ANOMALY DETECTION BASED IN BAYESIAN INFERENCE

A method is described for determining the presence of anomalies in a water system, the method comprising the steps of (a) periodically sensing at least one water system parameter at least one sensing location, (b) transmitting data representative of the sensed parameter to a processing unit, (c) performing a checking and cleaning operation on the data received at the processing unit, (d) performing a noise removal operation to the data received at the processing unit, (e) predicting a subsequent data value for the water system parameter, (f) determining a discrepancy between the sensed data and the predicted data value using Statistical Process Control, (g) feeding the processed discrepancy into a Bayesian-based inference system to determine a probability of the data discrepancy representing an actual system anomaly, and (h) raising a detection alarm if the probability exceeds a predetermined threshold.

---

**Title**: ANOMALY DETECTION BASED IN BAYESIAN INFERENCE

**Abstract**: A method is described for determining the presence of anomalies in a water system, the method comprising the steps of (a) periodically sensing at least one water system parameter at least one sensing location, (b) transmitting data representative of the sensed parameter to a processing unit, (c) performing a checking and cleaning operation on the data received at the processing unit, (d) performing a noise removal operation to the data received at the processing unit, (e) predicting a subsequent data value for the water system parameter, (f) determining a discrepancy between the sensed data and the predicted data value using Statistical Process Control, (g) feeding the processed discrepancy into a Bayesian-based inference system to determine a probability of the data discrepancy representing an actual system anomaly, and (h) raising a detection alarm if the probability exceeds a predetermined threshold.
ANOMALY DETECTION BASED IN BAYSIAN INFERENCE

The loss of large volumes of treated water from water supply systems is environmentally and economically damaging especially if water is pumped to reach customers. The present situation is also characterised by accelerated population growth, rapid urbanization, depletion and pollution of aquifers and more extreme environmental fluctuations due to the climatic consequences of global warming drastically increase pressure on water resources. Consequently, water utility companies can no longer tolerate inefficiencies in water distribution systems and the resulting loss of revenue associated with water losses through leaks and bursts. Furthermore, indirect savings, such as reduced damage to building foundations and roadways, diminished water treatment and sewage loading as well as reduction in potential dangers such as reduced fire-fighting capabilities and contaminated water, must also be taken into account.

The performance of water distribution networks has come under increasing scrutiny in recent years due to the growing emphasis on sustainable business practice. In the UK privatisation has put the spotlight on the water industry, resulting in increased public accountability, new government controls, and higher customer expectations. The combination of these factors has led water service providers to consider water loss in treated water supply systems as a key water supply system performance indicator in order to achieve higher levels of operational efficiency and improved levels of service.

Although the water industry has been investing heavily in instrumentation and control systems, current practices for data management and for extraction of meaningful information to inform understanding of distribution network performance are inadequate and time consuming. This is a consequence of the reliance on human intervention for data analysis and interpretation, which is inefficient for the volume and complexity of the gathered data. Hence, technological developments and new methodologies are essential to seek the efficiency in the local distribution of water aiming at a reduction of losses. Of particular interest now is how to process the data gathered to provide meaningful, reliable and constantly relevant information.

A number of investigations have been carried out that attempt to link hydraulic models with online data from telemetry systems. Real-time transient model-based methods
simulate in real-time the unsteady flow in a pipeline using real-time measurements at the pipeline boundaries. Leak detection is performed using deviation analysis or model compensated volume balance methods. State estimation approaches estimate the current values of state variables (pressure, flow, nodal demands), given fixed network parameters. These techniques are vulnerable to modelling errors and require detailed information about pipeline and instrumentation. Transient-based techniques have been extensively explored for leak detection and location. These techniques attempt to extract information about the presence of a leak from the measured pressure transient, based on: time analysis of the leak travelling wave, frequency and wavelet analysis and inverse transient analysis. These techniques have been described in a number of papers and so are not described herein in further detail. Although these techniques demonstrate certain potential they often rely on complex and not so accurate transient network prediction models.

The above techniques also require a high number of measurement points and a high sampling frequency of measurement. However, the cost of sensors and data acquisition equipment required for monitoring a large number of branches in a network is financially infeasible. As a result, these techniques are not widely used in practice. Generally, the current practice is to place sensors that measure and record flow and pressure "(e.g., every 15-30 minutes)" on the DMA (District Metering Areas) boundaries and one or few pressure sensors inside the DMA (usually at the critical point).


The present invention relates to a method suitable for use in achieving awareness of system performance and response using automatic and self learning data mining and analysis. The new methodology performs data fusion and combines effectively event information from all the available flow and pressure sensors deployed in the analysed DMA. The developed methodology provides online detection and location at DMA level of leak/burst and anomalies in flow/pressure variation as they occur. It minimises the time prior to awareness and enables water companies to react within required time horizon, help save water, decrease the damaging consequences to infrastructure and improve customer service through minimising the inconvenience of interruption. Furthermore, due to the Bayesian Inference System employed, it provides probabilities associated with each detection together with the accurate estimate of leak/burst or anomaly induced, flow and pressure magnitude variations. These values provide useful information for filtering and ranking alerts in water distribution networks consisting of many DMAs and for prioritising responses.

According to the present invention there is provided a method for determining the presence of anomalies in a water system, the method comprising the steps of:
(a) periodically sensing at least one water system parameter at at least one sensing location;
(b) transmitting data representative of the sensed parameter to a processing unit;
(c) performing a checking and cleaning operation on the data received at the processing unit;
(d) performing a noise removal operation on the data received at the processing unit;
(e) predicting a subsequent data value for the water system parameter;
(f) determining a discrepancy between the sensed data and the predicted data value using Statistical Process Control;
(g) feeding the processed discrepancy into a Bayesian-based inference system to determine a probability of the data-discrepancy representing an actual system anomaly; and
(h) raising a detection alarm if the probability exceeds a predetermined threshold.

The noise removal step preferably comprises a wavelets based technique.

The method preferably further comprises a data-cleaning step to accommodate, for example, missing and/or corrupt data.

The sensed parameter preferably comprises at least one of water pressure and a flow rate signal at the sensing location. Conveniently the parameter is sensed at intervals of, for example, approximately 15 minutes. A plurality of sensing locations are preferably used, and it will be appreciated that the quantities of data that must be processed in accordance with the invention are very high.

The step of determining the discrepancy between the observed or measured data and the modelled data conveniently makes use of a Statistical Process Control method. In such a method, a short term pressure and/or flow prediction model is used first to determine an expected data value. The discrepancy obtained between this value and the actual observed value is further processed by the Statistical Process Control method to determine whether a difference between the modelled data value and the sensed data value is significant. The model is preferably trained periodically, for example on a weekly basis, using the sensed data. A long term prediction model, for example using data representative of night-time flow and pressure measurements may alternatively or additionally be used.

The invention further relates to a computer programmed to implement at least part of the method of the invention.

The invention will further be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a diagram illustrating a Bayesian based leak, burst or other anomaly detection system in accordance with an embodiment of the invention;
Figure 2 is a chart indicating the success rate of the Bayesian-based leak/burst and anomalies detection system for the threshold probabilities of 0.5, 0.6 and 0.7 and two analysed cases: (1) 1 flow and 2 pressure sensors and (2) 1 flow sensor only;

Figure 3 is a chart indicating the number of detected anomalies uncorrelated to WMS repairs for BIS threshold probabilities of 0.5, 0.6 and 0.7; and

Figure 4 is a graph showing the measured flow and pressure, BIS alarm, WMS repair and customer contact information for the analysed burst event.

The approach used herein is based on the simultaneous analysis of both flow and pressure sensor data from measurements taken at a sensing location. The assembled data sets are pre-processed by a processing unit first to deal with the missing data or data from faulty sensors. Once this is done, noise is removed from the flow and pressure signals using wavelets, for example using the technique described in Donoho & Johnstone 1995 [Donoho, D. L. & Johnstone, I. M. 1995. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90, 1200-1224]. Because wavelet analysis affords a different view of data to those presented by traditional techniques, it can de-noise a signal without compromising the non-stationary or transitory characteristics of the original signal.

The de-noised signals are then used as training data for the Group Method of Data Handling (GMDH) data-driven model, for example as described in Farlow 1984 [Farlow S. J. 1984. The GMDH algorithm, in: S. J. Farlow (Ed.), *Self-Organizing Methods in Modeling: GMDH Type Algorithms*. Marcel-Dekker, New York. 1-24] which predicts future flow and pressure profiles. It has to be underlined that the data-driven model is treated as continuously adaptive; this means that the model is trained for each logger when the system is initialised and thereafter at some predefined time period in order to continually capture the system's current conditions. The retraining interval is currently one month as this was found to be a good compromise between the computational expense and the system variability. This way the methodology is readily applicable in practice.

The data-driven model is then used to identify discrepancies between predicted and observed flow/pressure values. These discrepancies can be seen as indicators that
something in the system has suddenly (short-term) become unusual and warrant attention. Statistical Process Control (SPC) theory, such as that described in Shewhart 1931 [Shewhart, W. 1931. Economic control of quality of manufactured product. Van Nostrand Reinhold, New York] is used to further increase confidence in the hypothesis that an abnormal event has occurred. This is achieved by applying a set of control rules to the detected discrepancies in successive time steps. In parallel with this short-term flow/pressure based SPC analysis, SPCs control charts are used to analyse the (long-term) variations of flow and pressure based on the night trend (flow and pressure measured between midnight and five am). This time window is used in order to capture the Minimum Night Flow (MNF) that in urban situations normally occurs during the early morning period, usually between around two and four am, although the exact timing varies from zone to zone. By monitoring average night flows and pressure, cumulative background leakage can become apparent in flow and pressure data and unusual changes in water volumes and pressures can be detected.

Finally, evidence of the unusual changes in flow/pressure from the short-term SPC discrepancy based analysis and evidence of the unusual changes in night flow/pressure from the long-term SPC trend based analysis are fed to the Bayesian Inference System that is used for raising alarms. Figure 1 provides a diagrammatic representation of the components of the system showing how data flows through them.

Steps of the method of the invention are described in greater detail below:

**Wavelets de-noising**

Wavelet analysis is a tool for studying the properties of signals simultaneously in the frequency and time domains. The application of interest for this work is the use of wavelets for data analysis, specifically for signals de-noising. The de-noising procedure applied here involves three steps:

1. Perform a discrete wavelet transform (DWT) of the signal.
2. Select a threshold and apply it to the detail coefficients.
3. Perform inverse discrete wavelet transform (IDWT) using the original approximation coefficients and the modified detail coefficients, such as as described in Mallat 1989 [Mallat, S. 1989, A theory for multi-resolution signal

Assuming additive white Gaussian noise, any signal \( y(t) \) can be represented by the summation of the original \( x(t) \) and the noise \( n(t) \):

\[
y(t) = x(t) + n(t) \quad (1)
\]

where time \( t \) is equally spaced. The objective of this model is to remove noise by suppressing the noise part of the signal \( y \) and to recover \( x \). The usual approach to noise removal models noise as a high frequency signal added to the original signal. Fourier transforms could be used to identify this high frequency and remove it by adequate filtering. However, when the original signal has important information associated with the same frequency as the noise, filtering out this frequency is likely to induce a noticeable loss of information of the target signal. Since flow and pressure time series data from a water distribution system contain numerous non-stationary or transitory characteristics (i.e., drifts, seasonal trends, abrupt changes) the wavelets approach is more appropriate due to the fact that the signal is studied using a dual frequency-time representation, which allows noise to be removed from signal frequencies that are likely to contain important information. After applying the discrete wavelet transform a suitable strategy for noise removal would consists in making the coefficients associated to the noise equal to zero. A number of methods for de-noising differ in the way these coefficients are tracked and taken out from the representation. The conceptual details of the method applied here can be found in Donoho & Johnstone mentioned hereinbefore. Finally thresholds are applied and IDWT is performed.

**Short-term Pressure/Flow Prediction Model**

The Group Method of Data Handling (GMDH) technique is used here to build a short-term pressure/flow prediction model. The model is built as a network of polynomial functional elements, which is self-organised in layers to represent complex relationship between the output (future flow and/or pressure values one step, i.e., 15 minutes ahead in time) and the independent input variables (24 hrs of past flow and/or pressure values).
A detailed description of both the theory and the applications of GMDH can be found in Farlow mentioned hereinbefore. To avoid overfitting the model, but without the use of a test set, the Prediction Squared Error (PSE) criterion described in Barron 1984 [Barron, A. R. 1984. Predicted squared error: a criterion for automatic model selection. in: S. J. Farlow (Ed.), Self-Organizing Methods in Modeling: GMDH Type Algorithms. Marcel-Dekker, New York. 87-103] may be used.

Compared to regression analysis and neural networks, GMDH has two main advantages. The first is that it automatically builds optimal networks without requiring the form for the model relationship or the network architecture to be specified in advance. The second is that it uses all available data to train the network. The latter is particularly important given the imperfect nature of the data obtained from water distribution system sensors (i.e., large chunks of missing data, data from faulty loggers, etc.).

Statistical Process Control


The general idea behind SPC is that while every process displays variation, some processes display controlled variation that is natural to that process (common causes of variation), while other processes display uncontrolled variation that is not present in
the process causal system at all times (special causes of variation). A process that
features only common-cause variation is said to be in statistical control.

SPC refers to a number of different methods to detect any abnormality in the process.
In this work control rules are applied to the discrepancies between predicted and
observed flow/pressure values while control chart are used to detect unusual changes
in the monitored night flows and pressure.

A simple control rule tests if a measurement falls outside a confidence region that is
plus and minus a defined number of standard deviation limits from the predicted value
or

\[ x_t - N \sigma_l \leq x_{\text{obs}_{i,t}} \leq x_t + N \sigma_u \]

where \( x_{\text{obs}_{i,t}} \) is the measured value at time \( t \), \( x_t \) is the predicted value at time \( t \), \( \sigma_t \) is the
standard deviation of the errors in observations from the historic record and \( N_l \) and \( N_u \)
are user defined values denoting the acceptable lower and upper confidence bounds,
respectively. When a point falls outside these limits, the process is said to be out of
control. Here, the out of control implies that a leak is developing or a burst has
occurred. Whenever the process is stated to be out of control, there is a chance that
the statement is in error. Therefore, the occurrence of values in successive time steps
outside of the desired control limits can provide more confidence about the existence of
an abnormal event. Here, a subset of the Western Electric rules for detecting out of
control situations is used (Western Electric Company 1956) [Western Electric Company

As mentioned above, control charts are used in this work to analyse the long-term
variations of flow and pressure based on the night trend. A control chart is a graphical
representation of certain descriptive statistics for specific quantitative measurements of
the process. It is used to study how a process changes over time and is very useful to
track unusual variations and trends. A control chart always has a central line for the
average, an upper line for the upper control limit and a lower line for the lower control
limit. These lines are determined from historical data. By comparing current data to
these limits, conclusions can be drawn about whether the process variation is in control
or affected by special causes of variation (i.e., leaks, bursts).
Bayesian Inference System

The Bayesian Belief Network (BBN) is a directed graph, together with an associated set of probability tables. The graph consists of nodes and arcs. The nodes represent variables, which can be discrete or continuous. The arcs represent causal/influential relationships between variables. The key feature of BBNs is that they enable modelling and reasoning about uncertainty. The main use of BBNs is for statistical inference. Inference is the process of updating probabilities of outcomes based upon the relationships in the model and the evidence known about the situation at hand. As evidence accumulates, the degree of belief in an outcome ought to change. With enough evidence, it should become very high or very low. After inference, the updated probabilities reflect the new levels of belief in (or probabilities of) all possible outcomes coded in the model. The beliefs computed after evidence is entered are known as posterior probabilities, because they reflect the levels of belief computed in light of the new evidence.

In this invention a Bayesian inference system (BIS) is used to infer the probability of a leak/burst or an abnormal event given: (i) the evidence of the short-term unusual variations in the flow and pressure time series from the SPC discrepancy based analysis, and (ii) the evidence of unusual changes in night flow/pressure from the long-term SPC trend based analysis. The implemented BIS is designed in such a way that is able to process different type of evidence (i.e. discrepancy-base/night trend-based) and evidence from multiple and different sensors (i.e. flow/pressure) in a synergistic way. Furthermore, due to its temporal structure, BIS allows reasoning over time by updating probabilities of outcomes over consecutive time steps. Two past time steps are considered here resulting in a maximum of 30 minutes required to raise the alarm (assuming a 15-minute time step). Obviously, the more past time steps are taken into account, the higher the detection reliability but also the longer the detection time. At present, the above value is chosen empirically after a limited number of trials and it seems to be working well (see case study results below).

Because the BIS methodology synergistically combines all the information at hand about the current status of the system it is able to provide rapid and reliable anomaly detection. If events are detected based on a user definable minimum probability, then an alarm is generated giving the DMA location, classification probability, start time and
an estimate of the leak/burst flow rate and induced pressure drops. The marginal probability of an alarm given the evidence can be expressed in mathematical terms as 

\[ P(\text{Alarm}=\text{on} \mid \text{evidence}) \]. Furthermore, it is possible to compute the marginal distribution for every other node in the Bayesian network and the joint distribution for node groups.

In conclusion, the BIS not only performs detection and location, but generates an abnormal event classification probability value that can be used, together with the flow rate/pressure drop estimate, for filtering and ranking alarms in water distribution networks consisting of many DMAs and thus prioritise responses.

Case Study

The historical data analysis reported here is aimed at illustrating the capabilities of the developed Bayesian-based data analysis techniques for leak/burst and anomalies detection. Note that all the tests performed here were done in an online context where the existing historical data was used to simulate the incoming 'real-time' data. The BIS developed was applied without any prior knowledge of the system and was recalibrated every month, as it would have been done in a real-life application.

Flow and pressure data from a UK DMA recorded between 1/2/2008 and 31/12/2008 were used. The data consists of time-stamped files of 15 minute readings from the pressure and flow logger at the inlet as well as one pressure logger at the critical point inside the DMA. To evaluate the effectiveness of the developed methodology, the Work Management System (WMS) data containing the information about repairs carried out in the network, were used as an indication that a leak/burst occurred. However, based on this information, the exact date and time when a leak/burst commenced cannot be confirmed. The WMS's operations considered for this analysis are tagged as 'main repair' (MR) and 'service connection repair' (SE). Customer contacts tagged 'Pressure/flow', 'Burst/Leak' and 'No Water' were also used to correlate the raised alarms related to other anomalies not followed by a pipe repair.

For this DMA there were 7 recorded WMS repairs during the 11-month period. However, two of these were recorded when the measurement instruments were not working (and so no data was available), thus they were not taken into consideration in the following analysis. It must be stressed that this is the only situation when the developed methodology is not applicable, i.e., when all instruments are off line at the
same time. In the case when at least one instrument is working, the methodology will continue to perform, although less accurately, as will be shown later. The case study also demonstrates that the BIS is capable of using data from all 'available' sensors at a particular time thereby increasing detection reliability.

The ability of the developed methodology to detect the confirmed leaks/bursts (i.e., the ones classified either as the 'main repair' or the 'service connection repair' in the WMS) is shown in Figure 2. The figure presents a function of three BIS threshold probability values (0.5, 0.6 and 0.7) for two different sensor configurations (1 flow and 2 pressure sensors in case 1, and 1 flow sensor only in case 2). Results shown in Figure 2 clearly demonstrate a very high detection rate (up to 100%) when all sensors are used and the reduced performance when only data from a single flow meter is used. Therefore, inference performed by synergistically combining information from multiple and different type sensors provides better and more reliable anomaly detection. Figure 2 also emphasises the importance of choosing the right BIS threshold probability value.

Furthermore, the methodology also has the potential to detect other anomalies that affected the network of which no records are found in the WMS repair database. These types of events can be related to anomalies in flow/ pressure variations such as illegal or unexpected water usage and/or unusual system activity (e.g., network rezoning or changes to valve arrangements). Note also that the system will raise alarms in cases where a sensor starts drifting or when a sensor starts recording values that are outside its usual range (e.g., due to some fault), thus confirming that the developed methodology is an efficient tool to support sensor management decisions.

The BIS ability to detect events which were not originally recorded in the WMS is shown in Figure 3, again as a function of the threshold detection probability. Whenever possible, the alarms raised were correlated to the customer contacts tagged either as 'Pressure/flow', 'Burst/Leak' or 'No Water'. The remaining alarms, i.e. the alarms that could not be correlated to the customer contacts represent potential ghost alarms.

The term potential is deliberately used here to indicate that due to the lack of relevant information it is not possible to confirm these alarms either as ghosts or real alarms. As can be seen from Figure 3, the threshold probability of 0.5 leads to the identification of
8 events that could be correlated to the aforementioned customer contacts with further
5 alarms as potential ghosts. As the values of threshold probability increases, the
number of events correlated to the customer contacts decreases but so does the
number of potential ghosts. This confirms that the BIS threshold probability value
needs to be chosen carefully.

One of the main advantages of the BIS based methodology is its fast detection time. In
fact, all alarms shown here are raised within the maximum of 30 minutes (i.e., two time
steps) after an anomaly occurred. In order to demonstrate this important feature of the
methodology, an example of a complex event is shown in Figure 4. The sequence of
events illustrates a situation in which the data analysis system correctly produced an
alarm for a burst which was confirmed later on by the manual inspection and recorded
in the WMS system.

The threshold probability used to raise an alarm is set equal to 0.5. Table 1 reports the
values of the classification probability at some chosen time and the corresponding
estimates of flow and pressure magnitude variation induced by the burst.

Table 1. Burst event. Estimates of flow and pressure magnitude variations induced by
the burst; date-time and probability of the generated alarms.

<table>
<thead>
<tr>
<th>Date/Time</th>
<th>Estimated event magnitude</th>
<th>Detection Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pressure1 [m]</td>
<td>Flow [l/s]</td>
</tr>
<tr>
<td>4/2/2008 11:45</td>
<td>0.45</td>
<td>3.9</td>
</tr>
<tr>
<td>7/2/2008 20:45</td>
<td>0.38</td>
<td>9.4</td>
</tr>
<tr>
<td>18/2/2008 11:45</td>
<td>0.5</td>
<td>10</td>
</tr>
</tbody>
</table>

As shown in Figure 4, four customer contacts tagged 'Burst/Leak' were recorded during
the analysed time period. The 'main repair' started on the 20th of February. As can be
seen from Figure 4, the water company received the first customer contact more than
29 hours after the alarm was raised (dark highlighted area) while the main repair was carried out only 16 days later (light highlighted area), probably, only after the second customer contact was received. It is important to observe that the probability of alarm increases with time. As shown in Table 1, the alarm probability calculated by the BIS is 0.53 on the 4th when small variations in the flow rate triggered the system to raise the alarm. This value increases to 0.63 on the 7th reflecting the increasing flow rates induced by the burst. Finally, the same value reaches 0.93 on the 18th when a large anomaly in the second pressure sensor occurs. This example demonstrates how a burst, even at its very early development stage, can be timely identified and repaired before it causes a major impact. It is expected that by minimising the time to detection, water companies will be able to: (i) react quickly to reduce the water lost, (ii) decrease the potential damage to the infrastructure and to third parties, and (iii) to improve the customer service (by minimising the interruption inconvenience).

Real-time monitoring of water distribution systems thus provides useful information for detecting potential leaks, bursts and other anomalies. This then requires a methodology capable of providing timely warnings required to mitigate the negative impacts of these events. The invention provides a proactive, automatic and self-learning data analysis system for the real-time location and detection of leaks/bursts/anomalies at DMA level. The system also enables the estimation of related flow rates and pressure variations generated by the above anomalies. These values together with alarms probabilities can be effectively used to prioritise events and responses.

The Bayesian-based system developed has been applied to the analysis of flow and pressure data in a UK DMA. The chosen DMA has deployed instrumentation that reflects the current best practice in the UK industry and has shown the ability to detect small to very large leaks/bursts/anomalies rapidly (within 30 minutes) and reliably.

Furthermore, based on the operator's experience and decision about the most suitable threshold choice, it has the potential to identify small precursor features that are important for monitoring the cumulative effect or development of leaks. The methodology also has the potential to advance the state-of-the-art in real-time water distribution system management and is readily transferable to practice.
Finally, note that the current values of BIS parameters (i.e., the conditional probability tables) have been determined experimentally. This can be improved further by using the past recorded and verified burst events to automatically calibrate these parameters, thus leading to a more reliable detection of the DMA bursts.

It will be appreciated that a number of modifications and alterations may be made to the methodology described hereinbefore without departing from the scope of the invention.
CLAIMS:

1. A method for determining the presence of anomalies in a water system, the method comprising the steps of:
   (a) periodically sensing at least one water system parameter at at least one sensing location;
   (b) transmitting data representative of the sensed parameter to a processing unit;
   (c) performing a checking and cleaning operation on the data received at the processing unit;
   (d) performing a noise removal operation to the data received at the processing unit;
   (e) predicting a subsequent data value for the water system parameter;
   (f) determining a discrepancy between the sensed data and the predicted data value using Statistical Process Control;
   (g) feeding the processed discrepancy into a Bayesian-based inference system to determine a probability of the data discrepancy representing an actual system anomaly; and
   (h) raising a detection alarm if the probability exceeds a predetermined threshold.

2. A method according to Claim 1, further comprising a step of requesting a maintenance action in the event that the probability exceeds a predetermined level.

3. A method according to Claim 1 or Claim 2, wherein the noise removal step comprises a wavelets based technique.

4. A method according to any of the preceding claims, wherein the sensed parameter comprises at least one of water pressure and flow rate at the sensing location.

5. A method according to any of the preceding claims, wherein the parameter is sensed at intervals of approximately 15 minutes.

6. A method according to any of the preceding claims, wherein a plurality of sensing locations are used.
7. A method according to any of the preceding claims, wherein a short term pressure and/or flow prediction model is used to determine an expected data value, and the discrepancy between the observed data value and the predicted data value is processed by the Statistical Process Control method to determine whether a difference between the modelled data value and the sensed data value is significant.

8. A method according to Claim 7, where the model is trained periodically using the sensed data.

9. A method according to Claim 7 or Claim 8, wherein a long term prediction model is provided.

10. A method according to Claim 9, wherein the long term prediction model uses data representative of off-peak measurements.

11. A method according to Claim 10, wherein the off-peak measurements comprises night time measurements.

12. A computer programmed to implement at least part of the method of the preceding claims.
Figure 2

- Detected (WMS's records)
- Undetected (WMS's records)
Figure 3

Figure 4
### A. CLASSIFICATION OF SUBJECT MATTER

<table>
<thead>
<tr>
<th>INV.</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>G06Q10/00</td>
<td><strong>G06Q10/00</strong></td>
</tr>
<tr>
<td>G06Q50/00</td>
<td><strong>G06Q50/00</strong></td>
</tr>
<tr>
<td>G01M3/26</td>
<td><strong>G01M3/26</strong></td>
</tr>
</tbody>
</table>

According to International Patent Classification (IPC) or to both national classification and IPC.

### B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

<table>
<thead>
<tr>
<th>Field</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>G06Q</td>
<td><strong>G06Q</strong></td>
</tr>
<tr>
<td>G01M</td>
<td><strong>G01M</strong></td>
</tr>
</tbody>
</table>

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched.

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**EPO-Internal**

### C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
</table>

Further documents are listed in the continuation of Box C

See patent family annex

- **A** document defining the general state of the art which is not considered to be of particular relevance
- **E** earlier document but published on or after the international filing date
- **L** document which may throw doubts on novelty claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- **O** document referring to an oral disclosure, use, exhibition or other means
- **P** document published prior to the international filing date but later than the priority date claimed
- **T** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- **X** document of particular relevance, the claimed invention cannot be considered to be obvious to the skilled person when the invention is taken alone from the document
- **Y** document of particular relevance, the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- **F** document member of the same patent family

Date of the actual completion of the international search: 4 August 2010

Date of mailing of the international search report: 10/09/2010

Name and mailing address of the ISA/

**European Patent Office, P B 5818 Patentlaan 2 NL - 2280 HV RUISWALD Tel (+31-70) 340-2040, Fax (+31-70) 340-3016**

**Authorized officer**

Hopper, Eva
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>J.C. Rougier and M. Goldstein: &quot;A Bayesian Analysis of Fluid Flow in Pipelines&quot;</td>
<td>1-12</td>
</tr>
<tr>
<td>X</td>
<td>Mary Lam: &quot;Benchmark of Probabilistic Methods for Fault Diagnosis&quot;</td>
<td>1-12</td>
</tr>
<tr>
<td>A</td>
<td>J.C. Rougier: &quot;Formal Bayes Methods for Model Calibration with Uncertainty&quot;</td>
<td>1-12</td>
</tr>
<tr>
<td></td>
<td>K. Beven and J. Hall (eds), Applied Uncertainty Analysis for Flood Risk</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Retrieved from the Internet: URL <a href="http://www.maths.bris.ac.uk/~mazjc/FRMbox2-4.pdf%3E">http://www.maths.bris.ac.uk/~mazjc/FRMbox2-4.pdf&gt;</a> [retrieved on 2010-08-03] the whole document</td>
<td></td>
</tr>
<tr>
<td></td>
<td>abstract paragraphs [0052], [0053], [0101], [0107], [0108], [0110]; claim 7</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>WO 2009/048340 A2 (TECWEL AS [NO]; LIE TERJE LENNART [NO]; ANDERSEN GUNNAR [NO]; INSTANES) 16 April 2009 (2009-04-16) abstract-</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>page 2, line 22 - line 33 page 14, line 6 - line 19</td>
<td></td>
</tr>
<tr>
<td></td>
<td>abstract column 6, line 19 - column 7, line 46</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Citation of document, with indication, where appropriate, of the relevant passages</td>
<td>Relevant to claim No.</td>
</tr>
<tr>
<td>----------</td>
<td>---------------------------------------------------------------------------------</td>
<td>----------------------</td>
</tr>
</tbody>
</table>
## INFORMATION ON PATENT FAMILY MEMBERS

<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 2003009239 A1</td>
<td>09-01-2003</td>
<td>NONE</td>
<td></td>
</tr>
<tr>
<td>US 6526358 B1</td>
<td>25-02-2003</td>
<td>NONE</td>
<td></td>
</tr>
</tbody>
</table>