# Model- vs. data-based approaches applied to fault diagnosis in potable water supply networks

M. À. Cugueró-Escofet, J. Quevedo, C. Alippi, M. Roveri, V. Puig, D. García and F. Trovò

## ABSTRACT

The problem of fault diagnosis in potable water supply networks is addressed in this paper. Two different fault diagnosis approaches are proposed to deal with this problem. The first one is based on a model-based approach exploiting a priori information regarding physical/temporal relations existing among the measured variables in the monitored system, providing fault detection and isolation capabilities by means of the residuals generated using these measured variables and their estimations. This a priori information is provided by the topology and the physical relations between the elements constituting the system. Alternatively, the second approach relies on a data-driven solution meant to exploit the spatial and temporal relationships present in the acquired data streams in order to detect and isolate faults. Relationships between data streams are modelled using sequences of linear dynamic time-invariant models, whose estimated coefficients are used to feed a hidden Markov model. Afterwards, a cognitive method based on a functional graph representation of the system isolates the fault when existing. Finally, a performance comparison between these two approaches is carried out using the Barcelona water supply network, showing successful and complementary results which suggest the integrated usage in order to improve the results achieved by each one separately.

**Key words** | cognitive systems, critical infrastructure systems, fault isolation, hidden Markov models, model-based fault diagnosis, time series

**M. À. Cugueró-Escofet** (corresponding author)
**J. Quevedo**
**V. Puig**
**D. García**
Safety and Automatic Control Research Center (CS2AC),
Polytechnic University of Catalonia (UPC),
Terrassa Campus, Gaia Research Bldg., Rambla Sant Nebridi, 22,
Terrassa, Barcelona 08222,
Spain
E-mail: *miquel.angel.cuguero@upc.edu*

**C. Alippi**
**M. Roveri**
**F. Trovò**
Department of Electronics and Information,
Politecnico di Milano,
Piazza Leonardo da Vinci,
Milano 32 I-20133,
Italy

## INTRODUCTION

Water networks are complex large-scale systems needing highly sophisticated supervisory and control schemes in order to satisfy a certain degree of performance when unfavourable faulty conditions are occurring. To deal with this problem, the use of a fault detection and isolation (FDI) system capable of detecting and isolating these faults (or events) is highly desirable, aiming to help the operators to identify which is the actual event occurring in the water network. The FDI problem applied to water networks has been extensively studied from various perspectives and at different levels (see, e.g., Lees 2000; Colombo & Karney 2002; Misiunas *et al.* 2006). On the one hand, at the district metered area (DMA) level, many FDI approaches have addressed the problem of leak/burst detection and isolation (see, e.g., Mounce *et al.* 2009, 2011; Mounce & Boxall 2010; Wu *et al.* 2010; Bicik *et al.* 2011; Palau *et al.* 2011; Perez *et al.* 2011, 2014; Xia *et al.* 2011; Romano *et al.* 2013, 2014a, 2014b; Sanchez-Fernandez *et al.* 2015; Veldman de Roo *et al.* 2015), where different approaches are applied. Sensor data validation and reconstruction when exploiting the temporal redundancy of the sensor measurements is also addressed in several works (see, e.g., Prescott & Ulanicki 2001; Filion *et al.* 2007; Quevedo *et al.* 2010; Eliades & Polycarpou 2012; Farley *et al.* 2012; Cugueró-Escofet *et al.* 2016). The problem of water quality monitoring concerning contamination event detection has also been extensively

doi: 10.2166/hydro.2016.218

addressed (see, e.g., Eliades & Polycarpou 2010). Moreover, regarding DMA monitoring (either for leaks or water quality), there is also a related problem involving optimal sensor placement in order to maximize the performance of the FDI algorithms applied. This problem has been studied separately both for leak detection and location (see, e.g., Ostfeld & Salomons 2004; Krause *et al.* 2008; Pérez *et al.* 2010; Wu 2011; Casillas *et al.* 2013; Cugueró-Escofet *et al.* 2015) and quality monitoring (see, e.g., Eliades & Polycarpou 2010). On the other hand, at the water supply level (i.e., the network connecting the water potabilization plants with the water distribution tanks) less research has been carried out (see, e.g., Ragot & Maquin 2006; Quevedo *et al.* 2014). The water supply networks, also referred to as trunk main systems, are regional networks used to supply water to the cities and villages of a certain region. This kind of network can be analysed using a flow-driven model, that is, using mass balance linear relations, alternatively to water distribution networks, which are generally modelled using pressure-driven models implying non-linear non-explicit relations. The use of mass balance relations for modelling regional water supply networks is appropriate because an actuator is typically installed in each pipe, which establishes its flow. Of course, the energy balances could also be formulated in this case, but this would add extra complexity which is actually not needed since the goal is to establish analytical redundancy relations (ARRs) between flow sensors. This is the case, e.g., in Quevedo *et al.* (2014), where the problem of sensor data validation and reconstruction (which is addressed for DMA networks in Quevedo *et al.* (2010) exploiting the temporal redundancy of the sensor measurements) has been extended to water supply networks, considering combined temporal/spatial redundancy models. In Ragot & Maquin (2006), a model-based FDI approach is applied to the Nancy water network, a city in the north-eastern French department of Meurthe-et-Moselle, in order to detect faults in the sensors. The present paper also moves towards the FDI application to a water supply network by proposing two different fault diagnosis approaches: a model-based approach using a priori information of the system, i.e., the physical relation between its elements, and a data-driven approach, which is able to exploit a priori information about the network topology to perform fault diagnosis but does not require any additional information about the physical models of the water supply network. According to the literature, model-based approaches rely on the concept of analytical redundancy (Blanke *et al.* 2006), which is based on the use of software sensors, i.e., models using available sensor historic records in order to estimate the desired sensor measurement, as an alternative to hardware-based approaches, which rely on the use of extra hardware sensors. Although hardware redundancy is desirable in critical elements, the use of the latter in large-scale water networks may be dramatically expensive because of the installation, calibration and maintenance actions to be performed on the system when considering this approach.

The fault diagnosis problem in critical infrastructure systems, such as potable water supply networks (PWSNs), involve the answers to some common questions formulated in general fault diagnosis problems, such as if there is a fault affecting the system (fault detection stage) or which is the actual faulty element in this system (fault isolation stage). Also, sometimes it is important to know the magnitude of the fault occurring in order to decide the importance of this fault and the corresponding actions to be taken. The novelty of this paper is not only to compare two well-accepted and promising general purpose fault diagnosis methods (one model-based, the other data-based), but also to determine the main features of each method and which is the best way to combine them in order to optimize the overall performance at both fault detection and fault isolation stages, when considering PWSNs as is the case here. Specifically, the Barcelona PWSN is used as the case study in this work. In ideal situations, the use of a model obtained from the physical relations, as considered in the first approach, should lead to the optimal solution. However, it may be noted that analytical models may be affected by several system practical issues, such as the potential uncertainty on the model parameters (e.g., actual tank surface), the difficulty of having an on-line well-calibrated model due to frequent network topology changes (caused by, e.g., new elements like tanks added or blocked pipes resulting from maintenance operations) and common changes in the consumers' demand behaviour, which are hard to determine in real-time operation. Hence, a data-based approach, as suggested in the second method, is also a useful and effective alternative to the use of analytical

models obtained from physical/temporal relations existing in the network.

The structure of the paper is as follows: the next section presents both the FDI model-based method combining both spatial and time series (TS) models, and the data-based approach based on a cognitive fault diagnosis system (CFDS) method exploiting hidden Markov models (HMMs). The case study, based on the Barcelona PWSN, is presented next. This is followed by a section in which fault isolation results obtained by each methodology are presented, compared and discussed. Finally, conclusions and ongoing work are outlined.

## FAULT DIAGNOSIS METHODOLOGY

In this work, two different well-accepted general purpose fault diagnosis methodologies are used to address the FDI problem in PWSNs. The first approach is based on checking the consistency between the observed and the nominal system behaviour by means of a set of physical/temporal parity relations (PTPR), which relates the measured system variables under normal (faultless) operation assumption of the monitored system. The novelty of the PTPR approach presented here relies on the combined usage of TS residuals with physical residuals, by means of a classical FDI residual-based approach which generally considers only physical parity relations. An inconsistency is detected when ARRs derived from models do not match the measurements, generating a non-null residual. Then, the fault diagnosis mechanism is activated in order to isolate the possible fault by matching the residuals against the fault signature matrix (see, e.g., Puig *et al.* 2006). The proposed FDI scheme is shown in Figure 1, where $u_i(k)$ is the $i$-th measured system input, $x_i(k)$ is the $i$-th measured system output, $\hat{x}_i(k)$ is the $i$-th predicted system output, $r_i(k)$ is the $i$-th system residual of the complete set of $n_r$ residuals and $\mathbf{s}(k)$ is the corresponding fault signature at time instant $k$. Further details on this general fault diagnosis scheme are given later in this section.

The second approach relies on the CFDS method presented in Alippi *et al.* (2013), that is able to exploit spatial and temporal relationships among measured system variables. The considered CFDS, which does not require any a priori information about the physical model of the network, is based on a two-layer hierarchical architecture to detect and isolate faults. In the first layer, a change-detection test (CDT) based on HMMs (see, e.g., Alippi *et al.* 2012) is able to detect a fault occurring in the system by checking the variations in the relationship between couples of data streams, while, in the second layer, a cognitive method based on a functional graph representation of the system is able to isolate the fault occurring. Here, we are proposing a modified version of the original CFDS, specifically crafted for this application, where the available topological information of the water network is integrated into the data-driven approach. For instance, Figure 2(a) represents all the relationships which may exists among data coming from the Can Guey subsystem (see section 'Case study') induced by the water flow physical phenomenon. This initial dependency graph in Figure 2(a) is reduced by taking into account

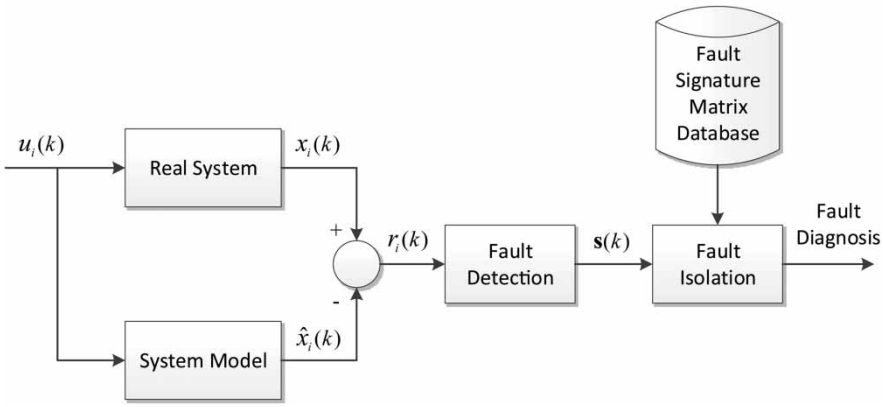


**Figure 1** | FDI block diagram.

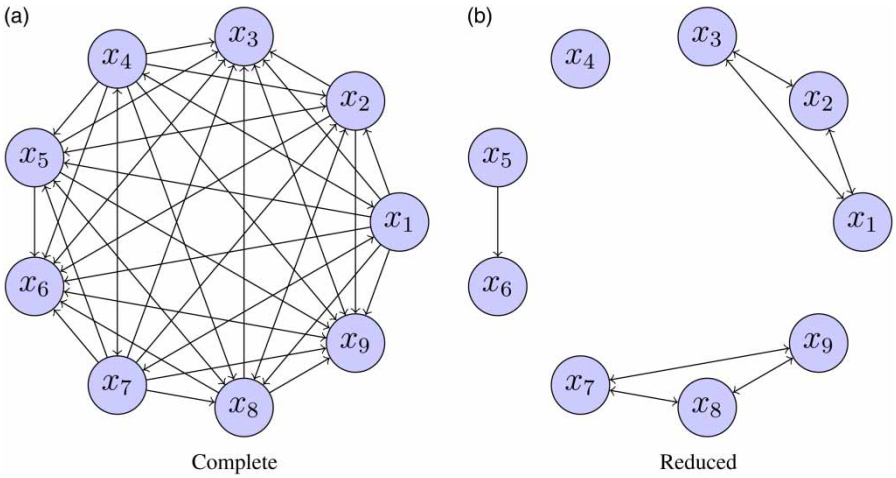


**Figure 2** | CFDS method dependency graphs of Can Guey subsystem.

the correlation analysis and the final outcome of the Can Guey subsystem (Figure 2(b)).

The methodologies presented here aim to detect and isolate faults of different kinds appearing in PWSNs, as discussed later in the section 'Fault scenarios'. These may well represent actual common hydraulic faults occurring and jeopardizing the performance of water networks, e.g., leaks, bursts or sensor communication faults, as further detailed in the same section. Generally, in order to apply these methodologies, the set of considered faults to be addressed should be defined beforehand. This allows to generate the set of relations or data-based models able to detect and isolate the set of specified faults.

## Method I: fault diagnosis based on PTPR

### Residual generation

The fault diagnosis method presented in this section evaluates the nominal residual $r_i(k)$ obtained from the difference between the system measurements and the model prediction, considering the model for the $i$-th subsystem expressed in input–output regressor form as follows:

$$r_i(k) = x_i(k) - \hat{x}_i(k) = x_i(k) - \phi_i^T(k)\theta_i \tag{1}$$

where $\theta_i$ are the nominal parameters obtained using a training dataset, $x_i(k)$, $i \in \{1, \ldots, N\}$ is the sensor measurement at time instant $k$, $N \in \mathbb{N}$ is the number of considered sensors, $\hat{x}_i(k)$ is the model prediction at time instant $k$ and $\phi_i(k)$ is the regressor vector of dimensions $n_{\theta_i} \times 1$, including inputs $(u_i(k), u_i(k-1), u_i(k-2), \ldots)$ and outputs $(y_i(k), y_i(k-1), y_i(k-2), \ldots)$. Considering the uncertainty (e.g., modelling errors, noise) the detection test involves checking the condition

$$r_i(k) \in [\underline{\tau}_i, \bar{\tau}_i] \tag{2}$$

where $\tau_i$ is the detection threshold. This detection threshold can be determined using statistical methods (Basseville & Nikiforov 1993) or set-membership approaches (Puig 2010). When using a set-membership approach the noise is assumed to be unknown but bounded, with a priori known bound. Then, the threshold can be obtained by propagating the uncertainty to the residual computation. In the case of statistical methods, the noise is assumed to follow a normal distribution with known mean value $\mu_i$ and standard deviation $\sigma_i$ (Ding 2013). Then, the threshold of the $i$-th residual can be determined as follows: $\tau_i = \mu_i + 3\sigma_i$, including the 99.7% of the values of a normal distribution according to the *3-sigma* rule. On the other hand, when using a set-membership approach, the noise is assumed to be unknown but bounded, with a priori known bound. Then, the threshold can be obtained by propagating the uncertainty to the residual computation (Puig 2010). Using either one or the other approach, the threshold in Equation (2) is determined to include the values of the whole residual

distribution in the faultless situation and, hence, it may be used for fault detection purposes. This threshold is also useful to provide prediction interval bounds for the data forecasting process, so test condition (2) can be equivalently expressed as follows:

$$x_i(k) \in [\underline{\hat{x}}_i(k), \bar{\hat{x}}_i(k)] \tag{3}$$

where $\bar{\hat{x}}_i(k) = \hat{x}_i(k) + \tau_i$ and $\underline{x}_i(k) = \hat{x}_i(k) - \tau_i$, respectively.

*Spatial consistency residuals*. The PTPR method is based on establishing mass balance equations for the water network constitutive elements. As an example, the mass balance expression for the $i$-th tank can be stated by means of the following discrete-time difference equation:

$$y_i(k+1) = y_i(k) + \frac{\Delta t}{A_i}(q_{in_i}(k) - q_{out_i}(k)) \tag{4}$$

where $y_i(k)$ is the tank level, $A_i$ is the cylindrical tank surface, $q_{in_i}(k)$ is the manipulated inflow and $q_{out_i}(k)$ is the outflow, which may include manipulated tank outflow and consumer demands, both given in $m^3/s$.

Similarly, in the water supply network nodes, the mass balance may be expressed as the static equation

$$\sum_i q_{in_i}(k) = \sum_i q_{out_i}(k) \tag{5}$$

where, similarly to Equation (4), $q_{in_i}(k)$ and $q_{out_i}(k)$ correspond to the inflow and outflow of the $i$-th subnet node, also given in $m^3/s$.

*TS residuals*. Additional residuals can be obtained considering that level in tanks and flow in demand sectors have a daily repetitive behaviour which can be modelled using a TS model. TS models take advantage of the temporal redundancy of the measured variables. A widely used method for signal forecasting is the Holt-Winters (HW) triple exponential smoothing approach (Winters 1960; Makridakis *et al.* 1998). This method, which is in wide use because of its simplicity and performance, may be presented in several different versions, e.g., additive or damped trend, additive or multiplicative seasonality, single or multiple seasonality. Here the additive single seasonality version is considered, which may be implemented as shown next for a forecasting horizon $\ell$:

$$\hat{x}_{ts}(k) = \hat{R}(k-\ell) + \ell\hat{G}(k-\ell) + \hat{S}(k-L) \tag{6}$$

where $\hat{x}_{ts}(k)$ is the TS model forecasted value, $\hat{R}$ is the estimate of the deseasonalized level (i.e., removing the seasonal effect $\hat{S}(k-L-\ell)$), $\hat{G}$ is the estimate of the trend and $\hat{S}$ is the estimate of the seasonal component, which may be respectively stated as follows:

$$\begin{aligned}\hat{R}(k-\ell) = {} & \alpha(x(k-\ell) - \hat{S}(k-\ell-L)) \\ & + (1-\alpha)(\hat{R}(k-\ell-1) \\ & + \hat{G}(k-\ell-1)) \quad 0 < \alpha < 1\end{aligned} \tag{7}$$

$$\begin{aligned}\hat{G}(k-\ell) = {} & \beta(\hat{R}(k-\ell) - \hat{R}(k-\ell-1)) \\ & + (1-\beta)\hat{G}(k-\ell-1) \quad 0 < \beta < 1\end{aligned} \tag{8}$$

$$\begin{aligned}\hat{S}(k-\ell) = {} & \gamma(x(k-\ell) - \hat{R}(k-\ell)) \\ & + (1-\gamma)\hat{S}(k-\ell-L) \quad 0 < \gamma < 1\end{aligned} \tag{9}$$

where $L$ is the season (i.e., daily here) periodicity, $\alpha$, $\beta$ and $\gamma$ are the HW parameters (level, trend and season smoothing factors, respectively) and $x$ is the measured value (i.e. $y_m$, $q_{in_m}$ or $q_{out_m}$, depending on the TS model considered). Hence, analysing the historic records of the measured values in a certain sensor, a HW model is derived and used to validate the current data acquired by this device.

## FDI scheme

The FDI methodology (Figure 1) is based on determining the *actual fault signature* $\mathbf{s}(k) = [s_1(k), s_2(k), \ldots s_{n_r}(k)]$ of the system for $n_r$ different residuals, as a result of the fault detection phase (see the section 'Residual generation') as follows:

$$s_i(k) = \begin{cases} 0, & \text{if } |r_i(k)| < \tau_i \text{ (no fault)} \\ 1, & \text{if } |r_i(k)| \geq \tau_i \text{ (fault)} \end{cases} \tag{10}$$

where $\tau_i$ is the threshold associated with the $i$-th residual. The actual fault signature is compared against the *theoretical fault*

*signature matrix* (FSM) $\Sigma$ that binary codifies the influence of each fault in the set of considered faults $f_1, f_2, \ldots f_{n_f}$ on every residual in the set of considered residuals $r_1, r_2, \ldots r_{n_r}$. This matrix has as many rows as residuals and as many columns as considered faults. If $\Sigma_{ij} = 1$, the $j$-th fault appears in the expression of the $i$-th residual; otherwise $\Sigma_{ij} = 0$. Assuming classical FDI fault hypotheses, i.e., single faults and no-compensation (exoneration), fault isolation consists in looking for a column of $\Sigma$ matching the actual fault signature $\mathbf{s}(k)$. More details about the algorithm implementation of this FDI scheme, including how to manage the temporal aspects of the binarized residuals, can be found in Meseguer *et al.* (2010) and Puig & Blesa (2013).

## Method II: fault diagnosis system based on the cognitive approach

The considered CFDS method is based on the ability to characterize the functional dependencies among the streams of acquired data, where each functional dependency models the temporal and spatial relationships between couples of data streams. The main characteristics of the CFDS are the ability to work without any a priori information about the physical models of the system and the possibility to isolate the potential faults by exploiting a functional graph representation of the system. Details about the considered CFDS can be found in Alippi *et al.* (2013). In this applicative scenario, we considered the possibility to include the a priori information about the topology of the water supply network. In fact, the physical phenomenon of the water flow induces a causality among the respective acquired data streams, allowing those relationships in which this causality principle does not hold to be discarded.

The overall system architecture is presented in Figure 3, in which it may be observed how once the parameter vectors have been estimated from data, they are used in the learning phase to create the dependency graph and learn the nominal state of the process. During the operational phase, newly estimated functional relationship parameter vectors are used to evaluate their statistical likelihood with respect to the learned nominal state through the HMM-CDT. Each relationship is evaluated over time using a HMM-based CDT. When a change in the estimated parameters (i.e., a fault) is detected, the cognitive fault diagnosis module is activated in order to extract information about it, e.g., the time instant when the fault is produced and the fault location.

In more detail, let $x_i$ be the $i$-th stream over the $N$ sources of data within the considered sensors network. Let $g_{(i,j)}, i, j \in \{1, \ldots, N\}$ be the functional relationship (in terms of transfer function) between the $i$-th and the $j$-th sensor. All the functional relationships present in the acquired data can be modelled as a *dependency graph* $\mathcal{G}$, where the nodes of this graph represent the $N$ sensors of the network, while the arcs between nodes represent the

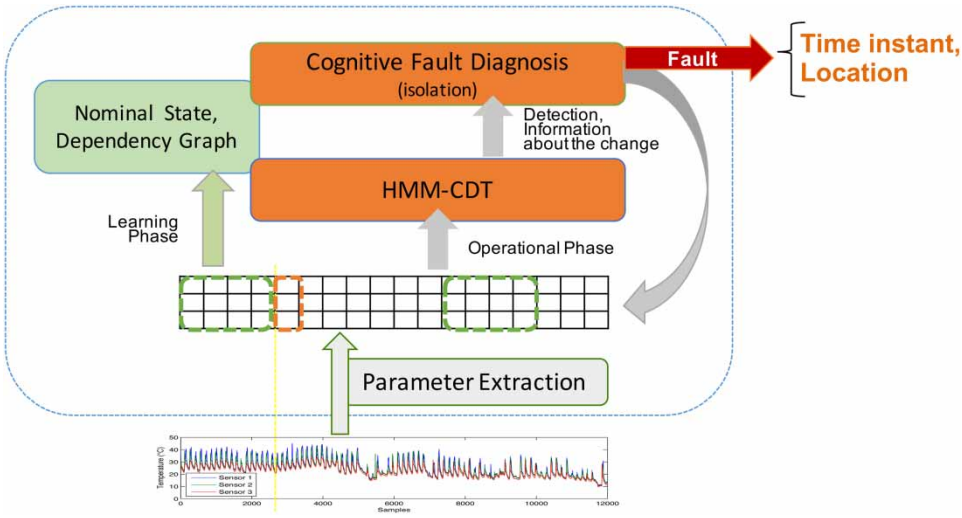


**Figure 3** | General architecture of the considered CFDS method.

functional relationships. We emphasize that generally, in real-world situations, not all the relationships are meaningful, since there might be nonexistent or very weak relationships between two streams of data. To address this issue, we only consider those relationships not contradicting the causality provided by the physical phenomenon and, subsequently, we resort to the use of the analysis of the linear correlation between $x_i$ and $x_j$ to define a level of dependency associated with $g_{(i,j)}$. More specifically, to remove weakly correlated streams of data, those data streams with cross-correlation peak absolute value below a certain threshold $\xi_{\min}$, with $0 \leq \xi_{\min} \leq 1$, have been removed (e.g., $\xi_{\min}$ can be set to 0.5 or larger, depending on the network complexity, to keep only highly correlated data streams). This led to the definition of a *reduced dependency graph* $G = \{V, E\}$, where $V$ and $E$ are the set of nodes representing the $N$ network sensors and the set of arcs representing relationships characterized by higher correlation (i.e., cross-correlation above $\xi_{\min}$), respectively. Figures 2(b) and 4 show examples of the reduced dependency graph for Can Guey and Orioles systems, respectively.

Functional relationships $g_{(i,j)}$ in $G$ are modelled either by a linear time-invariant (LTI) dynamic system or by a sequence of LTI dynamic systems following the HMM hypothesis (i.e., the Markov memoryless property of stochastic processes). Among the wide range of LTI dynamic systems, we focus on single-input single-output (SISO) models such as autoregressive with exogenous input (ARX) models, autoregressive moving average with exogenous input (ARMAX) models or output error (OE) models (Ljung 1999) in their predictive form, i.e., $g^{\theta}_{(i,j)}$ parametrized in $\theta \in \mathbb{R}^p$. Here, $\theta$ represents the parameter vector of the considered predictive models, while $p$ represents the cardinality of $\theta$.

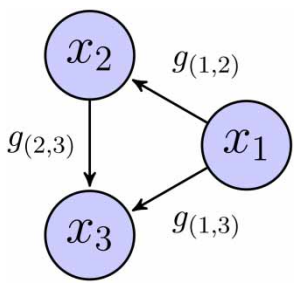


**Figure 4** | Example of a reduced dependency graph with three sensors (i.e., $x_1$, $x_2$ and $x_3$) and three relationships (i.e., $g_{(1,2)}$, $g_{(1,3)}$ and $g_{(2,3)}$).

Assuming that the data generating process satisfies the exponential stability for closed loop and following the hypotheses on $g^{\theta}_{(i,j)}$ stated in Ljung (1978), the theoretical results in Ljung (1978) grant that:

$$\sqrt{N_T}P^{-\frac{1}{2}}(\hat{\theta} - \theta^*) \sim \mathcal{N}(0, I_p) \quad \text{when } N_T \to \infty, \tag{11}$$

where $P \in \mathbb{R}^{p \times p}$ is the covariance matrix of the estimated parameter $\hat{\theta}$ of the model, $N_T$ is the length of the training set used to estimate $\hat{\theta}$ with the least square method, $\theta^*$ is the optimal configuration of the parameters within the chosen model family and $I_p$ is the identity matrix. From Equation (11) it may be stated that, given $N_T$ sufficiently large, the distribution of the estimates $\hat{\theta}$s follows a multivariate Gaussian with mean vector $\theta^*$ and covariance matrix $P$. This theoretical result led us to consider HMMs ruled by a mixture of Gaussians (GMM) to model the statistical behaviour of estimated parameters $\hat{\theta}$ over time. In more detail, the HMM is defined as:

$$\mathcal{H} = \{n, \mathcal{F}, A, \pi\}, \tag{12}$$

where $n$ is the number of states, $\mathcal{F} = \{p_1, \ldots, p_n\}$ is the set of probability density functions (PDFs) associated with each state, $A \in [0, 1]^{n \times n}$ is the state transition probability matrix and $\pi \in [0, 1]^n$ is the initial state distribution vector. It is worth noting that Equation (11) allows us to model the PDF associated with the nominal state by using a GMM. Let

$$p_i(\hat{\theta}|\Phi_i) = \sum_{k=1}^{K_i} w_{k,i}\mathcal{N}(\hat{\theta}|\mu_{k,i}, \Sigma_{k,i}) \tag{13}$$

be the GMM associated with the $i$-th state, where $K_i$ is the number of Gaussian mixtures for the $i$-th state, $w_{k,i}$ is the weight for state $i$ and Gaussian mixture $k$,

$$\Phi_i = [\mu_{1,i}, \ldots, \mu_{k_i,i}, \Sigma_{1,i}, \ldots, \Sigma_{k_i,i}] \tag{14}$$

with $\mu_{k,i}$ and $\Sigma_{k,i}$ the mean vector and the covariance matrix for state $i$ and Gaussian mixture $k$, respectively.

The analysis of the evolution over time of estimated parameters $\hat{\theta}$s by means of a HMM is the core mechanism of the HMM-based CDT. More specifically, for each

$g_{(i,j)} \in V$, the sequence of $\hat{\theta}$s are estimated by overlapping windows of $N_T$ observations. The HMM $\mathcal{H}_{(i,j)}$, associated with relationship $g_{(i,j)}$, is trained on the estimated parameters computed from the first $T_0$ observations, i.e., the training dataset. Then, after the training phase, the HMM-based CDT keeps on estimating the parameter $\hat{\theta}$ by overlapping windows of data, and let $\hat{\theta}^s$ be the parameter estimated on the $s$-th window. The log-likelihood of the HMM is as follows:

$$l_{(i,j)}(s) = P(\hat{\theta}^{s-\omega+1}, \ldots, \hat{\theta}^s | \mathcal{H}_{(i,j)}) \tag{15}$$

where $\omega$ is the considered log-likelihood window length. This value indicates how likely the sequence of parameters $\hat{\theta}^{s-\omega+1}, \ldots, \hat{\theta}^s$ has been generated by $\mathcal{H}_{(i,j)}$. When $l_{(i,j)}$ decreases below an automatically defined threshold (i.e., the minimum value assumed by the log-likelihood in training or validation scaled by a user-defined coefficient factor), a change in the relationship $g_{(i,j)}$ is detected. In other words, the goal of the HMM-based CDT is to inspect variations in the distribution of the estimated parameters $\hat{\theta}^s$ over time with regard to that learned during the initial training (performed on data up to $T_0$). Whenever, the trained HMM is no more able to explain the statistical behaviour of recently estimated parameters, the likelihood decreases and a change in the relationship is detected. This is the purpose of a CDT, whose mechanisms aim to detect discrepancies with regard to an (autonomously defined) nominal state. Other solutions, e.g., recursive ARX estimation, cannot be applied for change detection purposes since they are meant to track the system over time (hence adapting to variations over time) and are not able to inspect for variations. Details about the HMM-based CDT can be found in Algorithm 1 of Alippi *et al.* (2013) and in Alippi *et al.* (2012).

Once a change in one of the HMM-based CDTs is detected, the cognitive fault diagnosis layer is activated to isolate the fault within the system. The basic idea of this cognitive isolation mechanism is as follows: when a fault affects a sensor, all the relationships connected to that sensor should be affected by this change. Hence, by looking at the likelihood of all the relationships in $V$, we are able to identify the sensor of the system that has been affected by the fault, and thus to isolate it. Details about the cognitive level presented here can also be found in Algorithm 2 of Alippi *et al.* (2013).

## CASE STUDY

### Barcelona PWSN

The Barcelona PWSN (Figures 5 and 6), considered as a case study here, is distributed in 23 different districts covering 424 km$^2$ area and providing water to about three million end users. Due to the geographical topology of Barcelona and its surroundings, the water network in the metropolitan area is structured in 113 pressure floors. This allows two management levels to be established: the supply network and, at a smaller detail level, the distribution network that is divided in DMAs. The first level is in charge of the right choice of water sources that supply water to the system at each moment (quality and quantity of the water supply) and of the bulk water transferences. On the other hand, the second level is responsible for delivering water from the reservoirs to the end consumers, optimizing the pressure profiles in order to minimize losses due to leakage. This paper is mainly focused on the supply level. Water managed by this supply network is obtained from both surface and underground sources, including Ter (surface source) and Llobregat (both surface and underground source) as the most important ones in terms of usage and capacity. The water supplied by these sources is distributed through a network of around 4,645 km of pipes to 218 DMAs or demand sectors. The current AGBAR Control Centre has a tele-control system for the network management. The supervisory control system installed in the Control Centre is in charge of optimally controlling the whole network by taking into account operational restrictions and consumer demands, including about 400 controlled/monitored points, 63 storage tanks, 84 pumps and 46 valves. Looking at Figure 5, it can be noted that the Barcelona supply network is constituted by multiple similar tank subsystems which behave alike. Since considering the whole network would not allow the models and results obtained to be analysed in detail, the results presented in this paper are focused on the analysis of a single tank subsystem and a three-tank subsystem (Orioles and Can Guey in Figure 6). Similar results are

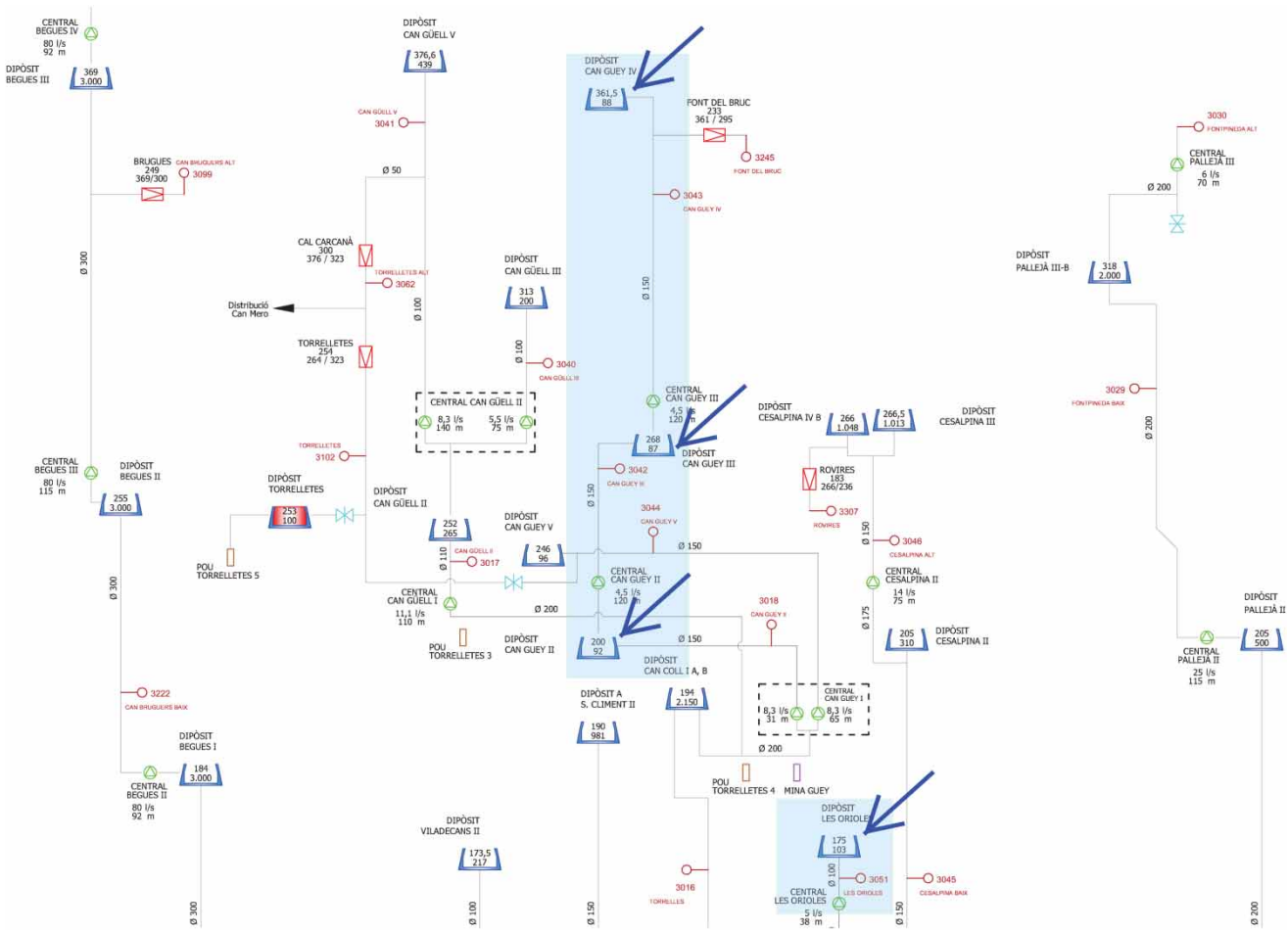


**Figure 5** | Barcelona PWSN detail, including Orioles and Can Guey subsystems (arrows).

also achieved when considering alternative tank subsystems in the same PWSN, since they follow the same structure. Regarding data management, the Barcelona telecontrol system receives real-time data mainly from the available flow meters, usually installed in the DMAs supply points, so their readings closely fit the actual DMAs water demand.

In this paper, the two fault diagnosis methods introduced in previous sections are tested using a simulation of the Barcelona PWSN. This simulation uses a model calibrated and validated with real data provided by the company managing the network. This model has been accepted by the company as a good representation of the actual water network behaviour, and is used for its operational control (Pascual *et al.* 2013), e.g., optimizing water production and transport costs, guaranteeing a minimum volume in the tanks for eventual emergencies and smoothing operations of the actuators to extend the life of the equipment. These objectives may be configured with the graphical user interface (GUI) of this simulator (Figure 7), which is developed with MATLAB/SIMULINK, a widely used numerical computing and programming platform in many research institutions and industrial enterprises, which makes it a convenient prototyping and development framework. Existing hydraulic well-known simulators such as EPANET (Rossman 2000) are pressure-driven, i.e., are not able to simulate the mass balance of the supply network (needed here) without providing additional physical parameters of this network. These additional physical details may not always be available or sufficiently accurate and are actually not needed for the mass balance computation used here, but for the pressure simulations. In contrast, the supply network simulator used

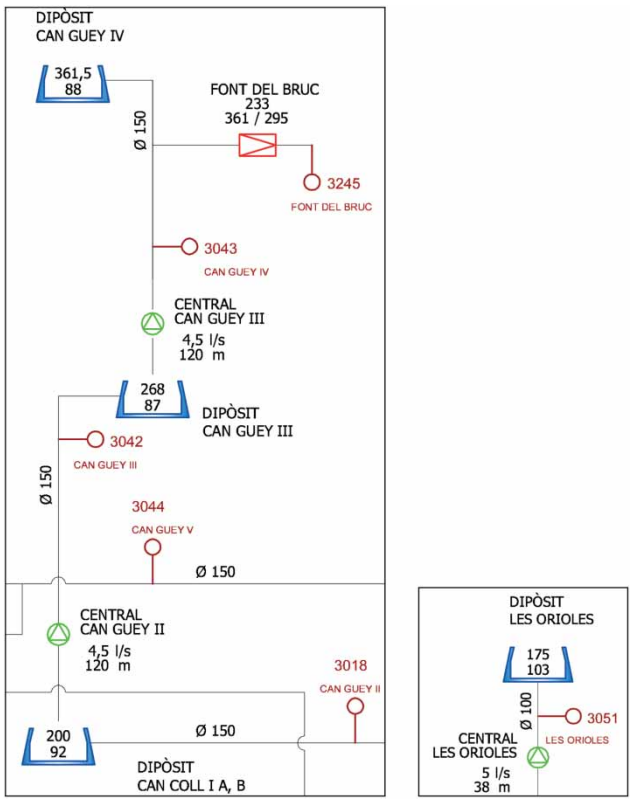


**Figure 6** | Orioles and Can Guey subsystems within Barcelona PWSN.

here utilizes topology and limited network element details (such as tank surfaces) and historic records of sensor measurements TS in order to provide the required mass balance information needed. Furthermore, this supply network simulator has been provided with additional ad hoc features useful for a different range of applications, such as for testing the control system (based on model predictive control) or a fault module, in order to simulate different kinds of hydraulic faults in different elements of the network, which are not available in more standard hydraulic simulators such EPANET. Specifically, the fault module allows providing synthetic scenarios of the network under study and designing and testing new control schemes and FDI approaches such as the ones presented here. In Figure 7, the whole network SIMULINK model is contained in the *Network* block, and is composed of different elements (blocks), such as tanks, nodes, pumps, valves and demands. Each demand of this supply network is actually a DMA of hundreds to thousands of users. Also, each actuator may integrate several pumps or valves working in parallel.

Without loss of generality, the results presented here are focused on two subsystems within the Barcelona water supply network (Figure 5), known as Orioles (Figure 8) and Can Guey (Figure 9), in order to illustrate the performance of the fault diagnosis methodologies presented. This part of the network includes the following elements:

- **Tanks:** *d175LOR, d200CGY, d268CGY, d361CGY*
- **Actuators with flow sensors:** *iOrioles, iCanGuey1d2, iCanGuey2, iCanGuey3*
- **Demands with flow sensors:** *c175LOR, c200CGY, c268CGY, c361CGY*
- **Level sensors:** *xd175LOR, xd200CGY, xd268CGY, xd361CGY*

## Residual definition

In Figures 8 and 9, $q_{in}$, $q_{out}$ and $y$ are the incoming tank flow, consumer demand and tank level, respectively, and $q_{in_m}$, $q_{out_m}$ and $y_m$ are the corresponding measured values. The corresponding discrete-time model equations, including the considered faults, are as follows:

- Tank (level):

$$y(k+1) = y(k) + \frac{\Delta t}{A}[q_{in}(k) - q_{out}(k)] \tag{16}$$

- Pump (flow):

$$q_{in}(k) = \Gamma_p(k) q_p(k) + f_p(k) \tag{17}$$

- Tank level sensor:

$$y_m(k) = \Gamma_{y_m}(k) y(k) + f_{y_m}(k) \tag{18}$$

- Pump flow sensor:

$$q_{in_m}(k) = \Gamma_{q_{in_m}}(k) q_{in}(k) + f_{q_{in_m}}(k) \tag{19}$$

- Demand sector flow sensor:

$$q_{out_m}(k) = \Gamma_{q_{outm}}(k) q_{out}(k) + f_{q_{outm}}(k) \tag{20}$$

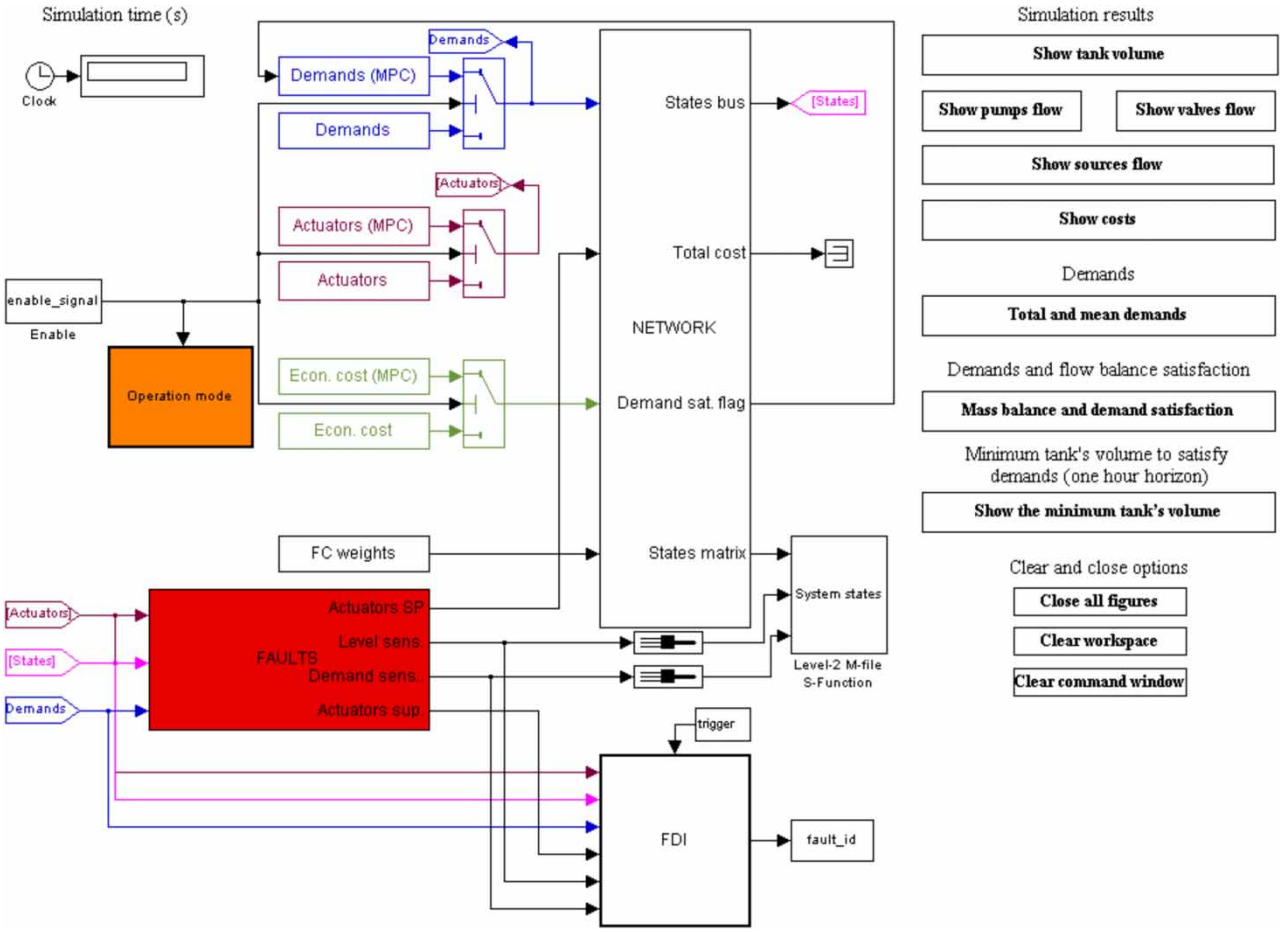


**Figure 7** | MATLAB/SIMULINK Barcelona supply network simulator GUI.

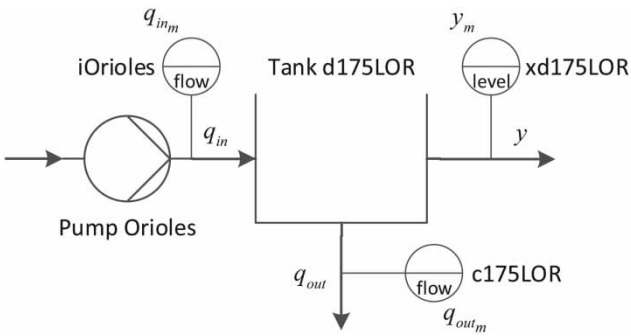


**Figure 8** | Orioles subsystem.

where $y(k)$ is the actual tank level, $y_m(k)$ is the measured tank level, $q_{out}(k)$ is the actual demand flow, $q_{out_m}(k)$ is the measured demand flow, $q_{in}(k)$ is the actual input tank flow, $q_p(k)$ is the set-point pump flow, $q_{in_m}(k)$ is the measured input flow, $\Gamma_c(k)$ is the multiplicative fault signal component related to element $c$, $f_c(k) = \beta_c(k)\chi_c(k)$ is the additive fault signal component related to element $c$, with $\beta(k)$ time profile and $\chi(k)$ behaviour, $\Delta t$ is the sampling time and $A$ is the cylindrical tank surface.

Spatial consistency residuals can be obtained using the mass flow model (Equation (16)) and the sensor measurements (Equations (17)–(20)) in a non-faulty situation. In particular, the following residual $r_{i,1}$ may be obtained using the tank model (Equation (16)) for the $i$-th tank subsystem using measured variables

$$r_{i,1}(k) = y_{m_i}(k) - \hat{y}_{sc_i}(k) \tag{21}$$

with

$$\hat{y}_{sc_i}(k) = y_{m_i}(k-1) + \frac{\Delta t}{A_i}\left[q_{in_{m_i}}(k-1) - q_{out_{m_i}}(k-1)\right] \tag{22}$$

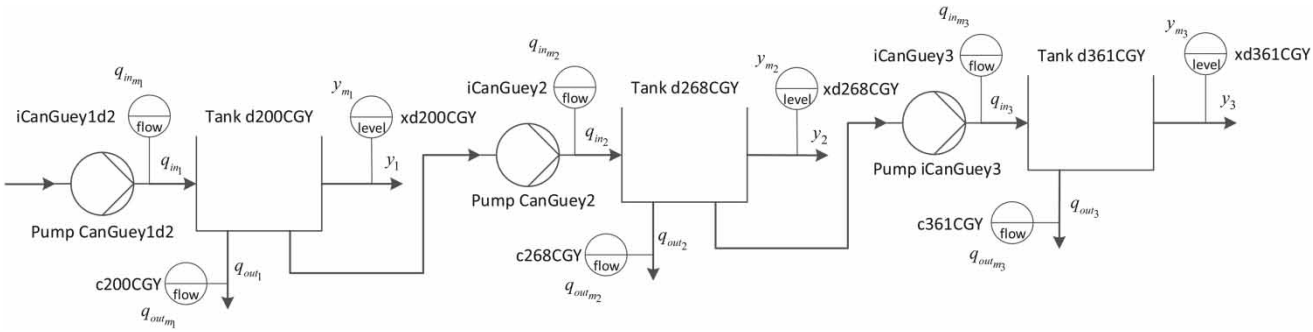


**Figure 9** | Can Guey subsystem.

where $\hat{y}_{sc_i}$ is the tank level spatial consistency estimation, $A_i$ is the cylindrical tank surface, $q_{in_{m_i}}$ is the manipulated measured inflow and $q_{out_{m_i}}$ is the measured outflow, both given in $m^3/s$.

Furthermore, for each input and output with periodic behaviour of the $i$-th tank subsystem, a TS HW model can be derived and the following ARRs may be obtained:

- Tank (level) TS:

$$\hat{y}_{ts_i}(k) = g(y_{m_i}(k-1),\ \ldots,\ y_{m_i}(k-L)) \tag{23}$$

- Demand sector flow TS:

$$\hat{q}_{out_{ts_i}}(k) = h(q_{out_{m_i}}(k-1),\ \ldots,\ q_{out_{m_i}}(k-L)) \tag{24}$$

- Pump flow TS:

$$\hat{q}_{in_{ts_i}}(k) = l(q_{in_{m_i}}(k-1),\ \ldots,\ q_{in_{m_i}}(k-L)) \tag{25}$$

where $g$, $h$ and $l$ are the HW TS expressions (Equations (6)–(9)) for the tank level sensor, sector demand sensor and pump flow sensor, respectively, for data exhibiting a periodicity of $L$ samples.

Using the TS models (Equations (23)–(25)), the following residuals are obtained:

$$r_{i,2}(k) = y_{m_i}(k) - \hat{y}_{ts_i}(k) \tag{26}$$

$$r_{i,3}(k) = q_{out_{m_i}}(k) - \hat{q}_{out_{ts_i}}(k) \tag{27}$$

$$r_{i,4}(k) = q_{in_{m_i}}(k) - \hat{q}_{in_{ts_i}}(k) \tag{28}$$

It may be noted that TS residuals in Equations (26)–(28) can only be applied for those sensors presenting periodic behaviour. This is applicable to all the faulty elements considered in this work and also to the major part of the elements of the network, but still there are some which are not expected to evolve periodically (e.g., pump stations not directly serving water to a demand sector). Hence, in order to apply PTPR fault isolation methodology using TS residuals, elements involved should be previously checked for periodic behaviour. As a counterpart, TS residuals may be computed with information provided by a single sensor, which may be advantageous in certain applications with non-obvious/non-existent model relation among sensors, e.g., intelligent sensors with embedded diagnosis unit (Alippi *et al.* 2012).

From residuals (21), (26)–(28) and Equations (17)–(20), the theoretical FSM for the subsystems considered (Figures 8 and 9) is presented in Table 1. In the latter, the sensitivity of each residual to each fault is detailed by means of a 0 (i.e., non-sensitive) or a 1 (i.e., sensitive) in the corresponding element of the matrix, obtaining a fault signature from each of its columns. Also, ordinal index $i$ is assigned for each tank subsystem as follows: $i = 1$ for d175LOR, $i = 2$ for d200CGY, $i = 3$ for d268CGY and $i = 4$ for d361CGY tank subsystem, respectively. Moreover, it may also be observed how spatial consistency residuals are used for fault detection here, since $r_{i,1}$ is sensitive to all the considered faults within each $i$ single tank subsystem, while TS residuals are employed for fault isolation purposes.

**Table 1** | Faults' signatures

| | $f_{y_m}$ | $f_{q_{out_m}}$ | $f_{q_{in_m}}$ | $f_p$ | $f_{y_{m1}}$ | $f_{q_{out_{m1}}}$ | $f_{q_{in_{m1}}}$ | $f_{p1}$ | $f_{y_{m2}}$ | $f_{q_{out_{m2}}}$ | $f_{q_{in_{m2}}}$ | $f_{p2}$ | $f_{y_{m3}}$ | $f_{q_{out_{m3}}}$ | $f_{q_{in_{m3}}}$ | $f_{p3}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $r_{1,1}$ | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{1,2}$ | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{1,3}$ | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{1,4}$ | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{2,1}$ | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{2,2}$ | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{2,3}$ | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{2,4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{3,1}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| $r_{3,2}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| $r_{3,3}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| $r_{3,4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| $r_{4,1}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| $r_{4,2}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| $r_{4,3}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| $r_{4,4}$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |

The resulting PTPR fault diagnosis scheme implemented in the Barcelona PWSN system, involving the different elements detailed in this section, is shown in Figure 10.

## Fault scenarios

The Barcelona PWSN simulator allows the introduction of faults of different kinds in distinct elements of this water

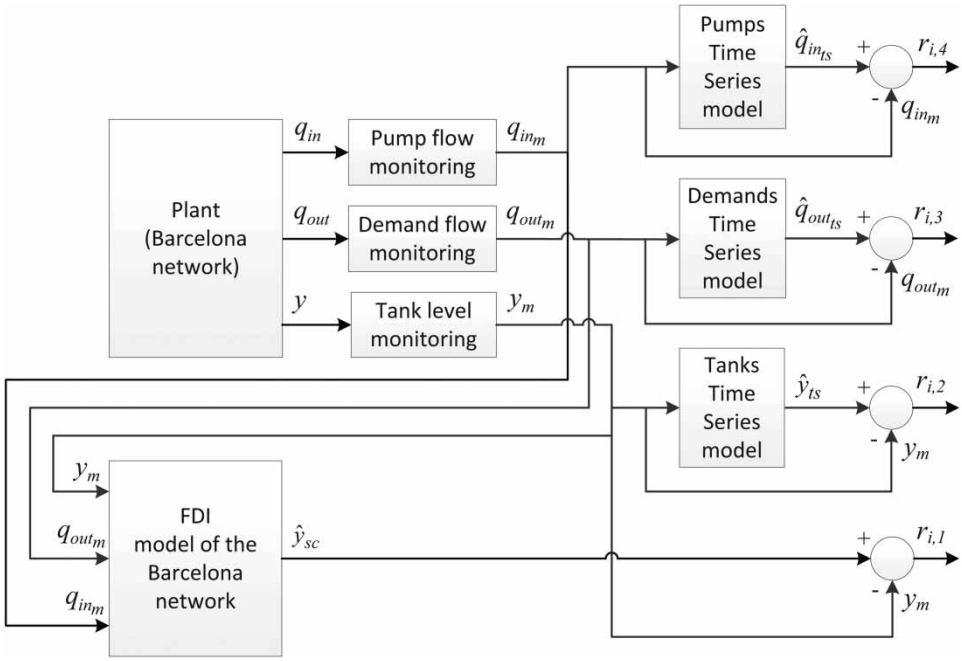


**Figure 10** | FDI scheme for Barcelona PWSN.

network. Here, faults of freezing, offset and drift nature are considered:

- **Freezing:** $\Gamma(k) = 0$ and $\chi(k) = \kappa$ for $k \geq k_f$
- **Offset:** $\Gamma(k) = 1$ and $\chi(k) = \eta$ for $k \geq k_f$
- **Drift:** $\Gamma(k) = 1$ and $\chi(k) = R(k - k_f)$ for $k \geq k_f$

where $\kappa$ and $\eta$ are constants, $R$ denotes the ramp function of a certain slope and $k_f$ is the time instant when the fault is occurring.

Moreover, the faults considered are either of abrupt or incipient nature, as defined by their time profile $\beta(k)$ as follows:

- Abrupt:

$$\beta(t) = \begin{cases} 0, & k < k_f \\ 1, & k \geq k_f \end{cases}$$

- Incipient:

$$\beta(t) = \begin{cases} 0, & k < k_f \\ 1 - e^{-\rho(k-k_f)}, & k \geq k_f \end{cases}$$

where $\rho > 0$ is the constant characterizing the evolution of the corresponding fault and $k_f$ is the time instant when the fault occurs. The faults presented in this section are meant to be generic, but may well represent actual common hydraulic faults occurring in water networks, e.g., leaks (which may be represented by offset/drift abrupt/incipient faults), bursts (which may be represented by offset abrupt faults) or sensor communication faults (which may be represented by freezing abrupt faults).

Different fault scenarios are defined in order to test and compare the methods presented here, all including random normally distributed measurement noise of ±1% full scale. The dataset considered to implement these fault scenarios lasts for seven months, with a sampling period $T = 1$ hour and a fault appearing at $t_f = 744 \cdot T$ in different elements of the Barcelona PWSN subsystems considered (Figures 8 and 9, respectively):

- iOrioles pump sensor ($f_{q_{in_m}}$) in Orioles subsystem
- c175LOR demand sensor ($f_{q_{out_m}}$) in Orioles subsystem
- iCanGuey1d2 pump sensor ($f_{q_{in_{m_1}}}$) in Can Guey subsystem
- iCanGuey2 pump sensor ($f_{q_{in_{m_2}}}$) in Can Guey subsystem
- iCanGuey3 pump sensor ($f_{q_{in_{m_3}}}$) in Can Guey subsystem
- c268CGY demand sensor ($f_{q_{out_{m_2}}}$) in Can Guey subsystem
- c361CGY demand sensor ($f_{q_{out_{m_3}}}$) in Can Guey subsystem

The way these faults affect the hydraulic network components is represented in the models (19) and (20). These fault scenarios are part of a fault benchmark used in the framework of the Seventh Framework Program European project iSense (ref. FP7-ICT-2009-6) as a collaborative dataset provided by the Polytechnic University of Catalonia to be used by all the partners involved. The parametrization of the faults involved in this benchmark is depicted in Table 2.

## Methods setting

The HMM-based CDT uses ARX linear models for the extraction of the parameters $\theta$. In the case of Orioles subsystem (Figure 8), the relationship patterns among the measured tank level, the input flow and the measured demands are modelled. The dependency graph is learned by considering all the binary relationships with autocorrelation greater or equal to $\xi_{\min} = 0.5$. The result is the graph presented in Figure 4 with $(x_1, x_2, x_3) = (y_m, q_{in_m}, q_{out_m})$. In the case of Can Guey subsystem (Figure 9), we only considered those relationships compatible with the causality given by the water flow phenomenon (i.e., those in Figure 2(a)) and selected those having autocorrelation greater or equal to $\xi_{\min} = 0.8$, where the higher threshold is due to the increased complexity of the network. Under faulty conditions, the considered relationships will exhibit changes that will depend on the kind and magnitude of the fault introduced, thus their monitoring is useful for fault isolation purposes.

Regarding the PTPR initialization, the first 13 days of data are used as training dataset to identify the model parameters, the next 13 days are used as validation dataset to obtain the corresponding fault detection threshold and the rest of the data is used as test dataset. Regarding the HMMs, also a total of 26 days are used for training and validation purposes: 23 and 25 days are used for training the

**Table 2** | Faults' parametrization and diagnosis results

| | | | PTPR method | | | | | CFDS method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Delay [# of samples] | | | | | Delay [# of samples] | | | | |
| Id. | Type of fault | Magnitude | Detection | Isolation | FP [%] | FN [%] | Iso. [%] | Detection | Isolation | FP [%] | FN [%] | Iso. [%] |
| 1 | Offset abr. iOrioles | 10% MFD | 2 | 4 | 0 | 7.34 | 22.22 | 4 | 4 | 7.44 | 94.52 | 5.48 |
| 2 | Offset abr. iOrioles | 25% MFD | 2 | 2 | 0 | 2.37 | 79.16 | 3 | 3 | 11.57 | 4.11 | 95.89 |
| 3 | Offset inc. iOrioles | 10% MFD | 12 | 23 | 0 | 3.04 | 4.16 | 35 | 35 | 12.40 | 82.19 | 17.81 |
| 4 | Offset inc. iOrioles | 25% MFD | 9 | 13 | 0 | 4.53 | 52.77 | 16 | 16 | 0.00 | 27.40 | 72.60 |
| 5 | Drift abr. iOrioles | 1% MFD | 9 | 13 | 0 | 2.71 | 73.61 | 8 | 8 | 13.22 | 10.96 | 89.04 |
| 6 | Drift abr. iOrioles | 10% MFD | 3 | 3 | 0 | 2.78 | 84.72 | 4 | 4 | 1.65 | 5.48 | 94.52 |
| 7 | Drift inc. iOrioles | 1% MFD | 12 | 23 | 0 | 3.54 | 59.72 | 18 | 18 | 9.92 | 24.66 | 75.34 |
| 8 | Drift inc. iOrioles | 10% MFD | 7 | 7 | 0 | 4.09 | 77.77 | 4 | 4 | 8.26 | 8.22 | 91.78 |
| 9 | Offset abr. c175LOR | 10% MFD | 1 | 3 | 0 | 0.02 | 65.27 | 3 | 3 | 0.00 | 4.11 | 95.89 |
| 10 | Offset abr. c175LOR | 25% MFD | 1 | 3 | 0 | 0.02 | 65.27 | 1 | 1 | 0.00 | 1.37 | 98.63 |
| 11 | Offset inc. c175LOR | 10% MFD | 10 | 21 | 0 | 0.29 | 34.72 | 47 | 47 | 0.00 | 64.38 | 35.62 |
| 12 | Offset inc. c175LOR | 25% MFD | 7 | 11 | 0 | 0.13 | 58.33 | 65 | 65 | 0.00 | 89.04 | 10.96 |
| 13 | Drift abr. c175LOR | 1% MFD | 7 | 7 | 0 | 0.11 | 59.72 | 33 | 33 | 0.00 | 45.21 | 54.79 |
| 14 | Drift abr. c175LOR | 10% MFD | 2 | 4 | 0 | 0.04 | 63.88 | 4 | 4 | 0.00 | 5.48 | 94.52 |
| 15 | Drift inc. c175LOR | 1% MFD | 10 | 15 | 0 | 0.25 | 52.77 | 39 | 39 | 0.00 | 53.42 | 46.58 |
| 16 | Drift inc. c175LOR | 10% MFD | 5 | 7 | 0 | 0.11 | 59.72 | 8 | 8 | 38.02 | 10.96 | 89.04 |
| 17 | Freezing abr. iOrioles | – | 7 | 12 | 0 | 7.22 | 8.33 | 17 | 17 | 10.74 | 68.49 | 31.51 |
| 18 | Freezing inc. iOrioles | – | 19 | 36 | 0 | 7.52 | 5.55 | 3 | 3 | 22.31 | 73.97 | 26.03 |
| 19 | Freezing abr. c175LOR | – | 11 | 14 | 0 | 16.79 | 6.94 | 33 | 33 | 0.00 | 45.21 | 54.79 |
| 20 | Freezing inc. c175LOR | – | 59 | – | 0 | 92.84 | 0 | 58 | 58 | 4.13 | 79.45 | 20.55 |
| 21 | Offset abr. iCanGuey1d2 | 15% MFD | 2 | 4 | 0 | 1.37 | 39.73 | 21 | 22 | 0 | 30.56 | 69.44 |
| 22 | Offset inc. iCanGuey1d2 | 15% MFD | 9 | 26 | 0 | 12.33 | 8.22 | 32 | 50 | 0 | 69.44 | 30.56 |
| 23 | Drift abr. iCanGuey1d2 | 15% MFD | 3 | 4 | 0 | 2.74 | 93.15 | 5 | 11 | 0 | 15.28 | 84.72 |
| 24 | Freezing abr. iCanGuey1d2 | – | 3 | 8 | 0 | 9.59 | 71.23 | 12 | 21 | 0 | 29.17 | 70.83 |
| 25 | Offset abr. iCanGuey2 | 15% MFD | 2 | – | 0 | 1.37 | – | 14 | 29 | 0 | 40.28 | 59.72 |
| 26 | Offset inc. iCanGuey2 | 15% MFD | 8 | 26 | 0 | 12.33 | 2.74 | 32 | 53 | 0 | 73.61 | 26.39 |
| 27 | Drift abr. iCanGuey2 | 15% MFD | 3 | 4 | 0 | 2.74 | 93.15 | 7 | 9 | 0 | 12.50 | 87.50 |
| 28 | Freezing abr. iCanGuey2 | – | 3 | 4 | 0 | 5.48 | 67.12 | 11 | 12 | 0 | 16.67 | 83.33 |
| 29 | Offset abr. c361CGY | 15% MFD | 2 | 4 | 0 | 1.37 | 93.15 | 3 | 3 | 0 | 4.17 | 95.83 |
| 30 | Offset inc. c361CGY | 15% MFD | 9 | 11 | 0 | 10.96 | 82.19 | 10 | 12 | 0 | 16.67 | 83.33 |
| 31 | Drift abr. c361CGY | 15% MFD | 3 | 4 | 0 | 2.74 | 93.15 | 3 | 3 | 0 | 4.17 | 95.83 |
| 32 | Freezing abr. c361CGY | – | 10 | 12 | 0 | 34.25 | 10.96 | 9 | 11 | 0 | 15.28 | 84.72 |
| 33 | Offset abr. c268CGY | 15% MFD | 2 | 4 | 0 | 1.37 | 93.15 | 1 | 1 | 0 | 1.39 | 98.61 |
| 34 | Offset inc. c268CGY | 15% MFD | 8 | 11 | 0 | 9.59 | 80.82 | 3 | 9 | 0 | 12.50 | 87.50 |
| 35 | Drift abr. c268CGY | 15% MFD | 3 | 4 | 0 | 2.74 | 93.15 | 2 | 2 | 0 | 2.78 | 97.22 |
| 36 | Freezing abr. c268CGY | – | 10 | 12 | 0 | 56.16 | 6.85 | 3 | 11 | 0 | 15.28 | 84.72 |
| 37 | Offset abr. iCanGuey3 | 15% MFD | 2 | 4 | 0 | 1.37 | 63.01 | 5 | 7 | 0 | 9.72 | 90.28 |

(*continued*)

**Table 2** | continued

| | | | PTPR method | | | | | CFDS method | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Delay [# of samples] | | | | | Delay [# of samples] | | | | |
| Id. | Type of fault | Magnitude | Detection | Isolation | FP [%] | FN [%] | Iso. [%] | Detection | Isolation | FP [%] | FN [%] | Iso. [%] |
| 38 | Offset inc. iCanGuey3 | 15% MFD | 8 | 28 | 0 | 28.77 | 12.33 | 28 | 29 | 0 | 40.28 | 59.72 |
| 39 | Drift abr. iCanGuey3 | 15% MFD | 3 | 4 | 0 | 2.74 | 93.15 | 4 | 4 | 0 | 5.56 | 94.44 |
| 40 | Freezing abr. iCanGuey3 | – | 3 | 10 | 0 | 26.03 | 64.38 | 9 | 13 | 0 | 18.06 | 81.94 |

MFD, maximum flow/demand.

Orioles and the Can Guey cases, respectively, while the remaining days are used to compute the threshold for detection and validation (Alippi *et al.* 2013). The orders of ARX models have been chosen by means of a validation procedure. The log-likelihood window length has been set to $\omega = 10$ and the batch size has been set to $N_T = 96$ and $N_T = 100$ for the Orioles and Can Guey subsystems, respectively.

### Figures of merit

The numerical results are presented by means of different figures of merit. The first stage in fault diagnosis deals with faultless vs. faulty situation discrimination. The performance achieved in this fault detection stage is measured by the next indices:

- **Detection delay:** Number of samples needed by the fault diagnosis method to detect a certain fault.
- **False positives (FP):** Percentage of test dataset faultless samples (i.e., not affected by a certain fault) that are determined as faulty by the fault detection method. FP corresponds to false alarms in FDI terminology (see Blanke *et al.* 2006).
- **False negatives (FN):** Percentage of test dataset faulty samples (i.e., affected by a certain fault) that are determined as faultless by the fault detection method within the 72 samples (i.e., 3 days) after a fault is produced. FN corresponds to missed alarms in FDI terminology (see Blanke *et al.* 2006).

Moreover, the second stage in fault diagnosis involves fault isolation and classification abilities of the FDI method. These may be quantified by different figures of merit, which are defined as follows:

- **Isolation delay:** Number of samples needed by the fault diagnosis method to isolate a certain fault.
- **Isolation index:** Percentage of test dataset faulty samples (i.e., affected by a certain fault) that are properly isolated within the 72 samples (i.e., 3 days) after a fault is produced, considering a certain fault scenario.

## RESULTS

In Table 2, fault diagnosis results achieved by both FDI methods are detailed. In Figures 11 and 12, isolation results obtained by PTPR method for faults Id. 2 and Id. 14 in Table 2 are also depicted. As may be seen in the figures, residuals' evolution fit the fault signatures corresponding to the respective faults occurring in the system when these results are attained, i.e., iOrioles actuator sensor fault ($f_{q_{in_m}}$ in Table 1) and c175LOR demand sensor fault ($f_{q_{out_m}}$ in Table 1). Similar isolation results have been obtained for faults included in Table 2 affecting similar network elements.

On the one hand, Table 2 shows generally better detection and isolation delay results achieved by PTPR method than by CFDS method, for the pump sensors of iOrioles (e.g., fault Id. 3, 4), iCanGuey1d2 (e.g., fault Id. 21, 22, 26), iCanGuey3 (e.g., fault Id. 37, 38) and the demand sensor c175LOR (e.g., fault Id. 11, 12, 13, 15, 19), with the exceptions of some faults with similar or better performance achieved by CFDS affecting iOrioles pump sensor (e.g., fault Id. 1, 2, 5, 6, 8), c175LOR (e.g., fault Id. 10), c268CGY and c361CGY demand sensors (e.g., fault Id. 29

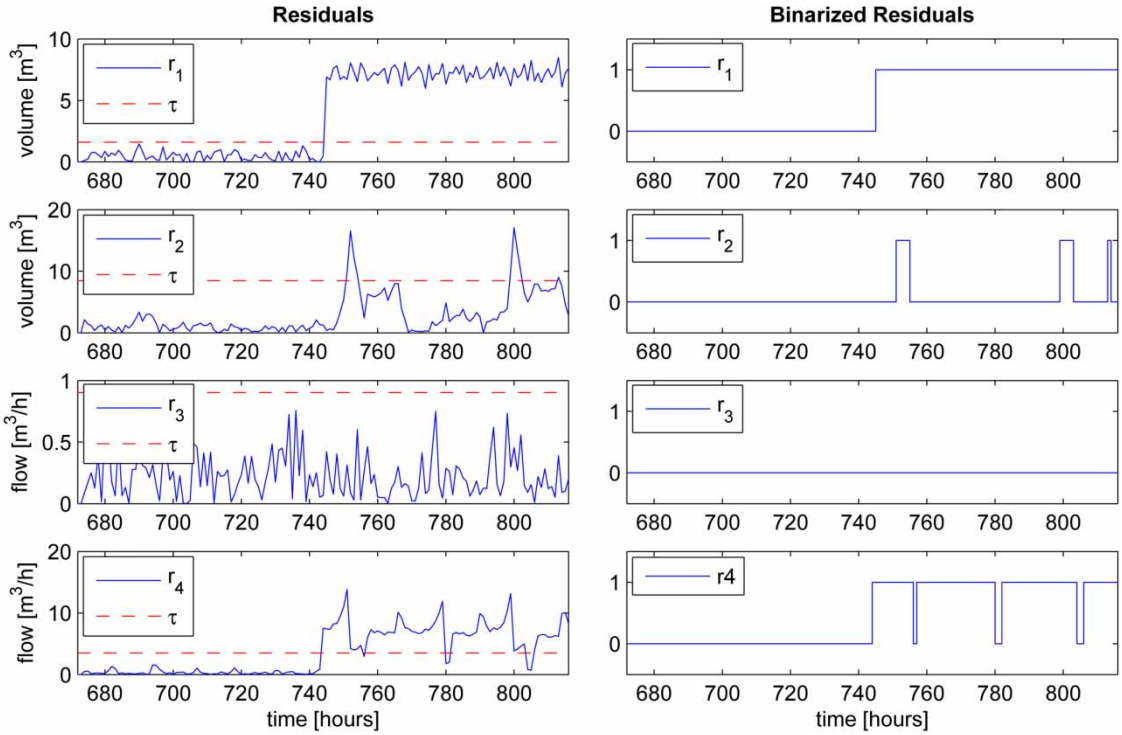


**Figure 11** | Fault Id.2 residuals, PTPR method.

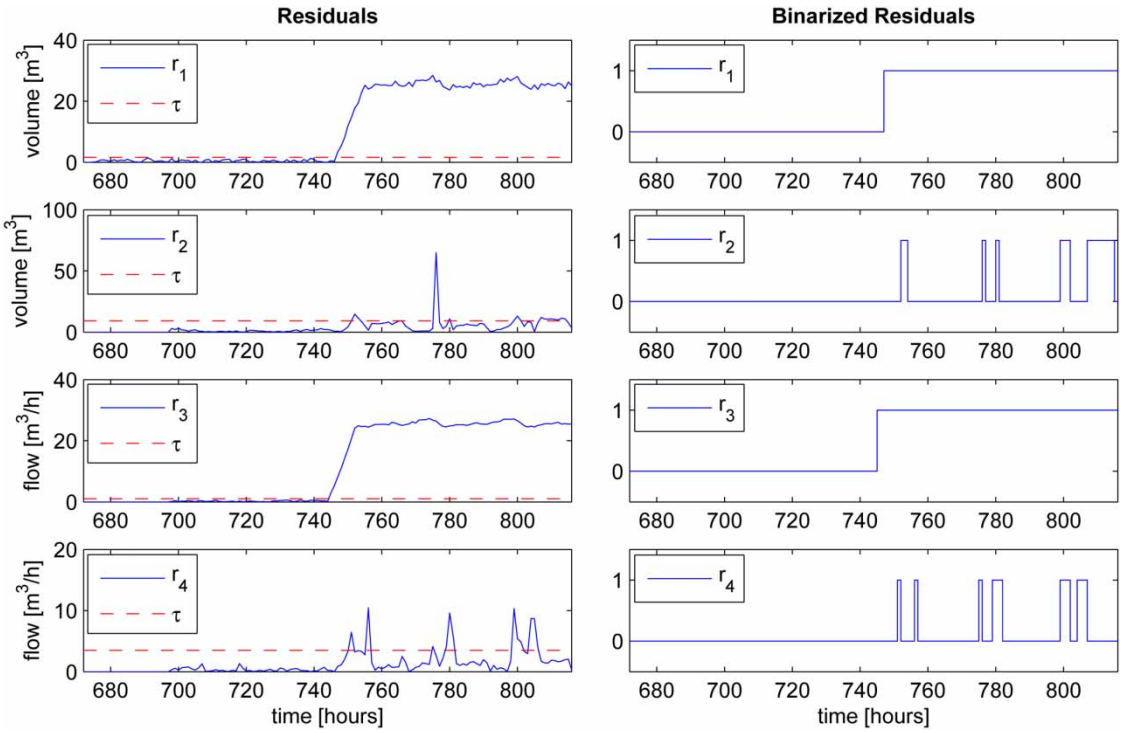


**Figure 12** | Fault Id.14 residuals, PTPR method.

to 36), freezing incipient faults affecting Orioles system (e.g., faults Id. 18, 20) or iCanGuey pump sensor (fault Id. 25) for isolation delay. Also, generally better FP and FN rates are achieved by the PTPR method, with some exceptions regarding freezing nature faults (i.e., faults Id. 20, 32, 36, 40) for FN rates. This overall detection and isolation behaviour leads to a generally quicker and more reliable diagnosis of the faulty component under study achieved by the PTPR method. On the other hand, the CFDS method grants better isolation rates in general (with the exception of faults Id. 1, 12, 13, 15, 23, 24, 27 where the PTPR method provides similar or better results, depending on the scenario considered), which makes it useful to confirm the isolated fault occurring in the system. It is worth noting that the CFDS method provides the performance in Table 2 without assuming any a priori information about the physical model of the system. In the case of faults with an increasing profile, the detection delays of the CFDS are worse than those obtained with the PTPR method. This is due to the fact that the HMM-based CDT is more sensitive to abrupt changes in the parameter distribution. Also, CFDS is generally characterized by higher FP index values. The reason for this behaviour is two-fold: nominal state approximation and process time invariance. First, the nominal model is estimated during an initial training phase that, in principle, could lead to inaccurate models, i.e., model bias due to, e.g., an incorrect selection of the family of models, the lack of enough data for training or the fact that training data do not excite the whole dynamics of the process. This undesired model bias tends to induce FP detections in the testing phase. Second, the process under monitoring could be intrinsically time-varying and not follow the Markov assumption. This leads to FP detection induced by an estimated model which is not able to fully describe the process.

## DISCUSSION

From the light of the results in the previous section, both FDI methods introduced present satisfactory performance for the fault scenarios considered, also showing some complementarity features which suggest possible integration in order to improve the overall fault diagnosis. Specifically, the PTPR method obtains generally lower detection and isolation delays, as well as better FP and FN rates. Hence, it is a good choice for early reliable isolation of the faults appearing in the system, while the CFDS obtains generally better isolation rates, which allows reliable confirmation of the fault being detected and isolated. Regarding the benefits of each method, on the one hand, PTPR is based on physical models describing normal behaviour and does not need to have data from all the possible fault scenarios to perform the fault diagnosis, in contrast to the CFDS method. On the other hand, the main drawback of the PTPR approach is the deep knowledge of the model structure and parameters required to successfully apply this methodology, which is not needed by the CFDS since it is a data-based approach. These facts further motivate the integration of both methods for FDI, taking advantage of the highlights which characterize each one separately.

## CONCLUSIONS

In this work, the application and comparison of two well-accepted general purpose fault diagnosis methods (one model-based, the other data-based) applied to a real PWSN located in the Barcelona area, is developed. Most of the works in the literature addressing fault diagnosis in water networks have treated the problem at the distribution level, but not at the supply level. However, water supply networks have characteristics which allow the application of techniques that cannot be applied to distribution networks. The first method is built upon a model-based approach exploiting a priori information regarding physical/temporal relations which exist between the measured variables within the monitored system, while the second aims at characterizing and detecting changes in the probabilistic pattern sequence of the data coming from this system. Some enhancements of the two approaches considered are introduced, such as the use of TS residuals, not generally considered in classical residual-based fault diagnosis schemes, which traditionally use physical residuals. Successful results have been achieved by both methods, showing good complementary conditions which suggest integrated usage in order to improve the results achieved by each one separately. These results have been tested using heterogeneous types of faults in representative subsystems within the network under study. Future work,

including additional analysis of the fault diagnosis results achieved, considering uncertainty such as noise and modelling/measurement errors derived by the examination of the system under study, is to be done in the ongoing works involving the Barcelona PWSN.

## ACKNOWLEDGEMENTS

This work has been partially funded by the Spanish Government (Ministerio de Economía y Competitividad) and FEDER through the Projects ECOCIS (Ref. DPI2013-48243-C2-1-R) and HARCRICS (Ref. DPI2014-58104-R), and EFFINET grant FP7-ICT-2012-318556 of the European Commission.

## REFERENCES

Alippi, C., Ntalampiras, S. & Roveri, M. 2012 An HMM-based change detection method for intelligent embedded sensors. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, pp. 1–7.

Alippi, C., Ntalampiras, S. & Roveri, M. 2013 A cognitive fault diagnosis system for distributed sensor networks. *IEEE Transactions on Neural Networks and Learning Systems* **24** (8), 1213–1226.

Basseville, M. & Nikiforov, I. V. 1993 *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, New York. Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.77.6896{\&}amp;rep=rep1{\&}amp;type=pdf. doi:10.1016/0967-0661(94)90196-1.

Bicik, J., Kapelan, Z., Makropoulos, C. & Savić, D. 2011 Pipe burst diagnostics using evidence theory. *Journal of Hydroinformatics* **13** (4), 596.

Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M. & Schröder, J. 2006 *Diagnosis and Fault-Tolerant Control*. Springer-Verlag, New York, USA.

Casillas, M., Puig, V., Garza-Castañón, L. & Rosich, A. 2013 Optimal sensor placement for leak location in water distribution networks using genetic algorithms. *Sensors* **13** (11), 14984–15005.

Colombo, A. & Karney, B. 2002 Energy and costs of leaky pipes: toward comprehensive picture. *Journal of Water Resources Planning and Management* **128** (6), 441–450.

Cugueró-Escofet, M. À., Puig, V., Quevedo, J. & Blesa, J. 2015 Optimal pressure sensor placement for leak localisation using a relaxed isolation index: Application to the Barcelona water network. In: *9th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS'15)*. Paris, France, pp. 1108–1113.

Cugueró-Escofet, M. À., García, D., Quevedo, J., Puig, V., Espin, S. & Roquet, J. 2016 A methodology and a software tool for sensor data validation/reconstruction: application to the Catalonia regional water network. *Control Engineering Practice* **49**, 159–172.

Ding, S. X. 2013 *Model-based Fault Diagnosis Techniques – Design Schemes, Algorithms and Tools*. Springer, London, UK.

Eliades, D. & Polycarpou, M. 2010 A fault diagnosis and security framework for water systems. *IEEE Transactions on Control Systems Technology* **18** (6), 1254–1265.

Eliades, D. & Polycarpou, M. 2012 Leakage fault detection in district metered areas of water distribution systems. *Journal of Hydroinformatics* **14** (4), 992–1005.

Farley, B., Mounce, S. & Boxall, J. 2012 Development and field validation of a burst localization methodology. *Journal of Water Resources Planning and Management* **139** (6), 604–613.

Filion, Y., Li, Z. & Buchberger, S. 2007 Temporal and spatial scaling of instantaneous residential water demand for network analysis. In: *World Environmental and Water Resources Congress 2007*. American Society of Civil Engineers, Tampa, FL, pp. 1–10.

Krause, A., Leskovec, J., Guestrin, C., Van Briesen, J. & Faloutsos, C. 2008 Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management* **134** (6), 516–526.

Lees, M. 2000 Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics* **2** (1), 15–34. Available from: http://www.iwaponline.com/jh/002/jh0020015.htm.

Ljung, L. 1978 Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control* **23** (5), 770–783.

Ljung, L. 1999 *System Identification*. Prentice-Hall, Upper Saddle River, NJ, USA.

Makridakis, S., Wheelwright, S. & Hyndman, R. 1998 *Forecasting Methods and Applications*. John Wiley & Sons, New York, USA.

Meseguer, J., Puig, V. & Escobet, T. 2010 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* **40** (5), 900–916.

Misiunas, D., Vítkovský, J., Olsson, G., Lambert, M. & Simpson, A. 2006 Failure monitoring in water distribution networks. *Water Science and Technology* **53** (4–5), 503–511.

Mounce, S. R. & Boxall, J. B. 2010 Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: a case study. *Water Science & Technology: Water Supply* **10** (3), 437–444.

Mounce, S., Boxall, J. & Machell, J. 2009 Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management* **136** (3), 309–318.

Mounce, S. R., Mounce, R. B. & Boxall, J. B. 2011 Novelty detection for time series data analysis in water distribution systems using support vector machines. *Journal of Hydroinformatics* **13** (4), 672.

Ostfeld, A. & Salomons, E. 2004 Optimal layout of early warning detection stations for water distribution systems security. *Journal of Water Resources Planning and Management* **130** (5), 377–385.

Palau, C., Arregui, F. & Carlos, M. 2011 Burst detection in water networks using principal component analysis. *Journal of Water Resources Planning and Management* **138** (1), 47–54.

Pascual, J., Romera, J., Puig, V., Cembrano, G., Creus, R. & Minoves, M. 2013 Operational predictive optimal control of Barcelona water transport network. *Control Engineering Practice* **21** (8), 1020–1034.

Pérez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. & Peralta, A. 2010 Leakage isolation using pressure sensitivity analysis in water distribution networks: Application to the Barcelona case study. In: *12th IFAC Symposium on Large-Scale Systems: Theory and Applications*, France. Vol. 9, pp. 578–584.

Perez, R., Puig, V., Pascual, J., Quevedo, J., Landeros, E. & Peralta, A. 2011 Methodology for leakage isolation using pressure sensitivity analysis in water distribution networks. *Control Engineering Practice* **19** (10), 1157–1167.

Perez, R., Sanz, G., Puig, V., Quevedo, J., Cugueró-Escofet, M. À., Nejjari, F., Meseguer, J., Cembrano, G., Mirats, T. J. & Sarrate, R. 2014 Leak localization in water networks: a model based methodology using pressure sensors applied to a real network in Barcelona [applications of control]. *Control Systems, IEEE* **34** (4), 24–36.

Prescott, S. L. & Ulanicki, B. 2001 Time series analysis of leakage in water distribution networks. *Water Software Systems – Theory and Applications.* **2**, 17–28.

Puig, V. 2010 Fault diagnosis and fault tolerant control using set-membership approaches: application to real case studies. *International Journal of Applied Mathematics and Computer Science* **20** (4), 619–635.

Puig, V. & Blesa, J. 2013 Limnimeter and rain gauge FDI in sewer networks using an interval parity equations based detection approach and an enhanced isolation scheme. *Control Engineering Practice* **21** (2), 146–170.

Puig, V., Stancu, A., Escobet, T., Nejjari, F., Quevedo, J. & Patton, R. 2006 Passive robust fault detection using interval observers: application to the damadics benchmark problem. *Control Engineering Practice* **14** (6), 621–633.

Quevedo, J., Puig, V., Cembrano, G., Blanch, J., Aguilar, J., Saporta, D., Benito, G., Hedo, M. & Molina, A. 2010 Validation and reconstruction of flow meter data in the Barcelona water distribution network. *Control Engineering Practice* **18** (6), 640–651.

Quevedo, J., Chen, H., Cugueró-Escofet, M. À., Tino, P., Puig, V., García, D., Sarrate, R. & Yao, X. 2014 Combining learning in model space fault diagnosis with data validation/reconstruction: application to the Barcelona water network. *Engineering Applications of Artificial Intelligence* **30**, 18–29.

Ragot, J. & Maquin, D. 2006 Fault measurement detection in an urban water supply network. *Journal of Process Control* **16** (9), 887–902.

Romano, M., Kapelan, Z. & Savic, D. 2013 Geostatistical techniques for approximate location of pipe burst events in water distribution systems. *Journal of Hydroinformatics* **15** (3), 634–651.

Romano, M., Kapelan, Z. & Savic, D. 2014a Automated detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management* **140** (4), 457–467.

Romano, M., Kapelan, Z. & Savic, D. 2014b Evolutionary algorithm and expectation maximization strategies for improved detection of pipe bursts and other events in water distribution systems. *Journal of Water Resources Planning and Management* **140** (5), 572–584.

Rossman, L. A. 2000 *EPANET 2 Users Manual*. US Environmental Protection Agency, Washington, DC, USA.

Sanchez-Fernandez, A., Fuente, M. & Sainz-Palmero, G. 2015 Fault detection with distributed PCA methods in water distribution networks. In: *23rd Mediterranean Conference on Control and Automation (MED)*, Torremolinos, pp. 156–161.

Veldman de Roo, F., Tejada, A., van Waarde, H. & Trentelman, H. L. 2015 Towards observer-based fault detection and isolation for branched water distribution networks without cycles. In: *European Control Conference (ECC)*. Linz, Austria. pp. 3280–3285.

Winters, P. R. 1960 Forecasting sales by exponentially weighted moving averages. *Management Science* **6** (52), 324–342.

Wu, Z. Y., Farley, M., Turtle, D., Kapelan, Z., Boxall, J. B., Mounce, S. R., Dahasahasra, S., Mulay, M. & Kleiner, Y. 2011 Chapter 9, Online monitoring and detection. In: Z. Wu (ed.). *Water Loss*. Bentley Systems, pp. 168–189.

Wu, Z., Sage, P. & Turtle, D. 2010 Pressure-dependent leak detection model and its application to a district water system. *Journal of Water Resources Planning and Management* **136** (1), 116–128.

Xia, L., Guojin, L. & Xinhua, Z. 2011 State estimation of municipal water supply network based on BP neural network and genetic algorithm. In: *International Conference on Internet Computing and Information Services (ICICIS)*. Hong Kong, pp. 403–406.

First received 31 October 2015; accepted in revised form 9 March 2016. Available online 10 May 2016