Novelty detection for time series data analysis in water distribution systems using support vector machines

Stephen R. Mounce, Richard B. Mounce and Joby B. Boxall

ABSTRACT

The sampling frequency and quantity of time series data collected from water distribution systems has been increasing in recent years, giving rise to the potential for improving system knowledge if suitable automated techniques can be applied, in particular, machine learning. Novelty (or anomaly) detection refers to the automatic identification of novel or abnormal patterns embedded in large amounts of ''normal'' data. When dealing with time series data (transformed into vectors), this means abnormal events embedded amongst many normal time series points. The support vector machine is a data-driven statistical technique that has been developed as a tool for classification and regression. The key features include statistical robustness with respect to non-Gaussian errors and outliers, the selection of the decision boundary in a principled way, and the introduction of nonlinearity in the feature space without explicitly requiring a nonlinear algorithm by means of kernel functions. In this research, support vector regression is used as a learning method for anomaly detection from water flow and pressure time series data. No use is made of past event histories collected through other information sources. The support vector regression methodology, whose robustness derives from the training error function, is applied to a case study.

Key words | data analysis, leakage, novelty detection, support vector machines, water distribution systems

Stephen R. Mounce (corresponding author)
Richard B. Mounce
Joby B. Boxall
Pennine Water Group,
Department of Civil and Structural Engineering,
University of Sheffield,
Sheffield S1 3JD,
UK
E-mail: *s.r.mounce@sheffield.ac.uk*

INTRODUCTION

Water utility companies are collecting ever-increasing amounts of data from sewerage and clean water distribution systems *via* loggers and telemetry systems. Interpreting this information is a challenging task due to:

- the large volumes of data
- missing data, data errors from sensors, and communications and data noise
- the parameter pattern being specific to a particular location, often with a predictable though changing fingerprint due to various forms of seasonality (daily, weekly and seasonal)
- inaccuracies in hydraulic models due to imprecise and outdated asset information, poor calibration and lack of system operational feedback.

UK industry practice is to separate water distribution systems (WDSs) into hydraulically isolated district meter areas (DMAs), which are generally permanent in nature (apart from occasional temporary rezoning). There has been recognition of zonal (DMA) monitoring as international best practice for monitoring and managing leakage in recent years (WHO 2001; Farley 2008). Although hydraulic measurements have historically been taken before DMAs were introduced, technological advances in flow metering and data capture and communications are facilitating identification of events in the distribution network. The size of zonal areas is of course a matter of debate, and consequently varying practices exist in different parts of the world. Janković-Nišić *et al.* (2004) recommended that the size of the monitored DMA

area should be smaller than it is typically in operational practice of water utilities in the UK, and further that it should be determined independently for every distribution network. Each DMA (consisting of approximately 200–2000 properties) is instrumented with a flow and pressure meter at the inlet. Most DMAs then typically have an additional pressure meter (DG2, regulated minimum pressure standard), which is usually located at the point of highest elevation in order to monitor the minimum pressure in a DMA. The flow meter is generally primarily used to derive information about leakage, though current research is also investigating utilizing pressure for leakage detection (e.g. Wu *et al.* 2010). Currently, the standard approach is based on using the input/output/night-line to calculate the water loss per DMA. A large change in this value (from day to day) or more gradually over a longer period will alert the water company to a problem in a specific DMA (Alegre *et al.* 2000).

Owing to the availability of cheaper telemetry (such as *via* SMS and GPRS), water utility companies are investigating using near real-time flow and pressure data directly for the detection of pipe rupture and other problems by applying simple flat-line alarm levels. Data analysis conducted in Mounce (2005) illustrated that pressure is a less reliable parameter than flow for abnormal event detection (through simulated burst by way of flushing), with the response of a particular meter more dependent on location. For example, a pressure meter at the inlet of a zone fed by a service reservoir is unlikely to display a significant drop in reaction to a burst within the zone. However, multiple pressure loggers within a DMA, especially if additional instrumentation is installed, which is becoming more common, gives the opportunity to gain additional information about event location (Farley *et al.* 2008). Research projects are exploring applying artificial intelligence and statistical techniques to improve upon flat-line alerts. Mounce *et al.* (2010a) describe an online system pilot implemented with a UK water company using an artificial neural network (ANN) and fuzzy logic system for detection of leaks/bursts at DMA level. Ye and Fenner (2010) present an automatic burst and leak detection algorithm based on the Kalman filtering of flow and pressure measurements. The technique was validated by application to a DMA with engineered tests and 10 DMAs with real burst events identified by customer contacts to the water company and pipeline reparation works. Several other methodologies have

been demonstrated on offline data. Akselaa *et al.* (2009) describe a method for leakage detection based on the self-organizing map (SOM) ANN. The data used consists of vectors of the flow meter readings and knowledge of reported leak locations used for training and validating the test results. Romano *et al.* (2009) describe a Bayesian-based system for application to flow and pressure data, which removes noise using wavelets and then uses the group method of data handling (GMDH) to predict future flow and pressure profiles for short-term changes and statistical process control (SPC) to monitor long-term changes.

Abnormality (i.e. deviation) from the normal pressure profile can result from a variety of causes. In fact, low or fluctuating pressure is the most frequently occurring operational problem. It is important for the water utility to be aware of low pressure, because, though not as severe as a complete interruption to water supply, inconvenience is caused to customers and there can be penalties applied by the regulator. Pressure abnormalities can result from the following:

- Low pressure can be caused by a large water loss resulting from a pipe burst. These are characterized by a sudden loss of water. Third-party damage can similarly cause such water loss.
- A power outage can affect a pump station and hence turn off one or more pumps. A number of failsafes generally limit this occurrence. Zones with storage will first experience a drop in pressure as water levels fall, and then a loss of service once the tanks are empty.
- Maintenance activities such as flushing, which is an important tool for helping operators to control distribution system water quality, may be the cause. At a high enough level and depending upon location, such activities can result in pressure drops within the DMA.
- Other operational events, such as tanks being taken offline, the adjustment of pressure zone boundaries, shutting down sections of the system or opening of cross-connections to adjacent zones, may also result in abnormal pressures.
- Logger and communications failures can of course also impact on the pressure data signal.

Some of these operational problems are described in more detail by Walski *et al.* (2001).

Time series novelty detection (or anomaly detection) is the process of automatic identification of novel (or abnormal)

events embedded in potentially large amounts of normal time series vectors. Novelty detection is a problem that data mining can address and is a useful approach for many applications where, for example, there is an abundance of normal data while abnormal data is scarce or when, even if abnormal data is available, abnormality is not easily defined. This last point is one of the particular challenges posed by WDSs applications suitable for machine learning approaches, as the classification of novelty is often unknown and usually uncertain in historic data sets. In addition, explicit models may be unavailable. There is generally no clear cut boundary between novel events and normal events in real-world applications, and interpretation of similarity of events can be subjective and task dependent. Application areas include, for example, intrusion detection systems (IDS) for protecting systems and their users on the internet (e.g. Mukkamala *et al.* 2002) and medical applications such as cancer detection (e.g. Tarassenko *et al.* 1995) and seizure analysis (Gardner *et al.* 2006). One way that abnormalities can be characterized is as outliers of the "normal" data.

In the field of water resources, Branisavljevic *et al.* (2009) explore online and offline data pre-processing techniques as a tool for improving anomaly detection methods (with the main aim to flag anomalous data before use in other models). After pre-processing, they applied flat-line thresholds and statistical tests to data from the Belgrade sewage system. Jarrett *et al.* (2006) explored data processing and anomaly detection techniques for data from WDSs, including control charting, time series analysis, Kriging techniques and Kalman filter techniques. They concluded that they did not find any one methodology in the literature which "answers all the questions". For water industry data, efficient techniques are needed that are data-driven, capable of self-learning, with limited exemplars and able to deal with often patchy, poor-quality data.

This paper presents the application of support vector regression (SVR) for novelty detection to time series collected from WDSs. Example detections of burst, artificial flushing and sensor failure events are given. A protocol is described for applying the approach to the application. Data from a water supply system in the UK was used for a case study that forms part of the Neptune Project. The Neptune Project is a £2.7 m UK research council and industrially sponsored project with seven academic and three industrial partners

(2007–2010). The core deliverable will be an integrated risk-based decision support system (DSS) for evaluating intervention strategies to inform decision-making for sustainable water system operation (Bicik *et al.* 2009). The analysis system described here is one component of this deliverable, providing alerts for further analysis and aggregation (http://www.neptune.ac.uk).

SUPPORT VECTOR MACHINES

A commonly encountered machine learning problem is to classify data into two or more groups. There exist many different approaches to classification, including statistical, naive Bayes and artificial neural networks (ANNs). The support vector machine (SVM) is a statistical learning theory based on machine learning methods. SVMs are widely used in the fields of bioinformatics, data mining, image recognition and hand-writing recognition. SVMs possess a number of similarities to ANNs. Both learn from experimental data and are universal approximators (i.e. they can approximate any function to any desired degree of accuracy with sufficient training data). Both can be applied to classification and regression. In contrast, they do differ by learning method: ANNs use back-propagation (or a similar gradient descent algorithm) whereas SVMs learn by solving a constrained quadratic optimization problem. This implies that there is a unique optimal solution for each choice of the SVM parameters. This is unlike other learning machines, such as standard ANNs trained using back-propagation. When applying an ANN, an appropriate structure (number of layers, neurons, etc.) is chosen and, keeping the confidence interval fixed in this way, the training error is minimized. An SVM keeps the value of the training error fixed (equal to zero or to some acceptable level) and minimizes the confidence interval (related to future generalization error). A key issue is dealing with the trade-off between under-fitting and over-fitting to the training data.

Suppose we have some training data $(\mathbf{x}_1, \mathbf{y}_1)$, $(\mathbf{x}_2, \mathbf{y}_2)$, …, $(\mathbf{x}_n, \mathbf{y}_n)$, where $\mathbf{x}_i \in \Re^k$ and $\mathbf{y}_i \in \{-1, +1\}$ indicates the class to which each vector $\mathbf{x}_i$ belongs. Any hyperplane can be expressed as the set of points satisfying the equation $\mathbf{w}^T\mathbf{x} = \mathbf{b}$, where $\mathbf{w} \in \Re^k$. Providing that the data can be separated into the two classes, we can chose $\mathbf{w}$ and $\mathbf{b}$ such

that $\mathbf{w^T x} - \mathbf{b} = 1$ for vectors on the margin in the class with $\mathbf{y}_i = 1$ and $\mathbf{w^T x} - \mathbf{b} = -1$ for vectors on the margin with $\mathbf{y}_i = -1$. Such vectors are called the support vectors and the width of the margin is $2/\|\mathbf{w}\|$, where $\|\mathbf{w}\|$ is the norm of the hyperplane normal weight vector. The SVM algorithm finds the optimal hyperplane separating the two classes. Figure 1 illustrates such a hyperplane dividing two data groups. Many other linear learning machines have been considered for the task of finding a hyperplane in a feature space that optimally separates two classes, but the SVM yields a unique solution that can be shown to minimize the expected risk of misclassifying unseen examples (Vapnik 1998). This is in contrast to the perceptron, which finds any solution.

The use of kernel functions allows all computations to be carried out in the input space, with no explicit computations in the higher dimensional space. SVMs are as powerful as nonlinear methods (such as ANNs) and statistically more robust than ANNs because of the error function used, which also limits the curse of dimensionality issue. Theoretical analysis and practical studies have shown that SVMs not only have a simple structure, but also have good generalization ability and the facility to provide a globally optimal solution (Trafalis 1999). In particular, when only limited samples are available, they tend to avoid over-learning and falling into local minima, which can be a weakness of other approaches. The *C*-support vector machine is the basic technique for classification, first proposed by Boser *et al.* (1992) and as described by Cortes & Vapnik (1995). Bishop (2007) provides more information on training and general background, and for algorithm implementation details see Chang
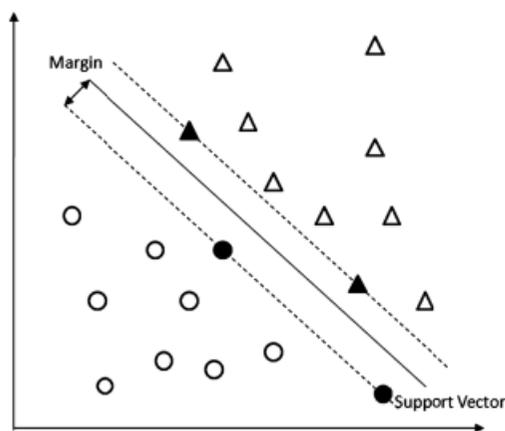
& Lin (2001). SVMs work well with limited size training data *N*, but solving the quadratic programming (QP) problem with increasing *N* is more difficult and memory intensive (Giustolisi 2004). Consequently several authors have proposed decomposition methods to solve QP or a linear programming (LP) approach designed to be more robust and faster in large-scale problems (e.g. Bennett 1999; Kecman & Hadzic 2000). The LibSVM library used in this work implements a sequential minimal optimization (SMO) algorithm to solve the SVM's quadratic programming optimization problem, which is usually more efficient than using a QP solver. SVMs have been applied in many areas of hydroinformatics over the last decade. For example, Yu *et al.* (2004) use embedding theory to create a state space reconstruction using hydrological time series and then use SVM for regression to predict future values of the time series. Moreover, they present an EC-SVM evolutionary methodology for optimal parameter selection. Giustolisi (2006) demonstrated their use for nonlinear regression of environmental data and presented a multi-objective genetic algorithm approach for optimizing the kernel parameter, input selection and ϵ-tube.

Support vector regression (epsilon SVR)

Although the SVM was originally developed for solving classification problems, it can be extended and successfully applied to regression estimation. Many applications deal with experimental data (training patterns, observations, samples, etc.) and unlike pattern recognition problems (where the desired outputs y_i are discrete values) we need to use real-valued functions. The general regression problem involves the learning machine being provided with l training data values from which it attempts to learn the functional mapping $f(x)$ as follows. A training data set $D = \{[\mathbf{x}_i \in \Re^n, \mathbf{y}_i \in \Re, i = 1, \ldots, l]\}$ consists of l pairs $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \ldots, (\mathbf{x}_l, \mathbf{y}_l)$, where the inputs $\mathbf{x} \in \Re^n$ are n-dimensional vectors and system responses $\mathbf{y} \in \Re$ are real valued. The SVM functional approximation is given by:

$$f(\mathbf{x}, \mathbf{w}) = \mathbf{w^T x} + b \tag{1}$$

$\mathbf{x}$ may be replaced with a transformation $\phi(\mathbf{x})$, e.g. for normalization. The prediction error is given by:

$$|\mathbf{y} - f(\mathbf{x}, \mathbf{w})| \tag{2}$$

Figure 1 | SVM classifying hyperplane.

Vapnik (1998) generalized the error function by introducing an ϵ-*insensitive loss function*. ϵ-SVR carries out the regression estimation by risk minimization, where the risk is measured by the ϵ-insensitive loss function:

$$|\mathbf{y} - f(\mathbf{x}, \mathbf{w})|_\epsilon = \begin{cases} 0 & \text{for } |\mathbf{y} - f(\mathbf{x}, \mathbf{w})| \le \epsilon \\ |\mathbf{y} - f(\mathbf{x}, \mathbf{w})| - \epsilon & \text{otherwise} \end{cases} \quad (3)$$

The error function is a key feature of SVR and the fact that it is linear is an important feature for statistical robustness with respect to outliers. This loss function only counts as errors those predictions that are more than ϵ away from the training (measured) data. The slack variables ξ and ξ^* represent the size of this excess deviation for positive and negative deviations respectively. This loss function defines an ϵ-tube around $f(\mathbf{x}, \mathbf{w})$ as shown in Figure 2. It allows the concepts of margin to be carried over to the regression case while maintaining useful statistical properties. The model produced by the support vector classification depends only on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin. Analogously, the model produced by SVR depends only on a subset of the training data, because the cost function for building the model ignores any training data close to the model prediction (within ϵ). The support vectors are those vectors that actually contribute to determining the approximating function. The objective is to minimize the empirical risk as well as $\frac{1}{2}\|\mathbf{w}\|^2$. Implementation of the optimization is described in detail in the literature (Chang & Lin 2001; Schölkopf et al. 2001). If $|\mathbf{y} - f(\mathbf{x}, \mathbf{w})|$ is sufficiently large (the difference
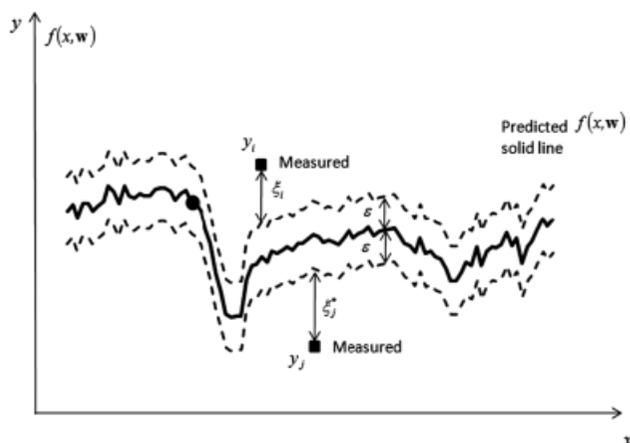

Figure 2 | Parameters for (one-dimensional) SVR.

between the observed value and the value predicted by the SVR regression model), this is classified as abnormal and termed a "surprise", as will be described in the methodology.

METHOD

SVR-based novelty detection for time series

SVR can be used to model and predict from a temporal sequence. A training set of samples can be constructed using a particular dimensional delay vector $\mathbf{m}$ of the last m observations, where the sampling time is taken as uniform with τ the lag time, i.e. the sampling rate:

$$\underline{x}_t = [x(t), x(t - \tau), \ldots, x(t - (m-1)\tau)] \text{ with target prediction}$$
$$y_t = x(t+1) \quad (4)$$

Ma & Perkins (2003) present a general framework for using a model for novelty detection with a given confidence level consisting of five key concepts. This scheme is then specialized to using SVR (summarized in Table 1). A model $\mathbf{M_x}(t_0)$ represents the knowledge about an underlying temporal sequence up to t_0, which in the case of SVR is constructed from available data $\underline{x}_t$.

Table 1 provides the basis of a methodology that allows the SVR to be used for novelty detection. The formulation of $p_{E_n}(|e_n(t_0)|)$ can be determined by the occurrences $O(t_0 + i)$, $i = 0, \ldots, n - 1$. If the occurrences in $\{O(t_0 + i), i = 0, \ldots, n - 1\}$ are identical independent Bernoulli variables, $E_n(t_0)$ becomes a binomial random variable and

$$p_{E_n}(|e_n(t_0)|) = \sum_{k=x}^{n} P_{E_n}(|e_n(t_0)| = k)$$
$$= \sum_{k=x}^{n} \frac{n!}{k!(n-k)!} q^k (1-q)^{n-k} \quad (5)$$

McKenna et al. (2007) adopt a similar approach for event detection in water quality measurement data to gather evidence over multiple consecutive time steps. A binomial event discriminator (BED) is presented that uses a failure model based on the binomial distribution to determine the probability of an event existing based on r outliers occurring

Table 1 | General and SVR-based framework for novelty detection with confidence

	General scheme	Using SVR
Matching function and value	Quantifies how well the model matches the temporal sequence. $$V(t_0) = F(\mathbf{M}_x(t_0 - 1, x(t_0)))$$	$V(t_0) = x(t_0) - \mathbf{M}_x(t_0 - 1) = x(t_0) - \hat{x}(t_0)$ i.e. the matching value is the residual of the regression function at t_0
Occurrence	An occurrence at t_0 is defined by $$O(t_0) \equiv I\{V(t_0) \notin (-\epsilon(t_0), \epsilon(t_0))\}$$ where $I\{\}$ is an indicator function and $2\epsilon(t_0) > 0$ is the tolerance width	Utilize the ϵ-insensitive loss function and define the tolerance width as 2ϵ for any t_0 where ϵ is the insensitivity parameter from SVR
Surprise	A surprise is observed if: $$O(t_0) = 1$$	Using the ϵ-insensitive loss function results in any sample that is a surprise to model $\mathbf{M}_x(t_0)$ being a support vector in the updated model $\mathbf{M}_x(t_0 + 1)$ (Smola & Schölkopf 1998)
Event and event duration	$E_n(t_0) = [O(t_0)\, O(t_0 + 1)\, \ldots\, O(t_0 + n - 1)]^T$ where n is the event duration and the number of surprises in event $E_n(t_0)$ is $$\|E_n(t_0)\| = \sum_{i=0}^{n-1} O(t_0 + i)$$ having density function $p_{E_n}(\|e_n(t_0)\|)$ where $\|e_n(t_0)\| = 0,\, 1,\, \ldots,\, n$	$E_n(t_0)$ is defined as a sequence of independent Bernoulli random variables with the same parameter: $$p_{E_n}(\|e_n(t_0)\|) = \binom{n}{\|e_n(t_0)\|} q(t_0)^{\|e_n(t_0)\|}(1 - q(t_0))^{n-\|e_n(t_0)\|}$$ where $\|e_n(t_0)\| = 0,\, 1,\, \ldots,\, n$
Novel event with confidence	Event $e_n(t_0)$ is defined as a novel event with confidence $c(t_0)$ if it satisfies (i) $\|e_n(t_0)\| > \max(h, E\{\|E_n(t_0)\|\})$ (E$\{\}$ is the mean and h is a fixed lower bound) and (ii) $p_{E_n}(\|e_n(t_0)\|) < 1 - c(t_0)$	$q(t_0)$ is the probability of an occurrence in the event $E_n(t_0)$ being a surprise and which can be approximated by $\hat{q}(t_0) = \frac{N_{SV}(t_0)}{D_{t_0}}$ with $N_{SV}(t_0)$ the number of support vectors in $\mathbf{M}_x(t_0)$ and D_{t_0} number of training samples of the form specified in (4)

within n time steps. The BED can be used on the results of any event detection algorithm that produces a binary result (success/failure) for every time step. They successfully applied the technique to 15 minute data for specific conductivity (SC), pH and oxidation reduction potential (ORP), demonstrating the potential to detect both events and baseline changes.

There are various approaches for selecting an appropriate value of lag m. A rule of thumb that Schaffer *et al.* (1988) recommend is to adopt a lag length of 10% to 30% of the periodicity of interest in the time series. However, one of the most commonly used techniques involves calculating the autocorrelation function for the time series, and then determining the first minimum of this function. In this context, the autocorrelation function is a measure of the linear dependence of a variable with a lagged (time-shifted) version of itself. A variant on this, mutual information, can be used, which is a measure of the general dependence of one variable and a lagged version of itself. The first minimum of the auto mutual information function can be used to find a good choice for a delay time (e.g. Fraser & Swinney 1986). Mounce (2005) found that in analysing a set of ten flow sensors with three months of 15 minute interval data each, this value varied between 12 and 19 (with mean 15.3). In essence, this means that most of the information regarding the next value of the time series is on average contained in approximately the previous four hours, i.e. $m = 16$.

The occurrences from Table 1 are not actually independent but under certain conditions we assume they are in order to make the detection; this is reasonable for normal training data. The approximation for $q(t_0)$ when using SVR is

reasonable and provides acceptable accuracy in the confidence level of a detection if the following conditions hold: Firstly, the regression function can sufficiently capture the dependent relationship in a temporal sequence. In practice, this is achieved by adequate model training. Secondly, all occurrences in an event should have approximately the same probability of being a surprise. This latter is sensible if the event duration n is not too large. A more robust approach can use a variant in which r different event durations are picked evenly from the range and each is applied to create a set of detection outputs, and a voting procedure can be applied to these to create a single generated output.

Implementation

The code developed to implement the scheme described utilizes LibSVM (Chang & Lin 2001), which is a library for support vector machines. It integrates C-SVM classification, nu-SVM classification, one-class-SVM, epsilon-SVM regression and nu-SVM regression. The implementation has been widely used in the academic research community and the developers were the winners of the EUNITE worldwide competition on electricity load prediction and also the IJCNN challenge (using SVR). The latest version 2.91 was released on 1 April 2010 and the MATLAB instantiation of the support vector algorithms was used for this application. The program developed implements the required data handling, normalization, formatting into time-delayed vectors (sparse format), SVM training and testing, and also realises the algorithmic procedure outlined in Table 1.

Within water distribution networks there is a distinct diurnal pattern (for both flow and pressure) and therefore the training in Equation (4) is performed for each time of day and day type (i.e. weekday, Saturday or Sunday, as usage is different between them). Hence there is a separate SVR model for each (time of day, day type) pair, which is trained using all instances of that (time of day, day type) pair from the training data. Although this can limit the amount of data available to train the separate models, fortunately SVR is not reliant on a vast amount of training data. The SVR scheme was implemented with 96 periods per day (i.e. 15 minute data) and with an embedding dimension of D, i.e. each value was predicted using only the previous D values. Hence there are 96 weekday models, 96 Saturday models and 96 Sunday

models. The model at time t is trained with all vectors of the form $[y_{t-D}, y_{t-D+1}, \ldots, y_{t-1}, y_t]$ in the training data for that day type. The SVR regression model finds an approximating function of the form:

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b \tag{6}$$

where $\mathbf{x} = [y_{t-D}, y_{t-D+1}, \ldots, y_{t-1}]$.

The original investigations had explored using a single model; however, improved results were obtained with a (time of day, day type) multiple model approach. Since this was not found to be a large computational overhead, the latter was utilized. For a single logger, and using three months data for training and six months for testing, the run time was 8.7 seconds on a dual core PC (2.2 GHz, 2 GB RAM). Although the application is not focussed on producing the most accurate possible regression, this was generally reasonable for stable data sets and found to be better for flow data because of this. For the historical data set presented in the Results section, the average R^2 value for flow was 0.86 for training data (calculated over three months) and 0.70 for test data (calculated over six months). For pressure, the average R^2 value for training data was 0.24 and 0.11 for test data. These lower R^2 values for pressure are consistent with the nature of the data (discussed in the Introduction), one specific factor observed in the pressure data was the effect of pressure reducing value (PRV) settings due to pressure management activities evident over such an extended period.

The SVR algorithmic parameters are set based on domain knowledge, empirical experimentation and with reference to Ma & Perkins (2003). Table 2 summarizes the parameters with recommended default values for this application.

The difference between the SVR model prediction and the actual value was divided by the average value for that time of day. When the modulus of the resulting value was greater than the tolerance width ϵ, this was classed as a "surprise". If enough surprises occur (as calculated using the proportion of surprises in the training data set) within a moving event window (of fixed size), this signals a novel event detection. More specifically, a surprise occurs when the difference between the observed and the predicted value is more than a given number, the tolerance width (chosen for the case study to be five) of standard deviations for that time of day, automatically calculated by the implementation. A fixed lower bound of

Table 2 | Recommended algorithmic parameter values

Parameter	Value
Epsilon-SVR cost parameter	0.01
Epsilon-SVR epsilon in the loss function	0.5 s.d.
Epsilon-SVR kernel function	RBF $K(x_i, x_j) = \exp\{\gamma \|x_i - x\|^2\}$
Epsilon-SVR kernel function gamma value	1
Embedding dimension D	4–16
Event window duration n	8
Tolerance width ϵ	5 s.d.
Confidence level C	99
Fixed lower bound of number of surprises h	$\sqrt[3]{n}$
Training period	4–12 weeks

number of surprises h (occurring in event duration) is an algorithmic parameter and the value can be varied depending on the application (hence the minimum time for detection h multiplied by the data sampling interval size). Unless this parameter is used, training data with few occurrences will result in a single surprise generating a detection, which for most applications is not desirable. Ma & Perkins (2003) suggest using $h = n/2$, though it was empirically discovered that this seems on the conservative side for water data, so the cube root of the event window size was used for the case study. The epsilon value in the ϵ-insensitive loss function of 0.5 standard deviations results in an ϵ-tube of width one standard deviation. If this value is set to zero, all vectors in the training data affect the function obtained from the SVR regression. This can over-fit the model to the training data. If ϵ is positive, only some vectors from the training data determine the function obtained from the SVR regression. The larger ϵ is, the fewer such (support) vectors. In the model, the value is set to half a standard deviation of the training data for that time of day.

The epsilon-SVR implementation from LibSVM was used with mainly default values for the algorithm and which were found empirically to be robust across a range of values. The definition of $\hat{q}(t_0)$ ensures that a more fluctuating profile will result in more surprises and hence a higher value on the

number of surprises required to be observed for a higher confidence detection.

RESULTS AND DISCUSSION

Experiments were conducted on real data from a distribution system to explore the performance of the system, and these results are now described. The aim of the system is not to detect short-term hydraulic transients (such as water hammer). Instead, the primary focus is to detect unusual flow and pressure fluctuations that can be detected from (typically) 15 minute sampled data.

Case study

The Harrogate and Dales (H&D) area in the North Yorkshire region in the UK consists of nearly 200 DMAs (excluding trunk main and industrial user DMAs) and includes approximately 122,000 properties. The water utility company installed 450 Cello Loggers equipped with a General Packet Radio Service (GPRS) communications infrastructure, to dramatically improve data transfer for both flow and pressure data. Data is communicated every 30 minutes and two readings are obtained (15 minute sampled data). An online AI system combining artificial neural networks and a fuzzy inference system (ANN/FIS) was upgraded in autumn 2009 to monitor 412 flows and pressures in the Harrogate and Dales area. It was decided to further trial the SVR-based methodology on measured data with a view to subsequent online implementation as an alternative alert type available to a decision support system (DSS). This DSS uses a methodology based on the Dempster–Shafer theory combining evidence from several independent sources/models (such as a pipe burst prediction model, a hydraulic model and a customer contacts model) to locate a pipe burst within a DMA (Bicik *et al.* 2010).

Example novel events

The SVR technique can be applied to flow and pressure data, and indeed potentially to measured water quality parameters. Associated with the alert, a $\pm$ average magnitude change for the occurrences in the window of first detection was calculated for the monitored data stream (in an online application

calculated for a newly generated alert). The ANN/FIS system is limited to calculating a burst size estimate for an increase in the existing observed flow (Mounce et al. 2007).

The first real event occurred in a single DMA, for which flow (but not pressure) was available at the inlet and a DG2 pressure point was also being measured within the DMA. Figure 3 plots the results of applying the SVR and classification algorithm in a simulated online manner to the measured data. Twelve weeks data was used to train the model and the test data consisted of approximately nine days containing the event. The peaks of the dotted lines indicate the positions and durations of the detected novel events for the test data.

This event was a major incident and is suspected to have been a temporary rezone, resulting in a flow and pressure profile change lasting approximately 20 hours. Although this resulted in a 70% increase on the maximum flow rate into the DMA, this was still insufficient to reach the high alarm level (18 l/s) on the SCADA system. In Figure 4, the first detection windows to result in a classification are graphed in more detail. The calculated number of occurrences (using Equation (5)) for the pressure SVR to result in 99.9% confidence was 2, while this was only 1 for the flow. A fixed lower bound on the number of surprises h in the event duration was used as previously outlined ($h = 2$ in the case of an event window size of 8). The engineering context here is that there is a requirement to minimize ghosts and provide more confidence

in alerts. Table 3 provides the details of the detections along with real alerts from an online AI alert system producing automated detections.

Examination of the raw data reveals that the event window of the first SVR detection for flow ($h = 2$) contains two abnormal data points. Hence theoretically a classification could be made within 30 minutes. Similarly the event window of the first SVR detection for pressure contains two abnormal data points, although the first abnormal pressure data point is 15 minutes previous to the event registering on the flow signal (instantaneous versus averaged data). Data is transferred every 30 minutes and two readings are obtained (15 minute sample rate). Further, export from the SCADA software system is limited to at most every hour, so detection within 30 minutes is somewhat hypothetical with the currently used technology. The ANN/FIS online system detections in Table 3 suffered from an overall delay of approximately 2.5 hours in total due to these issues and an additional FTP data transfer to the research prototype system. Also, data is not always available due to communications interruptions (as happened with the DG2 data in this example). However, it is evident from Table 3 that the SVR technique can result in faster detections than the previously developed AI system: in the case of the flow SVR analysis, an alert would have been generated over 8 hours earlier, even allowing for the data delay (ANN/FIS alert received time versus simulated SVR
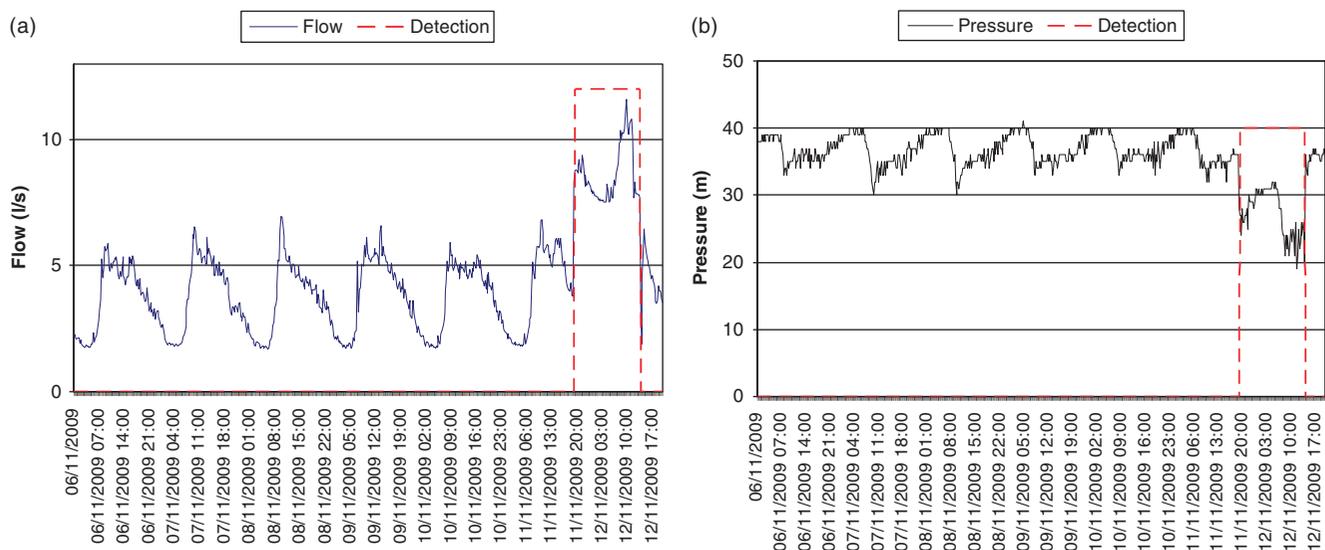

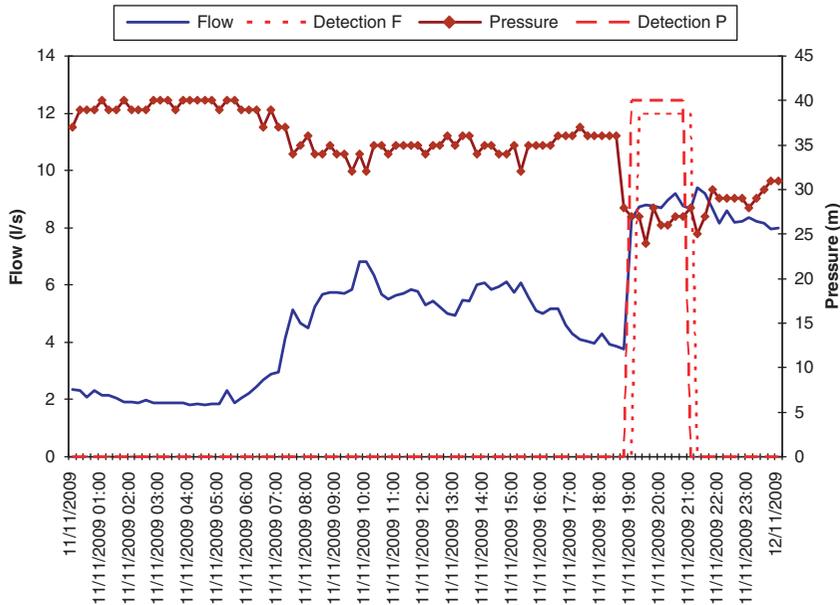**Figure 3** | Measured data for District Meter Area (DMA) with SVR detection.

Figure 4 | Detection windows for DMA A event.

alert received time). However, it should be noted that the AI system is operating on a larger window (12 hours) and has been shown to result in few ghosts for flow analysis (Mounce

Table 3 | ANN/FIS and SVR system detections for District Meter Area (DMA) A event

ANN/FIS (online system)	SVR (simulated online)
DMA: A	DMA: A
FLOW ALERT %CONFIDENCE: 99.00	*FLOW ALERT* %CONFIDENCE: 99.00
Received: 07:06 12/11/09	Simulated alert received: 19:15 11/11/2009
Dates: 11-Nov-2009 16:45:00 To 12-Nov-2009 04:45:00	Dates: 11-Nov-2009 19:00:00 To 11-Nov-2009 21:00:00
Size estimate: 4.5	Size estimate: $+3.8$ l/s
DMA: A	DMA: A
PRESSURE ALERT %CONFIDENCE: 99.00	*PRESSURE ALERT* %CONFIDENCE: 99.00
Received: 20:06 12/11/09	Simulated alert received: 19:00 11/11/2009
Dates: 11-Nov-2009 18:15:00 To 12-Nov-2009 00:15:00	Dates: 11-Nov-2009 18:45:00 To 11-Nov-2009 20:45:00
Size estimate: N/A	Size estimate: -7.8 m

& Boxall 2010). The setting of h for the SVR methodology is thus a trade-off between potential detection time and number of alerts generated (which, as the number grows, will be increasingly considered ghosts despite potentially exhibiting minor novelty). Although flat-line alarm levels can be applied to this type of time series data in an elementary manner, they can result in many ghosts, and updating their values to reflect current conditions is an issue.

Finally, the detection of two other types of novel events is demonstrated. Figure 5(a) graphs the successful detection of a hydrant flush from an in-zone DG2 meter, and Figure 5(b) shows the detection of a flow sensor failure. Note that the SVM approach as presented here identifies novelty only, and does not provide a classification on the type of novelty.

Historical analysis

A data set was assembled from the case study area for five typical DMAs (with different characteristics such as rural/ urban and size) in the water supply network. Each DMA had an inlet flow and pressure meter, and usually a DG2 pressure logger in the DMA. Data sets were assembled from operational meters for the period, and this resulted in a total of four flow meters and five pressure meters with both existing and usable data for analysis. In each case, the data consisted of time-stamped files of 15 minute readings from the pressure
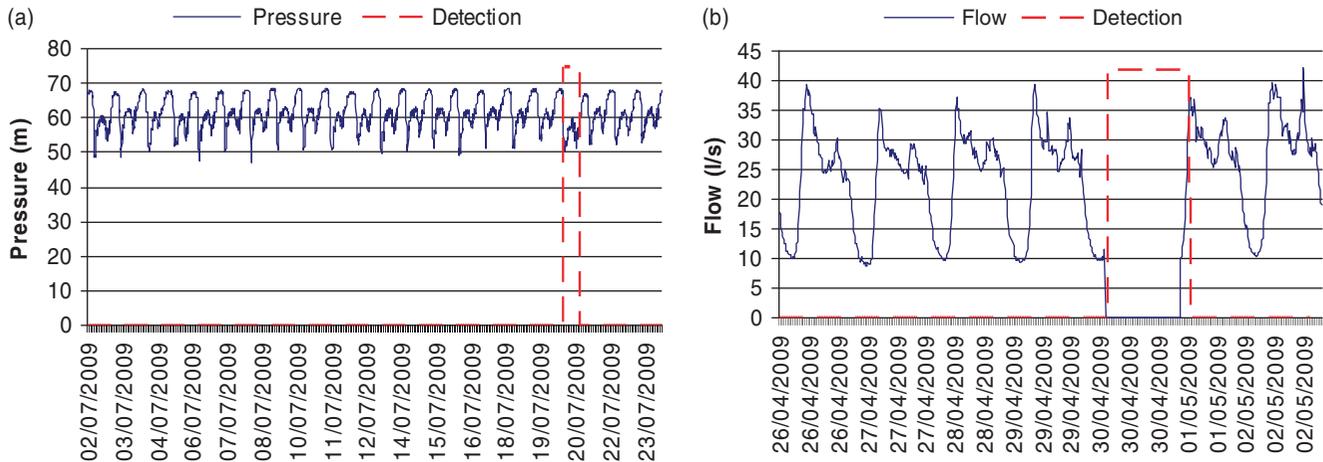
(a)

(b)

Figure 5 | (a) Hydrant flush detection (in-zone pressure, 99% confidence, −3.9 m). (b) Flow sensor failure detection (99% confidence, −22.2 l/s).

management and control telemetry system (PMAC), the work management system (WMS) mains repairs record and any associated customer contacts. A period of three months data (October–December 2008) was used for training for the period prior to the six month testing phase.

The software developed using the methodology described previously was applied to the data sets. A total of 46 detections were produced by the SVM system for the period 1/1/09 to 26/06/09. Of these:

- seven were correlated with WMS
- two were correlated to customer reports of bursts
- four were correlated to customer reports of low pressure/ no water
- 23 are "abnormal", i.e. large unusual demands or short-term increases in night-line (confirmed in the data)
- 10 are ghosts.

These results are shown graphically in Figure 6.

Pressure profiles are less sensitive to a burst or leak event. Fewer detections resulted from analysis of the historic pres-

sure data (though the majority of ghosts were from pressure analysis, perhaps reflecting the less stationary nature of pressure data used for training). A pressure effect will not always be seen in the DMA meter, for example where the pressure change can be supported by a service reservoir. The sensitivity of the pressure data to a burst or leak depends on the locality of the logger in the DMA (Farley *et al.* 2008), so a pressure sensor that is remote from the burst location is unlikely to produce a significant change from normal profile.

The "abnormal" classification covers situations where the system produced an alert and subsequent manual data investigation confirmed that an event of some type did occur, but for which there was no correlation with any further information from the utility company. Examples include large unusual demands (unknown industrial activity for example), and those related to known network maintenance or unexplained but significant short-term increases in night-line. The signature of some of these abnormal events such as large industrial demands, holiday consumption, filling of private fire tanks/ unauthorized filling of street cleaning
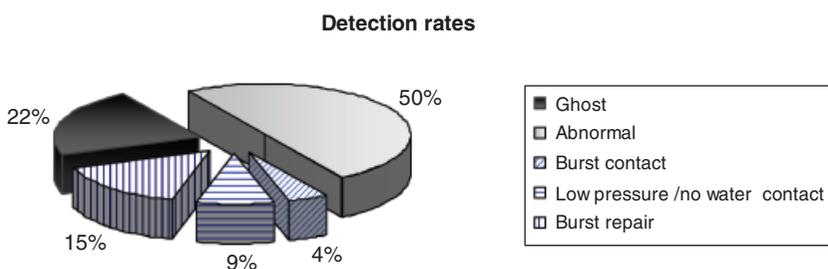
Detection rates

Figure 6 | Results for historical case study.

equipment and bowsers, and closure and opening of valves cannot be differentiated from bursts in this technique, i.e. type of novelty is not given. Figure 7 illustrates such an abnormal event detected in the historic analysis. Manual inspection reveals there is an apparent novel change in the data but with no corresponding confirmation in the WMS/contact records.

Figure 7 demonstrates the difficulty of matching with certainty the alarms/detections indicated by the modelling with real events. Obtaining exact and precise system information at any time is not straightforward. For this study, work management and customer contact records were available but no additional investigations were conducted by the water company. It is possible that in some cases real events go unnoted in the formal records, whilst considerable and sometimes unexpected changes in pressure and flow can be observed in the data. This is a root cause of uncertainty in calculating true and false detection rates. Overall, in the six month testing period, the WMS system recorded 18 burst repairs (MR35 and SE30). Seven distinct bursts corresponding to repairs were identified by the analysis system. However, non-detections of events corresponding to repairs should not necessarily be considered a false negative, as, when examining the data corresponding to the recorded repair, there is often no significantly abnormal profile change to detect.

Bursts detected (confirmed by WMS only) ranged in size from approximately 10% to 50% of the average daily maximum flow entering DMAs based on flow analysis. However, more subtle changes were also detected, for example for those only relating to customer contacts or unknown abnormal demands. The system detected engineered flushing events of known size (2 l/s) representing 6% of the maximum flow entering one DMA and 12% for another (Mounce *et al.* 2010b).

Testing on various measured data sets revealed the SVR detection technique is relatively insensitive to different choices of embedding dimension m (Equation (4)) and event window duration n. Hypothetically the SVR technique can detect novelty with high confidence, when well trained, in two data points (with 15 minute sampled data). Practically, however, there is a trade-off in parameter setting between detection time and number of alerts generated. For some temporal sequences, a Markov chain could be a better model than a binomial distribution and hence $p_{E_n}(|e_n(t_0)|)$ will be different. Exploration of optimal selection of parameters could also be useful rather than selection empirically.
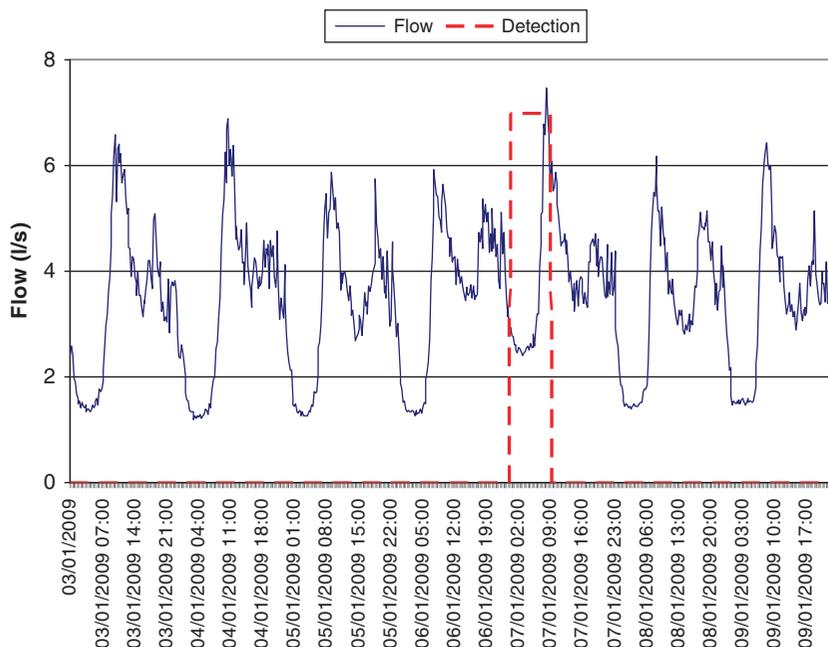

Figure 7 | Example "abnormal" detection ($+ 1.1$ l/s).

Further work on this system, as well as adaptation for online application, could explore automatic tuning of the system parameters, and evolutionary algorithms may lead to more optimal parameter selection. One issue for online implementation is that of system configuration changes requiring the need for the SVR system to be recalibrated. Rather than waiting for a full three-month training set to become available, it is suggested that models could be created with only several weeks data initially, although this would be flagged to indicate reduced confidence. Also, it would be interesting to explore the possibility of detecting gradually growing leaks. However, the challenge is understanding what the behaviour is in reality, such as a step change or more gradual change. A further future area of interest is combining novelty detection with pattern matching (Fletcher *et al.* 2008). A pattern matching approach is very proficient at finding known data patterns (hence classifying the type of novelty) and allowing the system to expand its knowledge base either autonomously or by supervised means.

Future operational applications will improve on the quasi real-time data that is currently being provided and allow data to be processed locally or remotely to provide exception alert generation or other information involving more than one signal. Benefits will accrue with increased coverage (with a mesh of flows and pressures enabling the location of incidents) but this will depend on progress with low-cost sensors and communications using peer-to-peer technologies or local hubs. The capture of such data will then need to be supported by data management (using more distributed middleware infrastructure such as DataTurbine (Tilak *et al.* 2007)) and receiving systems to make it accessible to downstream applications and actual use in the field through remote communications. This redundancy will provide additional knowledge. Visualization and aggregation of multiple data sources will be vital to support management of events and reduce their impact while improving the overall level of service. If this happens, opportunities exist to combine the data with other sensor information types (e.g. water quality) to support independent or related applications such as online network modelling and improved real-time response to customer issues (such as low pressures being related to a team working in the area).

CONCLUSIONS

This paper has presented the use of support vector machines for novelty detection for time series data analysis in WDS. A "bottom up", data-driven approach is suitable for such data, as often there is a lack of system knowledge (e.g. both with connectivity and in event information) and hands-on management at the DMA logger level is rarely a priority (such as configuration of settings of parameters and continuing maintenance). The primary focus of this paper has been on flow and pressure analysis. However, a generic methodology has been outlined which can be applied to any water domain parameter suitable for regression and for which novelty detection is a useful exercise.

The key conclusions of this work and directions for future work are as follows:

- The SVM-based approach presented can be used to achieve novelty detection from WDS time series data. Novelty can include a variety of events such as pipe bursts, hydrant flushing and sensor failure.
- The system was applied to a historical dataset consisting of five DMAs over a six-month testing period and has shown the ability to detect anomalies in flow and pressure patterns. Some 78% of detections were able to be correlated with operational information and manual data interpretation.
- It was demonstrated by example that the SVR methodology can provide faster alert generation than a previously developed ANN/FIS system.
- The system enables automated analysis, with limited manual setup, and the algorithmic settings were shown to be transferable and robust between DMAs and/or measurement parameter.
- The SVR technique thus shows potential for full online operation with a suitable scheme for data quality handling, training data selection and retraining schedule.

ACKNOWLEDGEMENTS

REFERENCES

Alegre, H., Hirner, W., Baptista, J. M. & Parena, R. 2000 *Performance Indicators for Water Supply Services*. IWA Manual of Best Practice, July 2000. IWA Publishing, London.

Akselaa, K., Akselab, M. & Vahalaa, R. 2009 Leakage detection in a real distribution network using a SOM. *Urban Water J.*, **6**(4), 279–289.

Bennett, K. 1999 Combining support vector and mathematical programming methods for induction. In *Advances in Kernel Methods – SV Learning* (ed. B. Schölkopf, C. Burges & A. Smola). MIT Press, Cambridge, MA, pp. 307–326.

Bicik, J., Kapelan, Z., Makropoulos, C. & Savic, D. A. 2010 Pipe burst diagnostics using evidence theory. *J. Hydroinformatics*. In press.

Bicik, J., Kapelan, Z. & Savic, D. A. 2009 Operational perspective of the impact of failures in water distribution systems. In *Proceedings of the World Environmental & Water Resources Congress 2009 (WDSA 2009), Kansas City, Missouri, 17–21 May, 2009*. ASCE, USA, pp 305–314.

Bishop, C. M. 2007 *Pattern Recognition and Machine Learning*. Springer, New York.

Boser, B. E., Guyon, I. & Vapnik, V. 1992 A training algorithm for optimal margin classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152. ACM Press. New York.

Branisavljevic, N., Kapelan, Z. & Prodsanovic, D. 2009 Online time data series pre-processing for the improved performance of anomaly detection methods. In *Integrating Water Systems* (ed. J. Boxall & C. Maksimovic). Taylor & Francis/CRC Press, London, pp. 99–103.

Chang, C. C. & Lin, C. J. 2001 *LIBSVM: a library for support vector machines*. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

Cortes, C. & Vapnik, V. 1995 Support-vector network. *Machine Learning*, **20**, 273–297.

Farley, B., Boxall, J. B. & Mounce, S. R. 2008 Optimal locations of pressure meter for burst detection. In *Proceedings of the 10th Water Distribution System Analysis Symposium*, South Africa, 17–20 August (ASCE online).

Farley, M., 2008. *Finding 'difficult' leaks*. International Water Association Specialist Group – Efficient Operation and Management. Available from: http://www.iwaom.org/datosbda/Descargas/54.pdf [Accessed 20th October 2010].

Fletcher, M., Liang, B., Smith, L., Knowles, A., Jackson, T., Jessop, M. & Austin, J. 2008 Neural network based pattern matching and spike detection tools and services – in the CARMEN Neuroinformatics Project. Special Issue on "Neuroinformatics". *Neural Networks*, **21**(8), 1076–1084.

Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Physical Review A*, **33**(2), 1134–1140.

Gardner, A. B., Krieger, A. M., Vachtsevanos, G. & Litt, B. 2006 One-class novelty detection for seizure analysis from intracranial EEG. *J. Machine Learning Res.*, **7**, 1025–1044.

Giustolisi, O. 2004 Sparse solution in training artificial neural networks. *J. Neurocomputing*, **56**, 285–304.

Giustolisi, O. 2006 Using multi-objective genetic algorithm for SVM construction. *J. Hydroinformatics*, **2**(8), 125–139.

Janković-Nisić, B., Maksimović, C., Butler, D. & Graham, N. 2004 Use of flow meters for managing water supply networks. *J. Water Resour., Plann. Manage.*, **130**, 171–179.

Jarrett, R., Robinson, G. & O'Halloran, R. 2006 On-line monitoring of water distribution systems: data processing and anomaly detection. In *Proceedings of the 8th Water Distribution System Analysis Symposium*, Cincinnati, 27–30 August. ASCE, USA, pp. 20061-21014.

Kecman, V. & Hadzic, I. 2000 Support vectors selection by linear programming. In *Proceedings of International Joint Conference on NN*, Como, Italy, vol. 5, pp. 193–198. IJCNN, Italy.

Ma, J. & Perkins, S. 2003 Online novelty detection on temporal sequences. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 613–618. ACM, New York.

McKenna, S., Hart, D., Klise, K., Cruz, V. & Wilson, M. 2007 Event detection from water quality time series. In *Proceedings of the World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat*, Tampa, FL, 15–19 May. ASCE, USA, pp. 1-12.

Mounce, S. R. 2005 *A hybrid neural network fuzzy rule-based system applied to leak detection in water pipeline distribution networks*. PhD thesis, University of Bradford, UK.

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

Mounce, S. R., Boxall, J. B. & Machell, J. 2007 An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation. In *Water Management Challenges in Global Change* (ed Ulanicki, B., Vairavamoorthy, K., Butler, D., Bounds, P. L. M. & Memon, F. A.). Taylor and Francis, AK Leiden, The Netherlands, pp 313–320.

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

Mounce, S. R., Farley, B., Mounce, R. B & Boxall, J. B. 2010b Field testing of optimal sensor placement and data analysis methodologies for burst detection and location in an urban water network. In *Proceedings of the 9th International Conference on Hydroinformatics*, Tianjin, China. Vol 2, pp. 1258-1265. Chemical Industry Press, Beijing, China.

Mukkamala, S., Janoski, G. & Sung, A. H. 2002 Intrusion detection using neural networks and support vector machines. In *Proceedings of IEEE International Joint Conference on Neural Networks*. IEEE Computer Society Press, USA pp. 1702–1707.

Romano, M., Kapelan, Z. & Savic, D. A. 2009 Bayesian-based online burst detection in water distribution systems. In *Integrating Water Systems* (ed. J. Boxall & C. Maksimovic). Taylor and Francis/CRC Press, London, pp. 331–337.

Schaffer, W. M., Olsen, L. F., Truty, G. L., Fulmer, S. L. & Graser, D. J. 1988 Periodic and chaotic dynamics in childhood infections.

In *From Chemical to Biological Organization* (ed. M. Markus, S. Müller & G. Nicolis). Springer, Berlin, pp. 331–347.

Schölkopf, B., Platt, J., Shawe-Taylor, J., Smola, A. J. & Williamson, R. C. 2001 Estimating the support of a high-dimensional distribution. *Neural Comput.*, **13**, 1443–1471.

Smola, A. J. & Schölkopf, B. 1998 *A tutorial on support vector regression*. NeuroCOLT Technical Report NC-TR-98-030, Royal Holloway College, University of London, UK.

Tarassenko, L., Hayton, P., Cerneaz, N. & Brady, M. 1995 Novelty detection for the identification of masses in mammograms. In *Proceedings of the Fourth International IEEE Conference on Artificial Neural Networks*, pp. 442–447.

Tilak, S., Hubbard, P., Miller, M. & Fountain, T. 2007 The Ring Buffer Network Bus (RBNB) DataTurbine streaming data middleware for environmental observing systems. In *Proceedings of the IEEE International Conference on e-Science and Grid Computing*, pp. 125–133.

Trafalis T. 1999 Primal–dual optimization methods in neural networks and support vector machines training. Tutorial Paper in *Support Vector Machines and Applications*, ACAI'99, Chania, Crete, Greece, 1999. ACAI, Greece.

Vapnik, V. 1998 *Statistical Learning Theory*. Wiley, New York.

Walski, T. M., Chase, D. V. & Savic, D. A. 2001 Chapter 9. *Haestad Methods: Water Distribution Modelling*. pp 333–396, Haestad Press, Waterbury, CT.

WHO 2001 *Leakage Management and Control – A Best Practice Training Manual*. Available online: http://www.who.int/docstore/water_sanitation_health/leakage/ch01.htm.

Wu, Z. Y., Sage, P. & Turtle, D. 2010 Pressure dependent leakage detection approach and its application to district water systems. *ASCE J. Water Resour. Plann. Manage.*, **136**(1), 116–128.

Ye, G. & Fenner, R. A. 2010 Kalman filtering of hydraulic measurements for burst detection in water distribution systems. *ASCE J. Pipeline Syst. Engg Pract.* In press.

Yu, X., Liong, S. Y. & Babovic, V. 2004 EC-SVM approach for real time hydrologic forecasting. *J. Hydroinformatics*, **6**(3), 209–223.