Kalman Filtering of Hydraulic Measurements for Burst Detection in Water Distribution Systems

Guoliang Ye, Ph.D.[1]; and Richard Andrew Fenner, Ph.D.[2]

Abstract: Automatic burst and leak detection in water distribution systems plays an important role in water saving and management. This research develops a novel burst detection method of using adaptive Kalman filtering on hydraulic measurements of flow and pressure at district meter area (DMA) level. Adaptive Kalman filtering is used to model normal water usage (or alternatively water pressure), so the residual of the filter (e.g., the difference between the predicted flow and the measured flow) represents the amount of abnormal water usage relating to the bursts (or newly occurred leaks) in the downstream network. The results from a series of engineered tests which simulated flushing show that the size of the bursts and leaks strongly correlates with the residual of the filter. Finally, the method was applied to data from several real DMAs in the north of England, and the results show that the detected bursts correspond well to known historical operational information such as customer complaints' records and work management (repair) data. The results suggest that flow measurement data are more sensitive to a burst or leak than the pressure measurement data.

DOI: 10.1061/(ASCE)PS.1949-1204.0000070

CE Database subject headings: Water pipelines; Water distribution systems; Leakage; Pipe flow; Hydraulic pressure; Signal processing; Kalman filters; Measurement.

Author keywords: Water pipelines; Water distribution systems; Burst; Flow; Hydraulic pressure; Signal processing.

Introduction

Water is one of the most important resources in the world and the need to achieve water saving is a growing concern. Water from urban water distribution systems can be extremely valuable especially in districts with very limited water resources. Even in water-rich districts, the loss of treated water from the distribution network not only affects the economics of water supply but also damages the environment, increases the carbon footprint of the operator, and can result in service disruption to the customer. Water loss is mainly caused by bursts (high volume, short duration events) and leaks (low volume, long duration events) in water distribution systems, often on aging and deteriorating assets. In cities such as London, which possess many aging pipelines, it is claimed that 30% of water in the distribution system is lost due to bursts and leaks (Brothers 2001; Colombo et al. 2009; Mounce et al. 2002).

To prevent water loss from the water in a distribution system, a fast response burst detection protocol is required. Manual onsite surveying using acoustic techniques or other methods is a direct way, but this is time-consuming and costly. Automatic detection through the analysis of data from permanently installed sensors is a more cost-effective method which can provide a rapid response to the onset of a burst event. Hydraulic measurements are promising for burst detection (Buchberger and Nadimpalli 2004; Khan et al. 2005a,b; Mounce et al. 2002, 2003; Mounce and Machell 2006; Mounce et al. 2007, 2010; Obradovic 2000; Yurdusev et al. 2009).

Flow and pressure measurements are currently recorded by U.K. water companies. For example, one water utility company recently installed a number of hydraulic sensors at district inlet valves. These sensors log the flow and pressure every 15 min (the current industry standard) and send information to the control center via general packet radio service (GPRS) in the existing mobile network every 30 min for data saving and analysis. Thus a burst detection based on such flow and pressure measurements requires no system hardware upgrade and consequently no additional cost if the utility has already purchased and installed those sensors.

Previous research (Khan et al. 2005a,b; Mounce et al. 2002, 2003; Mounce and Machell 2006) has studied the method of using artificial neural networks (ANNs) on flow and pressure data. A neural network with a mixture density network (Bishop 1995) was used to predict the probability distribution function (PDF) of the hydraulic parameters. The PDF was then combined with a fuzzy inference approach to detect bursts and leaks in long time series data. However, such ANN approaches have to be trained with three typical months' normal data and updated every month.

This paper presents a method of using adaptive Kalman filtering on hydraulic measurements. Compared to the conventional method of using ANN, this approach has the advantages of computational efficiency, rapid detection rates, and no requirements for large quantities of training data. The paper is organized as follows. The Kalman filtering application on the hydraulic measurements is described together with an explanation of how the

[1]Postdoctoral Research Associate, Centre for Sustainable Development, Dept. of Engineering, Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, U.K. E-mail: gy223@cam.ac.uk

[2]Senior Lecturer, Centre for Sustainable Development, Dept. of Engineering, Cambridge Univ., Trumpington St., Cambridge CB2 1PZ, U.K. (corresponding author). E-mail: raf37@cam.ac.uk

Note. This manuscript was submitted on February 2, 2010; approved on August 24, 2010; published online on August 31, 2010. Discussion period open until July 1, 2011; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Pipeline Systems Engineering and Practice***, Vol. 2, No. 1, February 1, 2011. ©ASCE, ISSN 1949-1190/2011/2(1)/14/9/$25.00.

Kalman filtered flow and pressure measurements can be combined to detect bursts. The method is applied to an engineered test with simulated bursts and then data from 10 other district meter areas (DMAs) with real bursts and leaks. Finally the overall performance of the method is evaluated and conclusions drawn.

Method

Kalman Filtering

Kalman filtering provides a complete statistical characterization of the current state of knowledge of a dynamic system, including the influence of all past measurements, by propagating the entire probability distribution of the variables it is asked to measure. Grewal and Andrews (2001) described it as using a model of the estimation problem that distinguishes between phenomena (what one is able to observe), noumena (what is really going on), and the state of knowledge about the noumena that one can deduce from the phenomena. To the extent that probability distributions represent knowledge of the real world, the cumulative processing of that knowledge represents a learning process. It is called a filter because the approach is related to the separation of "signals" from "noise" and estimates the independent variables as inverted functions of dependent (measurable) variables, where the variables of interest are also allowed to be dynamic, with dynamics that are only partially predictable.

Thus a Kalman filter is an optimal recursive data processing algorithm for the linear stochastic dynamic system (Bar-Shalom and Fortmann 1988; Bar-Shalom et al. 2001; Kalman 1960; Maybeck 1979). It utilizes all the information from the measurements and the state dynamic of the system and takes accounts of their uncertainties to produce an optimal estimate. The uncertainties are modeled as Gaussian functions, so if the time series suffers from Gaussian white noise Kalman filtering is the best data processor to estimate the real signal.

This work uses Kalman filtering to estimate the normal hydraulic parameter (flow or pressure) in the water distribution system, and hence a burst can be detected through the residual of the filter which is the difference between the estimation and the measurement. The hydraulic parameter here is assumed to follow a constant change, which means that the current value is set equal to the last value. This assumption will be further discussed below. Thus the state of the hydraulic parameter, X, at the time step of k from the process model can be modeled by a linear difference state equation with a white Gaussian process noise Q expressed by

$$X(k) = X(k-1) + Q \tag{1}$$

The measurement of hydraulic parameter, Z, with a white Gaussian measurement noise, R, can be modeled by

$$Z(k) = X(k) + R \tag{2}$$

R takes account of uncertainties in the data measurements, which in this problem relate to the inaccuracies and expected errors in the flow meters.

Assuming that at step k, $X(k|k)$ is the state estimation (or the optimal output of the filter), $X(k|k-1)$ is the state prediction from the process model [Eq. (1)], and $q(k)$ and $r(k)$ are the covariances of Q and R, the Kalman filter runs the following five steps (Bar-Shalom and Fortmann 1988; Bar-Shalom et al. 2001) to estimate the normal hydraulic parameter:

1. Predict the flow or pressure

$$X(k|k-1) = X(k-1|k-1) \tag{3}$$

2. Predict the flow or pressure covariance

$$p(k|k-1) = p(k-1|k-1) + q(k) \tag{4}$$

3. Calculate the Kalman gain

$$g(k) = \frac{p(k|k-1)}{p(k|k-1) + r(k)} \tag{5}$$

4. Update the flow or pressure estimate

$$X(k|k) = X(k|k-1) + g(k)[Z(k) - X(k|k-1)] \tag{6}$$

5. Update the flow or pressure covariance

$$p(k|k) = [1 - g(k)]p(k|k-1) \tag{7}$$

Once $X(k|k)$ has been estimated in Eq. (6) and the covariance $p(k|k)$ has been updated in Eq. (7), the filter moves to next time step $(k+1)$. $X(k|k)$ will replace $X(k-1|k-1)$ in Eq. (3) and then the filter runs recursively.

The parameters, $q(k)$ and $r(k)$, can be automatically estimated using the filter's innovation sequence, which refers to an adaptive Kalman filter (Mehra 1970). The innovation sequence is defined as the difference between the measurement and the predicted state [see Eq. (6)]

$$s(k) = Z(k) - X(k|k-1) \tag{8}$$

Finally $q(k)$ and $r(k)$ are estimated and updated iteratively from the following equations (Mehra 1970):

$$q(k) = g^2(k)c(k) \tag{9}$$

$$r(k) = c(k) + p(k|k-1) \tag{10}$$

where $c(k)$ is the innovation covariance computed by the autocorrelation of the innovation sequence $s(k)$ inside a moving estimation window of size M (Price 2003; Ye 2008)

$$c(k) = \frac{1}{M}\sum_{i=k-M+1}^{k} s^2(i) \tag{11}$$

The advantage of the adaptive Kalman filter is that it can adopt a correct weight between the process model and the measurement, which can automatically tune to yield an unbiased estimation.

Burst Detection through Residuals

Kalman filtering tries to estimate normal hydraulic parameters, so excessive flows and pressure variations caused by the bursts or newly occurred leaks relate to the abnormal part of the hydraulic parameters. They can be expressed by the residual of the filter which is the difference between the measurement and the filter estimation

$$R(k) = Z(k) - X(k|k) \tag{12}$$

Substituting Eqs. (6) and (8) into Eq. (12), the residual can be rewritten to relate to the innovation sequence of the filter

$$R(k) = s(k)[1 - g(k)] \tag{13}$$

The residual from Eq. (12) or Eq. (13) directly relates to the burst size when the algorithm applies to flow measurement. More generally, the residual can be further normalized by dividing by the measurement. In addition, since a burst in the downstream of a logger would increase the flow and decrease the pressure, the residual can be generalized and normalized by

$$R_f(k) = \frac{Z(k) - X(k)}{Z(k)} \tag{14}$$

$$R_p(k) = \frac{X(k) - Z(k)}{Z(k)} \tag{15}$$

where R_f and R_p are the normalized residuals for flow and pressure, respectively. Both positive residuals correspond to bursts. Negative flow residuals indicate that the actual measurement is less than that expected or predicted value. This reflects lower than anticipated demand for water at that time, and this may be affected by external factors such as air temperature, rainfall patterns, etc. Where DMAs are interconnected it may also indicate a pipeline fault upstream of the logger (an example will be shown in later sections).

Implementation of the Algorithm

The Kalman filter described in this work assumes a linear stochastic difference system, as expressed in Eq. (1). It means that the prediction for current state is assumed to be equal to the optimal estimate of last state. The normal flow and pressure measurements from real DMAs follow a typical diurnal variation with two cyclic peaks every day: one around breakfast time (05:00–08:00 h) and one during early evening (18:00–21:00 h). Thus the model as expressed in Eq. (1) might not be suitable due to rapid expected changes between adjacent data points at these times. However, if we consider the last state to be the same time during the previous week rather than the last data in the time series sample (i.e., the last 15 min), the current linear model is reasonable. This is because current flow and pressure values are more likely to be similar to those of the previous week at the same corresponding time step. Consequently, if considering 1 week as a cycle and the sampling rate as 15 min per sample, there are a total of 672 (10,080 min per week÷15 min per sample) independent Kalman filters in the algorithm. On this basis the hydraulic measurement Z is reconstructed into 672 sets of measurements for the Kalman filters and can be expressed as

$$Z_n(k) = Z(672(k-1)+n) \tag{16}$$

where Z is the original adjacent measurement; Z_n is the new n-dimensional measurement for the nth Kalman filter, and $n=1,2\ldots672$; k is the new state step for the Kalman filter, and $k=1,2,3\ldots672(k-1)+n=1,2,3\ldots$, is the state step in the original adjacent measurement.

The number of the Kalman filters depends on the sampling rate of the measurement. For example, there are 10,080 independent Kalman filters for the sampling rate of one sample per minute. Higher sampling rates will further increase the number of the Kalman filters. However, the algorithm is still efficient because the computation of each Kalman filter is extremely efficient [see Eqs. (3)–(11)]. In addition, the filter is recursive so only the last innovations at size of M [see Eq. (11)] and the last estimate and covariance need to be stored in the algorithm. The filter only needs to converge at the beginning of the data set. When the filter has converged the system knowledge has been attained and is represented by the updated covariances Q and R. Thus the filter only needs the current measurement for estimation, instead of a long sequence of previous weeks' data. Current experiments (described below) show that the computation time for processing 3 months of data is less than 2 s on a 3-GHz desktop with 3-Gbyte random-access memory.

To start the algorithm, $X(0|0)$, $p(0|0)$, $q(0)$, and $r(0)$ have to be initialized. These can be estimated from the initial conditions. $X(0|0)$ can be given roughly as the first measurement $Z(1)$, i.e., 1.01 times of $Z(1)$. This can give a low level of noise to the model prediction [Eq. (3)] which is usual in real cases. $p(0|0)$, $q(0)$, and $r(0)$ can be given as 1. This can let the filter take account of a level of uncertainty both in the process prediction and measurement [see Eqs. (4)–(6)]. If the measurement is likely to be true or confident (no bursts or leaks), we can adjust $X(0|0)$ to more closely approach $Z(1)$, increase $q(0)$, or decrease $r(0)$. This can speed up the filter convergence. However, $p(0|0)$, $q(0)$, and $r(0)$ are not suggested to be given zero values because it would make the filter 100% conform to the process prediction [Eq. (3)] or the measurement [Eq. (4)], and consequently may cause an unconverged result. The convergence process might last a couple of weeks to reach a stable output since the gap between each state is 1 week. However, once the filter is converged, the system knowledge (the covariances Q and R) will be updated automatically and hence the filter only needs current measurement for output.

Experiments and Results

Data Acquisition

To test and validate the adaptive Kalman filtering burst detection method, it has been applied to two kinds of data. The first kind of data is with known burst events manually simulated in engineered tests. The second kind of data is with natural burst and leak events which can only be assessed from the event records from the water company including customers' contacts and a knowledge of historic pipelines' repair work from maintenance records (for the purposes of validating the computational approach and the "events" it detects).

The in situ flow and pressure were measured by direct insertion probes (such as Intellisonde, Intellitect Water, U.K.) installed in the district inlet pipe by the water company in the north of England. The probe measured the flow and pressure at an interval of 15 min (to reflect the current U.K. industrial standard) and the data were transmitted to the water company every 30 min through a GPRS mobile network.

Engineered Tests

Fig. 1 shows the pipeline layout of the DMA used for engineered tests in the north of England. The DMA has a total main length of 17.8 km and has 897 domestic properties and 28 commercial properties. There are no consumers with an annual demand of over 5% of the zone flow. The DMA is isolated by 10 boundary valves and one internal valve. The engineered tests were designed to model large burst flows and were simulated by opening five hydrants to the environment. The pipeline distances from the inlet flow meter to the hydrants (from H1 to H5 in Fig. 1) are approximately 1 km, 1.1 km, 2.1 miles, 3.4 km, 2,7 km, and 5 km, respectively. The flow rates were not likely to cause low pressure to any customers, and the flush water did not disrupt the environment. Since the DMA inlet logger was located in a predominantly rural area and the GPRS coverage was poor, the flow data were manually downloaded. Fig. 2 shows the flow measurements and

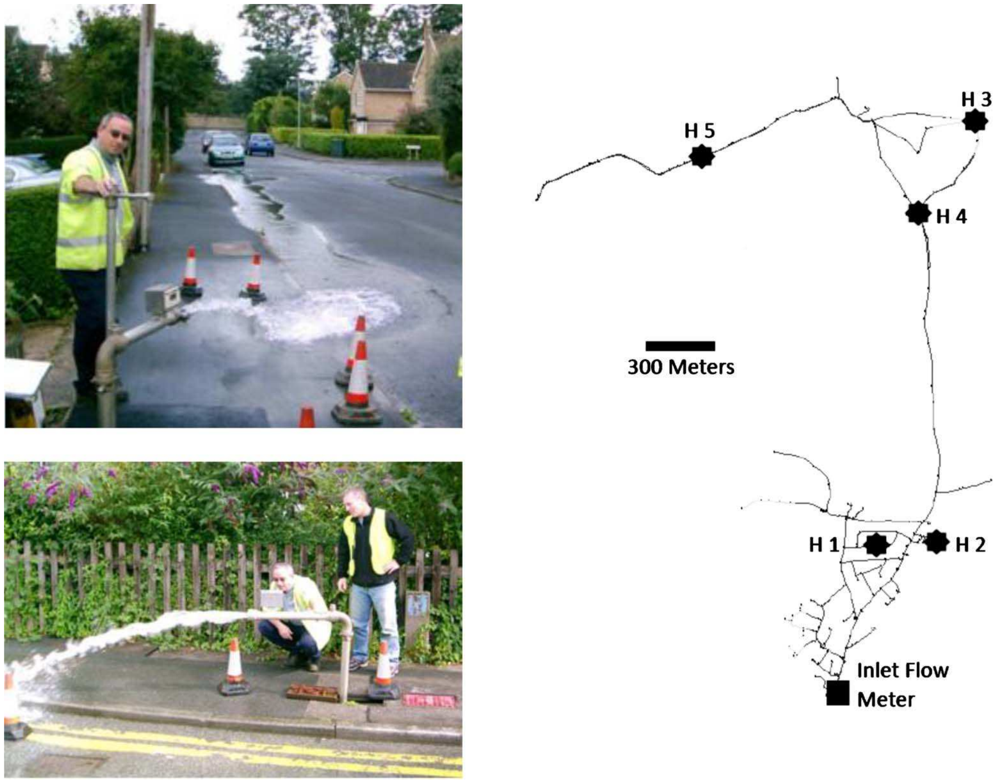


Fig. 1. Pipeline layout of the DMA in the engineered test; H1–H5 indicate the five hydrants used in the test

the Kalman filtering results (the estimated flow and the residual) on one normal day and two engineered test days.

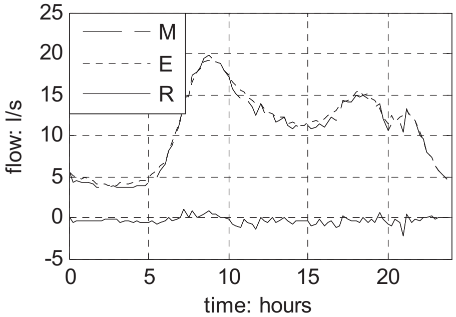
On July 20, 2008 [Fig. 2(a)], no engineered test was operated, which was a normal day. The estimated flow follows the measured flow, so the filter produces a low residual with a zero mean. On August 7, 2008 [Fig. 2(b)], an engineered test to model a large burst flow was simulated by opening five hydrants (H1–H5 in Fig. 1) to the environment in different locations within the DMA. The hydrants were opened one by one (from H1 to H5) and each hydrant was opened for about one and a half hours. The first hydrant (H1) was open at 7:30 and the last hydrant (H5) was closed at 14:20. This period is indicated by the gray region in Fig. 2(b). The average flow released from the hydrants in this testing period is 6.2 L/s. In Fig. 2(b), there is a great difference (residual) between the measured flow and the estimated flow from 7:30 to 14:30 which exactly matches the time of the engineered test (note that the flow data interval is 15 min). The average residual in this period is 5 L/s which is similar to the rate lost through the hydrants.

On August 8, 2008 [Fig. 2(c)], another engineered test was operated by opening four different hydrants (H1, H2, H4, and H5). Hydrant 5 was opened from 7:41 to 9:44 and Hydrant 4 was opened from 8:44 to 10:45. Hydrant 1 was then opened from 10:53 to 12:52 and Hydrant 2 was opened from 11:54 to 13:54. Fig. 2(c) shows the whole engineered test period from 7:41 to 13:54 indicated by the gray region. The residual shown on the figure matches the engineered test very well. The average residual in this period is 1.8 L/s which is also close to the average flow of 2.5 L/s from the hydrants.

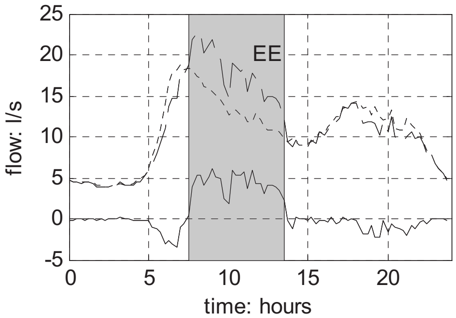
Small negative residuals can be produced during the early morning period from 5:00 to 8:00 [see Figs. 2(b and c)]. This is because for this DMA the last few weeks' flow rates during these hours were continually high. In these circumstances the Kalman filter may be unable to estimate accurately the true normal water usage and hence produces a bias (higher) estimate. Also, since water usage rapidly increases in the early morning as people rise and prepare for work, a slight shift in this pattern due to seasonal changes at this time of day (e.g., people getting up slightly later) can produce a large difference between the measured flow and the estimated flow leading to an unexpected residual.

Flow-Based Real Event Detection

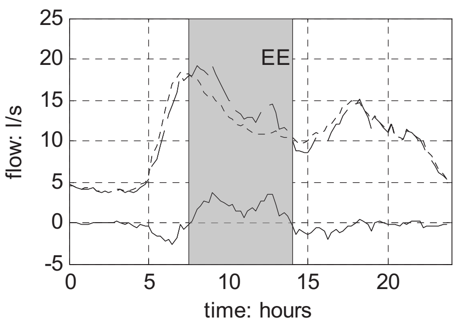
The Kalman filtering burst detection algorithm was applied to 10 DMAs' flow data randomly selected from a large supply area in the north of England. Each DMA's data contained flow measurements and four sets of DMA data contained pressure measurements. The burst detections identified by the Kalman filtering procedure were validated by the event information routinely recorded by the water company including customer contacts (complaints) and pipeline reparation work. Due to the uncertainties on these recorded events (customer complaints and reparation works), it could not be guaranteed that a real burst always occurred during a recorded event. In this study, a real burst event can only be inferred manually through the frequency of customer complaints and reparation works occurring in a short period of time. Occasional customer complaints may be caused by short-term fluctuations in industrial water usage and so they may not represent a real event. The simultaneous occurrence of customer complaints and reparation works is likely to indicate a real burst. In addition, the recorded dates of the customer complaints and reparation works can only approximately indicate the dates of burst since the customer may make contact with the water company some time after a problem has occurred and the reparation work may be actually started days after the job created appears on the maintenance record.



(a) 20 July 2008



(b) 07 August 2008



(c) 08 August 2008

Fig. 2. Engineered test results. Broken lines: flow measurement (M); dot lines: flow estimated by the Kalman filter (E); solid lines: residual (the difference between the measured flow and the estimated flow) (R). The gray regions indicate the engineered test periods.

Both flow and pressure data were measured at intervals of 15 min. In some DMAs there are periods of missing information (gaps in the record of hours or even weeks) due to temporary sensor failures and out-of-service mobile networks. When implementing the Kalman filter procedure, the algorithm automatically skips the gaps and leaves blank outputs.

Figs. 3 and 4 show the flow measurement time series at the top of the figures and the flow-based burst detection results on the bottom of the figures for the 10 DMAs reported in this study, comparing with the customer complaints and pipeline reparation work. In the bottom of each figure, the dark (black) bars indicate the periods of customer complaints to the water company concerning no water or low water pressure. The light (gray) bars indicate the periods of pipeline reparations in the downstream of the DMA.

The line plotted at the bottom of each figure shows the burst detection result with a magnitude related to a burst. A greater normalized residual more likely links to a burst event. These results were produced by the following procedures. First, a moving average window has been applied to $Z(k)$ and $X(k)$ in Eq. (14) on the data points over the last week to smooth the residual over time

$$R_f(k) = \frac{F\{Z(k)\} - F\{X(k)\}}{F\{Z(k)\}} \tag{17}$$

where F denotes the low pass filtering by the moving average window. Second, the negative residual was set to zero since it corresponds to lower than expected water usage patterns. To account for a certain level of this variation of the normal water usage, the value of the residual from Eq. (17) less than 0.01 was set to zero in practice

$$D_f(k) = \begin{cases} R_f(k) & \text{if } R_f(k) \geqslant 0.01 \\ 0 & \text{if } R_f(k) < 0.01 \end{cases} \tag{18}$$

where $D_f(k)$ is the burst detection shown as a line in the bottom part of each figure in Figs. 3 and 4. It also represents the estimation of the size of the burst.

In DMA 1, the residual was positive (nonzero) for an extended period from May 22, 2008 which was about 1 month prior to a series of recorded customer complaints and reparations. This would indicate a continual burst event and matched the frequently occurring customer complaints and recorded reparations.

In DMA 2, there was a data-missing gap from July 26, 2008 to September 5, 2008, although there were customer complaints in this period. The residual was positive coinciding with a burst event on October 30, 2008. This alarm exactly matched the customer complaint on October 30, 2008 and the reparation work started on October 31, 2008. The residual dropped to zero on November 2, 2008 when the reparation had fixed the burst.

In DMA 3, the residual increased from May 22, 2008 to indicate a burst, matching the customer complaints on May 20, 2008 and May 24, 2008. The following reparation started on June 10, 2008. The residual again increased on the same day and kept positive, indicating that the first reparation did not solve the problem until the second reparation on July 9, 2008 was carried out. From July 10, 2008 to October 15, 2008, the zero residual produced no alarms, confirmed by the normal water supply operation with no customer complaints and reparations in this period. On October 15, 2008, the residual increased to indicate a burst, also confirmed by the frequently occurring customer complaints until November 14, 2008 when the residual dropped to zero. Finally, customers complained on January 12, 2009 and January 27, 2009, but the residual did not produce any alarm here. This may not be a real burst event and the complaints may have related to issues such as water quality (discoloration).

DMA 4 is located contiguously immediately downstream of DMA 3 [see their relation in Fig. 5(a)]. In the early part of Fig. 3(d), there were frequent customer contacts and reparation works from March 28, 2008 to May 18, 2008. This was reflected in positive residuals indicating likely burst events in this period. After a normal period of operation in June and July, the residual increased to produce an alarm from August 26, 2008 to September 5, 2008, confirmed by the customer contacts on August 23, 2008 and August 30, 2008.

After that, there were frequent customer contacts and a reparation work during October, but there was no alarm produced. The problem in the water distribution system was caused by a burst or over water usage (as described earlier) in the upstream DMA [see Fig. 5(a) for the relation between DMA 3 and DMA 4]. The raw residuals [computed from Eq. (17)] shown in Figs. 5(b

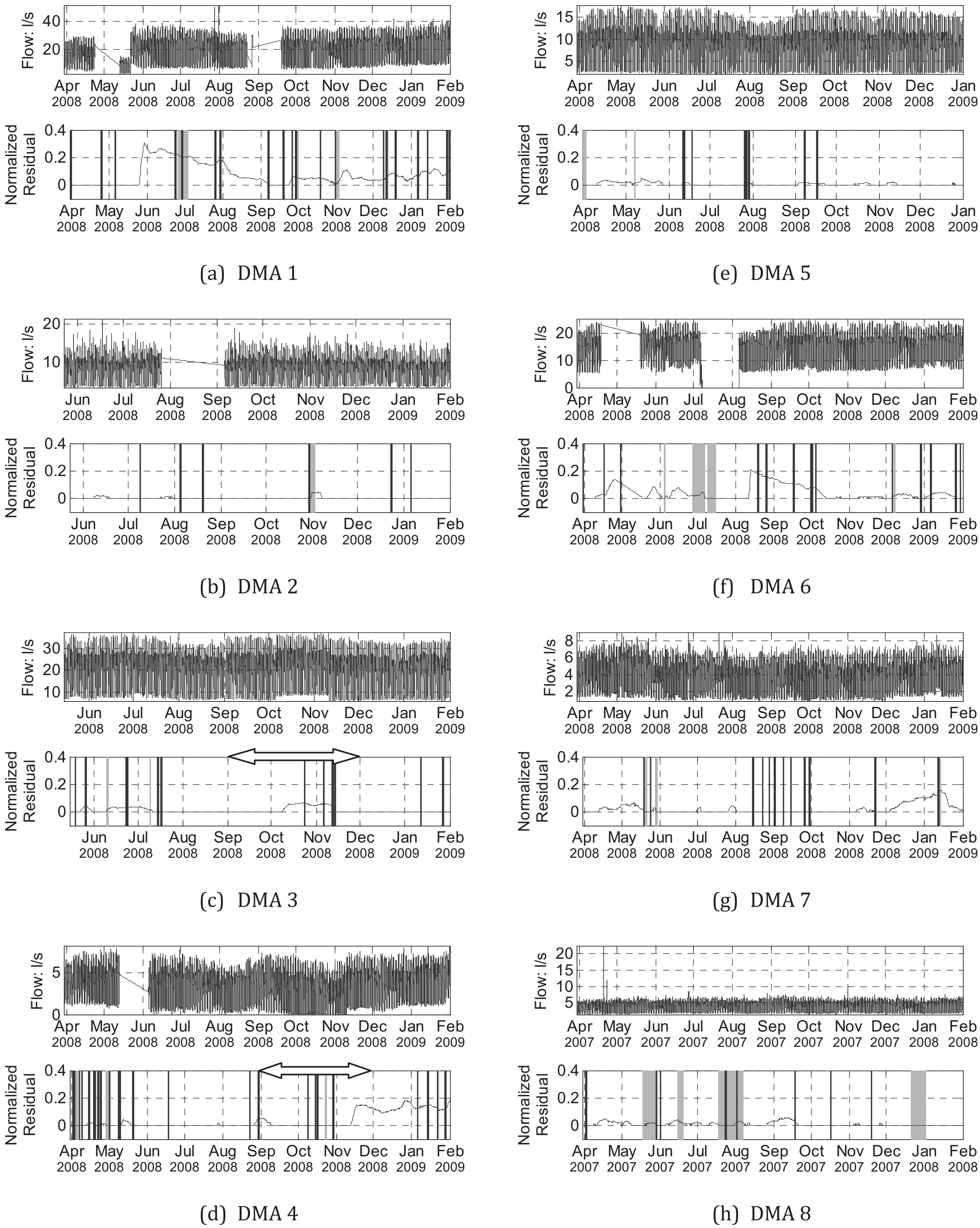
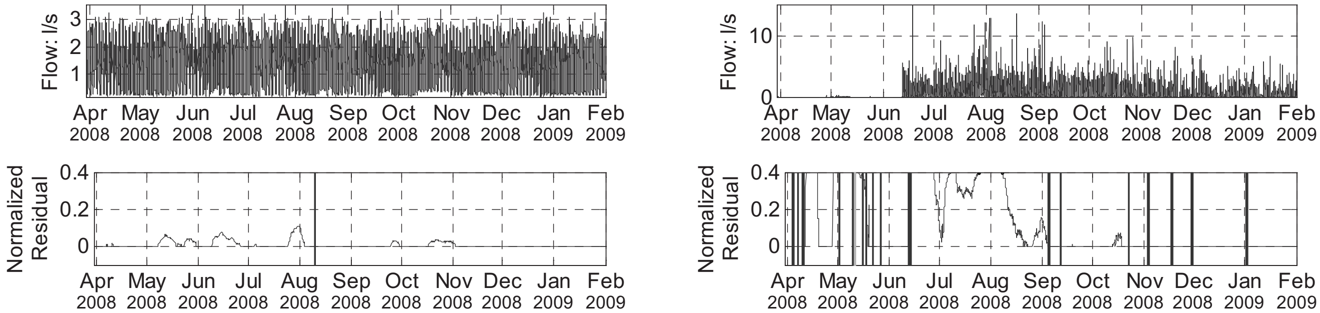


Fig. 3. Flow data and burst detection in DMA 1–DMA 8. The dark (black) bars indicate the time stamps of customer contact to the water company to report low water flow or no water. The light (gray) bars indicate the time stamps of pipeline reparation work. The original residuals (both negative and positive) during the time period of September–December 2008 indicated by the arrows in (c) and (d) will be shown in Fig. 5 to discuss their relation.



(a) DMA 9

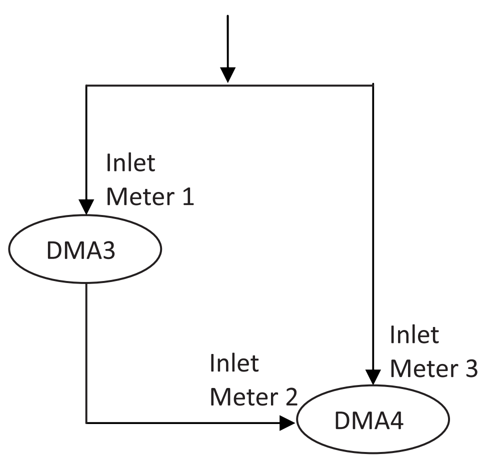
(b) DMA 10

Fig. 4. Flow data and burst detection in DMA 9 and DMA 10 with poor operational records/corrupted data signals. The dark (black) bars indicate the time stamps of customer contact to the water company to report low water flow or no water. The light (gray) bars indicate the time stamps of pipeline reparation work.

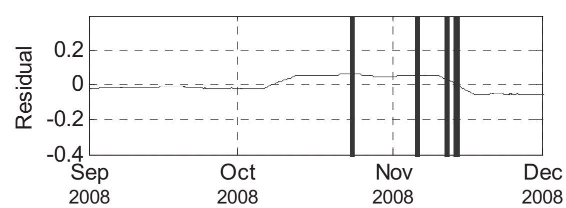
and c) confirm that more than normal quantities of water (positive residual) moved into DMA 3 but less than normal quantities of water (negative residual) moved to the downstream DMA (DMA 4). This explains the customer complaints in DMA 4 in this period.

Finally, the residual in Fig. 3(d) increased to alarm a burst from November 14, 2008 to January 31, 2009 (the end of the data), but this alarm period was only confirmed by the customer contacts and a reparation work in January 2009. The first part (before December 2008) of the alarm might be false and caused by the sudden increase of the inlet water from the upstream DMA 3 being restored to DMA 4 on November 14, 2008 when the problem in DMA 3 had been resolved. As the water flow had kept low for months before that date, the Kalman filter had compounded the level of this low water flow as normal. Consequently the Kalman filter produced a strong residual once the flow suddenly increased on November 14, 2008. Since this algorithm is self-supervised without any prior information confirming a burst or not, the Kalman filter considers all the previous water usage as normal and then automatically updates its knowledge. If the previous operations were abnormal for a long time, either overnormal flow (bursts) or undernormal flow (this case), the Kalman filter could produce false alarms. Where DMAs are interconnected in this way such a complex situation can arise, but when the output from the procedure is interpreted together the Kalman filter can cross confirm related events in adjacent DMAs. This highlights the importance of knowing meter location in interpreting results.

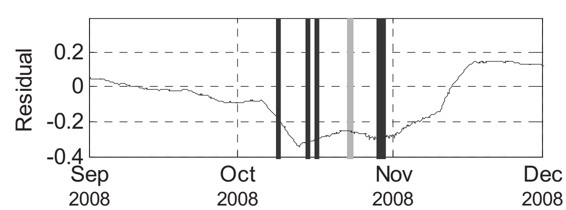
In DMA 5, DMA 6, DMA 7, and DMA 8 (Fig. 3), the residuals indicated bursts or leaks corresponded well with the customer contacts and reparation works. In DMA 9 (Fig. 4), positive residuals were produced to alarm bursts, but due to an incomplete event database for this DMA there was only one customer contact recorded on August 10, 2008 and so the detected events could not verified from the operational records. In DMA 10 (Fig. 4), flow measurements were frequently missing and hence led to poor results. These show the practical difficulties which can arise when either the data signal is corrupted or missing or the validation event information is poor, and in these circumstances little useful knowledge can be inferred.



(a) Relation between DMA 3 and DMA 4



(b) Original residual in DMA 3 (meter 1)



(c) Original residual in DMA 4 (meter 2)

Fig. 5. There is one inlet to DMA 3 and two inlets to DMA 4. In this work, Meter 1 was used to measure the inlet flow for DMA 3 and Meter 2 was used for DMA 4. Note that Meter 3 was not used. (b) and (c) show the raw residuals [computed from Eq. (17)] of DMA 3 (Meter 1) and DMA 4 (Meter 2) during the time period of September–December 2008 [indicated by the arrows in Figs. 3(c and d)]. They show that a burst in DMA 3 (upstream) can cause a high inlet flow in Meter 1 but a low inlet flow in Meter 2 to DMA 4 (downstream).

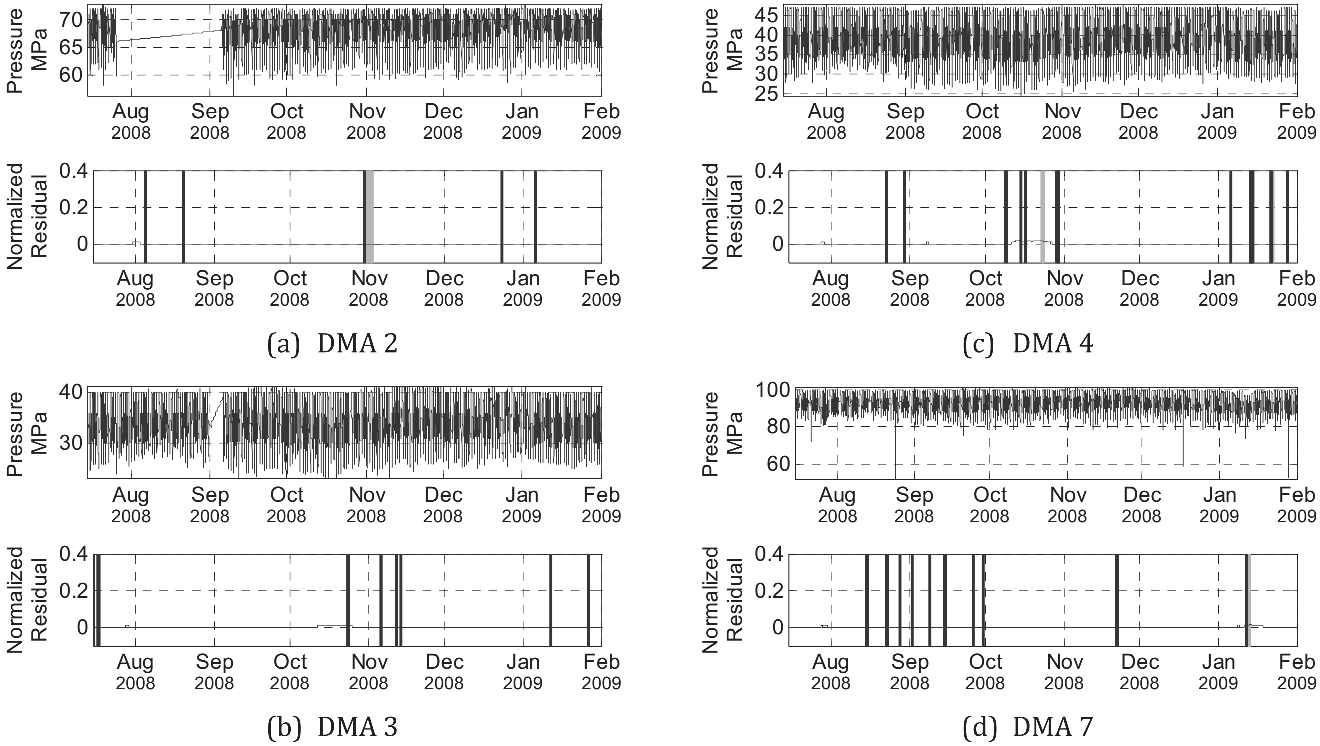


Fig. 6. Pressure data and burst detection in DMA 2, DMA 3, DMA 4, and DMA 7. The dark (black) bars indicate the time stamps of customer contact to the water company to report low water flow or no water. The light (gray) bars indicate the time stamps of pipeline reparation work.

Pressure-Based Real Event Detection

The algorithm was then applied to four sets of pressure data which were available from July 15, 2008 to January 31, 2009 in DMA 2, DMA 3, DMA 4, and DMA 7. Fig. 6 shows the pressure measurement time series at the top of each figure and the pressure-based burst detection results on the bottom of each figure. Again these can be compared with the customer contacts (black bars) and pipeline reparation work (gray bars). The burst detection results were produced following the same procedure as developed for the flow measurement in Eqs. (17) and (18). Generally, the pressure-based detection was not as sensitive as the flow-based detection, producing fewer alarms.

In DMA 2, the residual did not change to indicate the known burst around October 30, 2008 which was detected by the flow measurement approach and was also confirmed by customer contacts and reparation work. In DMA 3, the residual changed to suggest a burst from October 15, 2008 to October 25, 2008. This matched the alarm generated by the flow measurement and was confirmed by customer contacts and reparation work.

In DMA 4, there were frequent customer contacts and reparation work from October 09, 2008 to October 30, 2008. Fig. 3(d) showed that the flow measurement approach did not indicate a burst due to the overusage or burst in the upstream DMA causing less water to flow into DMA 4 and hence explaining the customer complaints. However, Fig. 6(c) shows that the pressure residual increased from October 11, 2008 to October 27, 2008, producing a burst alarm which matched these customer complaints.

In DMA 7, the residual increased from January 11, 2009 to January 17, 2009 to indicate a burst. This matched the flow-based detection shown in Fig. 3(g) and also was confirmed by the customer contact on January 12, 2009 and the reparation on January 13, 2009 and January 14, 2009.

A combination of using both flow and pressure data together for burst detection was tried through considering the residuals of flow and pressure and their difference as follows:

$$P(k) = \exp\left[-\frac{(R_f - R_p)^2}{\sigma|R_f R_p|}\right] \tag{19}$$

$p(k)$ ranges from 0 to 1 and represents the probability of occurring a burst. R_f and R_p are the normalized residuals of flow and pressure [see Eq. (17)]. σ is a positive number depending on the system. Finally, the burst detection is expressed by

$$D(k) = \begin{cases} R(k) & \text{if } R(k) \geqslant 0.01 \text{ and } P(k) \text{ is} \geqslant 0.5 \\ \text{else} & 0 \end{cases} \tag{20}$$

where

$$R(k) = \frac{R_f(k) + R_p(k)}{2} \tag{21}$$

The result showed that the combination does not improve the detection. This is because the prior information on the sensitivity of the pressure is lacking. The results presented here showed that flow measurement is more strongly related to a burst than pressure measurement, but it is still hard to quantify their relation. The data need to be carefully interpreted in relation to meter location and site specific circumstances.

Conclusion

This paper has presented an automatic burst detection algorithm based on the Kalman filtering of flow and pressure measurements. The Kalman filter tends to adaptively estimate the normal flow and pressure of the water distribution systems. Thus the filter's residual, the difference between the measurement and the filter output, can indicate abnormal water usage patterns. A positive residual in Eq. (14) or Eq. (15) indicates and excess flow of water or a pressure drop. The algorithm provides an alternative burst detection method to other approaches, with the advantages of computational efficiency, requirements for less training data, self-supervision, and fast response. However, the method seems better able to detect sudden bursts or gradually changing leaks than long-term stable leaks which may already exist in many systems. This is because without clear baseline data from a known non-leakage system or a nonleakage model as a reference, the algorithm assumes that historic data are from a normal operation. Thus it would be hard to detect stable leaks, unless the method is calibrated.

The algorithm has been validated by applying to a DMA with engineered tests and 10 DMAs with real burst events identified by customer contacts (or complaints) to the water company and pipeline reparation works. The engineered test results showed that the positive flow residuals match the simulated bursts very well. The real event results showed that the flow residual can detect most of the burst events. The pressure residuals seem less sensitive to a burst event. These usually produce fewer alarms but when an alarm is generated they coincide well with known events. The sensitivity of the pressure data to a burst depends on local pipe friction parameters and the location where the pressure is logged, so a pressure sensor which is remote from the burst location produces less sensitive measurement (Pudar and Liggett 1992). Pressure sensors in U.K. networks are often placed at topographically high points for the purpose of compliance monitoring to ensure that minimum pressures are maintained and these locations are often less useful for burst detection. Recent research (Farley et al. 2008) has been interested in optimizing the sensor location to produce more sensitive pressure measurement.

Although the high flow of water is likely to be caused by bursts downstream of the logger, further investigation is also necessary to confirm it. Besides bursts and leaks, such high flows can be affected by a sudden industry water usage at one or more consumption nodes (in the DMA) and also by the weather influencing water consumption patterns or possibly by changes in water distribution system reservoir water levels (outside of the DMA). Industry water usage is usually reported to the water companies and can be recorded. Further research needs to consider the effects of the air temperature and the rainfall information as partial explanatory external casual factors in the residual fluctuation. This will assist in the interpretation of the model output.

The pressure-based detection can provide additional information as a way of confirming the flow-based detection. However, the prior information of the relation between flow and pressure is lacking, which makes the combination more difficult. Further work is needed to investigate their relation when used in combination.

Acknowledgments

This work is part of the Neptune project funded by the U.K. Engineering and Physical Sciences Research Council, Grant No. EP/E003192/1. The writers would like to thank Mr. Ridwan Patel (Yorkshire Water Services) and Professor Joby Boxall, Dr. John Machell, and Dr. Steve Mounce (University of Sheffield) for conducting fieldwork and assistance with data sets.

References

Bar-Shalom, Y., and Fortmann, T. E. (1988). *Tracking and data association*, Academic, New York.

Bar-Shalom, Y., Li, X. R., and Kirubarajan, T. (2001). *Estimation with applications to tracking and navigation*, Wiley, Chichester.

Bishop, C. M. (1995). *Neural networks for pattern recognition*, Oxford University Press, Oxford.

Brothers, K. J. (2001). "Water leakage and sustainable supply—truth or consequences?." *J. Am. Water Works Assoc.*, 93(4), 150–152.

Buchberger, S. G., and Nadimpalli, G. (2004). "Leak estimation in water distribution systems by statistical analysis of flow readings." *J. Water Resour. Plann. Manage.*, 130(4), 321–329.

Colombo, F. A., Lee, P., and Karney, B. W. (2009). "A selective literature review of transient-based leak detection methods." *J. Hydro-environment Research*, 2, 212–227.

Farley, B., Boxall, J. B., and Mounce, S. R. (2008). "Optimal locations of pressure meter for burst detection." *Proc., Water Distribution Systems Analysis*, Kruger National Park, South Africa.

Grewal, M. S., and Andrews, A. P. (2001). *Kalman filtering: Theory and practice using MATLAB*, Wiley, New York.

Kalman, R. E. (1960). "A new approach to linear filtering and prediction problems." *Trans. ASME J. Basic Engineering*, 82, 35–45.

Khan, A., et al. (2005a). "Artificial neural network model for a low cost failure sensor: Performance assessment in pipeline distribution." *Int. J. Math. Comp. Sci.*, 1(1), 27–33.

Khan, A., et al. (2005b). "Performance assessment of leak detection failure sensors used in a water distribution system." *J. Water Supply: Res. Technol.-AQUA*, 54(1), 25–36.

Maybeck, P. S. (1979). *Stochastic models, estimation, and control*, Academic, New York, San Francisco, London.

Mehra, R. K. (1970). "On the identification of variances and adaptive Kalman filtering." *IEEE Trans. Autom. Control*, 15, 175–184.

Mounce, S. R., et al. (2002). "A neural network approach to burst detection." *Water Sci. Technol.*, 45, 237–246.

Mounce, S. R., et al. (2003). "Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system." *Inf. Fusion*, 4, 217–229.

Mounce, S. R., Boxall, J. B., and Machell, J. (2007). "An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation." *Water management challenges in global change*, B. Ulanicki, D. Butler, F. A. Memon, K. Vairavamoorthy, and P. L. M. Bounds, eds., Taylor and Francis, London, 313–320.

Mounce, S. R., Boxall, J. B., and Machell, J. (2010). "Development and verification of an online artificial intelligence system for burst detection in water distribution systems." *J. Water Resour. Plann. Manage.*, 136(3), 309–318.

Mounce, S. R., and Machell, J. (2006). "Burst detection using hydraulic data from water distribution systems with artificial neural networks." *Urban Water*, 3(1), 21–31.

Obradovic, D. (2000). "Modelling of demand and losses in real-life water distribution systems." *Urban Water*, 2, 131–139.

Price, E. J. (2003). "The use of residuals for adaptive signal processing." Ph.D. thesis, Dept. of Engineering Science, Univ. of Oxford, U.K.

Pudar, R. S., and Liggett, J. A. (1992). "Leaks in pipe networks." *J. Hydraul. Eng.*, 118(7), 1031–1046.

Ye, G. (2008). "Model-based ultrasonic temperature estimation for monitoring HIFU therapy." Ph.D. thesis, Dept. of Engineering Science, Univ. of Oxford, U.K.

Yurdusev, M. A., et al. (2009). "Water use prediction by radial and feed-forward neural nets." *Proc. ICE J. Water Management*, 162(WM3), 179–188.