Leakage fault detection in district metered areas of water distribution systems

D. G. Eliades and M. M. Polycarpou

ABSTRACT

Fault tolerance and security in drinking water distribution operations are important issues that have received increased attention in the last few years. In this work the problem of leakage detection is formulated within a systems engineering framework, and a solution methodology to detect leakages in a class of distribution systems is proposed. Specifically, the case where water utilities use standard flow sensors to monitor the water inflow in a District Metered Area (DMA) is considered. The goal is to design algorithms which analyze the discrete inflow signal and determine as early as possible whether a leakage has occurred in the system. The inflow signal is normalized to remove yearly seasonal effects, and a leakage fault detection algorithm is presented, which is based on learning the unknown, time-varying, weekly periodic DMA inflow dynamics using an adaptive approximation methodology for updating the coefficients of a Fourier series; for detection logic the Cumulative Sum (CUSUM) algorithm is utilized. For reference and comparison, a second solution methodology based on night-flow analysis using the normalized inflow signal is presented. To illustrate the solution methodology, results are presented based on randomized simulated leakages and real hydraulic data measured at a DMA in Limassol, Cyprus.

Key words | adaptive learning, leakage detection, water distribution networks

D. G. Eliades (corresponding author)
M. M. Polycarpou
KIOS Research Center for Intelligent Systems and Networks,
ECE Department,
University of Cyprus,
75 Kallipoleos Av.,
CY-1678,
Nicosia,
Cyprus
E-mail: *eldemet@ucy.ac.cy*

NOMENCLATURE

$C(k)$	The CUSUM feature signal
$e_r(k)$	Normalized flow estimation error
G	Update law gain matrix
h_s	Threshold for leakage detection
h_w	Threshold for night-flow leakage detection
K_ξ	Time instances set for seasonal component calculations
K_θ	Time instances set for update-law initial conditions calculations
K_φ	Time instances set for leakage magnitude identification after detection
k	Discrete time
k_d	Leakage detection time
l	Day index
l_d	Day of leakage detection
M	Minimum number of days for night-flow detection
M_δ	Number of days to compute the difference $\delta(\cdot)$
$N(\cdot)$	Set of discrete times corresponding to night-flows
N_z	Number of basis functions
N_μ	Window length for outlier detection
$n(k)$	Multiplicative uncertainty term of DMA inflow
$q^*(k),q(k)$	Sensor and filtered DMA inflow measurement
$q_r(\cdot)$	Normalized DMA inflow
$r(k)$	DMA inflow yearly seasonal signal
$s(k)$	DMA inflow weekly periodic signal
$s_r(k)$	Estimation of the weekly periodic DMA inflow signal
T	Leakage fault start time
t_a,t_b	Discrete times at which night-flow period begins and ends
$w(l)$	Average night-flow
α	Positive update law design parameter

doi: 10.2166/hydro.2012.109

$\beta(k\text{-}T)$	Time profile of a leakage fault with start time T
Δt	Sampling time step
$\delta(\cdot)$	Difference of the average night flow with the minimum average night flow for previous days
$\zeta(\kappa)$	Approximation structure
η	Average outflow of a small leakage
$\theta^*(k)$, $\theta_r(k)$	Optimal and estimated approximation parameters for $s(k)$
$\theta_0(k)$	The offset (DC) term of a Fourier series
μ	Outlier detection threshold
ξ_r	Vector of estimated parameters of the seasonal signal
$\rho(\cdot)$	Approximation function of the seasonal signal
$\varphi(k)$,$\varphi_0(k)$	Actual and estimated average leakage outflow

INTRODUCTION

The International Water Association (IWA) Task Force on Water Losses has defined a set of metrics for water utilities to audit how provided water is consumed or lost. In general, water intended for consumption is segmented into 'Authorized Consumption', which corresponds to the billed or unbilled authorized consumption, and to the 'Water Losses', which corresponds to the 'Apparent Losses', due to unauthorized use, metering inaccuracies or calibration issues, and to the 'Real Losses', due to leakages, breaks, etc. (Alegre *et al.* 2006). Water losses impose a huge economic burden; hence the reduction of both apparent and real losses is an important goal for most water utilities.

Leakage is a type of hydraulic fault, which may be caused due to pipe breaks, loose joints and fittings, as well as due to overflowing water storage tanks (Farley 2001). Some of these problems are prompted by the deterioration of the water delivery infrastructure, which is affected by age and high pressures. Leakage faults which occur within the water distribution network may correspond to slowly developing incipient faults, as well as to abrupt faults, which may require immediate attention. Leakages may cause consumer problems, health risks as well as financial losses (Farley 2001), therefore their early detection and repair, if possible, is imperative. Leakages are classified by water utilities as 'background' (small undetectable leaks for which no action to repair is taken, with single leakage outflow less than 0.25–0.5 $m^3\,h^{-1}$ at 50 m water head), 'unreported' (moderate flow-rates which accumulate gradually and require eventual attention), and 'reported' (high flow-rates which require immediate attention) (Lambert & Morrison 1996; Thornton *et al.* 2008). In practice, there may be a significant time delay between the time a leakage occurs to the time the leakage is detected and the time it is located and repaired (Thornton *et al.* 2008). From a systems and control framework, leakages can be classified as: (a) slowly increasing incipient faults, to describe breaks which in the beginning are small but may deteriorate further while their size increases; (b) stepwise abrupt faults, which appear at a certain time step and whose physical characteristics do not change (Isermann 1997; Zhang *et al.* 2002).

In 1980, the UK Water Authorities Association proposed the concept of District Metered Area (DMA) management methodology to monitor leakage in water distribution networks (WAA 1980). According to Farley (2001), the main benefits of using DMA management over standard centralized monitoring and control are reduced fault detection time, better leakage detection and location isolation, as well as low pressure regulation. This approach has been adopted by a large number of water utilities around the world. Each DMA connects to the water distribution network through one or more water supply pipes; at each supply pipe, real-time flow sensors, pressure sensors and pressure reducing valves are installed (Farley 2001; Morrison 2004).

A number of water utilities apply the *minimum night-flow analysis* at each DMA, to estimate the quantity of water loss due to leakages (Thornton *et al.* 2008). The utilities examine the inflow at a DMA during the minimum consumption hours, e.g. between 12–4 a.m., at which times demand flow and its variance are usually low and the leakage outflow is high due to high DMA pressures. Through manual observation of the minimum flows during the minimum consumption hours and by comparing them to measurements from previous days or to certain thresholds, the water utility operators may be able to detect an unreported leakage fault which has occurred within a DMA (Thornton *et al.* 2008). However, detection of small leakages by this method in practice may not be possible, due to unpredictable variations in consumer demands and measurement noise, as well as due to the effect of long-term trends and seasonality. In addition, it may not be possible to detect incipient leakage faults that may occur due to slowly increasing leakages.

The leakage detection problem was originally formulated as an optimization problem for computing the magnitudes and locations of leakages, based on some flow and pressure measurements; Pudar & Liggett (1992) considered steady-flows for solving this optimization problem, and Liggett & Chen (1994) extended the problem formulation to take into consideration transients, i.e. changes in pressures due to the fact that pressure waves travel through the distribution network with some velocity. They formulated an optimization problem to compute the pipe friction coefficients to calibrate the system model, as well as to compute the orifice area of the leakage models, based on the available data measurements. A review of the transient-analysis solution methodologies is provided in Colombo *et al.* (2009). According to Savic *et al.* (2009), however, transient analysis is not widely accepted by the practitioners due to various reasons, such as cost and lack of expertise. Case studies on physical systems have been published in Covas & Ramos (2010) and Soares *et al.* (2011).

Some methodologies consider the use of computational intelligence techniques for leakage detection. A fuzzy min-max neural network approach was considered in Gabrys & Bargiela (1999) to analyze the patterns of state estimates for detecting and identifying leakages. In Caputo & Pelagagge (2002), artificial neural networks are utilized for isolating the leakage location and magnitude based on pressure, flow measurements, and boundary conditions. Mixture density networks were utilized in Mounce *et al.* (2003) for learning and forecasting DMA inflows; in Mounce *et al.* (2010), a fuzzy inference scheme was used to provide detection confidence intervals. In addition, the flow and pressure sampling interval has been investigated in Mounce *et al.* (2011). Genetic algorithms have been considered (Vítkovský *et al.* 2000; Wu & Sage 2006; Wu *et al.* 2010) to calibrate the network model and estimate leakage parameters; also, probabilistic and statistical methodologies have been proposed (Poulakis *et al.* 2003; Buchberger & Nadimpalli 2004; Puust *et al.* 2006) as well as Bayesian inference (Romano *et al.* 2010) for this purpose. Furthermore, a detection approach based on Kalman filtering of hydraulic measurements at a DMA level was considered in (Ye & Fenner 2011). Results from the Wireless Water Sentinel project in Singapore (WaterWiSe@SG) demonstrated transient pressure waves wavelet analysis for leakage detection (Whittle *et al.* 2010; Srirangarajan *et al.* 2010). A burst monitoring methodology suitable for water distribution networks given a set of measurement stations installed in the network is proposed for the abrupt leakage fault detection problem within a DMA, when measuring pressures and transient waves with high-frequency sampling sensors installed at various locations in a DMA (Misiunas *et al.* 2005), analyzed with the Cumulative Sum (CUSUM) algorithim, which detected changes in the mean of the difference between two consecutive filtered estimates. Details on the general CUSUM algorithm can be found in Basseville & Nikiforov (1993).

In practice, most water utilities may not have high-frequency sampling sensors installed in their distribution networks for measuring transient pressure waves, to use the transient leakage detection algorithms. Furthermore, most water utilities may not have well-calibrated models of their water distribution system, as well as representative consumer demand models, to solve optimization problems for leakage detection and estimation. On the other hand, a number of water utilities measure flows and pressures at DMA inlets and outlets, and these measurements are used for detecting and estimating leakages, e.g. by using alarm thresholds, or simply by manual operator observation.

In this work, the problem of leakage detection is formulated in an automated fault diagnosis framework, and a solution methodology for leakage detection in a class of distribution systems is proposed. The goal is to design algorithms which analyze the discrete DMA inflow signal and determine as early as possible whether a leakage above some magnitude has occurred in the system. The novelty of this work is in the use of adaptive approximation-based methodologies for learning a normalized DMA outflow signal by updating at each time step the coefficients of a Fourier series, to capture the structured changes in the demand dynamics, and use the offset term of the Fourier series as the means for leakage detection. The proposed formulation is shown to perform better in detecting small-magnitude leakages than baseline approaches considered in this work, with fewer false negatives.

This paper is organized as follows: a mathematical model of the DMA inflows and the leakages is presented, along with key concepts and definitions; next, a solution methodology based on adaptive approximation is presented and compared to a leakage fault detection algorithm using night-flow analysis. Finally, randomized leakage faults are simulated on real data acquired from a water distribution system in Limassol, Cyprus, and the performance of the algorithms is evaluated; the paper concludes with some discussion and final remarks.

MATHEMATICAL MODEL

Consider a DMA which is connected to the main water supply network through a single pipe. To reduce the possibility of high pressure at the DMA entrance, a pressure-reduction valve controller is utilized to regulate the pressure at the point of entry within certain bounds. Following the pressure-reduction valve, the hydraulic characteristics of the water are monitored by on-line sensors, with sampling time Δt. At each discrete time instance $k \varepsilon 2\{0,1,\ldots\}$, the water flow and hydraulic head measurement time-series are recorded and stored using a SCADA system.

In general, according to statistical theory, time-series can be decomposed into linear and periodic signals, which is useful for forecasting (Box & Jenkins 1994; Cowpertwait & Metcalfe 2009). These techniques have also been applied to hydraulic signals (Yamauchi & Huang 1977; Zhou *et al.* 2002), and in addition, various computational intelligence approaches have been utilized for forecasting daily demands (Li & Huicheng 2010; Tiwari & Chatterjee 2011).

In the present work the DMA inflow signal is decomposed into two signal components: (a) long-term trends and yearly seasonal changes, and (b) weekly periodic changes. The long-term trend is typically a monotonic function which describes the increase (or decrease in some cases) of water consumption, e.g. due to population increase (or decrease). The seasonal changes describe the variation in water consumption as a result of seasonality within a year. The weekly periodic component describes the fluctuation of the signal throughout one week, which depends on the various social and economic characteristics of the consumers. In some studies, the weekly periodic component is simplified to a daily, 24-hour periodic component. In addition, high frequency variations are considered, which may be due to unpredictable consumer demands, transients, repairs and other network activities. In this work, the long-term trends are not considered in the problem formulation; however if it is known that trends appears in the signal, this can be incorporated into the proposed methodology.

Specifically, DMA inflows are modeled in this work using a multiplicative formulation with respect to the signal components (Yamauchi & Huang 1977). In accordance with the multiplicative model, the uncertainty and weekly periodic variation amplitudes increase (or decrease) as the yearly seasonal component increases (or decreases). In addition, the uncertainty variation amplitude increases (or decreases) as the weekly periodic component increases (or decreases); for instance, during the day when the demand is high, the uncertainty variance is high, whereas during the night when the demand is low, the uncertainty variance is low.

For the time step k, let $q(k)$ be the DMA inflow, $r(k)$ be the function which describes the yearly seasonal component of the flow, $s(k)$ be the function which describes weekly periodic water demand signal, and $n(k)$ be the multiplicative uncertainty component with zero mean normal distribution. Therefore, the multiplicative model of the flow signal $q(k)$ is described by

$$q(k) = r(k)s(k)(1 + n(k)) \tag{1}$$

Leakage faults increase the DMA inflow; mathematical models which describe the leakage flow with respect to the pressure at the leakage location have been proposed in various empirical studies (Germanopoulos 1985; Walski *et al.* 2004; Giustolisi *et al.* 2008). For example, the leakage flow $\varphi(k)$ can be modeled mathematically as a function of a positive discharge coefficient, the pressure at the leakage location and an exponent term which depends on the type of leaks (e.g. small or large leaks in plastic or metal pipes) (Lambert 2001). A leakage fault can start at some unknown time T; let $\beta(k-T)$ be the unknown time profile of the leakage fault $\varphi(k)$; for an abrupt fault, this may correspond to a step function which is $\beta(k-T) = 0$ for $k < T$ and $\beta(k-T) = 1$ for $k \geq T$. Therefore, the mathematical model of DMA inflow with leakage is given by

$$q(k) = r(k)s(k)(1 + n(k)) + \beta(k - T)\varphi(k) \tag{2}$$

An important consideration to take into account is the fact that the uncertainty term $r(k)s(k)n(k)$ in the flow model (2) may have a significantly larger variance than the variance corresponding to the leakage fault. On the other hand, the uncertainty term $r(k)s(k)n(k)$ has a zero average value, while at the same time, the leakage fault has a positive (non-zero) average value. This is a key characteristic which is exploited in the leakage fault detection algorithms presented in this work.

Furthermore, it is important to note that the yearly seasonal component is significant for the leakage detection methodology proposed in this work, as it is used to normalize the DMA inflow signal, to compute a signal with constant average value, so that it is possible to detect small leakage faults. In the opposite case, if the signal is not normalized, seasonal variations of large magnitude may cause a large number of false positives or false negatives in some leakage detection algorithms.

The yearly seasonal function $r(k)$ is in general unknown; an approximation of this signal can be computed based on the historical measurements. It is important to note that some known or unknown leakage faults may exist in the historical data; in this work it is considered that the hydraulic data may have some unknown leakage faults, but for the purposes of the detection algorithms, these are considered as normal consumption; as a result, the detection algorithm will be able to detect only new leakages.

Let K_ξ be the set of historical time instances considered for computing the seasonal component. Let $\rho(k;\xi_r)$ be the estimate of the unknown $r(k)$, which is computed off-line, where $\rho(\cdot)$ is a selected structure for the estimator function and ξ_r is a set of approximation parameters. In the case of yearly seasonality, $\rho(\cdot)$ can be formulated as a Fourier series with a small number of terms and fundamental period of a year, to capture the seasonal consumption changes. In this case, the parameter ξ_r is the estimated coefficient vector of the polynomial of the Fourier series, computed given sufficient recorded hydraulic measurements, by solving the least-squares optimization problem

$$\xi_r = \underset{\xi}{\operatorname{argmin}} \sum_{k \in K_\xi} (q(k) - \rho(k;\xi))^2 \tag{3}$$

The approximation signal $\rho(k;\xi_r)$, which is computed off-line, is used in the following sections to normalize the flow signals used in the leakage fault detection algorithms. Based on Equation (2), and assuming that $r(k) = \rho(k;\xi_r)$, the normalized DMA inflow $q_r(k) = q(k)/\rho(k;\xi_r)$ is given by

$$q_r(k) = s(k)(1 + n(k)) + \beta(k - T)\varphi(k)/\rho(k;\xi_r) \tag{4}$$

In case of incomplete yearly hydraulic data, a less accurate approximation of the seasonal model could be constructed based on the available data, which corresponds to higher uncertainty and may reduce the leakage detection sensitivity; the seasonal signal can be updated periodically, to improve the approximation accuracy.

LEAKAGE FAULT DETECTION METHODOLOGY

The proposed fault detection algorithm is comprised of two parts: the adaptive DMA inflow approximation and the leakage fault detection logic component.

Adaptive DMA inflow approximation method

In this subsection, the adaptive inflow approximation methodology is formulated based on the concept of learning adaptively the periodic signal corresponding to the normalized flow and utilizing that information in the detection logic component. The unknown weekly periodic signal $s(k)$ is approximated by $s(k) = \theta^*(k)^T\zeta(k)$, where $\theta^*(k)$ is an unknown parameter, and $\zeta(k)$ represents the approximation structure (regressor). In this work, $\zeta(k)$ is a vector of Fourier functions, while $\theta^*(k)$ are the ideal Fourier coefficients associated with the weekly periodic signal $s(k)$. In general, $\theta^*(k)$ is time-varying, representing slowly changing consumption characteristics due to variations in consumers behaviour, as well as due to various external events such as the daylight savings time changes, festivities, weather conditions, etc. Let $\theta_r(k)$ be the estimate of $\theta^*(k)$, which is updated at each time step k based on the available hydraulic measurements. Since the weekly periodic signal $s(k) = \theta^*(k)^T\zeta(k)$ is time-varying, it is useful to compute and update in time the parameter vector $\theta_r(k)$, so that $s_r(k) = \theta_r(k)^T\zeta(k)$ adaptively approximates the unknown $s(k)$. As approximation structure, a linearly parameterized

Fourier series with N_z terms is considered, where $\theta_r(k)$ is the parameter estimation vector of size $2N_z+1$; for Δt the sampling time step (measured in minutes) and $\omega_s = 2\pi/T_S$ where $T_S = (60 \cdot 24 \cdot 7)/\Delta t = 10080/\Delta t$ is the number of samples within a weekly period, the regressor vector $\zeta(k)$ is given by

$$\zeta(k) = [1, \cos(k\omega_s), \ldots, \cos(N_z k\omega_s), \sin(k\omega_s), \ldots, \sin(N_z k\omega_s)]^T \tag{5}$$

The normalized flow estimation error $e_r(k)$ is given by $e_r(k) = q_r(k) – s_r(k)$. The error $e_r(k)$ is useful in computing the parameter vector estimate for the next time instance, using the normalized projection algorithm, such that

$$\theta_r(k+1) = \theta_r(k) + G\frac{\zeta(k)}{\alpha + \zeta(k)^T\zeta(k)}e_r(k) \tag{6}$$

where $\theta_r(0) = \theta_r{}^0$ are the initial conditions, α is a positive nonzero constant selected by the designer to exclude the zero denominator case, and G is a learning gain diagonal matrix used to change the update step, whose bounds are analytically computed as $0 < G_i < 2$ for the i-th diagonal element of G, to guarantee learning stability (Astrom & Wittenmark 1995). If the learning gain is small, then the adaptation will be slow; however, if it is large, then the adaptation will be faster, but may over-react to random events (Farrell & Polycarpou 2006).

The initial conditions of the estimated parameter vector, $\theta_r{}^0$, at the detection algorithm activation time, can be computed off-line, by taking into consideration a set of historical measurements. Let K_θ be the set of the most-recent time instances for which the system operated without any known significant leakage faults. By solving a least-squares optimization problem, the initial condition vector estimation is computed, such that

$$\theta_r^0 = \underset{\theta}{\operatorname{argmin}} \sum_{k \in K_\theta} \left(q_r(k) - \theta^T\zeta(k)\right)^2, \tag{7}$$

where $q_r(k) = q(k)/\rho(k;\xi_r)$ is the normalized flow.

The update law (6) can be used to adapt $\theta_r(k)$ with respect to the changes in the consumption dynamics. However, in addition to the consumption, the update law can learn the unknown leakage fault dynamics which begin at discrete time T. Specifically, the leakage fault acts as a positive offset to the flow signal and its magnitude may vary according to the pressure variations at the leakage location; the update law will change the elements of parameter vector estimator, to approximate the new flow characteristics and reduce the estimation error, which corresponds to learning the leakage faults.

Note that when certain known exogenous factors affect the periodic consumption, the update law can be suitably modified to capture such prior knowledge. For example, in some countries the daylight savings time change occurs in spring and ends in autumn, by advancing the clock time by 60 minutes; as a result, all the social and economic patterns are shifted. Unless this is taken directly into consideration, time change, for instance, may appear as a large error and may trigger the leakage fault detection algorithm.

Leakage fault detection logic

Detection of a leakage fault is performed by monitoring change in the mean value of the parameter $\theta_O(k)$, i.e. the first term (the offset DC term) of the Fourier series which represents the normalized weekly periodic DMA inflow at each time step. A suitable sequential analysis algorithm for the detection of changes in the mean value of a signal is the CUSUM (Basseville 1988; Blanke *et al.* 2003); the assumption for using CUSUM as detection logic in this work is that a monitored signal $\theta_O(k)$ has a constant average value of 1. Let $C(k)$ be the CUSUM feature signal for the k-th time instance, where $C(0) = 0$; for $k > 1$ and since leakage faults increase the average value of the monitored signal, the one-sided CUSUM algorithm is thus given by

$$C(k) = \max\{0, C(k-1) + \theta_0(k) - 1 - \eta/\rho(k;\xi_r)\} \tag{8}$$

where $\eta > 0$ is the average outflow a small leakage, as specified by the operator. In general, the smaller the value of η, the larger the maximum values $C(k)$ takes; similarly, for larger values of η, the smaller the maximum values $C(k)$ takes. A detection alarm is triggered at the time instance k_d at which the metric $C(k)$ is greater than the threshold h_s, such that

$$k_d = \min\{k | C(k) \geq h_s\} \tag{9}$$

The selection of h_s is important, since the smaller the threshold, the more sensitive the detection algorithm is; however, it may also trigger false alarms. Note that the detection threshold h_s does not have a direct physical meaning, but it depends on the system's DMA inflow variance. In practice, h_s can be selected by applying the adaptive approximation and the CUSUM algorithm on a set of recent historical DMA inflow data, which are assumed to describe the normal operation, for a specific small leakage magnitude η, and calculate a threshold h_s which is greater than the maximum value of the feature $C(k)$ within that period; a conservative selection of h_s could reduce the number of false alarms, but could also miss some leakages of small magnitude.

Finally, part of leakage fault diagnosis methodology is to identify the magnitude of the leakage; the large uncertainties in the measurements, however, impose difficulties in the task. To compute an estimate of the average leakage outflow φ_0 considering the finite set of K_φ of time samples, corresponding to some period after the leakage is detected and for which the inequality $C(k) \geq h_s$ is true, the following algorithm is utilized

$$\varphi_0 = \frac{1}{|\mathrm{K}_\varphi|} \sum_{k \in \mathrm{K}_\varphi} [(\theta_0(k) - 1)\rho(k; \xi_r)], \tag{10}$$

which corresponds to the average flow increase due to a leakage, based on the DC term $\theta_0(k)$ of the Fourier series and the seasonal signal approximation $\rho(k,\xi_r)$.

Baseline night-flow analysis method

In practice, utility operators use night-flow monitoring to determine the presence of leakage faults. This is usually done by comparing the average of the minimum night flow with those of the previous days; however, this decision may be subjective to the utility operator assumptions. In addition, at some periods night flows may appear to be increasing (or decreasing), when this could be justified as a result of trends or seasonality. In the following, a baseline leakage fault detection algorithm is formulated based on the average night flow, which takes into consideration the DMA water inflow during low consumption hours, normalized with respect to the seasonal signal. This will also serve as a comparison to the adaptive approximation algorithm proposed in the previous section. The intuition behind using the night-flows is that since the flows during the night have smaller variations than during the day, and since the leakage losses will be larger because of the higher pressures in the system, it will be easier to detect leakage faults.

Let $w(l)$ be the average night-flow measured for the l-th period, which corresponds to 24 hours. Let Δt be the sampling time (in minutes) for measuring the flow. Let t_a and t_b be the discrete times at which the night-flow begins and ends. Considering that the first discrete time $k = 0$ corresponds to midnight of period $l = 0$, and $T_H = (24 \cdot 60)/\Delta t$ to the number of samples in one day, let $N(l) = \{t_a + lT_H, t_a + 1 + lT_H, \ldots, t_b + lT_H\}$ be the set of discrete times corresponding to night-flows of the l-th day. In general, t_a and t_b are constant; however, in some cases (e.g. due to the daylight savings time change) they may need to shift. At the l-th day and after the night-flow period has finished, the normalized average night flow $w(l)$ is computed by

$$w(l) = \frac{1}{|N(l)|} \sum_{i \in N(l)} q_r(i), \tag{11}$$

where $q_r(\cdot)$ is the normalized flow with respect to the estimated yearly seasonal signal $\rho(k;\xi_r)$. The feature signal used for detection is $\delta(l)$, the difference of the average night flow $w(l)$ with the minimum average night flow of the previous $M_\delta \geq 1$ days, such that

$$\delta(l) = w(l) - \min\{w(l - M_\delta), \ldots, w(l-1)\} \tag{12}$$

Let l_d be the day a leakage fault is detected, such that $l_d = \mathrm{argmin}_l \ \delta(l) > h_w$, where h_w is a detection threshold which is selected off-line by using historical measurements, such that to minimize false positives and maximize true positives. This approach gives rise to certain tradeoffs: setting h_w too low may cause a large number of false positive leakage fault alarms, while setting it too high may cause the detection algorithm to miss some leakage faults. In addition, due to the large uncertainties in the flow measurements, e.g. due to festivities or other events, this threshold could be exceeded even when no leakage fault has occurred in the system.

As an improvement, the algorithm can be modified to consider a window of difference measurements, so that detection occurs when the difference $\delta(\cdot)$ computed for each day within that period is greater than a certain threshold h_w, for at least M days. If a leakage fault has not been detected after M_δ days from its day of occurrence, the algorithm will consider the previous average night flow measurements as normal, and may not be able to detect the leakage fault in the future; this is a significant drawback of the night-flow based, leakage fault detection methodology.

SIMULATION EXAMPLES

In this section, simulation results are presented by applying the leakage fault detection solution methodology on historical hydraulic data taken from a DMA in Limassol, Cyprus, corresponding to 'Period A' of 241 days between 1/11/2006 and 30/6/2007, as well as 'Period B' of 115 days between 7/9/2007 and 1/1/2008. Within the period 30/6/2007 and 7/9/2007, no data were recorded due to a sensor failure. The hydraulic data were collected with a five-minute sampling time $\Delta t = 5$; a daily period is comprised of $T_H = 288$ samples. A pressure reduction valve at the DMA entrance regulates pressure at around 4.7 bar.

Outlier detection and replacement

Real sensor data are sometimes affected by outliers, i.e. data points which are inconsistent with the expected values (Pearson 2002). In the dataset considered in this work, three outliers were identified and replaced. These were detected and replaced as follows. Consider the following real subset of flow time-series in the dataset considered: $\{7.56, 7.56, 7.2, 7.92, 46.8, 7.2\}\mathrm{m^3\,h^{-1}}$; the measurement $46.8\ \mathrm{m^3\,h^{-1}}$ corresponds to an outlier. To detect and replace these outliers, the use of an online robust filter is considered (Menold *et al.* 1999). Let $q^*(k)$ be the sensor measurement at time k; if the inequality $|\mathrm{median}\{q^*(k-N_\mu+1), \ldots, q^*(k)\} - q^*(k)| > \mu$ is true, where N_μ is a window length considered and μ an outlier detection threshold, then an outlier is detected and is replaced by $q(k) = \mathrm{median}\{q^*(k-N_\mu+1), \ldots, q^*(k)\}$; otherwise, $q(k) = q^*(k)$. In the case when a large abrupt leakage fault occurs, it may be detected as an outlier, initially; however, as new sensor measurements arrive, the filtered signal will converge to the measured signal with some time delay, depending on the window length N_μ. The threshold μ can be selected by the designer off-line after computing the absolute differences based on historical data (which may already contain background leakages); for example, the threshold may be large enough so that it does not cause False Positive alerts when applied on the historical data. For the real data subset considered, for $N_\mu = 5$ and $\mu = 20$, $q^*(k) = 46.8\ \mathrm{m^3\,h^{-1}}$ is detected as an outlier since $|\mathrm{median}\{7.56, 7.56, 7.2, 7.92, 46.8\} - 46.8| = |7.56 - 46.8| = 39.24 > \mu$ is true; thus the measurement is replaced by the median value of the previous five measurements, i.e. $q(k) = 7.56\ \mathrm{m^3\,h^{-1}}$.

Seasonal signal estimation

The historical data were considered to compute the seasonal signal estimation $\rho(k;\xi_r)$, by solving the least-squares optimization problem, Equation (3), for finding the coefficients of vector ξ_r of a Fourier series with two terms. Specifically, for K_ξ the set of discrete times corresponding to the data available for the period 1/11/2006 and 1/11/2007, and assuming no monotonic trend to the signal, the seasonal signal is given by solving (3), as the yearly periodic Fourier series such that $\rho(k;\xi_r) = 32.66 - 0.21\cos(\omega k) - 0.04\cos(2\omega k) - 2.96\sin(\omega k) + 0.34\sin(2\omega k)$, where, for $T_Y = 365 \cdot T_H = 105120$ is the number of samples within a year and $\omega = 2\pi/T_Y = 5.9772 \cdot 10^{-5}$. This function corresponds to a sinusoidal signal; during its first period, this average DMA inflow function has a minimum at $\rho(28937) = 29.70\ \mathrm{m^3\,h^{-1}}$ (9/2/2007), and a maximum at $\rho(74470) = 35.79\ \mathrm{m^3\,h^{-1}}$ (17/7/2007).

Nominal flow signal decomposition example

To illustrate the DMA inflow signal decomposition, historical hydraulic flow data were acquired from a DMA in Limassol, Cyprus, for the period between 8/3/2007 and 15/3/2007; the flow measurements corresponding to this week have the smallest Root Mean Square Error (RMSE) with respect to the average weekly flow signal, which is computed from 'Period A' of the dataset. For $\Delta t = 5$ minutes, this weekly period corresponds to 2016 samples. The original signal is depicted in Figure 1(a) and the estimated seasonal signal for that week is depicted in Figure 1(b).

The weekly periodic function does not change significantly within the examined period; therefore, the coefficient vector θ_r of the weekly periodic approximation function $\theta^T\zeta(k)$ is computed by solving the following optimization problem

$$\theta_r = \underset{\theta}{\operatorname{argmin}} \sum_{k\in K_\theta} \left(\frac{q(k)}{\rho(k;\xi_r)} - \theta^T\zeta(k)\right)^2, \tag{13}$$

for which K_θ is the set of discrete times corresponding to that week $K_\theta = \{36576, \ldots 38592\}$. The approximation signal which corresponds to a Fourier series with 100 terms is depicted in Figure 1(c). Finally, the uncertainty estimation is depicted in Figure 1(d); this is calculated by

$$n(k) = q(k)/\left(\rho(k;\xi)\theta_r^T\zeta(k)\right) - 1 \tag{14}$$

The estimated uncertainty follows a normal distribution (as verified by the Lilliefors normality test), with 0.045 standard deviation.

Estimator performance

To evaluate the effectiveness of the estimation algorithm, the RMSE of the estimated and the actual signal for part of the historical data is calculated for various numbers of Fourier series terms $N_z \in \{10, 50, 100, 200\}$, as well as for various learning gains $G_i \in \{0.01, 0.1, 1, 1.9\}$. Let $T_S = 2016$ be the number of samples within a weekly period, $\omega_s = 2\pi/T_S = 0.0031$, and $\alpha = 0.01$ for the update law is considered. In this example the first 14 days of 'Period A' dataset are used to compute with Equation (7) the initial conditions of the parameter vector θ_r; then the update law (6) is activated for the next 55 days from the same dataset; afterwards, the update law stops and with these parameters the Fourier series signal is compared to the historical measurements; in particular considering the following 14 days of the dataset, the RMSE is computed and the results are given in Table 1. From the parameter examined, the Fourier series with $N_z = 100$ terms and a learning gain $G_i = 0.01$ achieves the smallest RMSE. These results demonstrate the performance of the estimation algorithm, and in addition, provide intuition on how to select its parameters when historical data are available. As the results demonstrate, small learning gains can generalize better than larger learning gains, but adapt more slowly. Furthermore, increasing the number of Fourier terms generally improves learning performance; however, after a certain point, learning cannot be improved, e.g. due to noise, as it is the case for $G_i = 0.01$ and $N_z = 200$. On the other hand, large learning gains may cause the algorithm to over-react to random errors, as it appears to be the case for $G_i = 1.9$. In that case, the estimated Fourier signal

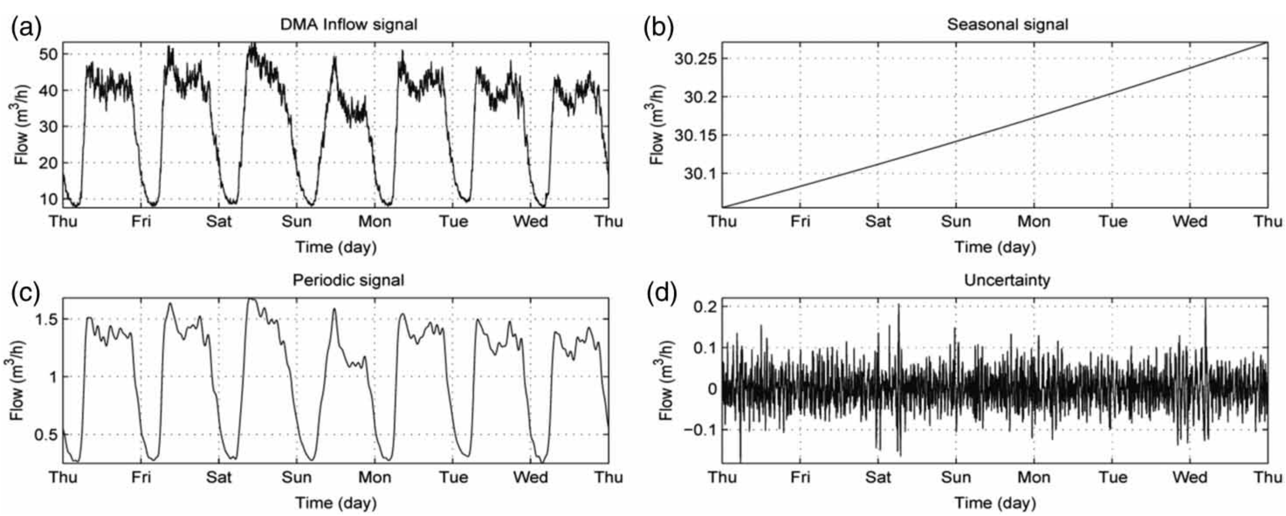

Figure 1 | Decomposition of the flow signal for the time span between Thursday, 8/3/2007 and Thursday, 15/3/2007.

Table 1 | RMSE of the estimated and the actual flow measurements for 14 days, for $N_z \in \{10, 50, 100, 200\}$ Fourier series terms and $G_i \in \{0.01, 0.1, 1, 1.9\}$ learning law gains

G_i	N_z 10	50	100	200
0.01	0.212	0.093	0.087	0.089
0.1	0.272	0.132	0.119	0.108
1.0	0.348	0.318	0.277	0.228
1.9	0.356	0.374	0.381	0.393

cannot approximate well the actual normalized weekly periodic dynamics, even though the number of Fourier terms increases.

Leakage detection using adaptive inflow approximation

In this example, the adaptive inflow approximation leakage detection methodology is evaluated. The first 14 days of 'Period A' dataset are considered to compute the initial conditions of the parameter vector θ_r; then, the update law is activated for the next 227 days from the dataset, with $N_z = 100$ Fourier series terms learning gain $G_i = 0.1$; these terms are not the best parameters with respect to Table 1, but they have been selected to facilitate both faster adaptation and efficient learning of the unknown, normalized periodic dynamics. Furthermore, the magnitude of change which corresponds to the average outflow of a small leakage is selected as $\eta = 0.5\ \mathrm{m}^3\,\mathrm{h}^{-1}$; therefore, the CUSUM feature signal is given by $C(k) = \max\{0, C(k–1) + \theta_0(k)–1–0.5/\rho(k;\xi)\}$.

As an indication of the detection threshold magnitude, the feature signal is computed based on the historical values. The reasoning behind the detection threshold selection, is not to trigger any False Positives (false alarms) with respect to recent historical data (e.g. of the last few months) which do not contain known hydraulic faults. For the last 100 days of 'Period A' dataset, for $\eta = 0.5\ \mathrm{m}^3\,\mathrm{h}^{-1}$, the maximum value of $C(k)$ is 35.7 for learning law gain $G_i = 0.1$. This maximum value can serve as a lower bound for the selection of the leakage fault detection threshold h_s. As a special case, different detection thresholds could be considered during festivity periods, which may correspond to known increased water demands.

In the following paragraphs the efficiency of the proposed algorithm for detecting randomized leakages simulated during 'Period B' is demonstrated. In this case study, four different thresholds are examined, for $h_s \in \{20, 30, 40, 60\}$. For each test, 1000 random leakages are simulated and added to the real flow data, as in Equation (2); in relation to the fault model, $\beta(k–T)$ corresponds to the unit step function such that $\beta(k–T) = 0$ for $k < T$ and $\beta(k–T) = 1$ otherwise, where T is a random burst time from a period between 21/9/2007 and 18/12/2007, with random magnitude $\varphi(k) \in [0.5, 3]\ \mathrm{m}^3\,\mathrm{h}^{-1}$. It is important to note that incipient fault evolution models could be considered as well, as real bursts are not always abrupt. In Table 2, the results based on the simulated fault scenarios are shown. These results confirm that the smaller the detection threshold, the False Positive instances (i.e. false alarms) increase, whereas the larger the detection threshold, the False Negative instances (i.e. misses) increase. The True Positive instances receive the best values for $h_s = 30$, and are reduced as the threshold increases.

The average detection delays are reported in Table 2, with respect to the True Positive alerts; the higher the detection threshold, the higher the average detection delay. Considering the case which corresponds to the detection threshold $h_s = 30$, the distribution of the detection delays for the 1000 simulated leakage scenarios is skewed to the right (skeweness metric is 2.4); specifically, 49.3, 78.0 and 85.4% of the simulated leakage scenarios were detected within the 1, 2, and 3 weeks, respectively.

After the leakage was detected, the leakage magnitude estimation algorithm (10) is activated, by considering the new hydraulic measurements. Based in the simulation examples for detection threshold $h_s = 30$, and considering hydraulic data for which $C(k) \geq 30$ within 1 week after the leakage fault detection, the average leakage magnitude estimation error (i.e. $\varphi(k)–\varphi_0$) is 0.21 with variance is 0.26; the

Table 2 | Performance of the adaptive learning leakage detection methodology for 1000 simulated random leakage faults with magnitudes [0.5,3] $\mathrm{m}^3\,\mathrm{h}^{-1}$ added to real inflow data, with start time between 21/9/2007 and 18/12/2007

h_s	True Positive (%)	False Positive (%)	False Negative (%)	Average Det. Delay (days)
20	25.0	75.0	0	6.0
30	94.5	0	5.5	9.8
40	90.6	0	9.4	11.2
60	85.3	0	14.7	12.0

estimation error values range between [–0.75, 1.41]. Moreover, analysis of the False Negative events for $h_s = 30$, shows that the events which were missed have an average leakage magnitude of $0.63\,\mathrm{m^3\,h^{-1}}$, and largest undetected leakage magnitude of $0.86\,\mathrm{m^3\,h^{-1}}$, which correspond to relatively small leaks.

Leakage detection using night-flow analysis

For night-flow leakage detection, a first step is to determine the time period of some minimum duration with corresponds to the smallest variance. In this work, a minimum of 3-hour duration was considered; based on historical data from the first 5 months of 'Period A', it was found that the smallest average variance was between 2 and 5 a.m. (standard time). Therefore, $t_a = 24$ and $t_b = 60$ were selected as the discrete times corresponding to the night-flow period (for the daylight savings time period this corresponds to $t_a = 12$ and $t_b = 48$); thus, the set of night-flow discrete times corresponding to the l-th day, during standard time, is given by $N(l) = \{24 + 288l, 25 + 288l, \ldots, 60 + 288l\}$, for $l = \{0,1,2,\ldots\}$.

Figure 2 depicts the average night-flows computed using the normalized flows for 'Period A' dataset. Even though the signal is normalized over the seasonal signal, the uncertainty and the extreme values in the signal pose difficulties in detecting leakages by using simple observation of the time-series from a human operator. Figure 2(b) illustrates this issue, in which a simulated leakage fault with average leakage outflow $0.5\,\mathrm{m^3\,h^{-1}}$ is added to the real normalized flow at day $l_0 = 115$. If a leakage fault occurred in the network at the l_0-th day, it would correspond to an increase in the average night-flow measurements; in this example, a leakage fault with outflow $0.5\,\mathrm{m^3\,h^{-1}}$ would correspond to an average increase of 0.015 in the signal $w(l)$ for $l \geq l_0$, which is relatively small with respect to the uncertainty magnitude.

To demonstrate the difference-based leakage fault detection algorithm, the random set of leakage faults constructed in the previous example is utilized and applied to the same period between 21/9/2007 and 18/12/2007. For Equation (12), a 1-week window of average night flows is considered, i.e. $M_\delta = 7$. The results in Table 3 demonstrate the performance of the night-flow analysis leakage detection algorithm, with various combinations of detection thresholds h_w; leakages are detected when the threshold is exceeded for at least M days. As an indication of the detection threshold magnitude, the feature signal $\delta(l)$ is computed based on the historical values corresponding to the last 100 days of the first part of the dataset; the maximum value of $\delta(l)$ is 0.107; in general this is an indication of the threshold magnitude. However, the results in Table 2 demonstrate some challenges in selecting the best parameters.

Considering the case when $M = 0$, i.e. the detection occurs as soon as the feature signal is greater than a threshold; for $h_w = 0.10$, there is a large number of True Positive alarms, however, there is also a large number of False Positive alarms, which is undesirable. For a larger

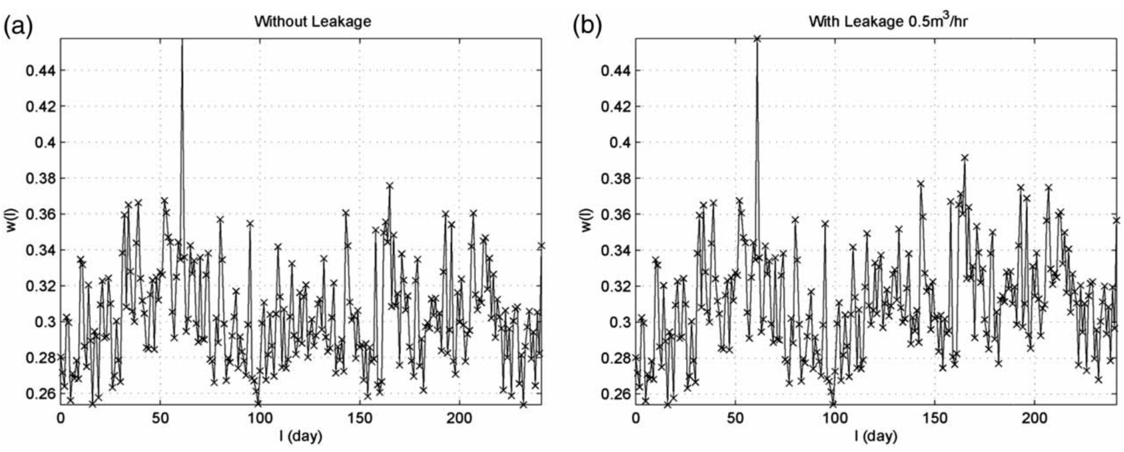

Figure 2 | Average night-flow measurements for the first 241 days in the dataset, (a) without leakage faults and (b) with a leakage fault of $0.5\,\mathrm{m^3\,h^{-1}}$ occurring after the 115th day. The large average night-flow value $w(61) = 0.46$ corresponds to Christmas day festivities.

Table 3 | Performance of the night-flow analysis leakage detection methodology for 1000 simulated random leakage faults with magnitudes [0.5, 3] $m^3 h^{-1}$ added to real inflow data, with start time between 21/9/2007 and 18/12/2007

h_w	M	True Positive (%)	False Positive (%)	False Negative (%)	Average Det. Delay (days)
0.10	0	79.5	20.5	0	14.7
0.11	0	53.3	0	46.7	1.9
0.07	1	73.0	0	27.0	2.0
0.06	2	68.7	0	31.3	2.5

threshold, $h_w = 0.11$, 53.3% of the simulated leakage faults are detected within on average 1.9 days, and 46.7% of the faults are undetected (False Negatives). In general, larger thresholds would reduce the True Positives and increase the False Negatives. Considering the proposed modification, for $h_w = 0.07$, with $M = 1$, the detection succeeds in detecting more than 70% of all the simulated leakages within on average 2 days, and the other faults are undetected. It is important to note that the 75% of the undetected leakages had magnitudes less that $1.5\,m^3\,h^{-1}$, whereas 75% of the detected leakages had magnitudes greater than $1.5\,m^3\,h^{-1}$. Similar results correspond in the case of $h_w = 0.06$ with $M = 2$.

It should be noted that the average detection delays reported in Table 3 are relatively shorter than those reported in Table 2; for the night-flow method, the average detection delays are calculated based mostly on large-magnitude leakages which have in general shorter detection delays (73% of all scenarios examined), whereas for the adaptive learning method, the average detection delays are calculated based on leakages of all magnitudes (94.5% of all scenarios examined), and reported detection delays appear longer due to the small leakages which may require weeks to be detected. In general, detection delay for the adaptive learning approach can be reduced by lowering the detection threshold (with the possibility of some False Positive alarms), or by increasing the adaptive learning gain parameter (which may increase sensitivity to noise).

CONCLUDING REMARKS

Fault tolerance in water distribution system is receiving increasingly more attention in the context of leakage detection, so that water utilities can be more effective in reducing water losses. Previous research has examined the problem by monitoring transient waves using high-frequency pressure sensors, or by solving an optimization problem when part of the distribution network model was known; however, most water utilities do not have high-frequency pressure sensors or system models available. A common practice is for water utility operators to routinely observe the inflow measurements in DMAs during night hours, sampled every few minutes, to detect leakage faults by using some standard leakage detection approaches, such as night-flow analysis, which are suitable for leakages of large magnitudes; however, small leakages or incipient faults may remain unnoticeable for some time, and finally be considered as normal consumption.

In this work, the problem of small-leakage detection in a DMA is addressed considering the availability of water inflow measurements, sampled every few minutes. The mathematical model considered which describes the DMA inflow with leakages is comprised of multiple components: a yearly seasonal, a weekly periodic and an uncertainty component, along with the leakage fault signal. An optimization problem is formulated for computing the seasonal signal over a set of historical flow measurements. To solve the leakage detection problem, an adaptive demand flow approximation methodology is proposed, in which an update law is used in learning the parameters of a Fourier series which estimate the changing weekly periodic consumption dynamics, as well as the leakages. The CUSUM metric is utilized to construct a feature signal based on the offset (DC) term of the Fourier series, and detection occurs when this feature signal exceeds a threshold selected offline. In addition to the proposed methodology, the use of night-flow analysis for leakage detection is discussed for comparison.

Simulation results for measuring the performance of the two methodologies for detecting randomized leakage faults are presented. The results demonstrate the ability of the proposed adaptive approximation methodology to detect leakages of smaller magnitudes, in comparison to the night-flow method examined. It is important to note that the drawback of the night-flow leakage detection methodology of considering measurements with leakage as normal if a leakage is not detected within a time window, does not affect the adaptive learning leakage detection methodology.

The selection of the detection threshold depends on historical data; however, if these data are affected by background leakages, these would be considered as normal and a high detection threshold would be selected; as a result, the proposed algorithm would be able to detect only new leakages. Moreover, in the case of a new DMA for which few historical hydraulic measurements are available, it is possible to initially make some conservative assumptions regarding the approximated yearly seasonal signal as well as the detection threshold; both of which could be updated periodically, utilizing the latest recorded data, to improve estimation accuracy.

In general, both methods are able to detect leakages with some time delay, which can be longer than some of the other methods discussed in the literature, e.g based on transient analysis (which require specialized sensors) or other algorithms which may be suitable for larger leakage faults or cause large number of False Positive alarms. Future work will examine the use of a sample of various leakage detection algorithms, to reduce False Positives alarms and to reduce detection time.

ACKNOWLEDGEMENTS

This research work has been funded by the European Commission 7th Framework Program, under grant INSFO-ICT- 270428 (iSense), and by the Cyprus Research Promotion Foundation's Framework Programme for Research, Technological Development and Innovation, co-funded by the Republic of Cyprus and the European Regional Development Fund. The authors would like to acknowledge the help of the Water Board of Limassol and Bambos Charalambous in particular for providing real data and discussions.

REFERENCES

Alegre, H., Baptista Jr., J. E. C. , Cubillo, F., Duarte, P., Hirner, W., Merkel, W. & Parena, R. 2006 *Performance Indicators for Water Supply Services*. International Water Association, London, UK.

Astrom, K. & Wittenmark, B. 1995 *Adaptive Control*. Addison-Wesley, Reading, MA, USA.

Basseville, M. 1988 Detecting changes in signals and systems – a survey. *Automatica* **24**, 309–326.

Basseville, M. & Nikiforov, I. 1993 *Detection of Abrupt Changes: Theory and Application*. Prentice Hall, Englewood Cliffs, NJ, USA.

Blanke, M., Kinnaert, M., Lunze, J., Staroswiecki, M. & Schröder, J. 2003 *Diagnosis and Fault-Tolerant Control*. Springer, Berlin, Germany; New York, USA.

Box, G. E. P. & Jenkins, G. M. 1994 *Time Series Analysis: Forecasting and Control*. 3rd edn, Prentice Hall, Englewood Cliffs, NJ, USA.

Buchberger, S. G. & Nadimpalli, G. 2004 Leak estimation in water distribution systems by statistical analysis of flow readings. *ASCE Journal of Water Resources Planning and Management* **130**, 321–329.

Caputo, A. & Pelagagge, P. 2002 An inverse approach for piping networks monitoring. *Journal of Loss Prevention in the Process Industries* **15**, 497–505.

Colombo, A. F., Lee, P. & Karney, B. W. 2009 A selective literature review of transient-based leak detection methods. *Journal of Hydro-Environment Research* **2**, 212–227.

Covas, D. & Ramos, H. 2010 Case studies of leak detection and location in water pipe systems by inverse transient analysis. *ASCE Journal of Water Resources Planning and Management* **136**, 248–257.

Cowpertwait, P. S. P. & Metcalfe, A. V. 2009 *Introductory Time Series with R*. Springer, New York, USA.

Farley, M. 2001 *Leakage Management and Control: A Best Practice Training Manual*. World Health Organization, Geneva, Switzerland.

Farrell, J. A. & Polycarpou, M. M. 2006 *Adaptive Approximation Based Control: Unifying Neural, Fuzzy and Traditional Adaptive Approximation Approaches*. Wiley-Interscience, Hoboken, NJ, USA.

Gabrys, B. & Bargiela, A. 1999 Neural networks based decision support in presence of un-certainties. *ASCE Journal of Water Resources Planning and Management* **125**, 272–280.

Germanopoulos, G. 1985 A technical note on the inclusion of pressure dependent demand and leakage terms in water supply network models. *Civil Engineering Systems* **2**, 171–179.

Giustolisi, O., Savic, D. & Kapelan, Z. 2008 Pressure-driven demand and leakage simulation for water distribution networks. *ASCE Journal of Hydraulic Engineering* **134**, 626–635.

Isermann, R. 1997 Supervision, fault-detection and fault-diagnosis methods – an introduction. *Control Engineering Practice* **5**, 639–652.

Lambert, A. 2001 What do we know about pressure: Leakage relationships in distribution systems? Proc. IWA System Approach to Leakage Control and Water Distribution Systems Management.

Lambert, A. & Morrison, J. A. E. 1996 Recent developments in application of 'bursts and background estimates' concepts for leakage management. *Water and Environment Journal* **10**, 100–104.

Li, W. & Huicheng, Z. 2010 Urban water demand forecasting based on HP filter and fuzzy neural network. *IWA Journal of Hydroinformatics* **12**, 172–184.

Liggett, J. A. & Chen, L.-C. 1994 Inverse transient analysis in pipe networks. *ASCE Journal of Hydraulic Engineering* **120**, 934–955.

Menold, P., Pearson, R. & Allgöwer, F. 1999 Online outlier detection and removal. In *Proc. Mediterranean Conference on Control and Automation*, Haifa, Israel.

Misiunas, D., Lambert, M., Simpson, A. & Olsson, G. 2005 Burst detection and location in water distribution networks. *Water Science and Technology: Water Supply* **5** (3–4), 71–80.

Morrison, J. 2004 Managing leakage by District Metered Areas: a practical approach. *Water 21* February, 44–46.

Mounce, S. R., Boxall, J. B. & Machell, J. 2010 Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *ASCE Journal of Water Resources Planning and Management* **136**, 309–318.

Mounce, S. R., Khan, A., Wood, A. S., Day, A. J., Widdop, P. D. & Machell, J. 2003 Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. *Information Fusion* **4**, 217–229.

Mounce, S. R., Mounce, R. B. & Boxall, J. 2011 Novelty detection for time series data analysis in water distribution systems using Support Vector Machines. *Journal of Hydroinformatics* **13**, 672–686.

Pearson, R. 2002 Outliers in process modeling and identification. *IEEE Transactions on Control Systems Technology* **10**, 55–63.

Poulakis, Z., Valougeorgis, D. & Papadimitriou, C. 2003 Leakage detection in water pipe net-works using a Bayesian probabilistic framework. *Probabilistic Engineering Mechanics* **18**, 315–327.

Pudar, R. S. & Liggett, J. A. 1992 Leaks in pipe networks. *ASCE Journal of Hydraulic Engineering* **118**, 1031–1046.

Puust, R., Kapelan, Z., Savic, D. & Koppel, T. 2006 Probabilistic leak detection in pipe networks using the SCEM-UA algorithm. Proc. ASCE Water Distribution Systems Analysis, Cincinnati, Ohio, USA.

Romano, M., Kapelan, Z. & Savic, D. A. 2010 Bayesian inference system for the fast and reliable detection of burst related leaks in real-time. Proc. Water Loss, Sao Paulo, Brazil.

Savic, D., Kapelan, Z. & Jonkergouw, P. 2009 Quo vadis water distribution model calibration? *Urban Water Journal* **6**, 3–22.

Soares, A.K., Covas, D. I. C. & Reis, L. F. R. 2011 Leak detection by inverse transient analysis in an experimental PVC pipe system. *Journal of Hydroinformatics* **13**, 154–166.

Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H. B. & Whittle, A. J. 2010 Water main burst event detection and localization. Proc. ASCE Water Distribution Systems Analysis, Tucson, Arizona, USA.

Thornton, J., Sturm, R. & Kunkel, G. 2008 *Water Loss Control*. 2nd edn, McGraw-Hill, New York, USA.

Tiwari, M. K. & Chatterjee, C. 2011 A new wavelet-bootstrap-ANN hybrid model for daily discharge forecasting. *Journal of Hydroinformatics* **13**, 500–519.

Vítkovský, J. P., Simpson, A. R. & Lambert, M. F. 2000 Leak detection and calibration using transients and genetic algorithms. *ASCE Journal of Water Resources Planning and Management* **126**, 262–265.

Walski, T., Bezts, W., Posluszny, E. T., Weir, M. & Whitman, B. 2004 Understanding the hydraulics of water distribution system leaks. Proc. ASCE World Water and Environmental Resources. Orlanto, Florida, USA.

WAA 1980 Leakage control policy and practice. Report No. 26. Water Authorities Association, London, UK.

Whittle, A. J., Girod, L., Preis, A., Allen, M., Lim, H. B., Iqbal, M., Srirangarajan, S., Fu, C., Wong, K. J. & Goldsmiths, D. 2010 WaterWiSe@SG: A testbed for continuous monitoring of the water distribution system in Singapore. Proc. ASCE Water Distribution Systems Analysis. Tucson, Arizona, USA.

Wu, Z. & Sage, P. 2006 Water loss detection via genetic algorithm optimization-based model calibration. Proc. ASCE Water Distribution Systems Analysis. Cincinnati, Ohio, USA.

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

Yamauchi, H. & Huang, W. 1977 Alternative models for estimating the time series components of water consumption data. *Journal of the American Water Resources Association* **13**, 599–610.

Ye, G. & Fenner, R. A. 2011 Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *ASCE Journal of Pipeline Systems Engineering and Practice* **2**, 14–22.

Zhang, X., Polycarpou, M. M. & Parisini, T. 2002 A robust detection and isolation scheme for abrupt and incipient faults in nonlinear systems. *IEEE Transactions on Automatic Control* **47**, 576–593.

Zhou, S. L., McMahon, T. A., Walton, A. & Lewis, J. 2002 Forecasting operational demand for an urban water supply zone. *Journal of Hydrology* **259**, 189–202.

First received 10 August 2011; accepted in revised form 5 March 2012. Available online 12 June 2012