A new approach to improve water loss control using smart metering data
D. Loureiro, H. Alegre, S. T. Coelho, A. Martins and A. Mamade

ABSTRACT
The control of water losses is a major concern in the sustainability of urban water utilities and in promoting the efficient use of this natural resource. Recent advances in telemetry technology provide high-resolution consumption data at the consumer level, allowing for a remarkable knowledge improvement on the different water balance components. However, few studies focus on systematic approaches for improving system operation and maintenance by processing and analysing large amounts of consumption data. This paper presents a new methodology to calculate real losses and apparent losses in distribution networks using data collected from telemetry systems. The methodology proposes a set of algorithms that are simple to implement. These algorithms were tested on different district metered areas (DMA) to improve understanding about water loss components and have already been included in commercial software. The results showed that these algorithms are robust and allow for accurately estimating the background leakage level (and unreported leaks and bursts), detecting earlier the occurrence of bursts and providing important insights into the type of illegal water uses. The use of these approaches reduced non-revenue water by more than 10% in the majority of the DMA tested. These findings are promising and demonstrate the strong potential of telemetry systems to reduce water losses and to improve the understanding of water uses.

Key words | apparent losses, real losses, smart metering, water consumption, water losses

INTRODUCTION
The control of water losses plays a major role in the sustainability of urban utilities and in promoting efficient use of this natural resource. Water losses, together with unbilled authorized consumption, make up the amount of non-revenue water in a distribution system (Farley & Trow 2003). Apart from causing significant economic, environmental and social costs, water losses increase the carbon footprint of the utility due to the water-energy nexus (Cohen et al. 2004).

Real water losses include leakage from pipes, service connections, joints and fittings, and leakage and overflows at storage tanks. Apparent losses include unauthorized consumption and metering errors (Lambert & Hirner 2000; Alegre et al. 2006). According to Farley & Trow (2003), the components of real losses are grouped into: background leakage, unreported leaks and bursts and reported leaks and bursts. Background leakage is the aggregation of sources of water loss from all fittings on the network that are too small to be detected; unreported leaks and bursts correspond to events associated with moderate flow rates and variable duration; and reported leaks and bursts correspond to events associated with a sudden flow increase and short duration. Average duration of reported leaks and bursts should take into consideration three components: awareness, location and repair time (Lambert & Morrison 1996).

Calculating water balance is the most common top-down approach for assessing the components of water losses. For a more successful control of losses, calculation of water balance should be carried out in watertight distribution network sectors, such as district metered areas (DMA). These are sectors with about 1,000–3,000 service connections, equipped to measure all inflows and outflows.
(Farley & Trow 2003). This method requires the measurement or estimation of water produced, imported, exported, consumed and lost, and typically covers a 12-month retrospective period. This low frequency of analysis is due to the fact that data at the consumer level are usually sparse, not continuous (i.e., available at monthly or lower frequencies) and typically rely on estimates of unmetered water uses. In contrast, continuous records of water produced, imported and exported are usually available. Recent advances in telemetry technology enable the acquisition of high-resolution consumption data (i.e., at less than 1 hour time steps) at the consumer level. With these type of data it is possible to reduce significantly the uncertainties associated with the different components of water balance, avoid using estimates, and detect abnormal consumption and unreported leaks and bursts more actively.

Bottom-up real loss assessment can be carried out using the Minimum Night Flow (MNF) analysis (García et al. 2006; Puust et al. 2010). MNF analysis can be considered the second part of the audit process using high resolution network flow data. This analysis is usually carried out during the early morning period and the estimation of background leakage requires subtracting legitimate night uses (i.e., exceptional uses, household uses and non-household) from the MNF (UK Water Industry 1994). Any significant difference between estimates of background leakage may be due to unreported leaks and bursts. Due to the inexistence of high resolution data for customer consumption, general simplified rules were established to estimate legitimate night uses (UK Water Industry 1994; Warren 2002). However, estimates of legitimate night uses can improve significantly with the use of customer telemetry systems.

Notwithstanding, telemetry systems generate large amounts of data and their processing is still challenging. Few studies focus on systematic approaches to process (i.e., validate, normalize, detect, clean outliers and combine) and analyse (i.e., pipe burst detection, pattern analysis) consumption data collected from telemetry systems with the purpose of improving network operation and maintenance (Loureiro 2010; IBM 2011; Urkullu et al. 2012).

This paper presents a new methodology to calculate the components of real losses (i.e., background leakage, unreported leaks and bursts and reported leaks and bursts) and apparent losses in distribution networks using data collected from telemetry systems. Although the study is based on well-established concepts of water balance calculation (Lambert & Hirner 2000), the methodology represents, relative to previous studies (IBM 2011; Urkullu et al. 2012), a step forward to control water losses using telemetry data. The methodology proposes a set of algorithms that are simple to implement to estimate water loss components. These algorithms were tested on different DMA to improve understanding about water loss components, and are already included in commercial software.

**METHODOLOGY**

To convert large amounts of consumption data collected from smart metering into easily understandable information, solid approaches for data handling should be adopted as depicted in Figure 1. After data collection, processing and storage, analytical methods may be applied to improve the knowledge of the different water loss components and operation management.

Water loss analysis at DMA level requires the collection of inflow/outflow data from existing flow meters, and consumption data from household and non-household water meters.

Total metered consumption is calculated combining telemetry data collected from the different water meters. Prior to combination, data provided by each water meter should be validated (e.g., consumption data above maximum or minimum values, reverse time series order, etc.), normalized to a common and regular time step, outliers should be detected and removed and missing data should be reconstructed.

Data reconstruction is due to the fact that gaps – time intervals without any records – may occur as a result of meter malfunction or issues with the acquisition and communication systems. The occurrence of gaps will disable the calculation of total metered consumption in those time intervals. Loureiro et al. (2012) have tested different methods for the reconstruction of missing data in the total metered consumption: (1) based on consumption statistics of each consumer, (2) based on consumption statistics of groups of consumers, and (3) based on consumption statistics of the total number of consumers. The authors have found that methods (1) and (2) were more accurate than method (3).
for predicting missing data. In this paper, method (1) was adopted for data reconstruction. Although this method may affect the computing performance if a large number of consumers are connected to smart meters (thousands of consumers), it is very simple to implement. Thus, periodically (e.g., once per month), hourly consumption statistics based on data from the last 3 months were calculated for each consumer and stored per weekday. Therefore, at every instant, and before data combination, consumers with gaps were identified and missing data were reconstructed based on the addition of the respective stored values of hourly median consumption at the same instant and weekday.

Once the total metered consumption without gaps was calculated, this component was subtracted from the total inflow in a DMA to obtain non-revenue water (assuming that billed unmetered and unbilled metered consumption are negligible). The calculation of these components on a daily basis can be very useful to understand consumption trends throughout the year (e.g., seasonal effects, the occurrence of pipe burst leakage or anomalous water uses). Nevertheless, to assess real losses and separate this component from apparent losses, it is essential to have at least hourly consumption data.

Regarding the assessment of real losses, the first step consisted in calculating background leakage and unreported leaks and bursts. This calculation was based on the value of non-revenue water at the instant of MNF (García et al. 2006; Puust et al. 2010). In this study, it was assumed that the component of apparent losses is insignificant at the instant of MNF and that real losses are constant throughout the day. This last assumption is acceptable since most of the pipelines in distribution DMA are oversized for regular, human consumption, due to fire-fighting requirements, and the pressure variation between night and day can be neglected. The occurrence of pipe bursts or anomalous non-authorized consumptions during the night period will occasionally overestimate the value of background leakage and unreported leaks and bursts.

Although the methodology to calculate background leakage and unreported leaks and bursts is based on the concept of MNF analysis, the approach used in this paper is new. In order to calculate accurately the daily value of background leakage and unreported leaks and bursts, it applies concepts of outlier region (Davies & Gather 1993) and uses robust statistics to avoid the influence of anomalous observations (Huber 2004). Thus, to detect these outliers in the estimation of daily background leakage, each value was compared with a pre-defined upper threshold value ($\tau$). This threshold value was defined based on robust statistics suggested by Menold et al. (1999) (median and median absolute deviation (MAD)) of background leakage and unreported leaks and bursts of the last 10 days and on the 90% percentile of the normal distribution, see Equation (1).

$$\text{Threshold value} = \text{Median} + \tau \cdot \text{MAD}$$ (1)
Therefore, each value of background leakage and unreported leaks and bursts was compared with this upper threshold value. If higher, it was labelled as an outlier and replaced by the respective median value of the last 10 days.

The proposed methodology also allows detection of events associated with reported leaks and bursts. After removing background leakage (and unreported leaks and bursts) from non-revenue water, a new algorithm, based on the identification of consumption periods whose initial and final instants cause a significant flow variation was proposed and tested. Therefore, a median filter (Schettlinger et al. 2010) was applied to a small window of previous observations to smooth the water loss time series (without the background leakage and unreported leaks and bursts). This smooth time series removes small variations from the non-revenue water time series and provides an expeditious way to detect events with high flow rates and short duration. The identification of instants with significant flow variation was carried out according to Equation (2).

$$\Delta_i = \frac{M_i - M_{i-1}}{Xm_{50}}$$  \hspace{1cm} (2)

where $\Delta_i$ is the variation in the $i$ instant, $M_i$ and $M_{i-1}$ are the median values in the $i$ and $i-1$ instants and $Xm_{50}$ is the respective monthly median value. The threshold value $\Delta_i > 2.5$ marked the start of a new event ($E$). The event $E$ was characterized by the median in the instant $i-1$ ($M_{i-1}$) and only ended when the initial situation was re-established, $M_i \leq E + \epsilon$, where $\epsilon = 0.1$. After identifying these events, they were classified in terms of duration and volume. An event was classified as a reported leak and burst when the duration was longer than 6 hours. This rule of thumb was based on the assumption that the duration of reported bursts always includes some time to perceive, locate and repair, which may vary between some hours to several days (Lambert & Morrison 1996). In addition, this minimum duration was adopted to distinguish appropriately these types of events from significant consumption events with short duration.

After removing these events, the remaining component of non-revenue water can essentially be associated with apparent losses. With the availability of high frequency data about this component (1 hour time step) it is possible to perform comparative analyses between apparent losses and total metered consumption in terms of daily patterns. The identification of similarities or differences between these patterns enables a more accurate detection of some types of illegal water uses. A daily demand pattern aims at describing the typical consumption variation throughout the day (Coelho 1988; Loureiro 2010). Each time step can be characterized by a set of robust statistics using a representative sample of historical data (e.g., 3 months).

### CASE STUDIES

The proposed methodology to separate background leakage and unreported leaks and bursts from reported leaks and bursts and to identify unauthorized consumption was tested on five DMA. These small sectors were selected for this study because they corresponded to networks where the existing flow meters and all the water meters were already equipped with telemetry (Table 1). Data from these DMA were used to test the algorithms that were afterwards

<table>
<thead>
<tr>
<th>DMA</th>
<th>District</th>
<th>Total number of customers</th>
<th>Flow meters</th>
<th>Water meters</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMA 1</td>
<td>Faro (Southern)</td>
<td>61</td>
<td>Single jet turbine</td>
<td>Volumetric</td>
</tr>
<tr>
<td>DMA 2</td>
<td>V. do Castelo (Northern)</td>
<td>378</td>
<td>Single jet turbine</td>
<td>Volumetric</td>
</tr>
<tr>
<td>DMA 3</td>
<td>V. do Castelo (Northern)</td>
<td>322</td>
<td>Single jet turbine</td>
<td>Volumetric</td>
</tr>
<tr>
<td>DMA 4</td>
<td>V. do Castelo (Northern)</td>
<td>870</td>
<td>Single jet turbine</td>
<td>Volumetric</td>
</tr>
<tr>
<td>DMA 5</td>
<td>Oporto (Northern)</td>
<td>162</td>
<td>Woltmann horizontal</td>
<td>Volumetric</td>
</tr>
</tbody>
</table>
implemented in commercial software. Consumption data with 1 hour time intervals were collected between January 2012 and February 2013, and transmitted daily using radio/GSM communications to a central server. Water meters were volumetric (Aquadis+ Model, class C and R 160), with nominal diameters of 15 mm predominantly \( Q_3 = 2.5 \text{ m}^3/\text{h} \) and \( Q_n = 1.5 \text{ m}^3/\text{h} \) and 20 mm \( Q_3 = 4 \text{ m}^3/\text{h} \). Network flow meters were single jet turbine (Flostar M Model, Class C, from R 250 to R 315) or Woltmann horizontal (Woltex Model, Class B) (Table 1).

**RESULTS**

**Non-revenue water**

Before separating water losses into different components, the daily variations of total inflow, total metered consumption and non-revenue water were calculated (Figure 2). Non-revenue water shows a different behaviour from total metered consumption and is characterized by smooth variation (average value 19 m\(^3\)/day), with occasional high consumption values (maximum value 76 m\(^3\)/day). In contrast, metered consumption has a significant tendency to increase between April and May. This increase may be due to garden watering and the beginning of the summer holiday period in this region. Clearly, these findings indicate that the availability of telemetry data with a daily frequency provides valuable information concerning metered consumption trends and the non-revenue water baseline, and allows an early detection of anomalous events (e.g., pipe bursts, unauthorized uses).

**Background leakage and unreported leaks and bursts**

The estimation of background leakage and unreported leaks and bursts was based on the value of non-revenue water at the instant of minimum night flow, and it was assumed to be constant throughout the day (Figure 3(a)) (average value 0.49 m\(^3\)/h). To detect possible outliers, each value was compared to an upper threshold value calculated according to Equation (1). If higher, it was labelled as an outlier and replaced by the respective median value of the last 10 days (Figure 3(b)). Detected outliers may be due to possible pipe bursts or unauthorized billed consumption. After removing outliers, background leakage and unreported leaks and bursts are almost constant in time and have a lower value (average value 0.17 m\(^3\)/h). These results emphasize the importance of approaches to assess this component of real losses accurately when data with high frequency (1 hour time step) are available and to confirm the smooth variation of the real losses component, as previously mentioned (UK Water Industry 1994).

**Reported leaks and bursts**

The identification of a possible pipe burst event is illustrated in Figure 4. After removing background leakage and unreported leaks and bursts from non-revenue water, the identification of the initial and final instants that cause a significant variation was carried out using a smooth time series. Smooth time series have the advantage of removing small variations from the non-revenue water time series (Figure 4), making the identification of possible pipe bursts easier. The event labelled as a pipe burst had a 1.96 day duration, an average flow of 1.14 m\(^3\)/h and a volume of 53.6 m\(^3\). Results have shown that these events vary significantly in duration, average flow and volume, as pointed out by Lambert & Morrison (1996), but are generally characterized by a significant initial flow increase, followed by a period with high flow and small variation. These findings should be explored in an extensive number of DMA.

![Figure 2](https://iwaponline.com/ws/article-pdf/14/4/618/415684/618.pdf) | Daily variation of total inflow, total metered consumption and non-revenue water in DMA 1.
Apparent water losses

After removing reported leaks and bursts, the remaining component of non-revenue water is mainly associated with apparent losses (unauthorized consumption and metering errors). Daily patterns for apparent losses, real losses and total metered consumption are compared for DMA 3 and 4 (Figure 5). In DMA 3, apparent losses are negligible when compared to the total inflow (0.7%), whereas in DMA 4 they represent a significant proportion (15%). In addition, the fact that daily patterns for apparent losses and total consumption in DMA 4 are quite similar, indicates that most of the illegal uses may be of the same type of authorized consumption (residential, in DMA 4). This straightforward approach is very important to support the utility in the definition of the best strategy to control illegal water uses. An important question for future studies is how to improve the knowledge on metering errors, the other component of apparent losses, using this type of data.

Validation of proposed algorithms

The algorithms presented in the previous sections were applied in commercial software (www.itron.com), followed by its test and validation using the support of two utilities during 3 months (i.e., from December 2012 to February 2013). These methods allow accurate detection of pipe bursts, authorized unmetered consumption and illegal uses. For example, the occurrence of a reported pipe burst (Figure 6(a)) between 4th and 5th December (duration and volume equal to 21 hours and 17 m$^3$, respectively) and illegal water use on 10th January (Figure 6(b)) were confirmed by the utility. During these tests, the water utilities have found DMA where non-revenue water exceeded 20% (DMA 3, DMA4, DMA 5), and implemented measures that allowed a reduction of more than 10% of non-revenue water (pipe bursts repair, illegal water uses). DMA 1 was not used in these tests. These results represent a step forward in the control of water losses using data from telemetry systems, since most of these components of losses are usually estimated as a percentage of total inflow (Farley & Trow 2003) and are based on scarce information.
CONCLUSIONS

Few studies focus on processing and analysing large amounts of telemetry consumption data to improve network operation and maintenance in water distribution systems. Although this study is based on well-established concepts of water balance calculation, the methodology proposed is new. It represents a step forward in comparison to previous studies where telemetry data were used to control water losses. It includes new algorithms to calculate real loss components (i.e., background leakage, unreported leaks and bursts and reported leaks and bursts) and apparent losses. To calculate the daily value of background leakage and unreported leaks and bursts accurately, concepts of outlier region using robust statistics were used. The proposed methodology also allows detection of events associated with reported leaks and bursts. A median filter, to smooth the water loss time series and detect these events, was applied. The characterization of apparent water losses was carried out in terms of daily patterns to allow comparison with the daily pattern of metered consumption. This comparison constitutes new and valuable information and enables the water utility to make some inferences about the type of water uses related with unauthorized consumption.

The proposed algorithms were implemented in commercial software and tested under operational conditions on a variety of distribution systems. Results have shown that the algorithms are robust and allow an accurate calculation of water loss components. These findings are promising and the algorithms and technology will be explored in an extensive number of DMA. In addition, they represent a step forward in the control of water losses, since most of these water loss components are traditionally estimated as a

---

**Figure 5** | Daily patterns of apparent losses, real losses and total metered consumption in (a) DMA 3 and (b) DMA 4 (based on the calculation of the median value in each time step).

**Figure 6** | Detection of anomalous consumption occurrences in DMA 5: (a) reported pipe burst, (b) detection of illegal consumption.
percentage of total inflow and based on scarce information. The use of these results as reference values in network areas with similar characteristics, but without smart metering, may provide important insights when carrying out standard water balances. Results also highlight the significant potential of smart metering to improve network operation, in addition to billing purposes.

ACKNOWLEDGEMENTS

This work was developed within the scope of project WIMS – Water Intelligent Management System. The authors would like to thank the project coordinators – Itron, for turning the results of this research into de facto market products, and their partnering utilities Indáqua Matosinhos and SMSB Viana do Castelo for testing and exploring the results under operational conditions on different distribution systems. The approaches explored in this work were partially tested and improved under the European Project iWIDGET and this research has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 318272.

REFERENCES


IBM 2011 Smart Water Pilot Study Report. IBM Research, Dubuque.


First received 11 December 2013; accepted in revised form 18 February 2014. Available online 7 March 2014