Politecnico di Torino

Master's Degree course in Civil Engineering

2025-2026

Master's Degree Thesis

Pipe Replacement in Water Distribution Networks: The Influence of Network Topology and a Preliminary Analysis of Traffic Impacts

Supervisor

Prof. Riccardo VESIPA

Candidate

Atahan UZUNER

Abstract

As Water Distribution Networks (WDNs) age, they become increasingly susceptible to leakage, requiring comprehensive renovation strategies. While essential, rehabilitating these networks in dense urban areas faces strict environmental and operational constraints. Consequently, simultaneous network-wide renovation is often infeasible; instead, a sequential replacement of pipes is required to minimize disruption to city dynamics. However, this process inevitably disturbs urban functionality due to water supply interruptions and road network closures. Given the growing scarcity of global water resources, optimizing both water losses and delivery efficiency during this replacement sequence is critical.

To address this, a Probability-Based Incremental Learning (PBIL) algorithm is employed. This thesis evaluates the algorithm's applicability and robustness across networks of varying sizes and configurations. Furthermore, road network interruptions are analysed to quantify the combined impact of water supply disturbance and reduced urban accessibility.

Acknowledgements

I would like to express my sincere gratitude to my supervisor Riccardo Vesipa for the invaluable guidance, encouragement, and support provided throughout this thesis. His insight, patience, and constructive feedback have been fundamental in shaping this work and in deepening my understanding of the subject.

My deepest gratitude goes to my family for their unconditional love, constant support, and endless belief in me throughout all my academic journey. Their patience and encouragement have given me strength throughout this challenging process.

Finally, I would like to thank my loved ones for their understanding, motivation, and presence during this period. Their support has meant a great deal to me and has helped me stay focused and determined until the completion of this thesis.

Contents

1 Introduction

Water distribution networks (WDNs) are essential urban infrastructure systems that play a fundamental role in modern life. These pressurized systems deliver water to users through interconnected pipes and junctions, and their proper operation requires the provision of adequate discharge, pressure, and water quality. The hydraulic behaviour of WDNs is governed by the principles of mass continuity and energy conservation.

Because water distribution networks are large-scale and capital-intensive infrastructures, their design, operation, and maintenance must be carried out carefully and efficiently. In addition, since these systems are mostly located underground within urban areas, their management is often technically challenging and closely connected to economic and operational constraints. Therefore, alongside hydraulic performance requirements, cost-effectiveness must also be considered in the planning and management of water distribution networks.

Due to population growth and per capita water consumption , the pattern of the demand of users has changed significantly. Since many water distribution networks were not originally designed for such major changes in demand, they often require adaptation or rehabilitation. Moreover, as networks age and user demands increase, pipes progressively deteriorate, leading to leakages. Moreover, since the water distribution networks are generally located underground in the urban areas, especially in the busy roads and crowded areas, their maintenance and monitoring is becoming a difficult challenge.

As water distribution networks age, they become increasingly susceptible to leakages in different parts of the system. In addition, because water distribution networks are generally located underground in urban areas, especially beneath busy roads and densely populated zones, their maintenance and monitoring are particularly challenging. This pressure increase enhances the leakages in the water distribution networks. These leakages may occur in different components of the networks such as pipes, valves, joints. As the leakages increase, not only the system is damaged but also the water availability for the users in the networks decreases. Although this may not appear to be a major problem in areas with abundant water availability,

when considered at the global scale, it becomes a significant issue. These losses accumulate to a significant amount in the days when water scarcity is so crucial. In addition to that, only problem is not related to reduced delivery of water due to the leakages, but also it is related to economic losses due to water leakages. These economic losses include energy losses during pumping of water, energy losses during treatment of the water. In addition to those economic losses, water safety becomes an important issue due to leakages. Since water distribution networks are located below the ground surface, as the leakages increase, potable water becomes more susceptible to contamination. In addition to those risks, a water distribution network needs to be reliable by supplying water continuously without frequent disruptions. Water leakages in the water distribution networks are also causing operational inefficiencies. As the leakages in the networks increase, local pressure in the network decreases. This pressure decrease results in insufficient water delivery to users which results in user dissatisfaction. As problems related to reliability and operational efficiency coupled with economic burden of the water losses, systematic pipe replacement and rehabilitation becomes unavoidable for a long lasting and reliable water distribution network.

Although complete rehabilitation of a water distribution network may theoretically restore the hydraulic performance of the system, such an intervention is rarely feasible in practice due to economic, social, and operational constraints. Consequently, renovation activities are typically performed progressively, replacing a limited number of conduits at a time while the system remains operational.

In the literature, optimization in water distribution networks has been extensively studied. Most existing optimization studies focus on problems such as pipe sizing in the design phase, optimal pump scheduling to reduce energy consumption, pressure management strategies, district metered area (DMA) partitioning, and optimal pipe replacement planning based on cost–benefit analyses. In many of these approaches, the objective is to determine *which* pipes should be replaced and when, often minimizing total life-cycle costs or maximizing hydraulic reliability over a planning horizon.

In many rehabilitation planning studies, the focus is primarily placed on the identification of which pipes should be replaced and at what stage of the planning horizon. These approaches generally evaluate the final configuration of the network after the rehabilitation has been completed. As a consequence, the intermediate hydraulic conditions that occur during the renovation process itself are often simplified or neglected.

In real urban systems, renovation is implemented sequentially: pipes are disconnected, replaced, and reconnected one at a time, or in small groups. During this progressive replacement process, the hydraulic characteristics of the network evolve step by step. Each intervention modifies local parameters such as pipe roughness, diameter, and leakage coefficients, which in turn affect pressure distribution, flow patterns, and leakage magnitudes in the remaining parts of the network. As a result, the order in which pipes are replaced may influence the magnitude and distribution of pressure deficits, demand deficits and leakage reductions during the renovation period of the water distribution networks.

This aspect introduces a dynamic and strongly nonlinear problem. The hydraulic state of the network after each replacement depends not only on the characteristics of the newly installed conduit but also on the residual condition of the pipes that are not replaced yet. Therefore, two different sequences of replacement that lead to the same final network configuration may produce significantly different intermediate hydraulic performances and, consequently, different levels of service disruption for users.

Despite the practical relevance of this problem, limited attention has been devoted to the optimization of the sequence of pipe replacement from a purely hydraulic perspective. Most previous studies prioritize long-term cost minimization or reliability improvement, while the short and medium term hydraulic impacts experienced during the renovation phase remain less explored.

Sequential optimization explicitly addresses this gap. Instead of optimizing only the final set of pipes to be replaced, the problem is formulated as the search for an optimal ordering of interventions that minimizes the cumulative negative hydraulic effects experienced by users during the entire renovation process. These effects may include pressure deficits at junctions, demand reductions, or other service-level problems.

By accounting for the progressive evolution of the hydraulic state, sequential optimization provides a more realistic representation of actual renovation practices. It allows utilities to not only decide what to replace, but also in which order to perform the replacements so as to mitigate adverse impacts on service quality.

Within this framework, the central research question addressed in this study can be formulated as follows: Is there an optimal sequence of pipe replacements during the renovation of a water distribution network that minimizes the negative hydraulic impacts on users?

To answer this question, three main steps are required:
(i) simulate the hydraulic response of the network to a generic sequence of progressive pipe replacements;
(ii) define and quantify, through an appropriate scalar performance indicator, the cumulative negative effects experienced by users during the renovation process;
(iii) identify, through an optimization procedure, the sequence of replacements that minimizes these negative effects.

This approach adapts concepts traditionally applied in infrastructure scheduling and transportation engineering to the context of water distribution network renovation, with a specific focus on hydraulic performance during the transition phase rather than only on long-term post-renovation conditions.

In order to evaluate the hydraulic consequences of progressive pipe replacement, it is necessary to define a mathematical representation of the water distribution network and of the operations that occur during renovation.

Renovation is not a single action but a succession of individual interventions. Each intervention typically involves operations such as the isolation of an existing conduit, the temporary modification of boundary conditions, and the activation of the newly installed pipe. These actions may locally modify hydraulic characteristics such as resistance, effective topology, and leakage parameters. Since these modifications occur progressively, the hydraulic state of the network evolves step by step during the renovation time interval. During this time interval, the system experiences a sequence of intermediate hydraulic states, each corresponding to a specific configuration of replaced and non-replaced pipes. The objective of the problem is not the final configuration itself, but rather the trajectory of configurations that the network follows throughout the renovation period.

The hydraulic behaviour of the WDN is described through the time evolution of pressures at junctions and flows in pipes. These quantities are governed by mass continuity at nodes and energy conservation along links. The network is modelled as a set of junctions connected by pipes, where each pipe is characterized by its length, diameter, and roughness. The head loss along pipes is computed according to standard formulations for pressurized flow.

An important aspect in aging networks is the presence of distributed leakages. Leakages are assumed to depend on local pressure and on pipe-specific coefficients that are based on the

material condition and deterioration. When a pipe is replaced, its hydraulic resistance and leakage parameters are modified, typically resulting in reduced head loss and lower leakage rates. These local changes may influence pressures and flows in other parts of the network, especially in systems with limited redundancy.

User demand is considered pressure-dependent. Each node is characterized by a required demand and a minimum service pressure. When the available pressure is above the required threshold, user demand is fully satisfied. If pressure falls below this threshold, a demand deficit occurs. This deficit can be quantified as the difference between required and delivered discharge, and it represents one of the main indicators of service degradation during renovation.

Two characteristics make this optimization problem particularly challenging.

First, the decision variable is not a static design configuration but an ordered sequence of operations. The solution space therefore consists of permutations of replacement actions, which grows combinatorially with the number of pipes considered.

Second, the system response is nonlinear. The hydraulic state after each intervention depends on the updated resistance and leakage properties of replaced pipes, as well as on the residual condition of the remaining ones. Consequently, two different replacement sequences that lead to the same final rehabilitated network may produce different intermediate pressure distributions and demand deficits.

The objective of the optimization is therefore to identify the sequence of pipe replacements that minimizes the cumulative hydraulic disturbance during the time interval, subject to hydraulic constraints and service requirements.

While previous investigations on sequential pipe replacement have mainly focused on specific real-world case studies, limited attention has been devoted to understanding how the structural characteristics of a network influence the hydraulic impacts of renovation. Real networks embed multiple interacting features, such as topology, demand distribution, infrastructure age, and operational rules, which make it difficult to isolate the role of individual structural properties.

To overcome this limitation, the present study adopts a synthetic network framework. Artificial networks with controlled size, connectivity, and geometric characteristics are

generated to systematically explore how topological features affect the hydraulic response during sequential pipe replacement. This approach allows decoupling structural effects from site-specific characteristics and provides a more general understanding of the relationship between network topology and renovation performance.

In addition to hydraulic impacts, pipe replacement works in urban areas inevitably interact with transportation systems, as excavation sites may affect road accessibility and traffic conditions. Although the primary focus of this thesis is the hydraulic performance during sequential renovation, a preliminary traffic analysis is conducted to explore how road network disruptions can be quantified. A shortest-path framework is implemented and tested on a real urban road network to evaluate potential traffic deviations associated with pipe replacement works. Due to time constraints, a full integration between hydraulic optimization and traffic modelling is not developed in this study. Nevertheless, the presented framework outlines a possible extension toward multi-domain infrastructure planning, where hydraulic and traffic impacts could be jointly considered.

The objective of this thesis is to investigate how the structural characteristics of water distribution networks influence the hydraulic impacts of sequential pipe replacement and the performance of sequence optimization strategies. By employing synthetic network models and systematic simulation campaigns, the study aims to provide a generalized understanding of the relationship between network topology and renovation performance. Furthermore, a preliminary traffic analysis is conducted to explore how pipe replacement operations may interact with urban mobility systems, laying the groundwork for future multi-domain infrastructure planning.

2 Modelling Framework

2.1 Hydraulic Modelling of Water Distribution Networks

Water distribution networks (WDNs) are complex infrastructure systems designed to deliver water from supply sources to consumers through interconnected pipes, pumps, valves and junctions. Because these networks are deeply embedded in urban environments, their operation must ensure reliable water supply while maintaining adequate pressure levels across the system.

The hydraulic behaviour of WDNs is governed by the conservation of mass and energy within the network. These principles determine how flow rates and pressures evolve throughout the system and how the network responds to operational changes. In particular, changes in pipe characteristics, topology, or operational status may significantly influence the hydraulic conditions within the network.

Hydraulic modelling therefore plays a fundamental role in analysing the dynamic behaviour of WDNs and evaluating the impact of infrastructure interventions such as pipe replacement or network renovation.

2.2 Mathematical Representation of the Network

A water distribution network can be represented as a system composed of links and nodes.

Links include pipes, that transport water between nodes. Each pipe is characterised by its length, diameter, roughness and hydraulic resistance, which influence the flow and energy losses along the conduit.

Nodes represent elements such as junctions, reservoirs and tanks. Junctions correspond to locations where water demand occurs, while reservoirs and tanks act as boundary conditions that regulate the hydraulic head within the network.

At each junction, the continuity equation must be satisfied so that the incoming and outgoing flows balance the water demand and possible leakages. The hydraulic behaviour of pipes is described through energy equations that account for friction losses and minor losses

along the conduits. These equations are solved simultaneously across the entire network to determine pressures at nodes and flow rates in pipes.

2.3 Continuity Equation

At each junction of the network, the continuity equation must be satisfied to ensure conservation of mass.

$$\sum_{i \in IN(j)} Q_i - \sum_{i \in OUT(j)} Q_i = D_j$$

where

- Q_i= flow rate in pipe i
- $IN(j)$= set of pipes entering node j
- $OUT(j)$= set of pipes leaving node j
- D_j= water demand at node j

The total outflow from a junction can be expressed as

$$D_j = d_j + L_j$$

where

- d_j= user demand
- L_j= leakage occurring at the junction.

This formulation allows the modelling of both **customer demand and leakages** in the hydraulic balance of the network.

2.4 Energy Equation and Head Losses

Energy conservation along pipes is expressed through the head loss equation.

$$H_{US,i} - H_{DS,i} = \frac{Q_i \mid Q_i \mid}{A_i^2 \cdot 2g}\left(\sum_{k=1}^{M_i} K_k + \lambda_i \frac{L_i}{D_i}\right)$$

where

- $H_{US,i}$and $H_{DS,i}$= upstream and downstream hydraulic heads
- Q_i= flow rate in pipe i
- A_i= cross-sectional area of the pipe
- g= gravitational acceleration
- K_k= minor loss coefficients
- λ_i= friction factor
- L_i= pipe length
- D_i= pipe diameter.

These equations describe the hydraulic behaviour of pipes and allow the determination of flow distribution and pressure levels across the network.

2.5 Leakage Representation

Leakages occurring in aging water distribution networks represent an important component of the hydraulic balance and must therefore be incorporated into the modelling framework. In this study, leakages along pipes are represented using a **pressure-dependent formulation** based on an orifice flow relationship.

The leakage contribution associated with pipe $i$and aggregated at junction $j$is expressed as

$$Q_{L,i,j} = 0.5\, \xi_{L,i}\, L_i\, P_j^{0.5}$$

where

- $Q_{L,i,j}$= leakage flow contribution from pipe $i$at junction j

- $\xi_{L,i}$= leakage coefficient associated with pipe i
- L_i= length of pipe i
- P_j= pressure at junction j

The leakage formulation can be rewritten as

$$Q_{L,i,j} = \chi_{L,i}\, P_j^{0.5}$$

where

$$\chi_{L,i} = 0.5\, L_i\, \xi_{L,i}$$

represents the equivalent leakage coefficient of the pipe.

This formulation implies that leakages occurring along a pipe are aggregated at the junctions connected to that pipe and are modelled using an orifice-type relationship. The coefficient $\chi_{L,i}$depends on half of the pipe length and on the hydraulic and structural characteristics of the conduit, such as material, diameter, and aging condition.

This pressure-dependent representation allows the model to capture the interaction between pipe leakages and the hydraulic state of the network.

2.6 Modelling with the EPANET- MATLAB toolkit

In this study, to understand the behaviour of the synthetic networks, EPANET-MATLAB toolkit is utilized to construct, modify and simulate the hydraulic behaviour of the networks.

The EPANET–MATLAB Toolkit is an open-source software interface that is developed by the KIOS Research and Innovation Center of Excellence at the University of Cyprus. This open source software creates a link between the MATLAB and EPANET. The toolkit can be used to access EPANET libraries directly from MATLAB, and to automatically generate and modify networks, parameters. The toolkit also allows users to run hydraulic simulations and automatically extract the results from EPANET into MATLAB.

The use of this toolkit was necessary for this study because, it provides full automation of the synthetic networks for different sizes, and modification of the networks automatically from MATLAB.

In this thesis, the toolkit was used to:

- Create synthetic square-grid network topologies
- Add junctions, and pipes to the network
- Assign pipe lengths, diameters, and roughness coefficients
- Define nodal demands and demand patterns
- Run hydraulic simulations
- Extract computed pressures and velocities
- Integrate simulation outputs into the optimization framework

The optimization algorithm was fully integrated with EPANET MATLAB toolkit through MATLAB. For each solution :

1. Pipe diameters were updated in the EPANET model.
2. Hydraulic simulation was executed.
3. Pressures and velocities were retrieved.
4. The objective function (construction cost + hydraulic penalties) was evaluated.

This integration enabled automated evaluation of huge number of different network configurations.

2.7 Sequential Pipe Replacement Modelling

The sequential pipe replacement modelling framework adopted in this study is based on the approach proposed by Vesipa and Avarzamani (2022), and is here adapted to the synthetic-network simulation environment developed in MATLAB.

During the renovation of water distribution networks, pipes cannot be replaced simultaneously across the entire system. Instead, renovation works are typically performed

through a progressive sequence of pipe replacements, where individual conduits are disconnected, replaced and reconnected over time.

Each replacement operation modifies the hydraulic characteristics of the network and may influence pressures, flows and leakages in surrounding areas. For this reason, the pipe replacement process must be explicitly represented within the hydraulic modelling framework.

In this study, the replacement of a pipe is modelled through three main phases:

1. **Preparation phase**
   Preliminary works are performed while the existing pipe remains operational. During this phase, the pipe maintains its original hydraulic properties.

2. **Disconnection phase**
   The pipe is temporarily disconnected from the network by closing the corresponding valves. From a modelling perspective, the pipe status is set to closed in the EPANET model, effectively removing the link from the network topology. In addition, leakages associated with the disconnected pipe are set to zero.

3. **Reconnection phase with updated pipe**
   After the installation of the new pipe, the link is reintroduced into the network. The pipe is assigned updated hydraulic parameters, such as the new diameter and roughness, while leakages associated with the pipe remain equal to zero.

This modelling approach allows the hydraulic behaviour of the network to evolve dynamically during the renovation process.

The timing of renovation operations is controlled through operational parameters implemented in MATLAB. These parameters determine the start time, duration and day of pipe disconnection during the replacement process.

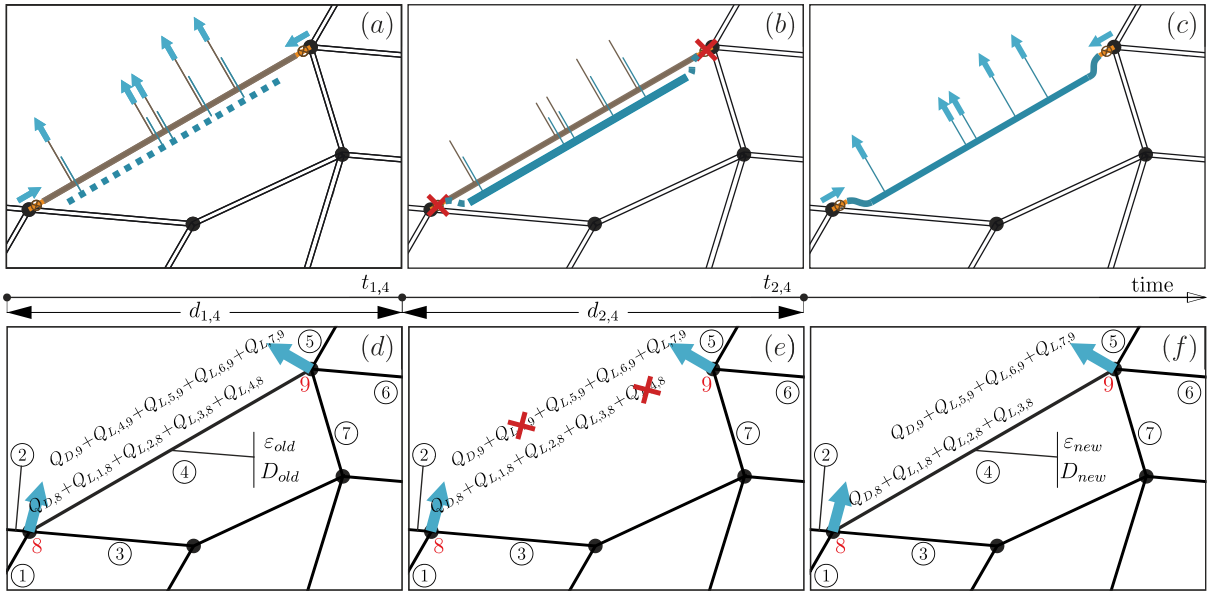

Figure 2.1 Phases of pipe replacement during network renovation and their representation in the hydraulic model. Source: Vesipa & Avarzamani (2022)

The phases illustrated in Figure 2.1 show how the physical pipe replacement process is translated into the hydraulic modelling framework. During the renovation process, the hydraulic state of the network evolves dynamically as pipes are progressively disconnected and replaced. These operations modify both the network topology and the hydraulic parameters of the system, which in turn influence pressures, flows, and leakages across the network. By explicitly modelling these transitions within the hydraulic simulation, it becomes possible to evaluate the impact of different pipe replacement sequences on the overall performance of the water distribution network.

The modelling framework described above provides the basis for evaluating and comparing alternative pipe replacement sequences, which are optimized in the following section using the PBIL algorithm.

3 Algorithms

3.1 Hydraulic Simulation Framework

The hydraulic simulation algorithm used in this work follows the general renovation-simulation logic presented by Vesipa and Avarzamani (2022), while being implemented here through the EPANET-MATLAB Toolkit and extended to synthetic water distribution networks.

Figure 2.2 illustrates the algorithm adopted to simulate the sequence of pipe replacement operations within the water distribution network.

In this study, the hydraulic simulations are performed using the EPANET MATLAB Toolkit. Differently from traditional approaches where the network is directly defined within the EPANET environment, the network topology is generated and controlled within MATLAB. The network topology is generated programmatically within MATLAB using the EPANET-MATLAB Toolkit. Nodes, pipes, lengths, diameters and connectivity relationships are defined directly in MATLAB, after which the hydraulic simulation is executed using the EPANET solver. The modelling framework adopted in this work relies on synthetic water distribution networks. These networks are generated within MATLAB by defining the number of nodes, pipe lengths, diameters, and connectivity relationships. The use of synthetic networks allows the structural characteristics of the system to be controlled while avoiding the complexity associated with real-world networks.

The algorithm simulates the hydraulic behaviour of the network while progressively modifying its topology as pipes are replaced.

Network initialization

At the beginning of the simulation, the synthetic network is generated within MATLAB. The network structure is defined through parameters such as:

- number of nodes
- pipe lengths
- connectivity relationships between nodes.

Once the network topology and hydraulic properties are defined, the corresponding EPANET input structure is automatically created using the EPANET MATLAB Toolkit.

The list of pipes to be replaced is then defined as a set R, representing the sequence of links that will undergo renovation.

Operational parameters for renovation works

Instead of using a detailed calendar-based schedule, the timing of renovation operations is defined through a set of operational parameters implemented in MATLAB. These parameters describe the duration of works and the timing of disconnections during pipe replacement.

In particular, the following parameters are used:

- Duration of the pipe disconnection during replacement (hours)
- Time of day when the disconnection begins
- Day of the renovation work when the pipe is disconnected
- Duration of the replacement works on the pipe.

These parameters allow the operational constraints of the renovation process to be reproduced without explicitly defining the calendar time of each individual operation.

Hydraulic simulation cycle

Once the network and operational parameters are defined, the hydraulic behaviour of the system is simulated over time.

At each simulation step, the EPANET solver computes the hydraulic state of the network. The main variables stored during the simulation include:

- pressure at junctions $P_j(t)$
- supplied demand at junctions $Q_j(t)$

When a pipe is scheduled to be disconnected according to the operational parameters, the status of the corresponding link is set to closed in EPANET. During this phase, the leakages associated with that pipe are set to zero since the pipe is temporarily removed from the network.

Once the replacement operation is completed, the pipe is reconnected to the network by setting its status to open. At this stage, the hydraulic characteristics of the pipe are updated by assigning the new diameter and roughness values associated with the renovated infrastructure. As the pipe is newly installed, the leakages associated with it remain equal to zero.

This procedure allows the hydraulic simulations to reproduce the progressive modification of the network as renovation works advance.

Simulation outputs

The simulation produces several time series describing the hydraulic performance of the system during the renovation period. In particular, the following variables are obtained:

- pressure at junctions $P_j(t)$
- supplied demand at junctions $Q_j(t)$
- leakage flows $Q_{L,j}(t)$.

These variables are used to evaluate performance indicators related to conditions experienced by users. One relevant indicator is the demand deficit, defined as

$$\Delta Q_j(t) = Q_{DU,j} - Q_{D,j}(t)$$

where $Q_{DU,j}$represents the demand requested by users and $Q_{D,j}(t)$is the demand actually delivered by the network.

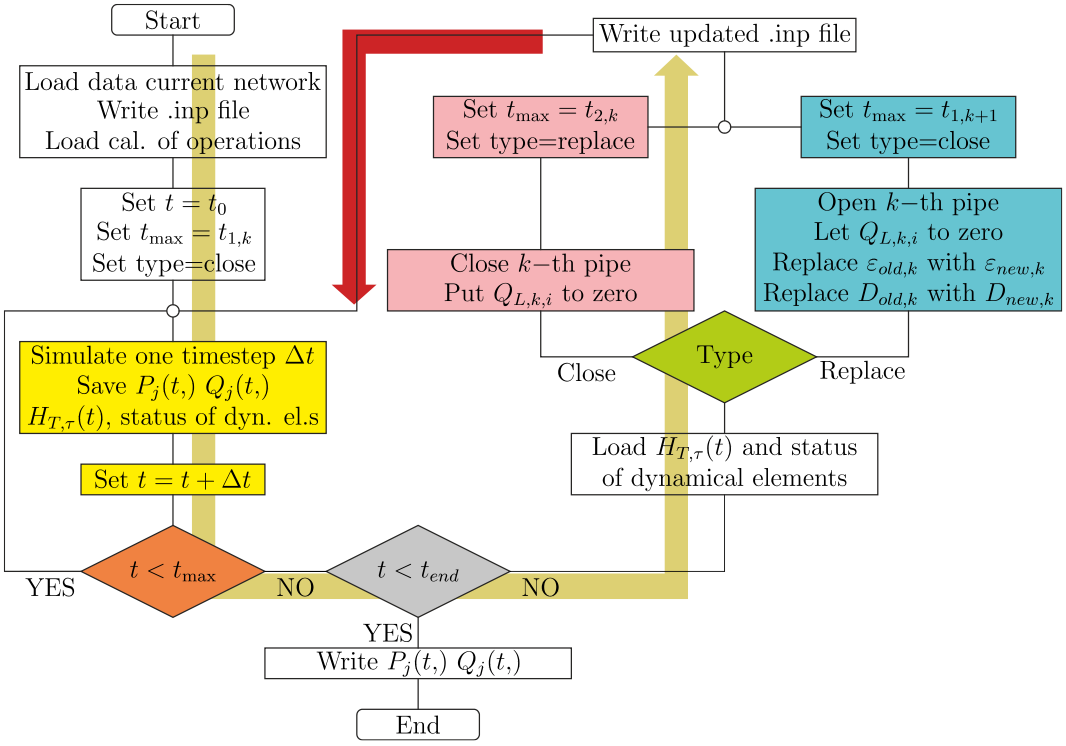

Figure 3.2 Algorithm adopted to simulate the sequence of pipe replacement operations in water distribution networks. Source: Vesipa & Avarzamani (2022)

The modelling framework described above provides the hydraulic simulation environment used to evaluate the performance of different renovation strategies. In this study, the sequence of pipe replacement operations is not predefined but is determined through an optimization procedure.

In particular, the Population-Based Incremental Learning (PBIL) algorithm is adopted

3.2 PBIL Algorithm for optimal sequence

For the optimization of the pipe replacement order, Population Based Incremental Learning (PBIL) algorithm is used. The PBIL algorithm is a stochastic guided search method that derives its directional guidance from the best solutions found from the previous iterations. Therefore, this optimization algorithm is utilized to find the best sequence for the pipe replacement and thus, to find the minimum cost for this procedure.

PBIL Algorithm Workflow

For the identification of the links, all links in the network are stored in the Vector *I* and the links which will be replaced are stored in the vector *R* with the number of change denoted as N_R . The number of phases for the link replacement is denoted as N_m, for this case N_m=N_R.

The steps of the pipe replacement sequence optimization are explained below.

Creation of the Probability Matrix

At the beginning, the probability matrix P^s of size N_R x N_R is created with the same probability of P^s equal to $1/N_R$. In this probability matrix, each element $P^s(k,m)$ represents the probability of whether the pipe k will be replaced during the m^{th} phase, where s is the current iteration number of the optimization algorithm.

Generation of the Sequence Vector

Each pipe $k*$ that is present in the network, has a phase *m* during which the pipe $k*$ will be replaced. This phase is determined by generating random sample values from the probability distributions $P^s(k,m)$ that are described above. For example, $J = [i_3,i_5,i_2,....,i_{NR}]$, where i_m represents the index of the pipe that will be replaced during phase m.

Generation of the Population

In this step, a population consisting of N_{rdm}sequence vectors J_ε(for $\varepsilon = 1: N_{rdm}$) is generated. These sequence vectors are created following the procedure described in the step above. Consequently, the overall population vector can be expressed as: J=[J_1,J_2,…,J_{Nrdm}].

Reconfiguration of the Sequence Vectors

The elements that are in the sequence vector $J_\varepsilon$must be unique in each sequence vector of each population. Therefore, in order to check this elements i_m are randomly selected from the vector R, and the algorithm checks whether repetitive elements are presentand changes

them with randomly selected elements. This procedure is repeated for every sequence vector of the population until each sequence vector includes each elements that are in the matrix *R*.

Cost estimation of the Sequence Vectors

For each sequence vector of the population, cost of the network is calculated. Among those sequence vectors, the one which corresponds to the lowest cost is denoted as *J**.

For this study, the cost is divided into 3 different scenarios.

Scenario 1 evaluates the performance from a service-oriented perspective, where the cost is associated with demand reduction and is expressed in liters.

Scenario 3 evaluates performance from a rule-oriented perspective, where the cost is related to pressure deficit.

Scenario 5 evaluates performance from an economic perspective, accounting for the cost of demand reduction associated with leakages.

Probability Matrix Update

At iteration $s + 1$, each row (indexed by k) of the probability matrix is updated as follows. For each element of *J*(k)* where $k = 1{:}N_R$, the algorithm checks whether the condition $J^*(k) = i_m$ is satisfied. This condition justifies that, the phase during which, in the best sequence identified so far, the k^{th} pipe is replaced – specifically, that it occurs in the m^{th} phase. If this condition holds true, the probability of that event is increased in the probability matrix as follows; $P^{(s+1)}(k,m) = (1 - LR) \times P^s(k,m) + LR$, where LR is the learning rate which is taken as 0.5 . When the condition is not satisfied, the probability of that event is reduced in the probability matrix as $P^{(s+1)}(k,m) = (1 - LR) \times P^s(k,m)$.

Probability Matrix Mutation

During the optimization process in order not to get stuck in local optimal solutions, mutations are used. For each row k of the probability matrix $P^{(s+1)}$, an element is selected randomly by the mutation probability f. After the direction of mutation is randomly determined as d_f= (1 or -1), mutations are applied on row k as follows.

For the location (k,m),

$P^{(s+1)}(k,m) = [P^{(s+1)}(k,m) \text{ x } (1 - (d_f \text{ x } \Delta f))] + (d_f \text{ x } \Delta f)$;

For locations other than (k,m) in row k ;

$P^{(s+1)}(k,m) = [P^{(s+1)}(k,m) \text{ x } (1 - (d_f \text{ x } \Delta f))]$;

Where Δf is the mutation shift, in this case it is taken as 0.1 .

For the conditions where $P^{(s+1)}(k,m) > 1$, the condition is set to $P^{(s+1)}(k,m) = 1$, and for the conditions $P^{(s+1)}(k,m) < 0$, the condition is set to $P^{(s+1)}(k,m) = 0$.

Algorithm Stop Criteria

When the number of iterations s has reached to a pre-set number of iterations NI (in this case it is taken as 100), the algorithm comes to an end and stops; otherwise, algorithm starts another iteration from the step of Generation of the Sequence Vector.

4 Creation of Synthetic Water Distribution Networks

4.1 Rationale for Using Synthetic Networks

In order to systematically analyse the behaviour of water distribution networks under renovation strategies, synthetic network models were adopted in this study. The use of synthetic networks provides several advantages compared to real networks.

First, synthetic networks allow full control over the structural characteristics of the system. Parameters such as network size, connectivity and hydraulic requirements can be modified easily, enabling the exploration of a wide range of configurations. Second, synthetic networks can include a large number of components, which allows the testing of the optimization algorithm under different levels of network complexity. Finally, synthetic networks facilitate the comparison of algorithmic performance across networks of different sizes and structural characteristics.

For these reasons, synthetic networks provide a controlled benchmark environment for analysing the interaction between network topology, hydraulic conditions and renovation strategies.

4.2 Network Topology Generation

The topology of the water distribution networks was generated using a square grid configuration. A regular square grid provides a simplified representation of urban distribution networks and allows the systematic investigation of the influence of size and connectivity parameters on the performance of the optimization algorithm.

The generated network consists of N_s junctions arranged along each side of the grid. The connectivity between adjacent junctions is controlled through a connection probability parameter P_c. The parameter P_c represents the probability that a junction is connected to its neighbouring junctions within the grid. By changing this parameter, networks with different levels of connectivity can be generated.

This probabilistic approach enables the creation of networks ranging from sparsely connected structures to highly interconnected grids, allowing the investigation of how network connectivity influences hydraulic performance and renovation strategies.

4.3 Pipe Length and Nodal Demand

The hydraulic loading of the network was controlled by assigning predetermined values for pipe lengths and nodal demands. For simplicity and consistency, the length of all pipes within a network configuration was assumed to be constant.

Two representative pipe lengths were selected in order to simulate different urban conditions:

- $L = 30\text{m}$, representing densely populated urban areas with short pipe lengths.
- $L = 90\text{m}$, representing less densely populated areas with longer pipe lengths.

Similarly, nodal demands were assigned predetermined values to represent different levels of water demands within the network. Two demand levels were considered:

- $q = 1$, representing low demand conditions
- $q = 3$, representing high demand conditions

To analyse the combined effects of network density and demand levels, two representative combinations were defined:

Combination 1:

- $q = 1$
- $L = 30\text{m}$

This configuration represents densely populated urban areas with relatively low water demand per node.

Combination 2:

- $q = 3$
- $L = 90\text{m}$

This configuration represents less densely populated areas with higher water demand per junction.

The main input parameters used to generate the synthetic networks and define the subsequent hydraulic and optimization analyses are summarized in Algorithm Box 4.1. A simplified MATLAB-style parameter definition is also reported in Appendix A.1.

Algorithm Box 4.1. Definition of input parameters

- Define topology parameters: connection probability P_c, network size N_s, and pipe length L
- Define hydraulic parameters: nodal demand q, pipe roughness ϵ_{old}, reservoir head H_r, minimum head H_{min}, and residual leakage level
- Define degradation parameters: percentage of degraded pipes, major leakage level, and insufficient-supply condition
- Define PBIL parameters: selected district, scenario number, population size, and maximum number of pipe changes per phase

Related MATLAB script: Appendix A.1 – Parameter initialization

4.4 Assignment of Pipe Diameters

Pipe diameters were determined through an optimization-based design procedure aimed at minimizing the overall construction cost while satisfying hydraulic constraints.

Diameters were selected from a commercial pipe catalogue. The cost of each pipe depends on its diameter and is expressed as a cost per meter. The total construction cost of the network is therefore computed as

$$Cost = \sum L_i \cdot C(D_i)$$

where

L_i= length of pipe i

$C(D_i)$= cost per meter associated with diameter D_i

However, the design must also satisfy hydraulic constraints related to pressure and velocity limits. Therefore, the objective function includes penalization terms associated with hydraulic violations.

Pressure constraint penalization

Pressure deficits at junctions are penalized based on the difference between the minimum required pressure and the pressure obtained from the hydraulic simulation.

$$\Delta H_i = \max(0, H_{min} - H_i)$$

The total pressure penalization is expressed as

$$P_{total} = \sum \alpha (\Delta H_i)^2 + \beta I(\Delta H_i > 0)$$

where

H_i= pressure at junction i

H_{min}= minimum required pressure

$I(\cdot)$= indicator function

$\alpha$and β= penalization coefficients.

Velocity constraint penalization

Velocity constraints were also considered to ensure acceptable hydraulic performance. Pipe velocities were required to satisfy

$$0.5 < V_i < 1.5$$

Velocities exceeding the upper bound were penalized using a linear function, whereas velocities below the lower bound were penalized using a logarithmic function.

4.5 Centrality-Based Diameter Prioritization

To support the preliminary diameter assignment, pipes were ranked using a tailored edge betweenness centrality routine implemented in MATLAB. In graph-theoretic terms, betweenness centrality measures how often an edge lies on the shortest paths connecting different parts of the network. Therefore, pipes with higher betweenness are interpreted as structurally more important links for the transfer of flow through the system.

The adopted routine is based on the interpretation proposed by Giustolisi et al. (2019), namely that in water distribution networks the analysis should focus on pipes (edges) rather than only on nodes, since pipes are the physical components governing conveyance and hydraulic relevance. In the implemented function, the network adjacency matrix is first extracted from the EPANET model and transformed into a symmetric graph representation. Betweenness centrality is then computed on the edge set, and the resulting pipe scores are normalized to obtain a relative ranking of link importance.

In addition, the implemented routine includes some WDN-oriented features. Pipe weights can be introduced through a term proportional to L/D^5, so that hydraulic resistance effects can be reflected in the centrality calculation when required. Moreover, fictitious nodes are added at reservoirs and tanks in order to better represent the role of sources and storage elements within the network structure. In this way, the centrality measure is not a purely generic graph metric, but a tailored indicator adapted to the characteristics of water distribution networks.

The resulting relative betweenness values were used here as a ranking criterion for the initial diameter allocation. Pipes with higher centrality were given priority for larger diameters, while less central pipes were assigned smaller diameters unless hydraulic constraints required otherwise. Thus, centrality was used as a structural guide for preliminary design, while the final feasibility of the network was still verified through hydraulic simulation and optimization.

4.6 Optimization Procedure for Diameter Assignment

Pipe diameters were assigned through an optimization-based procedure implemented in MATLAB. The aim of this step was to identify a feasible and cost-effective set of commercial diameters for all pipes in the synthetic network before the renovation analysis. The procedure

takes as input the network topology, the hydraulic requirements, and the economic data associated with the available commercial diameters.

First, the set of admissible diameters and their corresponding unit costs were defined from the commercial pipe catalogue. Additional hydraulic parameters, such as reservoir head, minimum pressure requirement, pipe roughness, and nodal demands, were then assigned to build the initial hydraulic description of the network. These data were combined with the connectivity and geometric information already obtained from the topology-generation step.

A population of candidate diameter configurations was then generated randomly. Each candidate solution represents one possible assignment of commercial diameters to all pipes in the network. For each configuration, the hydraulic performance and the total cost were evaluated through the objective function described previously. This produced an initial memory of feasible and non-feasible solutions, which constituted the starting point for the optimization stage.

To make the diameter allocation more structurally consistent, the candidate diameter vectors were not assigned randomly to pipes. Instead, the relative ranking of pipes obtained from the tailored betweenness centrality routine was used to reorder the diameter sets. In this way, larger diameters were preferentially associated with more central pipes, while smaller diameters were assigned to less central ones. Therefore, the diameter assignment was guided not only by random search, but also by the structural importance of the pipes within the network.

Starting from this initial population, the final diameter configuration was refined using the Harmony Search Algorithm (HSA). At each iteration, new candidate solutions were generated through memory consideration and pitch adjustment rules, and then evaluated by the hydraulic cost function. If a newly generated solution improved the current harmony memory, it replaced the worst existing configuration. Through this iterative process, the algorithm progressively identified diameter combinations with lower total cost and better hydraulic feasibility.

At the end of the procedure, the best diameter vector was selected and assigned to the network. The resulting pipe diameters were stored in the topology structure and exported to an EPANET input file, which was then used in the following stages of hydraulic simulation.

The main steps of the integrated diameter-design procedure are summarized in Algorithm Box 4.2. This procedure combines preliminary diameter assignment, centrality-based pipe prioritization, and Harmony Search optimization within a single computational framework. A simplified MATLAB-style skeleton of the procedure is reported in Appendix A.2.

Algorithm Box 4.2. Centrality-informed diameter assignment and optimization

- Read the commercial diameter set, cost data, and hydraulic parameters
- Generate an initial harmony memory of candidate diameter configurations
- Evaluate the initial candidates using the hydraulic cost function
- Compute tailored betweenness centrality and rank pipes according to structural importance
- Reorder candidate diameter sets according to the centrality ranking
- Generate updated candidates using Harmony Search rules
- Replace inferior harmonies and track the best solution
- Assign the final diameter configuration to the network

Related MATLAB-style skeleton: Appendix A.2

4.7 Representation of Insufficient Supply Conditions

To analyse the behaviour of degraded networks, insufficient water delivery conditions were also simulated. Two insufficient demand scenarios were considered:

- $q_{insufficient} = 0.2$, representing partial water delivery
- $q_{insufficient} = 0$, representing complete supply interruption

These conditions were simulated by replacing selected pipes with smaller diameter pipes, thereby reducing the hydraulic capacity of the network.

4.8 Leakage Assignment

Leakages were incorporated into the synthetic networks in order to represent the hydraulic effects of aging infrastructure.

Two types of leakages were considered.

Residual leakages

Residual leakages represent small and spatially distributed water losses occurring throughout the network. In the modelling framework adopted here, these losses were simulated by assigning emitter coefficients to the junctions of the network. The emitter values were not assumed to be identical for all nodes; instead, they were scaled according to the **junction type**, so that nodes connected to a larger number of pipes were assigned higher leakage multipliers. In this way, the residual leakage distribution reflects the local connectivity of the network rather than being uniformly imposed.

The base residual emitter coefficient was determined through an iterative calibration procedure. Starting from an initial guess, the emitter value was adjusted using a numerical solver until the total leakage fraction produced by the network matched the target residual leakage level prescribed in the hydraulic parameters. In all simulations, this target value was set equal to approximately **5% of total water losses**. Therefore, residual leakage was represented as a distributed background loss calibrated at network scale and allocated locally according to junction connectivity.

Large leakages

Large leakages were introduced to represent more severe and localized degradation conditions associated with selected pipes. In the adopted formulation, these leakages were assigned through the junctions connected to the degraded pipes, rather than being distributed over the whole network. For each selected degraded pipe, its end nodes were identified, and additional emitter contributions were imposed at those nodes. If a node was connected to more than one degraded pipe, the added leakage contribution increased accordingly. This allowed the leakage intensity to reflect the local concentration of deterioration around damaged links.

The large-leakage formulation was superimposed on the residual leakage pattern already assigned to the network. Thus, the original distributed background leakage was preserved, while an additional localized leakage component was introduced around the degraded pipes.

As in the residual-leakage case, the magnitude of this additional emitter contribution was not fixed a priori, but calibrated iteratively until the total leakage level matched the target degradation parameter. In the analysed cases, this target large-leakage level was set equal to 40%.

This distinction between residual and large leakages makes it possible to represent both the diffuse background losses typical of aging systems and the localized hydraulic stress associated with more severe pipe deterioration.

Network creation is summarized in Figure 4.1 .

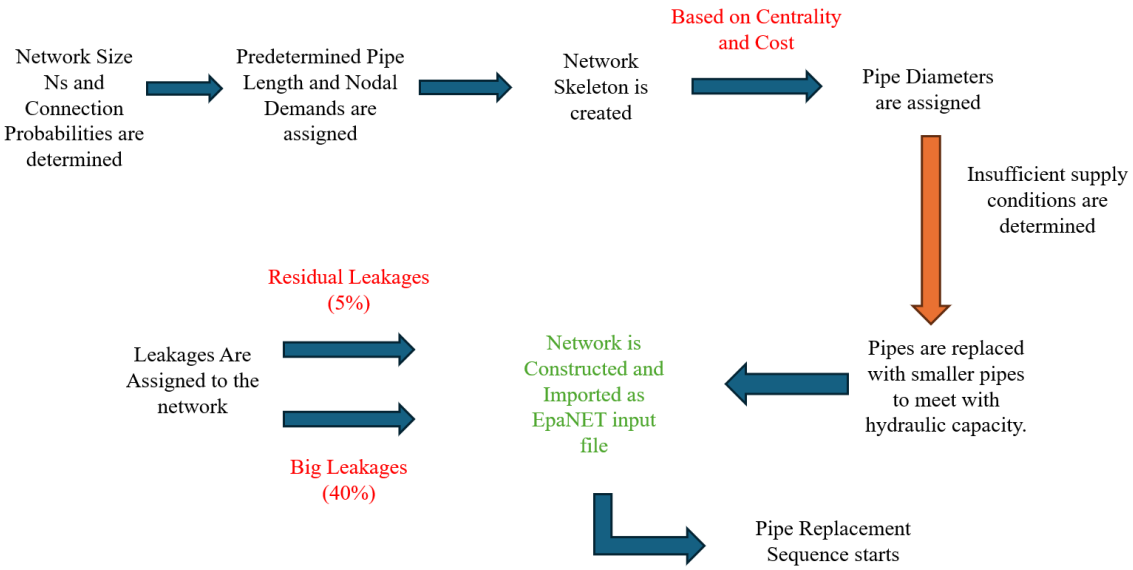

Figure 4.1, Network Creation Scheme

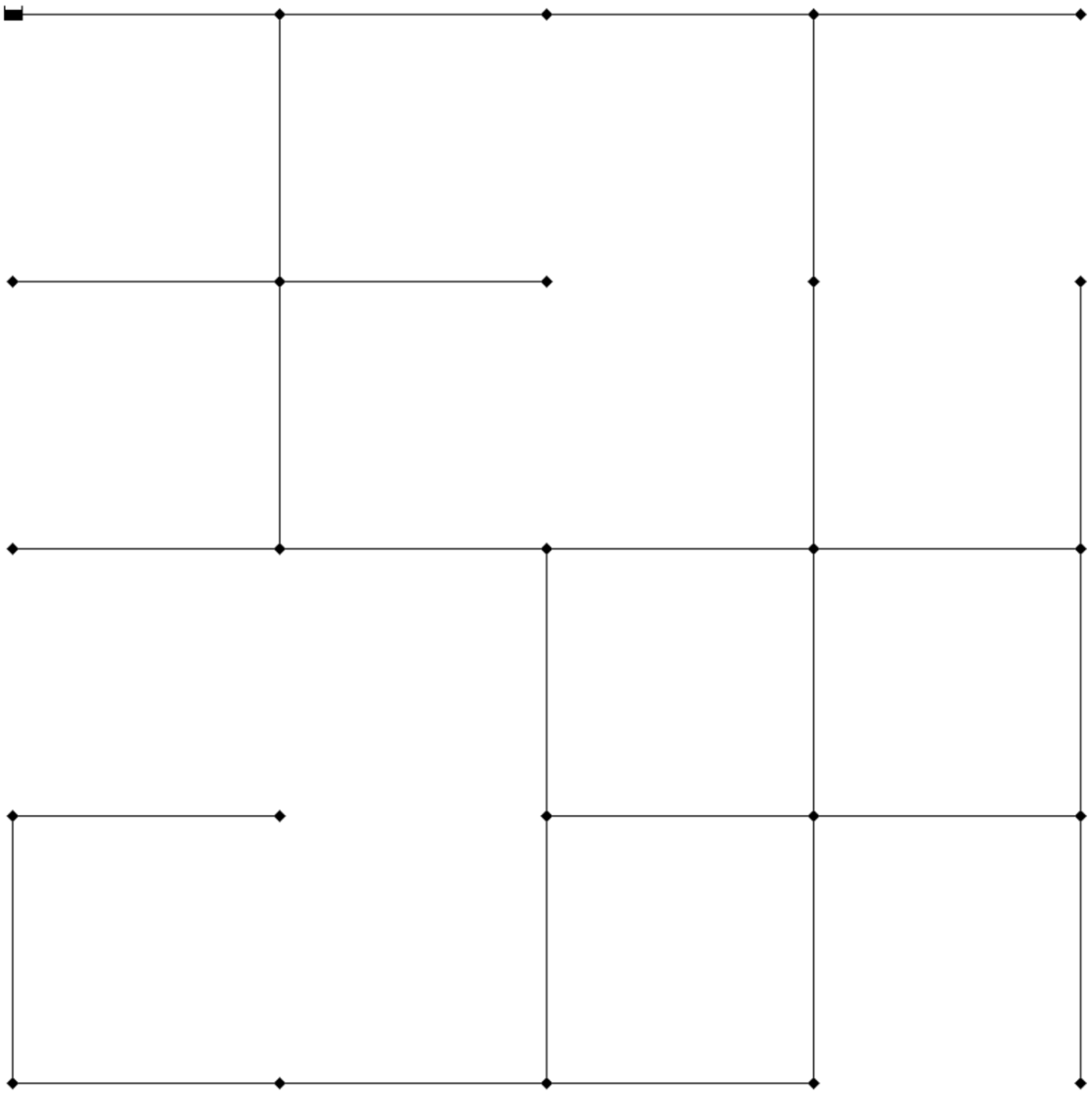

Figure 4.2 Example Network configuration in EPANET

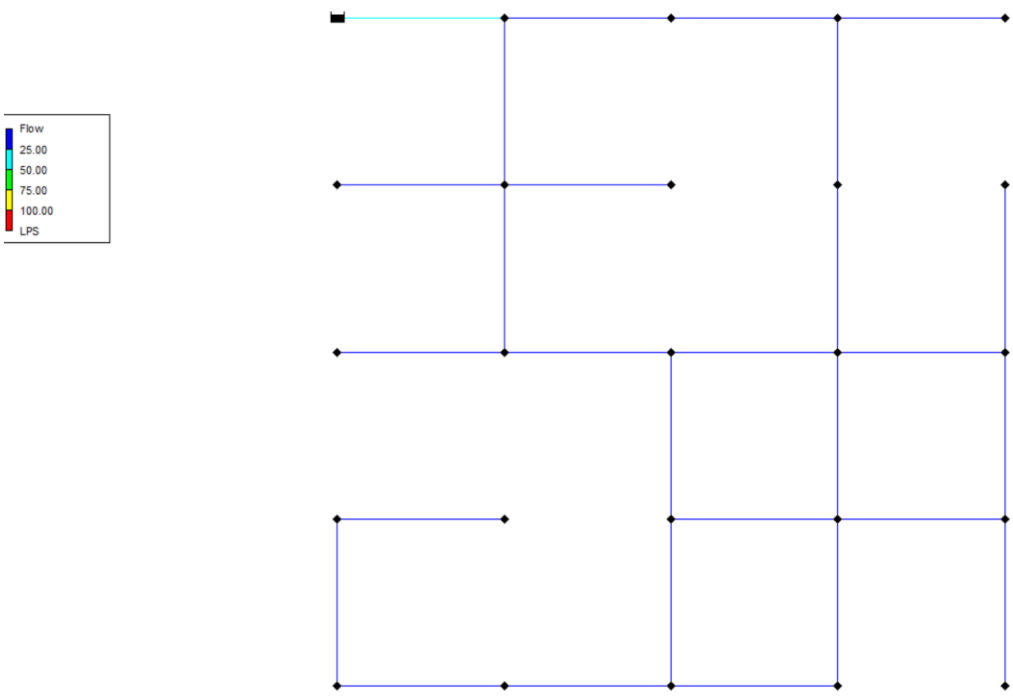

Figure 4.3 Example flow in Network configuration in EPANET

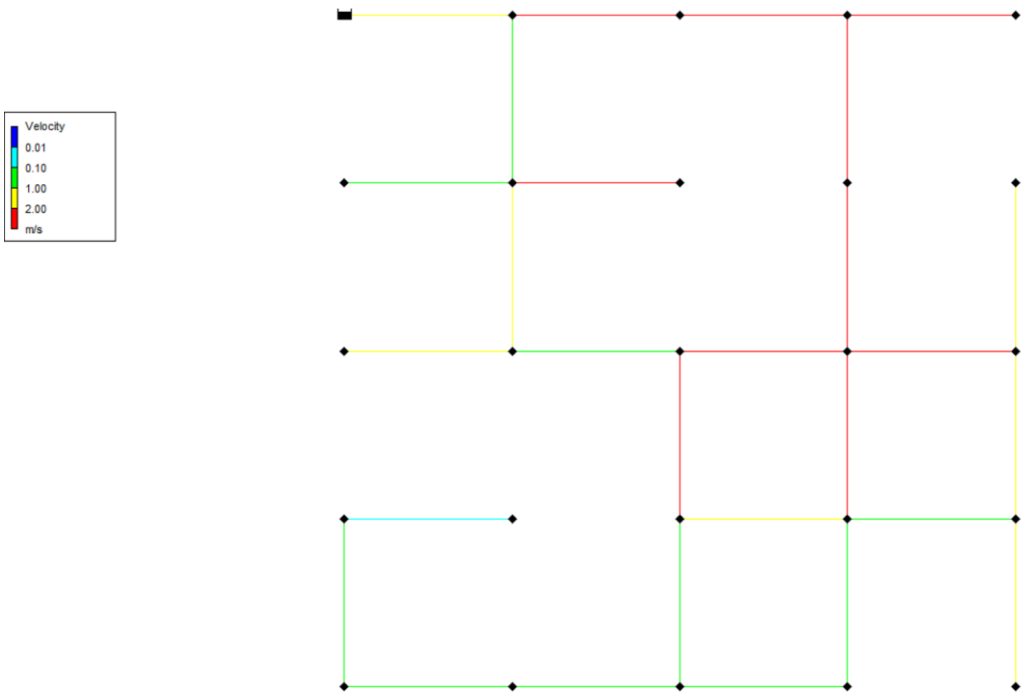

Figure 4.4 Example velocity representation in Network configuration in EPANET

5 Analysis of Network Behaviour

5.1 Convergence Behaviour

Before the evaluation of final optimization performance, the convergence behaviour of the algorithm in different network configurations are evaluated.

It should be noted that, the convergence behaviour of different costs are evaluated in different scenarios.

For each optimization run, the convergence history was obtained from the recorded cost evolution throughout all the iterations. The convergence iteration was then defined as the first iteration at which the cost became sufficiently close to the best cost achieved at the end of the run.

More specifically, let C $_{\text{final}}$ denote the minimum cost reached during the optimization process. A convergence threshold of 1% was adopted, and the convergence iteration was identified as the first iteration satisfying

$$C_i \leq C_{final}(1 + 0.01)$$

where $C_i$is the cost at iteration i.

This definition provides a practical measure of convergence speed, since it identifies when the algorithm first reaches a solution sufficiently close to the final optimal cost.

To compare convergence patterns across simulations with different iteration lengths and cost ranges, the cost histories were normalized. For each run, the convergence curve was scaled between 0 and 1 according to the minimum and maximum cost values observed during that run. The normalized curves were then interpolated onto a common iteration axis in order to allow direct comparison.

The convergence behaviour was analysed at three different specific conditions:

- Scenario-based analysis, to evaluate how the optimization dynamics vary across different objective cost function formulations
- Network-size-based analysis, to investigate the influence of $N_s$on convergence speed

- Connection-probability-based analysis, to assess the effect of $P_c$on the optimization dynamics

For each of these categories, average normalized convergence curves were computed by grouping all simulations belonging to the same class and taking the mean of the normalized convergence histories.

In addition to the curve-based analysis, convergence speed was also evaluated using the distribution of convergence iterations. These distributions were compared through boxplots for different scenarios, network sizes, and connection probabilities.

This combined approach made it possible to analyse the variability of convergence speed across different network configurations.

5.2 Performance Indicators of the Optimization

In order to evaluate the effectiveness of the optimization process across different network configurations, several performance indicators were defined. The purpose of this analysis was to quantify how effectively the optimization algorithm reduced the overall network cost during the optimization process.

To compare performance across different simulations, indicators were computed based on the evolution of the objective function between the initial and final stages of the optimization. Since the absolute cost magnitude may vary significantly across networks of different sizes and scenarios, ratio-based indicators were introduced in order to reduce the bias associated with scale effects.

- **Minimum reduction ratio (R_{min})**
- **Mean reduction ratio (R_{mean})**
- **Absolute minimum reduction** $(CostReduction)$

These indicators evaluate the improvement achieved by the optimization algorithm between the initial and final populations.

Minimum Reduction Ratio

The minimum reduction ratio $R_{min}$is defined as the ratio between the minimum value obtained by the last population and minimum value associated with the first population of the optimization process.

Since each population consists of 25 iterations, the first population corresponds to the first 25 iterations of the algorithm, while the last population corresponds to the final 25 iterations.

The ratio is therefore expressed as

$$R_{min} = \frac{Cost\ (\text{Last min})}{\text{Cost}\ (\text{First min})}$$

where

Cost(First min)= minimum cost obtained within the first population
Cost(Last min)= minimum cost obtained within the final population.

Mean Reduction Ratio

Similarly, the mean reduction ratio R_{mean}compares the average cost value between the last and first populations.

$$R_{mean} = \frac{Cost\ (\text{Last mean})}{\text{Cost}\ (\text{First mean})}$$

where

Cost (First mean) = mean cost of the first population
Cost (Last mean) = mean cost of the final population.

These ratios provide a normalized measure of the improvement achieved by the optimization algorithm.

For the comparison, lower ratios across the networks represent a better optimization for the algorithm. On the contrary, higher reduction ratios represent a limited optimization.

Absolute Minimum Reduction

In addition to the ratio-based indicators, the absolute minimum reduction (CostReduction) was also considered. This indicator represents the direct difference between the minimum cost value of the first population and the minimum cost value of the last population:

$$\text{CostReduction} = \text{Cost}(\text{First min}) - \text{Cost}(\text{Last min})$$

Thus, this measures the absolute improvement achieved by the optimization process in terms of the best solution found.

Performance indicators are indicated in the following figure.

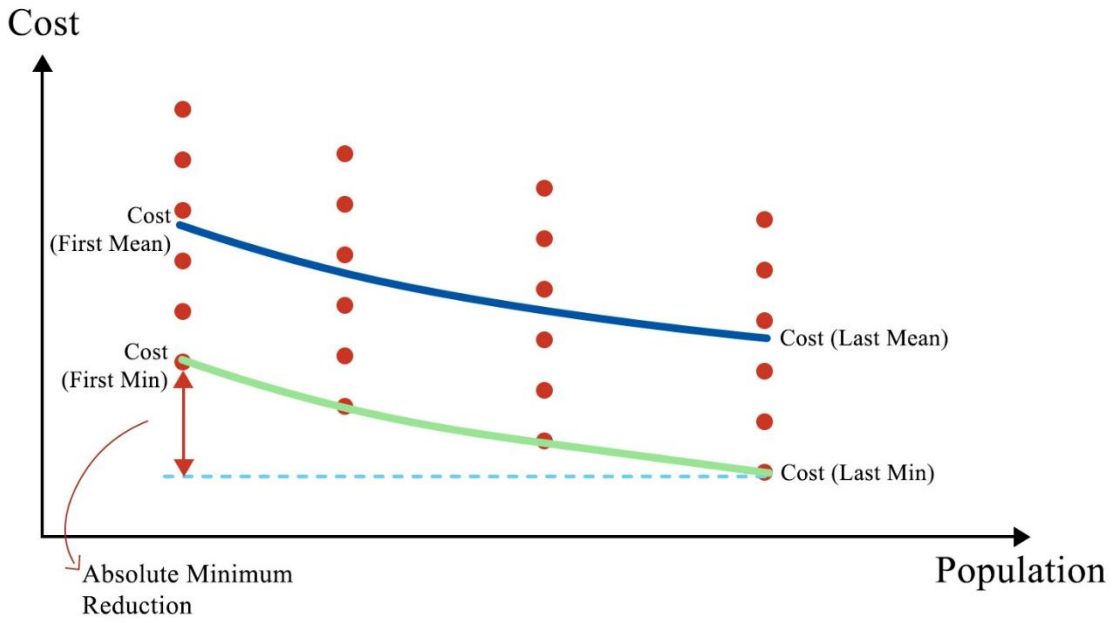

Figure 5.2 Performance indicators explanation

Overall, the ratio-based indicators provide a normalized measure of optimization improvement, whereas the absolute minimum reduction provides a direct measure of the magnitude of improvement. For the ratio-based indicators, lower values indicate a more effective optimization process, while values closer to 1 indicate limited improvement. For the absolute reduction indicator, higher values indicate a greater improvement between the initial and final stages of the optimization.

It should be noted that the cost values used in these indicators were not always associated with monetary units, since different objective-function scenarios were considered. The meaning of cost under each scenario is described in the Results chapter. In addition, for some comparative analyses and visualizations, logarithmic transformations of these indicators were used in order to improve interpretability across simulations with different magnitudes.

5.3 Heatmap-Based Trend Analysis

It was hard to understand the behaviour just based on the reduction ratio numbers. Therefore, to investigate the optimization across the large number of simulated configurations, heatmap visualizations were used. Heatmaps provide a convenient way to analyse the influence of key parameters on optimization performance.

The heatmaps were constructed by plotting the network size parameter $N_s$against the connection probability P_c. Each cell of the heatmap corresponds to a specific network configuration characterized by a specific combination of parameters.

Colour gradients were used to represent the magnitude of the reduction ratios. In these visualizations:

- lighter colours indicate lower ratios and stronger optimization performance
- darker colours indicate higher ratios and weaker optimization performance.

This representation allows the identification of regions within the large amount of simulations, where the optimization algorithm performs particularly well or poorly.

The analysis was conducted for networks ranging from 5x5 to 20x20 grid sizes, allowing the investigation of how increasing network size influences optimization performance.

Furthermore, separate heatmaps were generated for different network scenarios, including variations in pipe lengths and demand levels. In particular, comparisons were performed between the two different pipe length configurations defined in the synthetic network generation process:

- L= 30 m representing dense urban conditions
- L = 90 m representing less dense network structures.

In addition, the influence of insufficient demand conditions was analysed by comparing scenarios with:

- $q_{insuff} = 0.2$
- $q_{insuff} = 0$

With the help of visual comparisons provided by the heat maps, it was possible to identify the regions where the simulations performed better than the others.

5.4 Cluster-Based Behaviour Analysis

Although the heatmap visualizations allow the identification of a visual trend, further analysis was required to understand the behavioural patterns underlying the large number of simulations performed with different specific conditions. For this purpose, a cluster analysis was conducted to group simulations exhibiting similar performance characteristics.

Cluster analysis is an unsupervised learning technique that groups observations according to their similarity within a multidimensional feature space. In this study, the clustering procedure was applied to group network configurations into families exhibiting similar optimization behaviour.

The clustering analysis was performed using the following variables:

$$log_{10}(CostReduction),\ R_{min},\ R_{mean}$$

A logarithmic transformation was applied to the cost reduction values in order to stabilize magnitude differences between simulations and improve clustering performance.

Before clustering, all variables were normalized using **z-score standardization**, allowing the comparison of variables with different scales.

The clustering procedure was performed using the k-means algorithm with

$$k = 3$$

clusters. This choice allowed the identification of three distinct behavioural regimes within the simulation dataset.

5.4.1 Behavioural Regimes

The clustering analysis revealed three distinct performance regimes across the analysed network configurations.

Cluster 1 – High Efficiency Networks

The first cluster represents network configurations that shows strong cost reduction during the optimization process.

Cluster 2 – Intermediate Performance

The second cluster represents configurations exhibiting moderate optimization performance. These networks show some degree of improvement during the optimization process but do not reach the high efficiency observed in Cluster 1.

Cluster 3 – Saturation Behaviour

The third cluster corresponds to configurations in which the reduction performance of the algorithm was poor.

This classification helps reveal hidden structural patterns within the dataset and provides insight into how network parameters influence optimization behaviour.

6 Results

6.1 Overview of the Simulation Campaign

The optimization performance was evaluated across a wide parameters, including network size (N_s), connection probability (P_c), pipe length scenarios ($L = 30$ m and $L = 90$m), insufficient demand conditions ($q_{insuff} = 0$ and $q_{insuff} = 0.2$), and different objective scenarios. In total, 144 network configurations were simulated. Performance was assessed through reduction-based indicators and analysed using convergence plots, heatmaps and cluster-based visualizations. This chapter first examines convergence behaviour, then reduction performance, and finally the determination of behavioural regimes across wide range of simulations.

6.2 Performance Indicators and Visualization Framework

6.2.1 Reduction Ratios

As described in Section 5, the optimization performance is evaluated using the reduction ratios $R_{min}$and R_{mean}, which compare the cost levels reached at the end of the optimization process with those observed at the beginning. These indicators provide a normalized measure of how effectively the algorithm reduces the cost for different types of topologies. Lower values indicate stronger reduction performance, whereas higher values indicate limited improvement.

6.2.2 Heatmap Structure

Heatmaps were used to represent the global distribution of optimization performance over the parameter space. In these visualizations, the x-axis corresponds to connection probability (P_c), the y-axis corresponds to network size (N_s), and the colour scale represents the reduction ratio. Lower values correspond to stronger cost reduction, while higher values indicate weaker optimization performance.

6.3 Convergence Behaviour

The convergence analysis was conducted to evaluate how rapidly and how consistently the optimization algorithm approaches its best solution across different network configurations. The objective of this section is to compare how convergence behaviour changes with scenario, network size and connection probability.

6.3.1 Typical Convergence Pattern

Across all simulation groups, the optimization process exhibits a common convergence pattern. The early populations show high dispersion in the cost values, indicating a large scale exploration of the search space. As the optimization process continues, the cost curves shift downward, reflecting increasingly lower-cost solutions. In the later stages of the search, the curves flatten and form a convergence basin, suggesting that the algorithm approaches a stable solution region.

6.3.2 Convergence Dynamics

The average normalized convergence curves reveal meaningful differences across the analysed categories.

When grouped by scenario, Scenario 1 shows a slightly faster convergence behaviour than Scenarios 3 and 5, as its average curve drops more rapidly during the early stages of the optimization.

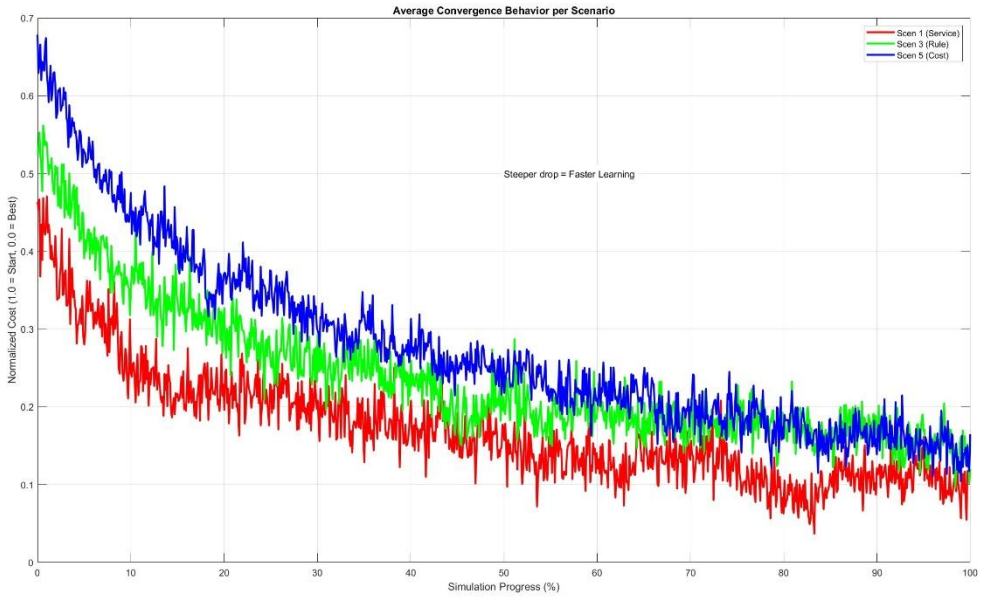

Figure 6.1 Convergence behaviour of different Scenarios

When grouped by network size, as can be seen from figure 6.2, smaller networks converge faster than larger sized networks. In particular, the 5x5 networks show the steepest initial reduction, while larger networks exhibit slower convergence and require more iterations to approach the final solution region.

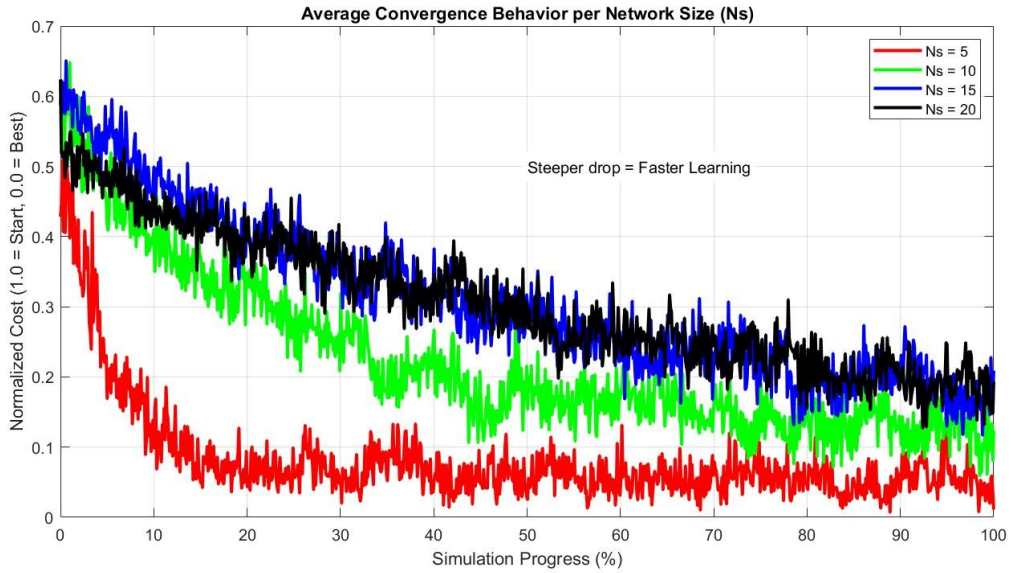

Figure 6.2 Convergence behaviour of different Network Size

When grouped by connection probability, no strong differences in the average convergence curves are observed. The curves remain relatively close across all P_c values, suggesting that connection probability does not really affect the convergence speed of the cost function.

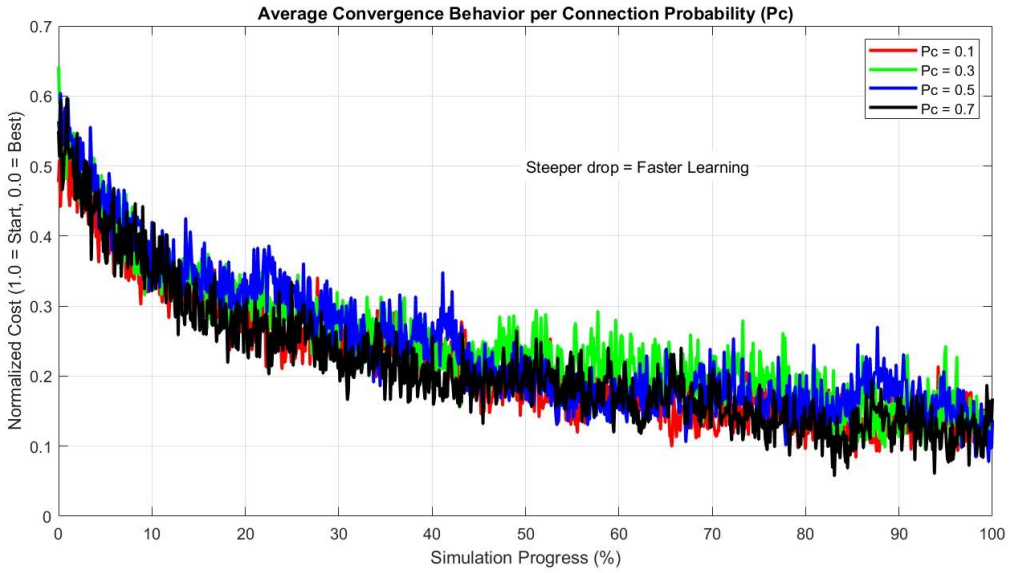

Figure 6.3 Convergence behaviour of different Connection Probabilities

6.3.3 Convergence Speed Analysis

To understand thoroughly the average convergence curves, convergence speed was also analysed through boxplots of convergence iteration. Convergence was defined using a 1% threshold from the final best solution.

Scenario-based convergence speed

The convergence speed grouped by scenario is shown in Figure 6.4. The boxplots indicate that the optimization behaviour differs across the three different cost scenarios. Scenario 1 generally exhibits lower convergence iteration values and a narrower spread compared to the other scenarios, suggesting that the algorithm tends to approach near-optimal solutions more rapidly under this formulation. Whereas scenario 3 and scenario 5 shows a slightly slower convergence patterns. These results are consistent with the normalized convergence curves represented in the Figure 6.1 .

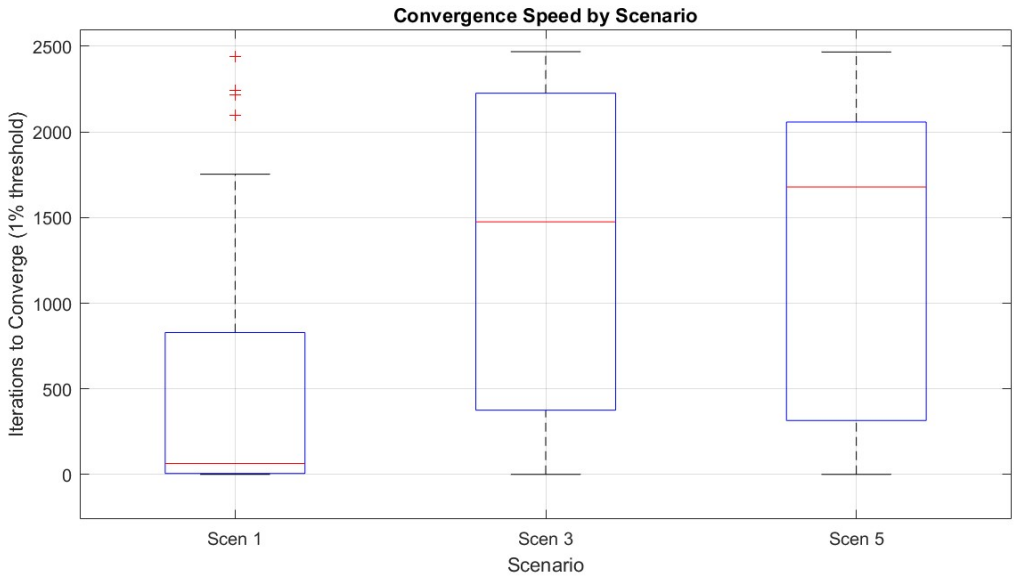

Figure 6.4 Convergence speed for different scenarios

Network-size-based convergence speed

The influence of network size on convergence speed is presented in Figure 6.5. A clear trend can be observed: smaller networks tend to converge more rapidly, while larger networks require more iterations to reach the convergence threshold. In particular, the boxplots show that the smaller network sizes are associated with smaller median convergence iterations and a more compact distribution, whereas larger networks show both higher average iterations and wider spreads.

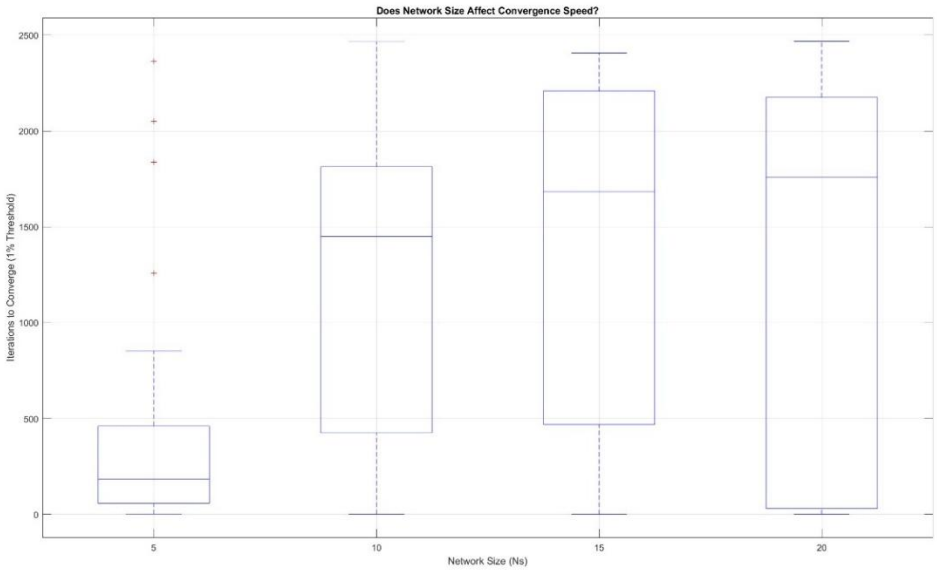

Figure 6.5 Convergence speed for different Network Sizes

Similar to the scenario case, the trend on the network size is also correlated with the average normalized convergence curves shown in Figure 6.2, where the curves for smaller networks exhibit a steeper early fall. This behaviour is expected, since larger networks generate a broader and more complex search space, making it more difficult for the optimization algorithm to identify favourable solutions quickly.

Connection-probability-based convergence speed

The convergence speed grouped by connection probability is reported in **Figure 6.6**. Compared to scenario type and network size, the influence of P_c on convergence speed appears to be too less significant. The boxplots show similar iterations to converge on the different connection probability levels.

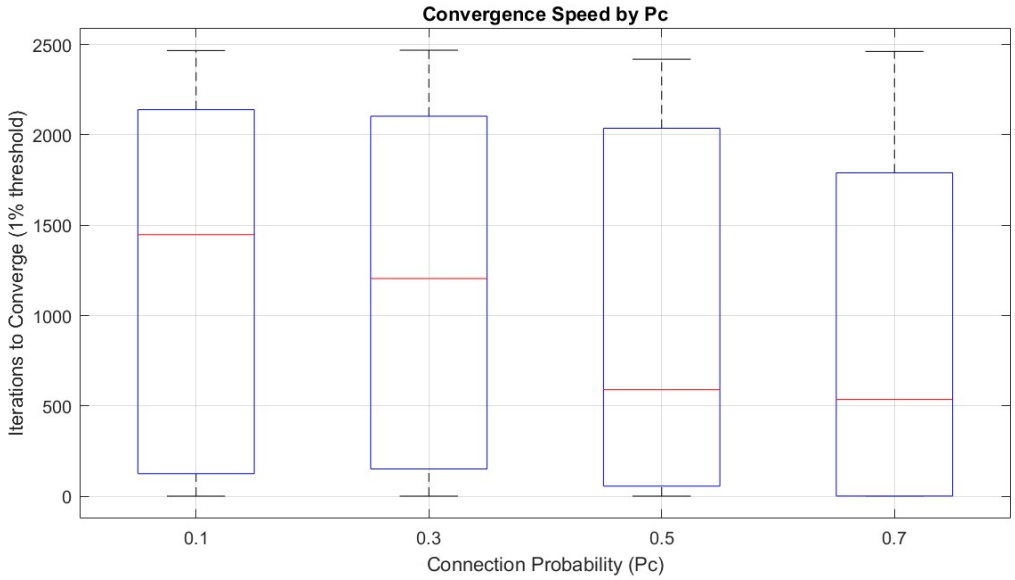

Figure 6.6 Convergence speed for connection probabilities

A similar observation can be made from the average normalized convergence curves in Figure 6.3, where the progression of the optimization between P_c-based groups are similar. This suggests that, within the explored parameter range, connection probability has a weaker effect on the convergence dynamics of the algorithm than the scenario type or the network size.

Overall, the convergence speed analysis indicates that the optimization process is primarily influenced by the type of objective function and the size of the network, whereas the role of connection probability is more limited. Scenario 1 and smaller networks tend to have faster convergence rates, while larger networks increase the number of iterations required to approach the final optimum.

This result is important because it shows that convergence behaviour is not controlled only by the optimization algorithm itself, but also by the structural and functional characteristics of the network under analysis.

6.4 Global Performance Reduction Analysis

6.4.1 Scenario 1 – Service-Oriented Objective

Scenario 1 evaluates the performance from a service-oriented perspective, where the cost is associated with demand reduction and is expressed in liters. The heatmaps show that the reduction performance depends on both network size and connection probability, with more favourable regions emerging in moderately sized and less connected networks.

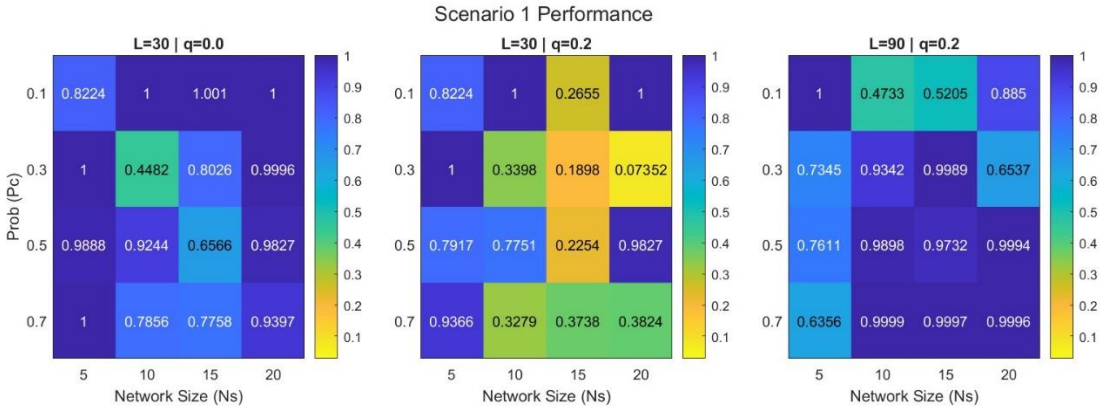

Figure 6.7 Heatmaps for Scenario 1, for R_{min}. Different panels refer to different configurations (length of pipes and demand)

6.4.2 Scenario 3 – Rule-Oriented Objective

Scenario 3 evaluates performance from a rule-oriented perspective, where the cost is related to pressure deficit. Compared to the other scenarios, this objective generally achieves stronger reduction performance over the explored parameter space.

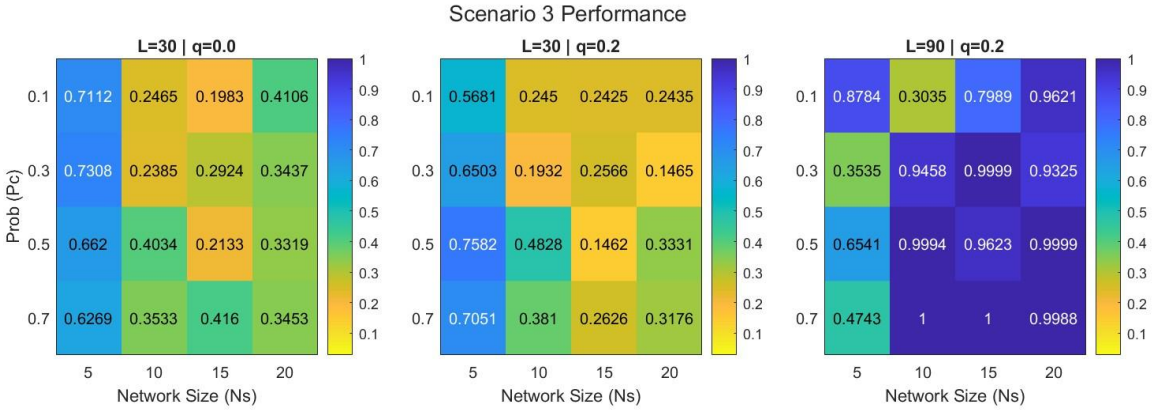

Figure 6.8 Heatmaps for Scenario 3, for R_{min}. Different panels refer to different configurations (length of pipes and demand)

6.4.3 Scenario 5 – Cost-Oriented Objective

Scenario 5 evaluates performance from an economic perspective, accounting for the cost of demand reduction associated with leakages. Among the analysed scenarios, this objective generally exhibits the weakest reduction performance.

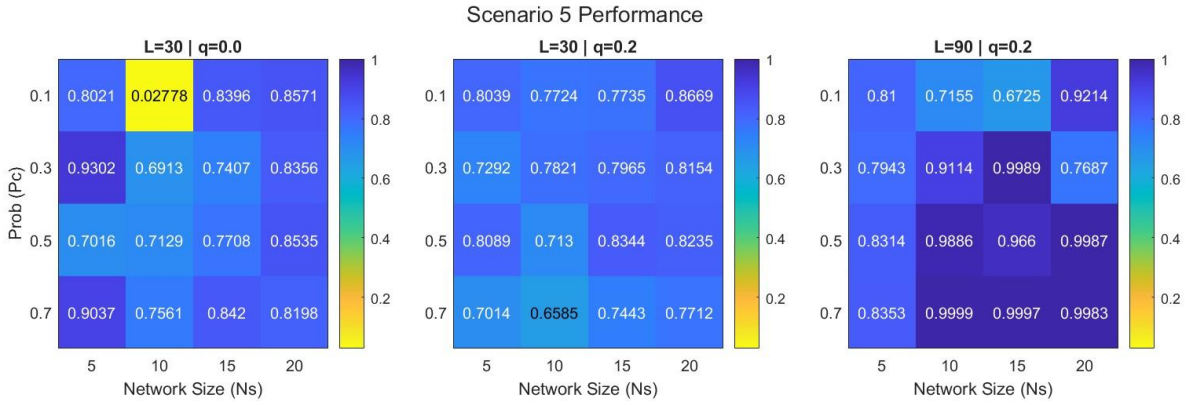

Figure 6.9 Heatmaps for Scenario 5, for R_{min}. Different panels refer to different configurations (length of pipes and demand)

6.4.4 Performance Comparison Across Scenarios

The averaged heatmaps of $R_{min}$and R_{mean}confirm clear differences among the three scenario types. Scenario 3 achieves the strongest overall reduction, Scenario 1 provides intermediate performance, and Scenario 5 shows the least reduction across the analysed configurations. This indicates that the optimization results are strongly influenced by the definition of the objective function.

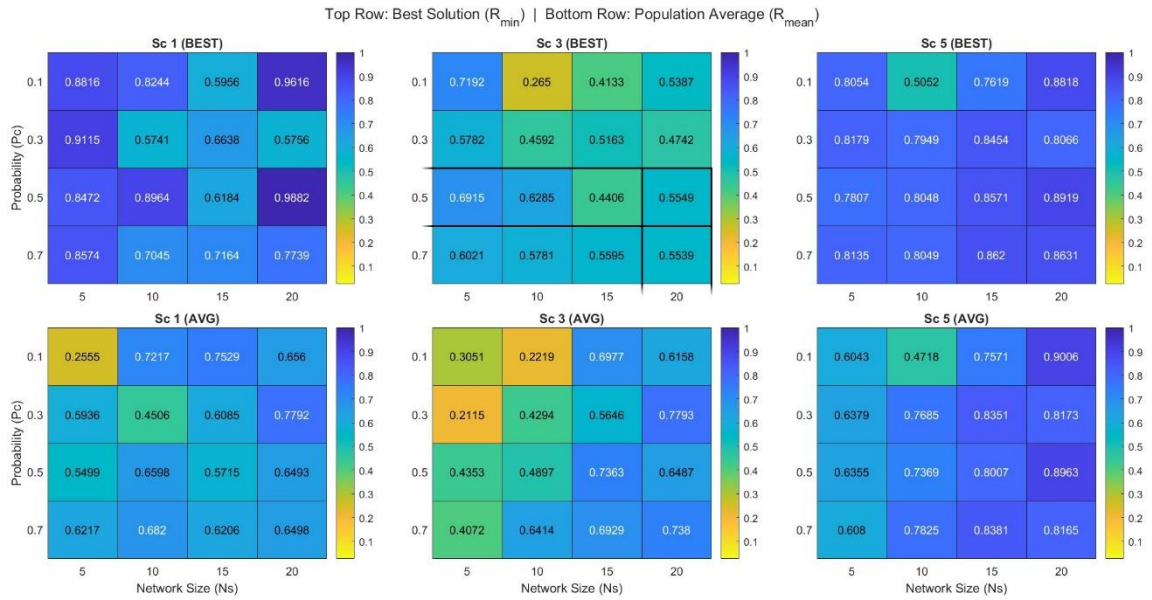

Figure 6.10 Heatmaps obtained considering all configurations (length of pipes and demand). Different panels refer to different scenarios (columns) and different ratios. The first row is the best achieved ratio; the second row is the average achieved ratio.

While the scenario comparison highlights the effect of different cost functions, the heatmaps also reveal a structural influence of Ns and Pc scenarios.

6.4.5 Structural Effects on Reduction Performance

The global heatmap analysis indicates that network structure plays a major role in optimization performance. Network size has a significant influence on reduction quality. Small networks provide limited structural diversity for optimization, while very large networks increase the complexity of the search space. As a result, moderate-sized networks tend to exhibit the strongest reduction performance. The role of connection probability is weaker but still contributes to the structural organization of the results. The comparison between $R_{min}$and R_{mean}also provides insight into convergence quality, since differences between best-solution and population-average performance indicate how widely the optimization improvement is distributed within the population.

6.5 Cluster-Based Behaviour Analysis

6.5.1 Global Cluster Structure

The clustering analysis reveals three distinct behavioural regimes across the simulations. In the three-dimensional feature space defined by reductions represented as logarithmic cost and minimum and mean ratios, the results organize into different groups rather than forming a continuous cloud. This indicates that the simulated networks naturally fall into different optimization behaviour families.

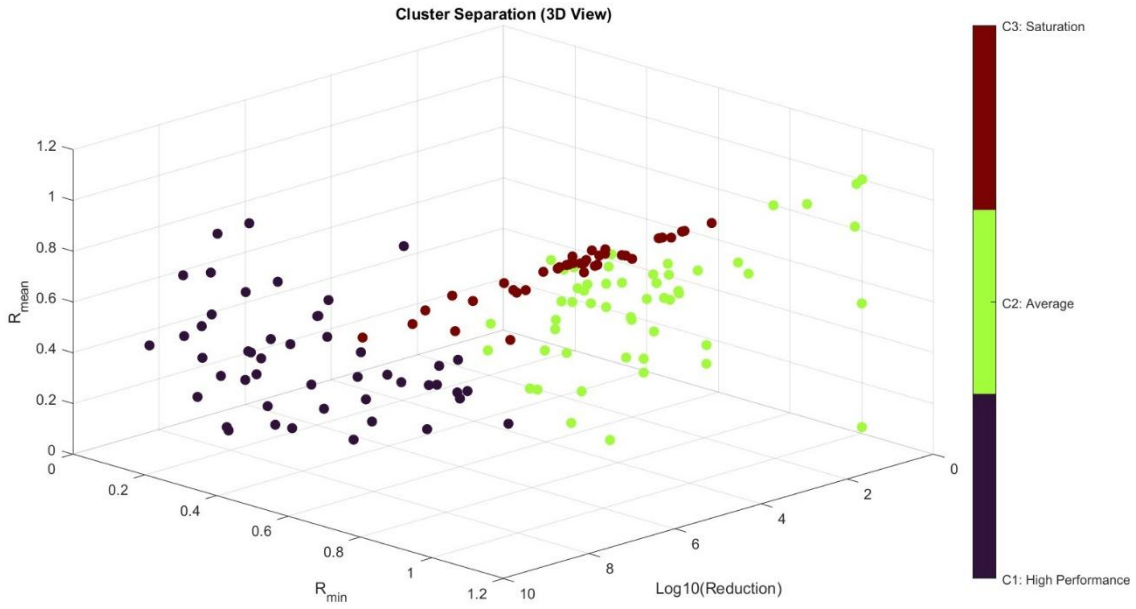

Figure *6.11 Cluster Separation*

6.5.2 Cluster Characteristics

Cluster 1 – High Efficiency Networks

Cluster 1 corresponds to the high-performance regime. As indicated by the algorithm character heatmap and the DNA bar chart, this cluster is associated with low connection probability, moderate network size, low pipe lengths, and low demand levels. At the same time, it shows the highest positive deviation in cost reduction and the lowest values of R_{min} and R_{mean}.

These characteristics indicate that the optimization algorithm is particularly effective in this regime. The low R_{min} and R_{mean} values show that both the best solution and the population average improve substantially during the search process.

An important observation from this cluster is that the best performance is not associated with the absolute smallest possible networks or the largest, but it was observed in moderate sized networks. This suggests that these configurations provide a favourable balance: the network is large enough to allow meaningful optimization, but not so large or interconnected that the search space becomes excessively difficult to explore.

Cluster 2 – Average Regime

Cluster 2 represents a regime of limited improvement. In the heatmap and DNA bar chart, this cluster is characterized primarily by small network size, moderate connection probability, and generally lower structural parameters. However, unlike Cluster 1, the cost reduction associated with this group is relatively weak.

This indicates that small networks do not necessarily produce the best optimization results. Although such networks are easier to analyse and simulate, they provide fewer structural alternatives for the optimization algorithm to alter. As a result, the achievable improvement remains limited.

In this cluster group, cost reduction came out as lower than the average cost, whereas the reduction ratios are more or less average. The reason of this is the fact that the networks in this cluster are small and the absolute cost has a limited capacity to decrease.

Cluster 3 – Saturation Regime

Cluster 3 represents the saturation regime. This cluster is associated with high connection probability, large network size, larger pipe lengths, and higher demand levels. In the heatmap and DNA chart, these variables all show strong positive deviations from the global mean. At the same time, R_{min} and R_{mean} are also relatively high, while the cost reduction is much weaker than in Cluster 1.

These results indicate that increasing network complexity does not lead to proportional improvement in optimization performance. On the contrary, as the network becomes larger and more connected, the search space grows rapidly and the optimization process becomes

increasingly difficult. In this regime, the algorithm is still able to identify improvements, but these improvements came out to be less.

For this reason, Cluster 3 can be interpreted as a diminishing-returns regime, where additional structural complexity leads to weaker relative optimization gains. This cluster therefore highlights an important practical result of the analysis: beyond a certain level of network complexity, the effectiveness of the optimization process tends to saturate.

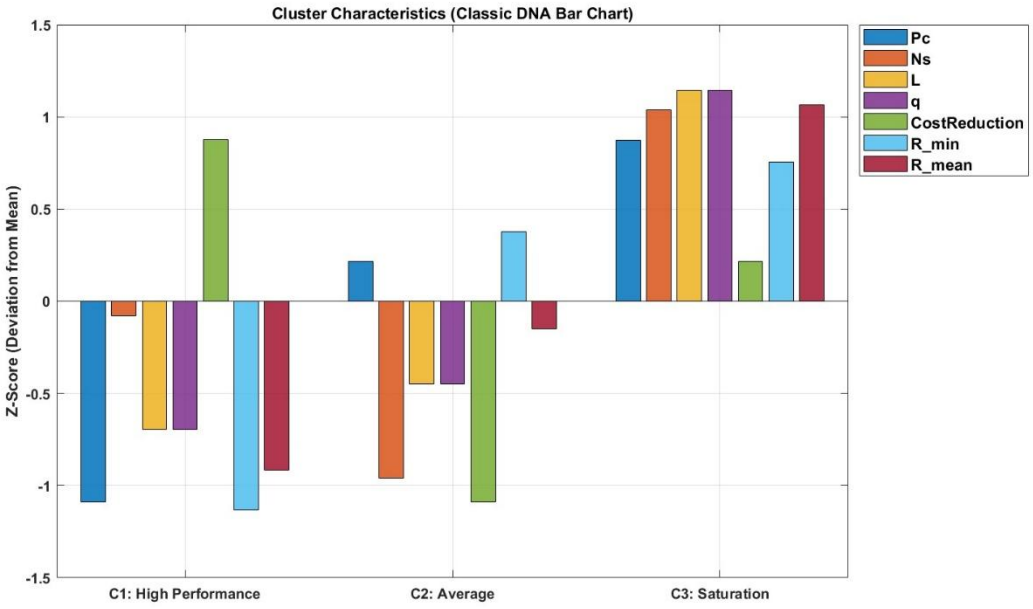

Figure 6.12 Cluster Characteristics

Role of the Scenario Distribution

The scenario distribution plot provides an additional interpretation of the clusters by showing how the three different scenarios are distributed across the regimes. This figure indicates that the identified clusters are not purely geometric features, but are also related to the way the scenario related optimization interact with the structural properties of the network.

The presence of different scenario contributions within each cluster suggests that optimization behaviour is jointly controlled by network structure and scenario. In particular, some scenarios

appear more frequently in the high-efficiency regime, while others are more represented in the average or saturation regimes. This supports the idea that the same network topology may exhibit different optimization behaviour depending on the cost scenario used.

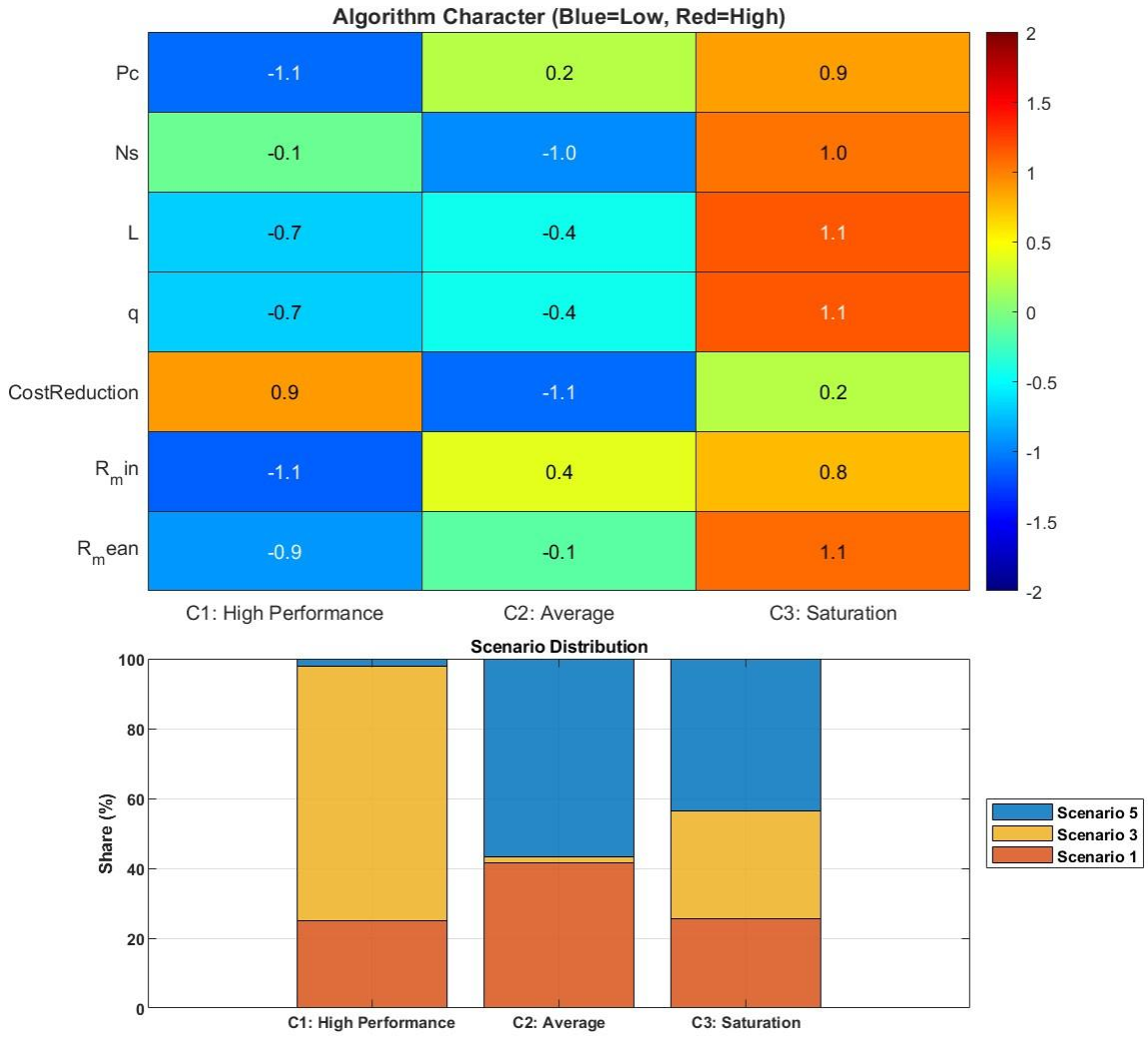

Figure 6.13 Algorithm character by clusters

As can be seen from Figure 6.13 that, High performance cluster mainly consists of Scenario 3 related objective. This outcome supports the results that are obtained from the heat map results from previous section.

This cluster-based interpretation is more deep and integrated because it moves the analysis beyond the individual simulation level and reveals broader structural patterns leading the optimization behaviour of the synthetic water distribution networks.

6.5.3 Spatial Distribution of Clusters

The cluster decision map confirms a clear structural organization in the N_s , P_c space. Cluster 1 is mainly associated with moderate network size. Cluster 2 is associated primarily with small networks. Cluster 3 is concentrated in large networks. This spatial pattern confirms that the topology of the synthetic network strongly influences the optimization regime observed in the results.

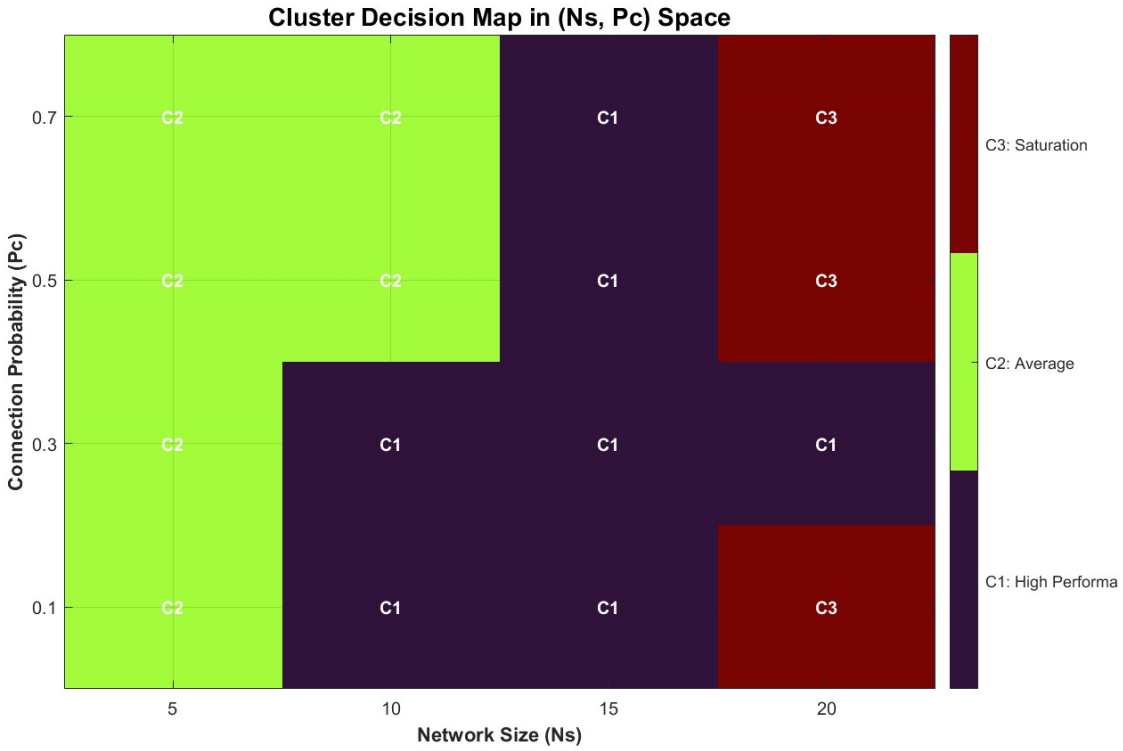

Figure 6.14 Cluster Decision Map

Overall, the results of the cluster analysis indicate that the optimization performance of sequential pipe replacement is governed by the combined effect of objective formulation and network structure. While the comparison of scenarios shows that different optimization priorities lead to different convergence speed and reduction performance, the analyses across N_s and P_c demonstrate that structural properties display a strong influence on the algorithms optimization capabilities. In particular, moderate-sized networks appear to provide the most balanced and strong reduction performance, allowing both meaningful cost reductions and consistent convergence behaviour, whereas very large and highly connected networks tend to exhibit more limited improvements. The cluster-based interpretation further confirms that the observed patterns are not random, but correspond to identifiable classes of hydraulic and structural response. Taken together, these findings show that the effectiveness of renovation planning should be evaluated with the optimization scenario and network configuration together.

7 Network Resilience Analysis

7.1 Problem Statement

Pipe replacement works in urban water distribution systems do not affect only the hydraulic performance of the network. In real urban environments, these interventions also occupy street space and may temporarily block roads, reduce accessibility, and alter the mobility conditions of the surrounding area. For this reason, the assessment of renovation strategies should not be limited only to hydraulic indicators, but should also consider the extent to which local accessibility is disrupted during construction works.

To address this issue, a shortest-path-based accessibility analysis was developed. The main idea is to compare the shortest travel paths available in the normal road network with those available after a road closure caused by pipe replacement works. In this way, the indirect impact of water infrastructure interventions on the urban transportation system can be quantified through the additional detour required to move between nearby streets. This provides a simple but effective measure of network resilience from the perspective of accessibility.

7.2 Network Model Construction

The road network used for the accessibility analysis was obtained from OpenStreetMap and refined in QGIS before being imported into MATLAB. During this preprocessing phase, the road geometry was cleaned, and the basic network structure was prepared for graph-based analysis. The road type information was also assigned, including the distinction between one-way and two-way streets where relevant. This information of the road type was very important for the correct implementation of the algorithm.

After preprocessing, the road network was represented as a graph. In this representation, road intersections are treated as nodes and road segments are treated as edges. The weight of each edge is defined by the corresponding Euclidean distance between connected intersections. This formulation allows the urban road network to be analysed in a mathematically consistent way and provides a direct basis for shortest path computations. Because the objective of the analysis

is to evaluate detours caused by local closures, a graph representation is particularly appropriate, since it transforms the accessibility problem into a pathfinding problem over a weighted network.

7.3 Road Closure Simulation

To simulate the impact of pipe replacement works on road accessibility, a specific road segment was selected as the location for the pipe replacement. This selected segment represents the street section occupied by the excavation and construction activity associated with pipe renewal. The corresponding road segment was then removed or restricted in the graph in order to reproduce the effect of temporary closure during works.

Two closure scenarios were considered.

In Scenario A, the road segment itself is closed to traffic, while the intersections at both ends of the segment remain open. This configuration represents a relatively mild disruption, in which the blocked street cannot be directly crossed, but adjacent junctions remain functional and turning movements are still possible.

In Scenario B, both the road segment and the connected intersections are considered closed. This represents a more severe disruption, where the intervention affects not only the road itself but also the immediate junction accessibility around the work zone. Moreover, the connection to the adjacent streets becomes a problem. For example, some of the one way roads cannot be accessed from the intersection, so that road is completely inaccessible.

The distinction between these two scenarios is important because the real impact of pipe replacement depends strongly on the extent to which the construction area occupies the surrounding street space.

Road closure rerouting simulation is summarized in the following figure.

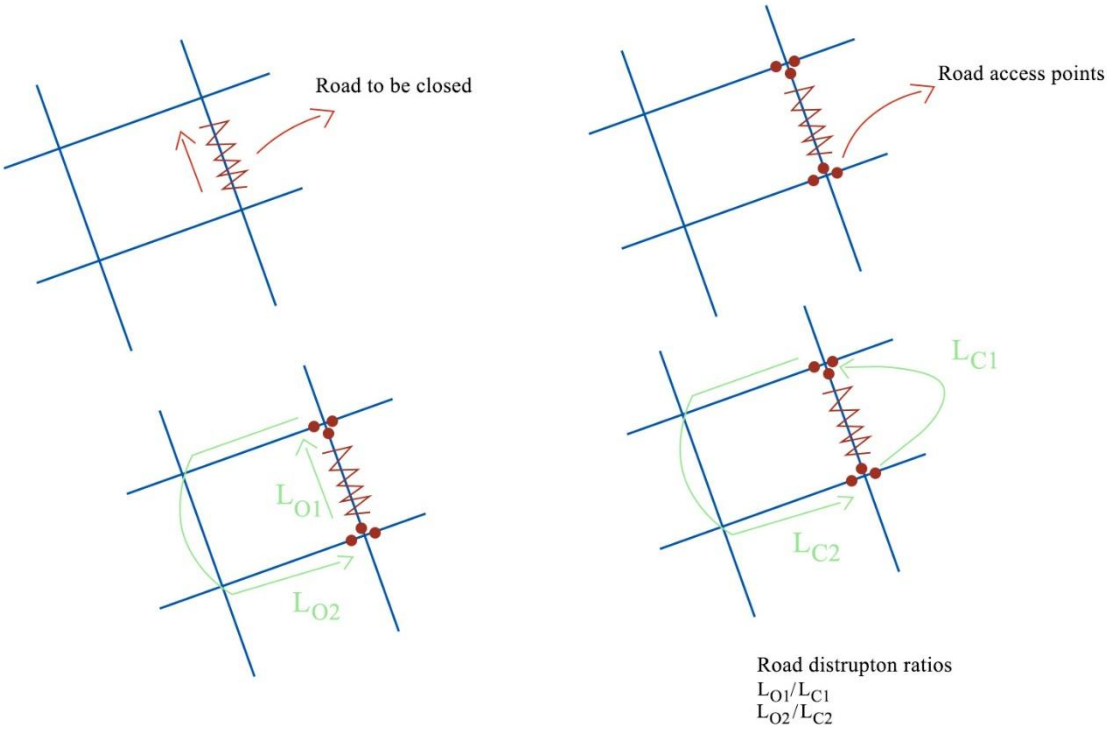

Figure 7.1 Road Closure Rerouting

7.4 Automated Accessibility Analysis Using Shortest Path Algorithm

To evaluate the local accessibility impact of the closure, an automated shortest path procedure was developed. The algorithm first identifies the road segment that is selected by the user. Then the algorithm identifies all valid road segments connected to the road intersections associated with the selected road closure area. In order to represent accessibility along the neighbouring streets, test points are placed at a distance of 1.0 m along the connected roads in all available directions. Linear interpolation is used to ensure that these points lie exactly on the road geometry, in order to prevent connectivity issues across the roads.

The graph is then modified according to the analysed condition. In the normal case, the test nodes are inserted into the existing road graph without changing the underlying connectivity. In the disrupted case, the critical road segment or the corresponding critical node is removed depending on the selected scenario based on the closure details. This enables the algorithm to compare the normal and disrupted configurations within the same graph-based framework.

Shortest paths between each possible pair of test points are then computed using Dijkstra's algorithm. For every pair, the algorithm compares the path length in the undisturbed network with the corresponding path length under closure conditions. The difference between these two values is interpreted as the accessibility penalty, or detour cost, introduced by the road closure.

In addition, the results are used to classify neighbouring points as either reachable or isolated. Reachable points are those for which a valid connecting path still exists after the disruption, whereas isolated points correspond to situations in which the target cannot be reached because the closure eliminates all connecting routes. This classification adds an additional resilience dimension to the analysis by distinguishing between disturbed accessibility and complete loss of accessibility.

7.5 Case Study for Network Resilience Analysis

To demonstrate the applicability of the proposed accessibility analysis framework, a real-world road network from the town of San Vito dei Normanni, Italy, was selected as a case study. The purpose of this application is not to build a complete traffic assignment model, but rather to show, through a realistic urban network, how localized pipe replacement works may influence road accessibility in the surrounding area. In this sense, the case study acts as a practical extension of the hydraulic renovation problem, illustrating that water network interventions may also generate temporary transportation impacts.

San Vito dei Normanni was considered suitable for this application because it represents a small urban settlement with a structured road network and a moderate level of traffic complexity with high numbers of one way roads. These characteristics make it appropriate for testing a graph-based accessibility framework, since the effects of a local disruption can be observed clearly without the additional complexity of a large metropolitan system. The road network data were obtained from OpenStreetMap and processed in QGIS to refine the geometry and remove unnecessary elements before being exported for analysis.

Figure 7.2 San Vito dei Normanni QGIS

7.5.1 Road Network Representation

The processed road network of San Vito dei Normanni was exported from QGIS as a shapefile and then imported into MATLAB. Once imported, the network was converted into a graph topology in which each road intersection is represented as a node and each road segment is represented as a weighted edge. The edge weights correspond to the distances between intersections. This representation makes it possible to analyse the urban road system using graph-based techniques and, in particular, to apply Dijkstra's shortest path algorithm in a direct and computationally efficient manner.

Figure 7.3 San Vito dei Normanni MATLAB

7.5.2 Road Closure Scenario

To simulate the impact of pipe replacement works on urban accessibility, a representative road segment located within the study area was selected and removed from the road network. The selected location is used as the closure point in the case study and represents a local street section affected by construction activities. The road network graph is then modified according to the procedure introduced previously in order to reproduce the temporary disruption created by the intervention.

This case study was analysed under the two previously defined closure configurations. In Scenario A, the road segment is blocked while the adjacent intersections remain open. In Scenario B, both the road segment and the intersections are closed. This comparison allows the analysis to capture not only the impact of blocking a street section, but also the additional accessibility loss that arises when the closure spreads to the surrounding junction space.

Figure 7.4 Selected Road to be closed

7.5.3 Accessibility Evaluation

The accessibility impact of the selected closure was evaluated using the automated shortest path method. Test points were generated along the road segments connected to the affected intersection in order to represent accessibility along the surrounding streets. The shortest path distances between these points were then computed under three different network conditions: the normal network, Scenario A with the road segment closed, and Scenario B with both the road segment and the intersections closed. By comparing these distances, the algorithm quantifies the detour required due to the disruption.

Figure 7.5 Closure Scenario A

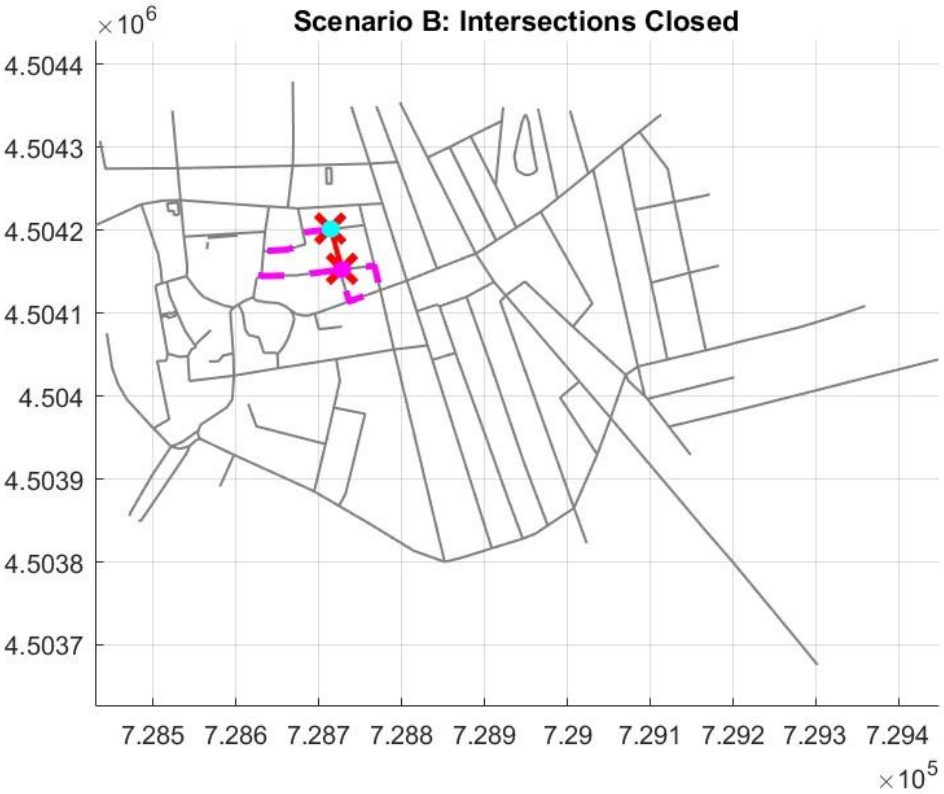

Figure 7.6 Closure Scenario B

The numerical results show that the impact of the closure depends strongly on the assumed scenario. When the intersections remain open in Scenario A, alternative routes are still available, and the accessibility loss remains moderate relative to the more restrictive case. For example, the routes A1→B1, A1→B2, A2→B1, and A2→B2 increase from an original distance of 52.6 m to approximately 226.5–230.5 m, corresponding to detour increases of about 330.6–338.2%. These values indicate that the closure forces a substantial rerouting, but the surrounding network still preserves functional connectivity.

A much stronger effect is observed in Scenario B, where the intersections are also closed. In this case, the loss of local junction accessibility significantly amplifies the detour. The most critical route, A1→B2, increases from 52.6 m to 372.1 m, corresponding to a detour increase of 607.4%. Similarly, A2→B1 and A2→B2 rise to 306.4 m and 293.1 m, corresponding to increases of 482.5% and 457.2%, respectively. These results show that the closure of intersections can have a much more severe impact than the closure of the road segment alone, because it eliminates nearby turning opportunities and forces circulation through much longer alternative paths.

At the same time, the results also indicate that the impact is not uniform in all directions. Some routes remain unchanged in Scenario A, such as B1→A1, B1→A2, B2→A1, and B2→A2, all of which show 0% increase under road-segment-only closure. Even under Scenario B, some paths remain less affected than the most critical cases, and B2→A2 remains unchanged at 128.4 m. This indicates that the surrounding road network still provides alternative connections for some directional movements. Therefore, the accessibility impact of pipe replacement works depends not only on whether a street is blocked, but also on the local topology of the road system and on the directional redundancy of alternative routes.

Route	Original (m)	Road Closed (m)	Road Closed (%)	Intersection Closed (m)	Intersection Closed (%)
A1 -> B1	52.6	226.5	330.6	226.5	330.6
A1 -> B2	52.6	228.5	334.4	372.1	607.4
A2 -> B1	52.6	228.5	334.4	306.4	482.5
A2 -> B2	52.6	230.5	338.2	293.1	457.2
B1 -> A1	132.4	132.4	0	324.2	144.9
B1 -> A2	130.4	130.4	0	202.3	55.1
B2 -> A1	130.4	130.4	0	250.4	92
B2 -> A2	128.4	128.4	0	128.4	0

Table 7.1 Road Detour Changes

Overall, the case study demonstrates that even a single localized road closure associated with pipe replacement works can generate substantial accessibility penalties in the surrounding urban network. The comparison between Scenario A and Scenario B further shows that modelling the status of the adjacent intersections is a critical aspect of the analysis, since the resulting detour may be significantly underestimated if only the road segment itself is considered closed. These findings highlight that rehabilitation planning in water distribution systems should not be evaluated solely in hydraulic terms. Even a localized water infrastructure intervention may affect local mobility conditions.

Appendix

Appendix A.1. Parameter initialization

The synthetic-network generation and optimization framework starts with the definition of topology, hydraulic, degradation, and PBIL parameters. These parameters control the network structure, hydraulic loading, leakage conditions, and optimization settings adopted in the analysis. A simplified MATLAB-style parameter definition is reported below.

```
% Topology parameters
parameters_topology.P_c = ...;          % Connection probability
parameters_topology.N_s = ...;          % Network size
parameters_topology.L   = ...;          % Pipe length
parameters_topology.random_gen = ...;   % Random seed / generation flag
% Hydraulic parameters
parameters_hydraulics.q                = ...;   % Nodal demand
parameters_hydraulics.epsilon_old      = ...;   % Pipe roughness
parameters_hydraulics.H_r              = ...;   % Reservoir head
parameters_hydraulics.H_min            = ...;   % Minimum required head
parameters_hydraulics.leakage_residual = ...;   % Residual leakage level
% Degradation parameters
parameters_degradation.percentage_pipes = ...;  % Fraction of degraded pipes
parameters_degradation.leakage          = ...;  % Major leakage level
parameters_degradation.insufficient     = ...;  % Insufficient-supply condition
% PBIL parameters
parameters_PBIL.choosen_district = ...; % Selected district
parameters_PBIL.scenario_num     = ...; % Optimization scenario
parameters_PBIL.N                = 25;  % Number of sequence vectors
parameters_PBIL.max_Pipe_change  = 1;   % Maximum number of pipe changes per phase
```

Appendix A.2. Centrality-informed diameter assignment and optimization

The following simplified MATLAB-style skeleton summarizes the main logic of the integrated diameter-design procedure. The full implementation includes hydraulic cost evaluation, centrality-based pipe ranking, and Harmony Search optimization

```
function topology2 = assign_diameters(...)
% Initialize Harmony Memory with candidate diameter sets
rand_index_M = randi(n_D, HMS, n_P);
% Evaluate initial candidates with the hydraulic cost function
for each candidate
    compute total_cost_a
end
% Compute tailored betweenness centrality and rank pipes
[btwrel_a, ~] = tailored_btwn(G, 0, 0);
[~, sort_list_btw] = sort(btwrel_a);
% Reorder diameter sets according to centrality ranking
for each candidate
    sort diameter indexes
    assign larger diameters to more central pipes
end
% Harmony Search optimization loop
for each iteration
    generate updated candidate using HMCR and PAR
    reorder according to centrality ranking
    evaluate updated candidate with cost_f
    replace worst harmony if improved
    update best solution
end
% Store final diameter configuration
topology.D_conduit_good = D_comm_a(best_indexes);
topology.best_indexes   = best_indexes;
topology.btwrel_a       = btwrel_a;
topology2 = topology;
end
```

References

[1] Avarzamani, M., & Vesipa, R. (2024). *Optimal sequence of replacements in the renovation of water distribution networks*. IAHR. https://www.iahr.org/library/info?pid=39323

[2] Eliades, D. G., Kyriakou, M., Vrachimis, S., & Polycarpou, M. M. (2016, November). *EPANET-MATLAB toolkit: An open-source software for interfacing EPANET with MATLAB*. In *Proceedings of the 14th International Conference on Computing and Control for the Water Industry (CCWI)* (p. 8). https://doi.org/10.5281/zenodo.831493

[3] Engelhardt, M. O., Skipworth, P. J., Savic, D. A., Saul, A. J., & Walters, G. A. (2000). Rehabilitation strategies for water distribution networks: A literature review with a UK perspective. *Urban Water, 2*(2), 153–170. https://www.sciencedirect.com/science/article/pii/S1462075800000534

[4] Giustolisi, O., Ridolfi, L., & Simone, A. (2019). Tailoring centrality metrics for water distribution networks. *Water Resources Research, 55*(3), 2348–2369. https://doi.org/10.1029/2018WR023966

[5] MathWorks. (n.d.). *Choose cluster analysis method*. MATLAB & Simulink Documentation. https://www.mathworks.com/help/stats/choose-cluster-analysis-method.html

[6] MathWorks. (n.d.). *Introduction to machine learning.* MATLAB & Simulink Documentation. https://www.mathworks.com/discovery/machine-learning.html

[7] Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems, 13*(1), 911–921. https://doi.org/10.12785/ijcds/130172