Control Engineering Practice, Volume 54, September 2016, pp. 70-80
DOI:10.1016/j.conengprac.2016.05.018

Canonical Variate Analysis for Performance Degradation under Faulty Conditions

C. Ruiz-Cárcel[a], L. Lao[a], Y. Cao[a], D. Mba[b]

[a] *School of Engineering, Cranfield University, Building 52 School of Engineering, MK430AL (UK)*
[b] School of Engineering, London South Bank University, UK.
Phone/ email: +44(0) 1234 754639 c.ruizcarcel@cranfield.ac.uk

Abstract

Condition monitoring of industrial processes can minimize maintenance and operating costs while increasing the process safety and enhancing the quality of the product. In order to achieve these goals it is necessary not only to detect and diagnose process faults, but also to react to them by scheduling the maintenance and production according to the condition of the process.

The objective of this investigation is to test the capabilities of canonical variate analysis (CVA) to estimate performance degradation and predict the behaviour of a system affected by faults. Process data was acquired from a large-scale experimental multiphase flow facility operated under changing operational conditions where process faults were seeded. The results suggest that CVA can be used effectively to evaluate how faults affect the process variables in comparison to normal operation. The method also predicted future process behaviour after the appearance of faults, modelling the system using data collected during the early stages of degradation.

Keywords: fault detection; fault diagnosis; system identification; multivariate analysis; canonical variate; experimental study

1. Introduction

Condition monitoring of industrial processes can minimize maintenance and operating costs while increasing the process safety and enhancing the quality of the product. In order to achieve these goals it is necessary not only to detect and diagnose process faults, but also to react to them by scheduling the maintenance and production according to the condition of the process. The technique presented here shows an application of Canonical Variate Analysis (CVA) to evaluate how detected faults affect the process variables in comparison with normal operation. This information can be used by plant operators to schedule optimal maintenance and production plans that consider the condition of the process.

The complexity of modern industrial plants has grown due to the use of devices of different nature (process, electrical and mechanical interfaces) interconnected by control systems. Physics-based methods have been traditionally used for process monitoring [1]. This type of methods provide accurate results as long as the mathematical model represents precisely the behaviour of the system. However, due to the complexity of modern facilities, the characterization by using first-principle models can be challenging and sometimes impossible [1]. These difficulties, together with the high degree of instrumentation and automation in modern industrial processes, have motivated an increment in the popularity of data-based methods in industrial monitoring applications.

In particular, multivariate algorithms such as the Principal Component Analysis (PCA) or the Partial Least Squares (PLS) and their dynamic variations (DPCA and DPLS) are extensively used nowadays for fault detection and diagnosis. The main advantage of this type of algorithms against traditional univariate methods is their capacity to capture the correlation between the different variables measured in the process [2]. Some of the main challenges associated with the application of data-driven methods are high-dimensional data, non-Gaussian distributions, non-linear relationships and dynamically varying operational conditions [1; 3-5]. Recent developments on these techniques have shown improved results for non-linear system monitoring [6-8].

Although the application of DPCA and DPLS has produced good results when applied to systems working under variable operating conditions, other algorithms such as Canonical Variate Analysis (CVA) can achieve a better representation of the system dynamics [3; 9; 10]. CVA aims to maximize the correlation between two sets of variables [2]. This method has been applied in the past for detection and diagnosis of process faults using computer simulated data [2-4; 9-12], data acquired from small test rigs [13], and data acquired from particular parts of a system [14]. Recently, the capabilities of CVA to detect and diagnose process faults were tested and compared with other monitoring techniques using experimental data [15]. This investigation proved the superior

Published by Elsevier. This is the Author Accepted Manuscript issued with:
Creative Commons Attribution Non-Commercial No Derivatives License (CC:BY:NC:ND 4.0).
The final published version (version of record) is available online at DOI:10.1016/j.conengprac.2016.05.018.
Please refer to any applicable publisher terms of use.

performance of CVA for the detection and diagnosis of faults in a real complex system working under varying operational conditions.

Another characteristic of CVA that is of special interest for this investigation is that it can be used to obtain a model of a dynamic system using purely process data [12]. This method has been successfully applied in the past for system identification using computer simulated data (including the Tennessee Eastman case study) [16; 17] and a pilot-scale distillation column [12] working under normal operating conditions. CVA has demonstrated to be more computationally efficient than other system identification techniques such as PLS as it is based on generalized singular value decomposition instead of iterative estimation of residuals from the previous step [17].

The objective of this paper is to assess the system identification capabilities of CVA in a real complex system affected by different faults, modelling the system not only under normal operation but also under faulty conditions to predict the behaviour of the faulty system. In [15] it was demonstrated that CVA can effectively detect and diagnose faults in a complex system working under dynamically varying operating conditions. In this paper CVA is used to estimate how the detected faults affect the variables measured in the system compared with normal operation, and also to model the behaviour of a faulty system. Once a fault is detected, if it is not critical and allows a safe process operation, the information provided by this algorithm can be used for the development of methods to improve the maintenance and production schedules. This improvement is based on the evaluation of the following two aspects: Firstly the impact of the fault in terms of operation safety, product quality, loading conditions, energy consumption, etc. compared with normal operation. Secondly the future behaviour of the faulty system for the different operating conditions expected.

In this investigation CVA was applied in the first place for fault detection and diagnosis, which provided information about the time at which the operators will be aware of the presence of a fault and what are the most affected variables. Secondly, once the starting point is known, CVA was applied to obtain estimations of the process measurements under normal and faulty conditions using process data. In this investigation performance degradation is studied as the difference between the process variables measured under faulty conditions and the corresponding model estimations assuming absence of faults. Additionally, the faulty system was modelled to predict its future behaviour under different operating conditions.

The process data was acquired from a multiphase flow facility in a university laboratory, which can be configured with different functionalities for teaching and research. For the work reported in this paper, the rig was configured to induce different faults, which are common in process plants such as pipeline blockage, leakage and valve sticking. The severity of the faults introduced was chosen to affect the measured variables while allowing a safe operation of the system. As this is a multiphase flow facility, an accurate mathematical model of the plant is not available. Therefore, it is impossible to provide a reliable mathematical description on what is happening. On the other hand, such a practical restriction is more representative of many industrial systems and highlights the importance of data driven approaches, such as those reported in this investigation. The novelty of this work relies on two key aspects: 1) The application of CVA for estimation of performance degradation under different fault conditions using real process data, and 2) the use of measurements acquired during the early stages of degradation to predict the dynamic behaviour of the faulty system.

The rest of the paper is structured as follows: chapter 2 describes the CVA methodology for fault detection and diagnosis in 2.1, and system identification in 2.2. A detailed description of the test rig is presented in 2.3, including details about the data sets acquired from the rig in 2.4. The results obtained from the data analysis are presented in chapter 3, including algorithm training in 3.1, fault detection and diagnosis in 3.2, performance degradation in 3.3 and prediction of faulty system behaviour in 3.4. Finally the work is concluded in chapter 4.

2. Methodology

2.1 CVA for fault detection and diagnosis

Fault detection and diagnosis in real complex systems using CVA was introduced in [15]. Although it is not the main objective of the application shown here, fault detection and diagnosis is the starting point of estimation of degradation in a faulty system. Most of the currently used multivariate monitoring algorithms such as CVA have a similar application procedure. These algorithms have to be trained using data acquired from the system working under normal operating conditions before they can be used for monitoring. The CVA mathematical application procedure has been described in several previous investigations such as [2; 3; 9; 10], consequently here only some fundamentals are introduced. The stages for the application of the method are:

1) Acquisition of training data from the system working in normal operating conditions.
2) Selection of tuning parameters and calculation of the transformation matrices and thresholds for the health indicators.
3) Fault detection: computing the health indicators at each time point by projecting the newly acquired data through the transformation matrices obtained with the nominal model. These indicators are compared with the threshold calculated in the previous step to differentiate between normal and faulty conditions.
4) Fault diagnosis: Once a fault is detected, it is possible to locate the source of the fault by looking at which measured variables are contributing most to the final value of the health indicators.

The target of CVA is to find the linear combinations of measured variables that maximize the correlation between two sets, the past and future sets of variables. For this purpose, the measurements stored in vectors $\boldsymbol{y_k} \in \mathbb{R}^m$ (each one containing information from m sensors) are amplified at each time point k by adding p previous and f future measurements in order to take into account time correlations. This operation generates the past and future observation vectors $\boldsymbol{y_{p,k}}$ and $\boldsymbol{y_{f,k}}$ respectively:

$$\boldsymbol{y_{p,k}} = \begin{bmatrix} \boldsymbol{y_{k-1}^T} & \boldsymbol{y_{k-2}^T} & \cdots & \boldsymbol{y_{k-p}^T} \end{bmatrix}^T \in \Re^{mp} \tag{1}$$

$$\boldsymbol{y_{f,k}} = \begin{bmatrix} \boldsymbol{y_{k}^T} & \boldsymbol{y_{k+1}^T} & \cdots & \boldsymbol{y_{k+f-1}^T} \end{bmatrix}^T \in \Re^{mf} \tag{2}$$

Vectors $\boldsymbol{y_{p,k}}$ and $\boldsymbol{y_{f,k}}$ are normalized to zero mean and unit variance to produce $\boldsymbol{\hat{y}_{p,k}}$ and $\boldsymbol{\hat{y}_{f,k}}$ respectively. The vectors obtained at different times k are arranged in columns to configure the past and future matrices $\boldsymbol{Y_p}$ and $\boldsymbol{Y_f}$:

$$\boldsymbol{Y_p} = \begin{bmatrix} \boldsymbol{\hat{y}_{p,p+1}}, \boldsymbol{\hat{y}_{p,p+2}}, \ldots \boldsymbol{\hat{y}_{p,p+M}} \end{bmatrix} \in \Re^{mp \times M} \tag{3}$$

$$\boldsymbol{Y_f} = \begin{bmatrix} \boldsymbol{\hat{y}_{f,p+1}}, \boldsymbol{\hat{y}_{f,p+2}}, \ldots \boldsymbol{\hat{y}_{f,p+M}} \end{bmatrix} \in \Re^{mf \times M} \tag{4}$$

The solution for the optimization problem to find the linear combination that best correlates both data sets can be obtained by using the Singular Value Decomposition (SVD) of the Hankel matrix $\boldsymbol{H}$:

$$\boldsymbol{H} = \boldsymbol{\Sigma}_{Y_f,Y_f}^{-1/2} \boldsymbol{\Sigma}_{Y_f,Y_p} \boldsymbol{\Sigma}_{Y_p,Y_p}^{-1/2} = \boldsymbol{UDV}^T \tag{5}$$

where $\sum_{\boldsymbol{A,B}}$ represents the sample covariance matrix of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$. The singular value decomposition provides two orthogonal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ and a diagonal matrix $\boldsymbol{D}$. Each of the elements γ_i in the diagonal of $\boldsymbol{D}$ represents the amount of correlation between the corresponding column vectors of $\boldsymbol{U}$ and $\boldsymbol{V}$. The reduced matrix $\boldsymbol{V_r} \in \mathbb{R}^{mf \times r}$ can be computed by selecting the columns of $\boldsymbol{V}$ related with the r highest eigenvalues from $\boldsymbol{D}$. This reduced matrix is used for the calculation of the transformation matrices $\boldsymbol{J}$ and $\boldsymbol{L}$ as $\boldsymbol{J}=\boldsymbol{V_r^T}\sum_{\boldsymbol{Y_p,Y_p}}^{-1/2}$ and $\boldsymbol{L}=(\boldsymbol{I}-\boldsymbol{V_r}\boldsymbol{V_r^T})\sum_{\boldsymbol{Y_p,Y_p}}^{-1/2}$. These matrices are used to project the acquired data into the low-dimensional space. The canonical variates $\boldsymbol{z_k}$ and the residuals $\boldsymbol{\varepsilon_k}$ are calculated as:

$$\boldsymbol{Z} = \boldsymbol{JY_p} \in \Re^{r \times M} \tag{6}$$

$$\boldsymbol{E} = \boldsymbol{LY_p} \in \Re^{mp \times M} \tag{7}$$

where the columns of $\boldsymbol{Z}$ and $\boldsymbol{E}$ are $\boldsymbol{z_k}$ and $\boldsymbol{\varepsilon_k}$ respectively at different time points k. These matrices contain the information required to calculate the health indicators in the retained and residual spaces. The Hotelling T^2 and Q statistics are used frequently for process monitoring. These indicators are computed at time point k as:

$$T_k^2 = \sum_{i=1}^{r} z_{k,i}^2 \tag{8}$$

$$Q_k = \sum_{i=1}^{mp} \varepsilon_{k,i}^2 \tag{9}$$

Faults are detected by comparing the value of these indicators with an upper control limit (UCL), which determines the threshold between normal and abnormal conditions. This threshold is calculated through a statistical analysis of the indicators during the training period for a given significance level α, such that $P(T^2<T^2_{UCL}(\alpha))= \alpha$ and $P(Q<Q_{UCL}(\alpha))= \alpha$ respectively. If the samples considered are normally distributed these control limits can be calculated using the Gaussian assumption. However, system nonlinearities can produce modelling errors that make this assumption invalid. Odiowei and Cao [3] developed a methodology that uses Kernel Density Estimations (KDE) to approximate the actual probability density function of the statistical indicators. This method was used in this investigation due to the non-linear nature of the process studied.

The isolation and identification of detected faults are necessary to understand the source of the problem and plan the best maintenance action according to the system condition. The use of contribution plots was suggested by Chiang et al. [2] to calculate the contribution of the different measured variables to the final value of the

health indicators. This procedure has been successfully applied by other researchers for fault identification [18-20] by representing the individual contributions in contribution plots. Once a fault has been detected, it is possible to prioritize the variables responsible for the fault based on their individual contributions. Plant engineers can use this information, together with their plant knowledge to determine the origin of the fault.

2.2 CVA for system identification

In addition to the detection and diagnosis of process faults, CVA can be used to produce a state-space model of the system using only process data. This method has been utilized in the past for system identification using simulated data or data acquired from pilot-scale rigs [12; 16; 17]. In particular Juricek et al. [17] presented a successful application of CVA for identification of the Tennessee Eastman challenge, which simulates a nonlinear, open-loop unstable and large-dimensional process which contains a mixture of fast and slow dynamics. The procedure of system identification using CVA requires a sequence of steps [12]:

1) Selection of the input (manipulated) and output (measured) variables for the model
2) Selection of the input excitation sequence
3) Collection of measurements using the selected input
4) Selection of the model structure (number of states and number of past and future lags considered)
5) Estimation of the parameters in the state-space equations (10) and (11)
6) Model validation

Given a set of inputs $\boldsymbol{u}$ and outputs $\boldsymbol{y}$, the model that represents the linear state-space is given by:

$$x_{k+1} = \boldsymbol{\Phi} x_k + \boldsymbol{G} u_k + w_k \tag{10}$$

$$\hat{y}_k = \boldsymbol{H} x_k + \boldsymbol{A} u_k + \boldsymbol{B} w_k + v_k \tag{11}$$

In this model $\boldsymbol{x_k}$ represents a r-order state vector, $\boldsymbol{\hat{y}_k}$ are estimated outputs, $\boldsymbol{w_k}$ and $\boldsymbol{v_k}$ are white noise and $\boldsymbol{\Phi}$, $\boldsymbol{G}$, $\boldsymbol{H}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are the state-space matrices. This model is non-time-variant, which means that it is suitable to estimate states and outputs of a system whose physical characteristics do not vary over time. According to [10; 21], the state vectors $\boldsymbol{x_k}$ can be replaced by the state estimates $\boldsymbol{z_k}$ obtained from (6) if the order of the model r is equal or greater than the actual order of the system. If the system outputs $\boldsymbol{\hat{y}_k}$ and inputs $\boldsymbol{u_k}$ are known and $\boldsymbol{z_k}$ can be obtained from CVA analysis using the method described in 2.1, the only unknowns of the system are the matrices $\boldsymbol{\Phi}$, $\boldsymbol{G}$, $\boldsymbol{H}$, $\boldsymbol{A}$ and $\boldsymbol{B}$. Larimore [21] proposed the next method based on multivariate regression for the calculation of these matrices:

$$\begin{bmatrix} \boldsymbol{\Phi} & \boldsymbol{G} \\ \boldsymbol{H} & \boldsymbol{A} \end{bmatrix} = \sum\left[\begin{pmatrix} \boldsymbol{z}_{k+1} \\ \hat{\boldsymbol{y}}_k \end{pmatrix}, \begin{pmatrix} \boldsymbol{z}_k \\ \boldsymbol{u}_k \end{pmatrix}\right] \cdot \sum^{-1}\left[\begin{pmatrix} \boldsymbol{z}_k \\ \boldsymbol{u}_k \end{pmatrix}, \begin{pmatrix} \boldsymbol{z_k} \\ \boldsymbol{u}_k \end{pmatrix}\right] \tag{12}$$

where $\sum[\boldsymbol{a},\boldsymbol{b}]$ denotes the sample covariance matrix for vectors $\boldsymbol{a}$ and $\boldsymbol{b}$.

This procedure was used in this investigation, in the first instance, to capture the behaviour of a system working under normal operating conditions. After a fault was introduced in the system, the degradation of the system's performance was evaluated as the difference between the actual measurements and the predictions produced by the model assuming normal operation (see 3.3). Secondly, data was acquired from the system during the early stages of degradation to generate a new model able to predict the behaviour of the faulty system for the different operating conditions forecasted in the future (see 3.4). Due to the limitation of using a non-time-variant model, this approach is valid for faults that remain stationary over time or faults that evolve slowly enough to be considered stationary during the prediction timeframe. The objective of this analysis is to observe how the fault affects the process when working under different loading conditions assuming that the severity of the faults allows the system to continue operating in suboptimal circumstances. The accuracy of the models was assessed by looking at the average normalized error e_j for each one of the m measured variables y_j:

$$e_j = \frac{1}{N}\sum_{k=1}^{T}\left|\frac{y_{k,j} - \hat{y}_{k,j}}{y_{k,j}}\right| \tag{13}$$

where $\hat{y}_{k,j}$ represents the estimated value of the measured variable y_j at time k, and N denotes the total number of observations acquired during a period of time T. The total averaged model error E is computed as:

$$E = \frac{1}{m}\sum_{j=1}^{m} e_j \tag{14}$$

2.3 Experimental set up

The data sets used for this investigation were acquired from a large scale 3-phase flow facility employed for the study of multiphase flow. It is designed to provide air, oil and water at controlled rates to a specific test section. The schematic representation in Fig. 1 shows a simplified representation of the facility. The test area consists of a gas/liquid two-phase separator (0.5 m diameter and 1.2 m high) at the top of a 10.5 m high platform. This device is connected with the rest of the equipment by pipelines with different bore sizes and geometries. The test area can be supplied with single phase of air, water and oil, or a mixture of those fluids, at required rates. The fluid mixtures are separated in an 11 m^3 horizontal three-phase separator on the ground level (GS500) after exiting the test area. In this separator the air is exhausted to the atmosphere and the emulsions of oil and water are separated and returned to their respective storage tanks (T200 and T100), with a capacity of approximately 12.5 m^3 each. Two coalescers (CW500 and CO500, both with a capacity of 1.5 m^3 approximately) warranty the fine separation or oil and water before returning to the tanks.

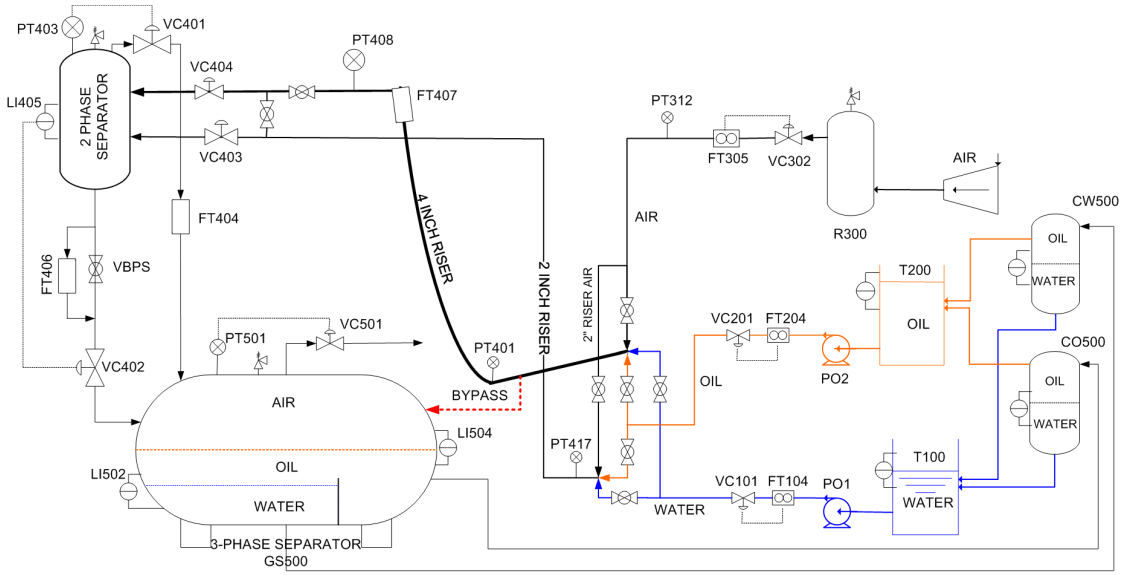

Fig. 1: Sketch of the three-phase flow facility (see TABLE 1 for tags description)

Air is taken from the atmosphere, compressed, and fed into an 8 m^3 vessel (R300). Before feeding to the system the air is filtered of droplets and particles and cooled for removing moisture. The air flow rate is controlled by valve VC302 according to the flow rate measurements from FT305. Water and oil are pumped independently through multistage Grundfos CR90-5 varying-speed pumps (PO1 and PO2) from storage tanks T100 and T200 respectively. Sensors FT104 FT204 provide measurements or water and oil flow rate respectively. Pneumatic valves VC101 and VC201 are used to control the water and oil flow rates respectively according to the user requirements. All control valves can be manipulated continuously from 0% (fully close) to 100% (fully open).

After the mixing point, the test area is supplied either through a 4” N.B. flow loop or a 2” flow loop. The 4” line is composed by a 10.5 m high catenary riser preceded by a 55 m long and 2° downward inclined pipeline, whilst the 2” line has a 40 m long horizontal pipeline connecting to a 10.5 m long vertical riser. There are manual valve manifolds in both ends of these flow lines to operate and isolate each other. The 4” line was used exclusively in all the experiments of this investigation, except for the study of a particular fault case (Case 3), where the 2” line was also involved.

TABLE 1: LIST OF MEASUREMENTS USED IN THIS STUDY

Output Variable	Location	Measured Magnitude	Units
y_1	PT312	Air delivery pressure	barg
y_2	PT401	Pressure in the bottom of the riser	barg
y_3	PT408	Pressure in top of the riser	barg
y_4	PT403	Pressure in top separator	barg
y_5	PT501	Pressure in 3 phase separator	barg
y_6	PT408	Diff. pressure (PT401-PT408)	barg
y_7	PT403	Diff. pressure (PT408-PT403)	mbarg
y_8	FT305	Flow rate input air	Sm^3/h
y_9	FT104	Flow rate input water	kg/s
y_{10}	LI405	Level top separator	m
y_{11}	FT407	Density top riser	kg/m^3
y_{12}	FT104	Temperature water input	°C
y_{13}	LI504	Level gas-liquid 3 phase separator	%
y_{14}	VC501	Position of valve VC501	%
y_{15}	VC302	Position of valve VC302	%
y_{16}	VC101	Position of valve VC101	%
y_{17}	PO1	Water pump current	A

TABLE 2: LIST OF INPUTS USED IN THIS STUDY

Input Variable	Magnitude	Unit
u_1	Air flow rate set point	Sm^3/h
u_2	Water flow rate set point	kg/s

Fieldbus based supervisory, control and data acquisition (SCADA) software Delta V [22] supplied by Emerson Process Management was used to manage the system. Time stamped data of different variables can be retrieved, processed and visualized. For this particular study all the data acquired was captured at a sampling rate of 1 Hz. The variables used include 17 process measurements (see TABLE 1) and two process inputs (air and water flow rate set point, TABLE 2).

During the experiments undertaken for this investigation a mixture of only air and water was used, and the pressure set point of the 3-phase separator was set to 1.0 barg. The variety in the nature of the selected variables (including pressure, flow rate, level, density, temperature, valve position and current measurements) and the different dynamics of these variables make the problem more challenging from the identification point of view. Additionally, the changing operational conditions and the non-linear nature of the multiphase flow process, together with the size and complexity of the test rig make this rig comparable to a small industrial facility.

2.4 Cases studied

Normal operation

As was mentioned in section 2.1, it is necessary to obtain data from the system working under normal operating conditions to obtain the transformation matrices and the thresholds of the health indicators from the CVA algorithm. These data are also necessary to build a dynamic model of the healthy process (2.2). The procedure used in this investigation to apply the CVA for fault detection and diagnosis is similar to the methodology used in [15], where CVA was used to detect and diagnose process faults in the same test rig used here. Three training data sets (T1, T2 and T3) were acquired from the system. The set points of air and water flow rates were deliberately modified at different moments during the tests, with the aim of capturing the system behavior under varying operating conditions. The duration of these sets was 11881 s, 3151 s and 4550 s respectively. The set T1 was acquired to train the CVA model while T2 and T3 were used for validation purposes. In order to cover a wide spectrum of operating conditions 20 different combinations of air and water flow rates (see TABLE 3) were tested in T1. The objective of this variety in the operational conditions is to ensure that the dynamics of the system are captured in all circumstances. The air and water flow rate set points chosen for T2 are also shown in TABLE 3 as they were the same selected for T1. Nevertheless, the set points selected in T3 were different from the set points used in T1 and T2, with the objective of validating the fault detection, diagnosis and system identification under operating points not tested during the training phase.

TABLE 3: TYPICAL SET POINT VALUES FOR AIR AND WATER FLOW RATES DURING T1 AND T2

Air flow rate (m^3/h)	75	100	125	150	
Water flow rate (kg/s)	0.5	1	2	3.5	6

Case 1: Sensor communication error/ Stuck valve

The first fault introduced simulates a sensor communication error in the pressure sensor PT501 (which measures pressure inside the 3-phase separator) or a stuck valve in VC501 (which regulates the pressure inside the 3-phase separator). In order to simulate that fault, the control loop that links PT501 and VC501 was broken by changing the valve operation to manual mode 3963 s after the test started. This test was carried out under varying operating conditions using different flow rate set points to those used for T1.

Case 2: Top separator air outlet blockage

The second fault simulates a partial pipe blockage in the air outlet line of the two phase separator. The fault was simulated by closing VC401 (which is normally 100% open) to 45% 3362 s after the test started. This test was also carried using different flow rate set points to those used for T1.

Case 3: Flow derivation through the 2" line

The last fault introduced simulates wrong system operation where part of the flow is derived through an alternative pipeline. The fault was simulated by opening the manual valves that allow the flow through the 2" line to the two phase separator in parallel with the flow in the 4" line 4550 s after the test started. Additionally, VC404 (normally 100% open) was fully closed 1629 s later to force the flow through the 2" line.

3. Results

3.1 Algorithm training and selection of tuning parameters for fault detection and diagnosis

In this particular application the system identification algorithm described in 2.2 is applied when a process fault has been detected and diagnosed using the methodology described in 2.1. In order to build the past and future vectors in (1) and (2) it is necessary to select the number of past and future lags considered (p and f) and the number of states retained (r). The autocorrelation function of the summed squares of all measurements can be used to estimate the number of lags that are significant in terms of correlation with past measurements [3]. Fig. 2 shows the autocorrelation function for the training data set T1 against a confidence bound of ±5%. According to this result the parameters p and f were set to 15 for this study.

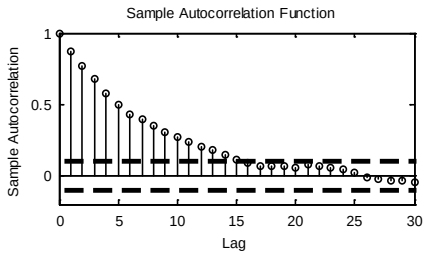

Fig. 2: Sample autocorrelation function for T1

From the different methodologies suggested for the calculation of the optimal number of states retained r, the most popular are those based on considering the dominant singular values in the matrix $\boldsymbol{D}$ [23] and methodologies based on the Akaike Information Criterion (AIC) [2]. Fig. 3 shows the normalized singular values obtained from $\boldsymbol{D}$ in (5) for data set T1. In this particular case there is not an evident number of dominant singular values as these values decrease slowly. In this case, if the number of retained states is set based on the rule of the dominant singular values it will derive in an unrealistic high order model [3]. The fault detection criterion used in this application, where both statistical indicators (T^2 and Q) are used at the same time for fault detection, makes the method robust to a range of states retained. Using this assumption those system variations not captured in the retained space will be captured by the residual space and vice versa.

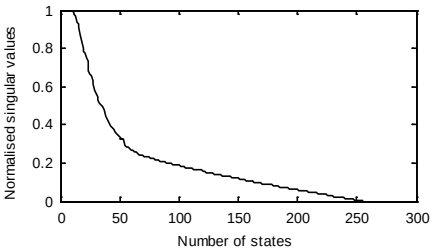

Fig. 3: Normalized singular values for T1

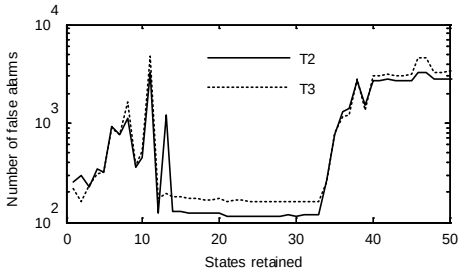

Fig. 4: Total number of false alarms for different number of states retained r

In order to check the capacity of T1 data set to represent the system dynamics accurately producing a low number of false alarms, this data set was used for training while T2 and T3 were used in the monitoring phase. The objective of this validation analysis is to select the optimal number of dimensions retained *r* that minimizes the total number of false alarms found in T2 and T3. A confidence bound of 99% was considered for the calculation of the UCL. After the analysis testing different values for *r* finally *r*=25 was adopted in order to minimize the false alarm rate in normal conditions (see Fig. 4), obtaining T^2 false alarm rates of 0.31% and 0.35% for T2 and T3 respectively, and 2.88% and 2.99% respectively for *Q*.

If parameter *r* is set too low it produces high false alarm rates because the dimension of the retained space is insufficient to accurately represent the system states, causing an increment of theT^2 threshold violations. On the contrary, too high values of *r* can cause the model overfitting the data [2], which also increases the false alarm rate. This behaviour was observed in the analysis (Fig. 4) and the total number of false alarms found followed a similar pattern for T2 and T3, although it was slightly higher for T3 probably because the operating conditions selected were different from those chosen in T1.

3.2 Results obtained for fault detection and diagnosis

The results obtained in terms of fault detection and diagnosis are sumarised in this section as the starting for the implementation of the proposed approach for estimation of performance degradation and prediction of faulty system behaviour. To avoid the problem of short false alarms before the introduction of the fault, the event of fault detection was considered when at least 60 consecutive samples of an indicator exceeded the UCL. For Case 3 two events of fault introduction are considered, the first when the 2” line valve was opened (3-1), and the second one when VC404 was closed (3-2). The fault detection results obtained from the three cases analysed are summarized in TABLE 4.

TABLE 4: FAULT DETECTION SUMMARY

Case	Fault start (s)	T^2			Q		
		Detection time (s)	False alarms	Failed detection	Detection time (s)	False alarms	Failed detection
1	3963	655	0.02%	31.31%	515	36.23%	20.15%
2	3362	510	0.26%	54.34%	459	13.83	63.81%
3-1	4550	-	-	-	1192	3.20%	11.85%
3-2	6179	8	0.34%	0.13%	-	-	-

All the faults were detected earlier by the *Q* statistic, which also produced a lower rate of failed detection in Case 1. On the other hand the T^2 statistic produced a very low rate of false alarms in all cases, showing a more reliable performance with no presence of faults. However, the failed detection rate for faults 1 and 2 is significantly high in both indicators. This is caused by the lack of detection between the fault introduction time and the detection time, as well as by the fluctuations of both indicators due to the changing operational conditions. In the case of fault 3, only the *Q* statistic was able to detect the fault when the 2” line was opened (3-1), and it was necessary to close VC404 and derive the flow completely through the 2” line (3-2) to make the fault visible for the T^2 statistic. It is important to notice that cases 1 and 2 were obtained under operating conditions that were not tested during the training phase. This demonstrates the capability of CVA to capture the system dynamics and detect abnormalities under a wide variety of operating conditions.

Once the faults were detected, contribution plots were used at the instant of fault detection to determine the origin of the fault. TABLE 5 summarizes the most significant variables in terms of fault contribution at the instant of fault detection. In Case 1 the most significant variable contributing to the T^2 statistic was the pressure measured in the 3-phase separator PT501 (y_5), caused by the lack of control action. Other pressure measurements

such as PT312, PT401, PT408 and PT403 were also affected; showing how the effects of a fault located at one point of the system disturb the rest of the process. The Q contribution plot points directly to valve VC501 (y_{14}) as the origin of the fault.

TABLE 5: VARIABLES WITH HIGHEST CONTRIBUTION AT DETECTION TIME

Contribution Ranking	Case 1		Case 2		Case 3	
	T^2	Q	T^2	Q	T^2	Q
1st	y_5	y_{14}	y_4	y_{10}	y_3	y_{11}
2nd	y_4	y_5	y_3	y_4	y_2	y_3

The most significant contribution for the T^2 indicator in Case 2 comes from the pressure measured inside the 2-phase separator (y_4), although the contribution of PT312, PT401 and PT408 are also significant due to the effects of the fault in the rest of the system. In the case of Q indicator, most of the contribution comes from PT403 (y_{10}), although the level measured inside the 2-phase separator (y_4) was also significant, pointing at a conflict in that region of the process.

The last case studied shows an important contribution of the top riser pressure PT408 (y_3) to the T^2 statistic, probably caused by flow being forced through the 2” line. The density measured in the top of the riser (y_{11}) is the most significant variable for the Q indicator, reinforcing the results provided by the T^2 contribution plot.

3.3 Performance degradation

The data acquired under normal operating conditions in data set T1 were used to build a dynamic model of the healthy system using the methodology presented in 2.2. These data include the 17 measurements listed in TABLE 1 and the two inputs presented in TABLE 2. This model was used to predict the system response assuming normal operation for the three faulty cases studied. The same input sequence $\boldsymbol{u}$ selected during testing was used to predict the response of the system. In order to select the optimum model order r and ensure model accuracy, the total averaged error presented in (14) was calculated for a range of values of r. T1 was used to build up the model, while T2 and T3 were used for validation. The objective of this analysis is to select the model order that minimizes the prediction error in T2 and T3, which is not necessarily the same order found in 3.1 as in that case it was calculated to minimize false alarm rate. Fig. 5 shows the results obtained from this analysis.

The averaged error for all variables in Fig. 5 grows as the order increases, mainly due to the inaccurate initial states estimation. This effect can be seen in Fig. 6, which represents the prediction of PT312 in T3 for model orders 2, 3 and 4 as an example. Fig. 6 shows that the model of order 2 produces very low oscillations in the initial estimations, but the model’s ability to respond to fast changes is quite limited due to the model restrictions. For model order 3 the initial oscillations are larger, but they are dissipated relatively quickly and the estimation provided is more flexible and able to represent better the system behaviour. In model order 4 (and larger) the initial oscillations are even higher, and the attenuation takes more time, affecting significantly the model error.

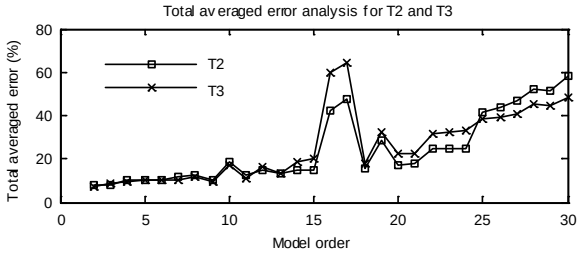

Fig. 5: Model order analysis for T2 and T3

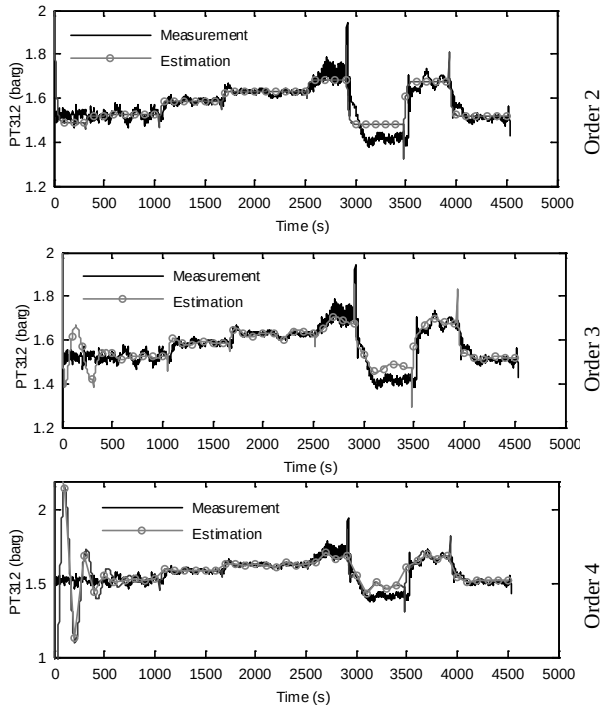

Fig. 6: Effect of model order over PT312 prediction accuracy for T3

For this investigation, model order 3 was selected in order to combine both flexibility to represent system dynamics and accurate initial estimations. Using this model structure, the average normalized error calculated for each measured variable in T2 and T3 is represented in TABLE 6. The modelling error obtained for each of the variables in T2 and T3 was of the same order of magnitude, proving that the model is able to represent the system working in conditions that were not tested during the training phase. The error is relatively much larger in variables number 7 and 11 due to the noise content of these particular variables, which is not accurately represented by the model. The total averaged error for the estimations of T2 and T3 was 8.42% and 8.97% respectively.

TABLE 6: AVERAGE NORMALIZED ESTIMATION ERROR FOR EACH MEASURED VARIABLE (%)

Set	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	Avg.
T2	1.86	1.92	2.19	2.16	1.32	4.70	47.37	3.32	2.20	11.13	39.32	4.58	0.64	3.09	3.56	2.05	5.17	8.42
T3	1.93	2.01	3.07	3.00	2.06	6.31	37.00	6.57	5.77	8.68	42.77	3.93	0.81	6.01	7.12	3.13	5.35	8.97

The validated model was used to provide estimations of the process variables assuming normal operation in each one of the faulty cases studied. The model representing the system under normal operation was fed with the same input sequence used during the total duration of each experiment. This allowed the prediction of the system outputs for the same operating conditions that were tested in the faulty cases, but assuming absence of faults. The objective of this analysis is to allow the process operators to evaluate the impact of the fault over the system variables and take into account the effects in terms of safety, efficiency and product quality when scheduling the production and maintenance plans according to the process condition.

Based on the results provided by the contribution plots (TABLE 5) the most affected variables in Case 1 were the pressure in the 3-phase separator PT501 and the position of the valve VC501. Fig. 7 represents the measurements observed for these variables in this faulty case and the estimations obtained using the normal operation model. It can be seen that after the fault introduction the changes in operating conditions generate variations in PT501 (which in normal conditions is maintained at 1barg) and the algorithm predicted the variations in VC501 assuming normal operation. The predictions obtained for both variables before the fault introduction were accurate as expected from the previous model validation results.

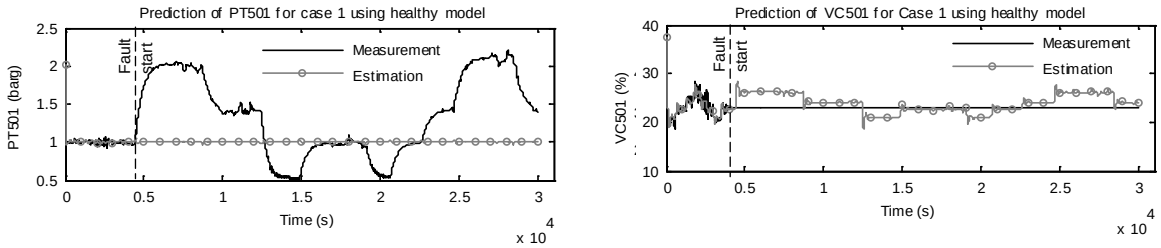

Fig. 7: Performance degradation in PT501 (left) and VC501 (right) in Case 1

The most significant variables for Case 2 in terms of contribution were the top separator pressure PT403 for T^2, and the top separator level LIC405 for Q. The results obtained for the prediction of these variables using a healthy model are represented in Fig. 8. The estimations were again accurate in both cases until the fault was introduced, and the most obvious effect of the fault was an increment in the pressure measured inside the 2-phase separator (PT403). Changes in the behaviour of the level measurement inside the 2-phase separator are not that evident due to the high level of noise in this signal, but a slight increment in the average level over the predicted value can be perceived after the fault introduction.

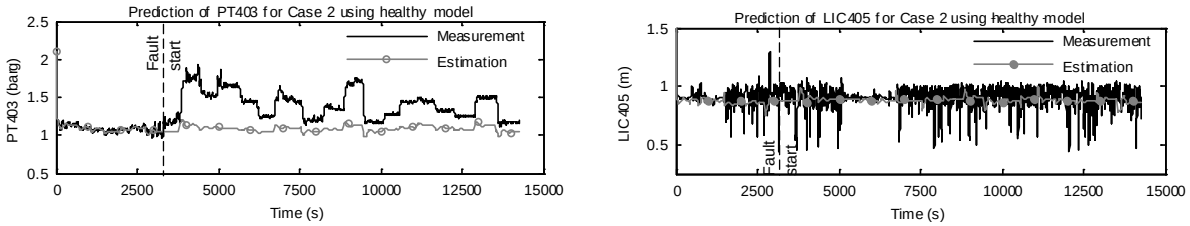

Fig. 8: Performance degradation in PT403 (left) and LIC405 (right) in Case 2

For Case 3, the most significant variables identified were the top riser pressure PT408 and the density measured by FT407 in the top of the riser. Fig. 9 shows the measurements of these two variables and the corresponding estimation using a healthy model. The initial opening of the 2” valve (Case3-1 start) did not have a significant effect on the measured variables, except for a slight increment in the oscillations of the PT408 signal. These oscillations are more evident around sample 5800 in both measurements, probably caused by the change in the water flow rate at that point. Closing VC404 to derive all the flow through the 2” line (Case 3-2 start) had a considerable effect on PT408 in comparison with the normal operation prediction, as well as on the density measurement which acquired an almost constant value of 998kg/m^3 due to the accumulation of liquid in the sensing region.

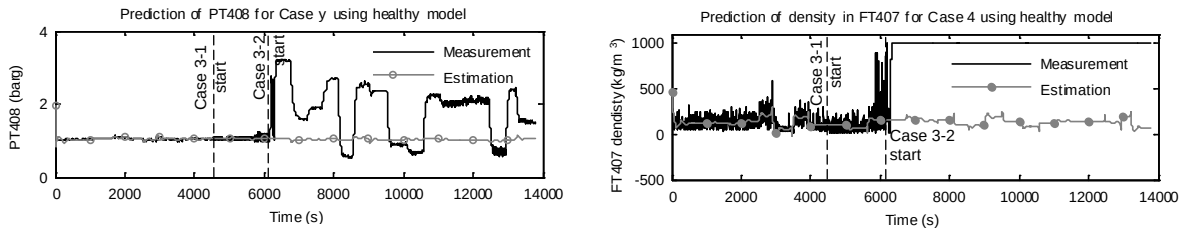

Fig. 9: Performance degradation in PT408 (left) and density measured in FT407 (right) in Case 3

3.4 Prediction of system behaviour under faulty conditions

The analysis carried out in the previous subsections described how process faults can be detected and diagnosed, and how the degradation of the system performance can be evaluated looking at the difference between measurements and model estimations assuming normal operation. If the fault severity is not critical, it can be the case that the optimal maintenance and operation strategy is to continue operating the system under faulty conditions until the next planned shutdown or until the spare parts and repairing equipment are available in the plant. In that case the plant operators will need to know how the faulty process will behave for the future expected operating conditions and determine how this behaviour affects the product quality, the safety of the plant, or the energy consumed by the equipment to evaluate the best way to operate the plant.

The model used in the previous section predicted the system outputs under normal operating conditions, but the faulty system can also be modelled using data acquired during the early stages of degradation once a fault has been detected. The future input sequences can be obtained from the production plan, from forecasts based on

historic data, etc. and then be used in the model to predict the system behaviour under that particular conditions for the time horizon required. As the model described in 2.2 is non-time-variant, this approach will be valid if the fault severity remains stationary, or if it evolves slowly enough that it can be considered stationary during the prediction horizon. For this analysis data acquired from the instant of fault detection was used to build a new model for each one of the faulty cases studied. The amount of data used to build each model needs to be sufficient to represent the system dynamics and produce accurate results. In order to ensure accuracy in the prediction the total averaged error was computed for different lengths of training data. The model was considered valid once the error of the model predicting the same data segment used for training was considered acceptable. Fig. 10 shows the evolution of the total averaged error with the number of samples selected for model training in each case studied. The evolution of the error is similar in all the cases studied; initially the error is large and, as the number of samples used for training is increased, the error converges to values of around 8.5%. These values are similar to the modelling error found under normal operating conditions.

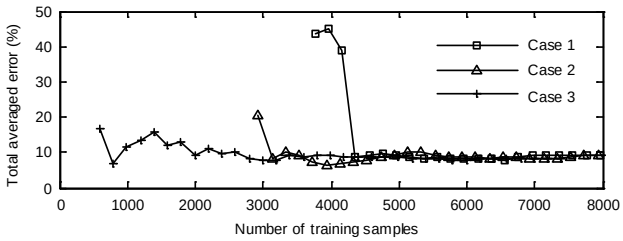

Fig. 10: Evolution of the total averaged error with the number of samples selected for model training

The total averaged error shown in Fig. 10 was calculated as the error of the model predicting the same data set portion used for model training. This allows the operator to evaluate in real time the model ability to perform a reliable prediction as the training set grows. One model was built online for each one of the cases studied using a limited amount of data acquired online after the fault detection. The number of samples used for training was selected according to the prediction error estimations shown in Fig. 10. The prediction horizon matches the number of remaining available samples in each data set, so that the model built with a limited amount of data acquired from the early stages of degradation can be used to predict future process behaviour under different operating conditions. TABLE 7 shows the normalized estimation error calculated for each variable of each case studied. The prediction starting point and the number of samples used to train the model are presented in columns 2 and 3 respectively.

TABLE 7: AVERAGE NORMALIZED ESTIMATION ERROR FOR EACH MEASURED VARIABLE (%)

Case	Prediction start (s)	Training samples	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}	y_{11}	y_{12}	y_{13}	y_{14}	y_{15}	y_{16}	y_{17}	Avg
1	5688	5000	3.05	3.21	4.47	4.41	4.71	2.87	27.54	1.19	1.82	7.85	41.36	2.21	1.17	-	1.54	1.39	5.25	7.13
2	3821	5000	6.19	6.38	8.51	8.28	2.71	2.73	31.55	2.34	0.63	8.93	32.67	1.07	1.75	7.28	2.99	1.50	6.69	7.78
3	6187	3000	5.45	6.21	11.20	2.87	2.61	3.08	64.62	1.36	5.88	14.03	77.46	1.08	1.62	5.99	1.87	5.11	6.80	12.78

All the cases showed similar error rates than the predictions obtained from modelling normal operating conditions, revealing that if enough data is used to train the model it is possible to represent the process behaviour under different operating conditions. Fig. 11 shows the prediction results under faulty conditions of the most significant variables of each case. All the predictions show a similar behaviour to the estimations obtained under normal operation, with some oscillations in the initial estimations but a good accuracy once the transient is extinguished. This result shows that it is possible to perform an accurate prediction using a small amount of data acquired during the early stages of degradation. This prediction will be accurate if the forecast of future inputs is reliable, but in the long term the accuracy of the model will be limited by the evolution of the fault. The predicted outputs can be used by plant operators to understand how the faulty system will react to different operating scenarios, and determine whether it is safe and appropriate to operate the system in these conditions.

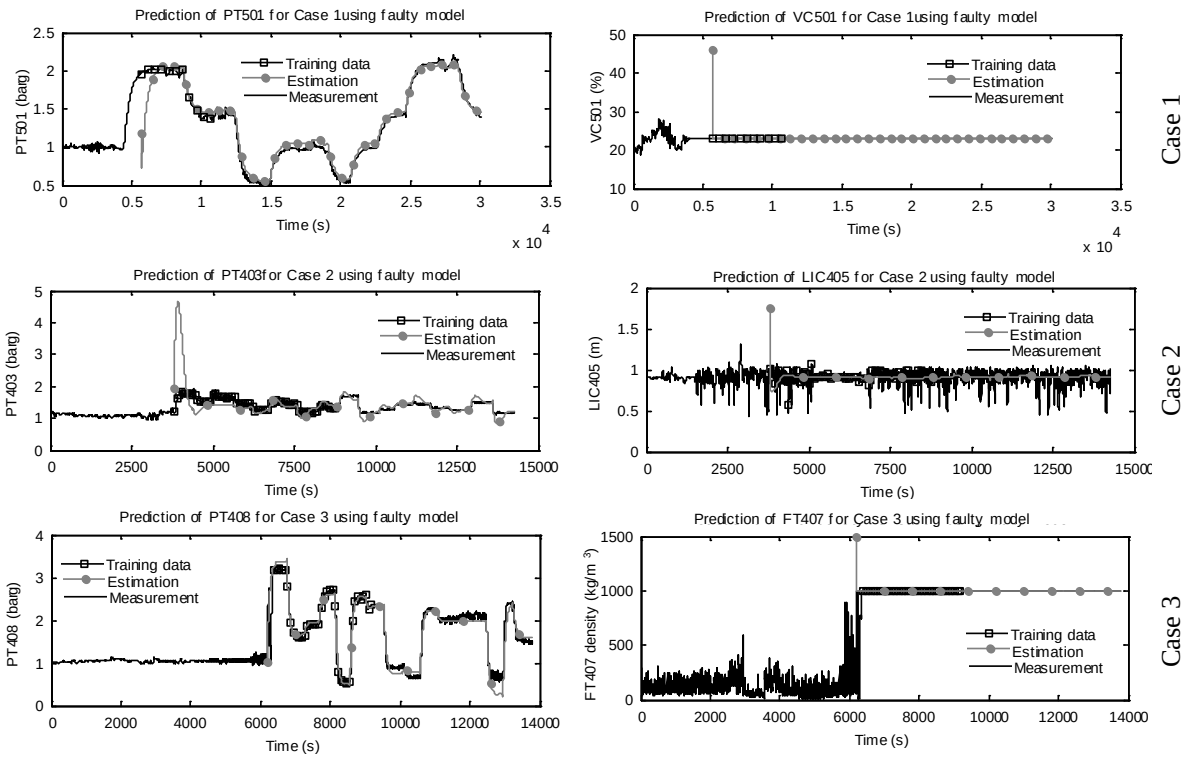

Fig. 11: Summary of prediction results under faulty conditions for the most significant variables in each case

4. Conclusion

Process data was acquired from an experimental large-scale multiphase flow facility to test the capabilities of CVA to estimate performance degradation and predict the behavior of the system working under varying operating conditions. This system is highly non-linear and was operated under changing operational conditions.

The three different faults introduced were successfully detected using the T^2 and Q indicators within a reasonable detection time, low number of false alarms and false negatives. The most significant variables affected by the fault were identified using contribution plots to help in the process of fault diagnosis. Once the faults were detected, CVA was used to build a state-space model that represents the system working under normal operating conditions. This model was used to observe the difference between the process measurements when working under fault conditions and the model estimations assuming normal operation. The estimations obtained were close to the process measurements before the introduction of the faults, and the differences observed afterwards can be used to quantify the impact of the fault on the different process variables. Secondly, a new model was built for each fault case studied using data acquired during the early stages of degradation after the fault introduction. This model was used to predict the faulty process behaviour under varying operating conditions.

The system identification procedure based on CVA was performed in a fast and stable manner, due to the numerical and computational benefits of singular value decomposition. The estimation error was always relatively low, allowing an accurate estimation of all process variables under normal and faulty operation. The model was able to represent the process dynamics working under varying operating conditions. However very fast oscillations and measurements with a high noise level were not accurately estimated. Additionally, the use of a non-time-varying model for prediction of faulty system behaviour limits its application to faults which severity evolves slowly over time compared with the prediction timeframe. The use of a time-varying model can overcome this limitation and will be the object of study in future research.

The results presented in [15] demonstrated that CVA can effectively detect and diagnose faults in real complex systems working under varying operating conditions. The methodology presented in this investigation goes one step further to demonstrate that CVA can also be used for system identification in a large and complex real facility. This technique allowed the estimation of performance degradation, which can be used by plant operators to develop optimal maintenance and production planning techniques that consider the condition of the process. In addition, the behavior of the faulty process was modelled using data acquired during the early stages

of degradation, allowing the operators to predict how the fault will affect the process for future operating conditions. This methodology advances the traditional condition monitoring procedure of fault detection and diagnosis by providing estimations of the impact of the fault on the system behavior. The capacity of using this information to react to faults with the optimal maintenance and production strategies can potentially lead to more efficient, reliable and profitable processes.

Acknowledgements

Financial support from the Marie Curie FP7-ITN project "Energy savings from smart operation of electrical, process and mechanical equipment– ENERGY-SMARTOPS", Contract No: PITN-GA-2010-264940 is gratefully acknowledged.

References

[1] Ge, Z., Song, Z. and Gao, F. (2013), "Review of recent research on data-based process monitoring", *Industrial and Engineering Chemistry Research,* vol. 52, no. 10, pp. 3543-3562.

[2] L.H. Chiang and E.L. Rusell and R.D. Braatz (2000), *Fault detection and diagnosis in industrial systems,* 1st ed, Springer, London UK.

[3] Odiowei, P. P. and Cao, Y. (2010), "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations", *IEEE Transactions on Industrial Informatics,* vol. 6, no. 1, pp. 36-45.

[4] Yang, Y., Chen, Y., Chen, X. and Liu, X. (2012), "Multivariate industrial process monitoring based on the integration method of canonical variate analysis and independent component analysis", *Chemometrics and Intelligent Laboratory Systems,* vol. 116, pp. 94-101.

[5] Chen, R. Q. (2013), "Advances in data-driven monitoring methods for complex process", *3rd International Conference on Applied Mechanics, Materials and Manufacturing, ICAMMM 2013;* Vol. 423-426, 24 August 2013, Dalian; China, pp. 2448.

[6] Pan, H., Wei, X. and Huang, J. (2010), "Fault feature extraction based on KPCA optimized by PSO algorithm", *IEEE International Conference on Industrial Informatics (INDIN),* pp. 102.

[7] Wang, G., Yin, S. and Kaynak, O. (2014), "An LWPR-based data-driven fault detection approach for nonlinear process monitoring", *IEEE Transactions on Industrial Informatics,* vol. 10, no. 4, pp. 2016-2023.

[8] Zhang, Y., Zhou, H., Qin, S. J. and Chai, T. (2010), "Decentralized fault diagnosis of large-scale processes using multiblock kernel partial least squares", *IEEE Transactions on Industrial Informatics,* vol. 6, no. 1, pp. 3-10.

[9] Juricek, B. C., Seborg, D. E. and Larimore, W. E. (2004), "Fault Detection Using Canonical Variate Analysis", *Industrial and Engineering Chemistry Research,* vol. 43, no. 2, pp. 458-474.

[10] Russell, E. L., Chiang, L. H. and Braatz, R. D. (2000), "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis", *Chemometrics and Intelligent Laboratory Systems,* vol. 51, no. 1, pp. 81-93.

[11] Lee, J., Yoo, C. K., Choi, S. W., Vanrolleghem, P. A. and Lee, I. (2004), "Nonlinear process monitoring using kernel principal component analysis", *Chemical Engineering Science,* vol. 59, no. 1, pp. 223-234.

[12] Schaper, C. D., Larimore, W. E., Seborg, D. E. and Mellichamp, D. A. (1994), "Identification of chemical processes using canonical variate analysis", *Computers and Chemical Engineering,* vol. 18, no. 1, pp. 55-69.

[13] Liu, D., Jiang, D., Chen, X., Luo, A. and Xu, G. (2012), "Research on fault identification for complex system based on generalized linear canonical correlation analysis", *2012 IEEE International Conference on Automation Science and Engineering: Green Automation Toward a Sustainable Society, CASE 2012,* 20 August 2012 through 24 August 2012, Seoul, pp. 474.

[14] Borsje, H. J. (1999), "Fault detection in boilers using canonical variate analysis", *Proceedings of the 1999 American Control Conference (99ACC),* Vol. 2, 2 June 1999 through 4 June 1999, San Diego, CA, USA, IEEE, Piscataway, NJ, United States, pp. 1167.

[15] Ruiz-Cárcel, C., Cao, Y., Mba, D., Lao, L. and Samuel, R. T. (2014), "Statistical process monitoring of a multiphase flow facility", *Control Engineering Practice,* vol. 42, pp. 74-88.

[16] Juricek, B. C., Seborg, D. E. and Larimore, W. E. (2002), "Identification of multivariable, linear, dynamic models: Comparing regression and subspace techniques", *Industrial and Engineering Chemistry Research,* vol. 41, no. 9, pp. 2185-2203.

[17] Juricek, B. C., Seborg, D. E. and Larimore, W. E. (2001), "Identification of the Tennessee Eastman challenge process with subspace methods", *Control Engineering Practice,* vol. 9, no. 12, pp. 1337-1351.

[18] Deng, X. and Tian, X. (2011), "A new fault isolation method based on unified contribution plots", *Proceedings of the 30th Chinese Control Conference, CCC 2011,* pp. 4280.

[19] Liu, J. and Chen, D. (2012), "Multiple Sensor Fault Isolation Using Contribution Plots without Smearing Effect to Non-Faulty Variables", *Computer Aided Chemical Engineering,* vol. 31, pp. 1517-1521.

[20] Ramírez, A. W. and Llinàs, J. C. (2011), "Fault diagnosis of batch processes release using PCA contribution plots as fault signatures", *ICEIS 2011 - Proceedings of the 13th International Conference on Enterprise Information Systems,* Vol. 1 DISI, pp. 223.

[21] Larimore, W. E. (1990), "Canonical variate analysis in identification, filtering, and adaptive control", *Proceedings of the IEEE Conference on Decision and Control,* Vol. 2, pp. 596.

[22] Emerson Electric Co. , *The DeltaV Digital Automation System* , available at: http://www2.emersonprocess.com/en-us/brands/deltav/pages/index.aspx (accessed 13 February 2014).

[23] Negiz, A. and Çinar, A. (1998), "Monitoring of multivariable dynamic processes and sensor auditing", *Journal of Process Control,* vol. 8, no. 5-6, pp. 375-380.

Cranfield University

CERES Research Repository **https://dspace.lib.cranfield.ac.uk/**

School of Aerospace, Transport and Manufacturing (SATM) Staff publications (SATM)

Canonical variate analysis for performance degradation under faulty conditions

Ruiz Cárcel, Cristóbal

2016-06-01

Attribution-NonCommercial-NoDerivatives 4.0 International

C. Ruiz-Cárcel, L. Lao, Y. Cao, D. Mba, Canonical variate analysis for performance degradation under faulty conditions, Control Engineering Practice, Volume 54, September 2016, pp70-80
http://dx.doi.org/10.1016/j.conengprac.2016.05.018

Downloaded from CERES Research Repository, Cranfield University