A Methodology for Incipient Fault Detection*

T. Escobet, V. Puig, J. Quevedo and D. Garcia

Abstract— **This paper proposes a fault detection methodology for incipient faults that combines different residual generation methods (observers and *l*-step ahead predictors) with different convergence velocity to the real output trying to benefit from the advantages offered by each one. The integration is based on generating a timed automaton, which combines the information extracted from each method in order to provide the best fault detection performance regarding incipient faults. The proposed methodology has as a main objective to detect as early as possible anomalies or incipient faults in system components. Nowadays, for many systems, early warnings contribute to increase system reliability, prevent major component failures and planning the necessary repair actions for several weeks (predictive maintenance). The application of this methodology will be illustrated in a case study based on a part of the Barcelona water network.**

I. INTRODUCTION

Component degradation can be characterised as an incipient fault. Faults and degradations in the components of the dynamic systems could manifest as a deviation of the system behaviour from the normal one required for operation. The fault detection of the incipient faults is more difficult than the case of abrupt faults because they evolve very slowly and their effects can be confused with noise and uncertainty. For this reason most of the methodologies used for detecting as early as possible anomalies due to incipient faults in components are based on the estimation of features (or condition indicators) from sensor data that are characteristic of the abnormal behaviour either in open-loop [1] [2] or in closed-loop [3]. Another problem of the incipient faults is that they can evolve very slowly (over a period of months or even years). This implies that to detect indicator drifts, long time periods must be used for evaluation. In real applications, this involves storing a lot of data if slow changes in the monitored system must be detected. To solve this problem, in [3], a hierarchical structure is proposed, with high-level indices based on the temporal and/or spatial combination of low-level indices, which include low-level information. The objective of these high-level indices is to compress the information.

Fault detection of incipient faults is still an open issue because of the difficulty associated with. In particular, one the main difficulties comes from the fact than some methods based on a closed-loop prediction as observers or Moving Average (MA) parity equations tend to track the system even when there is a fault. This problem has been studied in [4] where it has been shown that even in the case of abrupt faults the output prediction provided by observers (or MA parity equations) tends to track the faulty system in case of output sensor faults.

The way to avoid this problem is by playing with the observer dynamics or by enlarging the prediction horizon using *l*-step ahead predictors [5]. However, this can introduce the drawback that the fault detection is delayed or even some faults are not properly detected. Thus, there is a trade-off between the velocity of convergence of the model prediction to the real output and the fault detection performance.

This paper proposes a fault detection methodology for incipient faults that combines different residual generation methods (observers and *l*-step ahead predictors) with different convergence velocity to the real output trying to benefit from the advantages offered by each one. The integration is based on generating a timed automaton, which combines the information extracted from each method in order to provide the best fault detection performance regarding incipient faults.

The proposed methodology has as a main objective to detect as early as possible anomalies or incipient faults of the components of a large scale system (LSS). Nowadays, for many systems, early warnings contribute to increase system reliability, preventing major component failures and planning the necessary repair actions for several weeks (predictive maintenance). The application of this methodology will be illustrated on a case study based on a part of the Barcelona water network.

The structure of the paper is as follows: Section 2 presents the problem statement. Section 3 introduces the residual generation approaches considered to be used for incipient fault detection. Section 4 proposes how to integrate the different residual generation approaches in order to obtain the best performance regarding incipient fault detection. Section 5 presents the proposed case study and the results of application of the proposed methodology. Conclusions are provided in Section 6.

II. PROBLEM STATEMENT AND PROPOSED FDI ARCHITECTURE IMPLEMENTATION

A. Incipient faults

Fault detection and isolation (FDI) aims to carefully identify which fault can be hypothesized to be the cause of anomalous system behaviour. This problem has been studied from many points of view, using different techniques, and by many researchers, applying different schools of thought, theories and assumptions. An overview of techniques in this area is given in a series of review papers ([6]-[8]). A historical review of these techniques has been published recently in [9].

The diagnostic process aims to identify which fault can be hypothesized to be the cause of monitored events. A fault must be understood as an unexpected change in a component or the system behaviour. In the literature, most of works have proposed fault diagnosis schemes for abrupt faults, which are modelled as instantaneous changes in system behaviour at a point in time. However, degradations in system components are often modelled as incipient faults, which are slow drifts in

T. Escobet, V. Puig , J. Quevedo and D. Garcia are with Advanced Control Systesms (SAC) research group, Universitat Politècnica de Catalunya BARCELONATECH, Terrassa, Spain (e-mail: teresa.escobet@upc.edu, vicenc.puig@upc.edu, joseba.quevedo@upc.edu and diego.garcia@upc.edu).

system parameter values over time. In fact, the problem of diagnosis incipient faults is related with the problem of prognosis. For example, in [10], classical FDI techniques are proposed for incipient fault combined with a set of degradation models for incipient fault evolution prognosis.

One of the main problems associated with incipient faults is that the symptoms associated with are not detected at the same time instant since residual dynamics (and system degradation) require some time to evolve. Moreover, because the effect of incipient faults at the beginning is not that important, probably residuals will not be sensitive enough to be affected by the fault or to avoid tracking the faulty behavior. This can even be harder if several incipient faults appear at the same time.

B. Architecture implementation for FDI of incipient faults

This paper proposes a FDI methodology for incipient faults that at the same time allows continuing to address the case of abrupt faults.

The proposed FDI architecture is based on two modules as suggested in most model-based approaches. The first module is fault detection, implemented with a set of fault detection tests based on generating residuals using analytical redundancy relations (ARRs), which allows deciding whether a fault has occurred, and its apparition time instant. The second module is fault isolation, which is typically achieved through algorithms that determine a possible faulty component(s). Here, the implementation of this architecture, initially proposed by [11] for abrupt faults, will be extended to the incipient fault case. In the implementation proposed in [11], a timed discrete-event approach is introduced that improves the interface between fault detection and isolation considering the activation degree and the occurrence time instant of the diagnostic signals using a combination of several theoretical fault signature matrices that store the relationship knowledge between diagnosis signals and faults. The fault isolation module is implemented using a timed discrete-event approach that recognizes the occurrence of a fault by identifying a unique sequence of observable events (fault signals). The states and transitions that characterize such a system can directly be inferred from the relation between fault signals and faults.

III. RESIDUAL GENERATION FOR INCIPIENT FAULT DETECTION

A. Fault detection background

In general, the FDI procedure checks at every time the consistency between the observed and the normal system behavior using a set of analytical redundancy relations (ARRs), which relate the values for measured variables according to a model of normal operation (fault-free) from the monitored system. These ARRs can be derived with structural approaches, using algorithms as the proposed in [12][13]or [14], among others.

Given a model with a set of output observed variables, y_k, and the set of inputs, u_k, consistency tests can be derived from an ARR by generating a computational residual in the following way:

$$r_i = \Psi_i \left(y_k, u_k \right) = 0, \tag{1}$$

where Ψ_i is called the residual ARR expression. The set of ARRs can be represented as

$$\mathcal{R} = \left\{ r_i = \Psi_i \left(y_k, u_k \right) = 0, i = 1, \cdots, n_r \right\}, \tag{2}$$

where n_r is the number of ARRs obtained after applying the structural analysis.

Let $\mathcal{F}$ be the set of faults that must be monitored,

Definition 3.1. Detectable fault. A $f \in \mathcal{F}$ is detectable if its occurrence can be observed, or at least one of the residuals in the residual set (2) satisfies $r_i \neq 0$.

Using the set of computable ARR residuals (1), the fault detection module must check at each time instant whether or not they are consistent with the observations. Under ideal conditions, residuals are zero in the absence of faults and non-zero when a fault is present. However, modeling errors, disturbances and noise, in complex engineering systems are inevitable, and hence it appears the necessity of applying robust fault detection algorithms.

In the literature, there are different approaches to solve the problem of modeling errors, disturbances and noise. For example, statistical decision methods can be used when unknown dynamics and measurement noise are stochastically modeled [15]. Others works assume that disturbances/model errors and measurement noise are bounded and their effect is propagated to the residuals using, for example, interval methods [16]. Taking into account bounded uncertainties, the residuals (1) can be rewritten as follows:

$$r_i = \Psi_i \left(y_k, u_k, \delta_k \right), \tag{3}$$

with: $\delta_k \in D$, where D is the interval box $D = \left\{ \delta \in \mathbb{R}^{n_\delta} \middle| \underline{\delta} \leq \delta \leq \overline{\delta} \right\}$, that includes all uncertainties (i.e. disturbances/model errors and measurement noise). Then, fault detection is formulated as checking the consistency of (3) using a set-membership approach [17].

Definition 3.2 Consistency checking. Given the residuals described by (3) and a sequence of measured inputs u_k and outputs y_k of the real system at time *k*, they are consistent with measurements and uncertainty bounds if there exist a set of sequences $\delta_k \in D$ which satisfies $r_i = 0$.

Thus, according to Definition (3.2), a residual consistency is equivalent to check if $0 \in [r_i]$ where $[r_i]$ is the interval that bounds the effect of uncertainty in residual (3).

Definition 3.3. Fault detection. Given a sequence of observed inputs u_k and outputs y_k of the real system, a fault is said to be detected at time *k* if there does not exist a set of uncertainty sequences $\delta_k \in D$ to which the set of ARRs is consistent.

According to Definition (3.3) a fault is detected when $0 \notin [r_i]$. The information provided by the consistency checking is stored as fault signal $\phi_i (k)$:

$$\phi_i(k) = \begin{cases} 0 & if \ 0 \in [r_i] \\ 1 & if \ 0 \notin [r_i] \end{cases}. \qquad (4)$$

B. Residual computation approaches

From computation point of view, residuals (3) are generated as $r(k) = y(k) - \hat{y}(k, \delta)$, that is, by means of the difference between measured system outputs and estimated values $\hat{y} \in \Re^{n_r}$ obtained from rewriting residuals (3) as parity equations

$$\hat{y}(k, \delta) = y(k-1) + B(\delta)u(k) + \Delta(\delta) \qquad (5)$$

or state observers [4]

$$\hat{x}(k+1) = A(\delta)\hat{x}(k) + B(\delta)u(k) + L(y(k) - \hat{y}(k))$$
$$\hat{y}(k, \delta) = \hat{x}(k) + \Delta(\delta) \qquad (6)$$

where $A \in \Re^{n_r \times n_r}$ is the state matrix $B \in \Re^{n_r \times p}$ is the input matrix, $\Delta \in \Re^{n_r}$ is the noise uncertainty gain matrix and $L \in \Re^{n_r \times n_r}$ is the observed gain to be designed to guarantee stability for $\delta_k \in D$. Notice than when $L=A$, (5) and (6) are equivalent (see [4] for more details)

Abrupt faults correspond to changes that occur at time scales much faster than the nominal dynamics of the system. In order to detect this type of faults, fast dynamic residuals (as (5) and (6)) are required. Although these faults are relatively easy to detect because of the sudden large change they induce in the values of some variables. On the other hand, incipient faults require residuals with a slowly evolution over time in order to detect slow performance degradations or undesirable drifts. The detection of these slow changes can be achieved following the idea proposed in signal processing ([18]) where most of the methods for trend extraction involves approximating the signals with piecewise polynomial segments using for example fixed-width window approach or sliding windows. Thus, for example, parity equations can be evaluated in a time window on length m leading to a m-step ahead predictor

$$\hat{y}(k \mid k-m) = A_p(\delta) y(k-m) + B_p(\delta)\bar{u}(k) + \Delta(\delta), \qquad (7)$$

where $A_p(\delta)$ and $B_p(\delta)$ are the m-step ahead predictor matrices obtained considering output measurement at time $y(k-m)$ and inputs in the time window $\bar{u}(k) = [u(k) \quad u(k-1) \quad \cdots \quad u(k-m)]^T$.

An alternative to evaluating analytical model over a time windows is to consider filtered residuals as follows

$$fr_i(k) = \frac{\tau_i}{m} \sum_{\alpha=k-m}^{k} r_i(\alpha) + (1-\tau_i) fr_i(k-1), \qquad (8)$$

where fr is the smoothed residual, τ is a weighting scalar between the residual averages computed along the horizon of size m and the actual smoothed residual.

Both (7) and (8) allows to treat the temporal information regarding the fault effect in an aggregate way and leading to define a new set of residuals as follows

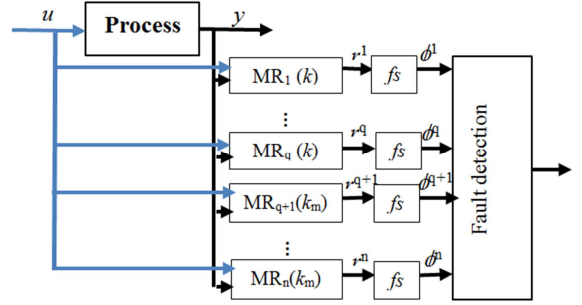$$R(k_m) = \Upsilon[r(k-m) \quad \cdots \quad r(k)], \qquad (9)$$

Figure 1. Proposed methodology.

where k_m is a multiple of the considered time window and Υ is a feature extraction function (such as mean, min, max).

IV. PROPOSED METHODOLOGY

A. Consistency check based on a bank of residuals

With the aim of providing a methodology that is able to deal with incipient and abrupt faults, a bank of residuals including observers and parity equations with sliding windows (or filters) is proposed (see Fig. 1). Each bank of residual generators, identified as MR_i, can use different sampling time in order to take into account different fault time evolutions.

In addition, each residual is evaluated against the fault detection test presented in (4) producing a fault signal, ϕ. This evaluation is performed in the block diagram identified as *fs*. Then, the fault signals are analysed in an integrated way as follows.

B. Fault detection module

The integrated analysis of the different signals is done via a timed automaton [19]. A timed automaton, denoted by G_t, is a six-tuple

$$G_t = (Q, \Sigma, C, T, I, q_0), \qquad (10)$$

where Q is the finite set of discrete states; q_0 is the set of initial states; Σ is the finite set of events (alphabet); C is the finite set of clocks, $c_1, ..., c_n$, with $c_i(t) \in \mathbb{R}^+, t \in \mathbb{R}^+$; I is a mapping that labels each state $q \in Q$ with some clock constraint in $\Phi(C)$; and $T \subseteq Q \times \Sigma \times \mathcal{L} \times 2^{\Phi_c} \times 2^C$, is the set of switches. A transition is defined by $T = \langle q, q', \sigma, \delta, \lambda \rangle$, where $q, q' \in Q$ are origin and destination T, $\sigma \in \Sigma$ is the event of T, $\delta \subset \Phi_C$ is called guard of T, and $\lambda \subset C$ is called reset of T. A guard (or an invariant) is satisfied when all its clock constraints are evaluated to true, otherwise it is unsatisfied. The automaton used for fault detection has three states: initial mode (M0), alarm mode (M1) and faulty mode (M2). The events are associated at each time instant with a logical combination of ϕ_i. Timed transition is a clock constraint used to specify the amount of time that may be spent in a state location; each one is determined using the model structure and the type of faults to be isolated.

Following the definition given by [20], a clock constraint is a formula "$h \sim k$", where h is a clock, $\sim\in\{<,>,\le,\ge\}$ and $k\in\mathbb{N}$. Let $\Phi_{\mathcal{C}}$ be the set of clock constraints using clocks of $\mathcal{C}$. The proposed time automaton G_t that implements the fault detection modules is presented in Fig. 2. Each node represents the system state, and each arrow linking q to q' and labelled by $\langle\sigma,\delta,\lambda\rangle$ represents a transition $\langle q,q',\sigma,\delta,\lambda\rangle$. An empty δ,λ is denoted by $-$. Initially, the G_t is in M0 and at any time it can execute the event σ_1. The execution of σ_1 is simultaneous with the reset of h and leads to M1, in which the event σ_2 is enabled when $h > h_0$ and the event σ_3 is enabled when $h \le h_0$. The execution of σ_3 resets h and leads to M0. The execution of σ_4 from M2 leads to M0 and resets h.

The logical conditions associated with σ_1, σ_2, σ_3 and σ_4 are:

- $\sigma_1 = 1$ if $\exists(\phi_i(k)=1\,\&\,\phi_i(k-1)=0)$ *for any i*
- $\sigma_3 = 1$ if $\phi_i(k)=0\ \forall i$
- $\sigma_2 = 1$ if $\exists\phi_i(k-j)=1$ *for all* $j=[0,h_l]$ and *any i*
- $\sigma_4 = 1$ when fault is repaired or diagnosed and accommodated.

In previous logical conditions, h_0 and h_l are parameters to be decided by the designer in order to increase the robustness of the fault detection module and to get an acceptable performance.

Notice than with these conditions a fault mode (M2) is activated when one of the fault signals is active at least h_l time instants. The transition from M2 to M0 can be considered as well by the designer in case that some fault tolerance mechanism is available or an incipient fault is detected allowing the automaton to be ready to detect new faults. In this way, multiple sequential faults can be detected. Once M2 is reached, fault isolation task is initiated. However, this paper does not detail this task being part of a future research.

V. CASE STUDY

A. Description of transport network

The Barcelona water network supplies water to approximately 3 million of consumers, distributed in 23 municipalities in a 424 km² area and it is monitored through a SCADA with a sampling period of 1 hour. The complete transport network has been modeled using: 63 storage tanks, 3 surface sources and 7 underground sources, 79 pumps, 50 valves, 18 nodes and 88 demands [21]. The case study used to illustrate the methodology proposed in this paper is based on a part of this network. It includes 6 tanks and 5 actuators, being the observed variables the pump flows, the tanks levels
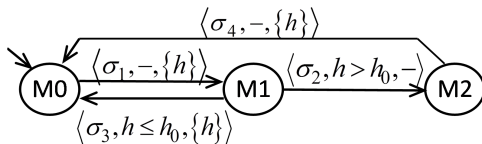

Figure 2. Time automaton implementation of the fault detection module.

and the demands.

The dynamic model of a water network is built by means of mass balances relating the stored volume (in m³) in tanks, x, with the manipulated tank inflows and outflows

$$x_i(k+1) = x_i(k) + \Delta t(\sum_i q_{in,i}(k) - \sum_j q_{out,j}(k)) \qquad (11),$$

where $q_{in,i}(k)$ and $q_{out,j}(k)$ correspond to the i-th tank inflow and the j-th tank outflow, respectively, in m³/s.

The case study model has 28 equations. The set of unknown variables is $\mathcal{X}=\{x_i,q_{in,i},q_{out,i},d_i\}$ while the set of known variables is $\mathcal{O}=\{u_i,y_j\}$ for $i=1,..,5$ and $j=1,..,15$, where d_i is the demand, u_j and y_j denote the measured input and output variables. The computational residuals have been obtained applying the algorithm proposed in [22]. The set of faults considered are located in actuators, f_{Pi}, flow transducers, f_{Fi}, level transducers, f_{Li}, and demand transducers f_{di}, for $i=1,\dots,5$. Each fault only affects a set of residuals as it can be seen from the Fault Signature Matrix (FSM) corresponding to the set of considered residuals and faults (Fig. 3).

The main aim of this case study is to show that the approach proposed in this paper is able to detect incipient faults in actuators and sensors.

B. Detection results

The bank of residuals used is presented in Table I. Notice that residuals present different temporal behaviour regarding the different fault scenarios.

	f_{P1}	f_{P2}	f_{P3}	f_{P4}	f_{P5}	f_{F1}	f_{F2}	f_{F3}	f_{F4}	f_{F5}	f_{L1}	f_{L2}	f_{L3}	f_{L4}	f_{L5}	f_{d1}	f_{d2}	f_{d3}	f_{d4}	f_{d5}
r_1	(+)1					(-)1														
r_2		(+)1					(-)1													
r_3	(+)1	(-)1										(+)1					(-)1			
r_4		(-)1						(+)1				(+)1					(-)1			
r_5	(+)1							(-)1				(+)1					(-)1			
r_6								(+)1	(-)1			(+)1					(-)1			
r_7		(+)1											(+)1					(-)1		
r_8								(+)1					(+)1					(-)1		
r_9			(+)1					(-)1												
r_{10}				(+)1					(-)1											
r_{11}				(+)1						(-)1										
r_{12}			(+)1	(-)1										(+)1					(-)1	
r_{13}				(-)1					(+)1					(+)1					(-)1	
r_{14}			(+)1						(-)1					(+)1					(-)1	
r_{15}									(+)1	(-)1				(+)1					(-)1	
r_{16}				(+)1	(-)1										(+)1					(-)1
r_{17}					(-)1					(+)1					(+)1					(-)1
r_{18}				(+)1						(-)1					(+)1					(-)1
r_{19}									(+)1	(-)1					(+)1					(-)1
r_{20}		(+)1													(+)1					(-)1
r_{21}								(+)1							(+)1					(-)1

Figure 3. Theoretical FSM using *binary* and *sign* information.

TABLE I. BANK OF RESIDUALS PROPOSED

Identifier	Detection strategy
PEm	Parity equation model with a sample time of 1h (5)
Lom	Luenberger Observed Model with L= 0.2 and a sample time of 1h (6)
WPEm	Windowed time parity equation with m=24 h (7)
SPEm	Smoothed parity equation model with m=24, τ= 0.05 and Υ equal to min/max feature extraction, equations (5), (8) and (9).
SWPEm	Smoothed windowed time parity equation with m=24, τ= 0.05 and Υ equal to min/max feature extraction, equations (7), (8) and (9)

Figure 4 shows the residual r_7 (computed using *PEm* and *LOm)* and signal fault evolution, ϕ_7, when an abrupt fault in pump, P2, has occurred, f_{P2}. The fault occurs at time 24 and all the residuals described in Table I have been evaluated. In this example, only noise uncertainty has been taking into account and the consistency test (8) is computed as follows

$$\phi_i(k) = \begin{cases} 0 & if \ r(k) = y(k) - \hat{y}_i(k) \in \Delta(\delta_i) \\ 1 & if \ r(k) = y(k) - \hat{y}_i(k) \notin \Delta(\delta_i) \end{cases} \quad (12).$$

All the fault signals ϕ_2, ϕ_3, ϕ_4, ϕ_7 computed are non-zero after the fault occurrence time are in agreement with the fault signature shown in Fig. 4. Table II shows fault detection delay (t_d) of all models considered (Table I), where fault detection delay is defined as the difference between the time of detection of a fault and the time of the fault occurrence. Notice that the detection times provided by *PEm*, *LOm* and *WPEm* ranges between 1 to 3, but neither of fault signals are activated simultaneously, and ϕ_7 computed by *PEm* is not persistent. Models *SPEm* and *SWPEm* (that consider a sampling time of 24 h) provide a detection times that is between 0 or 1 day, depending on the fault occurrence time.

Figure 5 shows residuals r_7, r_8 and fault signal ϕ_7, ϕ_8 evolution resulting from applying *PEm*, *LOm* and *WPEm* when an abrupt fault in sensor L2 (tank level sensor, f_{L2}) is occurred at time 29 h. In this scenario, all the fault signals are activated during a time interval, named as persistent indication time (Δt_d). This limited persistence time indication occurs in case of fault in sensors [4]. Notice from Table III that *PEm* and *WPEm* have the lower and greater Δt_d, respectively.

Figures 6 and 7 show residuals r_2, r_3, r_4, r_7 and the evolution of fault signals ϕ_2, ϕ_3, ϕ_4, ϕ_7 resulting from evaluating each one of the five residuals using the fault detection test (12) when an incipient fault in pump P2 (f_{P2}) has occurred at time instant 35 h. In case of *SPEm* and *SWEPm*, the minimum and maximum values extracted are drawn in red and blue, respectively. In this scenario, the residuals more affected by noisy measurements lead non-persistent fault indicators and their fault detection time has been indicated in brackets in Table IV. The effect of incipient faults become visible only after the magnitude of fault is increased above a certain threshold limit. This can cause some hazardous effects in the system.
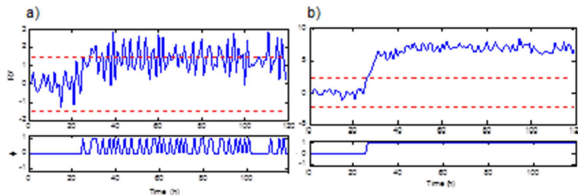

Figure 4. Evolution of r_7 and ϕ_7 with a) *PEm* and b)*LOm*.

TABLE II. FAULT DETECTION DELAY IN CASE OF AN ABRUPT F_{P2}

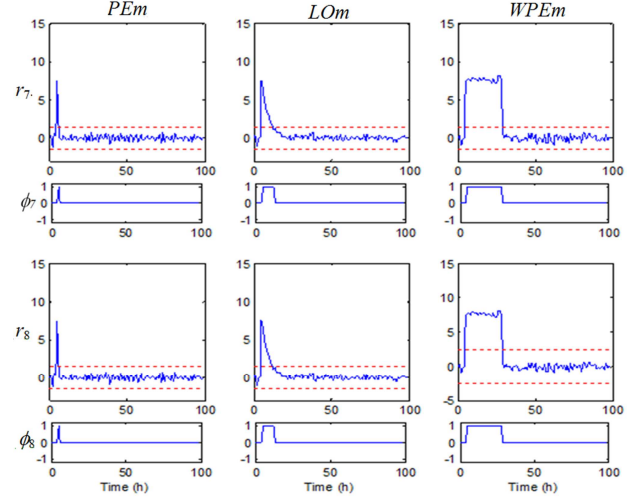| f_s	PEm t_d (h)	LOm t_d (h)	$WPEm$ t_d (h)	$SPEm$ t_d (days)	$SWPEm$ t_d (days)
ϕ_2	2	2	3	[0, 1]	[0,1]
ϕ_3	1	1	1	[0,1]	[0,1]
ϕ_4	1	1	1	[0,1]	[0,1]
ϕ_7	2 np^a	3	2	[0,1]	[0,1]

Figure 5. Evolution of r_7, ϕ_7, r_8 and ϕ_8 using *PEm*, *LOm* and *WPEm*.

TABLE III. FAULT DETECTION DELAY AND FAULT PERSISTENCE INDICATION TIME IN CASE OF AN ABRUPT SENSOR FAULT F_{L2}

f_s	PEm		LOm		WPEm	
	t_d	Δt_d	t_d	Δt_d	t_d	Δt_d
ϕ_7	1	1	1	12	1	25
ϕ_8	1	1	1	12	1	25

Table V summarises the activation delay of the timed automaton modes, M1 and M2, for each one of the presented scenarios. In this study, the automaton has been designed as described in Section 4 and with h_0=2 and h_1=2.

In case of an abrupt fault, the combination of the residuals *PEm*, *LOm* and *WPEm* allows to increase the persistence of the fault indicators and to reduce the detection time. Notice that when M2 is activated the fault could be isolated checking the FSM (Figure 4). In the cases in which the detected fault is identified as non-critical, it could be compensated applying fault tolerant control techniques. Moreover, an automaton at mode M0 for detecting the remaining faults could be synthetized.

On the other hand, in case of incipient faults, the fault sensitivity of a given residual structure depends on the inherent noise of the measured variables used. If the measured variables have a high noise index, the resulting residual will provide non-persistent fault indications that could lead to isolability problems in the isolation phase. This can be
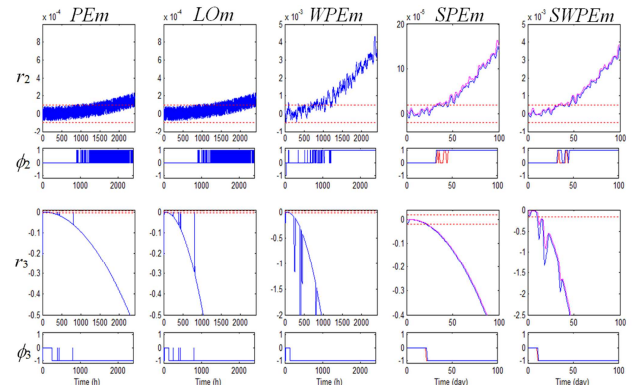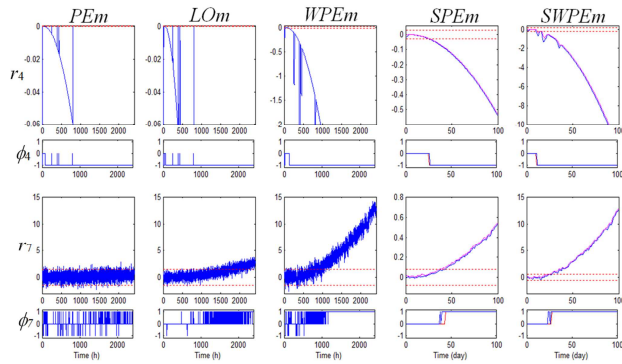

Figure 6. r_2 and r_3 evolution

Figure 7. r_4 and r_7 evolution

TABLE IV. FAULT DETECTION DELAY

f_s	PEm t_d (h)	LOm t_d (h)	WPEm t_d (h)	SPEm t_d (days)	SWPEm t_d (days)
ϕ_2	(900 *np*)	(905 *np*)	(66 *np*) 1192	31	(32 *np*) 45
ϕ_3	221	103	102	21	10
ϕ_4	44	24	102	26	11
ϕ_7	(77 *np*)	(1129 *np*) 2233	(1 *np*) 1020	40	26

TABLE V. M2 AND M3 ACTIVATION DELAY

	Scenario 1	Scenario 2	Scenario 3
M2_activation	1	1	24
M3_activation	3	3	25

noticed in the case of residuals *PEm*, *LOm* and *WPEm*. The problem of persistence could be solved by using residuals *SPEm* and *SWPEm* that act as filters, since by evaluating the information coming from them is easier to predict their evolution. These models also have the advantage that they allow tracking the system evolution with less data than the residuals that operate at an hourly basis. Combining the information coming from those residuals in the right way, it is possible to characterize whether an incipient fault is present and reset the automata to mode M0 to continue monitoring the system to detect new faults.

VI. CONCLUSIONS

This paper has proposed a fault detection methodology for incipient faults that combines different residual generation methods (observers, *l*-step ahead predictors and re-sampled models) with different convergence velocities to the real output trying to benefit from the advantages offered by each scheme. The integration is based on generating a timed automaton which combines the information extracted from each method in order to provide the best fault detection performance regarding incipient faults. The proposed methodology has as a main objective to detect as early as possible anomalies or incipient faults in system components. The fault detection performance has been illustrated in a Barcelona water network case stu. As further work, the integration of the proposed fault detection methodology for incipient faults with the fault isolation proposed in [11] will be studied allowing not only the diagnosis but also the prognosis.

REFERENCES

[1] I. Roychoudhury, G. Biswas, and X. Koutsoukos, "A Bayesian Approach to Efficient Diagnosis of Incipient Faults," in *Prod. 17th international Workshop on Principles of Diagnosis*, Peñaranda de Duero, Spain, 2006.

[2] Y. Zhang, and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual Reviews in Control*, vol. 32, no.2, pp. 229–252, 2008.

[3] V. Hölttä, and H. Koivo, "Quality index framework for plant-wide performance evaluation," *Journal of Process Control*, vol. 19, no. 7, pp. 1143-1148, 2009.

[4] J. Meseguer, V. Puig, T. Escobet, and J. Saludes, "Observer gain effect in linear interval observer-based fault detection", *Journal of Process Control*, vol 20, no. 8, pp. 944-956, 2010.

[5] J. Meseguer, V. Puig, T. Escobet, and S. Tornil, "Approximating Fault Detection Linear Interval Observers using l-Order Interval Predictors," in *Proc. of the European Control Conference*, Kos, Greece, 2007.

[6] V. Venkatasubramanian, R. Rengaswamy, and S. Kavuri, "A review of process fault detection and diagnosis - part i: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, pp. 293–311, 2003.

[7] V. Venkatasubramanian, R. Rengaswamy, and S. Kavuri, "A review of process fault detection and diagnosis - part ii: Qualitative models and search strategies," *Comput. Chem. Eng.*, vol. 27, pp. 313–326, 2003.

[8] V. Venkatasubramanian, R. Rengaswamy, and S. Kavuri, "A review of process fault detection and diagnosis - part iii: Process history based methods," *Comput. Chem. Eng.*, Vol. 27, pp. 327–346, 2003.

[9] S. X. Ding, *Model-based Fault Diagnosis Techniques*, Springer, 2008.

[10] M. Yu, D. Wang, M. Luo, and L. Huang, "Prognosis of Hybrid Systems With Multiple Incipient Faults: Augmented Global Analytical Redundancy Relations Approach," *IEEE Trans. Syst., Man, and Cybern - part A: Syst. and Hum.*, vol. 42, no. 3, pp. 1083-4427, 2011.

[11] J. Meseguer, V. Puig, and T. Escobet, "Fault Diagnosis Using a Timed Discrete-Event Approach Based on Interval Observers: Application to Sewer Networks," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 40, no. 5, pp. 900-916, 2010.

[12] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, *Diagnosis and Fault-Tolerant Control*. Springer, 2nd edition, 2006.

[13] M. Krysander, J. Åslund, and m. Nyberg, "An efficient algorithm for finding minimal over-constrained sub-systems for model-based diagnosis," *IEEE Trans. Syst., Man, Cybern. A*, vol. 38, no. 1, pp. 197–206, 2008.

[14] L. Travé-Massuyès, T. Escobet, and X. Olive, "Diagnosability analysis based on component supported analytical redundancy relations," *IEEE Transactions on Systems, Man, and Cybernetics-Part A*, vol. 36, no. 6, pp. 1146–1160, 2006.

[15] M. Basseville and I.V. Nikiforov, *Detection of Abrupt Changes : Theory and Application*. Prentice Hall, Inc., 1993.

[16] V. Puig, J. Quevedo, T. Escobet, F. Nejjari, S. De Las Heras, "Passive Robust Fault Detection of Dynamic Processes Using Interval Models," *Control Systems Technology, IEEE Transactions on* , vol.16, no.5, pp.1083,1089, Sept. 2008.

[17] S. Tornil-Sin, C. Ocampo-Martinez, V. Puig, and T. Escobet, "Robust fault detection of non-linear systems using set-membership state estimation based on constraint satisfaction," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 1, pp. 1-10, 2012.

[18] M. R Maurya, R. Rengaswamy, and V. Venkatasubramanian, "Fault diagnosis using dynamic trend analysis: A review and recent developments," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 2, pp. 133-146, 2007

[19] C. Cassandras, and S. Lafortune, *Introduction to Discrete Event Systems*, Springer, 2008.

[20] L. Ouedraogo, A. Khoumsi, and M. Nourelfath, "SetExp: a method of transformation of timed automata into finite state automata," *Real-Time Syst*, vol. 46, pp. 189-250, 2010.

[21] J. Quevedo, V. Puig, G. Cembrano, J. Blanch, J. Aguilar, D. Saporta, G. Benito, M. Hedo, and A. Molina, "Validation and reconstruction of flow meter data in the Barcelona water distribution network," *Control Engineering Practice*, Vol. 18, no. 6, pp- 640-651, 2010.

[22] A. Rosich, E. Frisk, J. Åslund, R. Sarrate, and F. Nejjari,"Fault Diagnosis Based on Causal Computations," *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, vol. 42, no. 2, pp. 371-381, 2012.