Anomaly Detection for Bivariate Signals

Marie Cottrell, Cynthia Faure, Jérôme Lacaille, Madalina Olteanu

Anomaly detection for bivariate signals

Marie Cottrell [1], Cynthia Faure [1,3], Jérôme Lacaille [2], and Madalina Olteanu [1,4]

1 - SAMM, EA 4543
Panthéon-Sorbonne University
90 rue de Tolbiac, 75013 Paris, France
http://samm.univ-paris1.fr
&
2 - Safran Aircraft Engines,
Rond Point René Ravaud, Réau, 77550 Moissy Cramayel, France
https://www.safran-aircraft-engines.com
&
3 - Aosis Consulting,
20 impasse Camille Langlade, 31100 Toulouse, France
http://www.aosis.net/
&
4 - MaIAGE, INRA
Paris-Saclay University
Domaine de Vilvert, 78352 Jouy en Josas, France
http://maiage.jouy.inra.fr

Abstract. The anomaly detection problem for univariate or multivariate time series is a critical question in many practical applications as industrial processes control, biological measures, engine monitoring, supervision of all kinds of behavior. In this paper we propose an empirical approach to detect anomalies in the behavior of multivariate time series. The approach is based on the empirical estimation of conditional quantiles. The method is tested on artificial data and its effectiveness is proven in the real framework of aircraft-engines monitoring.

1 Introduction

Detecting anomalies in univariate and multivariate time series is a critical question in many practical applications, such as fault or damage detection, medical informatics, intrusion or fraud detection, and industrial processes control. The present contribution stems from a joint work with the Health Monitoring Department of Safran Aircraft Engines Company. The motivation behind this collaboration was to find a judicious framework for mining the multivariate high-frequency data recorded on board computers during flights, and isolate unusual patterns, abnormal behaviors of the engine, and possibly anomalies. The issue of anomaly detection on flight data is not new, and some previous results of the joint work with Safran may be found in [2] and [18].

In a broader context, the literature on anomaly detection is quite abundant and was developed for decades in various fields: machine learning, statistics,

signal processing. The techniques used for addressing the matter use supervised or unsupervised learning, model-based algorithms, information theory or spectral decomposition. For quite an exhaustive review, the reader may refer to [3].

More particularly, we focus here on the issue of detecting collective anomalies or discords – an unusual subsequence of a time-series, in contrast to local anomalies which consist in unique abnormal time-instants – in a multivariate context. Detecting collective anomalies or unusual patterns in univariate time-series has been extensively studied, and we may cite, for instance, algorithms based on piecewise aggregated approximation [14], nearest-neighbor distances [11], Fourier or wavelet transforms [5], [16], Kalman filters [12], ... In the multivariate case, anomaly detection has to take into account both the multivariate aspect of the data, and the temporal span, the possibly existing correlations and dependencies. Whereas the initial approaches used time series projection [10] and independent component analysis [1] to convert the multivariate time series into a univariate one, or performed separate anomaly detection for each variable [13], global approaches have been developed only recently. Among these recent works, one may cite, for instance [6], who use a kernel-based method for capturing dependencies among variables in the time series, and [15] who use neural networks for isolating anomalous regions in a multivariate time-series.

This paper addresses the issue of detecting anomalies in a multivariate time series context. Unlike some of the cited literature above, we suppose the data is a set of multivariate time-series, which have already been segmented into patterns of unequal lengths, using some change-point detection technique. Our goal is to find the most unusual of them, and for doing so we take the unsupervised learning approach *(as defined by the AI)*. No hypothesis whatsoever is made on some underlying model, the only constraint is to suppose that one component of the time series, called *key variable* in the sequel, exists, may be distinguished, and its behavior strongly influences the behavior of the rest. The approach we introduce here may be briefly described as follows: first, the initial patterns of the *key variable* are summarized by a fixed number of numerical features, second, they are clustered into an optimal number of clusters, third, the multi-variate patterns are realigned and synchronized within each cluster, fourth, unusual patterns are extracted after computing confidence tubes from empirical quantiles in each cluster. This approach was introduced in [7], and the contribution of the present paper relies in the use of conditional first order quantiles for computing the confidence tubes, instead of quantiles computed at each time instant, independently of the past. As will be illustrated in the Experiments section, this conditional approach greatly improves the ratio of false positives detection.

For simplicity purposes, the bivariate case only is presented here, but the algorithm can be easily extended to higher dimensional data.

The rest of the paper is organized as follows: Section 2 describes the main steps of the proposed methodology, Section 3 contains results on simulated examples and a comparison between the previous version of the method and the modified one based on conditional quantiles, while Section 4 illustrates the method on real-life dataset stemming from flight data.

2 Methodology

Let $S_a = (X_a, Y_a)$ be a set of bivariate $\mathbb{R}^2$-valued time series, $a = 1, \dots A$. For each a, X_a and Y_a are of equal length l_a. Note that the lengths l_a can be different from one time series to another one. We assume that one of the two variables is a *key variable* (easier to observe, with a limited number of different behaviors, which influences the behavior of the other). This hypothesis is not very restrictive since in many processes, there is a measure which gives a first information on all the others variables (water temperature, blood composition, temperature of the core, etc. according to the application field.) This hypothesis leads us to define two successive levels of analysis: the first one deals with the key variable (say X_a) and the second one will further take into account the second variable (Y_a).

When assessing the possible existence of abnormal elements, a straightforward approach would consist in mixing together all signals $(X_a)_a$, compute an *average signal*, and say that all signals *far* from this average may be labelled as anomalies. However, there is one major issue with this approach, coming from the fact that the lengths l_a are different, so how does one actually compute an *average signal*? Furthermore, even if one was to find a way to define the *average signal*, there is no reason to summarize all signals behaviors by the average one. Hence, in order to have a better representation of the data, we choose to cluster signals X_a. Clustering will provide a limited number of homogeneous groups, and within each of them, one may define a representative signal.

2.1 Clustering

The difficulty of dealing with signals of different lengths is overcome as suggested in [9]: each signal X_a is replaced by a fixed-length vector composed of its relevant numerical features (length, midpoint value, median, variance, variances on the two halves, means of the two halves, ...). Let M be the number of relevant features for the set of $(X_a)_a$ time series. Any clustering algorithm may then be used on the feature-vectors data. In the following, let $C_1, C_2, \dots, C_I$ denote the clusters obtained on the feature vectors, where I is the number of clusters.

In the following, the clustering procedure consisted into first training a self-organizing map (SOM) with a large number of clusters, and second computing an optimal partitioning through an hierarchical agglomerative clustering (HAC) applied to the code-vectors computed by SOM. The optimal number of clusters is selected using an empirical criterion based on the percentage of explained variance. It is worth mentioning at this point that one may avoid summarizing time-series by a fixed number of features, and use some time-series dissimilarity measure (dynamic time warping, the distance defined in Eq. (1), ...) instead. In this case, relational or kernel SOM [17] may be used for clustering.

Once the clustering is trained, each cluster C_i contains a set of time-series X_a, say X_a^i for simplicity. They are grouped together based on the similarities of their extracted features, but may have different lengths. Hence, the next step of our methodology is to summarize each cluster by a *reference curve*, RC_i, which will serve hereafter for computing quantiles and for visualisation purposes.

2.2 Introducing reference curves for summarizing the clusters

The notion of *reference curve* for a set on univariate unequal time-series as defined hereafter was introduced in [8]. Let us briefly recall here the main steps of how does one compute it.

First, one has to define the dissimilarity between two curves with different lengths, say X_{a_1} and X_{a_2}, with lengths l_1 and l_2, and $l_1 < l_2$. If $X_{a_1} = (x_1, x_2, \ldots, x_{l_1})$, its extended version $\tilde{X}_{a_1}$ of length $l_2 + l_1 + l_2$ is

$$\tilde{X}_{a_1} = (\underbrace{x_1, x_1, \ldots, x_1}_{l_2} | \underbrace{x_1, x_2, \ldots, x_{l_1}}_{l_1} | \underbrace{x_{l_1}, x_{l_1}, \ldots, x_{l_1}}_{l_2}).$$

Note that if X_{a_1} is extracted from a longer time series, the extensions at left and at right may be done using the true values in the complete series. The dissimilarity between X_{a_1} and X_{a_2} is the defined as:

$$diss(X_{a_1}, X_{a_2}) = \min_{q \in 1, \ldots, l_1 + l_2 - 1} \frac{\| I_q(\tilde{X}_{a_1}) - X_{a_2} \|}{2 l_2} \ , \tag{1}$$

where $I_q(\tilde{X}_{a_1}) = \tilde{X}_{a_1}[q, q + l_2 - 1]$ is a l_2-long section of $\tilde{X}_{a_1}$ taken between indexes q and $q + l_2 - 1$, for $q = 1, \ldots, l_1 + l_2 + 1$.

Next, one computes the reference curve RC_i of a cluster C_i described by the curves X_a^i as being the one curve among the $|C_i|$ available which minimizes the sum of dissimilarities with respect to all curves in the cluster. Let L_i be the length of RC_i. Once RC_i is computed, all the curves in the cluster are realigned with respect to it. This step is achieved by applying a transformation which combines translation, completion and truncation, as described in the next section.

2.3 Time-series realignment within clusters

For realigning curves within a cluster, the idea is to use a similar approach to that in the previous section. We briefly describe it here, using the notations and approach in [8]. Consider a curve X_a^i in C_i with length l_a^i. X_a^i is then extended at its left by L_i constant values equal to its first value $X_a^i(1)$, and at its right by L_i constant values equal to its last value $X_a^i(l_a^i)$. The resulting curve is denoted $\hat{X}_a^i$. One may then compute

$$diss(X_a^i, RC_i) = \min_{q \in 1, \ldots, l_a^i + L_i + 1} \frac{\| I_q(\hat{X}_a^i) - RC_i \|}{2 L_i} \tag{2}$$

and

$$q_a^i = \arg \min_{q \in 1, \ldots, l_a^i + L_i + 1} \frac{\| I_q(\hat{X}_a^i) - RC_i \|}{2 L_i} \ , \tag{3}$$

where $I_q(\hat{X}_a^i) = \hat{X}_a^i[q, q + L_i - 1]$ is a L_i-long section of $\hat{X}_a^i$, computed between instants q and $q + L_i - 1$, for $q = 1, \ldots, l_a^i + L_i + 1$.

Each curve X_a^i of cluster C_i is thus replaced by $I_{q_a^i}(\hat{X}_a^i)$, denoted hereafter $\breve{X}_a^i$. In practice, this means replacing the initial unequal-length curves by a set of new curves, similar to the initial ones, and having all length L_i. The same *synchronization-transformation* is applied to the second signal Y_a^i : it is extended, translated, cut at the same indexes as X_a^i. The transformed signal is denoted by $\breve{Y}_a^i$ and has also the same length L_i. For the sake of simplicity, we denote by C_i the set of signals X_a^i as well as that of the transformed signals $\breve{X}_a^i$. The corresponding second components Y_a^i or $\breve{Y}_a^i$ define a set D_i. We denote by E_i the set of all the couples of transformed signals $(\breve{X}_a^i, \breve{Y}_a^i)$.

2.4 Anomaly detection

Let us recall at this point that our main goal is to detect possibly atypical curves in E_i. The anomalies can be related to the first component X, to the second one Y, or to both. Our approach consists in building quantile-based confidence tubes in each set C_i and D_i, for a given confidence level. The simplest way to do this is to compute point-by-point empirical quantiles, which is equivalent to supposing there is no time dependency in the data. Another solution is to take the past instants into account and consider rather conditional quantiles, as suggested in [19] and [4]. We describe next both approaches.

Point-by-point confidence tubes (CT method). Confidence tubes are computed in each cluster, for each set of realigned curves $(\breve{X}_a^i(t), \breve{Y}_a^i(t))$, where $t = 1, \ldots, L_i$. For a given confidence level $1 - \alpha$ (typically $\alpha = 5\%$), one denotes by $q^X_{t,\frac{\alpha}{2}}$ (resp. $q^Y_{t,\frac{\alpha}{2}}$) and $q^X_{t,1-\frac{\alpha}{2}}$ (resp. $q^Y_{t,1-\frac{\alpha}{2}}$) the α-quantiles computed for each time instant t.

The $1-\alpha$ confidence tube of the set $\breve{X}_a^i$ in C_i is defined with a lower bound curve given by $(q^X_{t,\frac{\alpha}{2}})_{t=1,\ldots,L_i}$ and an upper bound curve given by $(q^X_{t,1-\frac{\alpha}{2}})_{t=1,\ldots,L_i}$. The $1-\alpha$ confidence tube of the set $\breve{Y}_a^i$ in D_i are similarly computed.

With the previous definition of confidence tubes, we may now introduce the notion of *anomaly*. In the subsequent, a curve in C_i or D_i is considered as *anomalous* if at least $P\%$ consecutive instants are outside the corresponding confidence tube. P is generally to be tuned by the user; for the examples in this manuscript, its value was fixed to 10%.

Conditional quantiles (CQ method) Since data are time series and since one has strong reasons to suppose a dependency structure in time, another approach for computing empirical quantiles consists in taking the past values of the series into account. Conditional quantiles are also computed on the realigned curves $(\breve{X}_a^i(t), \breve{Y}_a^i(t))$, where $t = 1, \ldots, L_i$. For a given confidence level $1 - \alpha$ (typically $\alpha = 5\%$), one denotes by $\tilde{q}^X_{t,\frac{\alpha}{2}}$ (resp. $\tilde{q}^Y_{t,\frac{\alpha}{2}}$) and $\tilde{q}^X_{t,1-\frac{\alpha}{2}}$ (resp. $\tilde{q}^Y_{t,1-\frac{\alpha}{2}}$) the α-quantiles computed for each time instant t, conditionally on the very recent past, $t-1$. We only consider here a dependency structure of order 1, but more

sophisticated ones could be used, with an optimal selection of the number of lags.

The lower conditional quantiles will be computed by solving

$$\mathbb{P}\left(\breve{X}_a^i(t) \leq \tilde{q}^X_{t,\frac{\alpha}{2}}(x)/\breve{X}_a^i(t-1) = x\right) = \frac{\alpha}{2}\ ,$$

while the upper ones by solving

$$\mathbb{P}\left(\breve{X}_a^i(t) \geq \tilde{q}^X_{t,1-\frac{\alpha}{2}}(x)/\breve{X}_a^i(t-1) = x\right) = \frac{\alpha}{2}\ .$$

As the conditional distribution of $\breve{X}_a^i(t)$ conditionally to $\breve{X}_a^i(t-1)$ is generally unknown, the values in $t-1$, $\breve{X}_a^i(t-1)$, have to be discretized in order to have a sufficient number of values of $\breve{X}_a^i(t)$, conditionally to one given value of $\breve{X}_a^i(t-1)$.

Once the conditional quantiles are computed for each discretized value ${}^d\breve{X}_a^i(t-1)$ of $\breve{X}_a^i(t-1)$, a curve in C_i is detected as an *anomaly* if and only if the number of couples $(\breve{X}_a^i(t-1), \breve{X}_a^i(t))$ such that

$$\breve{X}_a^i(t) \notin \left[\tilde{q}^X_{t,\frac{\alpha}{2}}({}^d\breve{X}_a^i(t-1)), \tilde{q}^X_{t,1-\frac{\alpha}{2}}({}^d\breve{X}_a^i(t-1))\right]$$

is greater than a certain threshold P, fixed by the user (usually 10 %). The $1-\alpha$ conditional quantiles of the set $\breve{Y}_a^i$ in D_i are similarly computed.

3 An experimental example on simulated data

We first illustrate the proposed methodology and the interest of using conditional quantiles instead of point-by-point ones on a simulated example. 2,000 artificial bivariate time series were built, with an X-variable having one of the four shapes described in Figure 1. These shapes were inspired by the real-data from aircraft flights that will be described in the next section. For each time series, its length is randomly generated (between 500 and 3,000 time instants), as well as the instant where the slope changes (between the first and the second third of the series), the slope values. A Gaussian centered noise with a variance varying between 10 and 100 is also added for supplementary noise, and eventually the resulting curves are smoothed using a 5-degree polynomial.

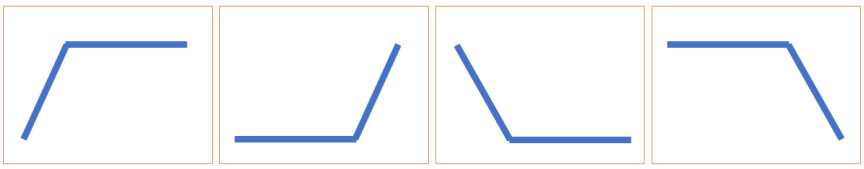
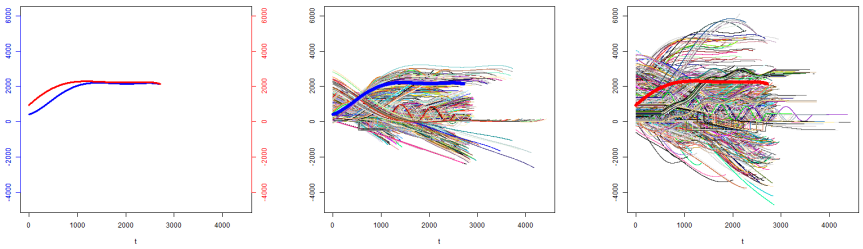


Fig. 1: Four shapes used for the simulated X-signals

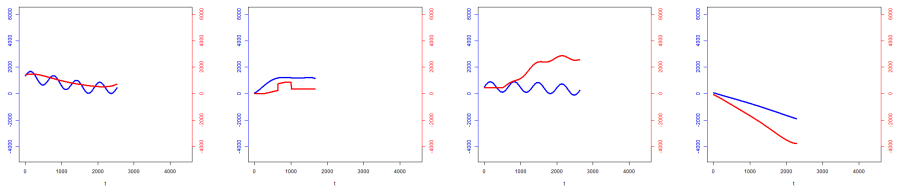
The Y-variables are simulated starting from X: for a given time series X_a, a change-point z is randomly selected, and a new time series is simulated with two random slopes before and after the change-point. The slopes are selected more or less in the same range as the X_a's. If necessary, Y_a is extended or cut so as to have the same length as X_a. Eventually, the resulting curves are also smoothed using a 5-degree polynomial. The resulting X-curves, Y-curves, and a bivariate example (X_a, Y_a) of simulated time series are illustrated in Figure 2.



(a) (b) (c)

Fig. 2: (a): Example of bivariate signal, X_a in blue and Y_a in red; (b): All signals X_a; (c): All signals Y_a

Furthermore, we added 50 anomalies to the simulated data. Four types of anomalies were introduced, as shown in Figure 3: sinusoidal, "hat"-shaped, and linear. A couple of variables (X, Y) can be anomalous in X only, in Y only or in both components.



(a) (b) (c) (d)

Fig. 3: Examples of atypical curves (X blue, Y red)

The proposed methodology was then applied to the simulated data: the curves were clustered, realigned within each cluster, and point-by-point and conditional quantiles were computed in each cluster. Eventually, anomalous curves identified by each of the two methods were extracted, and compared with the ground truth.

Clustering with SOM followed by HAC yielded five final clusters, denoted $C_1, C_2, \ldots, C_5$ for the X curves, and $D_1, D_2, \ldots, D_5$ for the Y curves. We recall here that D_i are defined by $D_i = \{Y_a / X_a \in C_i\}$, $i = 1, \ldots, 5$. The resulting clusters, which are globally homogeneous, are illustrated in Figure 4. The reference curves $RC_1, \ldots, RC_5$ computed according to the definition in Section 2.2, are drawn with solid red lines in Figure 4.

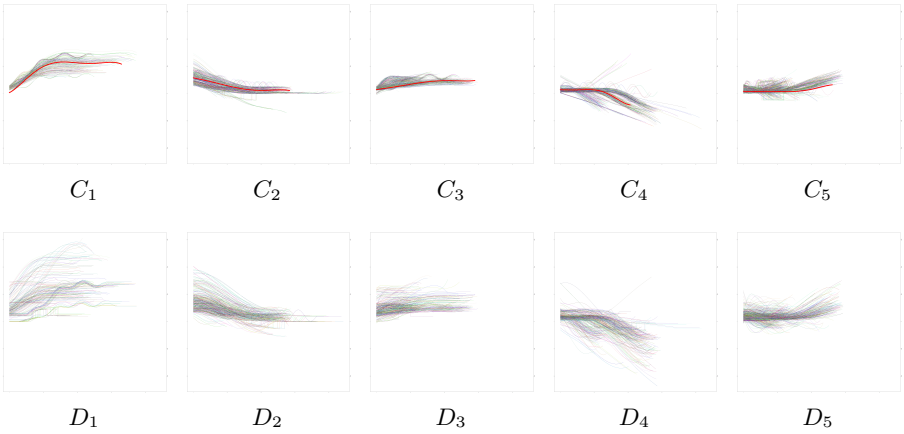


Fig. 4: Clustering (X-curves on top and Y-curves below)

Next, X and Y time-series within each cluster are realigned using the transformation in Section 2.3. We only illustrate here the results for the first cluster. Figure 5-a and 5-b contains the initial curves X_a^1 of cluster C_1 and their transformed $\breve{X}_a^1$. Similarly, Figure 5-c and 5-d displays the initial curves Y_a^1 of D_1 and their transformed $\breve{Y}_a^1$.

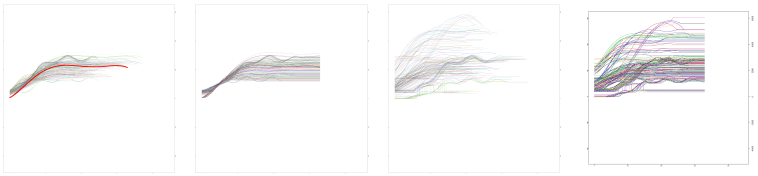
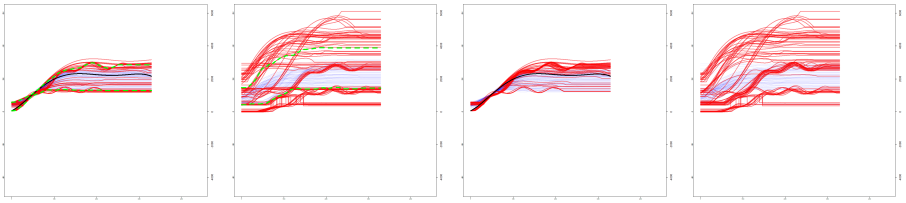


Fig. 5: Realignment and transformation of the curves in C_1 (a,b) and D_1 (c,d), initial curves in (a) and (c), transformed ones in (b) and (d)

Eventually, the last step consists in detecting the atypical curves in each cluster after having computed the confidence tubes (CT) with point-by-point

empirical quantiles, and empirical conditional quantiles (CQ) (see Section 2.4). Figure 6-a presents the results of the CT detection method for cluster C_1: all the X- curves in C_1 are drawn and the detected atypical ones are highlighted in red. The Y- curves and the detected atypical ones are displayed in Figure 6-b. In the same way, Figure 6-c and -d presents the result of the CQ detection method. At first glance, both methods seem to detect the same atypical curves.

The two approaches for identifying anomalous curves are then compared by computing their confusion matrices for each cluster. The confusion matrices for the CT and CQ methods are given in Tables 1 and 2. The following abbreviate notations were used: A for atypical curves, NA for normal curves, D for curves detected as atypical, ND for curves detected as normal. On the one hand, the CT method appears to detect a not negligible number of false alarms, in each cluster. On the other hand, with the CT method the number of false alarms decreases dramatically. The results are globally significantly improved.



(a) (b) (c) (d)

Fig. 6: Anomalies detected in cluster 1 with the CT method: (a) - on X; (b) - on Y. Anomalies detected in cluster 1 with the CQ method: (c) - on X; (d) - on Y. Normal curves are in blue, abnormal in red.

CT	i	1			2			3			4			5		
		D	ND	T	D	ND	T	D	ND	T	D	ND	T	D	ND	T
C_i	A	15	9	24	22	3	25	26	0	26	19	0	19	41	25	66
	NA	**27**	131	158	**20**	297	317	**11**	272	283	**11**	291	302	**30**	590	620
	T	42	140	182	42	300	342	37	272	309	30	291	321	71	615	686
D_i	A	35	15	50	24	7	31	13	1	14	19	7	26	23	16	39
	NA	**29**	103	132	**18**	293	311	**13**	282	295	**15**	280	295	**38**	609	647
	T	64	118	182	42	300	342	26	283	309	34	287	321	61	625	686
C_i & D_i	A	6	5	11	9	2	11	12	1	13	9	0	9	20	16	36
	NA	**11**	160	171	**10**	321	331	**10**	286	296	**10**	302	312	**16**	634	650
	T	17	165	182	19	323	342	22	287	309	19	302	321	36	650	686

Table 1: Confusion matrices for the CT method, in bold the number of false alarms

CQ	i	1			2			3			4			5		
		D	ND	T	D	ND	T	D	ND	T	D	ND	T	D	ND	T
C_i	A	23	1	24	24	1	25	25	1	26	17	2	19	45	21	66
	NA	**0**	158	158	**1**	316	317	**5**	278	283	**5**	297	302	**25**	595	620
	T	23	159	182	25	317	342	30	279	309	22	299	321	70	616	686
D_i	A	37	13	50	26	5	31	14	0	14	23	3	26	26	13	39
	NA	**13**	119	132	**13**	298	311	**12**	283	295	**15**	280	295	**35**	612	647
	T	50	132	182	39	303	342	26	283	309	38	283	321	61	625	686
C_i & D_i	A	8	3	11	9	2	11	12	1	13	7	2	9	19	17	36
	NA	**7**	164	171	**8**	323	331	**8**	288	296	**10**	302	312	**19**	631	650
	T	15	167	182	17	325	342	20	289	309	17	304	321	38	648	686

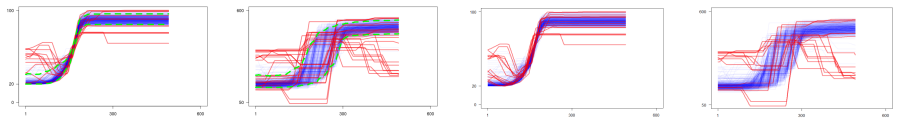
Table 2: Confusion matrices for the CQ method, in bold the number of false alarms

4 An application to real-world data

Let us now illustrate the method on a real dataset, containing bivariate time-series recorded during aircraft flights. Some of the following results are excerpted from C. Faure's PhD thesis [7], completed in collaboration with the Health Monitoring Department of Safran Aircraft Engines Company. In actuality, the real data contained much higher dimensional time series, since the sensors placed on the engines register more than 50 different signals. Here, for illustration purposes, we only considered the fan speed and the temperature inside the engine. The fan speed is the *key variable* X and the temperature is Y.

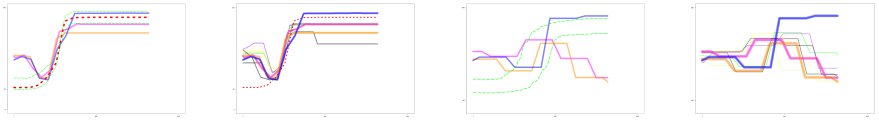
The data was initially made of 549 flights and 8 different engines, with a mean duration of 2.8 hours per flight. After having partitioned the flight data using some change-point detection algorithm, 4500 transient ascending phases (time-series with an ascending behavior) were extracted, clustered, and the rest of the methodology described in Section 2 was applied to them. Their lengths are comprised between 200 and 10,000 time units (8Hz).

We describe next the results obtained in one cluster only, that mainly contains take-offs. Figure 7 contains the atypical curves (red) with respect to the normal ones (blue), detected with the CT ((a) and (b) for the X and Y curves) and the CQ methods ((c) and (d) for the X and Y curves). The CT method detects 12 atypical X-curves, while the CQ detects 14 (6 common ones). As for the Y-curves, CT detects 24 anomalies, CQ 18, of which 14 anomalies are common. Let us stress also the case of bi-dimensional curves which are detected as atypical for X **and** for Y by both methods (see Figure 8). The CT method detects three couples which are atypical for X and for Y, while the CQ method finds ten couples, that include the first three. Again, the CQ method has better performances than the CT method, even though it is not possible to compute the confusion matrices here, since we have no a priori knowledge about the existence of the anomalies. In this real-world study, the experts have been able to bring a validation to our findings. The detected cases corresponded to some events that they could identify.



(a) (b) (c) (d)

Fig. 7: Anomalies detected in cluster 1 with the CT method: (a) - on X; (b) - on Y. Anomalies detected in cluster 1 with the CQ method: (c) - on X; (d) - on Y. Normal curves are in blue, abnormal in red.



(a) (b) (c) (d)

Fig. 8: Three atypical couples are detected by the CT method (X-curves in (a) and Y-curves in (b)). The CQ method detects seven more couples drawn in (c) and (d). The reference curve of the cluster is represented in red dots in (a) and (c). The confidence tubes defined by the CT method are in green dots.

5 Conclusion

We describe a complete methodology, based upon clustering, curve realignment and empirical quantiles computation, that allows one to detect abnormal elements in a large sample of time series with unequal lengths. When using conditional quantiles, the results are dramatically improved and the number of false alarms significantly reduced. We strongly believe that these results could be further improved by an optimal selection of the time lag in the conditional quantiles computation. From a practical point of view, the proposed methodology may be very useful in helping experts and engineers identify abnormal behaviors in the signals recorded during aircraft engines utilization. In the case of aircraft engine real data, companies may use this technique to increase the probability of detecting any kind of atypical, abnormal behavior of some recorded variable, in order to prevent any incident and to plan the maintenance events.

References

1. Baragona, R., Battaglia, F.: Outliers detection in multivariate time series by independent component analysis. Neural Computation 19(7), 1962–1984 (2007)
2. Bellas, A., Bouveyron, C., Cottrell, M., Lacaille, J.: Anomaly detection based on confidence intervals using som with an application to health monitoring. In: Villmann, T., Schleif, F.M., Kaden, M., Lange, M. (eds.) Advances in Self-Organizing Maps and Learning Vector Quantization, Proceedings of the 10th International

Workshop, (WSOM 2014). pp. 145–155. AISC, Springer-Verlag, Mittweida, Germany (July 2014)
3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. ACM Comput. Surv. 41(3), 15:1–15:58 (Jul 2009)
4. Charlier, I., Paindaveine, D., Saracco, J.: Conditional quantile estimation based on optimal quantization: from theory to practice. Computational Statistics and Data Analysis 91, 20–39 (2015), https://hal.inria.fr/hal-01108504
5. Chen, X.y., Zhan, Y.y.: Multi-scale anomaly detection algorithm based on infrequent pattern of time series. Journal of Computational and Applied Mathematics 214(1), 227–237 (2008)
6. Cheng, H., Tan, P., Potter, C., Klooster, S.: A robust graph-based algorithm for detection and characterization of anomalies in noisy multivariate time series. In: 2008 IEEE International Conference on Data Mining Workshops. pp. 349–358 (2008)
7. Faure, C.: Détection de ruptures et identification des causes ou des symptômes dans le fonctionnement des turboréacteurs durant les vols et les essais. Ph.D. thesis, Université Paris 1 Panthéon-Sorbonne (2018)
8. Faure, C., Bardet, J.M., Olteanu, M., Lacaille, J.: Design aircraft engine bivariate data phases using change-point detection method and self-organizing maps. In: Conference: ITISE - International work-conference on Time Series. University of Granada, Granada, Spain (September 2017)
9. Faure, C., Bardet, J.M., Olteanu, M., Lacaille, J.: Using self-organizing maps for clustering anc labelling aircraft engine data phases. In: Ed., M.C. (ed.) 12th International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+ 2017). pp. 1–8 (June 2017)
10. Galeano, P., Peña, D., Tsay, R.S.: Outlier detection in multivariate time series by projection pursuit. Journal of the American Statistical Association 101(474), 654–669 (2006)
11. Keogh, E., Lin, J., Fu, A.: Hot sax: efficiently finding the most unusual time series subsequence. In: Fifth IEEE International Conference on Data Mining (ICDM'05). pp. 226–233 (2005)
12. Knorn, F., Leith, D.J.: Adaptive kalman filtering for anomaly detection in software appliances. In: IEEE INFOCOM Workshops 2008. pp. 1–6 (2008)
13. Lakhina, A., Crovella, M., Diot, C.: Characterization of network-wide anomalies in traffic flows. In: Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement. pp. 201–206. IMC '04, ACM, New York, NY, USA (2004)
14. Lin, J., Keogh, E., Lonardi, S., Chiu, B.: A symbolic representation of time series, with implications for streaming algorithms. In: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery. pp. 2–11. DMKD '03, ACM, New York, NY, USA (2003)
15. Malhotra, P., Vig, L., Shroff, G., Agarwal, P.: Long short term memory networks for anomaly detection in time series. In: Verleysen, M. (ed.) European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2015). pp. 89–94. Bruges, Belgium (2015)
16. Michael, C.C., Ghosh, A.: Two state-based approaches to program-based anomaly detection. In: In Proceedings of the 16th Annual Computer Security Applications Conference. pp. 21–30. IEEE Computer Society (2000)
17. Olteanu, M., Villa-Vialaneix, N.: On-line relational and multiple relational som. Neurocomputing 147(1), 15–30 (2015)
18. Rabenoro, T., Lacaille, J., Cottrell, M., Rossi, F.: Anomaly detection based on indicators aggregation. In: International Joint Conference on Neural Networks (IJCNN 2014). pp. 2548–2555. Beijing, China (July 2014)

19. Samanta, T.: Non-parametric estimation of conditional quantiles. Statistics and Probability Letters 7, 497–412 (1989)