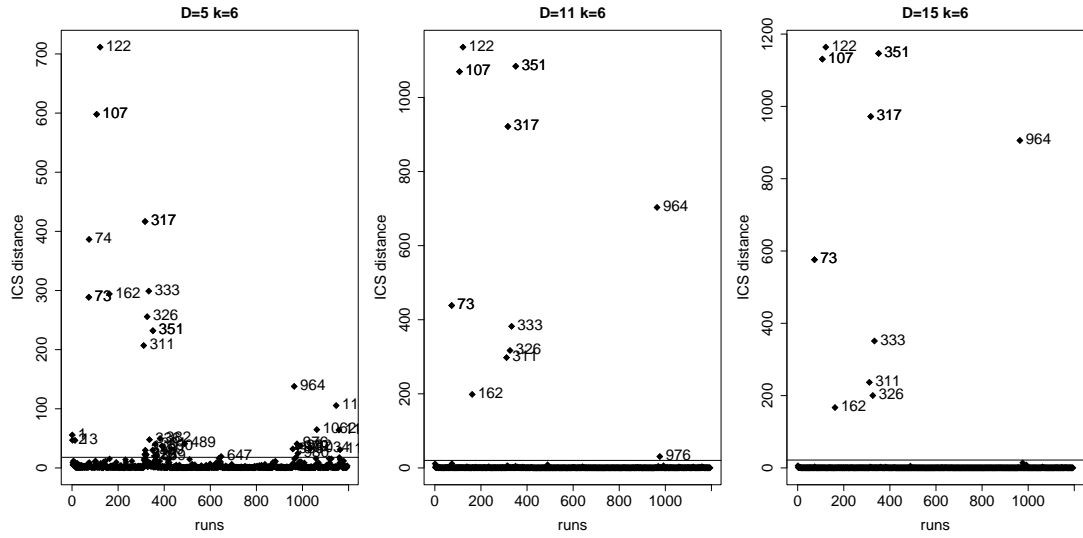


Figure 30: **Semiconductor data set** - Global ICS: square ICS distance with $k = 6$ and $D = 5$ (left), 11 (center) and 15 (right) coefficients for the Fourier basis.

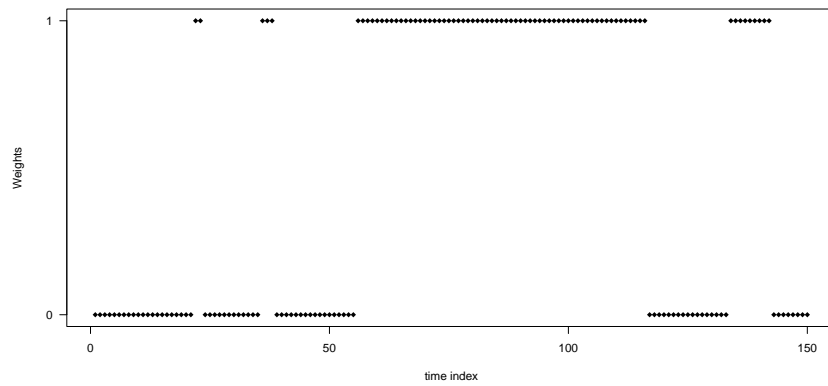


Figure 31: **Semiconductor data set** - Directional outlyingness method: 0/1 weights with a zero weight at time points where the method cannot run (50% of the points).