Data-Driven Fault Detection Scheme Using Errors-in-Variables Subspace Identification

J. Mina and S. Joe Qin
The Mork Family Department of Chemical Engineering and Materials Science
University of Southern California
925 Bloom Walk, HED 211, Los Angeles , CA 90089-1211
jmina@usc.edu, sqin@usc.edu
Tel: (1)-213-7404203

***Resumen*— This work presents a data-driven fault detection scheme designed in the framework of the so called parity space approach. The idea makes use of subspace identification methods to identify directly the residual generator instead of the process model. Particularly, it considers the identification of the residual generator from noisy input-output data using an errors-in-variables subspace identification method. The proposed scheme is illustrated by simulations in an interconnected tanks system.**

Keywords: Fault Detection, Parity Space, Subspace Model Identification, DPCA.

I. Introduction

The use of multivariate statistical methods such as principal component analysis (PCA) for fault detection and diagnosis has received significant attention in the last years. Originally proposed as a quality monitoring technique in (Jackson, 1991) and process monitoring method in (Kresta *et al.*, 1991), these methods are applied to a variety of process monitoring problems and sensor validation based on normal process data.

Given that the multivariate statistical monitoring methods make use of the correlation among the process variables, the statistical models are inherently steady state characterizations of the process which can cause false alarms in a dynamically operated process. To deal with dynamic processes, dynamic PCA (DPCA) is discussed by (Ku *et al.*, 1995) where time lagged variables are included in PCA. This extension, however, gives poor dynamic models as demonstrated in (Negiz and Çinar, 1997) and (Li and Qin, 2001).

To build a sound dynamic model from process data, subspace identification methods (SIM) are promising alternatives to PCA (Van Overschee and De Moor, 1996). Among these SIM algorithms are the canonical variate analysis, CVA (Larimore, 1990), N4SID (Van Overschee and De Moor, 1994), MOESP (Verhaegen and Dewilde, 1992), and the use of PCA, SIMPCA (Wang and Qin, 2002), (Wang and Qin, 2006). SIM offers consistent estimate of state-space models, e.g., the state space matrices A, B, C, D for multivariable dynamic systems with proper selection of the system order. A typical SIM contains two steps: (1) identification of the extended observability matrix and/or the state sequence; and (2) calculation of A, B, C, D.

While most SIM approaches appear as numerical algorithms, statistical properties such as consistency have been explored (Deistler *et al.*, 1995), (Jansson and Wahlberg, 1998), (Heij and Scherrer, 1999); as well as the error-in-variables (EIV) problem (Chou and Verhaegen, 1997), (Li and Qin, 2001).

The SIM formulation is attractive not only because of its numerical simplicity and stability, but also for its general state space form. More importantly, the SIM state space formulation allows the implementation of fault detection schemes ranging from totally model-based to totally data-based, which provides great flexibility for comparing and analyzing various methods.

The objective of this paper is to develop a data driven method for fault detection that uses a subspace identification algorithm under the EIV formulation. We propose to use PCA with instrumental variables to eliminate input and output errors.

Here, a direct estimation of the residual generator for fault detection purposes is obtained based on capturing the most of the normal process variations. This 'model' should be interpreted differently from the standard subspace identification methods which seek to estimate an unbiased model and requires external perturbation to excite all process dynamics.

The paper is organized as follows. Section II describes the parity space approach for fault detection. Section III presents the identification of the residual generator using SIM. Section IV shows simulation results on the interconnected tanks system. Finally, Section V gives conclusions to the paper.

II. Parity Space Approach for Fault Detection

II-A. Preliminaries

Although our method is data-driven, we start the process description with a state space model in EIV forms. Assume the model of the process is represented by the following LTI state-space model

$$\begin{aligned} \mathbf{x}(k+1) = &\quad \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}^0(k) + \omega(k) \\ \mathbf{y}^0(k) = &\quad \mathbf{C}\mathbf{x}(k) + \mathbf{D}\mathbf{u}^0(k) \end{aligned} \tag{1}$$

where $\mathbf{u}^0(k) \in \Re^l$, $\mathbf{y}^0(k) \in \Re^s$, $\mathbf{x}(k) \in \Re^n$, $\omega(k) \in \Re^n$ are noise-free inputs, noise-free outputs, state variables and process noise, respectively. Matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are system matrices with appropriate dimensions.

The available observations are input and output measurements $\mathbf{u}(k)$ and $\mathbf{y}(k)$:

$$\begin{aligned} \mathbf{u}(k) = \mathbf{u}(k)^0 + \eta(k) \\ \mathbf{y}(k) = \mathbf{y}(k)^0 + \nu(k) \end{aligned} \tag{2}$$

where $\eta(k) \in \Re^l$, $\nu(k) \in \Re^s$ are input and output noise.

We introduce the following assumptions:

1. All, the process noise $\omega(k)$, the measurement noise $\eta(k)$ and $\nu(k)$ are white noise, and they are statistically independent of the past noise-free input $\mathbf{u}^0(k)$, i.e.,

$$\mathbf{E}\left\{ \mathbf{u}^0(k) \begin{bmatrix} \omega(j) \\ \eta(j) \\ \nu(j) \end{bmatrix}^T \right\} = \mathbf{0} \quad \text{for } j \geq k \tag{3}$$

2. The three white-noise sequences can be correlated and their covariance is given by the following unknown matrix

$$\begin{aligned} &\mathbf{E}\left\{ \begin{bmatrix} \omega(k) \\ \eta(k) \\ \nu(k) \end{bmatrix} \begin{bmatrix} \omega(j)^T & \eta(j)^T & \nu(j)^T \end{bmatrix} \right\} \\ &= \begin{bmatrix} \mathbf{\Sigma}_{\omega\omega} & \mathbf{\Sigma}_{\omega\eta} & \mathbf{\Sigma}_{\omega\nu} \\ \mathbf{\Sigma}_{\omega\eta}^T & \mathbf{\Sigma}_{\eta\eta} & \mathbf{\Sigma}_{\eta\nu} \\ \mathbf{\Sigma}_{\omega\nu}^T & \mathbf{\Sigma}_{\eta\nu}^T & \mathbf{\Sigma}_{\nu\nu} \end{bmatrix} \delta_{kj} \end{aligned} \tag{4}$$

   where δ_{kj} is the Kronecker delta function.

By manipulating (1) and (2) we obtain

$$\mathbf{y}_f(k) = \mathbf{\Gamma}_f \mathbf{x}_f(k) + \mathbf{H}_f^d \mathbf{u}_f(k) - \mathbf{H}_f^d \eta_f(k) + \mathbf{H}_f^s \omega_f(k) + \nu_f(k) \tag{5}$$

where for an arbitrary time point k taken as the current time, the following vectors and matrices are defined.

$$\mathbf{y}_f(k) = \begin{bmatrix} \mathbf{y}(k) \\ \mathbf{y}(k+1) \\ \vdots \\ \mathbf{y}(k+f-1) \end{bmatrix} \in \Re^{sf} \tag{6}$$

is the future output extended vector. The vectors $\mathbf{u}_f(k) \in \Re^{lf}$, $\eta_f(k) \in \Re^{lf}$, $\omega_f(k) \in \Re^{nf}$ and $\nu_f(k) \in \Re^{sf}$ are defined similar to $\mathbf{y}_f(k)$. Additionally,

$$\mathbf{\Gamma}_f = \begin{bmatrix} \mathbf{C} \\ \mathbf{CA} \\ \vdots \\ \mathbf{CA}^{f-1} \end{bmatrix} \in \Re^{sf \times n} \tag{7}$$

is the extended observability matrix with $rank(\mathbf{\Gamma}_f) = n$, and

$$\mathbf{H}_f^d = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{CB} & \mathbf{D} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{f-2}\mathbf{B} & \mathbf{CA}^{f-3}\mathbf{B} & \cdots & \mathbf{D} \end{bmatrix} \in \Re^{sf \times lf} \tag{8}$$

$$\mathbf{H}_f^s = \begin{bmatrix} \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{CA}^{f-2} & \mathbf{CA}^{f-3} & \cdots & \mathbf{0} \end{bmatrix} \in \Re^{sf \times nf}, \tag{9}$$

are two block Toeplitz matrices.

II-B. Parity Space based Residual Generator

Introducing

$$\mathbf{z}_f(k) = \begin{bmatrix} \mathbf{y}_f(k) \\ \mathbf{u}_f(k) \end{bmatrix} \in \Re^{sf+lf}, \tag{10}$$

we can rewrite (5) into

$$\begin{bmatrix} \mathbf{I} & -\mathbf{H}_f^d \end{bmatrix} \mathbf{z}_f(k) = \mathbf{\Gamma}_f \mathbf{x}_f(k) - \mathbf{H}_f^d \eta_f(k) + \mathbf{H}_f^s \omega_f(k) + \nu_f(k) \tag{11}$$

To use (11) for fault detection, it is necessary to eliminate the state vector $\mathbf{x}_f(k)$. In the conventional approach of parity space (Chow and Willsky, 1984), the state is eliminated by pre-multiplying (11) by $\mathbf{\Gamma}_f^{\perp} \in \Re^{sf \times (sf-n)}$ which is the orthogonal complement of $\mathbf{\Gamma}_f$ and satisfies

$$\begin{aligned} rank\left(\mathbf{\Gamma}_f^{\perp}\right) = sf - n \\ \mathbf{\Gamma}_f^{\perp T} \mathbf{\Gamma}_f = \mathbf{0} \end{aligned} \tag{12}$$

The residual vector is then defined as

$$\mathbf{G}\mathbf{z}_f(k) = \mathbf{r}(k) \in \Re^{(sf-n)} \tag{13}$$

where

$$\mathbf{G} \equiv \mathbf{\Gamma}_f^{\perp T} \begin{bmatrix} \mathbf{I} & -\mathbf{H}_f^d \end{bmatrix}, \tag{14}$$

$\mathbf{G} \in \Re^{(sf-n)\times(sf+lf)}$, is the named *residual generator* and

$$\mathbf{r}(k) \equiv \mathbf{\Gamma}_f^{\perp T} \left(-\mathbf{H}_f^d \eta_f(k) + \mathbf{H}_f^s \omega_f(k) + \nu_f(k) \right) \tag{15}$$

is the residual which under nominal conditions depends only in the noise terms.

II-C. *Monitoring Index for Fault Detection*

If a sensor is faulty, its measurement will contain the normal values of the process variables and the fault component, this is represented as follows

$$\mathbf{z}_f(k) = \mathbf{z}_f^*(k) + \Xi_i \mathbf{f}_i(k) \tag{16}$$

where $\mathbf{z}_f^*(k)$ is the fault free portion of the variables, $\Xi_i \in \Re^{(sf+lf)\times l_i}$ is an orthogonal matrix representing the fault direction, and $\mathbf{f}_i(k) \in \Re^{l_i}$ is the fault magnitude vector.

Therefore, substituting (16) in (13), under fault conditions the residual results in

$$\mathbf{r}(k) = \mathbf{G}\left(\mathbf{z}_f^*(k) + \Xi_i \mathbf{f}_i(k)\right) = \mathbf{r}^*(k) + \mathbf{G}\Xi_i \mathbf{f}_i(k) \tag{17}$$

where

$$\mathbf{r}^*(k) = \mathbf{G}\mathbf{z}_f^*(k) = \mathbf{r}(k)|_{\mathbf{f}_i=\mathbf{0}} \tag{18}$$

is the residual under normal conditions.

Since $\mathbf{r}^*(k)$ contains the measurement terms which are Gaussian, then $\mathbf{r}^*(k)$ is also Gaussian (Anderson, 1984)

$$\mathbf{r}^*(k) \sim \aleph(\mathbf{0}, \mathbf{R}_{r^*}) \tag{19}$$

where $\aleph(\mathbf{0}, \mathbf{R}_{r^*})$ denotes a Gaussian distribution with mean $\mathbf{0}$ and covariance $\mathbf{R}_{r^*}$.

According with (Qin and Li, 2001) it can be defined a fault detection index for the residuals obtained in the parity space. So, given that $\mathbf{r}^*(k)$ is Gaussian, consequently

$$\mathbf{r}^{*T}(k)\mathbf{R}_{r^*}^{-1}\mathbf{r}^*(k) \sim \chi^2(sf-n) \tag{20}$$

where $\chi^2(sf-n)$ denotes the Chi square distribution with $(sf-n)$ degrees of freedom; by the other side, $\mathbf{R}_{r^*}$ is the covariance matrix of $\mathbf{r}^*(k)$ which according with (18), it is defined as

$$\mathbf{R}_{r^*} = cov\left(r^*(k)\right) = cov\left(\mathbf{G}\mathbf{z}_f^*(k)\right) = \mathbf{G}\mathbf{R}_{z^*}\mathbf{G}^T \tag{21}$$

where

$$\mathbf{R}_{z^*} \equiv E\left[\mathbf{z}_f^*(k)\mathbf{z}_f^{*T}(k)\right], \tag{22}$$

is the covariance matrix of $\mathbf{z}_f^*(k)$ and can be estimated from process normal data.

In conclusion, the fault detection index for the residual space is defined as follows

$$\begin{aligned} d(k) &= \mathbf{r}^T(k)\mathbf{R}_{r^*}^{-1}\mathbf{r}(k) \\ &= \mathbf{z}_f^T(k)\mathbf{G}^T\left(\mathbf{G}\mathbf{R}_{z^*}\mathbf{G}^T\right)^{-1}\mathbf{G}\mathbf{z}_f(k) \end{aligned} \tag{23}$$

whose nominal threshold UCL (upper control limit) is $UCL = \chi^2_\alpha(sf-n)$, where α is the confidence level. The reasoning for the fault detection is: if for a new sample $\mathbf{z}_f(k)$, $d(k) \le UCL$, then the process is in normal conditions, otherwise, a fault has been detected.

III. Identification of the Residual Generator

If the process matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ are known then the residual generator given in (14) can be constructed; however, if there is no knowledge of the system model we could resort to subspace identification methods.

In general, the SIM's make use of geometric tools like orthogonal or oblique projection, singular value decomposition, to determine the order, the observability matrix and/or the state sequence; then the extraction of matrices $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ is achieved through the solution of a least squares problem. However, for the purpose of fault detection we only need to identify directly the residual generator $\mathbf{G}$.

To identify $\mathbf{G}$ consistently from noisy input-output observations we use EIV subspace identification methods, specifically the SIMPCA-Wc proposed by (Wang and Qin, 2002), (Wang and Qin, 2006) to estimate $\mathbf{H}_f^d$ and $\mathbf{\Gamma}_f^\perp$.

From (13)-(15) and using the data matrices instead of the data vectors $\{\mathbf{z}_f, \eta_f, \omega_f, \nu_f\}$ it results

$$\mathbf{\Gamma}_f^{\perp T}\begin{bmatrix}\mathbf{I} & -\mathbf{H}_f^d\end{bmatrix}\mathbf{Z}_f = \mathbf{\Gamma}_f^{\perp T}\left(-\mathbf{H}_f^d\mathbf{\Upsilon}_f + \mathbf{H}_f^s\mathbf{\Omega}_f + \mathbf{V}_f\right) \tag{24}$$

where e.g.

$$\mathbf{Z}_f = \begin{bmatrix}\mathbf{z}_f(k) & \cdots & \mathbf{z}_f(k+N-1)\end{bmatrix} \in \Re^{(sf+lf)\times N} \tag{25}$$

Lets define the past measurements matrix $\mathbf{Z}_p$ as

$$\mathbf{Z}_p = \left[\begin{array}{c|c|c}\mathbf{y}_p(k) & \cdots & \mathbf{y}_p(k+N-1) \\ \mathbf{u}_p(k) & & \mathbf{u}_p(k+N-1)\end{array}\right] \in \Re^{(sp+lp)\times N} \tag{26}$$

where

$$\mathbf{y}_p(k) = \begin{bmatrix}\mathbf{y}(k-p) \\ \mathbf{y}(k-p+1) \\ \vdots \\ \mathbf{y}(k-1)\end{bmatrix} \in \Re^{sp} \tag{27}$$

is the past output extended vector. The vector $\mathbf{u}_p(k) \in \Re^{lp}$ is defined similar to $\mathbf{y}_p(k)$.

To eliminate the effect of noise asymptotically in (24) it is proposed the use of $\mathbf{Z}_p$ as instrumental variables given that the future noises $\{\mathbf{\Upsilon}_f, \mathbf{\Omega}_f, \mathbf{V}_f\}$ are independent of past data, (Chou and Verhaegen, 1997). So, post-multiplying (24) by $\frac{1}{N}\mathbf{Z}_p^T\mathbf{W}_c$ where

$$\mathbf{W}_c = \left(\frac{1}{N}\mathbf{Z}_p\mathbf{Z}_p^T\right)^{-\frac{1}{2}}, \tag{28}$$

it results

$$\mathbf{\Gamma}_f^{\perp T}\begin{bmatrix}\mathbf{I} & -\mathbf{H}_f^d\end{bmatrix}\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T\mathbf{W}_c = \mathbf{0} \tag{29}$$

The purpose of the weighting matrix $\mathbf{W}_c$ is in order to achieve the normalization of the instrumental variables.

Performing PCA on $\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T\mathbf{W}_c$ we obtain

$$\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T\mathbf{W}_c = \mathbf{P}\mathbf{T}^T + \widetilde{\mathbf{P}}\widetilde{\mathbf{T}}^T \tag{30}$$

and given that $\mathbf{\Gamma}_f^{\perp T}\begin{bmatrix}\mathbf{I} & -\mathbf{H}_f^d\end{bmatrix}$ is in the left null space of $\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T\mathbf{W}_c$ with

$$rank\left(\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T\mathbf{W}_c\right) = lf + n$$

then

$$\begin{bmatrix}\mathbf{\Gamma}_f^{\perp} \\ -\mathbf{H}_f^{d\,T}\mathbf{\Gamma}_f^{\perp}\end{bmatrix} = \widetilde{\mathbf{P}}\mathbf{M} = \begin{bmatrix}\widetilde{\mathbf{P}}_y \\ \widetilde{\mathbf{P}}_u\end{bmatrix}\mathbf{M} \tag{31}$$

where $\widetilde{\mathbf{P}} \in \Re^{(sf+lf)\times(sf-n)}$ and $\mathbf{M} \in \Re^{(sf-n)\times(sf-n)}$ is a nonsingular matrix.

By manipulating (31) the following relation is obtained which is independent of $\mathbf{M}$

$$-\widetilde{\mathbf{P}}_y^T\mathbf{H}_f^d = \widetilde{\mathbf{P}}_u^T \tag{32}$$

So, to estimate $\mathbf{H}_f^d$ from (32) this is rearranged as

$$\begin{aligned} -\widetilde{\mathbf{P}}_y^T &= \mathbf{\Phi} = \begin{bmatrix}\mathbf{\Phi}_1 & \mathbf{\Phi}_2\cdots\mathbf{\Phi}_f\end{bmatrix} \quad ; \quad \mathbf{\Phi}_i \in \Re^{(sf-n)\times s} \\ \widetilde{\mathbf{P}}_u^T &= \mathbf{\Psi} = \begin{bmatrix}\mathbf{\Psi}_1 & \mathbf{\Psi}_2\cdots\mathbf{\Psi}_f\end{bmatrix} \quad ; \quad \mathbf{\Psi}_i \in \Re^{(sf-n)\times l} \end{aligned} \tag{33}$$

thus, (32) becomes

$$\begin{bmatrix}\mathbf{\Phi}_1 & \mathbf{\Phi}_2\cdots\mathbf{\Phi}_f\end{bmatrix}\mathbf{H}_f^d = \begin{bmatrix}\mathbf{\Psi}_1 & \mathbf{\Psi}_2\cdots\mathbf{\Psi}_f\end{bmatrix} \tag{34}$$

Lets take the first block column of $\mathbf{H}_f^d$, see (8),

$$\mathbf{H}_{f_1}^d = \begin{bmatrix}\mathbf{D} \\ \mathbf{CB} \\ \vdots \\ \mathbf{CA}^{f-2}\mathbf{B}\end{bmatrix} \tag{35}$$

which can be estimated by least squares by rearranging (34) as an overdetermined system as follows

$$\begin{bmatrix}\mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \cdots & \mathbf{\Phi}_f \\ \mathbf{\Phi}_2 & \mathbf{\Phi}_3 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Phi}_f & \mathbf{0} & \cdots & \mathbf{0}\end{bmatrix}\mathbf{H}_{f_1}^d = \begin{bmatrix}\mathbf{\Psi}_1 \\ \mathbf{\Psi}_2 \\ \vdots \\ \mathbf{\Psi}_f\end{bmatrix} \tag{36}$$

Once $\widehat{\mathbf{H}}_{f_1}^d$ is estimated, $\widehat{\mathbf{H}}_f^d$ can be easily constructed.

By the other side, once $\widehat{\mathbf{H}}_f^d$ is known it can be defined

$$\mathbf{\Gamma}_f^{\perp T}\begin{bmatrix}\mathbf{I} & -\widehat{\mathbf{H}}_f^d\end{bmatrix}\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T = \mathbf{0} \tag{37}$$

and applying a SVD on $\begin{bmatrix}\mathbf{I} & -\widehat{\mathbf{H}}_f^d\end{bmatrix}\frac{1}{N}\mathbf{Z}_f\mathbf{Z}_p^T$ then $\widehat{\mathbf{\Gamma}}_f^{\perp}$ can be obtained as the left null space.

Finally, the residual generator $\widehat{\mathbf{G}}$ will be given as

$$\widehat{\mathbf{G}} = \left(\widehat{\mathbf{\Gamma}}_f^{\perp}\right)^T\begin{bmatrix}\mathbf{I} & -\widehat{\mathbf{H}}_f^d\end{bmatrix} \tag{38}$$

In the next section, the effectiveness of the proposed fault detection scheme is demonstrated by simulations on a three tanks hydraulic system.

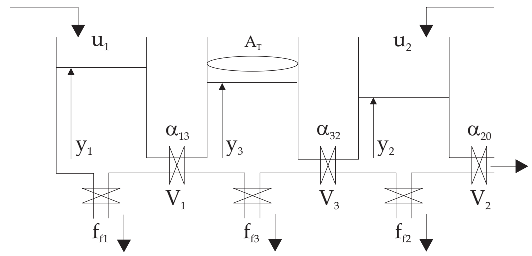

Figura 1. Interconnected Tanks System

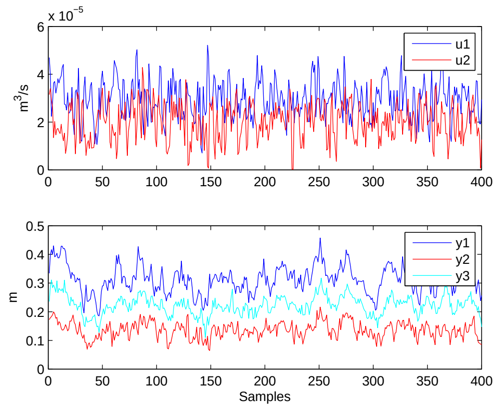

Figura 2. Nominal Input-Output Variables

IV. Interconnected Tanks System - Case Study

The interconnected tanks hydraulic system, described in Fig. 1, is composed of three cylindrical tanks, interconnected at the bottom by pipes with valves V_1 in the link between tanks 1 and 3, V_3 in the link between tanks 3 and 2, and V_2 in the link between tank 2 and the outside, which can be manipulated to emulate faults (e.g. pipe blockage). The system is feed by two inputs u_1 to the tank 1 and u_2 to the tank 2 which are measured as well as the output variables y_1, y_2 and y_3 which correspond to tanks levels. Additionally, each of the tanks is provided with a valve in the bottom which can be manipulated to emulate leakages, f_{fi}'s.

For the experiments, the system is simulated around the following operating point

Mean Values	Variance
$u_1 = 3\times 10^{-5}m^3/s$	$\sigma_{u_1}^2 = 6.5\times 10^{-11}$
$u_2 = 2\times 10^{-5}m^3/s$	$\sigma_{u_2}^2 = 6.5\times 10^{-11}$
$y_1 = 0.310m$	$\sigma_{y_1}^2 = 2.4\times 10^{-3}$
$y_2 = 0.130m$	$\sigma_{y_2}^2 = 8\times 10^{-4}$
$y_3 = 0.220m$	$\sigma_{y_3}^2 = 1.1\times 10^{-3}$

the value of the variances determines the normal variation in the input-output measurements. The nominal parameters considered are $\alpha_{13} = 1.002\times 10^{-4}$, $\alpha_{32} = 1.027\times 10^{-4}$ and $\alpha_{20} = 1.360\times 10^{-4}$; where α_{ij} is the flow constant for the corresponding pipe between tanks i and j. Additionally

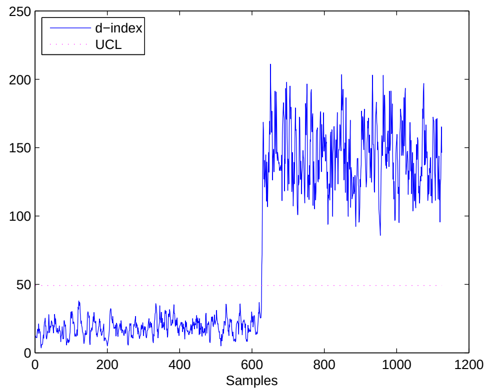

Figura 3. Bias in output sensor y_3

to the normal variations, the input and output variables are subject to zero mean Gaussian white noise.

Using the EIV subspace identification algorithm described in Section III, the residual generator $\mathbf{G}$ is identified from a set of 400 nominal observations of the variables measured every 60s, see Fig. 2. The future and past horizon are chosen as: $f = 7$ and $p = 20$. The variables are scaled to zero mean and unit variance to apply equal weighting to all variables. The control limit for a confidence level $\alpha = 0.01$ is $UCL = 49.18$.

The following operating conditions were evaluated:

1. Fault condition, bias of magnitude $0.1m$ in output sensor y_3. The fault occurrence time is at $40 \times 10^3 s$.
2. Fault condition, bias of magnitude $0.015l/s$ in actuator u_2. The fault occurrence time is at $40 \times 10^3 s$.
3. Fault condition, blockage of magnitude $20\,\%$ in the pipe which links tanks 1 and 3. The fault occurrence time is at $40 \times 10^3 s$.
4. Fault condition, leakage of magnitude $70\,\%$ in thank 1. The fault occurrence time is at $40 \times 10^3 s$.

The simulation results in figures 3-6, shows the effectiveness of the proposed scheme to detect not only sensor faults, but also process faults. It is important to note that for the simulations the fault magnitudes have been chosen with such values taken into account the high variability considered in the variables, see Fig. 2.

V. CONCLUSIONS

A data-driven fault detection scheme designed in the framework of the so called parity space approach has been presented. Here it is presented a method to directly identify the residual generator from noisy input-output data using an errors-in-variables subspace identification method, specifically, PCA with instrumental variables. The proposed scheme is illustrated by simulations in an

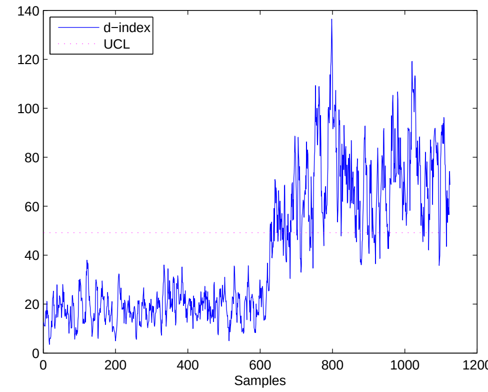

Figura 4. Bias in actuator u_2

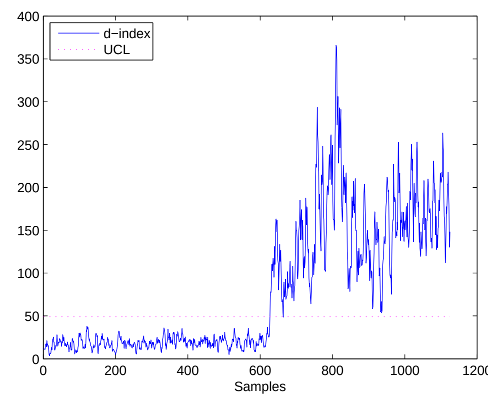

Figura 5. Blockage in the pipe which links tanks 1 and 3

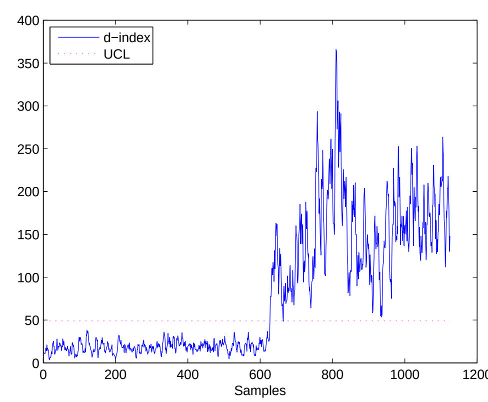

Figura 6. Leakage in tank 1

interconnected tanks system.

The use of multivariate statistical methods like PCA in the classical approach described by the process monitoring community has been successful for the fault detection task but is limited for the fault isolation task. In the case of fault diagnosis schemes based in the SIM formulation the fault isolation task can be easily achieved resorting to the conventional approaches of FDI based in models given that the SIM is able to get a model of the system under monitoring.

VI. ACKNOWLEDGEMENTS

J. Mina gratefully acknowledge the postdoctoral financial support by CONACyT Mexico.

REFERENCIAS

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. Wiley. New york.

Chou, C. T. and M. Verhaegen (1997). Subspace Algorithms for the Identification of Multivariable Dynamic Errors-in-Variables Models. *Automatica* **33**(10), 1857–1869.

Chow, E. Y. and A. S. Willsky (1984). Analytical Redundancy and the Design of Robust Failure Detection Systems. *IEEE Transactions on Automatic Control* **7**, 603–614.

Deistler, M., K. Peternell and W. Scherrer (1995). Consistency and Relative Efficiency of Subspace Methods. *Automatica* **31**, 1865–1875.

Heij, C. and W. Scherrer (1999). Consistency of System Identification by Global Total Least Squares. *Automatica* **35**, 993–1008.

Jackson, J. E. (1991). *A Users Guide to Principal Components*. John Wiley. New york.

Jansson, M. and B. Wahlberg (1998). On Consistency of Subspace Methods for System Identification. *Automatica* **32**(12), 1507–1519.

Kresta, J. V., J. F. MacGregor and T. E. Marlin (1991). Multivariate Statistical Monitoring of Process Operating Performance. *The Canadian Journal of Chemical Engineering* **69**, 35–47.

Ku, W., R. H. Storer and Ch. Georgakis (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* **30**(1), 179–196.

Larimore, W. E. (1990). Canonical Variate Analysis in Identification, Filtering, and Adaptive control. In: *Proceedings of the 29th IEEE Conference on Decision and Control, CDC'90*. Honolulu, Hawaii. pp. 596–604.

Li, W. and S. J. Qin (2001). Consistent Dynamic PCA Based on Errors in Variables Subspace Identification. *Journal of Process Control* **11**(6), 661–678.

Negiz, A. and A. Çinar (1997). Statistical Monitoring of Multivariate Dynamic Processes with State-Space Models. *AIChE Journal* **43**(8), 2002–2020.

Qin, S. J. and W. Li (2001). Detection and Identification of Faulty Sensors in Dynamic Processes. *AIChE Journal* **47**(7), 1581–1593.

Van Overschee, P. and B. De Moor (1994). N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Sstochastic Systems. *Automatica* **30**, 75–93.

Van Overschee, P. and B. De Moor (1996). *Subspace Identification for Linear Systems: Theory, Implementation and Applications*. Springer.

Verhaegen, M. and P. Dewilde (1992). Subspace Model Identification. Part I: The Output-Error State-Space Model Identification Class of Algorithms. **56**, 1187–1210.

Wang, J. and S. J. Qin (2002). A New Subspace Identification Approach Based on Principal Component Analysis. *Journal of Process Control* **12**(8), 841–855.

Wang, J. and S. J. Qin (2006). Closed-Loop Subspace Identification Using the Parity Space. *Automatica* **42**, 315–320.