Ανάλυση και αξιολόγηση αποτελεσμάτων εφαρμογής συστήματος

Παραδοτέο 5.4   **30/11/2020**

Πίνακας Περιεχομένων

1 Introduction

In the context of this deliverable, the analysis and evaluation of the results of the pilot tests and the performance evaluation of the data analysis techniques designed in the SmartWater2020 project will be presented. This includes the study of data verification and reconstruction methods, the assessment of hydraulic and quality status, the study of leak detection techniques using pressure sensors and AMR, the study of dynamic PRV control and finally, the analysis of the performance of the LoRaWAN network.

2 Data Reconstruction (FORTH)

2.1 Methodology

Compressive sensing (CS) provides a powerful framework for simultaneous sensing and compression, enabling a significant reduction in the sampling, computation, and transmission costs on a sensor node with limited memory and power resources. In the framework of SmartWater2020, CS is exploited as an efficient data, specifically pressure, compression mechanism, implemented at the edges of the network, whilst the decompression takes place at the control center where increased computational resources are available. A typical CS-based system consists of two distinct modules according to the functionality they perform, namely, the compressor and decompressor. The compression module (or encoder) is responsible for generating a reduced set of random measurements from the observed data.
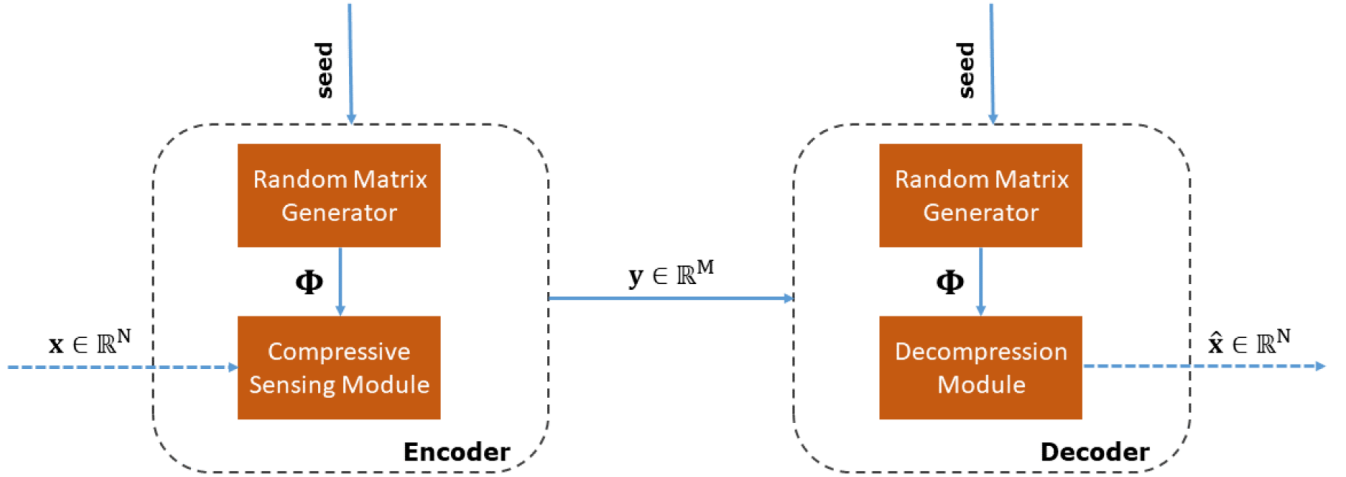

Figure 1: Flow diagram of a CS-based system.

More specifically, let $\mathbf{\Psi} \in \mathbb{R}^{N \times P}$ be a matrix whose columns correspond to a possibly overcomplete (i.e., $N < P$) transform basis. Let $\mathbf{x} \in \mathbb{R}^{N}$ be an observed discrete-time signal of N samples, which is associated to a transform coefficients' vector $\mathbf{a} \in \mathbb{R}^{P}$ over the basis $\mathbf{\Psi}$, as follows

$$\mathbf{a} = \mathbf{\Psi x} .$$

In terms of signal approximation, it has been demonstrated that, if a signal $\mathbf{x}$ is sparse or compressible in a basis $\mathbf{\Psi}$, then it can be reconstructed from a highly reduced set of M $<<$ N non-adaptive linear projections, where $M = \mathbf{O} \left(S \log \left(\frac{N}{S} \right) \right)$, where S is the signal's sparsity. From a practical perspective, instead of transmitting the originally observed N samples of $\mathbf{x}$, a sensor

reduces its consumed energy by only transmitting this significantly smaller number of M projections to the control center, where the original signal can be recovered with high accuracy for further processing. The random measurements vector $\mathbf{y} \in \mathbb{R}^M$ is generated simply as follows,

$$\mathbf{y} = \mathbf{\Phi x} \, ,$$

where $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is a measurement matrix, which must be incoherent with the sparsity basis $\mathbf{\Psi}$. In mathematical terms, let

$$\mu(\mathbf{\Phi\Psi}) = \max_{\substack{i=1,\ldots,M \\ j=1,\ldots,P}} (|\boldsymbol{\phi}_i^\top \boldsymbol{\psi}_j|)$$

denote the mutual coherence between $\mathbf{\Phi}$ and $\mathbf{\Psi}$, where $\boldsymbol{\phi}_i$ and $\boldsymbol{\psi}_j$ are the i-th row of $\mathbf{\Phi}$ and j-th column of $\mathbf{\Psi}$, respectively. The parameter μ serves as a rough characterization of the degree of similarity between the sparsity and measurement systems. The smaller the μ is, the more incoherent the two matrices are.

In practice, the system operator is responsible for defining the appropriate number of measurements by setting the value of the sampling rate (SR), which is simply the ratio of the number of random measurements over the original signal length, that is, $\mathrm{SR} = \frac{M}{N}$. Given that $M \ll N$, the computational and power savings of each sensor node stem from the fact that they process and transmit a highly compressed signal instead of the original $\mathbf{x}$. To be consistent with lossless compression, hereafter we also use compression rate (CR) as the input parameter to the compressive sensing, where CR = 1−SR.

Function 1: Edge compression overview

Function_1: *edge_compression(x, SR, seed)*

Inputs:

@x: The original pressure stream (bars)

@SR: The sampling ratio value ([0.1, 0.9])

@seed: The random matrix generator

Output:

@y: The compressed signal (bars)

By employing the M random measurements and given the S-sparsity property in the transform basis, the original signal $\mathbf{x}$ can be recovered by taking a number of different approaches. In our implementation, the NESTA algorithm (Matlab code available at https://statweb.stanford.edu/~candes/software/nesta/) is employed, which is shown to achieve a good trade-off between reconstruction accuracy and computational time. Focusing on the optimization problem to be solved for reconstructing the original data, NESTA solves the following synthesis-based problem,

$$\min_{\boldsymbol{\alpha} \in \mathbb{R}^P} \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \boldsymbol{\Phi}(\boldsymbol{\Psi}\boldsymbol{\alpha})\|_2 < \delta \, ,$$

where $\mathbf{a} \in \mathbb{R}^P$ is a sparse coefficient vector, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the l_1 and l_2 norm, respectively, and $\delta > 0$ is a small threshold ($\delta = 10^{-3}$ in our implementation). Having estimated the sparse coefficient vector, $\hat{\mathbf{a}}$, a reconstruction of the original signal is simply obtained by taking the inversetransform, that is,

$$\hat{\mathbf{x}} = \boldsymbol{\Psi}^{-1}\hat{\boldsymbol{\alpha}} \, .$$

The short-time Fourier transform (STFT), along with scrambled block Hadamard ensembles are utilized in our CS-based system in place of the sparsifying transformation $\boldsymbol{\Psi}$ and random measurement matrix $\boldsymbol{\Phi}$, respectively.

Function_2: *control_decompression(y, seed)*

Inputs:

@y: The compressed pressure stream (bars)

@seed: The random matrix generator

Output:

@x: The decompressed signal (bars)

Function 2: Control decompression overview

2.2 DEYAM Pilot

2.2.1 CS Vanilla-flavored

The subsequent performance evaluation utilized real pressure data recorded by the smart wate management infrastructure of the Municipal Enterprise for Water Supply and Sewerage of Malevizi. Incoming and outgoing flow pressure measurements are sent to the control center at a frequency of one sample (pressure expressed in Bars) per 15 min, yielding approximately 9000 observations per sensor. This sampling frequency suffices in order to enable real-time monitoring of the water distribution network of the municipality of Malevizi, which is divided into 10 zones, each monitored by a pressure sensor.

In our system, the reconstruction error is measured in terms of the signal-to-error ratio (SER) (in dB) between the original and reconstructed signals $\mathbf{x}$ and $\hat{\mathbf{x}}$, respectively, defined by

$$\text{SER}(\mathbf{x}, \hat{\mathbf{x}}) = 10 \log_{10} \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2}$$

The figure below shows the original along with the three reconstructed streams.
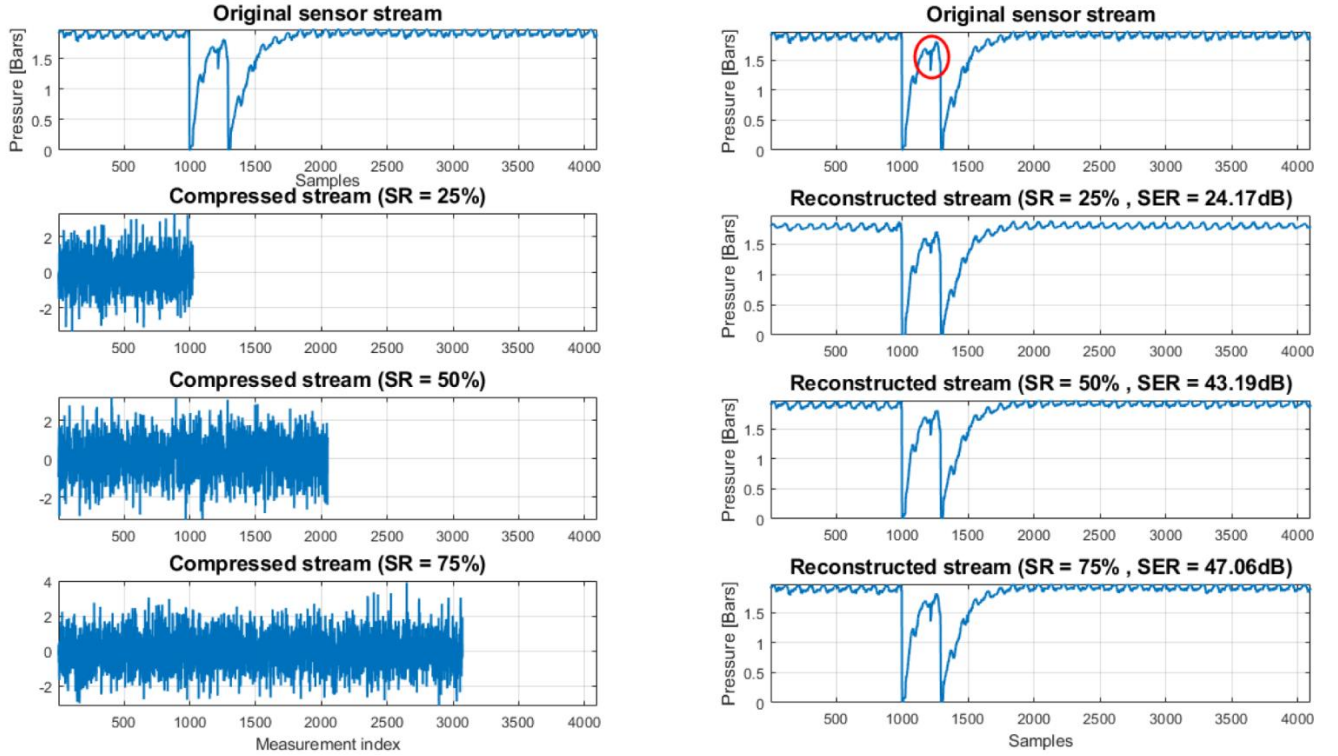
Figure 2: Original and reconstructed streams

As it can be seen, the reconstruction quality improves as the SR increases, as expected. Most importantly, the reconstruction is already accurate enough even for SR = 25%, except for some sharper details (see region in the red circle) that cannot be captured accurately when the number of random measurements M is small. Nevertheless, these details can be recovered very accurately as the SR slightly increases (see SR = 50%).

As a second illustration, the figure below shows a part of an original pressure stream of a sensor from DEYAM pilots, under abnormal network conditions.
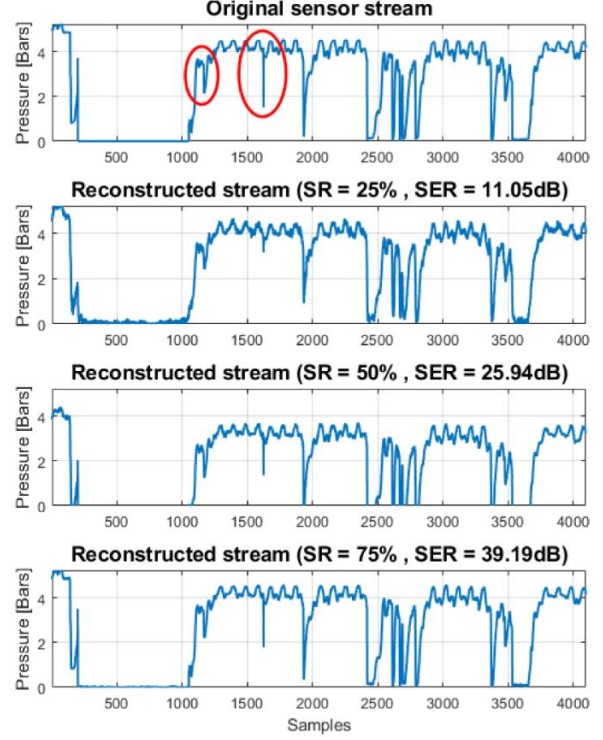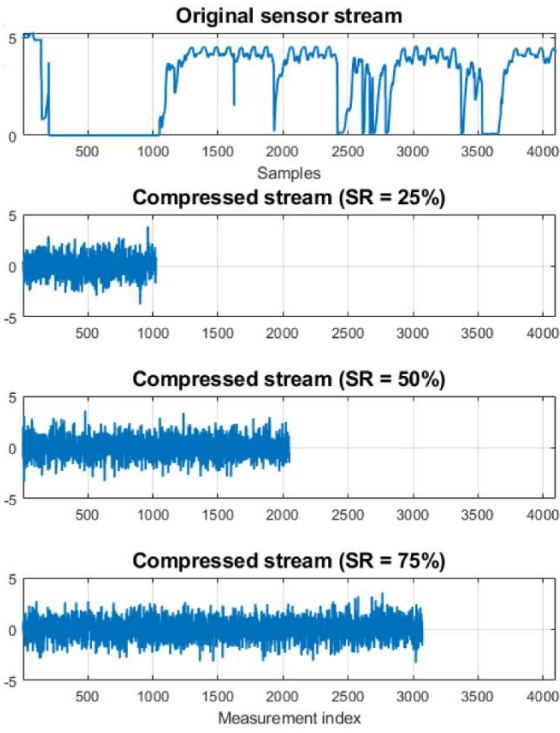
Figure 3: Part of an original pressure stream of a sensor from DEYAM pilots, under abnormal network conditions

2.2.2 CS weak encryption

In the subsequent analysis, we evaluate the weak encryption capability of a CS-based system by simulating the following adversarial scenario, as shown in Figure 4. Specifically, we assume that an adversary does not have access to the true original measurement matrix Φ that generated the compressed stream, but to a permutation of its rows. This scenario is simulated easily as follows,

$$\mathbf{y}_A = (\mathbf{P}_M \Phi)\mathbf{x}$$

where $\mathbf{y}_A$ is the random measurements vector generated by the adversary and $\mathbf{P}_M \in \mathbb{R}^{M \times M}$ is a permutation matrix which models the imperfect knowledge of the true Φ on behalf of the adversary. In the subsequent evaluation, the percentage of permuted rows of the original measurement matrix is defined by p ϵ [0.2, 0.4, 0.6, 0.8, 1], where, for each p value, $\lfloor pM \rfloor$ randomly selected rows are permuted while the remaining rows are left in the original position. When a legitimate system operator receives the compressed measurements, we assume that the permutation matrix is equivalent to the identity matrix, i.e., $\mathbf{P}_M = \mathbf{I}$.
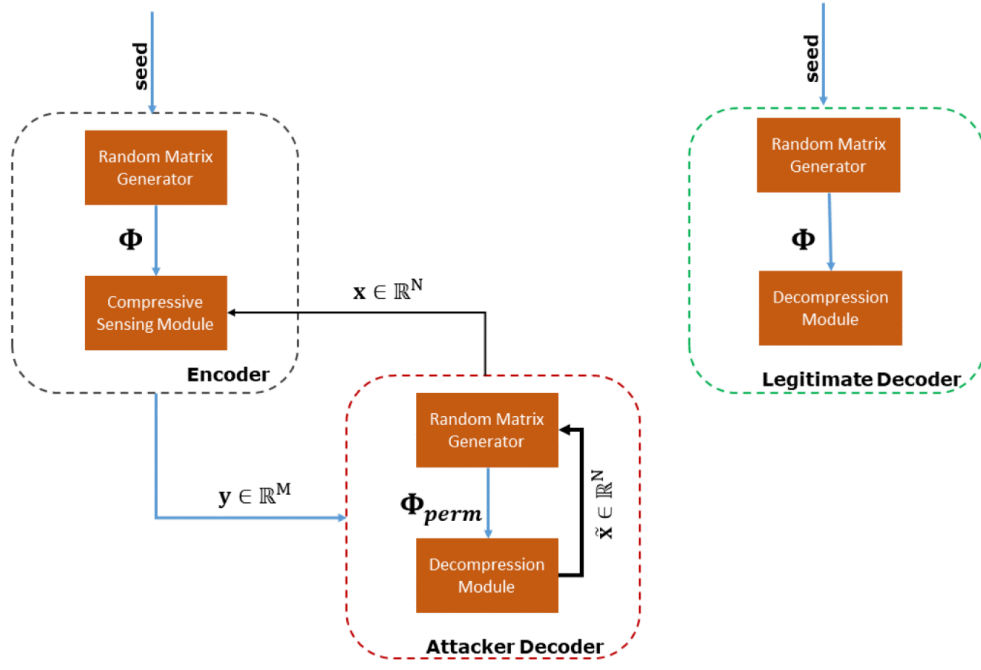
7

Figure 4: Adversarial scenario demonstrating the weak encryption property of a CS-based system.

The figures below show the reconstruction error, in terms of the achieved SER (in dB) averaged over all the pressure sensors, for sliding windows of length $N \in \{64, 128, 256\}$, as a function of p, for the three sampling ratios $SR \in \{25\%, 50\%, 75\%\}$ (or, equivalently, compression ratios $CR \in \{75\%, 50\%, 25\%\}$). Clearly, the reconstruction accuracy deteriorates dramatically, as p increases, for all the window lengths and sampling ratios, which verifies the weak encryption capability of CS. The difference in performance between the original and permuted Φ especially increases as the sampling ratio and window length increase. Furthermore, the larger the window length N and the smaller the CR (i.e., the higher is the SR), the better is the reconstruction performance (i.e., higher SER), as expected.
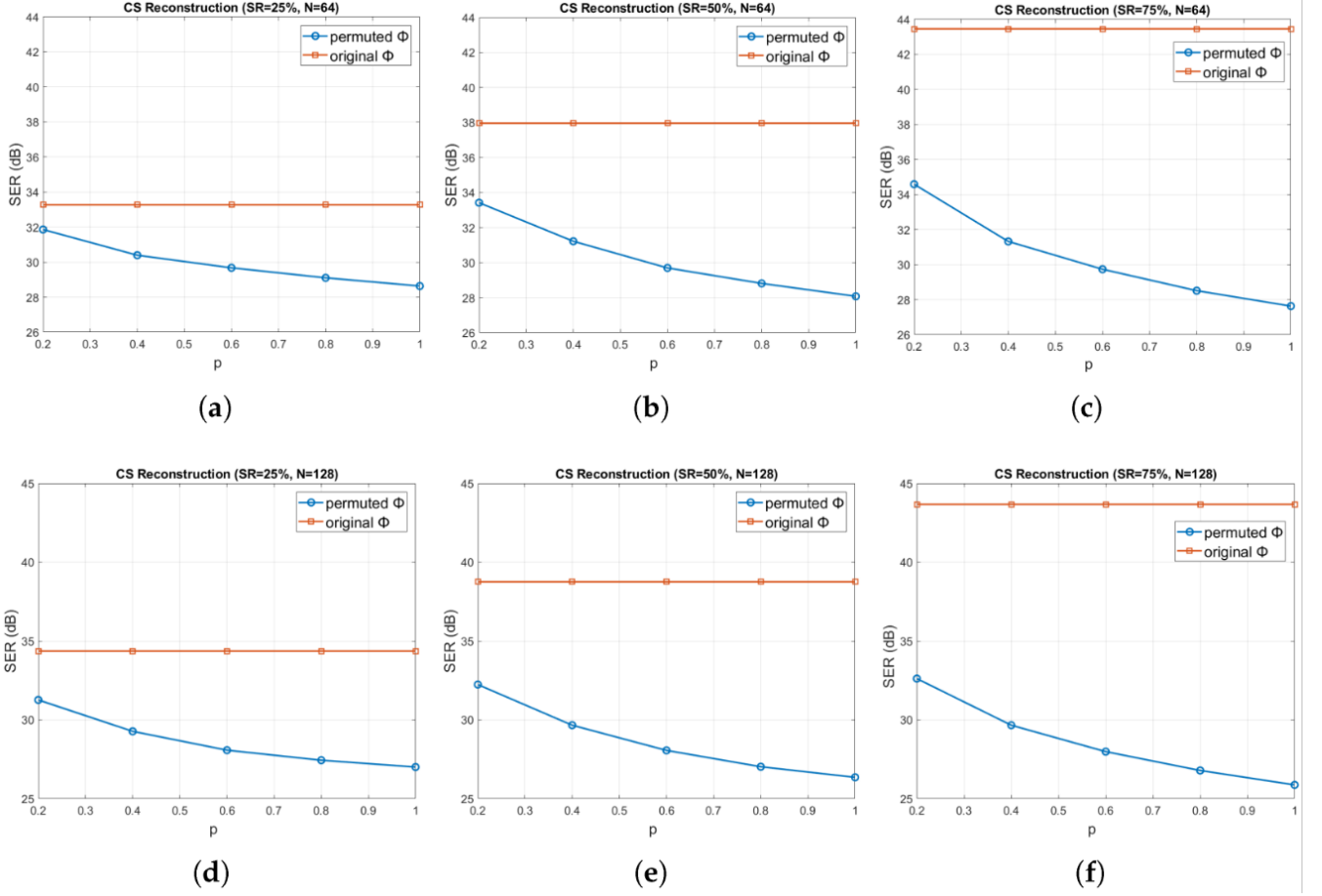
Figure 5: The reconstruction error, in terms of the achieved SER (in dB) averaged over all the pressure sensors, for sliding windows of length $N \in \{64, 128, 256\}$, as a function of p, for the three sampling ratios $SR \in \{25\%, 50\%, 75\%\}$ (or, equivalently, compression ratios $CR \in \{75\%, 50\%, 25\%\}$

2.2.3 CS Execution Time and Power Consumption

As an additional experiment, we evaluate the efficiency of the CS-based mechanism for data compression and transmission in DEYAM system and illustrate the execution speedup and energy consumption reduction it offers when compared against a well-established lossless compression method that is widely used in commercial solutions, namely, the LZ77 algorithm. We also quantify the energy savings achieved over the scenario of raw (uncompressed) sensor value transmission.

The experimental setup parameters are summarized in the table below.

Table 1: The experimental setup parameters

Parameter	Value
Sensor sampling frequency	1 sample every 15 min
Total number of pressure values in the monitored period	9984
Block size N	$\{64, 128, 256\}$
CS compression ratio CR_{CS}	$\{25\%, 50\%, 75\%\}$
Measurement matrix Φ	Hadamard
LZ77 dictionary hash table size	1 KB
Network stack	CoAP + UDP + IPv6 + 6LoWPAN
Physical and MAC layer	Non-beacon-enabled CSMA, IEEE 802.15.4
TX power	0 dBm
RF channel	26 (2480 MHz)

To evaluate the effect of different compression types to the performance metrics defined here, we followed a statistical-based approach. Due to lack of normality in our dataset (as reported by Shapiro–Wilk test), we applied the non-parametric Kruskal–Wallis test, followed by Dunn post hoc test for pairwise comparisons of compression types, in the case a significant difference in the means exists.

The figure below shows the average and standard deviation of Compression Execution Time (CET), over the total number of pressure blocks, for different compression types.
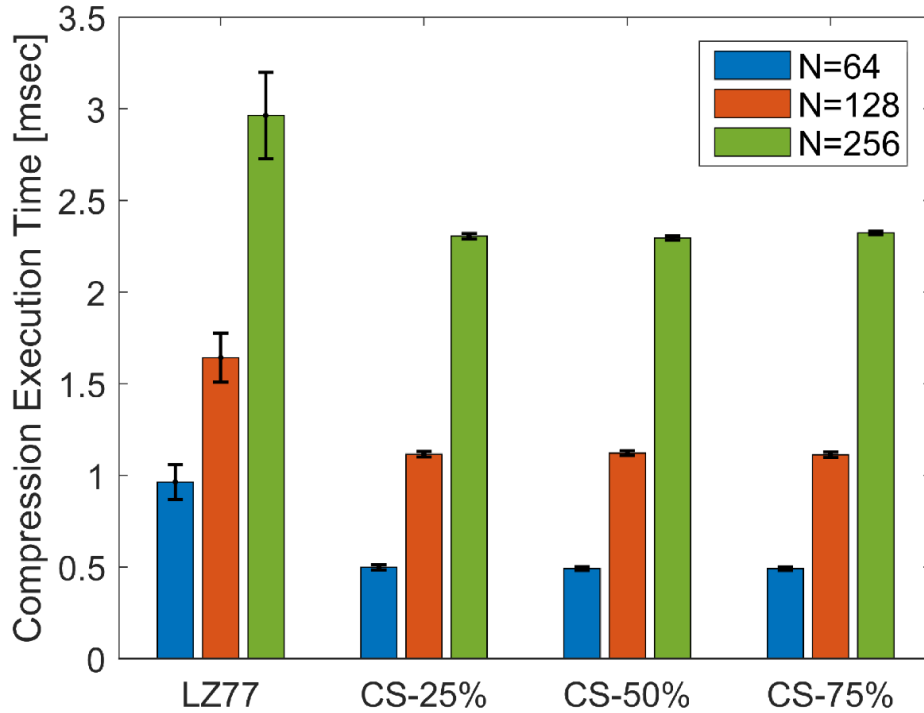

Figure 6: The average and standard deviation of Compression Execution Time (CET), over the total number of pressure blocks, for different compression types.

The average and standard deviation of Transmission Energy Consumption (TEC) is shown in the next figure, over the total number of pressure blocks, for different compression types.
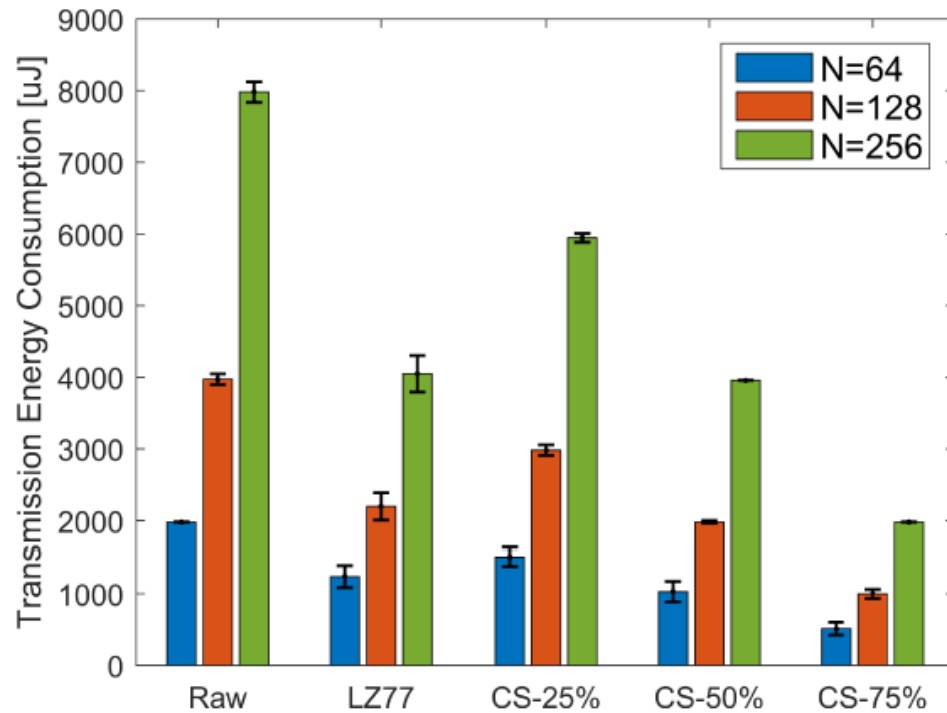

Figure 7: The average and standard deviation of Transmission Energy Consumption (TEC)

2.3 Conclusions

3 Hydraulic State Estimation (KIOS)

The "Hydraulic State Estimation" algorithm combines the available measurement data from the SCADA system with the WDN model data (network topology, asset data, etc.) and extrapolates the complete system state, i.e., water flows in every pipe, consumer waters demand, pressures at each node and tank water levels. A complete view of the network state supports the decision-making process and enables the efficient operation of these systems, improves customer service, and enables the early detection of emergency events, thus minimizing their impact. A demand calibration method is integrated in this algorithm in order to achieve a better matching between the measured and the predicted pressures and flows.

3.1 Methodology

The hydraulic state estimation process is carried out in the steps described below.

3.1.1 Creation and selection of hydraulic model

Hydraulic models of the network are created using GIS and CAD representations of the network provided by the water utilities. These include the available information regarding the topology and the hydraulic characteristics of each network component.

An estimated water consumption is allocated at each network node, which is calculated based on historical billing data. Moreover, the average water demand of each DMA is estimated using the DMA inflow measurements and then distributed to each building of this area based on its type, area and number of floors. Since information of the actual location of each consumption node is not available, each building's water demand was reallocated to the nearest node using "Voronoi polygons".

The initial (non-calibrated) hydraulic model of each DMA network are stored in a database and the user can select the appropriate model to perform state estimation.

3.1.2 Measurement data retrieval and validation

Hydraulic sensors have been installed at the inlet point of each DMA network (measuring pressure and flow) and at strategic locations inside the network (measuring pressure). Through a Graphical User Interface (GUI) the user can select the desired time period for which all measurements from hydraulic sensors installed in the selected water network are retrieved from the SCADA system (Figure 8: Graphical User Interface for selecting the desired time period for hydraulic-state estimation (Figure 8).

The retrieved sensor data are then pre-processed in order to detect outliers which may be the result of communication of sensor faults. Missing data points due to outliers or communication faults are recreated using interpolation of the available data points.

3.1.3 Hydraulic model calibration using sensor data

The available sensor measurements are first used to calibrate the available hydraulic model.

The first calibration step is to calibrate consumer demands. Due to the absence of knowledge of the actual consumption patterns at every node of the network, the inlet flow measurement and AMR measurements (wherever they are available) are used to generate and assign demand patterns to the corresponding consumption nodes of the model. The demand pattern time-step is the same as the sensor measurements time-step. The demand pattern pat_i for node i with base

demand bd_i at discrete time-step k and given the inlet flow measurement q_{in} is calculated as follows:

$$\text{pat}_i(k) = \frac{bd_i}{\sum_j bd_j} q_{in}(k)$$

Moreover, the inlet pressure measurements is used to create and assign a pressure pattern at the inlet point of the DMA network model.

The second calibration step is to calibrate model parameters by incorporating into the procedure the available pressure measurements. Using the assigned inlet pressure and demands for each node, the hydraulics of the DMA network are solved using the EPANET hydraulic solver to calculate the flows and pressures at every pipe and node of the network respectively.

Pressures measured by sensors are then compared with the corresponding simulated pressures in order to evaluate the accuracy of the model. By using the difference between the measured and estimated pressures, a calibration of network parameters is performed to increase model accuracy which includes:

1. Redistribution of base demands.

2. Checking model structure for closed pipes.

3. Calibration of model parameters such as pipe roughness and node elevations.

Note that if calibration of network parameters has been performed for an adequate amount of data, subsequent applications of the state estimation algorithm will not require the performance of this step.

3.1.4 Hydraulic State Estimation

Using the calibrated hydraulic model, the flows in pipes and pressures at nodes are estimated throughout the network using the acquired sensor measurements and by solving the network hydraulic equations using EPANET.

Function: *Hydraulic_State_Estimation(model, data)*

Inputs:

@model: Water network model

@data: Flow and pressure measurements received from SCADA

Output:

@Calibrated_model: Inp. file of the calibrated water network

Function 3: State estimation algorithm inputs and outputs

Platform Communication:

- Communicates with EPANET database and retrieves the selected input (.inp) file.

- Communicates with SCADA database and retrieves the required data for the selected period and the selected DMA/network.

3.2 WBL Pilot

The proposed hydraulic-state estimation methodology was applied on DMA 136 of the WDN of Limassol. The results of this pilot run are shown using sensor data from a one-week period between 12/10/2020 - 18/10/2020, however, any time period with available sensor data could have been selected.

Initially the user selects the available network model from the database and states the time period for which to perform hydraulic-state estimation using measurements (Figure 8). In this case, the one-week period between 12/10/2020 - 18/10/2020 is selected.
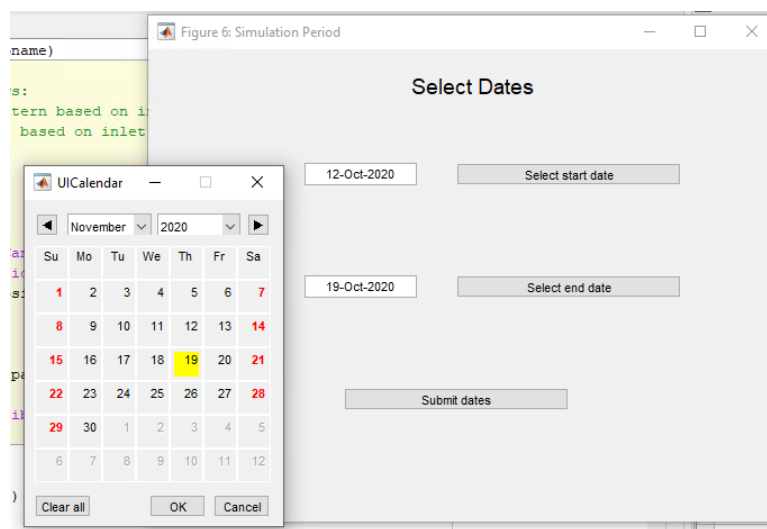

Figure 8: Graphical User Interface for selecting the desired time period for hydraulic-state estimation

The algorithm connects with the online server receiving data from the SCADA system and obtains the required measured data for the period 12/10/2020 - 18/10/2020 in order to execute the hydraulic state estimation (Figure 9).

The required data for the first step of the algorithm are the inlet flow and pressure of DMA 136 and any available AMR measurements. In this DMA, there are 2 AMR sensors located at the KEAN factory measuring its consumption and can be used for a better calibration of the hydraulic model (Figure 9).

Figure 9: Measured inlet flow, inlet pressure and KEAN factory consumption of DMA 136 received from installed sensors for the period 12-18/10/2020.

Consumer demand calibration in the hydraulic model is performed by using the inlet flow and inlet pressure measurements of DMA 136 (Figure 10) and the KEAN factory consumption (Figure 11).

These measurements were downloaded from the SCADA system and used to calibrate the demands in the hydraulic model, as seen in Figure 12.
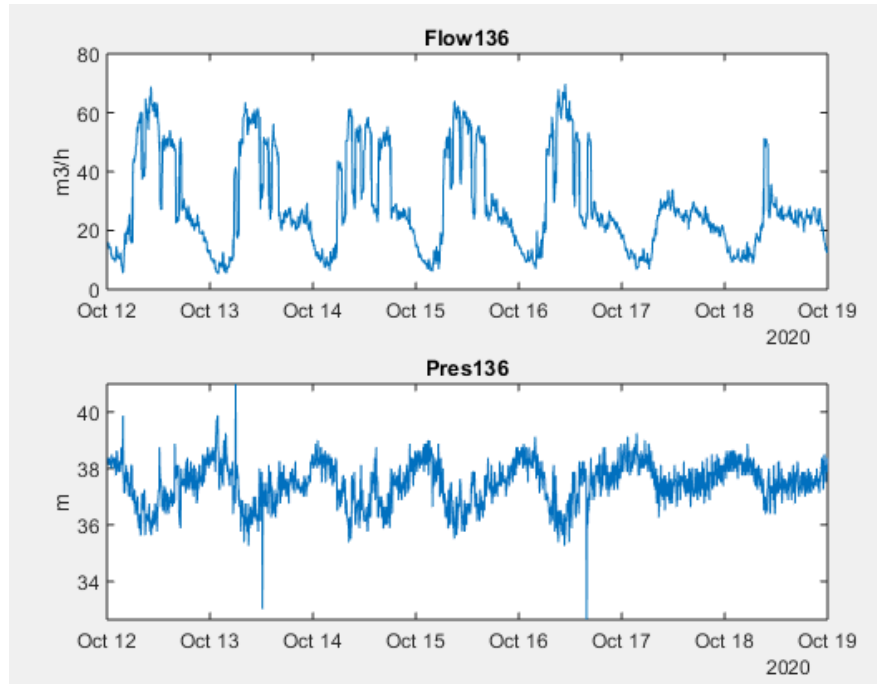
15

Figure 10: Measured inlet flow (a) and pressure (b) of DMA 136 for the period 12-18/10/2020
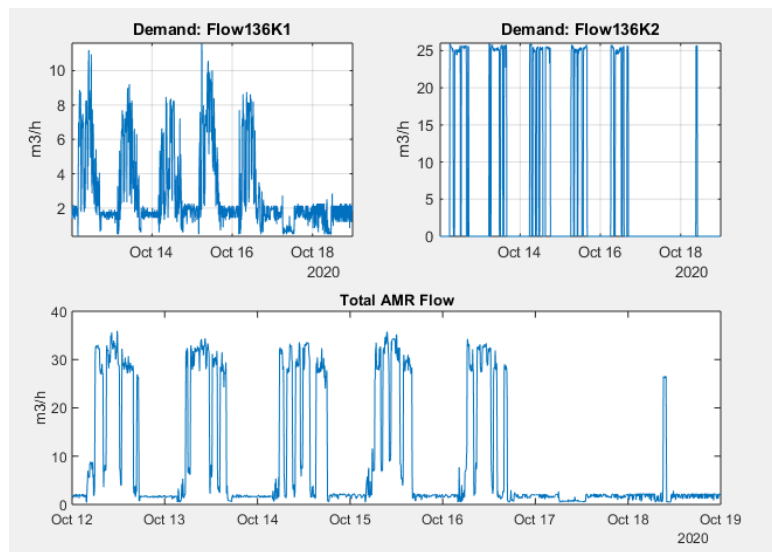

Figure 11: Measured consumption (m3/h) by two AMR sensors located at KEAN factory in DMA 136 for the period 12-18/10/2020 (a)-(b), and the total consumption at KEAN (c).
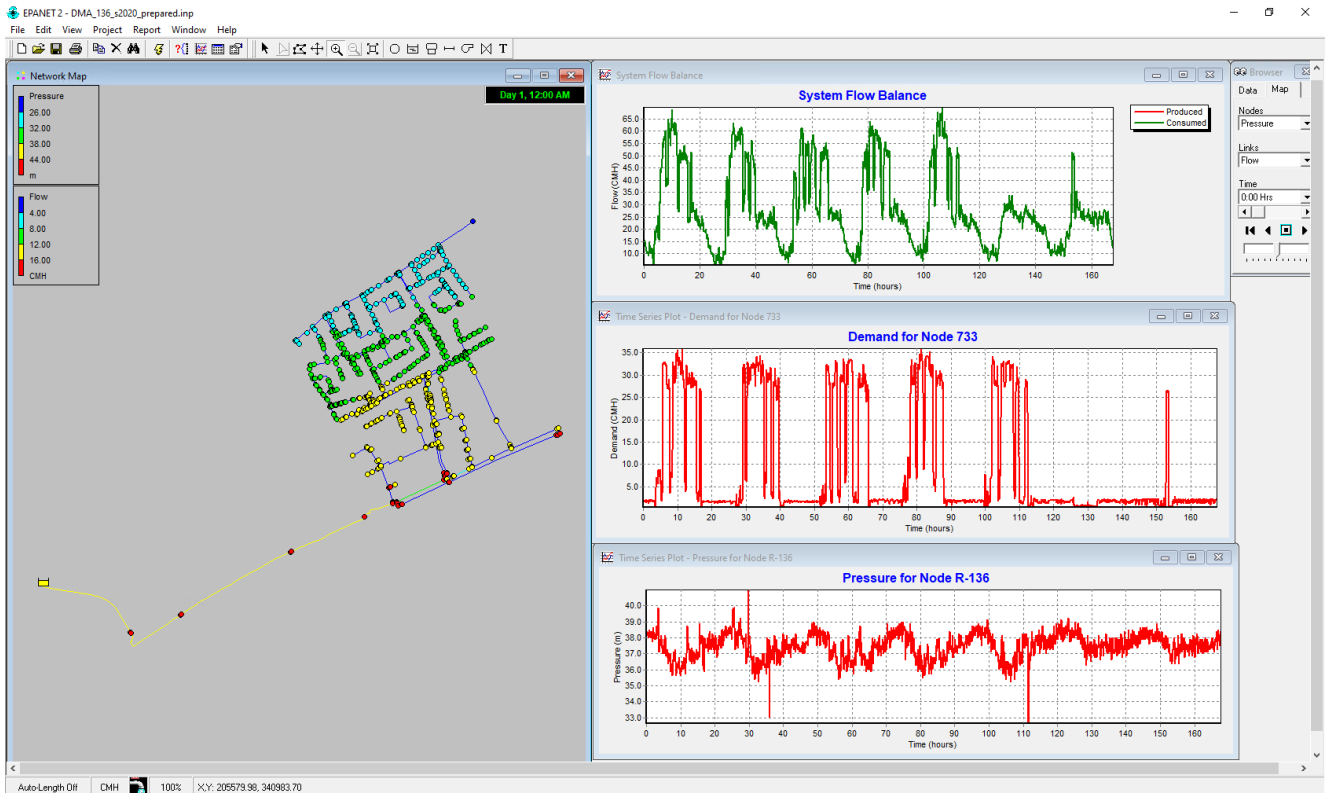
Figure 12: Calibrated hydraulic model of DMA 136 in EPANET, augmented with measurements from the inlet and KEAN obtained during the period 12-18/10/2020.

The next step of this methodology was to compare the simulated and measured pressures from sensors located in this network which were not used in the demand calibration process in order to evaluate the accuracy of the calibrated model. A GIS map with the locations of the installed pressure sensors in DMA 136 is shown in Figure 13.

In Figure 14 the measured (SCADA) and simulated (EPANET) pressure at nodes with installed sensors located at DMA 136 are presented. As can be seen in four cases (Pres136P1, Pres136P2, Pres136P4 and Pres136P5) the model accuracy is very high while in two cases (Pres136P3 and Pres136P6) is moderate. Regarding "Pres136P6" the pressure difference is following the same pattern meaning that either the actual sensor or a model parameter (node elevation) should be calibrated. "Pres136P3" represents the pressure sensor installed near the KEAN factory where AMRs are also placed. According to the results, the simulated pressure matched with the measured after some time when the actual sensor was calibrated manually. Based on information from the WBL the specific sensor needs manual calibration since it is affected form the large pipe flow of the pipe where it is installed due to the factory operation.
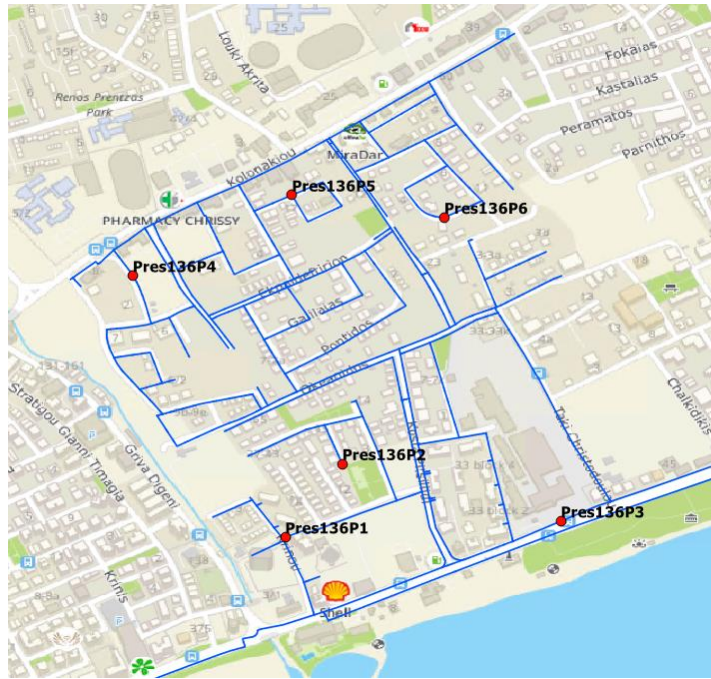
17

Figure 13: GIS map showing the location of the pressure sensors installed in DMA 136
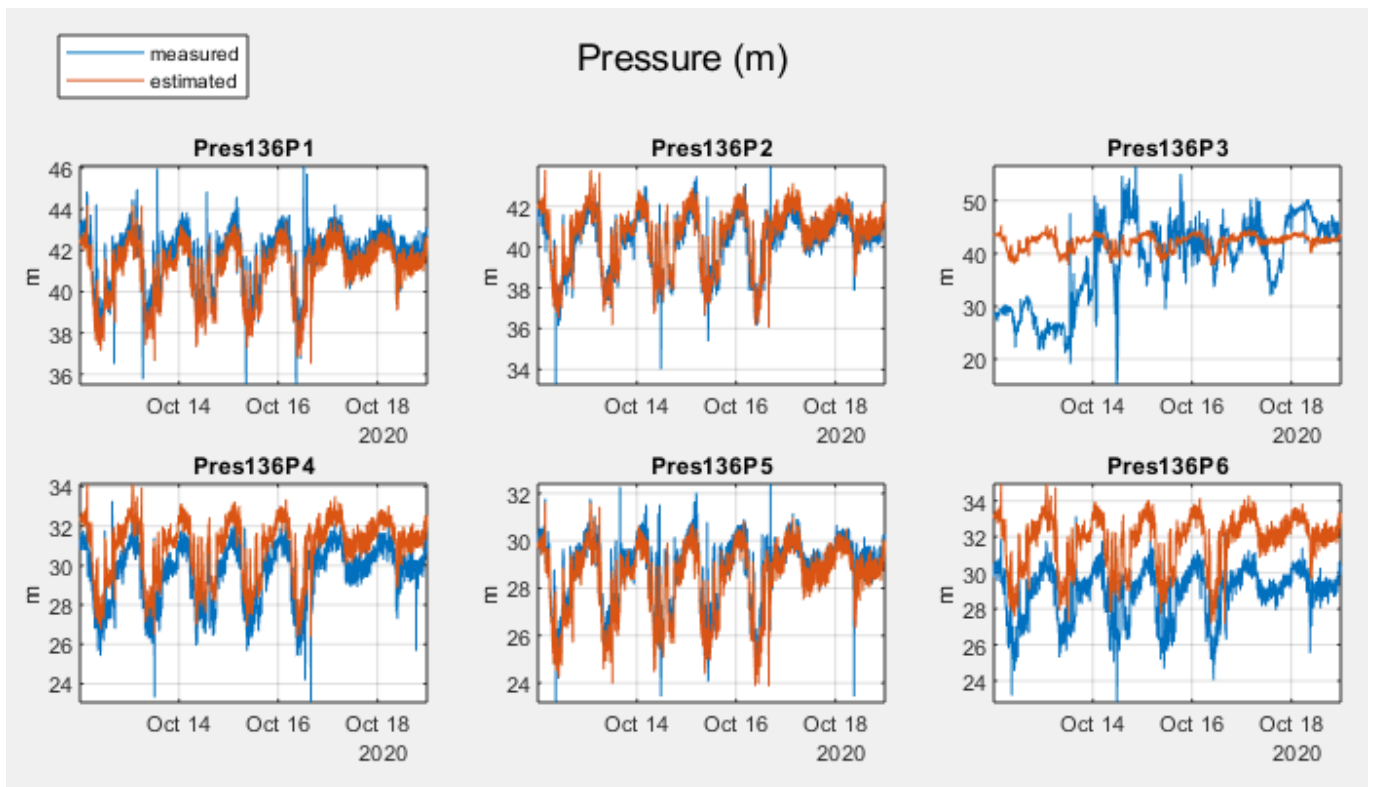

Figure 14: Measured (SCADA) and estimated (EPANET) pressure from various sensors located in DMA 136 for the simulation period 12-18/10/2020

3.3 WDD Pilot

The proposed methodology was applied at the main pipeline of the WDD network from Larnaca to Dekheleia area (Figure 15). The objective is to create a hydraulic model of the network and estimate the complete system state. The network operates under different conditions depending on the water needs of the supplied areas. Water is primarily pumped from the Tersephanou water treatment plant, however in periods of water shortage the Dekheleia desalination plant supplement the water demand. As a result, when the desalination plant operates there can be flow reversal from Dekheleia to Larnaca in the main pipelines.

Sixteen (16) AMR flow sensors (Figure 16) and six (6) pressure sensors (Figure 17) were installed along the pipeline pilot section. A list of the installed sensors in the WDD network is provided in Table 2. The flow sensors are located at the outlets of the main pipeline which supply water to nearby villages. In addition, two more pressure sensors are located at the Leivadia Pumping Station which are connected to the server through the SCADA system.

Figure 15: Map showing the main pipeline of the WDD connecting cities and villages located at south-east Cyprus. Circled is the pilot area.

Table 2: Sensor location, type and ID

| Name	Latitude	Longitude	UID	Type
DAV_54A_Chlorine_Sensor	35.00918	33.71370	Anzio_Chlorine	Chlorine
Leivadia_Chlorine_Sensor	34.96728	33.64248	Leivadia_Chlorine	Chlorine
FM_KaloChorio_Sensor	34.92597	33.54118	W001	Flow
FM_AradippouC_Sensor	34.94795	33.56950	W002	Flow
FM_AradippouA_Sensor	34.94985	33.57478	W003	Flow
FM_AradippouB_Sensor	34.95095	33.57900	W004	Flow
FM_Avdelero_Sensor	34.97185	33.58307	W005	Flow
FM_Troulloi_Sensor	34.97765	33.62106	W006	Flow
FM_Kelia_Sensor	34.97775	33.62076	W007	Flow
FM_LivadiaA_Sensor	34.96738	33.63815	W008	Flow
FM_LivadiaB_Sensor	34.97016	33.64259	W009	Flow
FM_OrokliniB_Sensor	34.97400	33.64981	W010	Flow
FM_OrokliniA_Sensor	34.98280	33.66009	W011	Flow
FM_PylaB_Sensor	35.00119	33.68145	W012	Flow
FM_PylaA_Sensor	35.00710	33.70418	W013	Flow
FM_DekehliaA_Sensor	35.00989	33.71109	W014	Flow
FM_DekehliaB_Sensor	35.00990	33.71110	W015	Flow
FM_KeliaTroulloi_Main_Sensor	34.96221	33.62606	W016	Flow
DAV_36_Pressure_Sensor	34.94193	33.55619	DAV_36	Pressure
DAV_39B_Pressure_Sensor	34.94933	33.57375	DAV_39B	Pressure
DAV_46_Pressure_Sensor	34.97074	33.64617	DAV_46	Pressure
DAV_49_Pressure_Sensor	35.00004	33.67688	DAV_49	Pressure
DAV_54A_Pressure_Sensor	35.00918	33.71370	DAV_54A	Pressure
DAV_Tremithos_Pressure_Sensor	34.89772	33.53957	DAV_Tremithos	Pressure
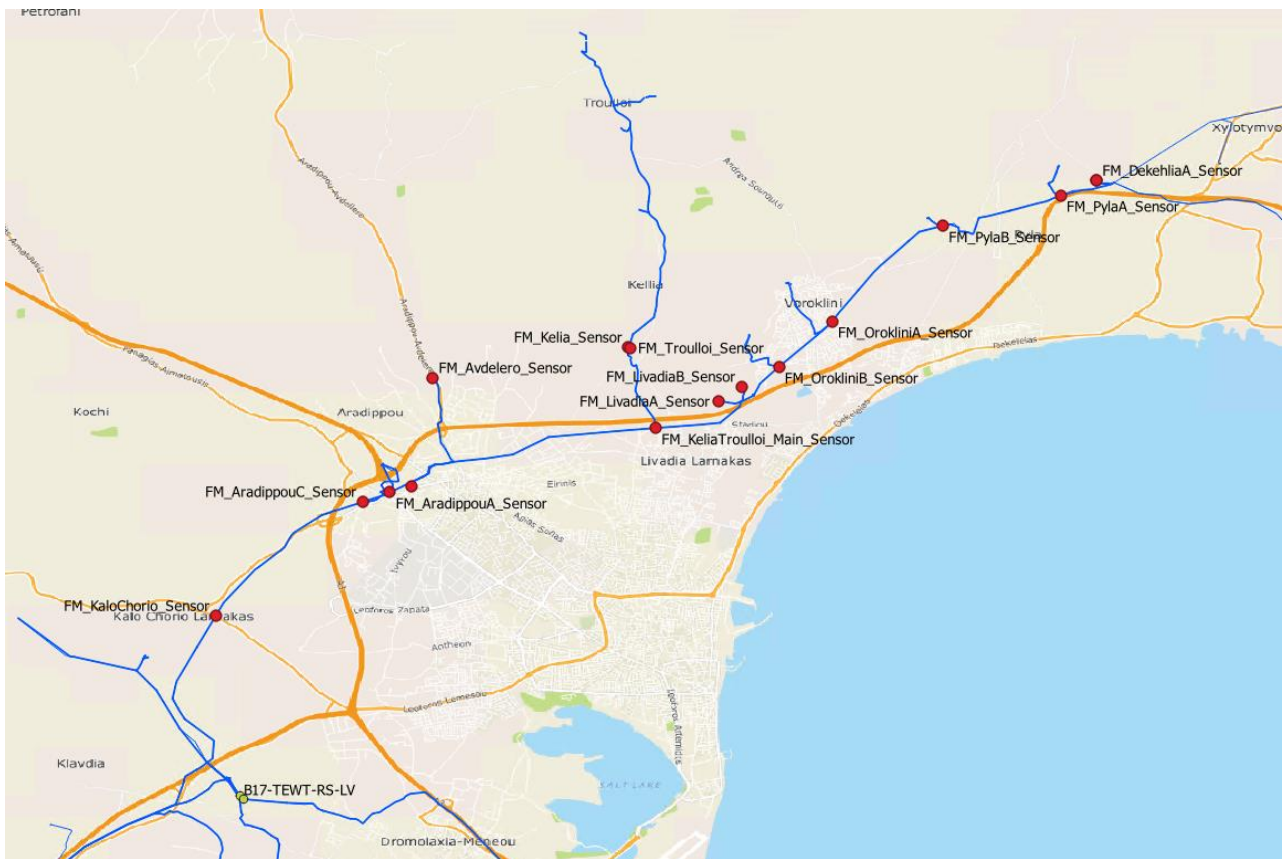
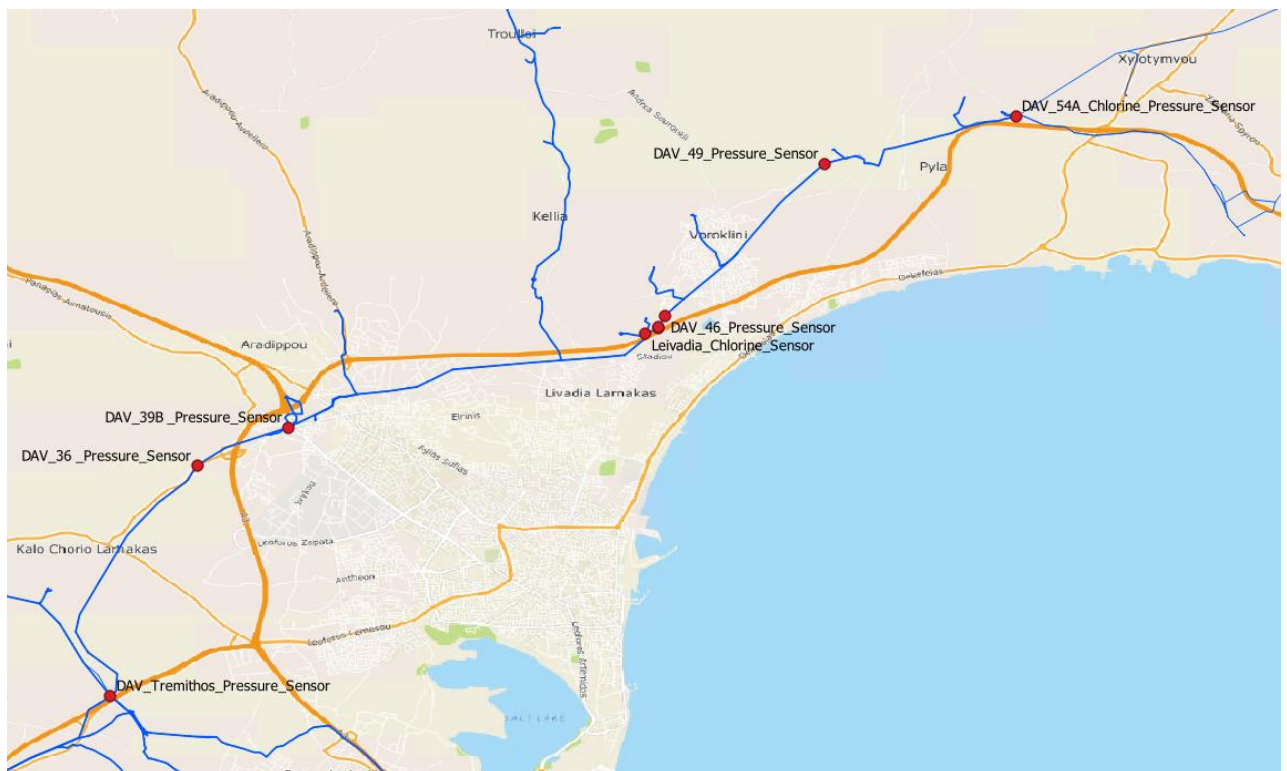*Figure 16: Location of AMR flow sensors in the WDD network*

Figure 17: Location of pressure sensors in the WDD network

The hydraulic state estimation was utilized for the one-week period between 24/08/2020 - 31/08/2020. The algorithm connects with the online server receiving data from the SCADA system and obtains the required measured data for the period 24/08/2020 - 31/08/2020 in order to execute the hydraulic state estimation. The data used for this pilot are the measured flows from each AMR (Figure 18) and the pressure measurements from the pump station located at Leivadia area (Figure 19). During the pilot phase period some of the AMRs were malfunctioning so an estimation of the hourly consumption was performed based on historical data.
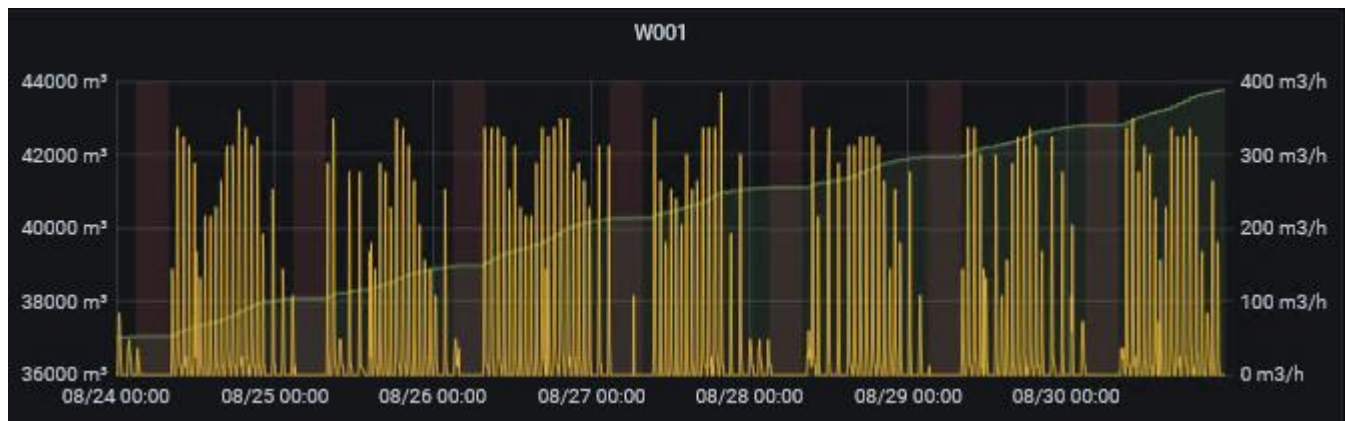

Figure 18: Measured flow from one of the AMR sensors located at WDD network for the period 24-31/08/2020.

Figure 19: Measured upstream and downstream pressure at the Leivadia Pump Station for the period 24-31/08/2020. Pump station was not operational during that time according to the general operational scenario of the entire network.

Initially, consumer demand calibration is performed using both flow measurements retrieved from the SCADA system and estimated consumptions (Figure 20). Then both the pressure and flow measurements are used to estimate the flow across the main pipeline and to calibrate the EPANET model (Figure 21).
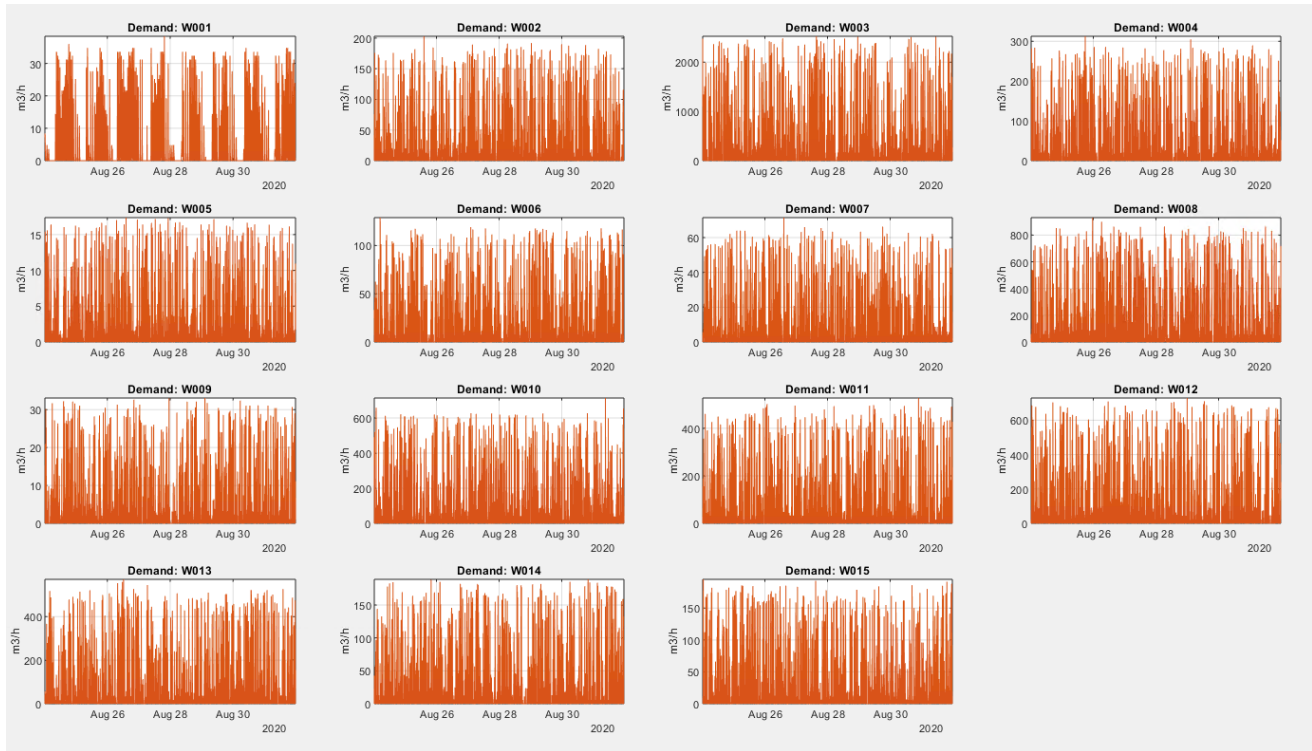
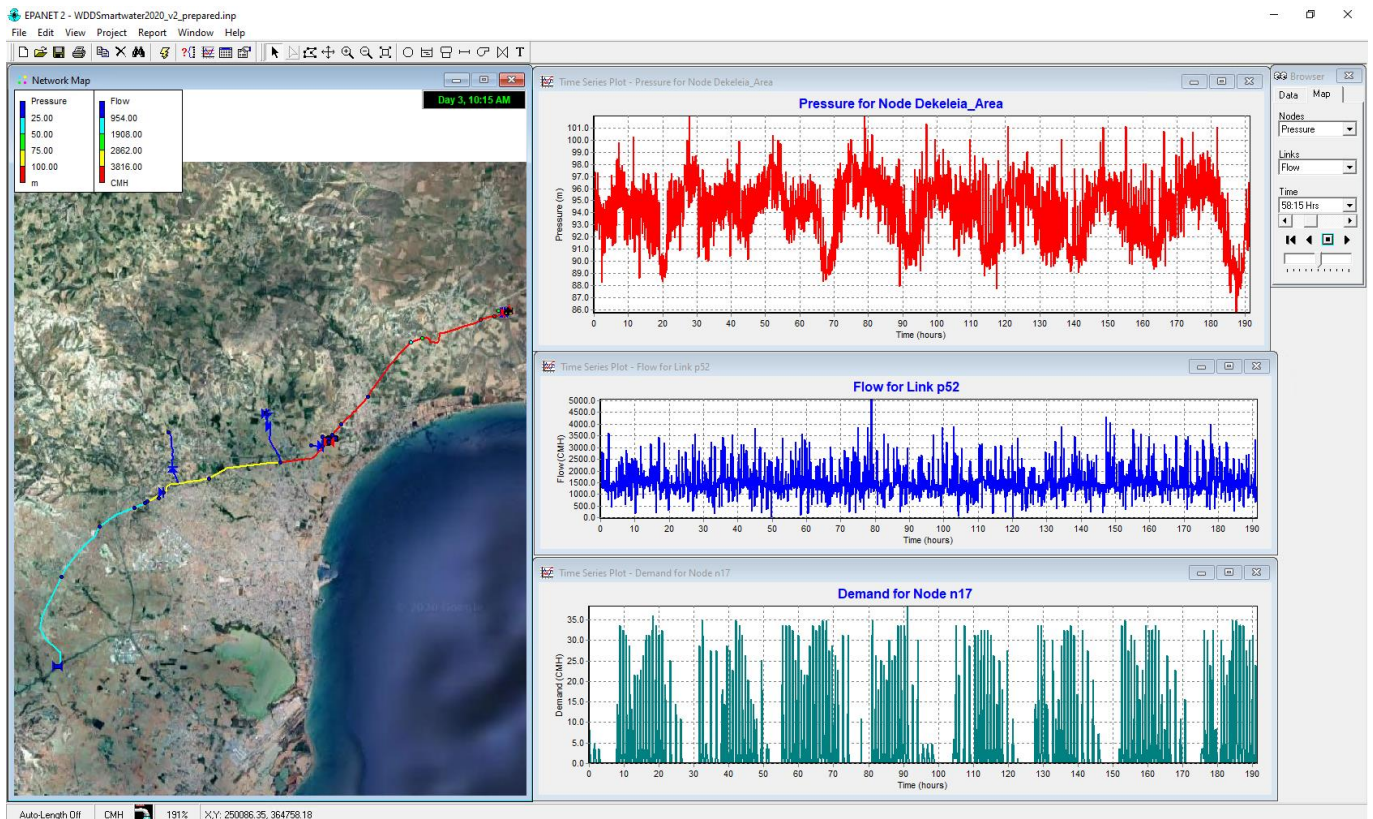*Figure 20: Estimated consumption (m³/h) of the AMR sensors based on historical data.*

Figure 21: Calibrated hydraulic model of WDD main pipeline in EPANET, augmented with measurements from the AMR and pressure sensors during the period 24-31/08/2020.

3.4 Conclusions

The results from the proposed methodology for hydraulic state estimation shows a high accuracy between the estimated and measured data. The main objective which was to predict and estimate the variation and periodic pattern of pressures and flows during selected periods was achieved. Differences between measured and estimated data were noticed indicating that calibration of the following parameters must be carried out more often:

- Sensors

- Model base demands

- Model pipe parameters (Status and roughness)

Sensors should be monitored by the corresponding water authorities and preserved in a good condition. The operation scheme including which pipes are open or closed should be updated regularly and specifically once a change is made on the network by the water authorities. Regarding the base demand – uncertainty, a solution could be the placement of additional AMRs on large consumers in order to measure a part of the actual network consumption.

4 Quality State Estimation (KIOS)

'Water Quality State Estimation' refers here to the process of estimating chlorine concentration using measurements from chlorine sensors installed at specific locations in the WDN to extrapolating the chlorine concentration at every location of the network. This is achieved using a water-quality model of the network in conjunction with a hydraulic model, since the hydraulic dynamics in these systems affect the water-quality dynamics, as illustrated in Figure 22. An accurate hydraulic-state estimate is imperative to achieve accurate estimation of chlorine concentration and this is why the hydraulic-state estimation methodology described in Section 3 is used to create a calibrated hydraulic model and then proceed to the calibration of the water-quality model.
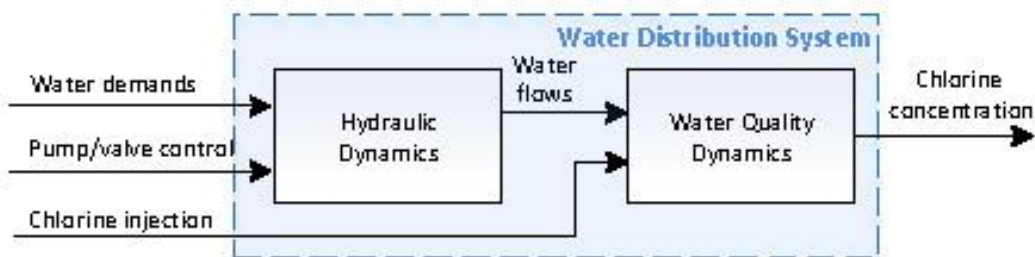

Figure 22: Quality State Estimation process

4.1 Methodology

The water-quality state estimation process is divided in steps and described below.

4.1.1 Hydraulic State Estimation

The methodology in Section 3 is used to acquire the hydraulic-state estimates for the selected period of time. This is available in the form of a calibrated hydraulic model derived using hydraulic sensor measurements. This model contains approximate values of pipe parameters related to water-quality such as bulk chlorine reaction rate and pipe wall reaction rates.

4.1.2 Chlorine measurement data retrieval and validation

Chlorine measurements from sensors located in the selected water network are retrieved through the platform. Through a Graphical User Interface (GUI) the user can select the desired time period for which all measurements from chlorine sensors installed in the selected water network are retrieved from the SCADA system. Chlorine sensors are typically located at the inlet point of each DMA network and at strategic locations inside the DMA network.

The retrieved sensor data are then pre-processed in order to detect outliers which may be the result of communication of sensor faults. Missing data points due to outliers or communication faults are recreated using interpolation of the available data points.

4.1.3 Water quality model creation and calibration

The calibrated hydraulic model is used as a basis to create a calibrated water quality model.

First, chlorine measurements at the inlet of the DMA network are augmented into the model by assigning the measured chlorine pattern at the inlet point of the model. To solve the first-order hyperbolic partial differential equations which describe chlorine transport and reaction in a WDN, the Lagrangian numerical approximation method implemented by the EPANET software is used. A water-quality simulation is performed using EPANET, for which the chlorine concentration at nodes and in pipes of the network is calculated based on the chlorine measurements at the inlet and the approximated reaction rate values that existed in the water-quality model.

The second step is to calibrate the model reaction rates using the chlorine measurements that exist inside the network. This is achieved by calculating the difference between the estimated and measured chlorine concentration at sensor locations and then adjusting accordingly the bulk and wall reaction rates of the model until this difference is minimized.

Note that if calibration of network parameters has been performed for an adequate amount of data, subsequent applications of the state estimation algorithm will not require the performance of this step.

4.1.4 Water-quality state estimation

Using the calibrated hydraulic and water-quality models, the chlorine concentration in pipes and at nodes are estimated throughout the network using the EPANET hydraulic and water-quality solver.

Function: *Chlorine_State_Estimation(model, data)*

Inputs:

@model: Hydraulic Calibrated water network model

@data: Chlorine measurements received from SCADA

Output:

@Calibrated_model: Inp. file of the calibrated water network

Function 4: Quality State Estimation algorithm inputs and outputs

Platform Communication:

- Communicates with EPANET database and retrieves the selected inp file

- Communicates with SCADA database and retrieves the required data for the selected period and the selected DMA/network

4.2 WBL Pilot

The proposed water-quality-state estimation methodology was applied on DMA 133 of the WDN of Limassol. The results of this pilot run are shown using sensor data from a 5 – day period between 19/11/2020 - 23/11/2020, however, any time period with available sensor data could have been selected.

Initially, the algorithm requests the execution of hydraulic-state estimation for the desired period 19/11/2020 - 23/11/2020 and acquires the calibrated hydraulic model of DMA 133 for that period. The inlet flow and pressure data which were used to create the hydraulic model for this period are shown in Figure 23.
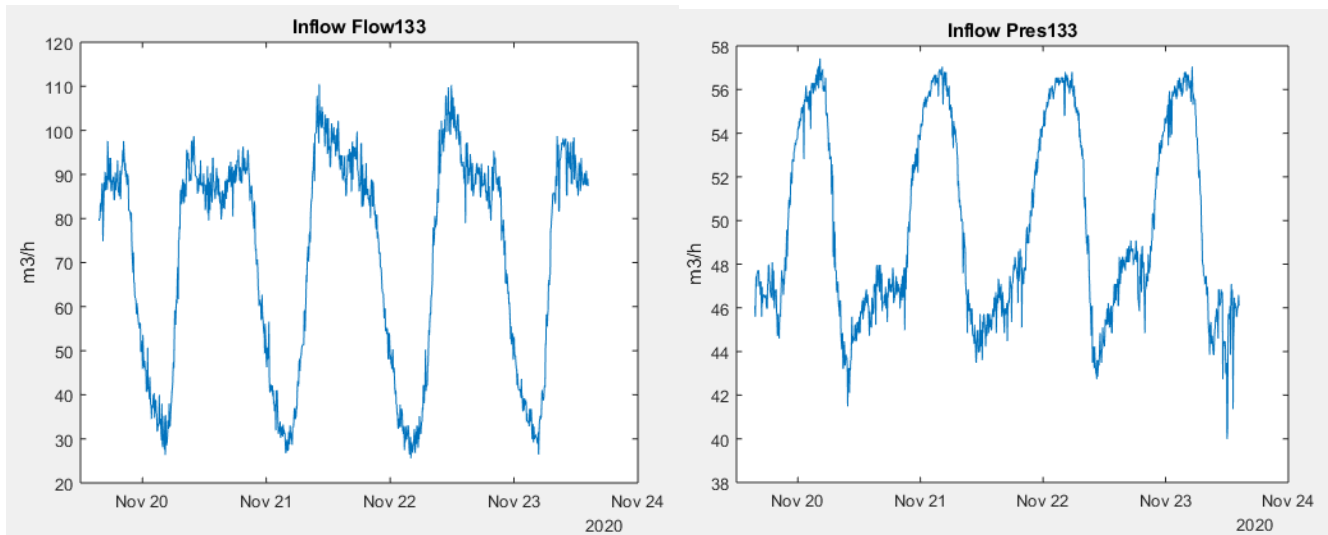

Figure 23: Measured inlet flow and pressure of DMA 133 for the simulation period 19- 23/11/2020

The algorithm then connects to the online server which hosts the SCADA system, obtains the chlorine measurements for the same time period and performs data validation and outlier detection. Due to the constant calibration needed by chlorine sensors, the retrieved measurements were compared and calibrated based on historical data and measurements from other sensors located at the same main pipeline. The acquired data of the inlet chlorine concentration of DMA 133 are visualized in Figure 24.
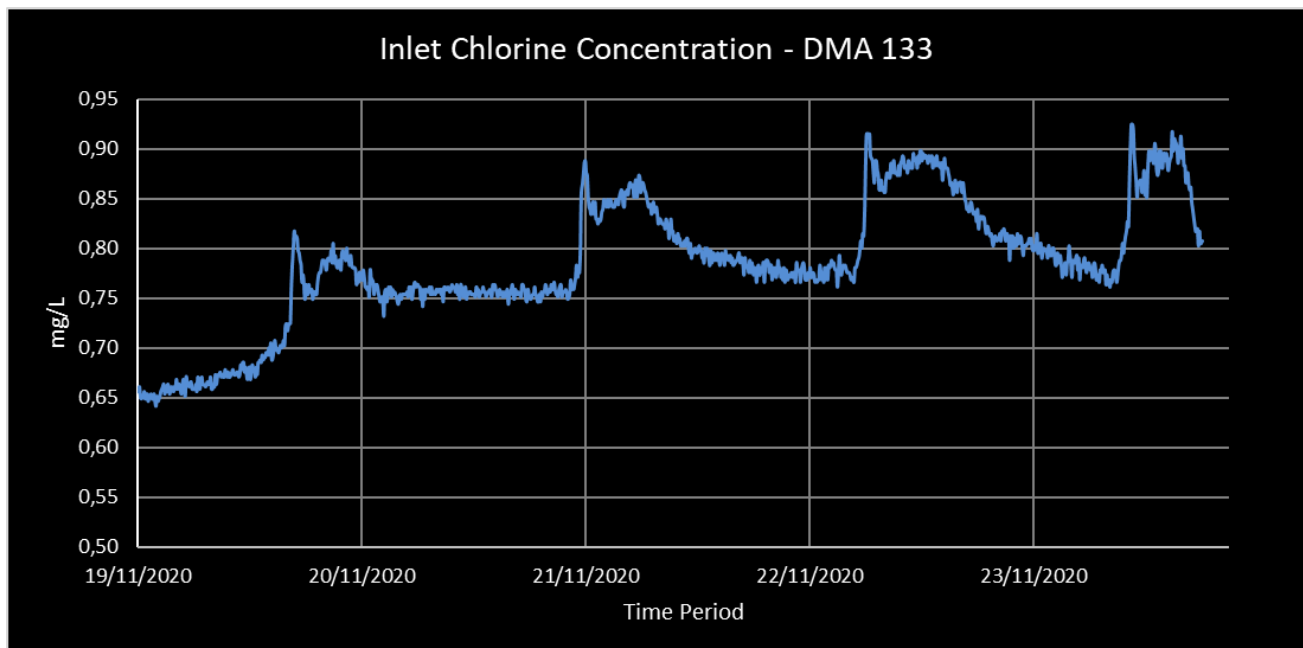
Figure 24: Measured inlet chlorine concentration of DMA 133 for the period 19/11/2020 - 23/11/2020.

The inlet chlorine concentration (Figure 24) is augmented with the available water-quality model of the network which also contains the calibrated hydraulic parameters. The resulting model is simulated and the estimations are illustrated in the EPANET software interface (Figure 25). It can be observed that using the simulated inlet chlorine concentration (Figure 25, upper graph), the chlorine concentration at given downstream nodes (4, 165, 239, 491, 572) is estimated. Moreover, the simulation results can be used to visualize the chlorine profile in multiple locations in the network (Figure 25, lower graph), to gain insight into the chlorine decay process in the network.
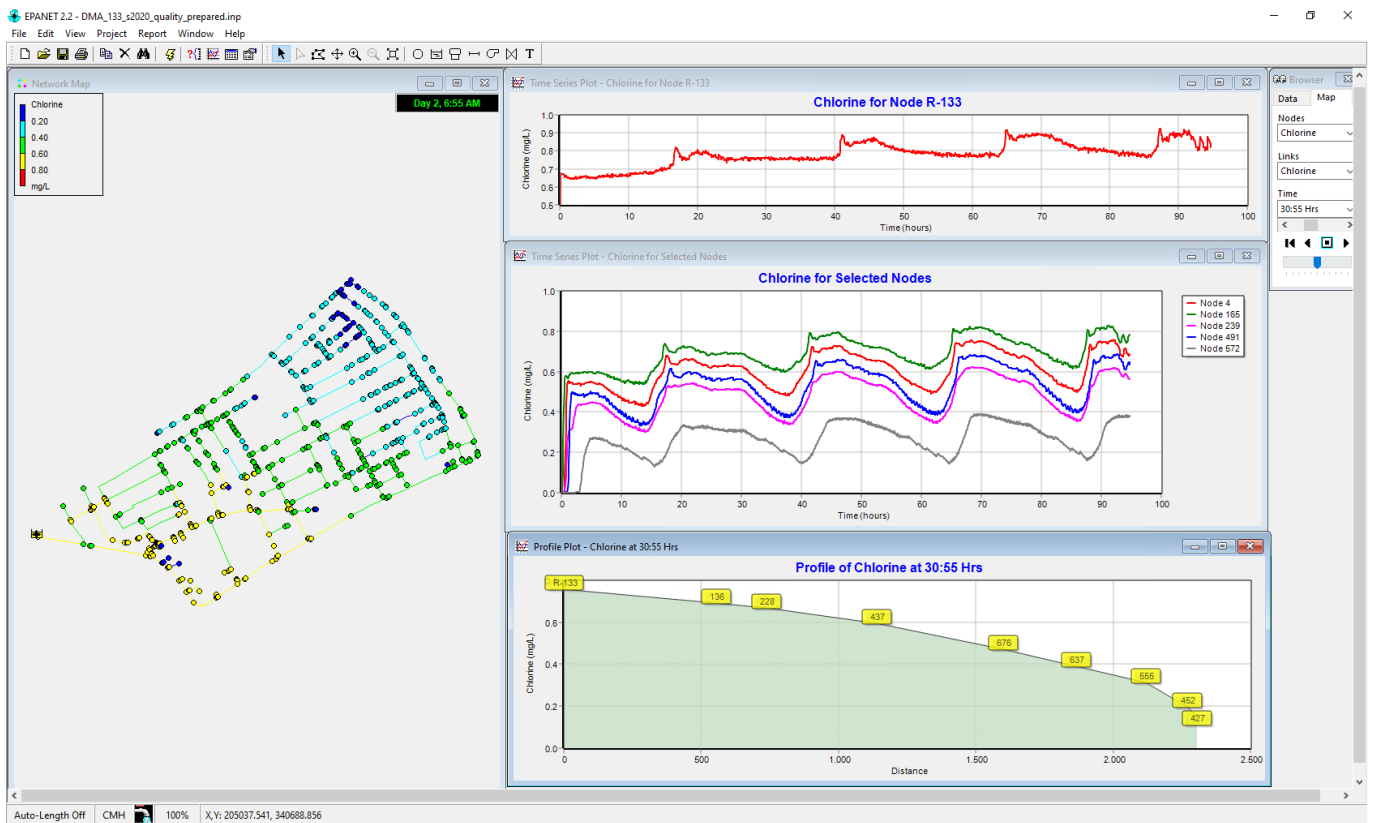
Figure 25: EPANET model with a complete quality system state of DMA 133 for the simulation period 19-23/11/2020.

The final step of this methodology was to compare simulated and measured chlorine data from sensors located in the network and were not used in the creation of the water-quality model. Figure 26 shows a GIS map with the piping system of DMA 133 and the location of the installed chlorine sensors.

In Figure 27 the measured (SCADA) and simulated (EPANET) chlorine concentration at nodes with installed sensors located at DMA 133 are presented. In all cases the estimated data are following the same pattern with the measurements, indicating the ability of the algorithm to predict the variation of chlorine across the network and throughout time. Regarding sensor "Chlor133P4" a difference between the measured and estimated data is observed meaning that either the actual sensor or a model parameter (bulk/wall coefficient or reaction rate) should be calibrated. Sensors "Chlor133P5" and "Chlor133P6" faced technical issues and were excluded from the process.

29

Figure 26: GIS map showing the location of the chlorine sensors (red) installed in DMA 133 and the locations where field measurements were taken (yellow)
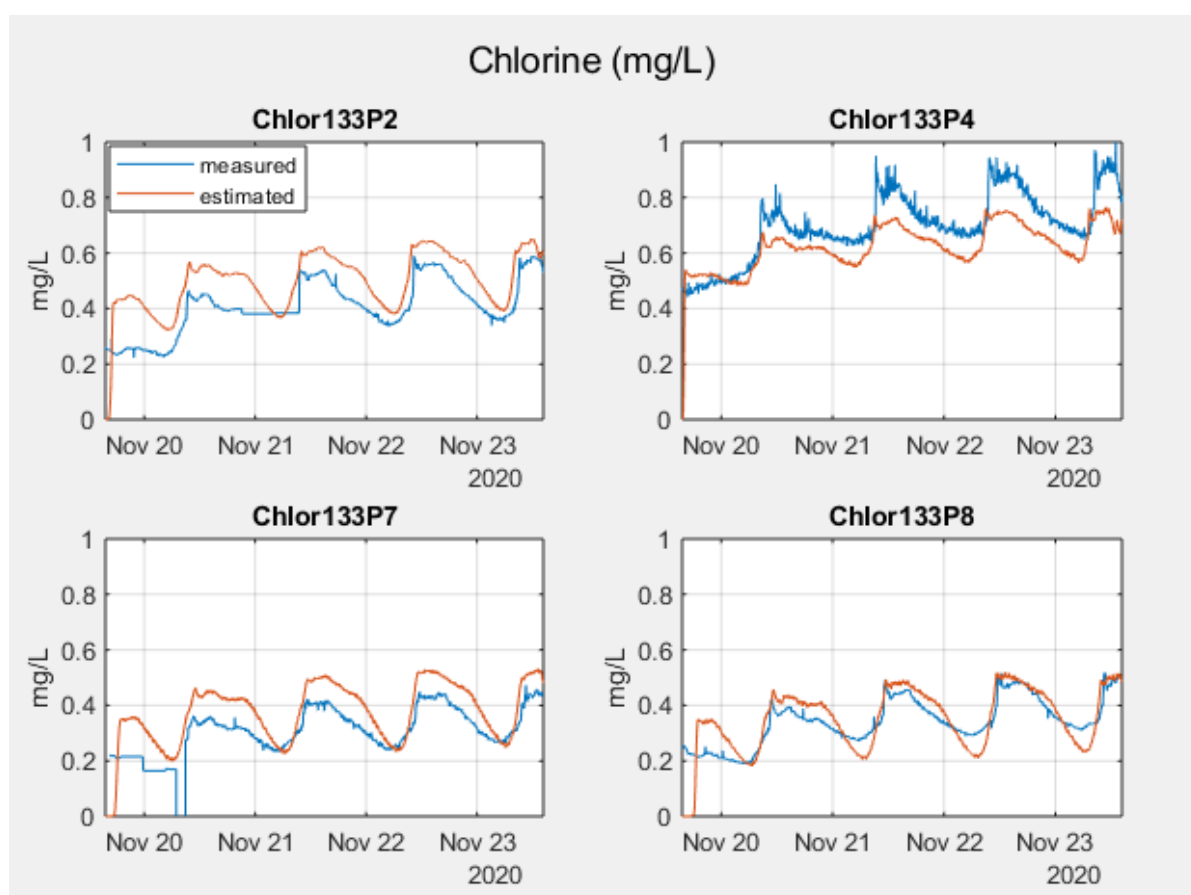

Figure 27: Measured (SCADA) and estimated (EPANET) chlorine from various sensors located in DMA 133 for the simulation period 19-23/11/2020.

Beside the comparison between SCADA and EPANET estimated data, field measurements were taken from specific locations (Figure 26) in the network in order to evaluate the accuracy of the quality state estimation application. Four measurements were taken (Table 3) from each location on 20/11/2020 between 8:20 am – 12:52 pm.

Table 3: Field chlorine measurements taken from several locations on DMA 133 on the 20/11/2020.

DMA133								
Point	Cl_2	time 1	Cl_2	time 2	Cl_2	time 3	Cl_2	time 4
DMA133								
4	0.23	8:20	0.29	9:55	0.39	11:20	0.35	12:15
165	0.43	8:30	0.43	10:00	0.43	11:25	0.42	12:20
239	0.40	8:40	0.49	10:10	0.50	11:30	0.48	12:25
491	0.37	8:50	0.48	10:15	0.48	11:35	0.45	12:32
572	0.35	8:57	0.53	10:20	0.42	11:42	0.49	12:40
657	0.34	9:05	0.37	10:25	0.47	11:48	0.43	12:45
510	0.30	9:10	0.33	10:30	0.35	11:53	0.34	12:48
306	0.32	9:15	0.37	10:35	0.34	12:00	0.36	12:52

In Figure 28, measured and estimated data of each location are presented. Nodes 239, 491, 657, 510 and 306 show a relatively good fit between measurements and simulated results while in nodes 4, 165 and 572 a deviation is noticed. In this case as well, either the actual sensor or a model parameter (bulk/wall coefficient or reaction rate) on specific pipes should be calibrated.
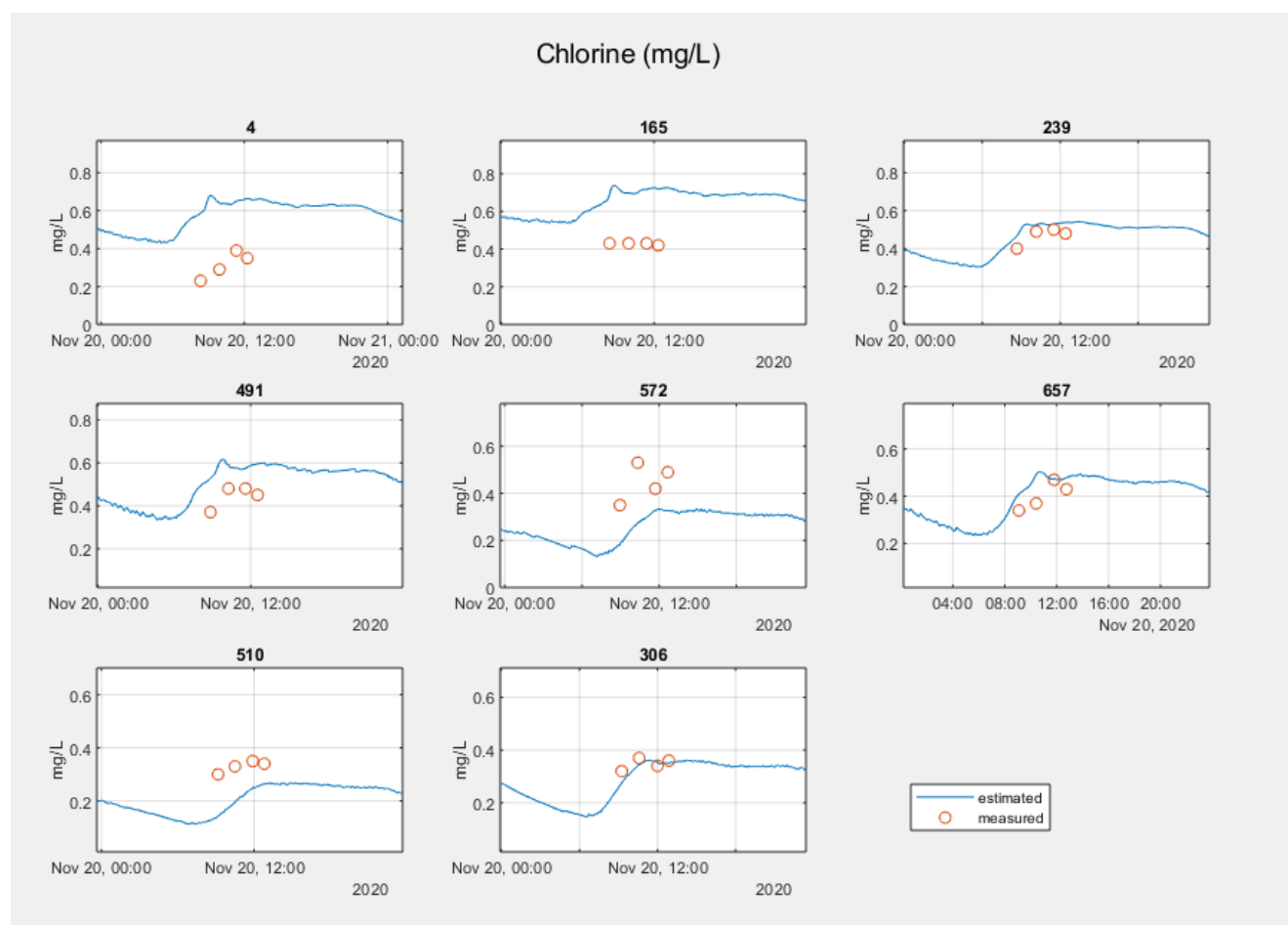

Figure 28: Measured (field measurements) and estimated (EPANET) chlorine from various locations in DMA 133 on 20/11/2020.

4.3 WDD Pilot

A quality state estimation pilot was performed in the Water Development Department (WDD) network using from a total of four (4) quality sensors which were installed across the network (Figure 29). The sensors comprise of two (2) chlorine concentration sensors and two (2) multiparameter sensors measuring chlorine concentration, pH, temperature, pressure and conductivity, as illustrated in Figure 30.

Figure 29: GIS map of the locations of chlorine and multi-parameter sensors in WDD network.

Figure 30: Measurement time-series from a multi-parameter sensor installed in the WDD network

Figure 31: Chlorine concentration measurements from the "Tersefanou" multiparameter sensor for the period 24/08/2020 - 31/08/2020.

For the estimation of the water-quality state of the pilot area, measurements from the "Tersefanou" multiparameter sensor were used, for the time period between 24/08/2020 - 31/08/2020.

First hydraulic-state estimation of the pilot network is performed for the indicated period, as described in Section 3, and the hydraulic model is calibrated.

Then, the water-quality model is calibrated using the chlorine concentration measurements from the inlet of the network of the examined period. During the pilot phase period, water flow from Dekheleia (Anzio Camp Area) to Larnaca in the main pipelines. The chlorine sensor DAV_54A located at Anzio Camp Area thus the inlet point of the network was not operational that time. However, an estimation of the inlet chlorine concentration was achieved using measurements from the "Tersefanou" multiparameter sensor, downloaded from the SCADA system (Figure 31) and historical data. A water-quality model with calibrated hydraulic parameters and including the inlet chlorine concentration is created as an EPANET input file and presented in Figure 33.

Using the EPANET model, a water-quality simulation of the network is performed by solving the network hydraulics and water-quality dynamics. The chlorine concentration at any part of the network can then be estimated, e.g., Node n17 of the model as shown in Figure 33. A profile plot of the chlorine concentration in multiple locations of the network was also constructed showing the decay of chlorine in the network (Figure 33, lower graph).
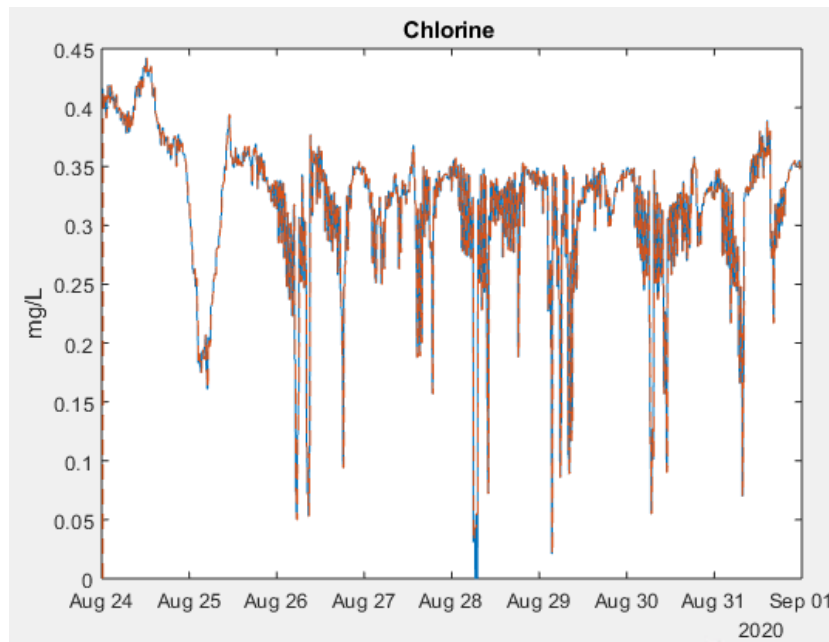
Figure 32: Inlet chlorine concentration of WDD network between 24/08/2020 - 31/08/2020.
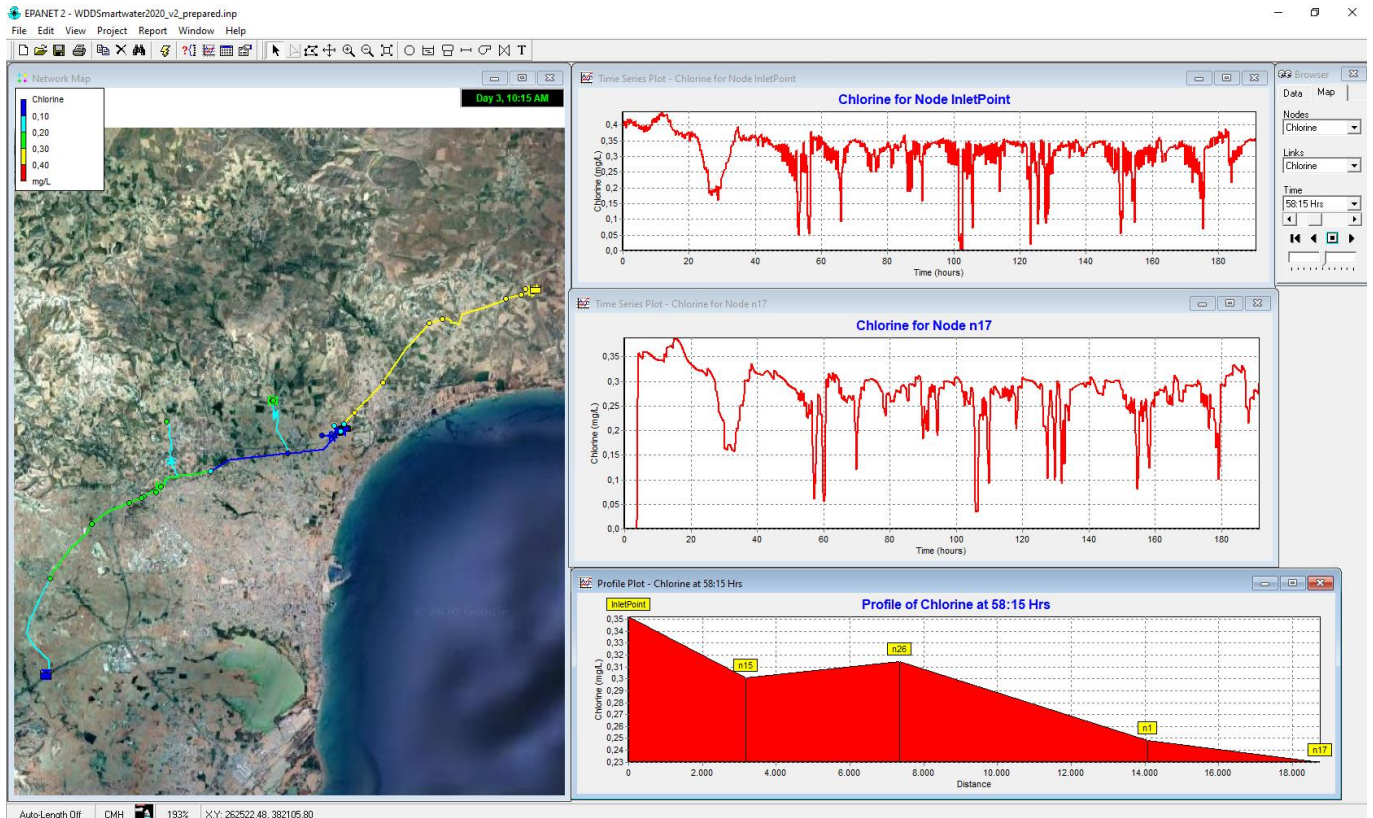

Figure 33: Hydraulic and water quality model of the WDD network in EPANET, augmented with measurements from the inlet chlorine measurements during the period 12-18/10/2020. 24-31/08/2020.

The water-quality parameters measured from the multiparametric sensors were analyzed in order to identify correlations between them which would help add insight into the measured quantities. In Table 4 we can observe the time-series correlation of measured parameters, where it is identified that there is a high inverse relationship between the free chlorine in water and the water temperature. This is explained by the fact that the decay rate of chlorine increases when temperature increases. Moreover, we observe a high correlation between temperature and conductivity,

which is explained by the fact that many salts present in water are more soluble in higher temperatures, thus increasing the water conductivity. A high correlation is also observed between conductivity and chlorine because when chlorine is introduced into the water, the quantity of electrolytes or total dissolved solids in the water rises, which in turn raises the conductivity of the water. Finally, the pH levels of the water appear to have low correlation with the other parameters.

Table 4: Time-series correlation of parameters measured from multiparametric sensors

Parameters	pH	Conductivity	Chlorine	Temperature
pH	100 %	0.007 %	18.53 %	-18.40 %
Conductivity	0.007 %	100 %	-48.44 %	63.34 %
Chlorine	18.53 %	-48.44 %	100 %	-74.29 %
Temperature	-18.40 %	63.34 %	-74.29 %	100 %

- PCA/t-SNE -> 2D

4.4 Conclusions

The results from the proposed methodology for quality state estimation shows a high accuracy between the estimated and measured data. The main objective which was to predict and estimate the variation and periodic pattern of pressures and flows during selected periods was achieved. Differences between measured and estimated data were noticed indicating that calibration of the following parameters must be carried out more often:

- Chlorine sensors
- Model pipe parameters (bulk and wall coefficients)

Sensors should be monitored by the corresponding water authorities and preserved in a good condition. Model pipe parameters should be calibrated more often since the bulk and wall coefficients used in this methodology depend from the condition each pipe which can be estimated by the corresponding age and material.

5 Detection and Localization of Leakages and Irregular Flow

In this section three different methodologies for leakage diagnosis and detection of irregular flow are presented and applied on three different pilots. The methodologies and the conditions for which they are applied are described below and results from the pilots runs are presented.

5.1 Leakage detection and localization using pressure measurements

The leakage detection and localization using pressure measurements methodology is applied in DMA network setups of which the inlet flow and pressure are measured and additional pressure

sensors are installed in strategic locations inside the DMA. The algorithm is able to detect abnormal pressure behavior which signals the existence of a leakage in the network and using information about the topology and location of the pressure sensor is able to roughly determine the area in which the leakage may exist. The methodology is based on adaptively approximating the unknown, time-varying, weekly periodic difference between the measured and estimated pressure at sensing locations, by updating the coefficients of Fourier series basis functions (Eliades and Polycarpou, 2012).

5.1.1 Methodology

The first step of the methodology is to use the hydraulic-state estimation methodology described in Section 3 to obtain a calibrated hydraulic-model of the network based on data received for the previous seven days (one week) of the day under examination. These data are considered to represent the "healthy" state of the system where no leakages exist.

Since this model does not exactly predict the measured pressures (see Figure 14), an upper bound on the model uncertainty needs to be estimated in order to be used for fault-diagnosis. This is achieved by first generating the difference between the measured and estimated pressure of each sensor i , hereby referred to as the error signal $e_i^w(k) = p(k) - \hat{p}_i(k)$, for the one week period before the day of interest.

By observing the error signal, we form the assumption that it exhibits weekly periodicity which may be caused by the pressure dependence of water demands during this one week period. Due to this assumption, a suitable approximation structure for providing an estimate of the error signal for each measurement $\hat{e}_i^w(k)$ is a Fourier Series of the form:

$$\hat{e}_i^w(k) = \alpha_0 + a_1 \sin(k \, w_f) + \cdots + \, a_n \sin(n \, k \, w_f) + \beta_0 + \beta_1 \cos(k \, w_f) + \cdots + \, \beta_n \cos(n \, k \, w_f),$$

with weekly periodicity, i.e., the period:

$$T_f = 7 \, \cdot 24 \cdot 60/t\_h \, ,$$

$$w_f = \frac{2\pi}{T_f},$$

where t_h is the measurement and hydraulic time-step in minutes. The Fourier Series coefficients are calculated by minimizing the square difference $e_i^w(k) - \hat{e}_i^w(k)$.

The error difference $e_i^w(k) - \hat{e}_i^w(k)$ is then statistically analyzed to find the mean μ_i and standard deviation σ_i. Typically, the error difference will have a zero mean. The upper bound on the error $e_i^w(k) - \hat{e}_i^w(k)$ is defined as five standard deviations from the mean, indicated here by $\bar{\eta}_i = 5\sigma_i$. The condition for "normal" pressure readings $p_i(k)$ is then given by using the following adaptive threshold:

$$|e_i(k) - \hat{e}_i^w(k) + \mu_i| \le \bar{\eta}_i$$

$$|p_i(k) - \hat{p}_i(k) - \hat{e}_i^w(k) + \mu_i| \le \bar{\eta}_i$$

The next step is to receive the pressure sensor data for the day of interest, which for better results should be the next day of the one-week period used to create the adaptive threshold. A test is then performed which assesses if, for all time steps and for all sensors, the pressure readings satisfy the normal operation condition. The pressure readings are considered healthy (leakage free) if:

$$|p(k) - \hat{p}_i(k) - \hat{e}_i^w(k) + \mu_i| \le \bar{\eta}_i, \forall i, k$$

36

Using additional time-series analysis, an alert is created for every sensor that does not satisfy the above condition for at least N number of time-steps during the examined period.

The **leakage localization step** is then performed offline by using the information of the location of pressure sensors for which an alert was issued. Using the available hydraulic model, multiple leakage scenarios are simulated until the best leakage location and magnitude is found which fits the observed pressure variations.

Function: *Leak_Pressure_Fourier_Algorithm (model, data, M_timesteps)*

Inputs:

@model: Calibrated water network model (.inp file)

@data: Flow measurements received from SCADA

@M_timesteps: Number of detection time steps for which an Alert is activated

Output:

@Alert: Node ID/sensor ID, measured-modelled pressure fault, datetime

@plot: A graph of the network with the high-risk nodes/pipes highlighted

---

Function 5: Leakage detection using pressure measurements and Fourier series algorithm outline

Platform Communication:

- Communicates with EPANET database and retrieves the selected calibrated inp file

- Communicates with SCADA database and retrieves the required data for the selected period and the selected DMA/network

5.1.2 Application on the BattLeDIM competition

The BattLeDIM competition was designed as a realistic benchmark for methodologies applied to networks with pressure sensors installed. The "L-Town" network used in this competition and the location of the pressure sensors is shown in Figure 34.
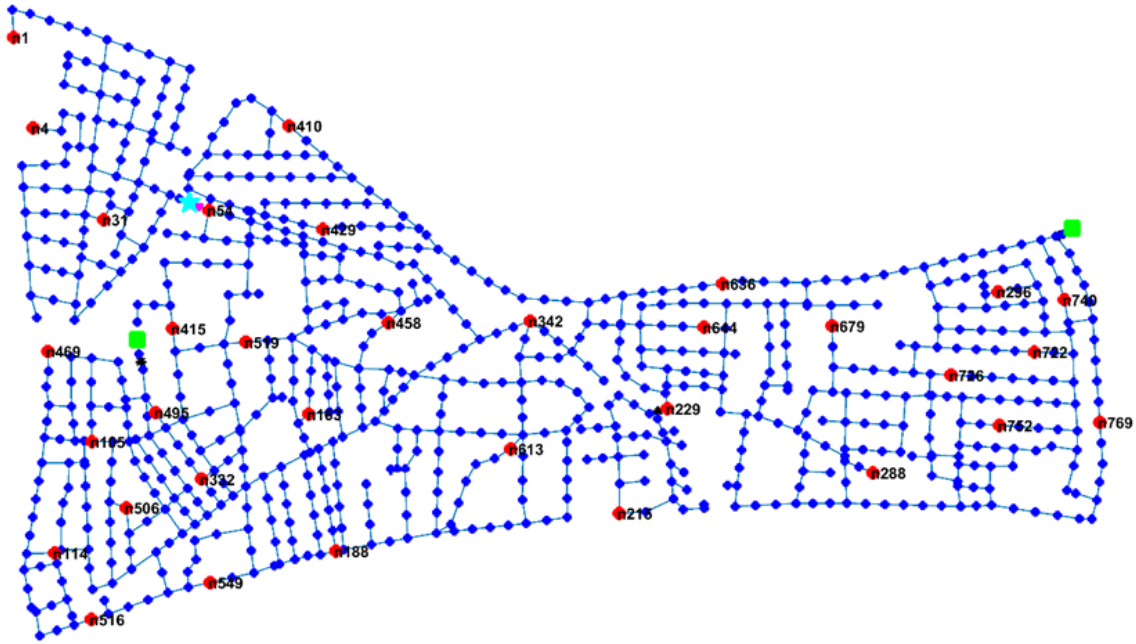
Figure 34: Location of 33 pressure sensors installed in "L-Town", BattLeDIM competition network

The methodology described in Section 5.1 was tested on the BattLeDIM benchmark by first defining a leakage of known characteristics, time and location to be simulated. The configuration file for producing the artificial data which will be provided by the virtual SCADA system and the location of the leakage are illustrated in Figure 35.
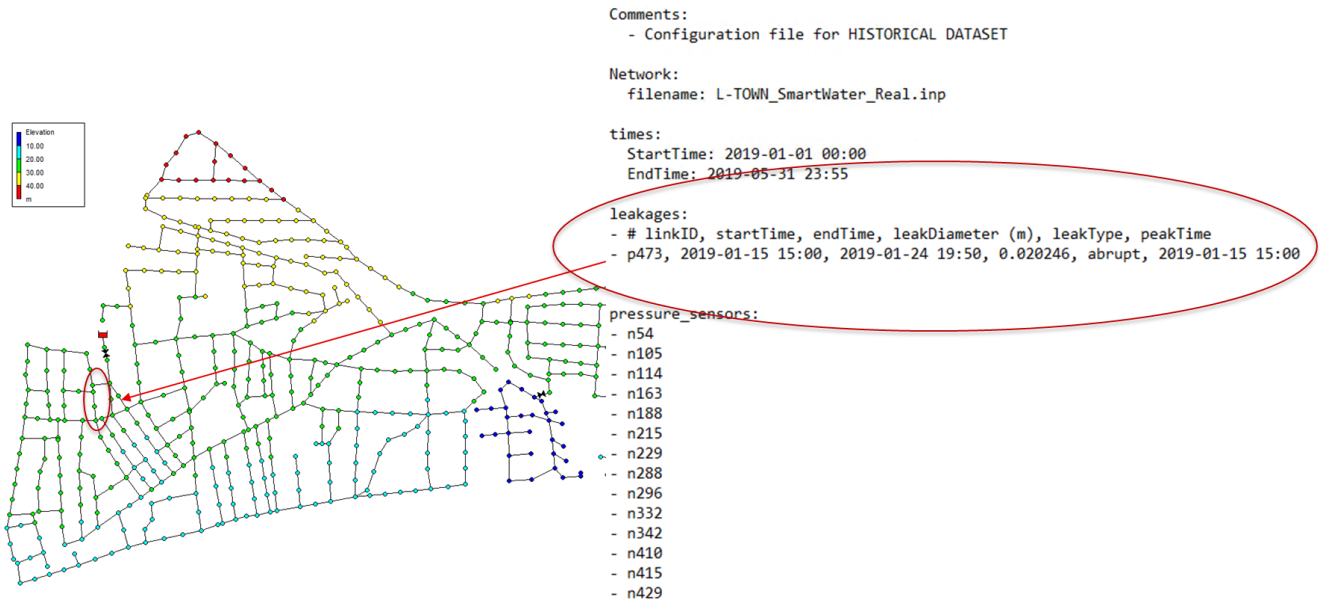


```
Comments:
  - Configuration file for HISTORICAL DATASET

Network:
  filename: L-TOWN_SmartWater_Real.inp

times:
  StartTime: 2019-01-01 00:00
  EndTime: 2019-05-31 23:55

leakages:
  - # linkID, startTime, endTime, leakDiameter (m), leakType, peakTime
  - p473, 2019-01-15 15:00, 2019-01-24 19:50, 0.020246, abrupt, 2019-01-15 15:00

pressure_sensors:
  - n54
  - n105
  - n114
  - n163
  - n188
  - n215
  - n229
  - n288
  - n296
  - n332
  - n342
  - n410
  - n415
  - n429
```

Figure 35: The BattLeDIM configuration file and the location of the artificial leakage on the network

The first step of the methodology was to learn the uncertainty of the provided hydraulic-model of the network. The difference between estimated and measured pressures at sensor locations is calculated for the one week period prior to the leakage occurrence. The difference is then approximated using Fourier Series (FS) as basis functions and by calculating the most suitable coefficients. The results of this procedure for the pressure sensor at node n506 is illustrated in Figure 36 (Left).

38

The next step is to calculate an upper bound on the error of the approximation using FS and use this a threshold for the detection of leakages that may occur. The pressure sensor data are then received from the virtual SCADA system for the day when the leakage occurs. It can be observed from Figure 36 (Right) that the expected difference between measurements and estimation deviates significantly for the sensor at n506 after the occurrence of the leakage. This deviation violates the defined threshold for this sensor, as illustrated in Figure 37, and thus the leakage is detected.

The localization phase takes into account the threshold violations observed at all the pressure sensors. As it can be observed from Figure 38, a deviation which violated the threshold occurred at a number of pressure sensor locations in the network. By further analyzing the frequency and magnitude of threshold violations at each location and using hydraulic simulations an area of the network is identified as possibly containing a leakage, as shown in Figure 39, which indeed contains the network pipe on which the leakage was induced.
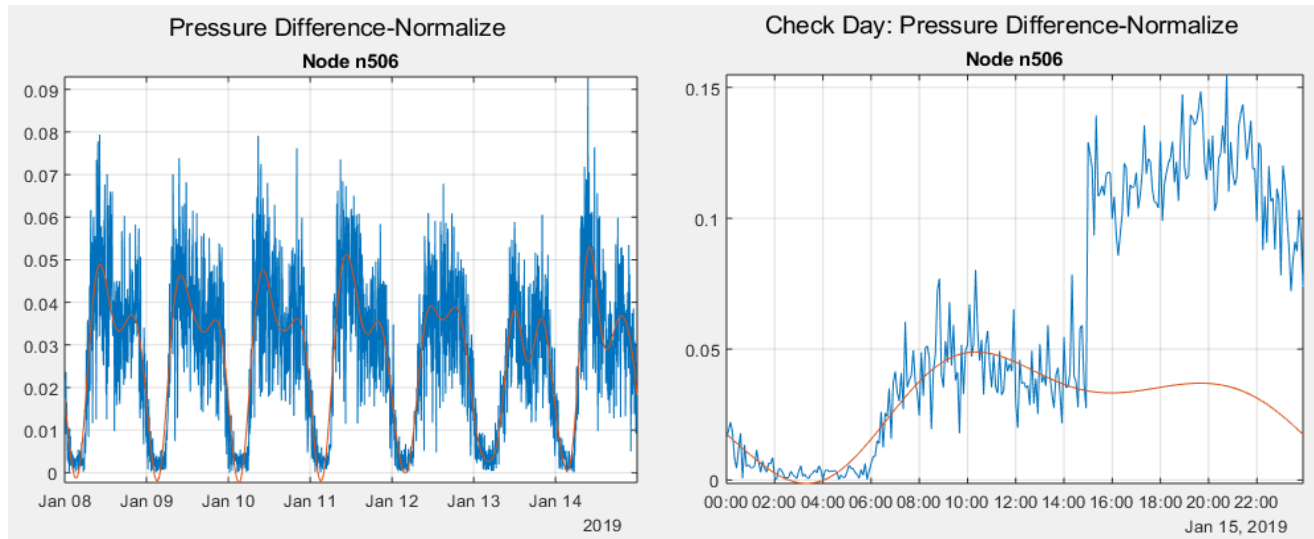

Figure 36: Left: Approximated (red line) and actual (blue line) difference between measured and estimated pressure of Node n506 for a one-week period before the leakage. Right: The same signals during the day of the leakage.

	1	2	3	4	5	6	7
	Node_ID	Sensor_ID	Error	Threshold	Error2	Threshold2	Date
1	'n506'	'n506'	0.0951	0.0502	0.1291	0.1216	'15-Jan-2019 15:00:00'
2	'n506'	'n506'	0.0906	0.0502	0.1245	0.1216	'15-Jan-2019 15:05:00'
3	'n506'	'n506'	0.1057	0.0502	0.1393	0.1216	'15-Jan-2019 15:20:00'
4	'n506'	'n506'	0.1011	0.0502	0.1356	0.1216	'15-Jan-2019 17:20:00'
5	'n506'	'n506'	0.0881	0.0502	0.1228	0.1216	'15-Jan-2019 17:30:00'
6	'n506'	'n506'	0.0868	0.0502	0.1216	0.1216	'15-Jan-2019 17:35:00'

Figure 37: The values of the error and the violated threshold for the pressure sensor at node n506 after the occurrence of the leakage
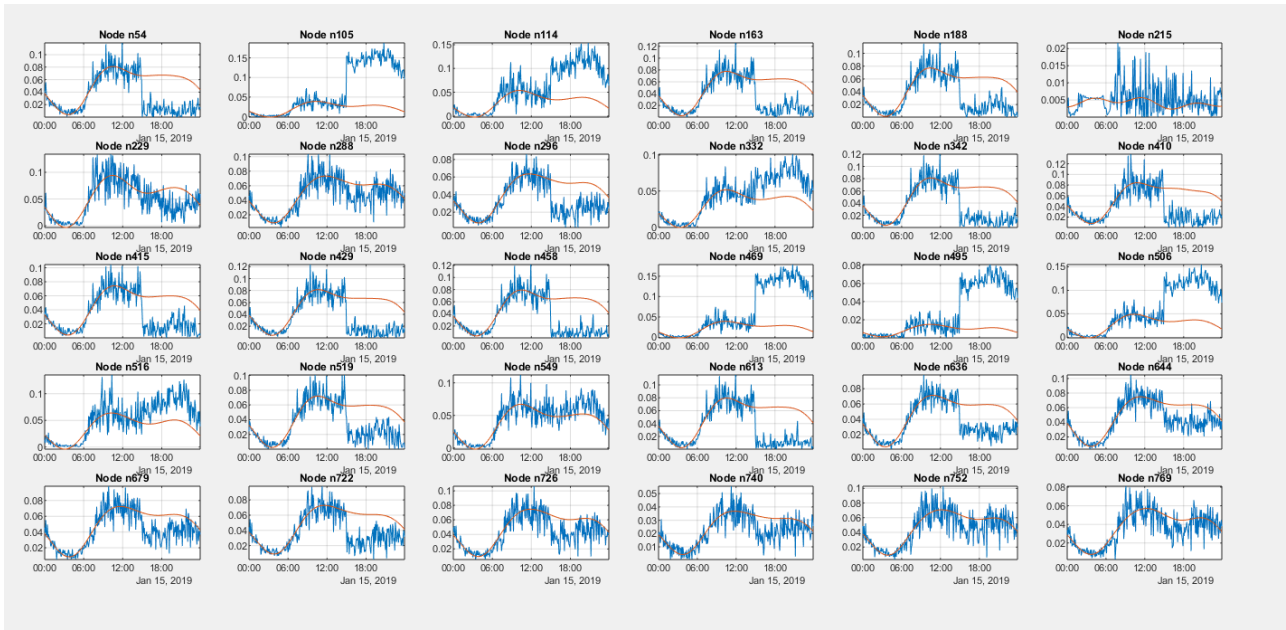
Figure 38: Approximated (red line) and actual (blue line) difference between measured and estimated pressures from 30 sensors during the day of the leakage.
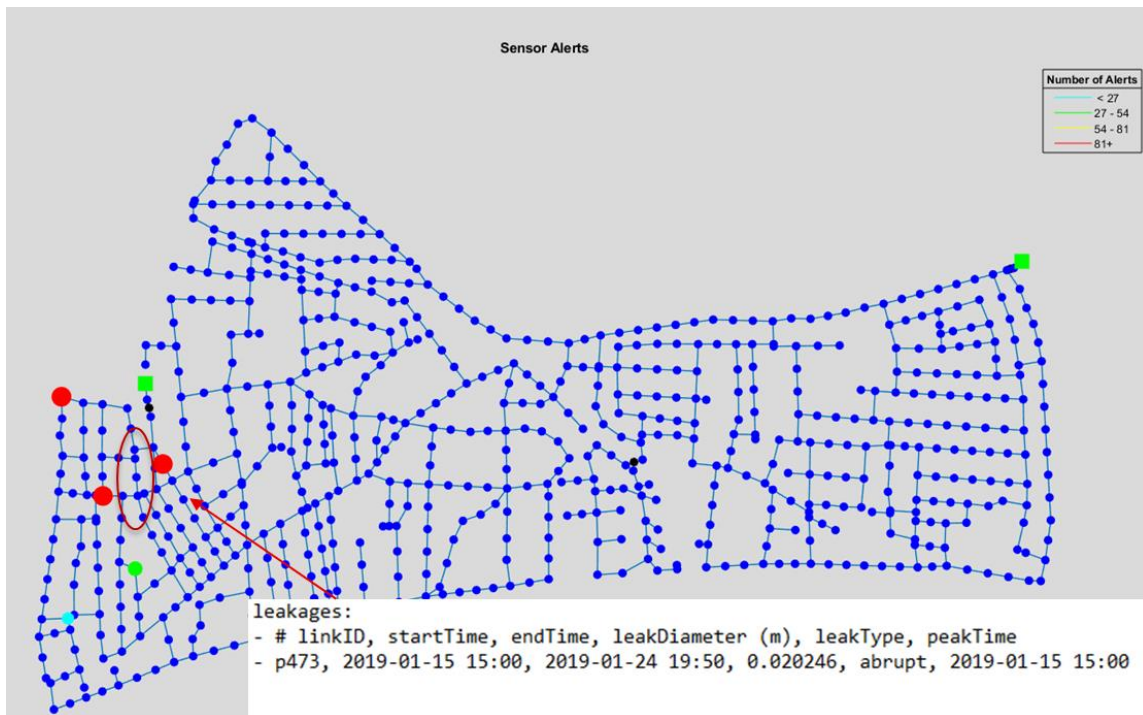

Figure 39: The area of the network identified to contain a leakage after analyzing the signals from all the pressure sensor.

5.1.3 WBL Pilot

The proposed algorithm is integrated in SmartWater2020 platform and activated every day. A table with alerts was created in the platform and is updated every day. An alert was triggered on "Pres136P3" sensor on 13th of November indicating abnormal pressure. Figure 40, retrieved through the algorithm, shows the weekly pressure variation signal of each senor based on previous week's data. Figure 41 shows the pressure variation signal during the examined day compared to the corresponding data of the same day of the previous week. As can be seen, the two

datasets regarding "Pres136P3" sensor are diverging. An alert is triggered, and the sensor locations are highlighted on the network map (Figure 42).
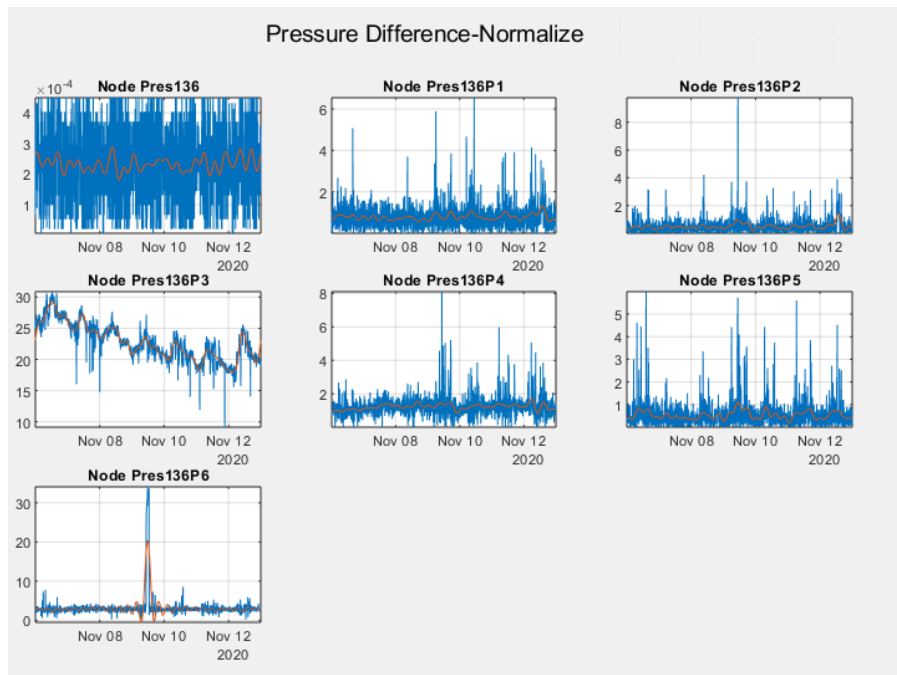

Figure 40: Pressure weekly periodic signal of DMA 136 sensors between 6-12/11/2020
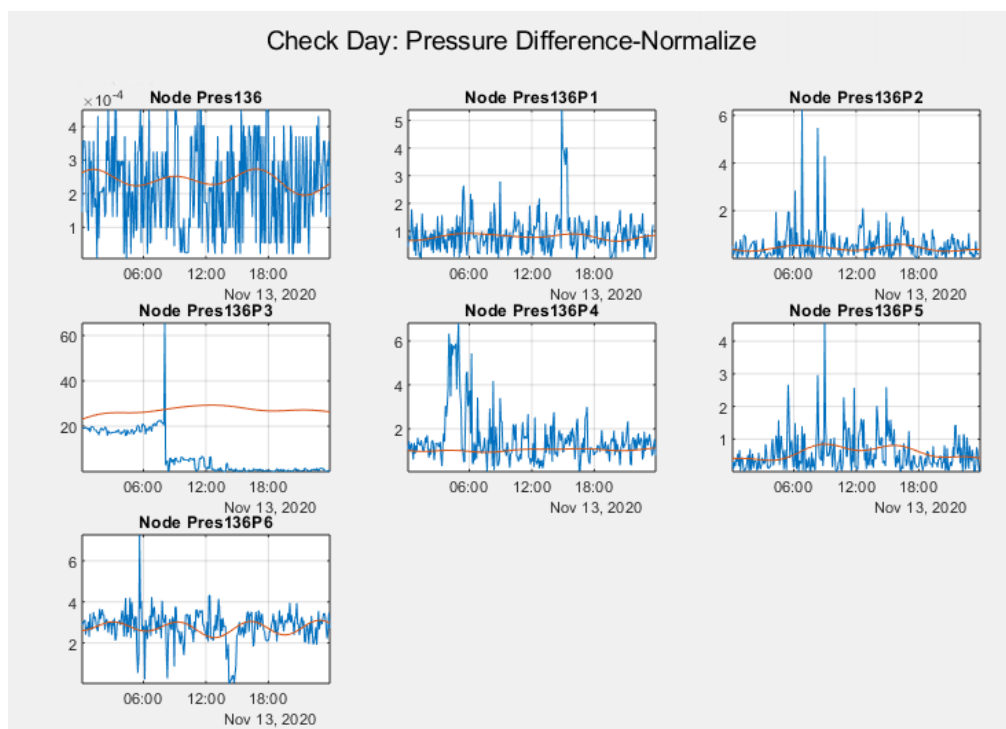

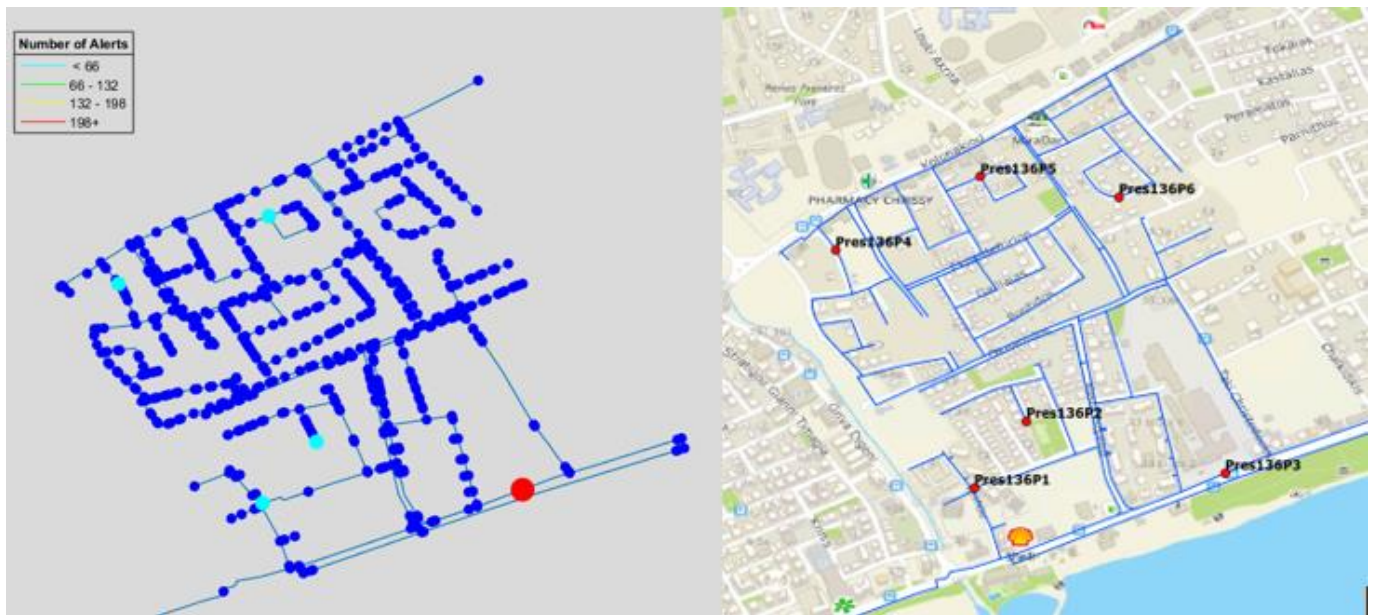*Figure 41: Pressure periodic signal of DMA 136 sensors on 13/11/2020*

Figure 42: Left: Sensor locations with abnormal pressure behavior. Right: GIS map showing the location of the pressure sensors installed in DMA 136

Figure 43 presents the pressure measurements of sensor "Pres136P3" retrieved from SCADA system which indicate that on 13th of November a pressure drop from 70 m to 45 m took place. As explained by the WBL, the pressure dropped as a result of sensor calibration.

Figure 43: Measured pressure of Pres136P3 sensor of DMA 136 for the period 7-18/11/2020.

5.2 Irregular flow detection at consumer level AMRs

In the Larnaca Water Board (LWB) pilot area, AMRs were installed which monitor a variety of consumers (industrial, commercial and residential) and provide measurements of the consumed water volume every one hour. Data are received every 12 hours as a package including all measurements from AMRs. For these consumer-level measurements, a methodology was developed for identifying irregular flow which may be due to a number of reasons such as leakage, abnormal consumption due to a social event or construction work, or even a forgotten running faucet.

5.2.1 Methodology

A real-time detection algorithm is developed which retrieves data of the past eight (8) days and is activated every 12 or 24 hours depending on the user settings. The data of the first 7 days serve

as "historical data" in order to create specific thresholds while the data of the 8th day to detect leakages during that day.

As a first step, the algorithm calculates the minimum flow of each day of the "historical data" period of 7 days. In this case, due to the uniqueness of each consumer (some consumers represents industrial industries which operate for a specific period both during the day and night) the algorithm estimates the minimum flow of the entire day and not just during the night time as it is common for leakage detection algorithms which assess the Minimum Night Flow (MNF).

The second step is defining a threshold by computing the average and the standard deviation of the extracted minimum flow of the "historical data" period. The threshold is based on the following equation (Claudio *et al.*, 2015):

$$\text{Threshold} = MA + a*MSD;$$

where MA is the average, MSD is the standard deviation and a is a statistical quality metric for detection performance assessment (Farah and Shahrour, 2018).

Finally, data are received for the "check-day", or the 8th day of the examined data, and the minimum flow is calculated. The minimum flow is compared with the threshold created in the previous step and sends a warning when this is violated, in which case a flow anomaly is detected. The warning includes the AMR ID, the minimum flow of the day and the time it was recorded.

Platform Communication:

- Communicates with SCADA database and retrieves the required data (AMR measurements) for the selected period and the selected DMA/network

5.2.2 LWB Pilot

The proposed algorithm is integrated in SmartWater2020 platform and activated once a day when the data from AMRs are received. A table with alerts was created in the platform and is updated every day when data are received.

Figure 44 illustrates an example of abnormal consumption in the LWB network. An alert was triggered on "water-meter 97" located at a public school showing that after the 1st of October abnormal consumption is detected. It is observed that the minimum consumption increased from approximately 45 m³/h to 1300 m³/h and remained at the same levels in the following days including during the weekend when public schools are normally closed. Eventually, the consumption dropped back to normal level after 10 days. Based on information received from the LWB, the event that caused this abnormal consumption was either a faucet which was left running or a leakage which was subsequently fixed by the consumer.
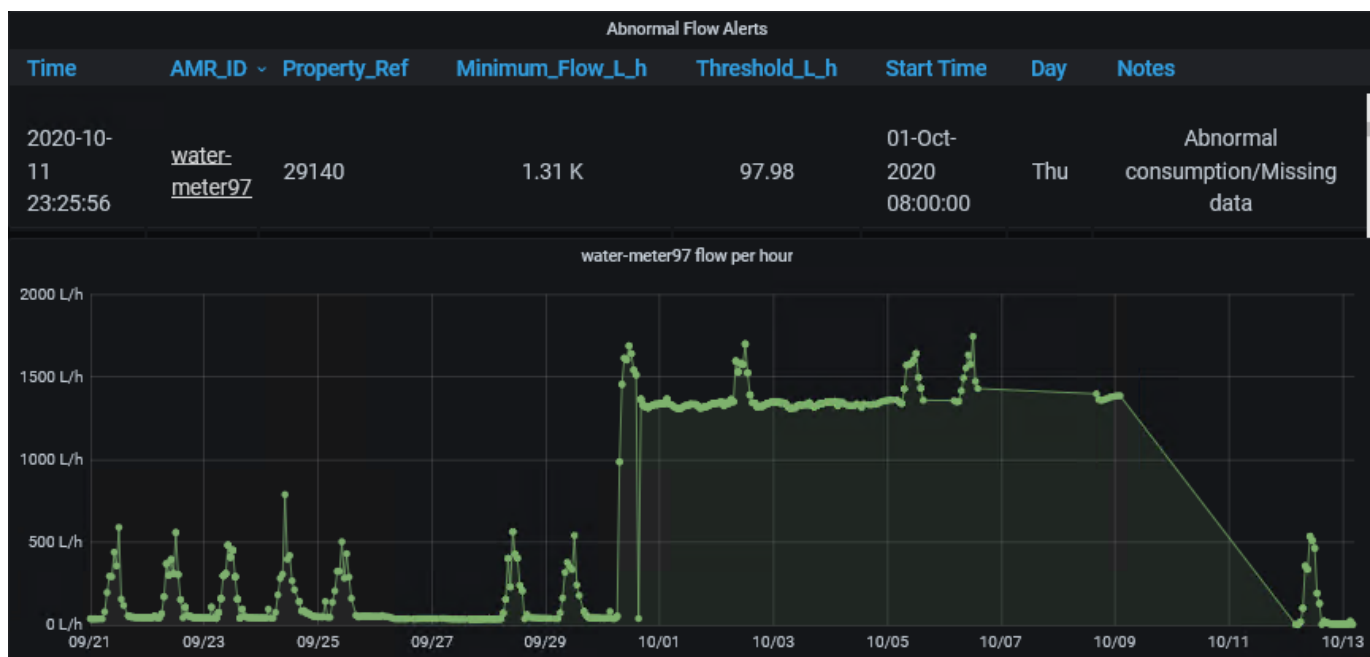
43

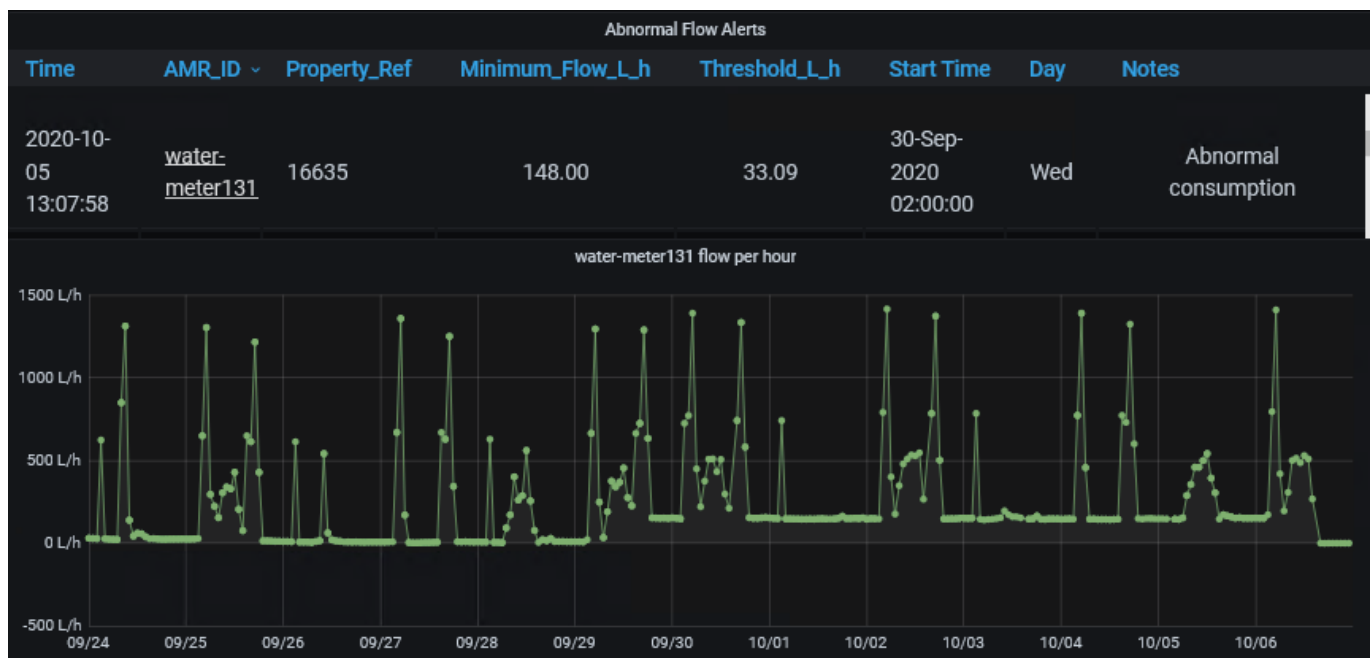| Abnormal Flow Alerts							
Time	AMR_ID ∨	Property_Ref	Minimum_Flow_L_h	Threshold_L_h	Start Time	Day	Notes
2020-10-11 23:25:56	water-meter97	29140	1.31 K	97.98	01-Oct-2020 08:00:00	Thu	Abnormal consumption/Missing data

Figure 44: Alert table and measurements received from "water-meter 97" AMR, during a period when an irregular flow is detected.

In a second example (Figure 45) the minimum consumption of "water-meter 131" located also at a public school increased from approximately 10 m³/h to 145 m³/h on 30th of September. In this case, according to information gathered by LWB, the school was undergoing construction works during the examined week which may explains the increase in water consumption.

Abnormal Flow Alerts							
Time	AMR_ID ∨	Property_Ref	Minimum_Flow_L_h	Threshold_L_h	Start Time	Day	Notes
2020-10-05 13:07:58	water-meter131	16635	148.00	33.09	30-Sep-2020 02:00:00	Wed	Abnormal consumption

Figure 45: Alert table and measurements received from "water-meter 131" AMR, during a period when an irregular flow is detected.

5.3 Irregular flow detection at utility supply level

The volumetric measurements of a utility supply area of the WDD network were used to test a methodology which determines if there is a leakage in that area. The measurements are the inflow of the area water supply tank. The conversion of volume measurements to flow measurements, illustrated in Figure 46, reveal a switching behavior which is due to the switching of valve status (open/close) when the water in the tank goes below a certain level.

5.3.1 Methodology

5.3.2 WDD Pilot

- **AF for consumer tanks**
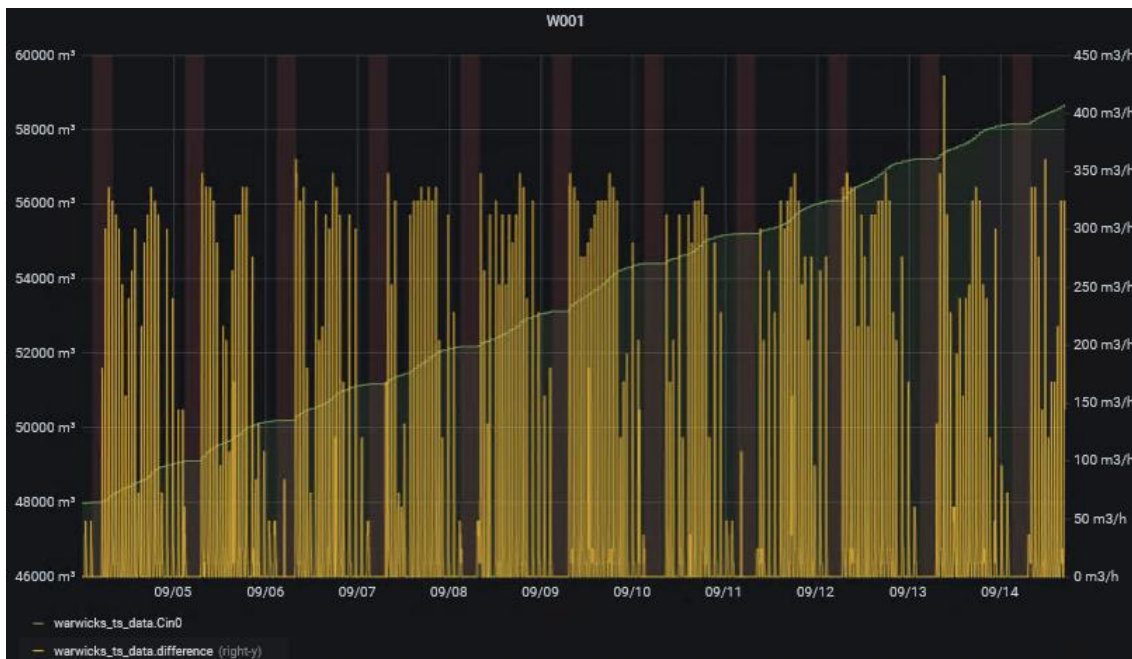
Figure 46: Water volume measurements from a supply area of the WDD network

5.4 DEYAM Pilot

5.4.1 Methodology

This feature aims to detect abnormal behaviors and promptly generate alerts to system administrators. The reliability of alerts increases when data uncertainty is included in the analysis, due to the fact the sensor measurements are often unreliable and inaccurate.

Having calculated the individual uncertainties for all potential sources of uncertainty, the combined typical uncertainty is given by
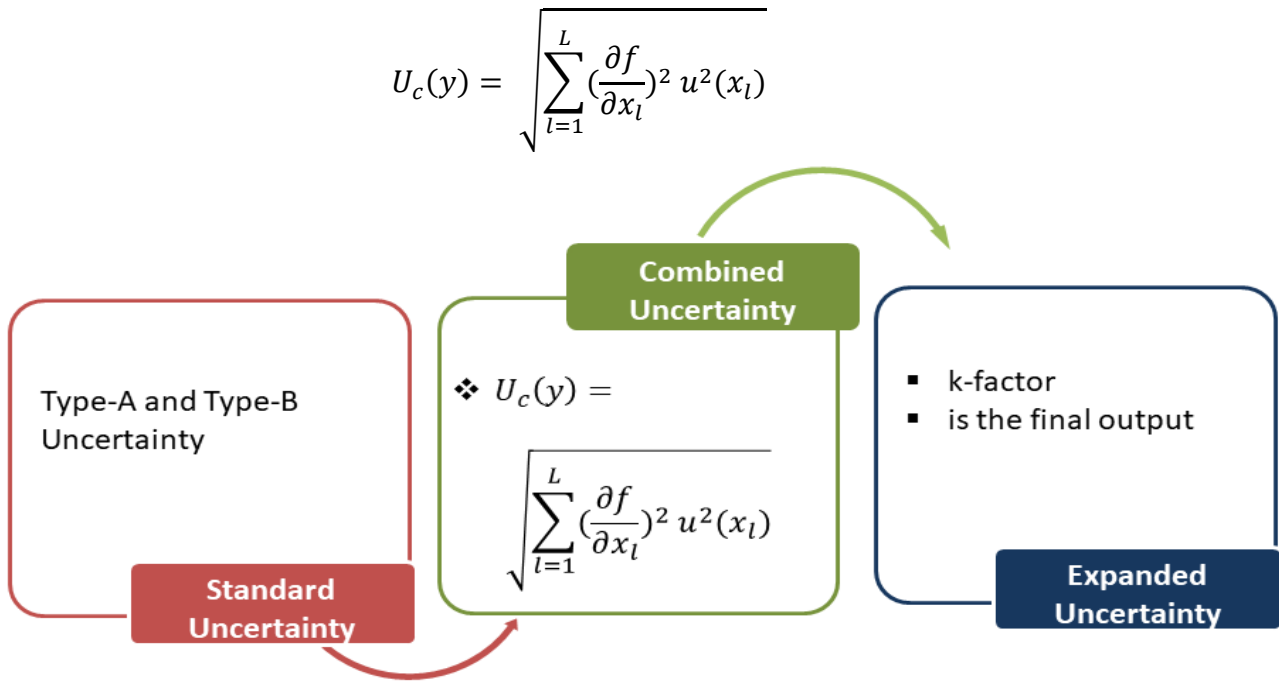
$$U_c(y) = \sqrt{\sum_{l=1}^{L} (\frac{\partial f}{\partial x_l})^2 \, u^2(x_l)}$$

Figure 47: The process of quantifying data uncertainty.

Multiplying the combined standard uncertainty by a parameter K, which is determined by the desired confidence level of the Gaussian distribution, gives the final output in the process of quantifying data uncertainty.

As a next step, we examine the compliance of the processed data augmented by the corresponding uncertainty against predefined normal operating limits. These limits are typically set empirically and can be related to environmental conditions or to the structural strength of the sensors. The following figure shows the different cases that arise during this examination.
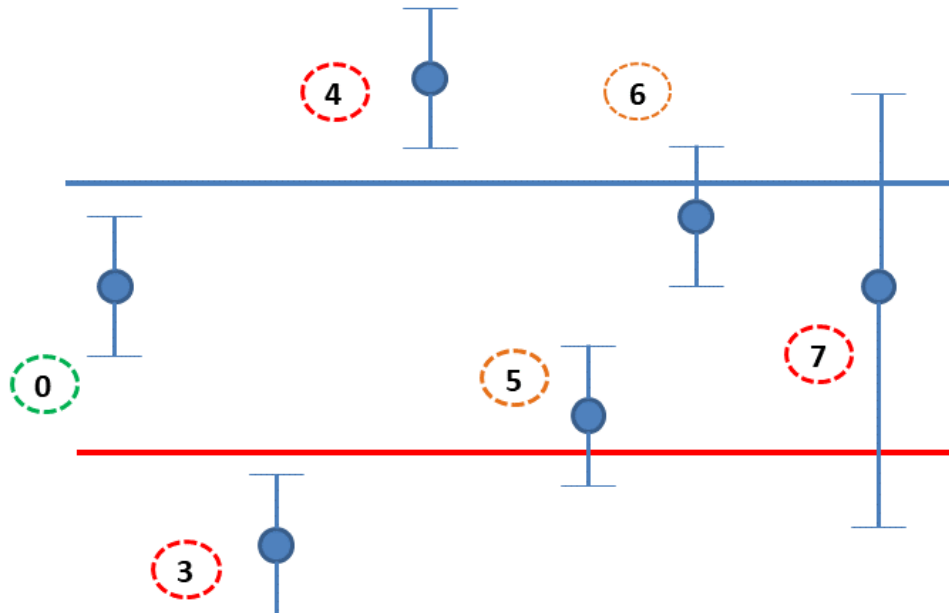
Figure 48: predefined normal operating limits

Accordingly, the actions to be taken depending on the category of the warning, are shown below. For convenience, the warnings are divided into three categories based on their importance and validity. The smooth operation of the system is indicated in green, an orange alert indicates that

there is a possibility of abnormal behavior given the uncertainty of the data, whilst a red alert indicates that an anomaly has been detected.
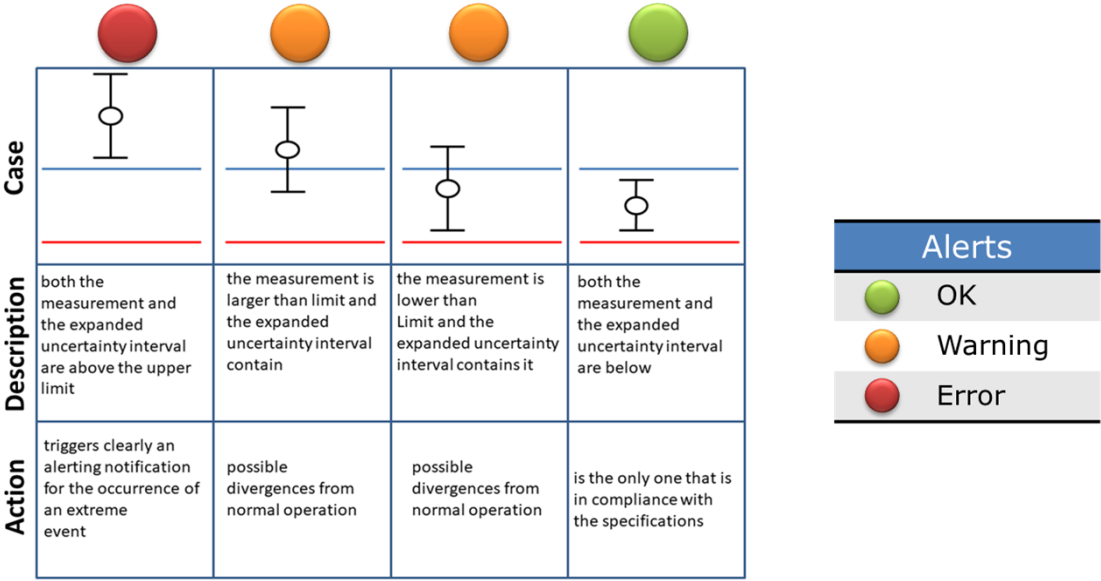

Figure 49: Actions to be taken depending on the category of the warning

The estimated alerts are further visualized in a simple and intuitive way, to facilitate fast decision making by the system operator, as shown below for the data provided by DEYAM. The visualization module has been developed in Grafana combined with an SQL database.

Figure 50: Estimated alerts developed in Grafana combined with an SQL database

5.5 Conclusions

In this sector, 4 different methodologies of abnormal pressure and flow detection are applied on 4 different case studies. The first methodology included the use of pressure measurements from pressure sensors installed in strategic locations inside the DMA. The proposed algorithm which is based on predicting the weekly periodic difference between the measured and estimated pressure showed to be very efficient since it detected accurately the time and the area of a leakage using the BattLeDIM benchmark. Moreover, this algorithm showed capable of detecting abnormal pressure behavior in the WBL network.

In the LWB pilot area, the proposed algorithm uses AMR measurements to identify irregular consumption which may be due to a number of reasons such as leakage, abnormal consumption due to a social event or construction work, or even a forgotten running faucet. As shown before, this methodology worked perfectly since it detected various types of irregular consumption within the LWB water network.

6 Pressure regulation for leakage reduction

6.1 Methodology

The following algorithm was developed as a pressure management tool which can be used eventually for leakage reduction. It is designed for DMA level water networks with one water source (one water input) which is regulated by a PRV. The algorithm calculates the optimal PRV setting at each time step by taking into consideration the daily consumption data and it is based on a trial-and-error methodology. The criterion for the selection of the optimal PRV setting is to achieve a minimum acceptable pressure in the water network/DMA during a selected time period.

The *PRV_setting* algorithm is applied in leakage models (INP. file). The leakages are modeled as emitters which are associated with junctions and simulate the leakage flow through an orifice based on the following equation (Rossman, 2000):

$$q = C\,p^{\gamma}$$

where q = flow rate, C= discharge/emitter coefficient, p=pressure and γ= pressure exponent (usually 0.5). Since the total inflow and the actual consumption of each DMA are known, the leakage flow is calculated. Following, the emitter coefficient is estimated based on the average leakage flow and pressure of each consumption junction (as a first step, the leakages are applied only in consumption junctions). The initial base demands which represented the entire inflow are reduced in order to represent only the actual consumption (Function 6).

Function: *Set_EmitterCoef(Model, ActualConsumption)*

Inputs:

@model: Water network model (.inp file), weekly consumption patterns, time duration and time step

@RealConsumption: Actual Consumption (m^3/h)

Output:

@Leakage_model: Water network model (.inp file) including leakages as emitters

---

Function 6: Outline of water model development including leakages

The user selects a range of input pressures which are assigned to the PRV and states the desirable optimal pressure range or the minimum acceptable pressure of the network. For example, the available pressure range for the PRV is set to 40 – 60 m (this pressure range is usually based on previous PRV settings or the knowledge of the user regarding the operation of the network) and the minimum acceptable network pressure is 20 m. The algorithm estimates the control setting of the PRV for each time step which must be between 40 – 60 m and will achieve at least 20 m pressure in the network (all nodes). The control setting which achieves network pressure closest to 20 m is selected. In case, the desirable network pressure cannot be achieved a warning message is issued.

Monte Carlo simulations are also integrated in the algorithm in order to take into consideration demand, diameter, pipe length and roughness uncertainties. For computational reasons the input pipe and the PRV of the DMA are simulated as a Reservoir.

The required input data for the *PRV_setting* algorithm are stated below:

- Calibrated leakage - hydraulic model of the DMA under study (EPANET input file). It includes the topology and the parameter values of the network (roughness, length and diameter for pipes; positions and characteristics for valves; demand patterns for junctions). The calibrated hydraulic model is created using the "Hydraulic State Estimation" algorithm.

- The minimum acceptable network pressure

- The proposed PRV pressure range

The output of the *PRV_setting* algorithm are stated below:

- A vector of PRV settings which can appear in a table and can be computed on demand for a specific period (e.g. a period of one week with a 30 minute time step thus a vector of 336 PRV settings).

- The optimized EPANET network file with the new PRV settings.

- The water losses before and after the optimization

- The reduction of CO_2 emissions and water cost.

Function: *PRV_setting(model, iters, hmin, hmax, Pmin, Pmax)*

Inputs:

@model: Water network model (.inp file), weekly consumption patterns, time duration and time step

@Iters: Number of Monte Carlo simulations

@Pmin: Minimum acceptable network pressure

@Pmax: Maximum acceptable network pressure

@hmin,hmax: Range of proposed PRV settings

Output:

@OptimalPRVSettings: PRV setting for each time step

@*Optimized_model:* Inp. file with new PRV settings

@*TotalVolumeLeakage_1:* Leakage volume (m^3) for initial model.

@*TotalVolumeLeakage_2:* Leakage volume (m^3) for optimized model.

@*CO2Emmision:* CO_2 emissions reduction (kg CO_2)

@*SavedWaterCost:* Cost reduction (€)

@*SavedWaterVolume:* Water volume reduction (m3)

Function 7: Optimal PRV setting selection algorithm outline

6.2 WBL Pilot

The proposed methodology was applied at the DMA 131 of the WDN of Limassol. The inlet flow of DMA 131 varies between 40-90 m^3/h with a minimum night flow close to 37-40 m^3/h and the inlet pressure varies from 45 to 50 m. In Figure 51 and Figure 52 the flow and pressure of DMA 131 during an average week (20/7/2020 – 26/7/2020) are presented, respectively. The objective was to reduce to inlet pressure from 45-50 m to 40-43 m and examine the changes on the water flow. The key performance indicators (KPIs) were used in order to estimate the impact and this experiment:

1. Average Water Cost (€/m^3): 0.8
2. CO2 Emission (kg CO2/m^3): 0.137

Figure 51 - Inlet Flow of DMA 131
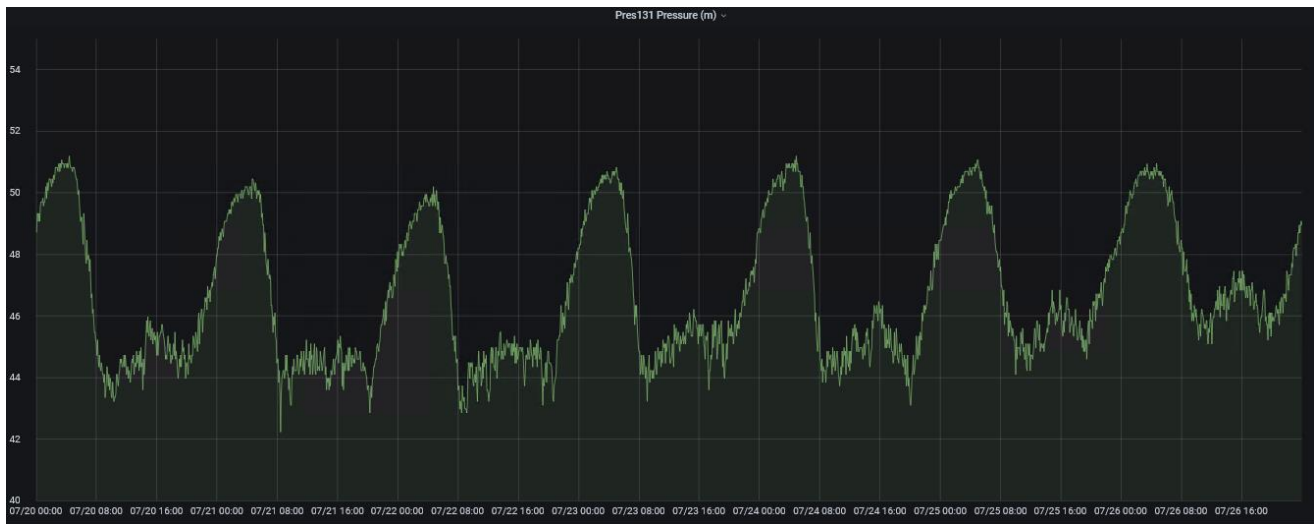
Figure 52 - Inlet Pressure of DMA 131

6.3 Application of methodology on a digital twin of the network

The application of the proposed methodology on a simulation of the network model which was calibrated using real data was performed in three steps, described below.

6.3.1 Step 1: Hydraulic model calibration

The 'hydraulic state estimation' algorithm presented in previous sections was utilized and a calibrated water model for the period between 20/7/2020 – 26/7/2020 was obtained (Figure 53). Notice that the simulated inlet flow and pressure match the corresponding measurements retrieved through the installed sensors (Figure 53).
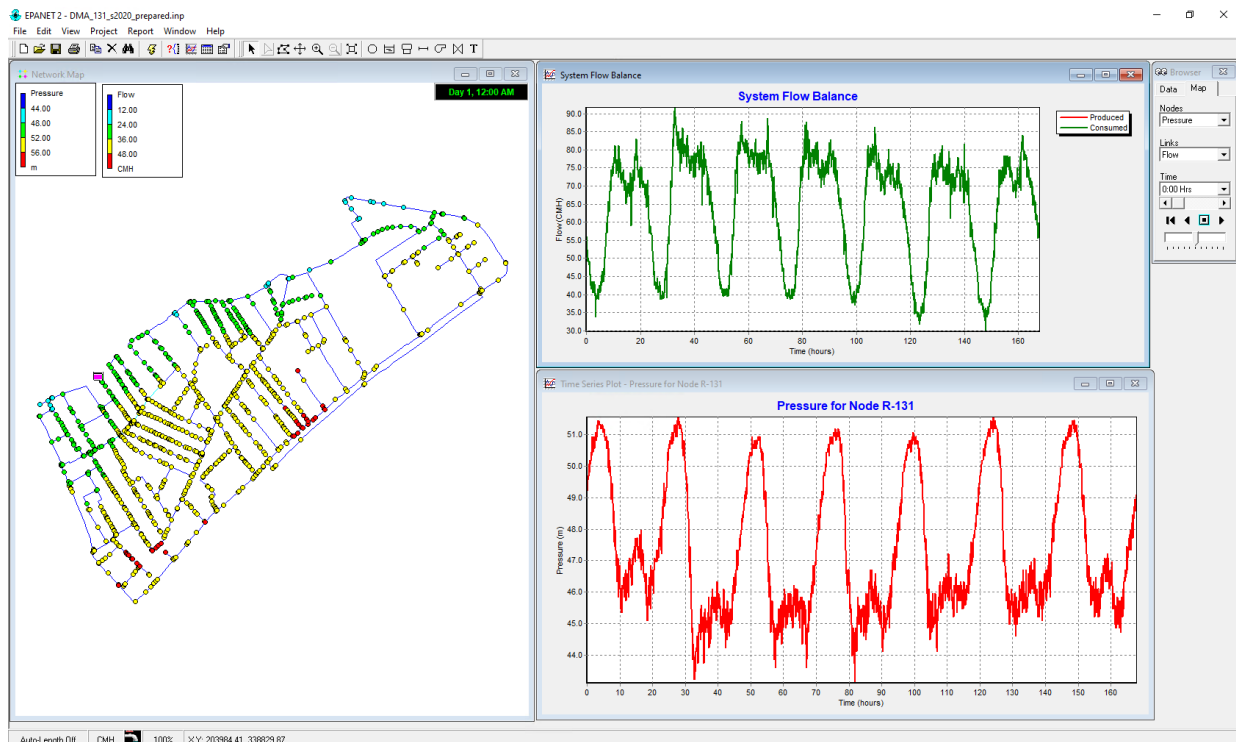
Figure 53 - Calibration of the DMA 131 network model using measurements

51

6.3.2 Step 2: Incorporating background leakages in the model

The second step of this methodology is to create a water network model including leakages. According to historical data and information from the WBL it was estimated that the actual average consumption of this area is approximately 45 m^3/h while the leakages are estimated to 20-30%. With this information, the calibrated model (Figure 53) was re-calibrated based on (Function 1) and a model including both the actual consumption and leakages was developed (Figure 54). Due to a variety of uncertainties such as the actual consumption and the varying leakage percentage throughout the day, a slide difference between the night flow of the two models is observed. However, this does not affect the overall aim of this pilot which is to show how the inlet flow is affected by pressure control.
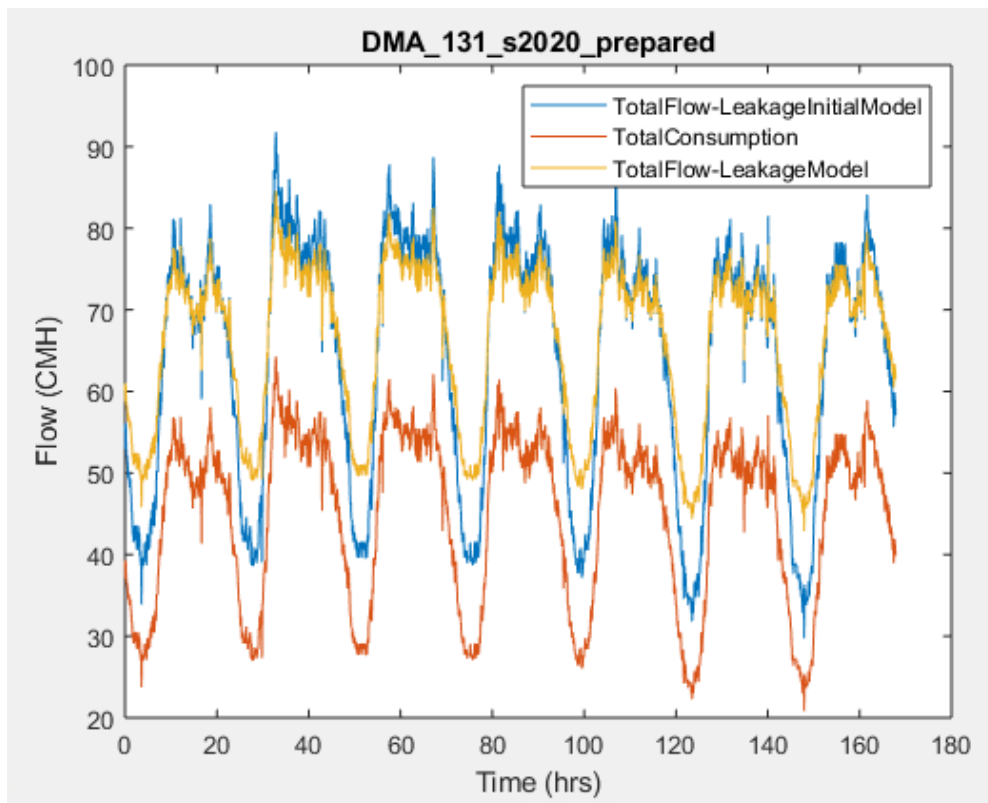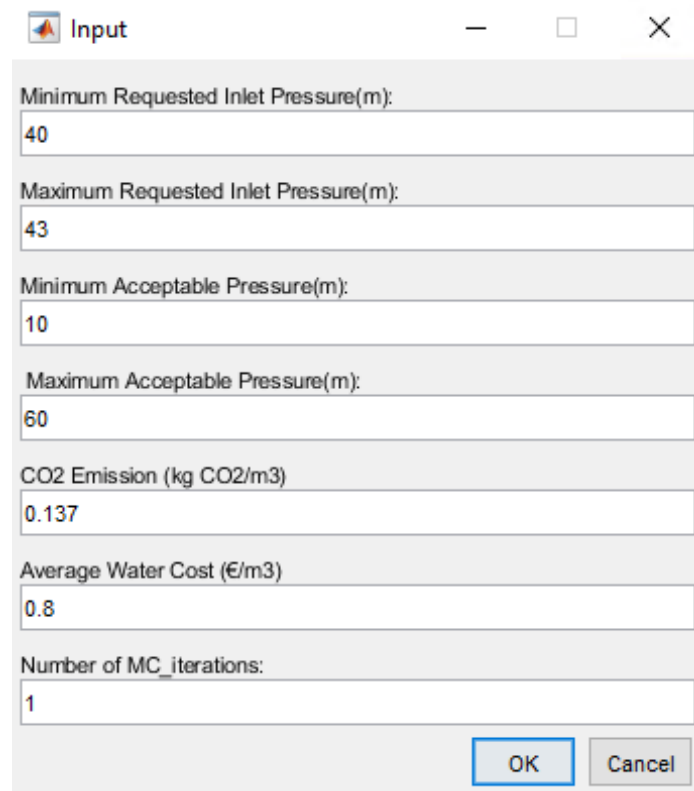
Figure 54 - DMA 131 system inflow when incorporating leakages into the model

6.3.3 Step 3: Application of pressure control algorithm

After obtaining the "leakage model", the pressure control algorithm (Function 2) is ready for use. The user introduces the required input data, as illustrated in

- minimum requested inlet pressure (m) – PRV setting
- maximum requested inlet pressure (m) – PRV setting
- minimum acceptable pressure (m) – Network pressure guideline
- maximum acceptable pressure (m) – Network pressure guideline
- Average Water Cost (€/m^3) - KPI
- CO2 Emission (kg CO2/m^3) - KPI
- Number of Monte Carlo Simulations – Simulation Option

Figure 55 – Input parameters of pressure control algorithm

The results of the simulated pilot phase are presented in Figure 56 and Figure 56. Figure 56 shows the total flow, the actual consumption and the leakage flow of the selected DMA before and after the application of pressure control. According to the simulated results **the minimum night flow is reduced by approximately 5 m³/h** due to the same reduction in the background leakages. The corresponding inlet pressure is set to 40 m during the entire simulation period. This reduction is translated in terms of the proposed KPIs (Table 5) as follows:

- 238 m³ of water saved per week

- 190 € saved per week

- 32 kg CO_2 reduction per week

> ~**5 m³/h reduction** in background leakages

> Estimated background leakages:
>   ▪ **without** pressure control
>   ▪ **with** pressure control
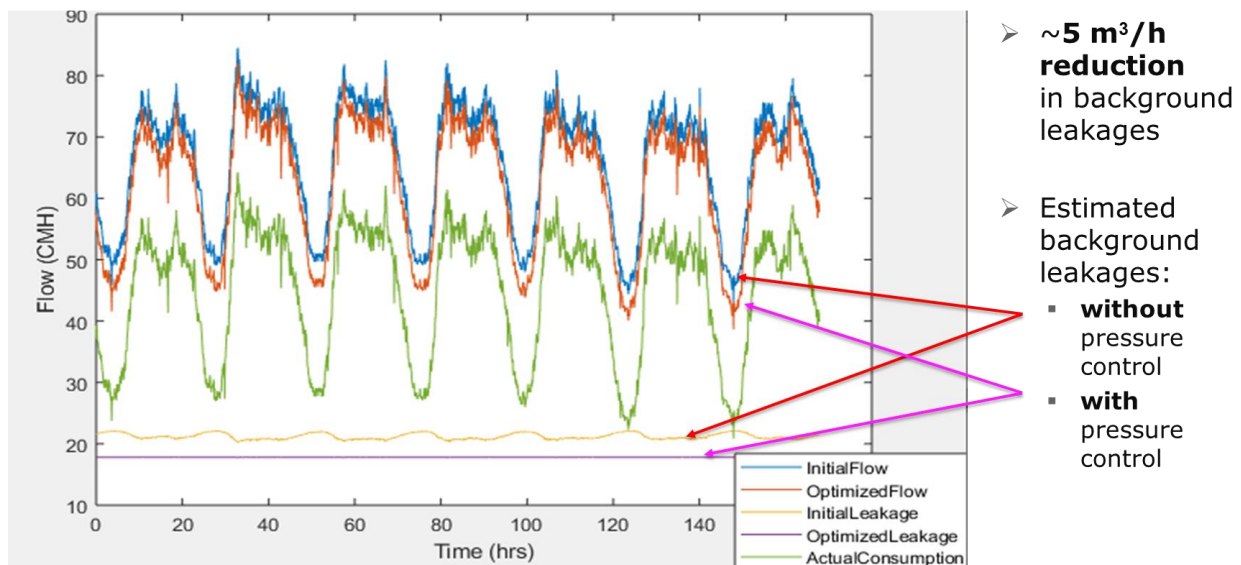
Figure 56 - Total and leakage flow of a DMA with and without pressure control.

Table 5: Key performance indicators of pressure control methodology.

Inflow volume (m³)		Leakage volume (m³)		Saved Water (m³)	Saved Cost (€)	CO₂ reduction (kg CO₂)
Without control	With control	Without control	With control			
11109	10871	3563	3325	238	190	32

6.4 Pilot phase: Application on the actual water distribution system

On 2/9/2020 the WBL proceed to the pilot phase of this methodology by reducing the inlet pressure of DMA 131 from 45-50 m to 40-43 m. The results of this application were noticed immediately. The minimum night flow was reduced by 10 m³/h from 37 m³/h to 27 m³/h. This can be observed in Figure 57, which show the inlet pressure and flow of DMA 131 from 28/8/2020 to 7/9/2020 while Figure 58 and Figure 59 present the flow and the pressure during the following week.

Figure 57 - Reduction of minimum night flow in DMA 131 when applying the pressure control algorithm
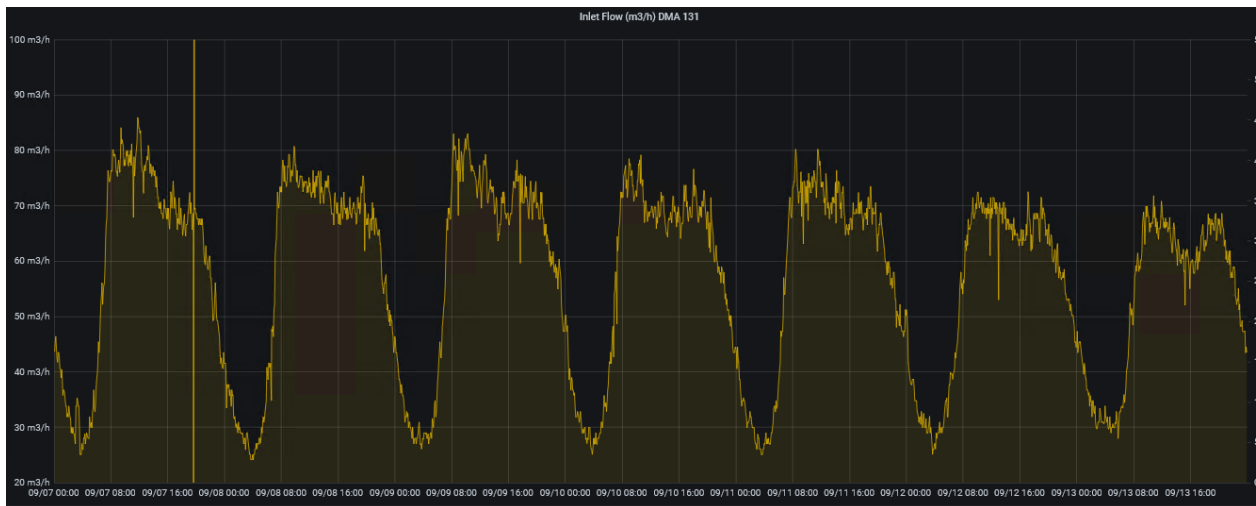

Figure 58 - DMA 131 Inflow during the week when the pilot was performed
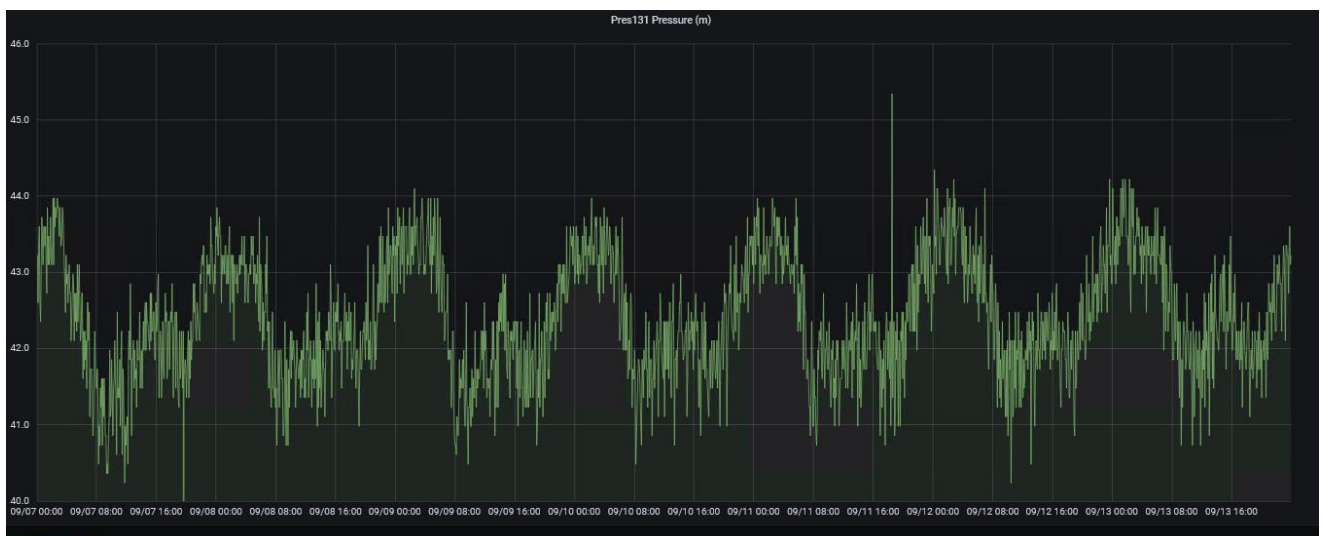

Figure 59 - DMA 131 Inlet pressure during the week when the pilot was performed

6.5 Conclusions

The pressure regulation methodology using PRVs which is developed and presented in the previous section showed that it can be a very helpful and efficient tool regarding the reduction of background leakages. The first part of the pilot where a digital twin of the network was used showed that a reduction of 5 m^3/h during the night period can be achieved with this methodology. The results from the application of this method on the actual network were even better since a 10 m^3/h reduction on the minimum night flow from 37 m^3/h to 27 m^3/h was achieved. Initially, this indicates that model calibration is needed for the first part of the pilot. However, the most important outcome is that this methodology can be used as leakage reduction tool.

The results can be considered far more promising and encouraging if someone considers that with that 5 m^3/h night flow reduction the water utilities can save up to 238 m^3 of water, 190 € and 32 kg CO_2 per week.

7 LoRaWAN Performance Evaluation

7.1 Methodology

In order to evaluate the performance of a LoRaWAN deployment a series of metrics may be used. The metrics are based (a) on the physical layer properties which are specific to the LoRa protocol, such as data rate, number of receiving gateways per node, and (b) on the physical layer properties in general, such as received signal strength, SNR, etc.

Moreover, the simulated coverage vs the real one is going to be compared in terms of RSS per geographical point and gateway connectivity.

Statistical analysis is going to be performed in all the available data. The available data include a per packet received at the LoRaWAN server record which has 21 attributes. We concentrate on the following five attributes which provide meaningful information regarding the physical layer performance of the LoRaWAN infrastructure, Device name, Gateway name, Data Rate, SNR, and RSSI.

We decided to calculate, statistically analyze, and visualize the following series of data, which will provide the most essential feedback regarding the performance of the LoRaWAN deployment:

1. Number of AMR positions without predicted coverage
2. Average received signal strength difference per AMR position
3. Number of communicating AMRs
4. Number of Packets per AMR
5. Number of Packets per Gateway
6. Number of Gateways per AMR
7. Distance of Gateways per AMR
8. Received Signal Strength
9. Signal to Noise Ratio
10. Data Rate distribution

Finally, we have depicted in the cumulative distributions the 90th percentile, such as to have an instant view of the performance for most AMRs.

7.2 LWB Pilot

The pilot of the Larnaca Water Board included the installation of 10 LoRaWAN gateways and 360 AMRs-water meters in the wider metropolitan area of the city of Larnaca in Cyprus. In Figure 60 the installation points of the AMRs are depicted. The main goal was set to be the LoRa network coverage of all the AMRs. Secondly the distribution of the gateways should be such that the majority of Larnaca's metropolitan area is covered and at the same time a significant overlap of the coverage of each gateway exists for redundancy.

Figure 60: Installation points of AMRs / LoRa end nodes

After the installation of all the gateways and the AMRs and the proper tunning of the system, a continuous recording of all the received packets information was established. Therefore, from June 2020 up to almost end of November 2020 (a period of almost 6 months), a total of 944255 packets' information has been recorded. This amount of data provides a representative indication of the LoRaWAN's infrastructure performance.

The installation points of the gateways were chosen from a selection of available building rooftops. The available rooftops were convenient to use, but LoRaWAN-wise they didn't have any strong or desirable advantage. The final installation points were the ones that provided the maximization of coverage, based on extended simulations.

For the coverage simulation we used two different propagation models, (a) Ericsson and (b) COST Hata. Both the terrain layout and the buildings have been taken into account. Various parameters of the models were chosen such as to have a more pessimistic prediction of the coverage. Depicted in Figure 61 are the simulation results projected on the Google Earth map of Larnaca metropolitan area.
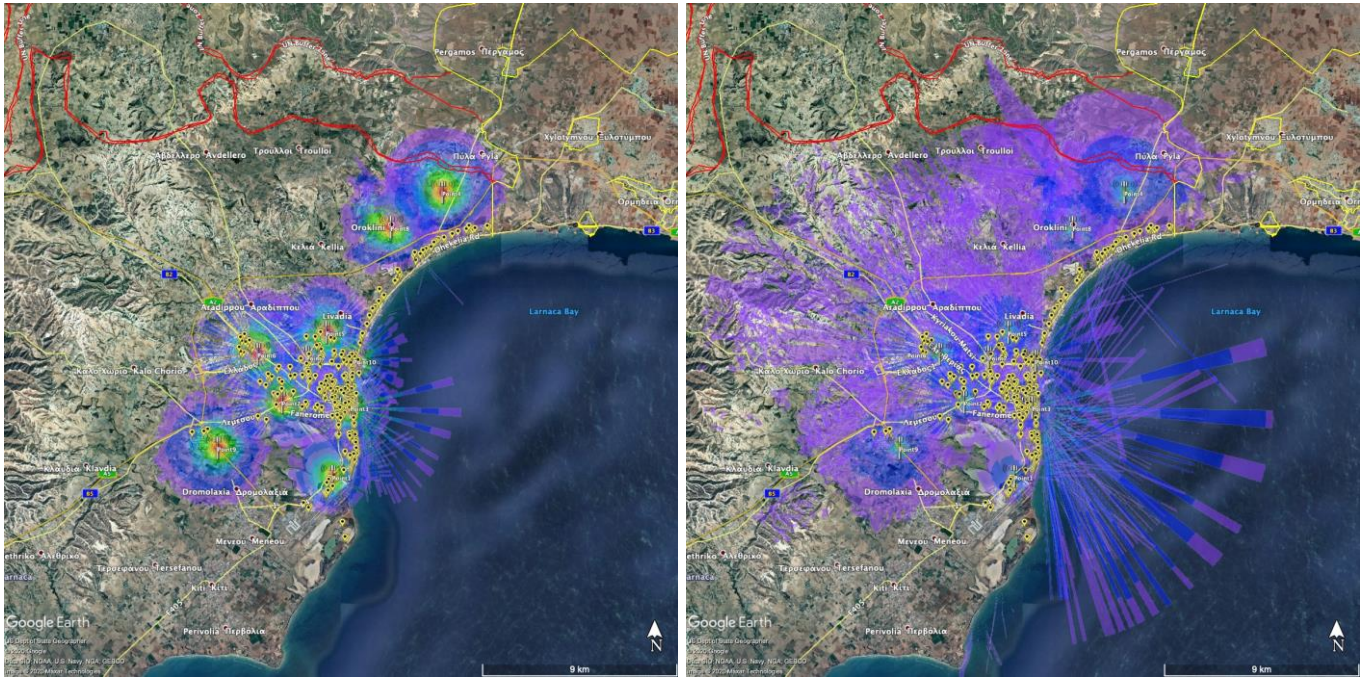
Figure 61: Simulated LoRaWAN coverage of the two propagation models [Ericsson left, COST hata right]

7.2.1 Number of AMR positions without predicted coverage

Based on the traces we verify the diversion of the simulated coverage compared against the real world measurements. The number of links observed are 1406, because each AMR may have connections with multiple gateways. Out of the 1406 links the simulation has predicted only the 485/522 (Ericsson/COST Hata model respectively). This shows that the simulation is rather pessimistic in terms of coverage area, which was expected due to the values used for various parameters of both models.

7.2.2 Average received signal strength difference per AMR position

For the links that have been predicted by the simulation the average difference from the real measurements is, respectively for Ericsson/COST Hata model, 15.7/6.2 dB.

7.2.3 Number of communicating AMRs

In the field 346 out of the 360 AMRs have been installed. Out of the 346 installed 328 have send packets from June to November (about 95%). There are less AMRs communicating the last

month, but unfortunately it has not been investigated whether there is a problem in the communication or device malfunction/vandalism.

7.2.4 Number of Packets per AMR

The expected number of packets per AMR is 64 for the plain water meters and 720 for the digital ones, over a 30 days period, thus over 384 and 4320 packets respectively for 6 months. The results show quite larger numbers than these, due to the fact that the same packets are received from more than one gateway, for a large percentage of the AMRs. We 've chosen to use the data before the de-duplication process in order to have a better overview. As shown in Figure 62 the 90% of AMRs have over 600 packets, over 50% more than the expected ones.
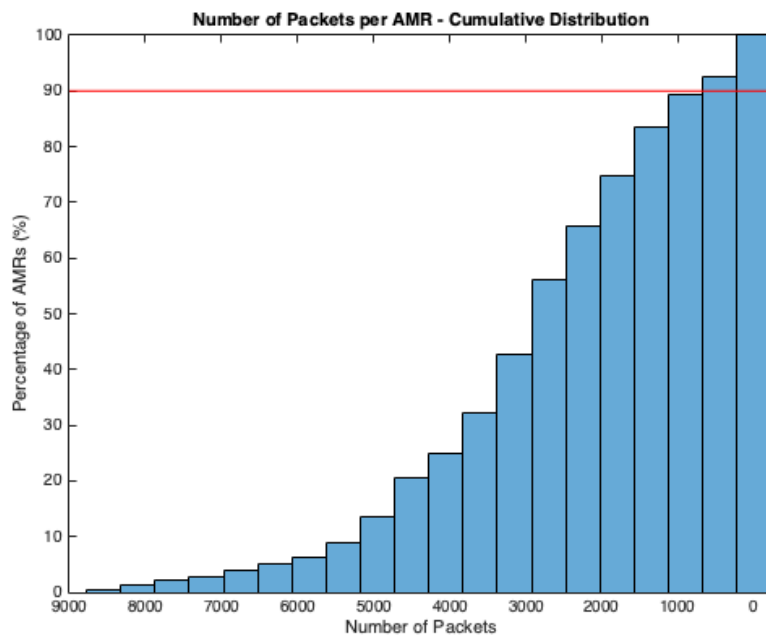
Figure 62: Cumulative distribution of the number of received packets per AMR

7.2.5 Number of Packets per Gateway

In Figure 63 we represent the distribution of the received packets to each gateway. It is obvious that there are a few gateways [1,2,3,10] that have captured the majority of the packets. These gateways are either better placed or closer to more AMRs. A good example of the importance of placement/location are the gateways 10 and 7, although they are close to each other they have a large difference in the number of the received packets, because gateway 10 is placed higher than gateway 7 [52 vs 33 m above sea level - 45 vs 15 m above ground level].
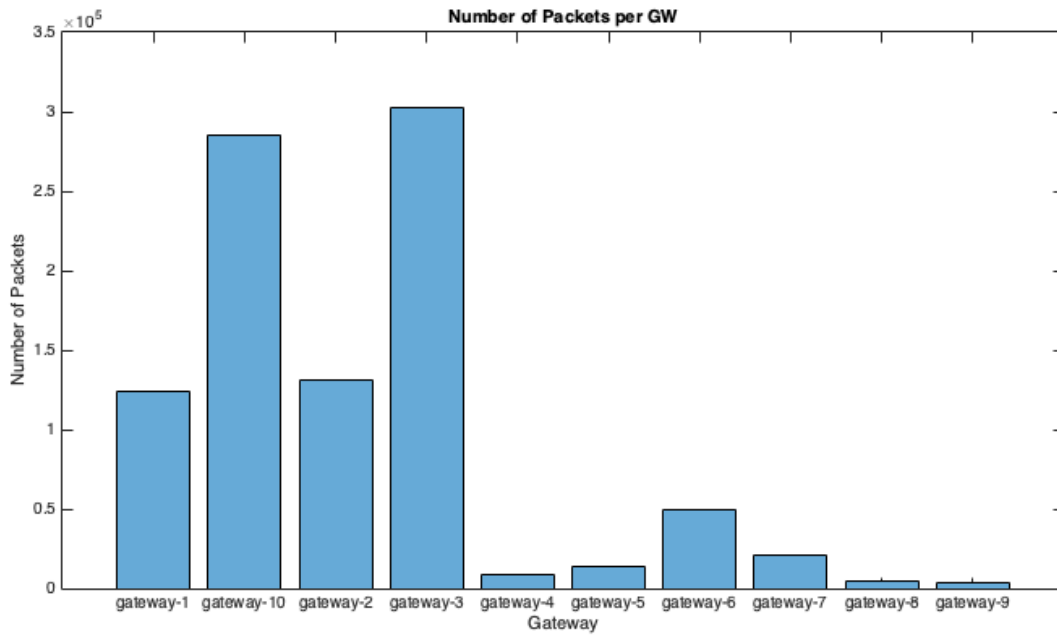
Figure 63: Number of packets received by each Gateway

7.2.6 Number of Gateways per AMR

As LoRaWAN permits the concurrent reception of packets by multiple gateways, providing essentially redundancy of the form of immediate failover, we observe in Figure 64 that the 90% of the AMRs is received by at least two gateways. Four or more gateways cover the 50% of the AMRs.
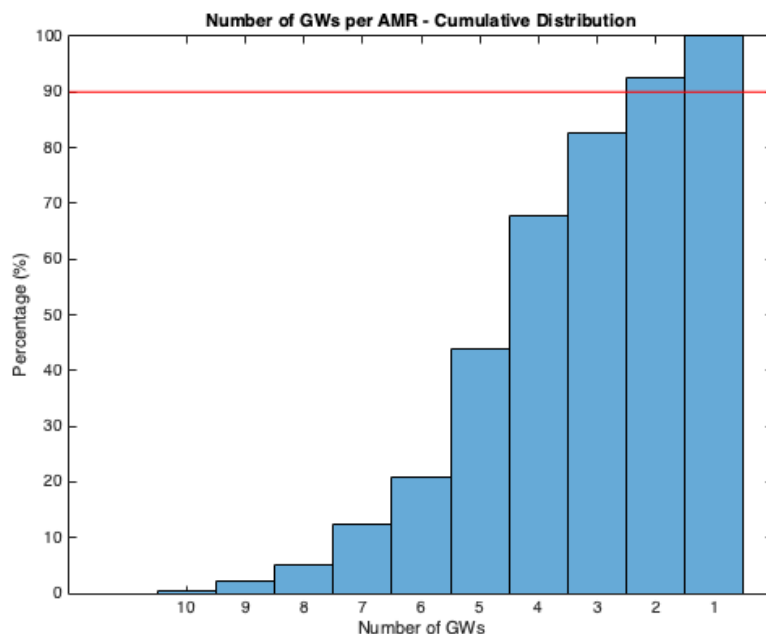

Figure 64: Cumulative distribution of the number of Gateways receiving packets per AMR

7.2.7 Distance of Gateways per AMR

Another interesting statistical result is the distribution of distances between the AMRs and their respective receiving gateways. Six kilometers or less is the distance for the 90% of AMRs, but

there is a small percentage of AMRs that achieve reception to over 10 km of distance, which is quite remarkable taking into account that they are placed just few centimeters above ground, Figure 65.
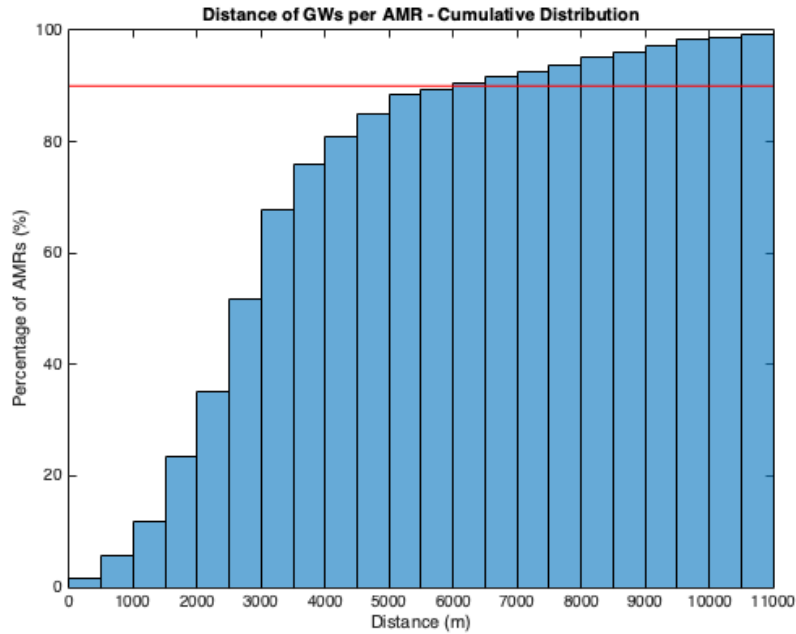

Figure 65: Cumulative distribution of receiving Gateways' distance from AMRs

7.2.8 Received Signal Strength distribution

The average received signal strength per AMR has been calculated and in Figure 66 the distribution is depicted. The average received power of 90% of all packets is -112 dBm or better (Figure 67), with an average close to -100 dBm and a standard deviation of about 7 dBm per AMR.
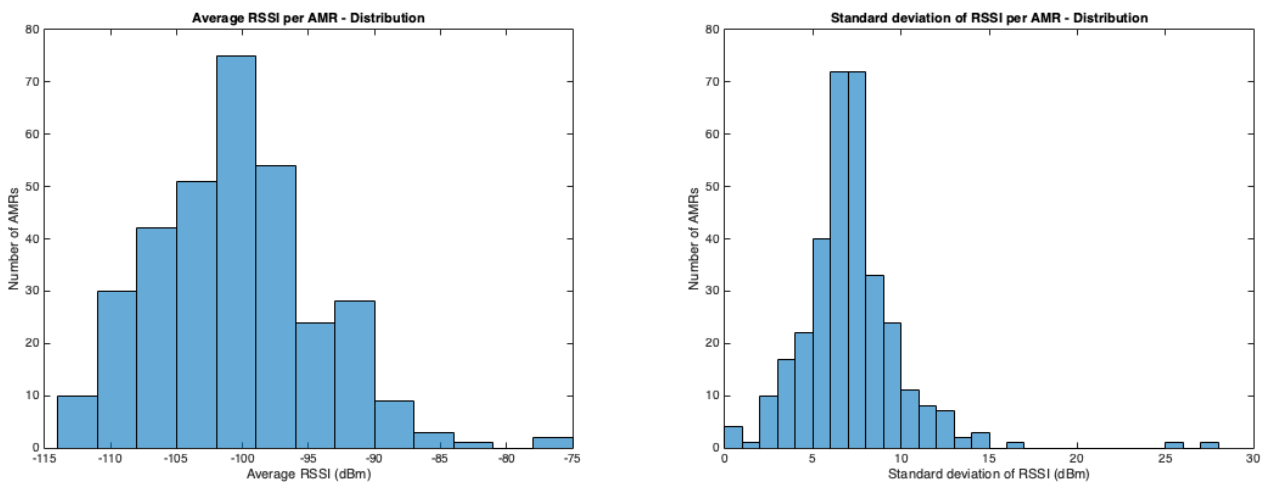
Figure 66: Distribution of the average received signal strength (left) and the respective standard deviation (right) per AMR
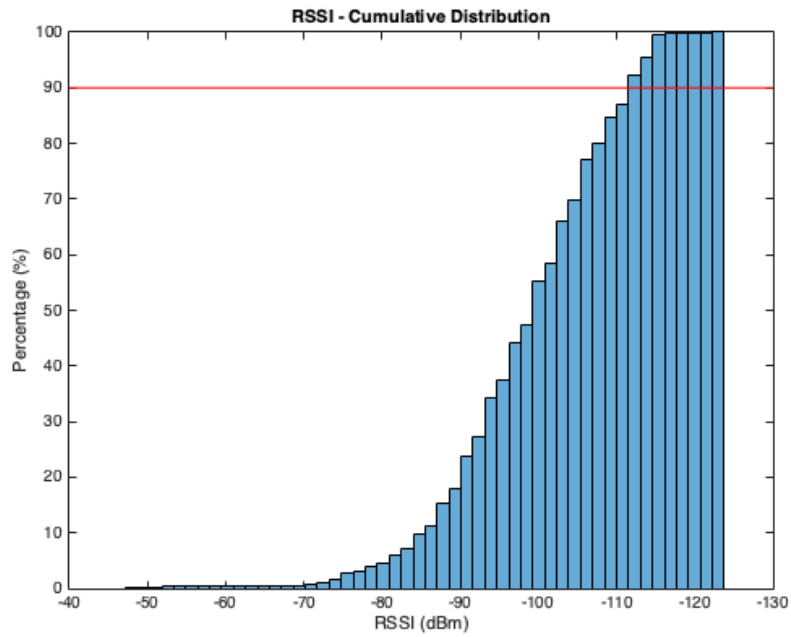
Figure 67: Cumulative distribution of received signal strength of all packets

7.2.9 Signal to Noise Ratio distribution

The Signal to Noise ratio (SNR) cumulative distribution in Figure 69 shows that 90% of the received packets have an SNR of -21 dB or better. Less than 20% of the packets has a positive SNR, which shows the robustness of the physical layer, as most of the competitive technologies require a positive SNR for reception.
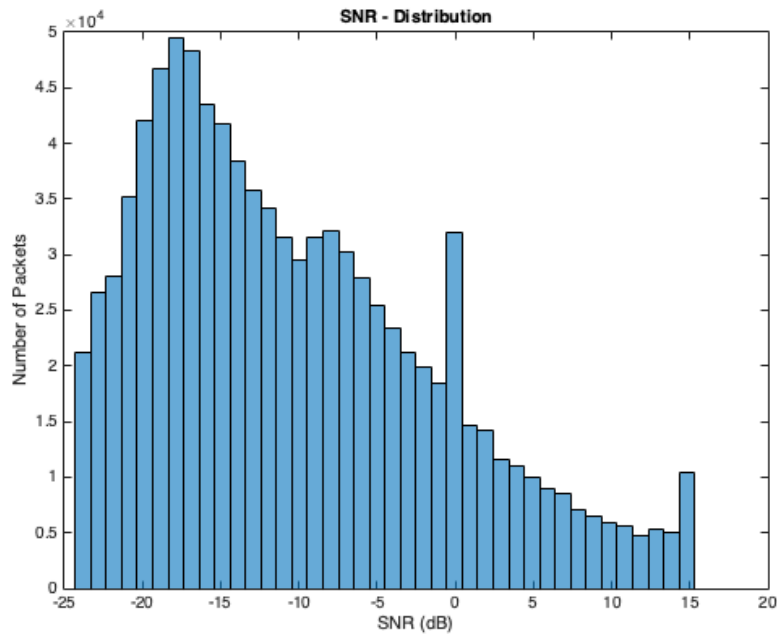

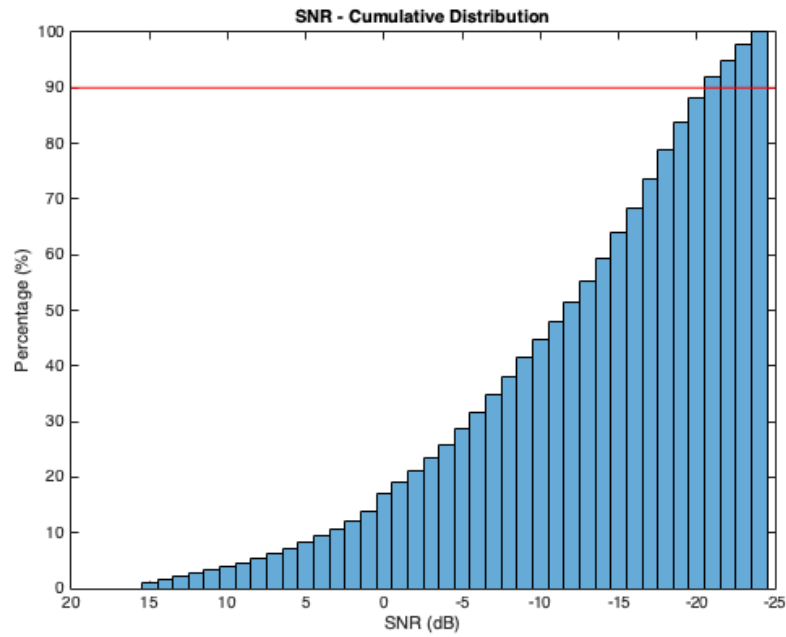*Figure 68: Distribution of the SNR of all received packets*

Figure 69: Cumulative Distribution of SNR of all received packets

7.2.10 Data Rate distribution

The data rate distribution in Figure 71 shows that the majority of the packets are received at the lowest data rate (DR0), which is expected from the long distances and the low SNRs. Also, there is a significant percentage at data rate 5 (Figure 70), for the AMRs that are close to a gateway.
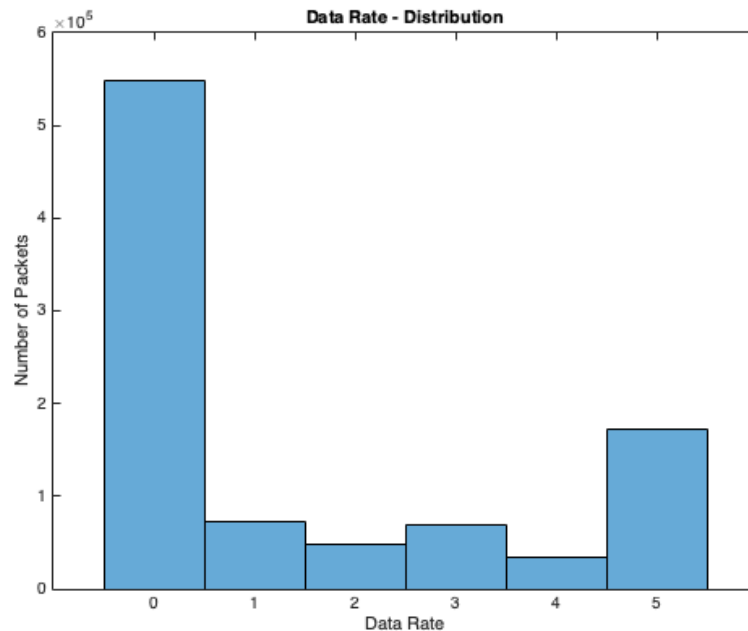

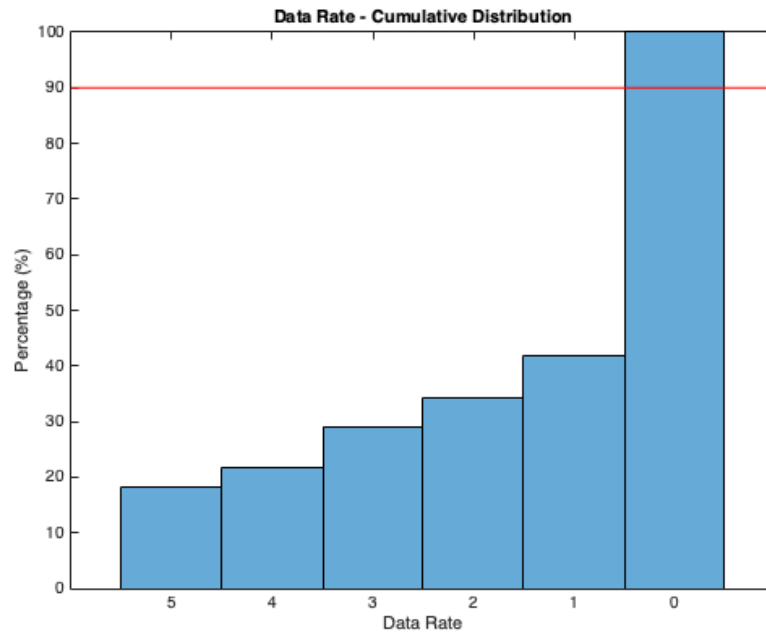*Figure 70: Distribution of Data Rates of all received packets*

Figure 71: Cumulative distribution of Data Rates of all received packets

7.3   Conclusions

Results from the operation of the LWB's LoRaWAN infrastructure demonstrate the advantages, the robustness, the performance, and the suitability of the specific technology for the remote monitoring of large-scale water networks. The installation locations of the gateways could have been placed at more optimal locations, however the infrastructure remains efficient and ready to support a larger deployment of AMRs in the metropolitan area of Larnaca.