UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA DELL'INFORMAZIONE

Household water leak detection using smart metering generated water consumption data

Relatore
Prof. Laurenti Nicola

Laureanda
Cesaro Greta Francesca

ANNO ACCADEMICO 2023-2024

Data di laurea 19/03/2024

Abstract

Residential water leaks pose a significant challenge, leading to water waste and financial burdens for both consumers and suppliers. Traditional leak detection methods, such as visual inspections, acoustic detections, and the use of district metered areas (DMAs), often fall short in terms of efficiency and proactivity. Smart metering infrastructure offers a promising solution to this persistent problem. This research explores the advantages of smart meters that provide real-time, high-resolution water usage data, enabling the detection of consumption patterns and leak indications. The case of Italy exemplifies the urgency for improved water management, with high per capita consumption and considerable water loss before reaching consumers. Italian households consume nearly double the European average, and in 2022, 42% of drinking water was lost due to leaks. The adoption of smart meters in Italy could revolutionize water management by reducing waste, saving resources, and alleviating financial impacts on distributors and consumers alike. The objective of this thesis is to build upon prior literature on predictive modeling of household water demand in order to develop a scalable model capable of accurately forecasting residential consumption patterns and detecting leaks in the water distribution network, solely from existing smart meter data. The model is based on an analysis of real smart meter readings collected across a broad region in northern Italy, thereby eliminating the need for costly additional sensor equipment or manual on-site inspection. By leveraging the extensive smart meter infrastructure already in place, the approach aims to provide a widely implementable and economically viable solution for predictive demand modeling and leak identification in residential areas.

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

Residential water leaks are a pervasive problem, often going unnoticed for extended periods, which can lead to substantial water waste and increased costs for both consumers and suppliers.

Current methods for leak detection in residential settings primarily include visual inspections, acoustic methods, and the use of district metered areas (DMAs). Visual inspections are often time-consuming and rely heavily on the experience of the inspector, while acoustic methods, which detect the distinct sounds of water escaping under high pressure from pipe leaks or cracks, require the installation of dedicated sensors along the pipe network or require an operator to be within 1-1.5 meters of the leak with a detection device. DMAs, which involve segmenting the water network to monitor flow and pressure, can be effective but often require significant infrastructure changes and investment. Despite their utility, these methods can be reactive rather than proactive, failing to provide immediate and continuous monitoring of water usage patterns.

The implementation of smart metering infrastructure provides a dual opportunity to predict daily water demand and to detect leaks without incurring additional costs for metering and supplier companies. Smart meters, with their ability to provide real-time, high-resolution water usage data, have opened up new avenues for enhancing water management practices. By continuously monitoring water consumption at a finer temporal resolution, smart meters generate large volumes of data that, when analyzed statistically, can reveal consumption patterns, identify anomalies, and signal the presence of leaks.

In Italy, residential water consumption statistics reflect a compelling case for the necessity of improved water management. According to a study carried out by the Eurispes Institute on the state of water in Italy, Italy places third in Europe in the ranking of countries with the greatest water availability, behind only Sweden and France; yet, at the same time, it is the country with the highest per capita consumption of drinking water and the second highest consumption in agriculture. The daily average household water consumption in Italy stands approximately between 150 and 240 liters per person,but this statistic is heavily skewed by the amount of water wasted or lost before it even reaches the consumer. The price of drinking water in Italy is among the lowest in Europe per cubic meter, yet in recent years, as prices have increased and supply diminished, many are making an effort to decrease their water consumption. Despite this, people may be using a lot more than they think. On average, Europeans consume 125 liters of water a day, while Italians consume 236 liters. [1] According to the Italian National Institute of Statistics 42 per cent of drinking water was lost in 2022, meaning that a significant percentage of treated water was lost to leaks before it even reached consumers. This not only represents a loss of a precious resource but also a financial loss to water suppliers and an unnecessary increase in water bills for consumers. This thesis posits that smart metering-generated water consumption data, when elaborated and analyzed effectively, can enable the early detection of household water leaks without incurring additional costs for metering infrastructure or supplier companies.

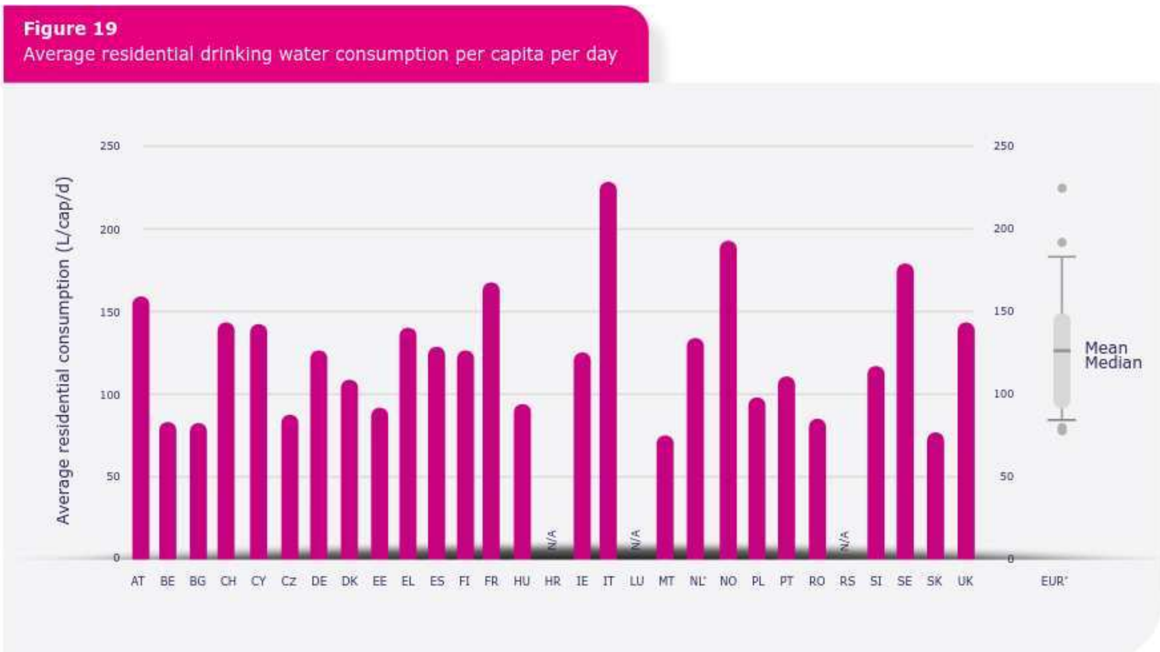

Figure 1.1: Average residential drinking water consumption per capita per day in each European country[2]

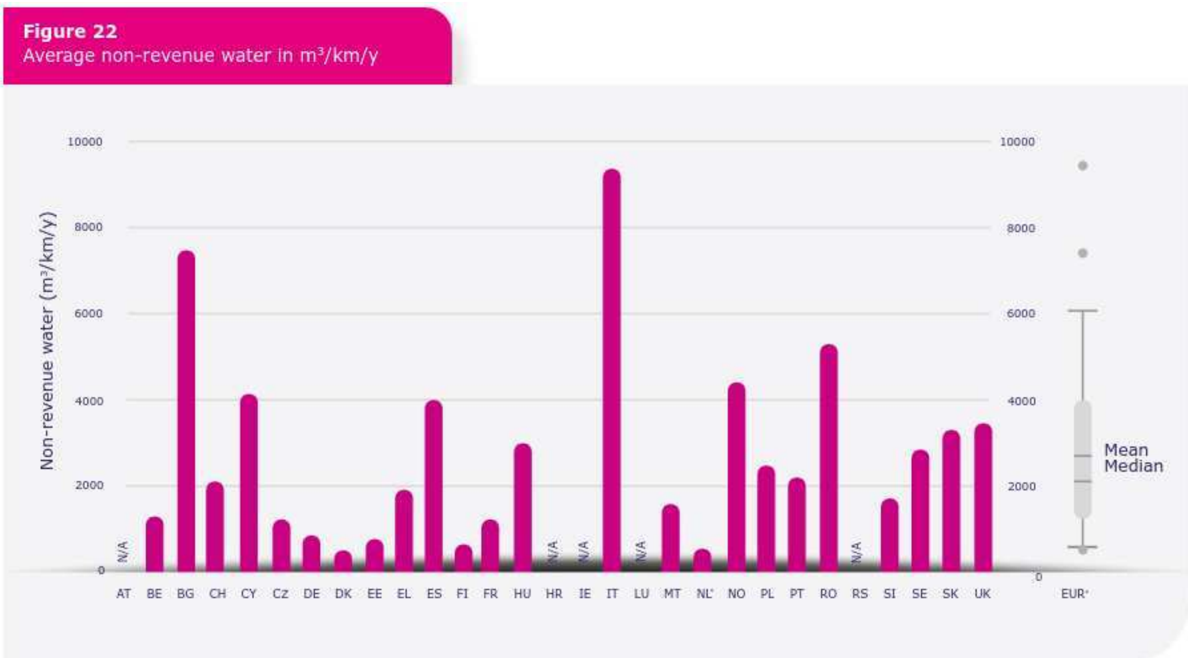

Figure 1.2: Average non-revenue water in each European country[2]

1.2 Thesis structure

Chapter 2 – Background

This chapter provides an overview of water sub-metering systems that use the wireless M-Bus communication protocol to enable automated meter readings. These systems are commonly used for multi-tenant buildings to allow per-unit water billing based on actual consumption. The typical architecture consists of individual water meters with integrated wireless M-Bus connectivity installed in each unit, MUC or IoT gateways that collect wireless transmissions from the meters, and backend software to manage the meter data. The meters used must comply with regulations like the EU's Measuring Instruments Directive for billing purposes. The wireless M-Bus standard defines various communication modes optimized for different applications, two-way communication between meters and data collectors, and open data formats like OBIS codes to ensure interoperability. The gateways receive transmissions from meters at regular intervals and forward the encrypted data payloads via cellular connectivity to the back office system. This enables completely automated and accurate monitoring and billing of water consumption for each tenant. Overall, wireless M-Bus is a key enabling technology for sub-metering systems due to its optimized architecture for high-density meter environments.

Chapter 3 – Methodology

This chapter delineates the methodological framework adopted for the research, outlining the original procedural strategy intended and the subsequent adjustments made during its implementation. It elaborates on the actions taken regarding the collected data, such as the volume of data gathered, the sources from which it was acquired, and the specific details that were either omitted or preserved to ensure anonymity without compromising the data utility for further analysis. The chapter also details the data processing methods used to prepare the dataset for the construction of a predictive model. This will include a thorough explanation of the criteria for data selection and exclusion, such as the removal of readings that would be anomalous from individual residences and the exclusion of meters with insufficient daily readings, which were deemed inadequate for a detailed hour-by-hour analysis. Additionally, the challenges faced during the research will be discussed, particularly those that necessitated changes in the methodology, along with the reasons behind these changes. A particular focus will be placed on the inability to replicate a consumption curve identical to that of the reference model, along with a discussion of the potential factors contributing to this discrepancy.

Chapter 4 – Experiments and results analysis

This chapter will explore in detail possible implementations of the predictive model for leak detection in the water pipe network of residential areas and which kind of approach to take to effectively determine when water leaks are happening without incurring in additional costs; for example, how the water demand curve of a real meter compares to the one generated by the predictive model and how significant of a difference between the two would guarantee an accurate detection without requiring an on–site operator.

Based on this premise, this chapter will also touch on an unsuccessful trial which had the purpose of simulating the water consumption of a single household across the span of a month using a water pump regulated by a programmable valve and monitored by a smart meter. The failure of this experiment was not due to a fault in the predictive model, as it was never applied, but rather due to the inability of the equipment to faithfully replicate household water consumption.
Lastly, it will give an overview of the results obtained and a brief analysis.

Chapter 5 – Conclusion

This chapter aims to do an overall conclusion of the work done and some final observations.

Chapter 2

Background

2.1 Water Meters

Water meters are legally relevant devices for measuring water flow for billing purposes. As such, they are subject to the European Measuring Instruments Directive (MID) Directive 2014/32/EU[3], which applies to measuring instruments used for legally relevant measurements. The purpose of the MID is to harmonize the requirements that these instruments must meet to be eligible for sale in the European Union. These requirements relate to metrological performance, physical and mechanical characteristics, and legal obligations. The directive aims to protect the market and consumers who use measurements from these instruments daily, and is a mandatory requirement for the CE certification marking. Therefore, all water meters used for billing, including sub-metering water meters, must be MID compliant to reliably measure water flow and consumption.

Water flow meters for monitoring systems typically fall into four main categories:

1. Positive Displacement Meters (PD Meters): PD Meters represent a class of water metering devices predominantly utilized in residential and small commercial environments. Notable for their high accuracy at low to medium flow rates, these mechanical meters incorporate a measuring component that moves in a manner directly proportional to the water volume traversing the meter. An internal magnet is subsequently engaged, driving the register to precisely record the volume of water that has traversed the meter.

2. Velocity Flow Meters: Velocity flow meters quantify the velocity and rate at which water passes through the meter. These mechanical meters transform the flow's velocity into a volumetric measurement to accurately assess consumption. Within the category of velocity flow meters, there are three distinct sub-types: single-jet and multi-jet water meters,

turbine water meters, and compound water meters. Each subtype is designed to cater to specific flow rate conditions and measurement requirements.

(a) Single-jet and multi-jet water meters employ blade rotation driven by water jets to measure flow, with an integrated strainer grid protecting against clogging. They are effective for low flows and small pipes, often used domestically for secondary billing, monitoring, and batch control, along with industrial applications. Single jet meters use one port to create a jet that impacts a turbine, transmitting motion to a display to measure volume; the simple, reliable design suits sub-metering roles. Multi-jet meters utilize multiple ports surrounding a chamber, evenly distributing force across an impeller to retain accuracy at low flows, making them ideal for large industrial installations monitoring high usage in real-time.

(b) Turbine Flow Meters: Turbine-type meters are well-suited for large diameter pipes with low flow rates and high volumes. They operate by using angled rotor blades that rotate in response to fluid movement, either clockwise or counterclockwise. Blade speed is measured by attaching a magnet to the rotor which allows accurate speed calculations regardless of flow direction. The responsiveness of the angled blades makes turbine meters an effective option for measuring low velocity flows in high volume applications.

(c) Compound water meters employ both turbine and positive displacement technologies to accurately measure flows with rapid, wide fluctuations in water demand. They utilize two meters - a large one for high flows and a smaller one for low flows. The readings from each meter are summed to determine total usage.

3. Electromagnetic Water Meters: Electromagnetic meters, or mag meters, are highly accurate volumetric flow meters containing no moving parts, which reduces maintenance and repair costs. They operate by using a magnetic field to route liquid through a pipe. As the liquid flows through this field, a voltage signal proportional to the flow rate is generated. The faster the flow of water, the greater the voltage, which is then translated into a reading.

4. Ultrasonic flow meters employ ultrasound technology to precisely measure the velocity of fluids in real-time, thereby enabling accurate data collection and avoiding billing disputes. With no internal moving components, these meters also have an extended operational lifetime. Two primary ultrasonic metering methods exist: Transit Time and Doppler. Transit Time ultrasonic meters allow a high degree of accuracy even at low flow rates. Their versatile installation allows them to retrofit into spaces with strict spatial constraints. Doppler ultrasonic meters utilize wrap-around sensors external to the pipe itself, rendering

them relatively inexpensive and well-suited for applications involving flows like mining and sewage where solids could be present. While Doppler technology is less accurate for clean liquids, it still provides sufficient approximate readings.

2.2 About sub-metering

Water metering can be categorized into three main types: Direct metering, master metering, and sub-metering. Direct metering involves every flat or property having its own meter, with the utility provider responsible for both the service and the billing based on actual usage. Master metering, on the other hand, uses a single meter for an entire block of flats, with the utility billing the building owner or property manager, who then divides the bill among the residents using a predetermined method, which can sometimes lead to unfair cost distribution. Sub-metering, also known as automatic meter reading (AMR) refers to the practice of measuring water usage at a more granular level than what is required for utility billing; this could include tracking water consumption in individual units within a building complex, apartment building, or even down to specific systems or devices within a facility, enabling tenants to be billed based on their actual consumption.

Sub-metering offers numerous benefits for both tenants and property managers. Tenants gain from a more equitable billing system that encourages mindful water usage by allowing them to directly influence their utility bills through personal consumption choices, which can lead to reduced overall utility costs. Property managers benefit from the ability to verify utility bills and promptly detect damage within the water network. Sub-metering facilitates the monitoring of water consumption patterns, enabling the early identification of leaks and maintenance needs, potentially curbing water loss and damage. When implemented effectively, sub-metering can lower utility expenses and minimize administrative tasks, as it allows for the automatic calculation and billing of individual utility charges to each resident, streamlining the process for property management.

An application of sub-metering can be demonstrated in an apartment building that has a single cold-water meter provided by a water distribution company. As illustrated in Figure 2.1, this example building contains two stairwells, each with two floors, and three apartments per floor. To enable correct water billing for each apartment, a communicating master sub-meter is installed downstream of the utility water meter, and individual smart sub-meters are placed in each apartment unit. The purpose of this sub-metering arrangement is to monitor the water consumption of each apartment and bill the residents accordingly based on usage. The master sub-meter provides a reference measurement to validate the aggregate consumption of all apartments matches the total water usage measured by the utility company's main building meter.

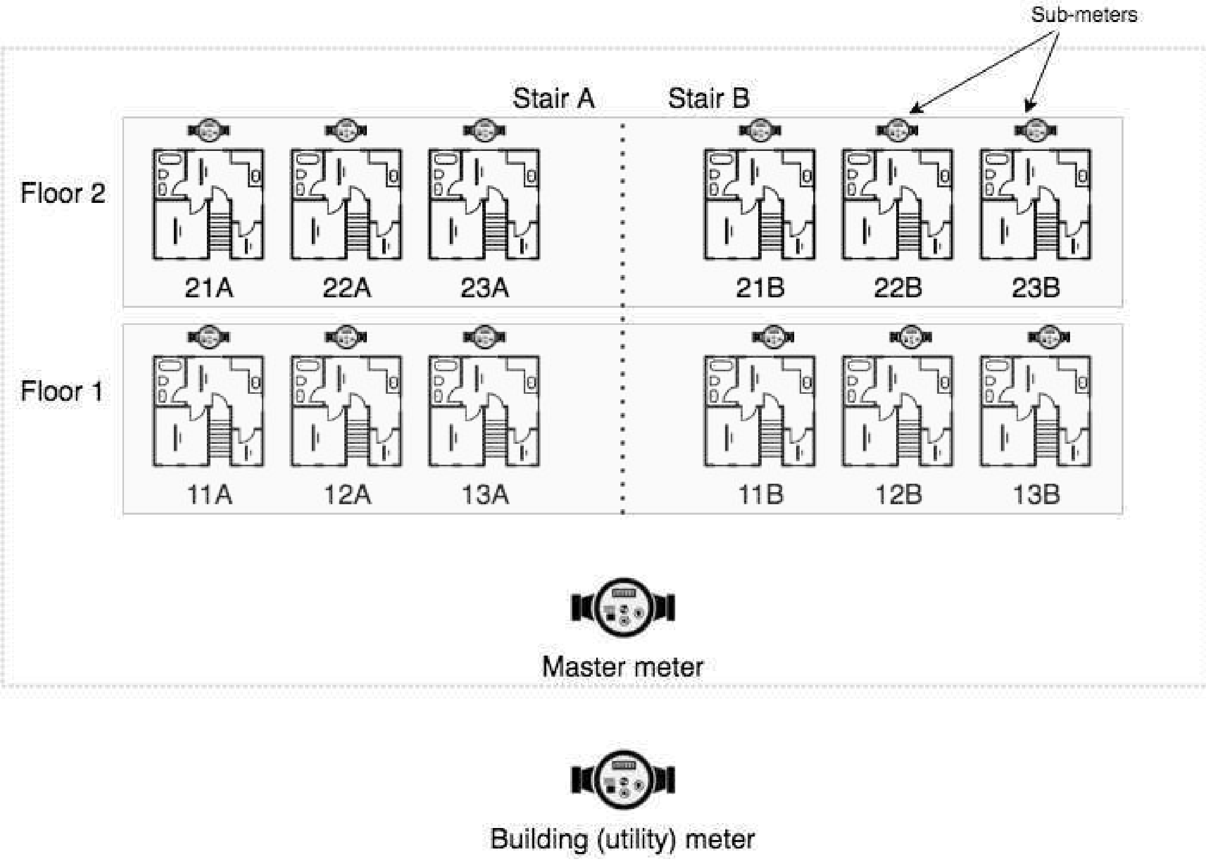

Figure 2.1: Schematic diagram of an example application of submetering in an apartment building.

2.2.1 Water sub-meters

The water meters utilized in this study were primarily manufactured by Maddalena S.p.A. and consisted of two distinct models:

1. SJ PLUS-EVO[4]: a single-jet traditional meter, equipped with a clip-on wireless MBUS communication module, RADIO EVO[5], operating at 868 MHz
2. ELECTO SJ[6]: a single-jet meter with an integrated wireless MBUS communication module, operating at 868 MHz.

Both of these meters support a fully compliant wireless MBUS protocol stack and OMS application stack including security (cryptography) features.

Figure 2.2: The SJ PLUS-EVO meter is available in cold or warm water versions. The model used for the thesis is the one for cold water measurements.

Figure 2.3: The ELECTO SJ single-jet water meter

2.3 Data acquisition architecture

Metering data is transmitted from the metering device to a data collector (MUC, concentrator, or IoT gateway) where it is stored and subsequently forwarded to the head-end system and meter data management system operated by the sub-metering service provider. The data collector can be permanently installed within the building or utilized as a mobile readout device by human operators conducting walk-by or drive-by data collection. For mobile collectors, operators either walk or drive in proximity to each property address to gather readings using a data receiver that retrieves the transmitted data from the built-in radio transmitter connected to each meter.

A key enabler of submetering technology is the use of a wireless communication protocol shared between the metering device and data collector that is based on an open standard for automated meter read-outs. The most widely adopted wireless protocol for smart metering applications in Europe is Wireless M-Bus, also known as Wireless Meter-Bus, a European standard that specifies communication between utility meters, data loggers, concentrators, and smart meter gateways. Originally developed as a standard system for networking and remotely reading utility meters across Europe, Wireless M-Bus is now also being utilized as the foundation for Advanced Metering Infrastructure (AMI) and is one of the primary communication protocols deployed for smart gas metering in Italy[7], [8]. The Wireless M-Bus protocol is designed for battery-powered devices like water meters to operate autonomously for up to 10 years. This extended lifespan is achieved by minimizing radio usage, keeping the module in a low-power mode whenever possible, and optimizing awake and transmit procedures. Following this principle, the communication is always initiated by the smart meter, not the MUC, which is more suitable for prolonging the sensor battery life.

Smart metering devices require robust, long range wireless communication. The frequencies most commonly utilized are around 868MHz, 434MHz, and 169MHz, which represent license-free bands in Europe that enable superior radio wave propagation compared to 2.4GHz. Operating within these license-free bands allows wireless radio signals to reach challenging locations such as underground meters or those inside buildings with multiple obstructions and walls. Additionally, the use of license-free bands provides cost benefits to utility companies implementing smart metering solutions, as licensing is not required.

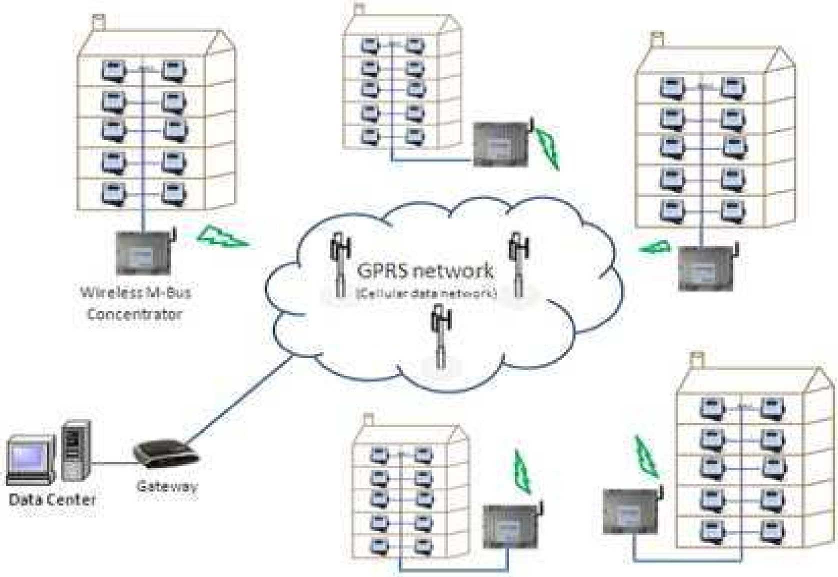

Figure 2.4: An automated meter reading system typically includes the following hardware components: Utility meters, transmitters, repeaters, a gateway, and billing software.

2.4 Wireless MBUS

The traditional OSI model shown on the left hand side of Figure is commonly used to depict stack architecture for various protocols where each layer has specific functions.

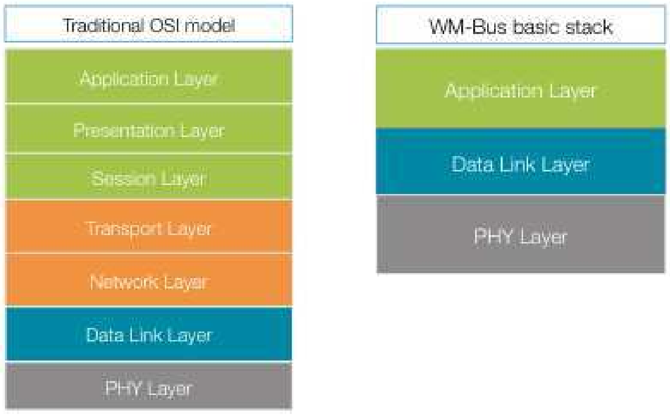

Figure 2.5: The traditional OSI model, compared to the WM-BUS basic stack

Originally, Wireless M-Bus utilized slightly different and less stringent requirements for the protocol layers. The core layers consist of the application layer, data link layer, and physical layer. However, more recent implementations of the protocol have started incorporating additional transport and network layers to enable advanced security features such as authentication as well as routing capabilities for larger scale networks.

2.5 International Harmonization

With the M/441 mandate, the European parliament empowered the standards organizations European Committee for Standardization (CEN), European Committee for Electrotechnical Standardization (CENELEC), and European Telecommunications Standards Institute (ETSI) to develop an interoperable communication framework for smart metering across Europe. To promote adoption and interoperability, the Wireless M-Bus protocol and its application layers are defined by a set of formal technical standard documents. Specifically for Wireless M-Bus in the sub GHz frequency bands, there are several key documents including: ETSI EN 300-220, which delineates permitted frequencies, bandwidths, emission limits, etc. for all sub GHz wireless products, not just Wireless M-Bus; and EN13757, published in 8 parts by CEN and more specifically focused on M-Bus, addressing each part of the solution spanning the physical layer to the application layer[9].

2.6 Water Metering application protocol

The Open Metering System (OMS) specification[10] was developed to meet the need for interoperable solutions for meter reading and a unified approach across different media types including electricity, gas, heat, and water. The OMS specification is founded on established standards with primary communication based on the M-Bus standard (wired or wireless), EN 13757. The specification defines a Multi Utility Communication (MUC) device that acts as an intelligent data concentrator between the automated meter management (AMM) back office system (for billing or other functions) and the metering and actuator devices.
OMS delineates three types of communication flows: primary communication between meters and the MUC; secondary communication as an extension of the primary communication utilizing simple repeaters or multi-hop routing; and tertiary communication between the MUC and the back office AMM system.

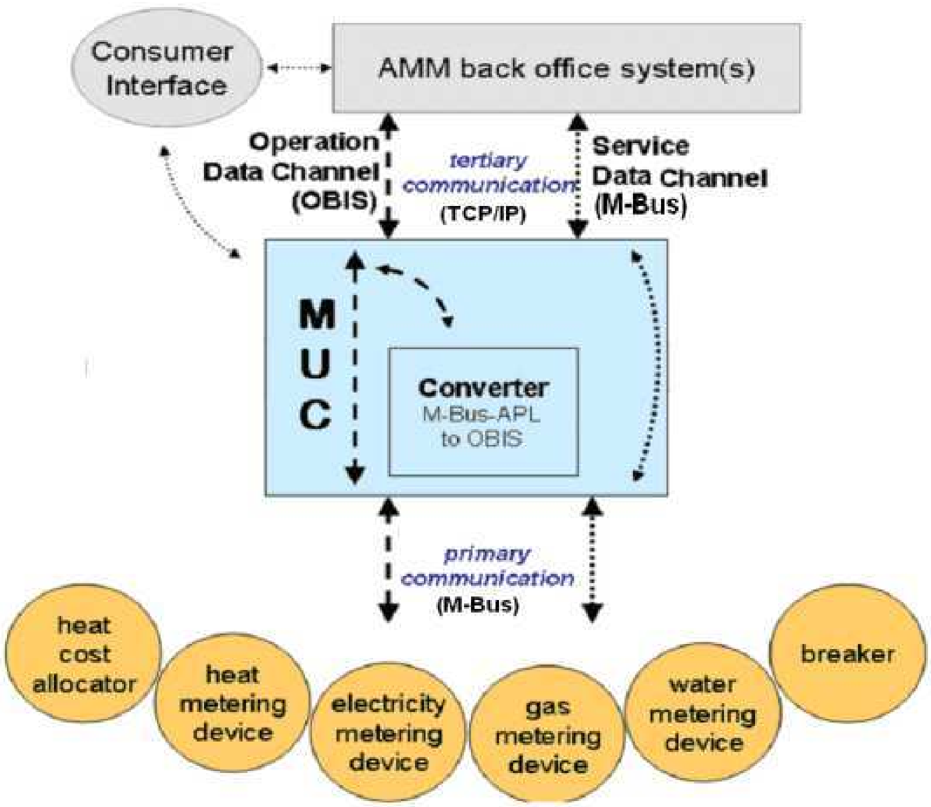

Figure 2.6: A schematic representation of the Open Metering Standard (OMS) specification and the three types of communication flows.[11]

The data format specified for standard Open Metering System applications is Object Identification System (OBIS) coded values[12], [13]. OBIS is Part 61 of IEC62056 which defines identification codes for commonly used data items in electricity metering equipment, and has been adapted for other utilities including gas and water. OBIS codes are 6-byte numbers that uniquely identify an object in a logical device. They identify data items in energy metering equipment through a hierarchical structure utilizing six value groups A through F: Group A specifies the medium (for example, 0=Abstract Objects, 1=Electricity, 7=Gas, 8=Cold Water); Group B indicates the channel; Group C denotes the physical value (Current, Voltage, Energy, Liters, Cubic Meters); Group D identifies processing types or algorithms applied to quantities in Groups A and C; Group E indicates further processing or classification of quantities in Groups A to D[11]. The MUC translates the M-Bus application data format into OBIS before sending it to the AMM via the operations data channel, while also supporting direct M-Bus formatted data on a separate service data channel from MUC to AMM.

2.7 MUC/Gateway

The MUC/IoT gateways used in this study are the WebdynEasy WMBUS Hub gateways from French company Webdyn.

Figure 2.7: A Webdyn WHMBUS Hub gateway

The WebdynEasy is a standalone data collector for meters and wireless M-Bus sensors that can operate on batteries without external power supply. The purpose of the WebdynEasy WM-Bus hub is to gather data from Wireless M-Bus-enabled devices including meters (water, gas, electricity) and sensors (temperature, humidity, etc.). This energy-efficient technology can achieve hub battery lifespan exceeding 10 years. WebdynEasy WM-Bus enables configuration of private communication networks. WM-Bus communication is point-to-point in a non-operated manner. Meters and sensors regularly transmit data frames that are received by the WM-Bus hub, which stores the raw encrypted or transparent data contained in the frames. The WebdynEasy WM-Bus hub periodically sends a BSON-formatted file containing the aggregated raw data from all sensors to a server via FTP using its integrated LTE-M/NB-IoT/2G modem.

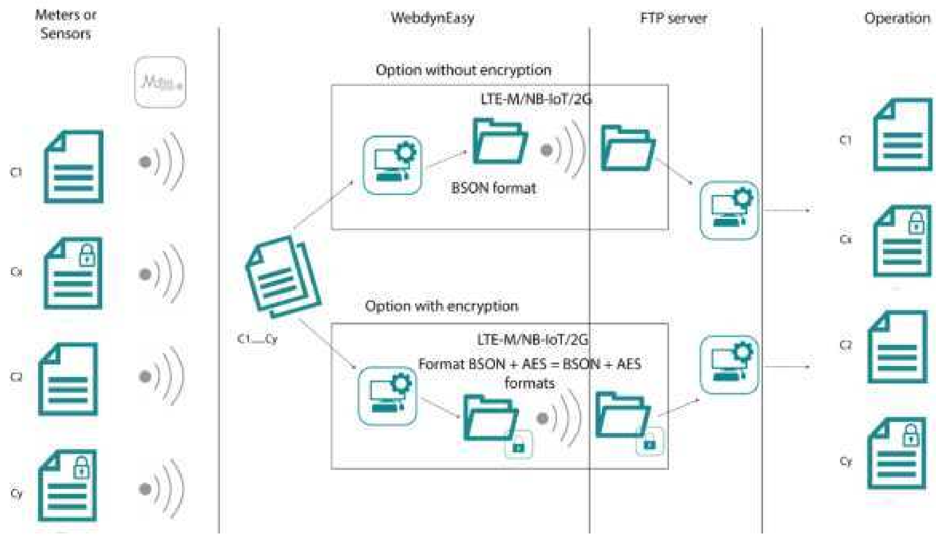

Figure 2.8: A schematic diagram of the WebdynEasy VMBus communication network[14]

The BSON format, an abbreviation for Binary JSON, represents a binary-encoded form of serialization that resembles JSON documents. Similar to JSON, BSON allows for nesting documents and arrays within other documents and arrays. Additionally, BSON includes supplementary features that enable the encoding of data types not included in the standard JSON specification[15].


```json
{
    "uid": "WE_1234",
    "source": "schedule",
    "TS": 1560068897,
    "frameCount": 2,
    "data": [
        {
            "T": 1560066602,
            "R": 199,
            "F": {
            "$type": "00",
            "$binary": "fAviGIskafDwA4E0UkO26w=="
                }
        },
        {
            "T": 15600645887,
            "R": 178,
            "F": {
            "$type": "00",
            "$binary": "+plAP1iAwE0df2KBY1N7Iodr/LAFS2CrXTrUsY3wy731VN113/9UmO\CiglwR"
                }
        }
        ],
"crc": 0
}
```


Figure 2.9: An example of a BSON frame with meter data represented as a JSON object

Chapter 3

Methodology

3.1 Approach

This project originally started during my internship at smart metering software company Inkwelld Data Ltd. and was inspired by an analysis of statistics on water usage and waste across Italy, which revealed that leaks in aging pipe infrastructure were responsible for immense volumes of wasted water annually. I was presented with the task of devising a cost-effective solution for detecting water leaks in residential buildings using only data from smart water meters. The wealth of information being generated by smart metering systems presents an untapped opportunity to extract actionable insights about potential leaks, without requiring any additional hardware investments by utilities or metering companies.

For the purpose of this thesis, the first step was obtaining access to a dataset of real smart meter readings from wireless MBUS systems installed in homes across Northern Italy. My goal was to leverage this data to build a robust statistical model capable of predicting expected normal water usage and consumption patterns in residential buildings on a daily basis. Once developed and validated, this model could then be used to analyze incoming smart meter data for discrepancies between actual and predicted daily water demand. I hypothesized that when actual consumption deviated significantly higher than the levels expected by the model, it would provide a strong indication of a potential leak on the property. Detecting when actual water use diverges meaningfully from usual patterns could enable utilities to proactively alert customers to likely leaks and ensure timely repair.

Overall, this project presented a meaningful opportunity to create value by devising a clever solution that relied exclusively on mining insights from previously underutilized data.

In the following sections, I will elaborate on the steps taken to anonymize and organize the raw data received, then process it into a format suitable for building a predictive model. The data processing approach follows a similar methodology to Rudy Gargano, Carla Tricarico, Giuseppe del Giudice and Francesco Granata's study "A stochastic model for daily residential water demand"[16], which I used as a basis for creating the statistical model.

3.2 Developing Stochastic Models of Daily Water Demand: Theoretical Framework

A robust methodology for developing a statistically and mathematically accurate predictive model for water demand necessitated reviewing existing academic literature on the subject. The aforementioned study[16] provided the most applicable framework, including clear delineation of the techniques utilized to develop their predictive model, enabling replication of their approach. In order to manage the variability in water demand, this paper puts forth a novel approach of creating a single versatile probability distribution, termed the mixed distribution (MD), that can be applied uniformly across all times of day, rather than considering multiple separate distributions for each time period. Furthermore, the relationships for estimating the standard deviation parameter of the proposed mixed distribution are expanded upon in detail, highlighting the dependence on the number of users and meters under examination.

3.2.1 Mixed distribution parameters

Firstly, the value that will represent actual water demand in the distribution is made dimensionless by creating a specific coefficient, termed mean demand coefficient and referred to with the symbol μ_{CD}(t), that is found as the ratio between the average water demand registered at time t (μ_Q(t)) and the average daily water demand for an entire day (μ_Q)[16].

$$\mu_{CD}(t) = \frac{\mu_Q(t)}{\mu_Q} \tag{3.1}$$

The variation coefficient parameter of the MD at time t (CV(t)) is obtained in relation to the mean demand at time t and the number of users, which in my own research I interpreted as individual meters. According to study, the variation coefficient is inversely proportional to both the number of users and the demand coefficient at time t. This relationship is characterized by the following equation[16]:

$$CV(t) = 0.1 + \frac{6}{\left(\frac{1}{4}\mu_{CD}(t) \cdot N_{us}\right)^{3/4}} \tag{3.2}$$

The probability distribution of null water demand, referred to with the symbol F0, was estimated in the reference paper through elaboration of field data. As neither the dataset nor the specific methodology utilized for estimating this specific parameter were made available by the authors, the results presented in the aforementioned publication were reused as a basis for subsequent calculations and elaborations within this thesis, in lieu of access to the primary data and analytic procedures.

The final parameter essential for the construction of the mixed distribution is the coefficient corresponding to the non-zero water demand, which exhibits a significant dependency on the number of users, and therefore meters, considered in the analysis.

3.2.2 The importance of sample size

The importance of utilizing an adequately large sample size is emphasized in the reference study, as it is noted that this simplifies the estimation of water demand when demand is nonzero. Specifically, with a sufficiently large sample, the mixed distribution coefficient associated with nonzero water demand can be treated as a single random variable that follows a unique random variable that is continuously positive.

3.2.3 The importance of sample size

The most salient finding from the reference paper, as it pertains to the current thesis, is the derivation of the equation to model demand flow at each discrete time step t. This equation[16], which is contingent on the previously discussed parameters, is as follows:

$$Q(t) = \mu_Q \left[1 - \frac{sqrt3}{\pi} ln \frac{1 - F(t)}{F(t) - F_o(t)} CV(t) \right] \mu_{CD}(t) \tag{3.3}$$

where F(t) represents a continuous positive random variable, with the sole restriction that F(t) must be greater than F0(t) for all values of t.

3.3 Dataset Acquisition and Refinement

As previously mentioned, the data-set is a sample of meter readings from wireless MBUS systems based on the digital twin concept by Inkwell Data Ltd and deployed in residential buildings in Northern Italy, taken between the dates of August 11th 2023 and December 27th 2023.
Figures 3.1 and 3.2 below present a subset of the initial readings prior to any further analysis or interpretation, with names and addresses obscured.

	A	B	C	D	E	F	G	H	
1	gateway_identifier	read_time	condominium_id	sub_condominium_id	immobile_id	meter_identifier	type	has_errors	meta
2	fb16165a-7070-11eb-94e7-005056411cd8	2023-12-27 09:19:05.000	19338	1df52120-1dbb-4800-a3c8-5d682e5be398	eb91f234-6887-42d9-adef-8840295edf9a	75b7aa5a-333e-420a-b64e-838523efc552	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
3	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:17:58.000	19526	52988711-4171-4858-95a8-56312cdc2404	d8eafb18-f1d2-4ce5-b4b4-04bd5bbf05b1	838892ce-7071-11eb-ba5c-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
4	faf6d1b4-7070-11eb-a74a-005056411cd8	2023-12-27 09:16:19.000	39	ab1f2ae9-9ad2-4cb8-b11c-23dd77414da3	2b437a60-a64d-4c74-84fa-4336f21d5197	4a31afc0-94a7-43f4-87e6-d20de5432952	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
5	faf6d1b4-7070-11eb-a74a-005056411cd8	2023-12-27 09:15:56.000	39	84ee78da-adb6-4f3a-a5d6-a7a0b225b1ab	610efb32-21f1-4bae-9590-cc77ce5a2a64	c7e0e273-74c3-4efe-bbef-bc182bce394f	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
6	faf6d1b4-7070-11eb-a74a-005056411cd8	2023-12-27 09:15:56.000	39	ab1f2ae9-9ad2-4cb8-b11c-23dd77414da3	2b437a60-a64d-4c74-84fa-4336f21d5197	444e26af-0729-437e-8c86-1ad0449d21e7	AF	true	{"errors": {"backflow_is": 0, "overflow_is": 0
7	f2a4fe00-7070-11eb-af06-005056411cd8	2023-12-27 09:15:45.000	19219	fba2b705-ebd4-45a8-af09-45eb045bdee1	e8415f1e-2939-44b8-8675-714d243b838f	09a23d54-d9ad-49ea-ba7e-76614dbf0793	AF	true	{"errors": {"firmware_check_sum_error": 0, "
8	f2a4fe00-7070-11eb-af06-005056411cd8	2023-12-27 09:15:36.000	19219	fba2b705-ebd4-45a8-af09-45eb045bdee1	05d7d234-7e76-40cd-a775-3e030d6f2a99	0509e297-1181-4336-982c-98c0a4d879ec	AF	true	{"errors": {"firmware_check_sum_error": 0, "
9	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:15:27.000	19526	52988711-4171-4858-95a8-56312cdc2404	287807c0-7d1f-44a0-9573-307f53a1ce84	83962ce0-7071-11eb-b061-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
10	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:14:46.000	19526	52988711-4171-4858-95a8-56312cdc2404	854a6a7a-15e0-4c64-8a7b-686a4c9fabbc	83a7360c-7071-11eb-834c-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
11	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:14:07.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	193a7dbd-0227-4e3b-9d33-44ac050b5b3	93ba1406-1bac-4363-b575-6eac36b145c1	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
12	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:14:03.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	48b2d39d-33e8-44a4-bdae-b9293765275	224c069a-cd88-41b7-b5f3-95708684ff79	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
13	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:14:00.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	48b2d39d-33e8-44a4-bdae-b9293765275	464916fd-6dd3-485f-a383-a157c085f01b	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
14	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:59.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	6b40c5de-ac5c-4810-a52e-9b75d804cfea	ee970af5-3a25-420e-889e-493f7b503d60	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
15	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:55.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	fd33b231-ef51-4306-9a91-cdd1cde97e88	00e61343-dd77-446e-ae39-729b61cdeb72	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
16	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:50.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	b0de6bcd-6c78-4ffd-9875-899d64f2a9e4	ba13d1b9-f08f-47a6-8c17-954c75bf5d37	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
17	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:49.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	e1f96982-d730-4678-9588-5dda1b7816a1	24a44d2f-d38c-4d69-9313-64853315d4b6	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
18	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:48.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	e1f96982-d730-4678-9588-5dda1b7816a1	d8544a53-7934-4452-8441-0ea953822858	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
19	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:46.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	178a268c-f43a-4ca2-ac14-91bd6e7c74ae	ae7343dc-f4df-430b-87d6-041eba64b749	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
20	e2bafcc4-7070-11eb-a4d3-005056411cd8	2023-12-27 09:13:41.000	19356	b1fcb3c8-dac0-45ea-b186-1902310d7890	a572ba7c-b88d-427c-b6ef-f33eb6d35cae	16531ae6-e90d-4cd4-89dc-d19d873d154f	ACS	false	{"errors": {"backflow_is": 0, "overflow_is": 0
21	f8bfe1d8-7070-11eb-9ca2-005056411cd8	2023-12-27 09:13:29.000	19197	7cf99fed-fb1a-45b1-b17a-80302b6b4c7f	e63043bb-0cf0-4bbe-ab17-4c904623dd92	f9431494-7071-11eb-a59f-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
22	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:13:25.000	19526	52988711-4171-4858-95a8-56312cdc2404	e0ccf7bd-7ef7-46ff-9576-91484d034616	8344519a-7071-11eb-afed-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
23	f8bfe1d8-7070-11eb-9ca2-005056411cd8	2023-12-27 09:13:24.000	19197	6dd51147-4a43-4581-b853-796c67f25563	baedd997-dfbf-4cda-92df-c0c397e491ab	26232a80-7072-11eb-8602-005056411cd8	AF	true	{"errors": {"backflow_is": 0, "overflow_is": 0
24	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:12:45.000	19526	52988711-4171-4858-95a8-56312cdc2404	e3ca7117-f6b9-43ea-a9f3-d38b28e40582	83555b48-7071-11eb-8ecc-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
25	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:12:44.000	19526	52988711-4171-4858-95a8-56312cdc2404	3064a1dc-8871-4fd2-90e7-6f9e49afad9b	8335f6c2-7071-11eb-accd-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
26	ec328b14-7070-11eb-8da8-005056411cd8	2023-12-27 09:12:07.000	19526	52988711-4171-4858-95a8-56312cdc2404	d8eafb18-f1d2-4ce5-b4b4-04bd5bbf05b1	838892ce-7071-11eb-ba5c-005056411cd8	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0
27	f9fd73a8-7070-11eb-be06-005056411cd8	2023-12-27 09:11:45.000	19348	0e40c5f3-d867-4ab0-b7a1-646fcded5f4e	18cc5126-3af4-41a0-a568-321ef45ed514	c8d0953e-11f9-441e-8eac-b45fb2b60b61	AF	false	{"errors": {"backflow_is": 0, "overflow_is": 0

Figure 3.1: A list of readings collected on December 27th, 2023. (1)

	K	L	M	N	O	P	Q	R	S
1	id	sub_condominium	immobile	floor	flat_number	measure_unit	condominium	meter_id	values
2	69040561	[illegible]	[illegible]	Piano 3	CIV08_INT10	m3	[illegible]	1242377	{"volume": 402.118}
3	69040612	[illegible]	[illegible]	Piano Terra	PP DX	m3	[illegible]	1041774	{"volume": 165.195}
4	69040466	[illegible]	[illegible]	Piano 1	2	m3	[illegible]	1951	{"volume": 370.173}
5	69040467	[illegible]	[illegible]	Piano 2	4	m3	[illegible]	1938	{"volume": 1.21}
6	69040465	[illegible]	[illegible]	Piano 1	2	m3	[illegible]	1952	{"volume": 1.486}
7	69040464	[illegible]	[illegible]	Piano Terra		m3	[illegible]	1041453	{"volume": 5549.99}
8	69040463	[illegible]	[illegible]	Piano Terra		m3	[illegible]	1041454	{"volume": 9966828.71}
9	69040474	[illegible]	[illegible]	Piano Terra	P2 DX	m3	[illegible]	1041777	{"volume": 107.028}
10	69040473	[illegible]	[illegible]	Piano Terra	P2/P3	m3	[illegible]	1041778	{"volume": 462.54900000000004}
11	69040374	[illegible]	[illegible]	Piano 3	P.3	m3	[illegible]	124	{"volume": 68.31400000000001}
12	69040375	[illegible]	[illegible]	Piano 2	P.2SX	m3	[illegible]	883	{"volume": 211.67000000000002}
13	69040373	[illegible]	[illegible]	Piano 2	P.2SX	m3	[illegible]	121	{"volume": 199.04500000000002}
14	69040372	[illegible]	[illegible]	Piano Terra	PT	m3	[illegible]	106	{"volume": 204.178}
15	69040371	[illegible]	[illegible]	Piano 3	P.3	m3	[illegible]	890	{"volume": 231.79}
16	69040370	[illegible]	[illegible]	Piano 1	P.1	m3	[illegible]	111	{"volume": 137.833}
17	69040369	[illegible]	[illegible]	Piano Terra	PT	m3	[illegible]	105	{"volume": 34.389}
18	69040368	[illegible]	[illegible]	Piano Terra	PT	m3	[illegible]	865	{"volume": 148.123}
19	69040366	[illegible]	[illegible]	Piano 4	P.4CD	m3	[illegible]	894	{"volume": 167.562}
20	69040367	[illegible]	[illegible]	Piano Terra	0	m3	[illegible]	102	{"volume": 289.488}
21	69040360	[illegible]	[illegible]	Piano Terra	SX	m3	[illegible]	1040465	{"volume": 485.093}
22	69040471	[illegible]	[illegible]	Piano Terra	1	m3	[illegible]	1041779	{"volume": 1715.432}
23	69040361	[illegible]	[illegible]	Piano 2	SX	m3	[illegible]	1040461	{"volume": 414.92900000000003}
24	69040470	[illegible]	[illegible]	Piano interrato -1	S.SUO	m3	[illegible]	1041773	{"volume": 712.091}
25	69040472	[illegible]	[illegible]	Piano Terra	PP SX	m3	[illegible]	1041775	{"volume": 186.63400000000001}
26	69040469	[illegible]	[illegible]	Piano Terra	PP DX	m3	[illegible]	1041774	{"volume": 165.195}
27	69040350	[illegible]	[illegible]	Piano Terra		m3	[illegible]	1243840	{"volume": 5.197}

Figure 3.2: A list of readings collected on December 27th, 2023. (2) The figure has been split into two parts for more accessible viewing.

The raw data table contains many columns that were not needed for this analysis, such as gateway identifier, specific addresses, names, and most identification numbers aside from the meter IDs (meter_id column). Before removing any readings, I used the 'type' column to extract only cold water meters (tagged as AF, Acqua Fredda), the only ones relevant to this project. The extraneous columns were then removed, along with all readings from other types of meters. After this first elaboration I was left with 5.021.584 total readings from 7690 meters.

Figure 3.3 presents a table that has been refined to include only the data points pertinent to this project. These essential data points are the precise date and time of each reading, the meter ID, the measurement unit, and the corresponding value recorded by the meter at the specified time.

	time	id	unit	value
1	11-Aug-2023 03:18:24	1042452	'm3'	262.2120
2	11-Aug-2023 03:18:24	1241820	'm3'	188.0830
3	11-Aug-2023 03:18:30	1241823	'm3'	343.4690
4	11-Aug-2023 03:18:31	1042336	'm3'	283.3990
5	11-Aug-2023 03:18:38	1916	'm3'	204.4310
6	11-Aug-2023 03:18:41	1651	'm3'	68.9010
7	11-Aug-2023 03:18:44	1242362	'm3'	56.9170
8	11-Aug-2023 03:18:48	1041997	'm3'	377.6760
9	11-Aug-2023 03:18:51	1907	'm3'	142.9090
10	11-Aug-2023 03:18:55	1241356	'm3'	2.2131e+04
11	11-Aug-2023 03:19:00	1901	'm3'	64.6950
12	11-Aug-2023 03:19:00	1917	'm3'	2.0090
13	11-Aug-2023 03:19:19	1042327	'm3'	183.5950
14	11-Aug-2023 03:19:20	1041808	'm3'	478.4560
15	11-Aug-2023 03:19:20	1241302	'm3'	458.5270

Figure 3.3: After removing extraneous columns, the filtered dataset was imported into Matlab, a platform better suited for the quantitative data processing and modeling required for this research.

With a clean dataset containing only the necessary data points, I could begin deriving the parameters required to generate my own mixed distribution for water demand forecasting. First, I sorted the data by meter identification number and timestamp so that all readings for a given meter would be clustered together chronologically. I then wrote a function named get_daily_readings that accepts the full dataset and a meter ID as inputs and returns a subset of the data containing readings only for the specified meter. Calling this function on the first row of the clean data table allowed it to iterate through and extract daily readings for all meters in the dataset.

```
function [meter_selection] = get_daily_readings(ordinati_con_ore,meter_selection_index)
    %find id on indexed row
    meter_id = ordinati_con_ore.id(meter_selection_index);

    %select all readings from meter:meter_id
    id_singolo = (ordinati_con_ore.id == meter_id);

    %use selection to create new timetable from single meter data
    meter_selection = ordinati_con_ore(id_singolo,:);
end
```

With the ability to analyze each meter individually, I developed a function to derive water demand from the water readings. Since the readings for a given meter comprise a monotonically increasing time series, demand over the full five-month period could be computed by a simple subtraction of each reading from the prior reading. The function find_delta accepts the data for a single meter, as extracted by the get_daily_ function, as input. It returns a new column containing the calculated demand value for each data point, which can then be appended to the original data table.

```
function [total_demand, meter_selection] = find_delta(meter_selection)
    total_demand = 0;
    for i = 2:1:height(meter_selection)
        current_reading = meter_selection.value(i);
        previous_reading = meter_selection.value(i-1);
        delta = current_reading-previous_reading;

        meter_selection.demand(i) = delta;
        total_demand = total_demand+delta;
    end
end
```

3.4 Development of a predictive model using the acquired data

3.4.1 Extracting the mixed distribution parameters: μ_{CD}(t)

With water demand data calculated for each time step, I could begin deriving the parameters needed for the mixed distribution model, specifically μ_{CD}(t). To compute μ_{CD}, which would become another column in the dataset, I first required the average daily water demand and the average demand at each time step for every meter based on the equation[16]:

$$\mu_{CD}(t) = \frac{\mu_Q(t)}{\mu_Q} \tag{3.4}$$

Daily demand was obtained by subtracting the first reading of the day from the last, as with the per-time step demand. I added a column to the dataset containing the daily demand values for each meter. I also constructed a separate table, mean_meter_demand, to contain the average daily demand calculated across all days for each meter ID. With daily demand values calculated, I could determine the average water demand coefficient (μ_{CD}(t)) for each time step. I created a function called find_mean_by_hour that takes as input the readings for a single meter, sorted by hour to group readings from the same time of day across multiple dates, along with the table of mean daily demands per meter. The function first identifies the correct average daily demand value for the given meter based on its ID. It then averages the water demand recorded by that meter for each hour of the day. μ_{CD} is calculated for each hour as the ratio between the mean demand at that hour and the overall average daily demand. The function outputs a table with one row per hour of the day, associating each hour with the corresponding μ_{CD} value for a single meter.

```
function [average_table, mu_cd_table] = find_mean_by_hour(single_meter_readings, mean_meter_demand)
    i = 1;
    current_hour  = single_meter_readings.hour(1);
    previous_hour = current_hour;
    average = 0;
    average_table = zeros(24,3);
    average_table(:,1) = [0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22; 23];
    avg_count = 0;
    avg_index = 1;
    meter_id = single_meter_readings.id(1);

    mu_cd_table = zeros(24,2);
    mu_cd_table(:,1) = [0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15; 16; 17; 18; 19; 20; 21; 22; 23];


    daily_average = find_average_by_id(meter_id,mean_meter_demand);




    while i< height(single_meter_readings)

        while current_hour == previous_hour
            average = average + single_meter_readings.demand(i);
            avg_count = avg_count+1;

            i = i+1;
            if i >= height(single_meter_readings)

                break;
            else
                current_hour = single_meter_readings.hour(i);

            end

        end



        average = average/avg_count;

        if average  >= 0
            average_table(previous_hour+1, :) = [previous_hour, average, avg_count];

            if daily_average == 0

                mu_cd_table(:,:) = [];
                break;
            else
                mu_cd = average/daily_average;

                mu_cd_table(previous_hour+1, :) = [previous_hour, mu_cd];
            end
        end
        previous_hour = current_hour;

        avg_count = 0;
    end

end
```

Figure 3.4: single_meter_readings is a timetable obtained by sorting the elements of the timetable extracted by get_daily_readings by hour instead of date, mean_meter_demand is a table which contains, associated to each meter id, the mean daily demand for that meter.

Finally, I merged the μ_{CD} tables for all meters, excluding those with null average hourly demand values for all 24 hours. The null values indicated that these meters were inactive during the data collection period. The resulting consolidated table contains 24 rows representing each hour of the day, and 6,751 columns for the number of active meters remaining after the exclusion process. This final table provided the complete set of μ_{CD} parameters across all active meters and times of day required as inputs to the mixed distribution model.

After performing the data processing steps detailed in the preceding paragraphs, the dataset took on the structure shown in Figure 3.5. This table provides an overview of the final dataset format, including the original raw meter readings as well as the derived features such as per-time step demand, daily demand, and μ_{CD} values calculated for each meter at each hour of the day.

time	id	unit	value	hour	m_cd	daily_demand
11-Aug-2023 04:...	19	'm3'	85.5500	4	0	0
11-Aug-2023 16:...	19	'm3'	85.5500	16	0	0
11-Aug-2023 20:...	19	'm3'	85.6140	20	1	0.0640
12-Aug-2023 00:...	19	'm3'	85.6140	0	0	0
12-Aug-2023 04:...	19	'm3'	85.6140	4	0	0
12-Aug-2023 07:...	19	'm3'	85.6140	7	0	0
12-Aug-2023 11:...	19	'm3'	85.6180	11	0.0435	0
12-Aug-2023 15:...	19	'm3'	85.6380	15	0.2174	0
12-Aug-2023 19:...	19	'm3'	85.7060	19	0.7391	0
12-Aug-2023 23:...	19	'm3'	85.7060	23	0	0.0920
13-Aug-2023 03:...	19	'm3'	85.7060	3	0	0
13-Aug-2023 07:...	19	'm3'	85.7060	7	0	0
13-Aug-2023 11:...	19	'm3'	85.7140	11	0.4444	0
13-Aug-2023 15:...	19	'm3'	85.7240	15	0.5556	0
13-Aug-2023 19:...	19	'm3'	85.7240	19	0	0
13-Aug-2023 23:...	19	'm3'	85.7240	23	0	0.0180
14-Aug-2023 03:...	19	'm3'	85.7240	3	0	0
14-Aug-2023 07:...	19	'm3'	85.7240	7	0	0
14-Aug-2023 11:...	19	'm3'	86.5710	11	0.6016	0
14-Aug-2023 15:...	19	'm3'	87.1320	15	0.3984	0
14-Aug-2023 19:...	19	'm3'	87.1320	19	0	0
14-Aug-2023 23:...	19	'm3'	87.1320	23	0	1.4080
15-Aug-2023 03:...	19	'm3'	87.1320	3	0	0
15-Aug-2023 07:...	19	'm3'	87.1320	7	0	0
15-Aug-2023 11:...	19	'm3'	87.1320	11	0	0

Figure 3.5: An extract from the final data table, displaying all data points associated with meter id 19 in the first three days of data collection

3.4.2 Outlier Identification and Removal

A critical step in preparing the data for calculating distribution parameters was removing any outlier meters not associated with single household usage. The dataset contained numerous meters for entire condominium complexes rather than individual households. Since the mixed distribution aims to model household-level demand, these non-single residence meters had to be excluded to prevent skewing the results. Obtaining a homogeneous sample of only single household meters was essential for the distribution to reliably represent residential water usage.

To identify and discard outliers, I employed a simple filtering process based on maximum reasonable hourly and daily consumption levels for a single household. Specifically, consultation with Inkwell Data established thresholds where any meter registering over 4 m^3 of hourly usage would be deemed an outlier unrepresentative of the target single household population. Using this criteria, I filtered the full dataset to discard all meters exhibiting outlying consumption exceeding the 4 m^3 hourly threshold. This filtration isolated the dataset to meters aligning with expected normal ranges for a single household.

Additional dataset refinement included excluding meters with insufficient data, defined as those transmitting on average less than five readings per day. With the distribution intended to predict hourly water demand, meters lacking robust hourly information could not adequately contribute to modeling the parameters.

After completing the data filtering steps to remove outliers and meters with insufficient data, the final refined dataset retained a total sample of 5438 meters from which to derive the mixed distribution parameters.

3.4.3 Extracting the mixed distribution parameters: CV(t)

The calculation of the CV(t) parameter relied heavily on the prepared table containing the mean demand coefficients (μ_{CD}) for all meters, as these coefficients are part of the CV(t) equation. I estimated CV(t) at a 15 minute time step, resulting in an array of 96 elements representing each 15 minute interval. To compute each CV(t) value, the corresponding μ_{CD} values were extracted from the μ_{CD} table at four values per hour, or four per table row. These could be drawn from any random column, and thus any random meter, since the table already excluded dissimilar meters, retaining only residential household meters with comparable behaviors. Therefore, using μ_{CD} values from different meters for each 15 minute increment did not introduce significant discrepancies.

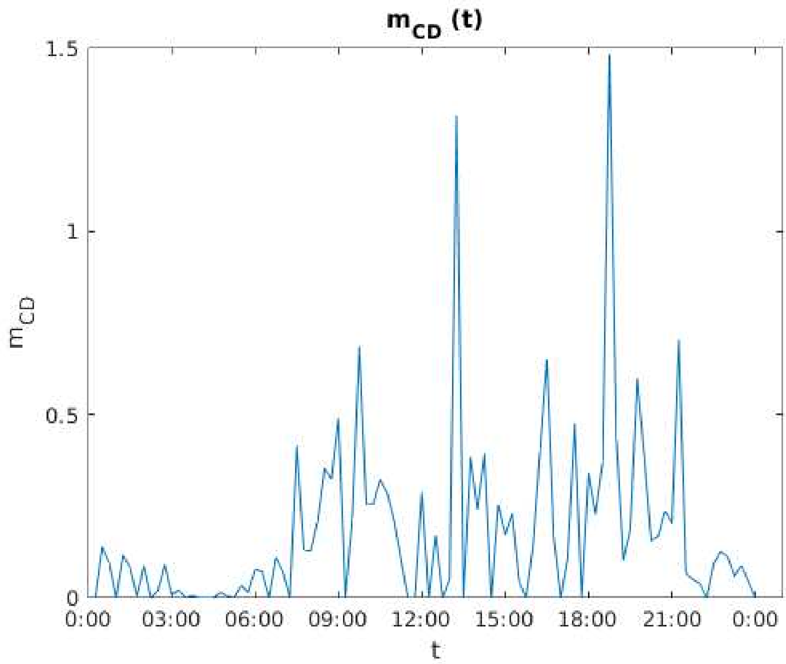

Figure 3.6: An example of extracted μ_{CD} values

To obtain the 96 μ_{CD} values needed from the larger table, I created a function called create_coefficients_arrays that extracts the requisite μ_{CD} data points and also retains a record of the μ_Q values for the meters from which each μ_{CD} was sampled. Maintaining this link between the extracted μ_{CD} coefficients and the corresponding meters' μ_Q parameters is necessary for computing the final predictive water demand curve equation.

```
function [coeff_array, mu_q] = create_coefficients_arrays(hourly_table, mean_meter_demand)
    hour = 1;
    i = 1;
    while hour <= 24
        for i = i:1:i+3
            random_extraction_index = randi([1, length(hourly_table)]);
            coeff_array(i) = hourly_table(hour, random_extraction_index);
            mu_q(i) = mean_meter_demand(random_extraction_index, 2);
        end
        i = i+1;
        hour = hour+1;
    end
end
```

Figure 3.7: The create_coefficients_arrays function enabled efficient extraction of the needed µcd data while preserving the meter information required for the final calculation.

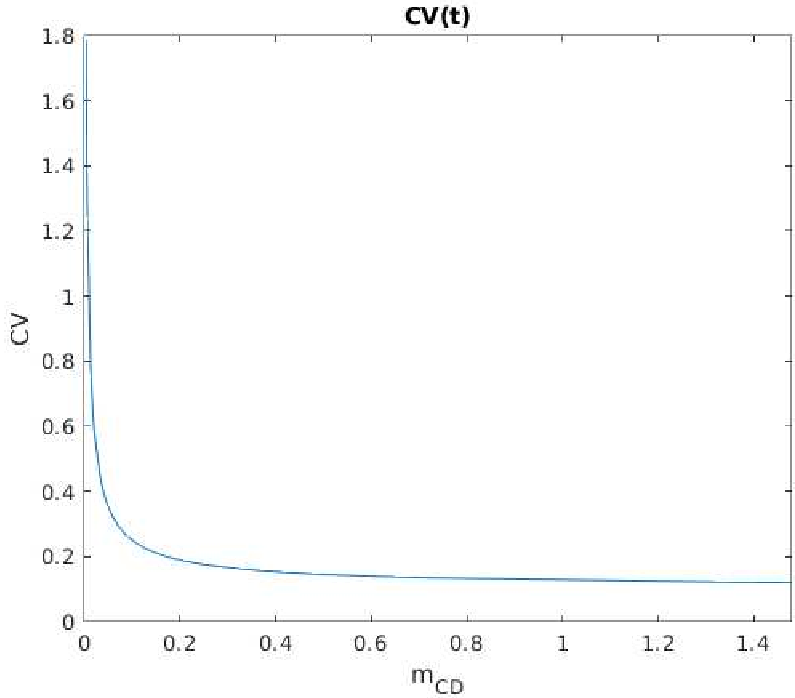

Figure 3.8: The variation coefficient calculated using the μ_{CD} values represented in Figure 3.6

3.4.4 The predictive model

With all required parameters extracted through preprocessing of the initial dataset, and incorporating the fixed parameter for the distribution of null water demand from the reference study, I was equipped to construct my own predictive model for the daily water demand pattern of a single meter (household). This was accomplished using Equation (3.3). The code excerpt below depicts the process to compute and plot the estimated daily demand curve for a sample meter.

```
Q = zeros(1,4*24);

U = Fo + (1-Fo).*rand(1,4*24);


A = sqrt(3)/pi;

for daily_interval = 1:1:4*24
    B = (1-U(daily_interval))/(U(daily_interval)-Fo(daily_interval));
    log_B_CV = log(B*CV(daily_interval));
    C = 1-A*log_B_CV;

    if  C >= 0
        Q(daily_interval) = mu_q(daily_interval)*mu_cd_real(daily_interval)*C;
    else
        Q(daily_interval) = 0;
    end
end
```

The predictive modeling enabled generation of estimated high-resolution daily water demand profiles at the individual meter level based on the statistical distributions derived from the raw meter data.

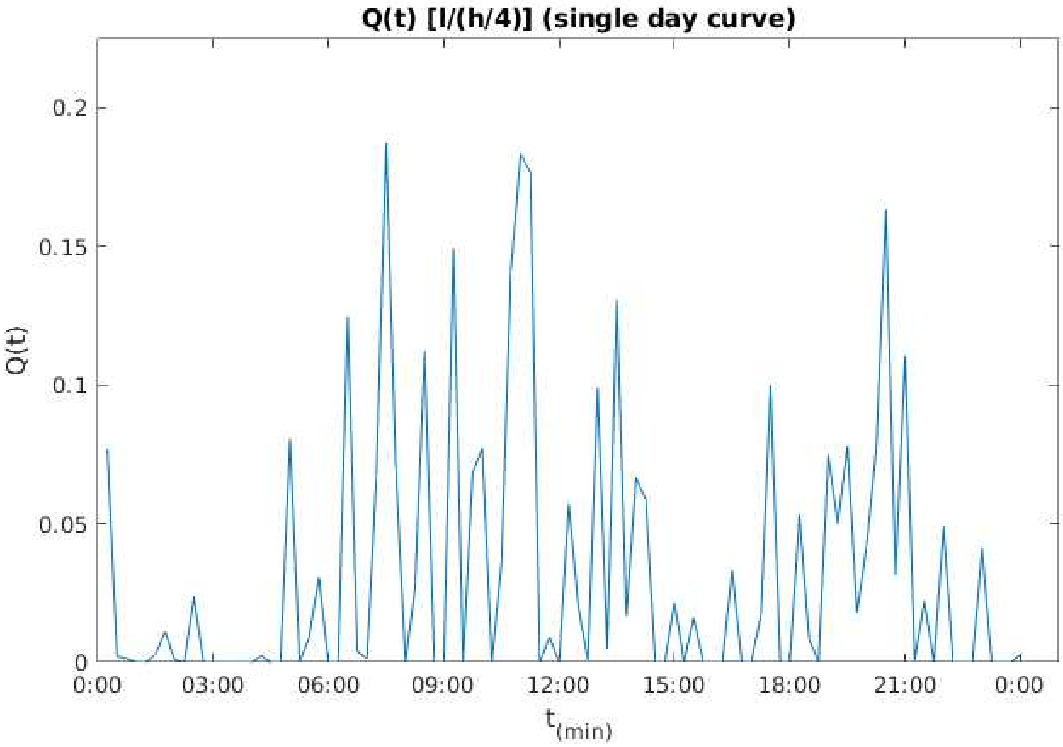

Figure 3.9: Predicted daily water demand pattern produced by the model for a sample meter, using the parameterized equation and data processing approach described.

Chapter 4

Experiments and Results Analysis

This chapter will focus on the analysis of the results obtained, the parameters estimated from the collected and refined data, and the final predictive model for water demand. Specifically, the chapter will evaluate the model's accuracy by comparing it with the model detailed in the referenced study[16].

4.1 Estimation of mean demand coefficient μ_{CD}(t)

While my estimated model appears visually similar, it shows generally lower values and more variation between time steps compared to the distribution in the reference study. Several factors may contribute to these differences. My μ_{CD} array was constructed by extracting four values per hour from randomly selected meters, since no meter had more than one reading per hour. However, the study utilized data with a much higher, minute-level sampling frequency, as noted in this quote:

> These two monitoring systems were realized specifically for studying the water demand, hence they gave reliable time series for the water request. In addition, CE and Fr monitor- ing systems, as the field laboratory of PSG, provide a fine description of the water demand during the day because the time step Δt equals 1 min. ([16])

The finer temporal resolution and localized scope of the study's data may have enabled a smoother and more consistent distribution. Additionally, my analysis used over five thousand meters across a broad geographic area, while the study data came from a smaller set of 596 meters in a single town. The larger and more diverse sample in my work may account for the greater variability observed. Nonetheless, the similar general shape provides evidence that my parameterized model effectively captures the underlying daily water demand pattern, providing validation of my modeling approach.

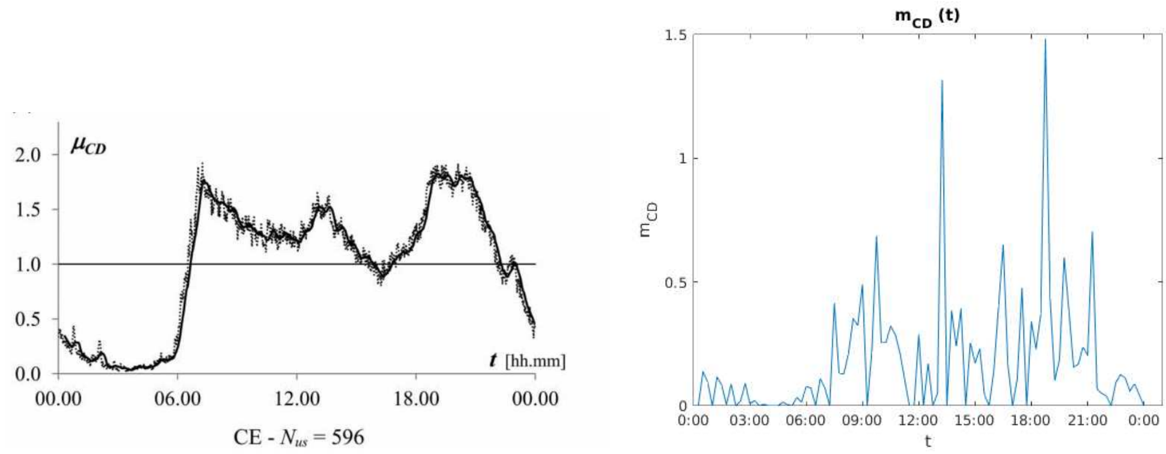

Figure 4.1: The estimated mean demand coefficient obtained by the reference study (left) compared to my own estimation (right)

4.2 Estimation of variation coefficient CV(t)

Similar to the μ_{CD} parameter, my estimated variation coefficient (CV) curve exhibits visual similarity to the reference study, approaching an asymptotic value around 0.1. However, my CV curve shows lower values overall. As noted in the reference study, CV decreases as the number of users (meters) increases. Therefore, the lower CV values in my model can be attributed to the much larger meter sample size.

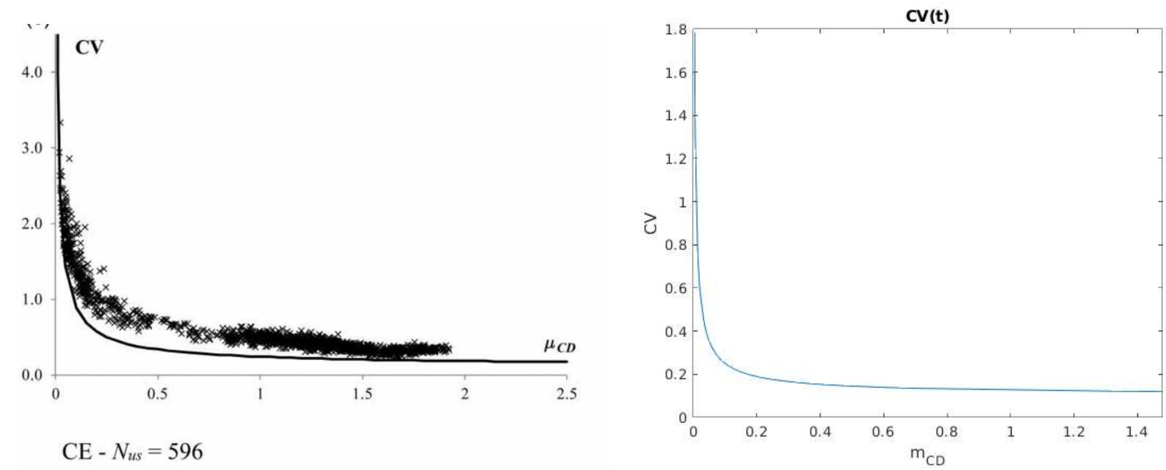

Figure 4.2: The estimated variation coefficient obtained by the reference study, plotted along with the discrete values of CV calculated on the source dataset (left) compared to my own estimation (right)

4.3 Single day water demand estimation

The visual similarity between my predicted daily water demand curve and the curve from the reference study further supports the validity of my data processing and modeling approach. Both curves exhibit near-zero estimated demand during the early morning hours from 0:00-5:00, followed by increasing peaks after 6:00 and again in the evening from 18:00-21:00. The finer one-minute sampling frequency of the reference study accounts for the different appearance of their curve compared to my sparser 15-minute estimates. Additionally, the greater variability between my adjacent time steps stems from the same factors discussed regarding the μ_{CD} parameter extraction. Namely, the way the μ_{CD}(t) values array was constructed, larger meter sample size and geographic distribution introduce variability not present in the original data. Nonetheless, the consistent daily demand profile demonstrates that my model successfully captures the underlying household behavior patterns evidenced in the study, despite discrepancies in data collection methodology.

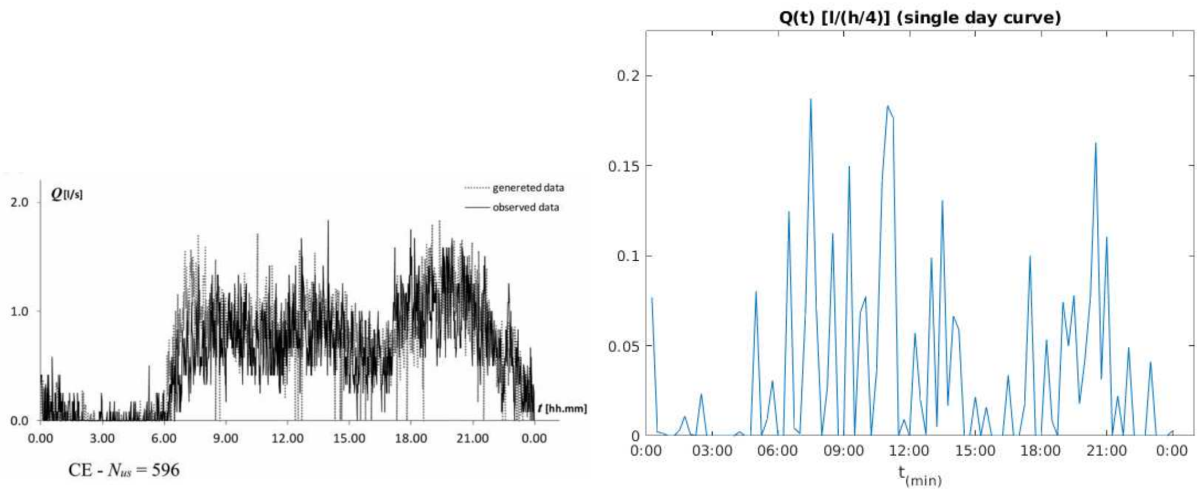

Figure 4.3: The estimated daily water demand curve generated by the reference study (left) compared to my own predictive model (right)

Additionally of note are the numerous zero or near-zero values present in my estimated daily water demand curve. A likely explanation stems from the 15-minute sampling frequency chosen for the prediction model. Each μ_{CD} value represents integrated demand over a 15-minute interval. Within a single household, it is probable that many 15-minute periods exhibit zero demand at such fine resolution. Decreasing the temporal resolution (i.e. increasing the sampling period) would reduce the likelihood of zero values as more non-zero demand events are captured in the longer measurement intervals.

4.4 Leak detection: Threshold and possible experimentation

With a robust model now developed to generate expected household water demand, the next challenge is transforming this model into a tool for water leak detection that can provide a preliminary determination of whether a specific household's water meter is detecting a leak.

I aim to approach the problem of leak detection in this manner because the generated model is readily adaptable to any large residential area and imposes no additional costs on distributors or metering companies. Unlike the model described in the quote from paragraph 4.1, which required installing sensors and equipment, the current model demands no new infrastructure or on-site operators.

The proposed solution for detecting a leak on a specific residential water meter involves generating predicted water demand data for the same number of meters in the area using the predictive model. The mean and standard deviation of the expected water demand value at each time step can then be calculated across all generated demand curves.

A threshold must be established whereby readings from the actual meter that persistently exceed that level would indicate a leak is likely occurring. In consultation with Inkwell Data, it was deemed reasonable that meter readings consistently exceeding 2-5 standard deviations above the mean curve of the generated data throughout the day suggest a leak with enough certainty to warrant an on-site inspection.

Figure 4.5 depicts the mean curve of the water demand generated for the same number of meters as in the final dataset. It also shows the threshold corresponding to a demand excess continuously equal to two standard deviations above the mean at each time step. This threshold provides the lowest acceptable level of accuracy. Also included is the figure from the reference study showing generated water demand for multiple meters concurrently, with the mean calculated at each time step. Despite differences in data collection and handling to determine distribution parameters, visual comparison shows the mean demand curve across multiple meters calculated using the current model is similar to that of the reference study. This similarity persists in spite of the variations in data acquisition and processing between the two approaches.

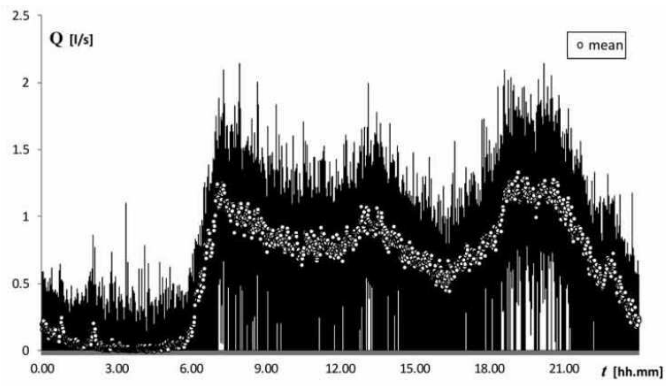

Figure 4.4: Simulated water demand for multiple meters with the mean demand at each 1-minute time step overlaid, as described in the reference paper[16]

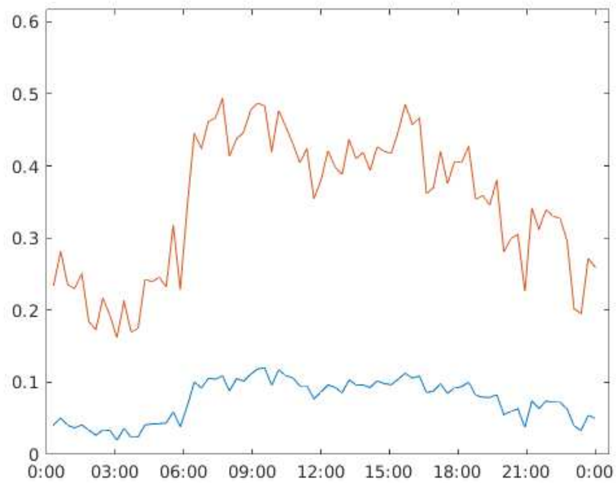

Figure 4.5: Mean water demand across generated daily water demand curves using my prediction model (blue) and minimum threshold for leak detection (orange)

4.5 Attempted testing and possible future work

In order to validate the results with empirical data, an experimental simulation of household water consumption was constructed using a simple laboratory apparatus consisting of a water tank, a circular pipe system to circulate water from the tank through a smart meter, and an electrovalve to regulate water flow, programmed to mimic household usage patterns. The experiment was designed to simulate daily water usage under two conditions: normal usage without any leaks, and normal usage with an induced leak. By collecting data sets for both conditions and comparing them to the daily water demand curves generated by the predictive model, it was hypothesized that the model could detect the presence of a leak based on the data from the deliberately disrupted system.

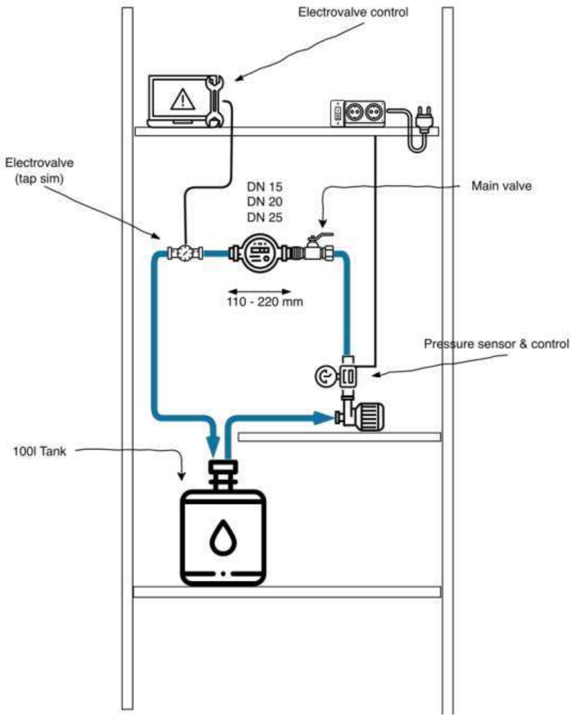

Figure 4.6: Schematic diagram of the laboratory setup

Upon implementation, the experimental setup proved inadequate for effectively simulating household water usage patterns. The electrovalve system did not allow for sufficient flexibility and precision in regulating water flow rates to accurately mimic the variable demands of actual domestic usage throughout a typical day (e.g. the difference between brief hand washing and extended showering events, or toilet flushing). As such, the apparatus could not adequately validate the simulation results against empirical data as intended. However, this preliminary experiment provides a foundation for future efforts to refine the laboratory simulation of household water consumption in order to empirically substantiate the model results obtained in this thesis. Further work should focus on enhancing the responsiveness and accuracy of the water flow control system to more realistically replicate the range of water usage events in a household setting over time.

Chapter 5

Conclusion

While current leak detection methods are reliable, they remain costly and, as evidenced by the persistently high water loss rates in Italy's pipe networks (exceeding 40%), are often implemented too late or infrequently to be maximally effective. This thesis sought to develop a lower-cost, proactive leak detection approach that would overcome the need for expensive, specialized equipment or time-consuming, localized manual inspection.

The predictive model presented here builds upon existing water demand forecasting methods, but utilizes real smart meter data, already collected regularly across Italy for billing purposes, to calculate the parameters for a mixed probability distribution of demand. This allows prediction solely from widely available, routinely gathered data, without requiring additional infrastructure expenditures.

The central aim was to demonstrate the greater utility of smart meter data for early identification of leaks across the water distribution network, beyond its current use for billing alone. Given Italy's high annual water loss, rising prices, and ranking as the top water consumer in Europe, improved leak detection enabled by comprehensive smart meter data analysis could significantly improve water efficiency and curb waste from pipe leaks nationwide.

Bibliography

[1] *https://www.euronews.com/green/2023/06/06/italy-wastes-enough-water-for-43-million-people-a-year-can-it-fix-its-leaky-pipes*.

[2] eureau.org, *Europe's Water in Figures, An overview of the European drinking water and waste water sectors, 2021 edition*.

[3] *https://single-market-economy.ec.europa.eu/single-market/european-standards/harmonised-standards/measuring-instruments-mid*$_e n$.

[4] *https://www.maddalena.it/en/product/sj-plus-evo/*.

[5] *https://www.maddalena.it/en/product/radio-evo/*.

[6] *https://www.maddalena.it/en/product/electo-sj/*.

[7] *https://www.cittametropolitana.mi.it/export/sites/default/ambiente/doc/Community/Archivio*$_l ifelon$ $ANIE_S UOM5.pdf$.

[8] *https://smg-anie.it/2018/02/21/anie-csi-position-paper-l-uso-delle-reti-a-169-mhz-per-smart-metering-gas/*.

[9] http://www.cen.eu/cen/Sectors/Sectors/Measurement/Documents/M441.pdf, *European Commission, Standardization mandate to CEN, CENELEC and ETSI in the field of measuring instruments for the development of an open architecture for utility meters involving communication protocols enabling interoperability*. 2009.

[10] https://oms-group.org/fileadmin/files/download4all/omsSpezifikationen/Generation5/OMS Spec$_V ol1_G eneral_v 241.pdf$, *OMS specification Volume 2 – Primary Communication Communication between meter or actuator and the gateway*.

[11] https://oms-group.org/fileadmin/files/download4all/omsSpezifikationen/Generation5/OMS Spec$_V ol1_A nnexA_F 241.pdf$, *Open Metering System Specification – Vol. 2 – Annex A*.

[12] https://webstore.iec.ch/searchformq=62056, *IEC 62056-1-0:2014 Smart metering standardisation framework and IEC 62056-6-1:2017 Object Identification System (OBIS)*.

[13] https://oms-group.org/en/open-metering-system/oms specification, *OMS specification Volume 1 – General Part Open Metering System Overview*.

[14] *https://www.webdyn.com/product/webdyneasy-w-m-bus/*.

[15] *https://bsonspec.org/*.

[16] G. d. G. Rudy Gargano Carla Tricarico and F. Granata, *A stochastic model for daily residential water demand*. WA Publishing, 2016.