Identification Techniques for Chemical Process Fault Diagnosis

Silvio Simani

Abstract— **The paper presents the application results concerning the fault diagnosis of a dynamic process using dynamic system identification and model–based residual generation techniques. The first step of the considered approach consists of identifying different families of models for the monitored system. In particular, it is selected the most accurate identified model able to describe in the best way the dynamic behaviour of the considered process. The next step of the fault diagnosis scheme requires the design of output estimators e.g., dynamic observers or Kalman filters) which are used as residual generators. The proposed fault diagnosis and identification scheme has been tested on a real chemical process in the presence of both sensor, actuator, component faults and disturbance. The results and concluding remarks have been finally reported.**

I. INTRODUCTION

Since the early 1970's, the problem of reliable fault diagnosis in dynamic processes has received great attention and a wide variety of robust approaches has been proposed and developed. Recently, different analytical redundancy–based methods have been developed to diagnose faults in linear, time-invariant, dynamic systems and a wide variety of model–based approaches has been proposed [1].

There are different model–based approaches to the fault diagnosis problem [2], namely parameter identification [3], parity equations [4], methods in frequency [5] or in state–space domain, such as diagnosis observers [6] and Kalman filters [7].

Even if analytical redundancy methods have been recognised as a powerful and effective technique for detecting faults, the generation of robust residuals is a critical issue because of the presence of unavoidable modelling uncertainty. The main problem regarding the reliability of fault diagnosis schemes consists of the modelling uncertainties which are due, for example, to process noise, parameter variations and non–linearities.

Model–based methods use a model of the monitored process in order to produce the symptom or residual generator. If the system is not complex and can be described accurately by the mathematical model, fault detection is directly performed by using a simple geometrical analysis of residuals.

In real industrial systems however, since the modelling uncertainty is unavoidable, the design of a robust fault diagnosis scheme should consider the modelling uncertainty with respect to the sensitivity of the faults. Several papers addressed this problem. For example, optimal robust parity relations were proposed in [4], and the threshold selector concept was introduced in [8].

S. Simani is with the Dipartimento di Ingegneria of the Università di Ferrara, Via Saragat, 1. 44100 Ferrara - ITALY. ssimani@ing.unife.it

One other promising approach is the decoupling between disturbances and residuals achieved by means of a proper observer scheme and design [9], [1]. This approach requires the knowledge of a model of the process under investigation and, in particular, of the disturbance distribution matrix. Thus, modelling [9] or identification [10], [11], [12] procedures can be defined to estimate the disturbance distribution matrix.

This work aims to define a comprehensive methodology for the diagnosis of actuator, component and sensor fault of an industrial process by using an output estimation approach [11], in conjunction with residual processing schemes which may include a simple threshold detection [9].

Two main aspects of the proposed methodology should be underlined. Firstly, the Fault Detection and Isolation (FDI) model–based approach does not require any physical knowledge of the process under observation. A linear mathematical model of the input–output links are, in fact, obtained by means of identification schemes which use Equation Error (EE), Errors–In–Variables (EIV) and State–Space (SS) models [10], [11].

In the case of the EIV identification technique, it is based on the Frisch scheme methodology [13], [11]. This approach gives a reliable model of the plant under investigation, as well as providing variances of the input-output noises [14], [11]. Secondly, in this work linear prototypes for the design of linear output estimators [11] have been developed instead of complicated non–linear models obtained by modelling techniques in connection with non–linear observers. In fact, as the feature of system supervision is to monitor the operation and performance of the system with respect to an expected point of operation, linear system methods are still very valid.

In particular, in this paper, the complete procedures of model identification and residual generation for fault diagnosis have been tested on the real data acquired from a chemical process. The results coming from the real data tests are reported and commented.

The paper is organised as follows. In Section II the problem statement is given and described from a mathematical point of view. The fault diagnosis scheme is then presented in Section III. In Section IV, the chemical industrial process used to test the proposed methodology is presented and the results concerning the diagnosis of faults are also reported. Finally, conclusions reported in Section V close the paper.

II. PROBLEM FORMULATION

This section addresses the mathematical modelling of the system under diagnosis for the problem of model–based fault diagnosis. Let us suppose that a number of N samples

0-7803-8335-4/04/$17.00 ©2004 AACC

can be acquired form the monitored system depicted in Figure (1). Such time sequences of data can represent the input and the output variables $\mathbf{u}(t) \in \Re^r$ and $\mathbf{y}(t) \in \Re^m$ of the process, with $t = 1, \ldots, N$, respectively.

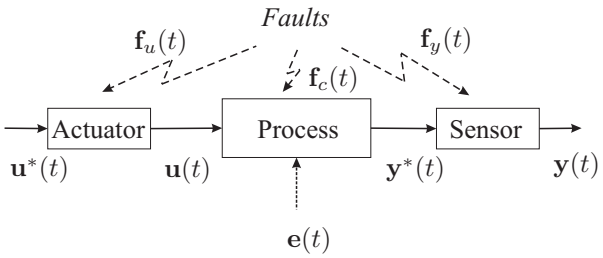

Fig. 1. The monitored system.

In the general framework of linear systems, in this paper we consider the description of the plant (1) and its input–output measurements by means of a discrete–time, time–invariant, input–output dynamic model with the following structure:

$$\mathbf{y}(t) = \mathrm{F}(z)\ \mathbf{u}(t) + \mathrm{G}(z)\ \mathbf{e}(t), \quad t = 1, \ldots, N. \tag{1}$$

The entries of the discrete–time transfer matrices $\mathrm{F}(z)$ and $\mathrm{G}(z)$ are parametrised as rational functions of polynomials in the variable z, the coefficients of which are the model parameters to be identified [10]. Since the *equation error term* $\mathbf{e}(t)$ is introduced to describe the disturbance (un–modelled dynamics) affecting the model, the system (1) is often called *equation error* (EE) *model* structure. The variable z represents the forward shift operator, i.e. $z\ y(t) = y(t+1)$ and it is consistent with the conventional definition of the z–transform.

Depending on the structure of the transfer matrices $\mathrm{F}(z)$ and $\mathrm{G}(z)$, the family of the EE models can describe the classes of the so–called ARX (Auto–Regressive eXogenous), ARMAX (Auto–Regressive Moving Average eXogenous) and the most general Box–Jenkins (BJ) structures [10].

The input–output models represented by the system family (1) can be transformed into a state–space formulation, in which a first–order difference equation exploits an auxiliary state vector $\mathbf{x}(t) \in \Re^n$ [10]. Hence, the following representation can also be considered:

$$\begin{cases} \mathbf{x}(t+1) &= \mathrm{A}\ \mathbf{x}(t) + \mathrm{B}\ \mathbf{u}(t) + \mathrm{H}\ \mathbf{e}(t) \\ \mathbf{y}(t) &= \mathrm{C}\ \mathbf{x}(t) + \mathbf{e}(t)\ , \quad t = 1, \ldots, N, \end{cases} \tag{2}$$

where A, B, C and H are matrices of appropriate dimensions that can be obtained by direct identification procedures, *e.g.*, the subspace approaches [15], [16].

Since the vector $\mathbf{e}(t)$ appears explicitly as in (1), the SS representation (2) is known as the *innovation form* of the state–space description.

Finally, another set of models which can be used for identification purpose is represented by the EIV systems. According to this theory, it is assumed that the monitored system can be described by a linear, discrete–time, time–invariant, dynamic model of the type:

$$\mathbf{y}(t) = \mathrm{F}(z)\ \mathbf{u}(t), \quad t = 1, \ldots, N, \tag{3}$$

where the transfer matrix $\mathrm{F}(z)$ consists of polynomial rational function of z representing the link between the input and the output measurements.

As depicted in Figure (1), the input and the output variables $\mathbf{u}^*(t)$ and $\mathbf{y}^*(t)$ are usually measured through actuator and sensors. Generally, sensor and actuator measurements are affected by additive noise, that can be modelled as:

$$\begin{cases} \mathbf{u}(t) &= \mathbf{u}^*(t) + \tilde{\mathbf{u}}(t), \\ \mathbf{y}(t) &= \mathbf{y}^*(t) + \tilde{\mathbf{y}}(t). \end{cases} \tag{4}$$

According to the EIV model theory, the variables $\tilde{\mathbf{u}}(t)$ and $\tilde{\mathbf{y}}(t)$ are generally described as white, zero-mean, uncorrelated Gaussian noises [13], [14], [11].

Since the error vector $\mathbf{e}(t)$ does not appear explicitly in the EIV models as in (1), the uncertainty is represented by the noise terms $\tilde{\mathbf{u}}(t)$ and $\tilde{\mathbf{y}}(t)$ and their variances, that have to be identified [14]. Hence, it is assumed that $\mathbf{u}(t)$ and $\mathbf{y}(t)$ are the only available measurements from the real process.

These model sets (1), (2) and (3) belong to the most commonly used ones in practice and we have therefore reason to present and use them since both explicit algorithms for parameter identification and analytic results are available [15], [10].

The model description in Eqs. (1), (2) and (3) assumes fault–free system operations and working conditions. As depicted in Figure (1), additive fault occurrence can be modelled by means of the following relations:

$$\begin{cases} \mathbf{u}(t) &= \mathbf{u}^*(t) + \mathbf{f}_u(t) \\ \mathbf{y}(t) &= \mathbf{y}^*(t) + \mathbf{f}_y(t) \end{cases} \tag{5}$$

where $\mathbf{f}_u(t)$ and $\mathbf{f}_y(t)$ are the actuator and sensor faults, respectively.

These vectors may be modelled by step and ramp signals in order to describe the presence of bias or drift on the measurements (abrupt and slowly developing faults). Signals $\mathbf{u}(t)$ and $\mathbf{y}(t)$ represent the input and output measurements, respectively, which have been used for the fault detection task. Therefore, by neglecting actuator and sensor dynamics, under fault–free assumptions (1), $\mathbf{u}(t) = \mathbf{u}^*(t)$ and $\mathbf{y}(t) = \mathbf{y}^*(t)$.

On the other hand, the case of *component faults* $\mathbf{f}_c(t)$ cannot be described by Eqs. (5). On the other hand, by assuming general detectability conditions [9], faults affecting output measurements $\mathbf{y}(t)$ can be successfully detected by monitoring both $\mathbf{u}(t)$ and $\mathbf{y}(t)$ signals. In particular, in some cases, the fault $\mathbf{f}_c(t)$ could be described as:

$$\mathbf{x}(t+1) = \mathrm{A}\ \mathbf{x}(t) + \mathrm{B}\ \mathbf{u}(t) + \mathbf{f}_c(t) \tag{6}$$

where the fault is represented as the case when some condition changes in the system rendering the dynamic relations (2) invalid.

The orders and parameters (structures) of the EE and SS models (1) and (2) can be estimated from the measured data $\mathbf{u}(t)$ and $\mathbf{t}(t)$ by means of automatic identification procedures available in the *System Identification Toolbox* in Matlab environment [15], [10]. On the other hand, the estimation of EIV models (3) was presented in [14] and was achieved by a software program implemented in Matlab environment by the same author [11].

Among all the systems presented above, the aim of the paper consists of selecting the most accurate identified model which is able to describe the measured data $\mathbf{u}(t)$ and $\mathbf{y}(t)$ in the "best possible" way. Since the essence of an identified model is its prediction aspect, we can introduce the following performance index:

$$J = \sum_{i=1}^{m} \frac{1}{N} \sum_{t=1}^{N} \left(\hat{y}_i(t) - y_i(t)\right)^2 \qquad (7)$$

representing the sum of the mean square errors between the i–th output vector $\hat{y}_i(t)$ predicted by the different MIMO models (1), (2) and (3) and the corresponding i–th output measurement $y_i(t)$.

It is worthwhile noting how another very effective way of evaluating the adequacy and flexibility of the identified models consists in their use for performing complete simulations (i.e. using only the initial samples of the predicted outputs) and in comparing the obtained predictions with the measured output samples. This procedure gives the best results when applied to sequences different from those used to identify the model. The mean square prediction error (7) between the measured outputs and the ones obtained by simulation can be used to compare the different identified models.

III. RESIDUAL GENERATION

The problem treated in this work regards the diagnosis of faults on the basis of the knowledge of the measured sequences $\mathbf{u}(t)$ and $\mathbf{y}(t)$. The structure of the fault detection device is depicted in Figure (2).

The symptom or residual generation $\mathbf{r}(t)$ is implemented by means of dynamic observers or Kalman filters, driven by $\mathbf{u}(t)$ and $\mathbf{y}(t)$, in order to produce a set of signals from which it will be possible to diagnose faults associated to actuators, components and sensors. As depicted in Figure (2), the symptom evaluation refers to a logic device which processes the redundant signals generated by the first block in order to unequivocally detect any fault occurrence.

Fault diagnosis is therefore achieved through the processing of the residual signals $\mathbf{r}(t) = \mathbf{y}(t) - \hat{\mathbf{y}}(t) = \mathbf{y}(t) - \mathrm{C}\ \hat{\mathbf{x}}(t)$. The are obtained comparing the system measurements with the dynamic observer or Kalman filter predictions designed on the basis of the identified model of the process under diagnosis.

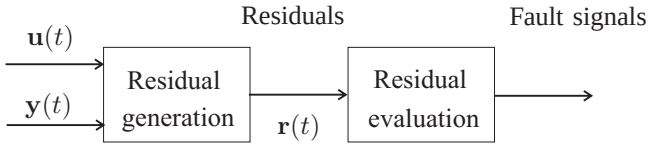

Fig. 2. Logic diagram of the residual generator.

As an example, a dynamic observer for the SS model has the following structure:

$$\hat{\mathbf{x}}(t+1) = \mathrm{A}\hat{\mathbf{x}}(t) + \mathrm{B}\mathbf{u}(t) + \mathrm{K}\big(\mathbf{y}(t) - \mathrm{C}\ \hat{\mathbf{x}}(t)\big) \qquad (8)$$

$\hat{\mathbf{x}}(t)$ being the observer state vector. The observer eigenvalues are often chosen in order to maximise fault detection promptness and to minimise the occurrence of false alarms.

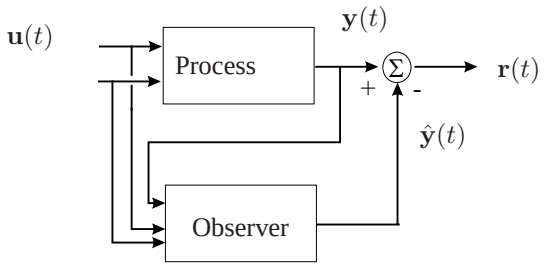

Fig. 3. The observer (filter) residual generator scheme.

On the other hand, for the Kalman filter design, the essential difference regards the choice of the feedback matrix K which is computed by solving a Riccati equation. The solution of this equation requires the knowledge also of the variance matrices of the input and the output noises, which can be identified by means of the dynamic Frisch scheme [13], [14], [11].

The proposed FDI scheme is applied to a Continuous Stirring Tank Reactor (CSTR) process [17], the dynamic behaviour description of which has been achieved by using a model obtained from identification procedures.

IV. CHEMICAL PROCESS FAULT DETECTION

The aim of the study presented in this paper is to develop a general procedure for the diagnosis of faults in a chemical process by means of identified models of the process under investigation. In particular, the monitored process is a real Continuous Stirring Tank Reactor, where the reaction between reactant and product is exothermic.

The main input variables ($r = 3$) are: the reactor jacket inlet temperature $T_{\text{in}}(t)$ [K], the reactor temperature $T(t)$ [K] and the reactor cooling water rate $q(t)$ [$\frac{m^3}{min}$]. The main output ($m = 4$) measurements are: the reactor jacket outlet temperature $T_{\text{out}}(t)$ [K], the product percentage conversion $C(t)$ [%], the number average molecular weight $N_m(t)$ [$\frac{g}{\text{mol}}$] and weight average molecular weight $W_m(t)$ [$\frac{g}{\text{mol}}$].

The process objective is to maintain constant the reactor polymer production by controlling the main input variables in despite of the unmeasurable disturbance, *i.e.*, the reactor impurity concentration and fouling $\mathbf{d}(t)$. The importance of this case study is that there are many examples of reactors in industry like polymerisation reactor [17]. The CSTR with cooling jacket is shown in Figure (4).

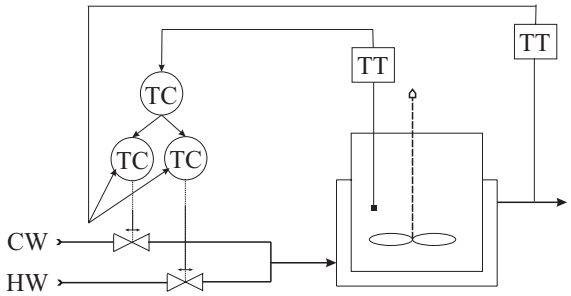

Fig. 4. Schematic of the CSTR process.

Hence, the process has $r = 3$ control inputs, $\mathbf{u}(t) = \left[T_{\text{in}}(t), T(t), q(t)\right]$, while the output measurements ($m = 4$) are $\mathbf{y}(t) = [T_{\text{out}}(t), C(t), N_m(t), W_m(t)]$. These actual signals can be acquired from the real plant depicted in Figure (4).

The disturbance vector $\mathbf{d}(t)$ represents reactor impurities and fouling. Constant physical properties and constant boundary pressures of all input and output streams are assumed. Both process normal operating time series and faulty data (with different amount of impurities and fouling) have been measured from the real process. A sampling rate of 0.5s. was used to acquire a number of $N = 240$ actual data sequences. The measurements acquired from the actual chemical process have been modified for proprietary reasons.

Therefore, according to Section (II), several families of multiple input–multiple output (MIMO) models (three inputs and four outputs) have been identified by using a batch sequence of normal operating data. Each MIMO model of type (1), (2) and (3) is driven by $\mathbf{u}(t)$ and provides the prediction of the output $\hat{\mathbf{y}}(t)$ for $t = 1, \ldots, N$.

Table (I) shows the performances of the different identified models by reporting the values of the J index (7) with respect to the *identification* data. Each model has been tested also in different operating conditions and the output reconstruction errors J are compared in Table (I). Several time series of batch data from by reactor corresponding to different amounts of reactor impurities and fouling (*validation* data) have been also exploited in order to validate the ARX, ARMAX, BJ, SS and EIV models.

On the basis of the simulation results summarised in Table (I), a MIMO SS model can be chosen to describe with the "best accuracy" the monitored process dynamics.

The CSTR process data contains several faults. Some of these faults are known (actuator $f_u(t)$ and sensor $f_y(t)$), and other are unknown (component or system $f_c(t)$). Abrupt

TABLE I

CSTR MODEL PERFORMANCES J WITH IDENTIFICATION AND VALIDATION DATA.

Model	**Order**	J **(Ident.)**	J **(Valid.)**
ARX (EE)	5	0.1203	0.4631
ARMAX (EE)	3	0.0067	0.0161
BJ (EE)	3	0.0826	0.0996
SS	4	0.0034	0.0081
EIV	5	0.1082	0.3511

fault dynamics can be associated with a step change in process variables. On the other hand, slow developing faults can be associated with an increase in the variability of some process variables, *e.g.*, a slow drift in the reaction kinetics.

In this work different fault cases have been considered: *(a)* the reactor jacket inlet temperature $T_{\text{in}}(t)$ (sudden actuator fault $f_u(t)$), *(b)* the reactor jacket outlet temperature $T_{\text{out}}(t)$ (incipient sensor fault $f_y(t)$) and *(c)* the process $f_c(t)$ fault (reactor impurities and fouling) concerning the product percentage conversion $C(t)$ have been considered in the following.

Therefore, in such fault scenario, in order to successfully perform the fault detection task, three process measurements $T_{\text{in}}(t)$, $T_{\text{out}}(t)$ and $C(t)$ are exploited.

The residual $\mathbf{r}(t)$ generation has been performed according to the fault diagnosis scheme presented in Section (III). The dynamic observers for the residual signal $\mathbf{r}(t)$ computation can be designed on the basis of the most accurate identified SS model of the process under diagnosis. Observer eigenvalues have been selected with a trial and error procedure in order to maximise the fault residual sensitivity and to minimise the false alarm occurrence [11]. On the other hand, the Kalman filters design requires the identification noise covariance matrices affecting the input–output measurements. They have been estimated by exploiting the Frisch scheme identification method [14], [11].

As an example, Figure (5) represents fault–free and faulty residual $r(t)$ abrupt change for the case in which the additive *actuator* fault $f_u(t)$ affects the reactor jacket inlet temperature $T_{\text{in}}(t)$ commencing at the sample 50 ($t = 25$s.).

On the other hand, Figure (6) represents the healthy and the faulty residual $r(t)$ slow variation for the case in which the additive *sensor* fault $f_y(t)$ affects the reactor jacket outlet temperature $T_{\text{out}}(t)$ starting at the sample 150 ($t = 75$s.).

Finally, Figure (7) represents the fault–free and the faulty residual $r(t)$ changes when a *reactor system* fault $f_c(t)$ affects the product percentage conversion $C(t)$. Such a process fault $f_c(t)$ is due to the formation of reactor impurities and fouling.

It is worthwhile noting that, in general, in order to achieve the maximal fault detection capability, the measurement corresponding to the most sensitive output $y(t)$ to a fault signal has to be selected. Moreover, with reference to this case study, the monitored signals are enough to accomplish

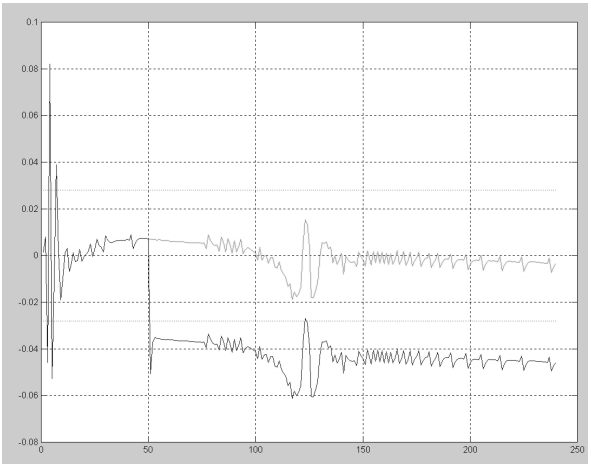

Fig. 5. Fault case *(a)* concerning the reactor jacket inlet temperature $T_{\text{in}}(t)$.

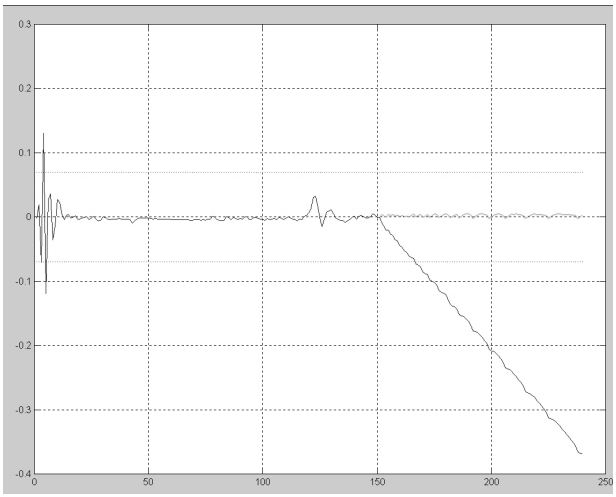

Fig. 6. Fault case *(b)* regarding the reactor jacket outlet temperature $T_{\text{out}}(t)$.

fault isolation, as well [11].

V. CONCLUSION

The complete design procedure for FDI in actuators, components and sensors of an industrial process was described in this work. The fault diagnosis was performed by using a bank of dynamic observers or Kalman filters. Faults on the component of the system, actuator and sensors were therefore considered. The suggested method did not require any physical knowledge of the process under observation since the input–output links were obtained by means of an identification scheme, which uses EE, EIV, and SS models. In the situation of EIV models, the identification technique (Frisch scheme) gave the variances of the input–output noises, which are required in the design of the Kalman filters.

Such a procedure was applied to the real data acquired

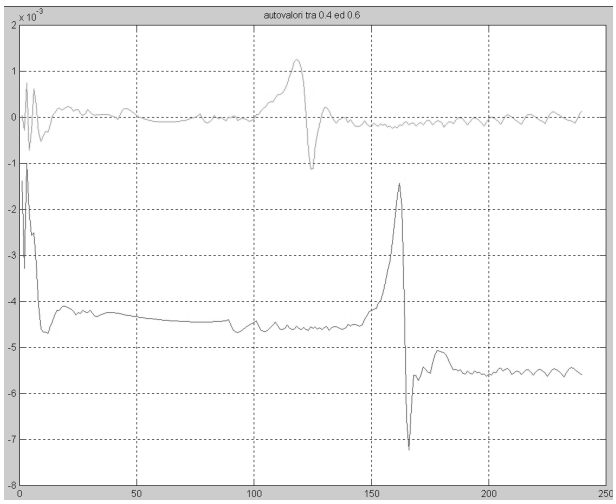

Fig. 7. Fault case *(c)* affecting the product percentage conversion $C(t)$.

from an industrial chemical process. In order to analyse the diagnostic effectiveness of the FDI system in the presence of changes or drifts in measurements, faults described also by ramp functions were generated. The results obtained by this approach indicated that the minimal detectable faults on the system actuator, component and sensor are of interest for the industrial diagnostic applications.

The main aspect of this work was the use of linear system identification and modelling methods, although the system considered was non–linear. This is considered important to avoid the complexities that would otherwise be inevitable when non–linear models are used. There is certainly an increasing interest in the use of non–linear methods (non–linear observers, extended Kalman filters, fuzzy-logic methods, etc). However, as the feature of system supervision is to monitor the operation and performance of the system with respect to an expected point of operation, linear system methods are still very valid. Deviations from expected behaviour can be used to monitor system performance changes as well as system component malfunctions.

REFERENCES

[1] R. J. Patton, P. M. Frank, and R. N. Clark, eds., *Issues of Fault Diagnosis for Dynamic Systems*. London Limited: Springer–Verlag, 2000.

[2] P. M. Frank, S. X. Ding, and B. Köpper-Seliger, "Current Developments in the Theory of FDI," in *SAFEPROCESS'00: Preprints of the IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes*, vol. 1, (Budapest, Hungary), pp. 16–27, 2000.

[3] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," *Automatica*, vol. 12, pp. 601–611, Nov. 1976.

[4] J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*. New York: Marcel Dekker, 1998.

[5] X. Ding and P. M. Frank, "Fault detection via factorization approach," *Syst. Contr. Lett.*, vol. 14, no. 5, pp. 431–436, 1990.

[6] P. M. Frank, "Fault diagnosis in dynamic systems using analytical and knowledge based redundancy: A survey of some new results," *Automatica*, vol. 26, pp. 459–474, May 1990.

[7] L. Xie, Y. C. Soh, and C. E. de Souza, "Robust Kalman filtering for uncertain discrete-time systems," *IEEE Transaction on Automatic Control*, vol. 39, pp. 1310–1314, 1994.

[8] A. E. Emami-Naeini, Akhter, and S. Rock, "Effect of model uncertainty on failure detection: the threshold selector," *IEEE Trans. on Automatic Control*, vol. 33, no. 2, 1988.
[9] J. Chen and R. J. Patton, *Robust Model–Based Fault Diagnosis for Dynamic Systems*. Kluwer Academic, 1999.
[10] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, N.J.: Prentice Hall, second ed., 1999.
[11] S. Simani, C. Fantuzzi, and S. Beghelli, "Diagnosis techniques for sensor faults of industrial processes," *IEEE Transactions on Control Systems Technology*, vol. 8, n. 5, pp. 848–855, September 2000.
[12] C. Fantuzzi, S. Simani, and S. Beghelli, "Robust fault diagnosis of dynamic processes using parametric identification with eigenstructure assignment approach," in *CDC'01* (I. CSS, ed.), (Orlando, Florida, U.S.A), pp. 155–160, 2001, 40th IEEE Conference on Decision and Control, December, 4–7 2001.
[13] R. Frisch, *Statistical Confluenece Analysis by Means of Complete Regression Systems*. University of Oslo, Economic Institute, publication n. 5 ed., 1934.
[14] S. Beghelli, R. P. Guidorzi, and U. Soverini, "The Frisch scheme in dynamic system identification," *Automatica*, vol. 26, no. 1, pp. 171–176, 1990.
[15] L. Ljung, *System Identification Toolbox User's Guide*. The MathWorks, Inc., Natick, MA, USA, 1995.
[16] P. van Overschee and B. De Moor, *Subspace Identification for Linear Systems*. Kluwer Academic Publisher, 1999.
[17] E. L. Russell, L. H. Chiang, and R. D. Braatz, *Data-Driven Techniques for Fault Detection and Diagnosis in Chemical Processes*. Advances in Industrial Control, London, UK: Springer–Verlag, 2000.