Source of Uncertainty in Water Supply Pipeline Leak Detection Using District Meter Area Data

Thewodros G. Mamo¹, Ilan Juran²

Department of Civil & Urban Engineering Polytechnic Institute of New York University¹, ²
6 Metro Tech Center, Brooklyn, NY 11202

Abstract: The use of multi parameter sensors for the day to day operational and management activities become an essential component of leading water utility companies in the effort to modernize their water distribution pipeline networks systems. With the widespread deployment of multi-parameter sensors to monitor water distribution pipeline networks operational activities, allows vast amount of data to be collected, analysed, and acted upon in the shortest periods of time. These multi-parameter sensors not only respond to the change of operational pattern to produce data, they also embed with computing and communication capabilities. These systems are able to store, process locally and transfer data they produce to the water utility companies’ main database. During these processes there are strong possibilities that data tends to become uncertain. These uncertainties can be originated from different components of multi-parameter sensors used in DMA, such as SCADA, AMR, AMI, data collection error, measurement precision limitation, data sampling error, outdated source, data acquisition and transmission error…etc. Multi-parameter sensors used inside DMA like any other devices are subject to wear and tear, or mal functioned or system failure that leads to inaccuracy with time. Therefore the primary goal of this paper is to transfer knowledge among water utility professionals, by highlighting the potential sources and types of uncertainty DMA data used for water distribution system (WDS) pipeline leak detection, and address them using an appropriate uncertainty analysis tools to determine a more accurate and reliable result by increasing DMA data quality.

Keywords: Water Supply Data Bases, SCADA, AMR, AMI, DMA, Multi-Parameter Sensors; Sources of Uncertainty, Uncertainty Quantification, Uncertainty Calibration, Uncertainty Reduction

1. Introduction

A range of techniques for quantifying and reducing uncertainty have been developed. Most of common and widely used approaches have focused on methodologies that quantifying and reducing parameter uncertainty [1] [4] [8] [13] [15]. methodologies including parameter optimization procedures such as generalized likelihood uncertainty estimation (GLUE) or probabilistic approaches, such as genetic algorithms (GA) or Markov Chain Monte Carlo (MCMC) may be best applied where data are available for model calibration and evaluation[31][44]. In the case of data are not available, limited, or restricted to expert opinion, or there are uncertainty regarding the possibility of future events, possibility theory and Evidence theory may form more appropriate frameworks for representing uncertainty. Evidence theory forms a more appropriate framework for combining different sources and types of data to reduce system uncertainty [6] [8] [16] [21] [24]. Recent advance findings, such as total error analysis and implicit uncertainty methodologies have helped not only parameter uncertainty within probabilistic approaches but also towards accounting input uncertainty, model structural uncertainty, and output evaluation data uncertainty. These recent advances, however, require more data to constrain and understand the effect of different sources of uncertainty on model performance [41] [45].

2. Sources of Uncertainty during Pipeline Leak Detection Using DMA Data

Various sources of uncertainty during analysis of pipeline Leak detection using District Meter Area (DMA) can arise, but in general those errors can be categorized as random error due to measurement; Model error due to uncertain parameters; and error due to data acquisition and transmission. [24][39][44]. below figure 2.1 shows the summary of source of uncertainty, and snap sample explanations of each category.

![Figure 2.1: Summary of sources of uncertainty](image)

2.1 Source of Uncertainty Due to Measurement Error

2.2 Data from Inaccuracy Meter Reading

A water meter is a device used to measure the volume of water usage. Like all measurement tools, calibration is required on a regular basis to ensure reliable measurements are obtained. Meter management systems must be implemented to ensure reliable data acquisition [32][43][44]. Meter management begins with correct meter selection and installation. Inaccuracy meter measurement errors occur due to incorrect meter selection and flow tantalizers may not register correctly for several reasons, that includes [24][25].

1) Incorrect selection, installation, positioning, and orientation.
2) Low Flow errors (under-reading).
3) Increased error by to age, corrosion and wear of meter components.
4) Increased error due to wear from accumulated flow volume.

Water meters, like any other devices, are subject to wear and tear, and hence loss of accuracy with time. The amount of water lost due to metering inaccuracies at low flows can be significant and varies greatly depending on meter type used, billing index, presence of private storage elevated tanks and post-meter leakage flow rates. For example, most of old metering models could hardly register flows at 0.996 gph. If water distribution pipeline connection between meter are leaking at 0.996 gph, and if we assume there are 15,000 leaking pipes at the connection in a distribution pipeline network, then annual lost revenue could be 600,000$ assuming a tariff of 3.58$ per 100cf, and also the data available for water consumption will also impaired [27][28][33][44].

Water demand/consumption varies continuously over time depending on daily, weekly, seasonal and long-term such as population changes and future system performance factors. These influence customer usage patterns, habits and demand requirements. Even if modeling water demand requires baseline demand data, exposed to demand multipliers and peaking factors, which may be influenced by time-varying values and/or steady state factors, modeling attempts to predict water consumption, or leakage analysis based on inaccurate meter reading data will leads to inaccurate results [44][45].

2.1.2 Due to Improper DMA Meter Selection and Performance

Selecting and installing water meter not only requires detail technical assessment of the entire WDS, but also need thorough evaluation of the capabilities and needs have trained operator. Operator’s technical knowledge required for metering ranges from very basic ability to read a meter to write down the rates of diversion and total use to a high level of technical ability for automated systems that enable user to operate and control the WDS from almost any location such as Supervisory Control and Data Acquisision (SCADA). Many older water systems have little or no metering between the master production meters, and the customer meters. Therefore system operators require information about the flows within their distribution systems and at specific locations throughout the system in order to better understand water use, quantify available system capacity, and compute the amounts of none revenue water (NRW) in different parts of system [24][32][45]. Knowledge of NRW sources, is required as part of the process for managing leak detection, repair and planning water main replacement programs. Once a meter is in place, it can provide additional information for fine-tuning model calibration. Sub-metering uses meters placed at selected points throughout the system, to determine water use within the service area due to the cost and difficulty of installation. Commonly adopted locations selected for metering points are pump-stations and PRV pits Pressure Reducing Valves, at pressure zone boundaries and in pipelines, which carry virtually all flow into an area [24][27][32][36][39].

Sizing the meters is primarily a problem of understanding the range of flows that the meter will experience. The meter needs to be selected to pick up both the high and low range flows. Previously, the operator would estimate the range of flows based on both the number, and type of customers or on readings from a temporary meter. Even temporary meters require time consuming excavation, installation, in-situ calibration and uncertainty as to the validity of the readings, which may or may not have been installed at the right time, for capturing extreme events or the full range of usage flow patterns and characteristics, due to the temporary nature of their installation [12][25][29]. Meter selection can then be based on pipe size and type of meter. The required flow range operation also varies greatly depending on whether the flow operates continuously small operating flow range or intermittently e.g. Pump station which has the full flow range from zero flow to maximum operational flow. The selection of a meter depends on the nature of the flow, the site and operator preference [33][42]. The pipe network model can be used to provide information on the range of flows once the meter is in place, the model can be improved with information from the flow meter [23][24][28][33].

Once the range of flows has been determined, the type of meter can be selected. Small flows such as those in 100 to 150mm pump discharge lines can be metered by turbine meters, equipped with some type of pulse counter that produces flow rate information in analog form, suitable for AMR equipment. For greater flows, more commonly used meters are electromagnetic (Mag) meters, differential head meters venture, orifice, flow tube, and nozzle types or ultrasonic meters.

Differential head meters are usually the most reliable and least expensive, and can be run without power. Unfortunately they are limited to unidirectional flow and can produce significant head loss as the velocity increases. Advances in technology have produced several types of PRVs, which can also serve as flow meters [1][40][44] Thornton highlights the following considerations required by the person responsible for the selection of meter sizes and types for use as production, DMA or customer meter applications: Size of main, Flow range, Head loss at peak flow rates, Reverse flow requirements, Accuracy and repeatability, Data communication requirements, Cost of the meter, Cost of ownership, maintenance and replacement requirements, and Water utility preference or preferred supplier agreements [1][24][40].

3. Predicting Water Meter Accuracy
Most problems in operations research and engineering involve establishing the relationship between two or more variables. Regression analysis is the statistical technique that is often used for such types of problems [17] [20] [26] [30] [31]. An important aspect of predictive models is to be able to predict how condition will deteriorate over time. Water meter accuracy degradation is a function of many variables and it is not easy to predict meter accuracy degradation rate with certainty. However, it is important to understand the meter accuracy degradation process in every metering strategy. Many researchers have assumed a linear relationship between accuracy and age or cumulative volume through the meter for domestic small meters [3] [5] [10] [29] [33] [41]. In general, the dependent variable Y may be related to k independent variables by a multiple linear regression model. The form of the regression model is:

\[ y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_m X_{mi} + \epsilon_i \] ……….. 1

Where \( \beta_j \) (j = 0, 1,…, k) are the regression coefficients and \( \epsilon_i \) is the random error term. Assuming that regression models are performed for specific meter models same manufacturer, same meter size, same metering technology and including other variables implicitly apart from the totalized registered volume usage 3.58 US $ per 100 cf. Which is explicitly included, and then the model takes the form:

\[ y_i = \beta_0 + \beta_1 X_1 + \epsilon_i \] …………………………… 2

Where \( X_1 \) is the totalized registered volume through the meter (proxy for meter age); the term “\( \epsilon_i \)” value is used as an adjustment factor to account for different system characteristics and how they impact on meter accuracy.

4. Uncertainty Due to Data Acquisition and Transmission

Data tends to be uncertain in many applications [1] [4] [8] [13] [15]. Uncertainty can originate from diverse sources such as data collection error, measurement precision limitation, data sampling error, obsolete source, network latency and transmission error. Selecting an appropriate system for obtaining meter data is the first essential step towards determining and monitoring none revenue water (NRW). Once NRW is identified, the next stage is implementing strategies for the purpose of minimizing this NRW or the so called water loss [39]. To manage NRW obtaining metered data is an essential element for determining water use [39]. The ability to obtain this data is constantly improving due to technology advancements, which enable data to be recorded, communicated and archived. Customer meters continuously register water flowing through them, however, meter readings are traditionally gathered on a periodic basis monthly or quarterly to determine water consumption for the previous period for billing purposes [2][9]. Rapidly developing technologies more enable this data to be gathered more frequently, even continuously, via data logging or fixed network automatic meter reading (AMR) systems.

Data acquisition is the key process, which distinguishes AMR from MMR. Water Providers have a range of AMR options. The simplest option, automates the traditional MMR process, which still passes ‘Walk-by’ or ‘Drive-by’ each customer meter on a defined meter route, but significantly speeds the process and eliminates data-entry errors. Advanced AMR options consist of Advanced Metering Infrastructure (AMI), which enables multiple daily reads, at specified times, through a fixed communications network. These readings are transmitted to a central billing system, using modern communication technology. This provides real-time system monitoring functionality, reporting and system management [27][31][36][39].

Fixed network AMR systems enable customer consumption to be recorded as frequently as every few minutes. This provides the water provider with a detailed profile of the consumption variation (diurnal patterns) throughout the day. This data can be used to indicate leakage in customer premises or DMAs. Water consumption profiles can then be developed to assist modeling calibration and operational needs such as infrastructure and supply management. Therefore, the customer meter now provides many additional purposes other than the fundamental purpose of generating accurate water bills. It is critical that the meter population is maintained at a high level of functionality, reliability and accuracy. AMR offers less susceptibility to data handling errors, compared to manual meter reading, however, both methods are incur errors, depending on the size of the customer population, method of meter reading, regulations and policies. Several researchers [39] [43] [44] have recommended the following indicators to be considered for closer investigation when determining data handling errors:

- Accounts without actual meter readings for one year or longer
- Accounts which show zero consumption for more than two consecutive billing cycles
- Accounts suddenly evidencing a significant drop or increase in consumption after a stable usage history
- Accounts with confirmed AMR equipment failures
- Accounts known to have suffered from manual meter reading inaccuracy from one or more meter readers confirmed to be
replacement, as part of a meter management, repair and meter error becomes excessive and is scheduled for meter readings to allow for performance reduction, until the calibrating meters currently in service to correct the actual volume measured). Factors can then be determined to regularly testing meter performance (for various meter age and Trends analysis can be performed for the meter fleet, by

4.1 Transfer and Systematic Data Handling Errors

Water providers manage meter data for thousands of customers. Systematic data handling inaccuracies can be easily hidden within the volume of bulk data. The following steps are typically performed in the order shown, to capture meter data to a historical archive. Customer meter registers water flow, routine meter reading taken, manually or automatically, Meter readings are transferred to customer billing. Customer consumption is shown on water bill and archived and aggregate consumption data is summarized on reports [23][25]. In addition to the above causes, these are some of the areas where data handling during any of the above steps, can introduce errors into the data, such as:

- Data transfer errors
- Manual meter reading errors
- Automatic meter reading equipment failure
- Data analysis errors
- Use of poorly estimated volumes in lieu of meter readings
- Customer billing adjustments granted by manipulating actual metered consumption data
- Poor customer account management (accounts not activated, lost or transferred erroneously)

Three primary meter problems have been identified [5][33][39] which could contribute to apparent water losses, those are:- due to customer meter inaccuracies which gradually declining of the meter’s mechanical accuracy, due to wear and age, meter or meter reading device may fail or stop altogether and Meters may not be of the proper size or type to accurately register the full range of water flows encountered in a given customer supply. Mechanical Wear Errors:- Loss of meter accuracy due to the mechanical wear of meter components is caused by [33]:

- Aggressive water quality
- High rates of accumulated flow measured
- Chemical or residual buildup
- Abrasive materials in suspension (such as sand)
- Air running through the meter after a system outage

Trends analysis can be performed for the meter fleet, by regularly testing meter performance (for various meter age and volume measured). Factors can then be determined to calibrating meters currently in service to correct the actual meter readings to allow for performance reduction, until the meter error becomes excessive and is scheduled for replacement, as part of a meter management, repair and replacement program. Zero Consumption Errors:- Meters which show no registration (since the last reading) may be due to zero consumption or may be the result of the register failing and not registering the full volume of water consumed between meter reading and billing cycles. This can introduce large amounts of Apparent Losses and lost revenue Non-Revenue Water [3][34][36][39][40]. Improper Meter Selection Errors: - Many brands of meters are known to have their own inherent inaccuracies and performance characteristics, which increase with the age of the meter and the accumulated volume of water measures [4][24][31][32].

5. Uncertainty Due to Model

Models can simply produce numerical results. Model analysis and presentation can be improved by integration with commercially available geographic information systems (GIS) systems, to provide interactive manipulation and visual display of results, trends and spatial interaction and information [5][33][38]. Modelers need reliable and accurate data inputs, in order to create and calibrate meaningful models for the purpose of determining water demand patterns [44]. These results are necessary for determining future infrastructure planning and development. Model data must be accurate, reliable and up-to-date, especially during periods of reduced water availability, such as prolonged periods of drought.

The popularity of geographic information systems (GIS) among water providers enables GIS systems to be utilized for storing and manipulating demand data. This data is obtained using meters. Therefore, meters and their resulting measurements form the backbone of the entire water management system [5][12][33]. Use of meter data (readings) for revenue collection is generally the main priority for water providers, however, current technological developments, environmental factors e.g. water scarcity, population growth, etc, and government regulation is the driving force behind using and obtaining meter data more frequently for a real-time analysis of water use, for the purposes of system monitoring and management, in addition to billing requirements. In general, there are three different types of model uncertainty, which incorporate the model system components: Structural uncertainty, Parameter uncertainty and Data uncertainty [6][8][9][10][25][29][37].

5.1 Structural Uncertainty

Refers to errors in the mathematical representation of reality that result from system conceptualization abstraction, numerical representation, and discretization of a model in space and time. The system boundary (B), and model equations (f) are both part of the model structure. Structural uncertainty is a form of epistemic uncertainty, which can be reduced as
more information becomes available to constrain understanding of a system, and enhance model representation. However, as models can never be confirmed as 'true', structural uncertainty will never be eliminated. Structural uncertainty /Error is widely known as systems are often simplified for reasons other than epistemic uncertainty (e.g. computational and data constraints lead to simpler system representation). Such errors, whilst known to exist, are often not accounted for fully/explicitly as they are difficult to quantify [9][10][11][12][14][16][37].

5.2 Parameter Uncertainty

Parameters uncertainty reflects uncertainty for different values of variables used in equations to represent model system components for example pipe roughness. Parameter uncertainty may be a form of both random uncertainty due to natural randomness in the process, and scientific uncertainty due to limited data and knowledge. Nodal demands in WDN are a form of random uncertainty as demand varies temporally throughout the day. Scientific uncertainty in model parameter values often results from the discretization of model equations in time and space, resulting in an inability to consistent with the scale of observations with model parameters. Many model parameters e.g. roughness are often effective as they cannot be observed directly in nature, and are estimated indirectly via calibration. Parameter uncertainty can result in large errors in model predictions, and of all forms of uncertainty, therefore they have received widest attention [12] [16] [25] [29].

5.3 Data uncertainty

These refers to uncertainty in the quantities used to define initial conditions, model inputs and observations used to evaluate model predictions either system states or outputs. Such uncertainty can result from either instrumentation error that fails to accurately and precisely record the quantity of interest [7][16][41], or result from the spatial and/or temporal miss-match between the scale, resolution of observation, and that required predicted by the model. Measurement uncertainty can be both predictable and unpredictable in nature [24][38].

6. Uncertainty Quantification

Water lost due to metering inaccuracy is a function of the proportion of water consumed at different flow rates [12][18][20][24][29].

\[
\text{CAWL I} = \frac{\sum_{n=1}^{n} P w e_n (1 + r')^{-1}}{t}
\]

Where, \(\text{CAWL I}\) is the cost of annual water loss due to inaccurate meters, n is the total number of meters for each meter model, \(r'\) is the real discount rate 6%, \(P\) is the price of water 3.58 US $ per 100 cf, and assumed to be constant throughout the analysis period, \(Q\) is the average annual volume of water registered through domestic meters, t is the number of years the meter is in service, \(E_i\) is the average weighted meter error during the useful life of the meter over the low flow ranges below \(Q\) and k is a discount factor for the time the meter is registering other flows 10%.

7. Uncertainty Calibration

Calibration may be defined as the method by which parametric uncertainty in models is reduced [20][22][41][44]. During these process uncertainties including parameter uncertainty have received the greatest attention. Assuming calibration of any proposed model, f, is typically confronted with a vector in time or space of observed system behavior

\[
\text{E} = (z_{min}, z_{max})
\]

which may represent both system output, and system states. The vector of residuals \(E_i\) is defined as the difference between \(Y\) and \(Z\) in the case of system outputs [10][200]:

\[
E_i = y_i - \theta z_i x_i B_i U_i - z_i, \quad i = 1, \ldots, n
\]

These traditional approaches have sought to minimize the vector of residuals to zero by adjusting model parameters, without considering structural uncertainty and input data uncertainty. Initial approaches to reduce parameter uncertainty through calibration in WDN models were based on trial and error procedures [9][33][35][36][43], which by manually adjusting model parameters, seek to maximally reducing an objective though often subjectively chosen function, such as the standard least squares problem (E):

\[
\min E = \sum_{i=1}^{n} E_i
\]

Manual calibration has also been applied extensively in WWTP model calibration and in this context is termed the process engineering approach. The process is effectively a local search process of the parameter hypercube, which may fail to find all well performing parameter sets. The engineer therefore requires expert process knowledge and experience for manual calibration [17][19][20][24], explicit calibration approaches have also been applied that solve the steady state mass balance and energy equations for the WDN, where unknown parameters are solved using the same number of equations [31][44]. Where sufficient measurements are not available to constrain calibration parameters an under-determined problem, parameters need to be grouped to make the problem at least even-determined. The explicit
methodology is limited for three reasons. The posed calibration problem must be at least even-determined. Measurements are assumed 100% accurate and data errors are not considered. Uncertainty in estimated parameters cannot be quantified. Both manual and explicit calibration approaches are considered to only have historical significance \cite{7}\cite{19}\cite{20}\cite{24}, and have largely been superseded by implicit optimization techniques in model calibration that are more flexible in dealing with uncertainties.

8. Uncertainty Reduction

Mathematical modeling of WDS gives unlimited opportunities in the design, maintenance and operating phase of the water supply life-cycle. Usually, it is developed using a system of deterministic, nonlinear equations that are solved numerically. Input values pipe lengths, pipe diameters, nodal demands, etc. are just one part of the model configuration. One of the ways to reduce uncertainty is to determine the input parameters and network topology more accurately. For some quantities, such as pipe length and pipe diameter, it is just a matter of a better field survey. But for nodal demands the uncertainty is harder to express and measure, especially if network leakage is represented as nodal demand pressure-related or constant. Nodal demand is a conceptual parameter and its determination is usually a product of the engineer’s knowledge and the information available. Sometimes, increasing the information used in nodal demand determination will not provide a more accurate and reliable result. To overcome this problem, engineers can use additional information about nodal demand that is easier to determine with sufficient reliability.

For example, the total of all nodal demands in the network or part of a monitored network, DMA has to be equal to the total network inflow. Total network inflow is much easier to determine, and it is usually in the data that already exist. To start solving the posed problem it is necessary to determine the representation for each uncertain variable. There are three ways of representing uncertain variables \cite{17}\cite{19}\cite{20}\cite{24}. The first, the simplest, is in the form of intervals. This form only provides information about the boundaries of all possible variable values. The second, statistical distribution provides the probabilities of the possible values. The third is the fuzzy set that can be considered as a set of intervals with the membership level of each interval.

9. Conclusion

Water distribution operational data has always been the benchmark for effective decisions making process, with an increasing number of leading water utility companies are taking advantages of using these tremendous amounts of operational data coupling with information technology solutions in their operational and management practices. This paper has highlighted the potential sources of uncertainty in water distribution system (WDS) pipeline leak detection using district metered area Data. It is also very important to understand these different source and types of uncertainty, to address them using an appropriate uncertainty analysis tools to determine a more accurate and reliable result by increasing data quality.

References

[6] Biao Qin, Yuni Xia, Fang Li. DTU: A Decision Tree for Classifying Uncertain Data. In the Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2009.


