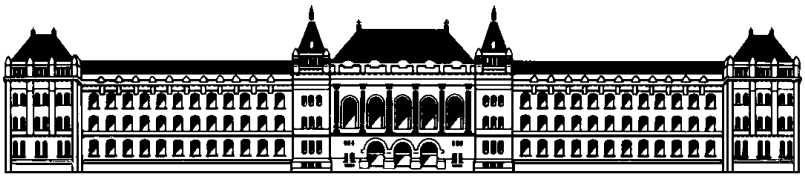

Budapest University of Technology and Economics
Faculty of Mechanical Engineering
Department of Hydrodynamic Systems

Vulnerability and sensor placement analysis of water distribution networks

Booklet of PhD dissertation

Written by: Richárd Wéber
mechanical modelling engineer

Supervisor: Dr. Csaba Hős
associate professor

1 April 2021, Budapest

1 Introduction

As clean drinking water is a fundamental demand of the population, water distribution networks (WDNs) are one of the most elementary infrastructures of modern settlements from small villages to large metropolises. The raw water can be extracted by drilled wells nearby rivers and lakes, or from karst layers nearby mountainous countryside. A typical arrangement is illustrated in Figure 1. The water, extracted by wells requires treatment, that can include filtering with different sizes of grids, and adding chlorine for killing pathogens and other parasites. The WDN transports the clean drinking water to demand locations, e.g. residential areas where people live or factories where the water might be used for production. Figure 1 also indicates the basic parts of WDNs, that is, a pipeline system connected through nodes (grey) with several water tanks (purple) and pumps (yellow). The focus of the dissertation is analysing the WDNs. Proper operation of WDNs is essential from the viewpoint of the inhabitants' health, living standards and industrial production.

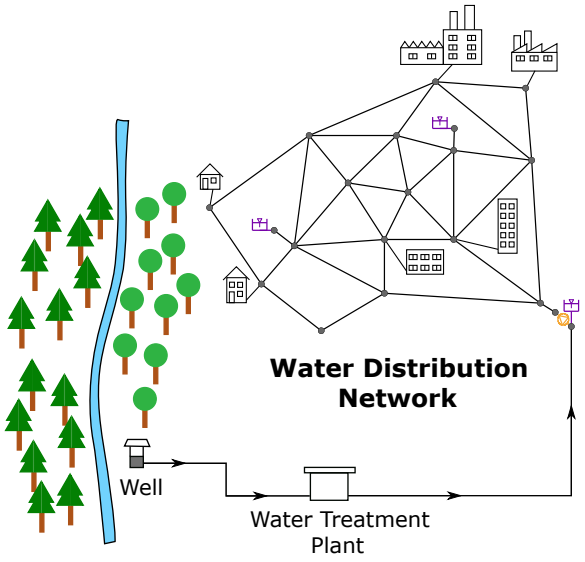

Figure 1. Typical arrangement of serving clean drinking water to inhabited area.

The complexity of these systems increases quickly with the number of demand (consumption) points (that is, the size of the area served) resulting in the challenging task of instrumentation and supervision of such large systems. Hydraulic models have been applied for the description of WDNs for decades, nowadays due to the large computational resources and enormous databases, even the smallest details can be resolved in a mathematical model, thus it might contain thousands or ten-thousands of pipelines. For increasing the accuracy, calibration can be performed using real-life measurement data. One of the emerging challenges is that as *the number of measurement devices - pressure loggers and/or flow meters - is far less than the number of potential measurement points* (typically fire hydrants for pressure), one has to optimise (in some sense) the layout of

the measurements. Another issue with the increasing complexity is that the number of parameters to be calibrated in a WDN computer model escalates with its size and the calibration process requires both large computational effort and highly sophisticated algorithms. Moreover, predicting the robustness or *the vulnerability of a WDN during the design phase or in the case of working networks is still a challenging task* nowadays. Even with the help of a large Geographic Information System (GIS) database and a detailed, properly calibrated hydraulic model, the consequences of a pipe burst is yet difficult to predict accurately.

The fundamental equations of the hydraulic modelling are introduced briefly. Although the flow in a pipeline system is necessarily three-dimensional, during the modelling of large WDNs, it is assumed to be *one-dimensional*, which, on the one hand, provides adequate accuracy, on the other hand, ensures favourable computational time. Furthermore, the flow is considered *incompressible and steady-state*. Even though an in-house C++ software (`STACI`) was used for simulations, the underlying equations are the same as those implemented in `EPANET` [1] that is the most common solver in the industry and also in research areas.

The *mass conservation law* is solved for each node

$$\sum_{i \in \text{in}} Q_i - \sum_{j \in \text{out}} Q_j = c \tag{1}$$

where it is assumed that the fluid is incompressible (i.e. liquid). Q stands for the volume flow rate, the first term with "in" notation indicates the inflows, while the second with "out" the outflows and c is the nodal demand (consumption). The other set of equations describes the relationship between the pressure difference across an edge (e.g. pipe, pump, valve) and the volume flow rate, and formally, it is given by

$$\Delta p = f(Q) \tag{2}$$

where Δp stands for the pressure difference between the two ends of the edge and the function f is typically a nonlinear function of the volume flow rate. The exact formula depends on the actual type of the edge. In overall, there is a set of algebraic, nonlinear equations with thousands of unknown variables, that can solved efficiently applying Newton's technique and sparse LU decomposition. [2, 3]

2 Calibration and sensor placement

2.1 Introduction

The general purpose of calibration is to reduce the discrepancy between the mathematical model and the measurements taken on the actual system by adjusting model parameters. In the field of WDNs, this typically means pressure measurements on fire hydrants; although in some cases there are built-in flow meters as well. Such calibration requires two steps:

- designating the sampling points, i.e. the nodes at which pressure values are measured,
- and a calibration algorithm that identifies the designated model parameters.

The calibration and the sensor placement problem focuses on the accuracy of the model output i.e. the purpose is to minimise the discrepancy between the measured pressure distribution and the output of the calibrated model by adjusting the model parameters. Note that the calibration and sampling techniques are independent of each other and can be applied separately. Since the mathematical model contains several independent groups of parameter, e.g. roughness, nodal demand or pipe diameter. In this work, I am focusing on the *roughness coefficients* of the pipelines, even though, other parameters can be just as important or even more critical. Although the proposed calibration technique strictly considers the roughness coefficients, *the methods can be further extended to calibrate different parameters.* Finally, last decades brought numerous different approaches for answering both questions, and the literature has a lack of the evaluation of their performance; *I introduce a novel approach for comparing different sensor placements objectively* in the aspect of their accuracy in the outcome of the calibration.

The accurate approximation of the complex fluid mechanical phenomena lying behind the effect of the wall roughness is a challenge for engineers, especially, if the proposed method must be computationally efficient and clearly applicable. During the last decades, mostly two roughness models are applied to estimate the pressure loss due to pipe wall roughness in the case of WDNs: the *Darcy-Weisbach model* and the *Hazen-Williams correlation.* The former one has more theoretical foundation and it is also more general, the latter one is a purely empirical formula, widely used mainly in the United States and provides an accurate prediction only for water and in the turbulent regime. Both models contain a parameter: the Darcy-Weisbach model includes the relative roughness coefficient (ε), and the Hazen-Williams the C-factor. Even if these parameters of newly installed pipelines can be accurately estimated, during years due to the sedimentation, regular calibration is inevitable.

2.2 Mathematical tools

For the methods two important mathematical tools is introduced first. The *sensitivity* is the variation of the quantities of interest (notably pressure and/or flow rate values) with

respect to (wrt) a parameter; say, the pipe roughness parameter μ which can be either the roughness coefficient (ε) or the C-factor. That is, the variation of the pressure at the i-th node wrt to a small change in the parameter μ of the j-th pipeline is

$$S_{i,j} = \frac{\partial p_i}{\partial \mu_j}. \tag{3}$$

This quantity can be computed efficiently for each node-edge pair by utilizing the LU decomposition [3] of the Jacobian, and in overall the sensitivity matrix can be built up.

Based on the pressure-roughness sensitivity matrix, I define an indicator for the nodes, that represents how much a node pressure is influenced by every pipeline roughness parameter. Mathematically, this means the column sum of the absolute values of the matrix elements, i.e.

$$S_{\mathrm{node},i} = \sum_{j=1}^{N_{\mathrm{edges}}} |S_{i,j}|. \tag{4}$$

$$\tag{5}$$

These variables forecast the increase (or decrease) of a nodal pressure in the case when the roughness of every pipe is (equally) increased.

The *distance* is a mathematical tool originating from graph theory. A WDN can be interpreted as a graph where the pipelines, valves, pumps etc. are the edges (links) and their intersections (nodes) are the vertices. In graph theory, a path is defined as a route through links between two nodes (possibly far away from each other), see e.g. [4]. Clearly, in a looped network there could be several different paths between two particular nodes and, in the simplest case, the shortest path or distance is the one with the minimum number of edges. However, the distance or length can also be defined in the case of a weighted graph, see [4]. For computing the distance, the C++ library `igraph` was used, see [5].

2.3 The calibration technique

The last decades brought numerous techniques for the calibration of WDNs applying omnifarious approaches from the Levenberg-Marquardt algorithm [6], through the particle swarm optimization [7] until the Bayesian-based, two-level Markov chain Monte Carlo-Particle Filter method [8]. Instead of detailing these approaches, I give a short summary emphasizing which properties and ideas are typical.

- First, the most popular parameter to be calibrated is the roughness parameter [9, 10] or the nodal demand [11, 12] (or demand pattern), sometimes both of them simultaneously [13, 8].

- In order to decrease the number of unknown parameters, engineers tend to create several groups, then assign each group only one unknown. In the case of the roughness parameter, such grouping is typically based on the pipe material, while in the case of

demand, the grouping is often based on the geographical distribution of the nodes. [13, 10, 7]

- Next, a scalar objective function is defined, which is often the discrepancy between the measurement and the model output or the difference between the model parameters and the "real" ones. [14, 15]

- Finally, engineers select a sophisticated optimization techniques to minimize the objective function, i.e. minimize the difference between the measurement and the model. In this aspect, heuristic methods have spread in the last decade utilizing the constantly increasing computational performance [8, 9]. However, most of them still requires a serious amount of CPU time.

My main goal for the calibration is to develop a computationally efficient yet accurate technique in the sense that the output of the model (the pressure values) should be as close to the measurements as possible. This also means that the parameters might not be as close to the "real" ones, but determining the exact roughness parameters properly is also more difficult, and the purpose of a model is to achieve accurate model outputs. *In overall, my primary aim is to optimize the model output and not the parameters themselves.* Ensuring these aspects and based on the experience of the literature, there are two important features of the proposed calibration technique.

- It assigns independent roughness coefficient to each pipeline i.e. grouping is not applied as in e.g. [16]. The advantage of grouping similar pipes (based on material, age etc.) and then assigning one single friction coefficient to all pipes within a group reduces the problem to a large extent, but it might require more input data that might not be available or uncertain. This also means that the number of independent measurements is typically lower with a magnitude than the unknown model parameters.

- It is iterative, yet no heuristic optimisation (e.g. genetic algorithm, simulated annealing) is required.

The proposed calibration method is part of Thesis 1.

2.4 The sensor placement strategy

Due to the rapid development of the measurement devices, nowadays smart sensors can be used in WDNs for various purposes: leakage detection [17], chlorine decay modelling [18], contamination detection [19] or, as in this dissertation, roughness parameter calibration. Several different approaches were defined recently originating from graph theory based metrics [20], geostatistical tools [21] or general hydraulics based methods [11]. Yet, *the most decisive direction in the literature about the sensor placement is to decrease the measurement error propagation to the model parameters, or to the output of the model.*

This approach is generally applied in different fields as well; however, in terms of hydraulic modelling, it was originally presented in [22].

Based on the experience from the literature and intuitively, a "good" sensor placement fulfils two requirements. First, sensitive nodes are used for sampling. As a counterexample, it is inefficient to deploy pressure loggers close to reservoirs, where the pressure is primarily defined by the reservoir water level. However, sensitive nodes tend to accumulate at certain locations within a WDN, thus simply choosing the most sensitive nodes often results in the accumulation of sampling points close to each other [23, 24]. The problem with close or neighbouring samplings is simply that the measurements are not independent of each other, which leads us to the the second requirement: the measurement points should be "far enough" from each other to be "as independent as possible".

This leads me to apply two concepts: the *nodal sensitivity* ($S_{\text{node},i}$) and the *hydraulic distance* (δ). The first one is stemming from the sensitivity matrix (see Equation (3)). The latter one is the distance of the actual node from the nodes of prescribed pressures (e.g. reservoirs, tanks) or sampled pressures, measured by weighting the edges by frictional pressure loss, i.e. the pressure difference without the geodetic head difference. A node can then be characterised by *the product of its nodal sensitivity and the hydraulic distance*. Note that if a sensitive node is designated for measuring, the nodes surrounding it (probably also with high sensitivity) will become of low importance as they are close to a master node. Similarly, a node far away from the master nodes but with low sensitivity is not suitable for calibration. This technique is sequential in the sense that it adds sampling points one by one; after a node is designated for sampling, only the distances must be recomputed that is computationally cheap. Nevertheless, *the strategy is straightforward and free of iteration or heuristic optimisation.* Since the method is based on these two quantities, I call it *hydraulic distance with sensitivity* i.e. HDS and the method is described in Thesis 2.

2.5 Comparison of sensor placement techniques

In this section, I present a detailed comparison of different sensor placement strategies including the proposed HDS technique and other approaches available in the literature. I have chosen the most common approaches from the literature, that is, the most widely cited ones. The proposed calibration technique is used for every method during the comparison. This ensures that once the sampling nodes are designated, the calibration process will be the same for all methods. *Even though I performed this comparison for all of the available real-life and articifial networks, I am presenting the results of two of them only, which are showing a representative picture, and also visualizing more networks would not change the overall conclusions.* One of them being the ky2 network [25, 26], while the other is a real-life network from Sopron, namely SN-26. The proposed method for objective comparison of sensor layouts is described in Thesis 1.

The left-hand side of Figure 2 depicts the results for the artificial ky2 network, while the right-hand side shows the results for SN-26. Different colours belong to different sensor placement techniques, the proposed method is indicated as HDS, while the rest of the

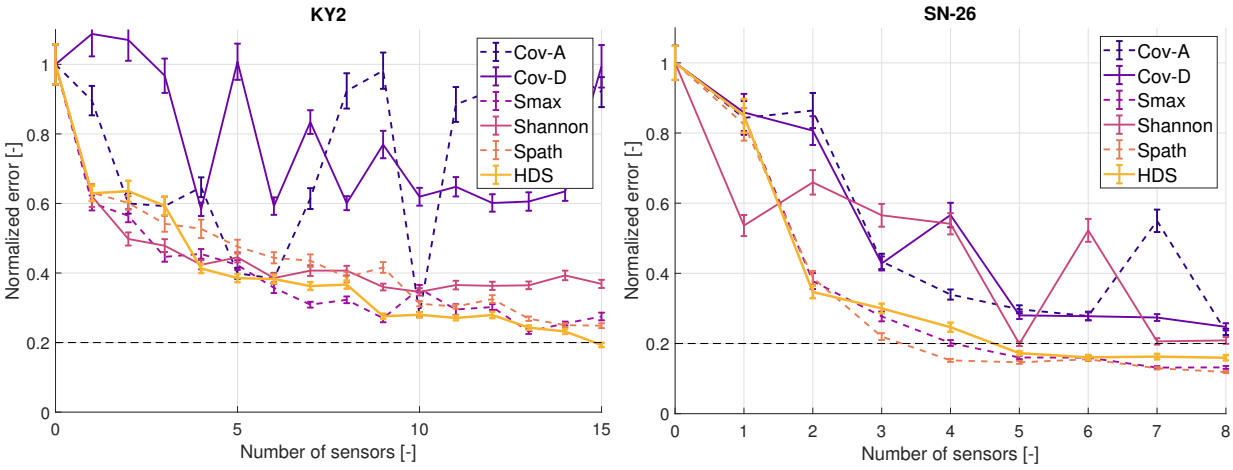

Figure 2. Normalized error of the calibration with standard deviation of the average as error bars as a function of the number of sensors for the ky and SN-26 networks.

methods are originated from the literature. I depicted the average pressure error defined in Thesis 1, which is the discrepancy between the model output and the measurements, including unsampled nodes too, against the number of sensors. I also indicated the standard deviation of the average, based on a hundred calibration runs using error bars. One sampling method is considered superior if it converges faster to zero. The dashed, horizontal line highlights the number of sensors needed to decrease the error of the initial "uncalibrated" error to its 20%. In the case of ky2 (see the left-hand side of Figure 2) four techniques converge similarly, but HDS is slightly ahead of Smax. Checking the left-hand side of the figure, system SN-26, HDS, Smax and Spath techniques are relatively close to each other, the latter one slightly outsmarts the rest.

Finally, it is essential to highlight the CPU time that is required to find an optimal layout. Table 1 contains the runtime values in seconds for both Sopron Networks, C-town and ky2. It can be seen that the covariance-based methods (Cov-A, Cov-D) require a massive amount of CPU time since they contain expensive matrix operations (e.g. pseudoinverse, determinant etc.). Contrary, the HDS and Spath techniques are computationally inexpensive, since they are free from iteration or optimization.

Network	Cov-A	Cov-D	Smax	Shannon	Spath	HDS
C-town	6750	3200	40	20	4	<1
ky2	4000	11800	32	30	6	<1
SN-12	400	100	10	10	3	<1
SN-26	8750	2066	20	14	10	<1

Table 1. CPU time in seconds to calculate optimal measurement layouts for different methods.

Thesis 1

For water distribution networks, the following technique is an objective comparison of different sensor placement layouts for the calibration of pipe roughness values.

1. Fix the number of pressure measurement points: N_{sampled}.
2. Perform a hydraulic analysis with nominal (estimated) friction values and demands. Save all nodal pressure data ($\mathbf{p}_{\text{nom}}$).
3. By means of a sampling layout design technique, choose N_{sampled} sampling nodes.
4. Analyse the effect of initial roughness values via Monte Carlo method, i.e. perturb them by a random number with normal distribution (0 expected value and a prescribed relative standard deviation, e.g. 20%). Perform calibration using the perturbed frictions as initial values. The steps of calibration are the following.
   (a) The sampling nodes, the output values at these nodes and the estimated parameter vector are input values.
   (b) Run a single hydraulic simulation, then calculate the sensitivity matrix ($\mathbf{S}$), that is
   $$S_{i,j} = \frac{\partial p_i}{\partial \mu_j}.$$
   (c) Extract the reduced sensitivity matrix $\mathbf{S}_r$ including only the sensitivity of the sampled pressures, i.e. delete the rows corresponding to the unsampled nodes.
   (d) Compute the pressure difference between the hydraulic model output (obtained with the estimated parameters) and the nominal state
   $$\Delta\mathbf{p} = \mathbf{p}_{\text{mod}} - \mathbf{p}_{\text{nom}}. \tag{6}$$
   (e) Solve the linear system $\mathbf{S}_{\text{r}}\Delta\boldsymbol{\mu} = \Delta\mathbf{p}$ for an update of the roughness coefficient distribution using singular value decomposition.
   (f) Update the parameter vector and repeat from point 2 until $\Delta\mathbf{p}$ is sufficiently small.
5. Evaluate the error, that is the norm of the difference between the nominal and calibrated model pressure at all nodes, mathematically
   $$e_p = \sqrt{\frac{\sum_{i=1}^{N_{\text{nodes}}} (p_{\text{mod},i} - p_{\text{nom},i})^2}{N_{\text{nodes}}}}. \tag{7}$$
6. The above procedure must be repeated until sufficiently large number of results are available for a reliable statistical evaluation of the average and the standard deviation of e_p errors. The recommended number of computations is at least 100.

This analysis can be performed with a successively increasing number of sensors (N_{sampled}), then a plot can be drawn where the horizontal axis represents the number of sensors, and the vertical one depicts the nominal error e_p. Whichever layout technique tends to zero faster is superior to the others.

Related publications: [J2], [C1], [C2], [C3], [C4].

Thesis 2

The Hydraulic Distance with Sensitivity (HDS) technique is capable of determining the sensor placement for the roughness calibration of a mathematical model of a water distribution network ensuring high accuracy with negligible computational time. The HDS method consists of the following steps.

1. Provide an initial guess for the roughness values. If possible, use the available data (e.g. pipe material, age).

2. Designate the intake nodes i.e. reservoirs and tanks as first measured nodes, if they are not equipped with pressure loggers by default.

3. Run a single hydraulic simulation with nominal operational conditions (e.g. average demands), calculate the sensitivity matrix $\mathbf{S}$ and the nodal sensitivities $\mathbf{S}_{node}$, mathematically
$$S_{i,j} = \frac{\partial p_i}{\partial \mu_j} \qquad \text{and} \qquad S_{\text{node},i} = \sum_{j=1}^{N_{\text{edges}}} S_{i,j},$$

4. For each node, compute the hydraulic distance δ_i (the smallest cumulative pressure loss) from the already designated nodes and prescribed pressure nodes (i.e. reservoirs and tanks). Calculate the product of the hydraulic distance δ_i and nodal sensitivity $S_{\text{node},i}$ for each node i.

5. Find the maximum value of $\delta_i S_{\text{node},i}$ and designate the corresponding node as a new sampled node.

6. Repeat from step 4 until the desired number of nodes have been selected.

Related publications: [J2], [C1], [C2], [C3], [C4].

3 Topological analysis of the segment graph

3.1 Introduction

Creating a proper graph representation of WDNs is not straightforward, if the focus is on examining the effect of an accidental pipe burst. The traditional approach, where the pipelines (also pumps, valves) are the edges, while their intersections are the nodes cannot describe the real effect of a pipe failure, since in these cases typically more than one pipeline must be isolated for the depressurization (see top right side of Figure 3). Therefore, a different idea is applied, where *the isolation (ISO) valves are the edges, and the smallest islands which can be segregated are the nodes*. This representation is called the *segment graph* and it was introduced in [27] (although it was called graph of segments). For the sample WDN, see the left bottom side of Figure 3. A single pipe loss can be modelled as a loss of a node in the segment graph as a result that every element of these islands have to be segregated for the maintenance period.

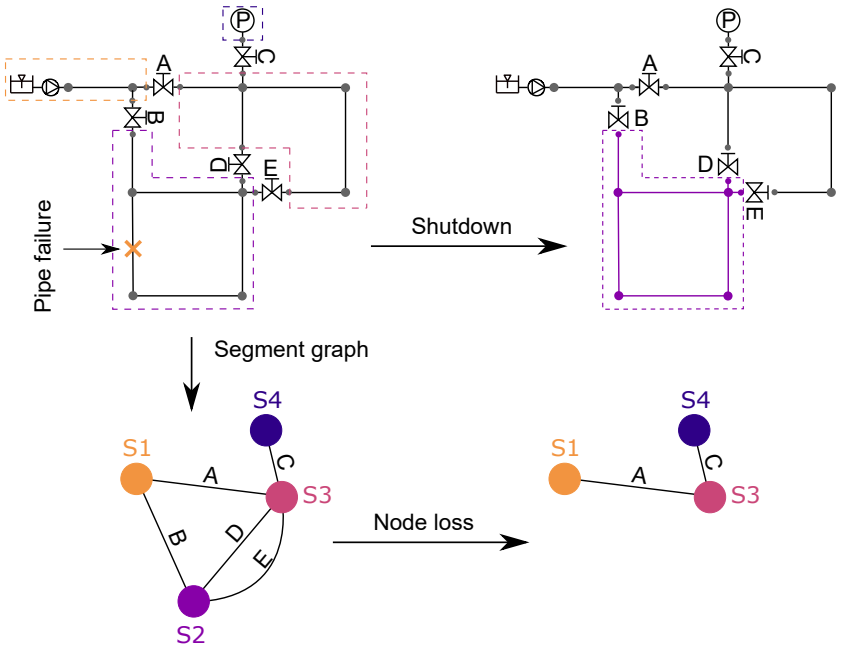

Figure 3. Demonstrating a shutdown using a sample WDN on the top side. The segment graph representation can be observed on the bottom side.

"Traditional" tools of complex network theory is applied to analyse the segment graphs of real-life WDNs. There are two important special graphs in the literature: random graph [28] and scale-free [29]. The former one is considered robust as most nodes have the same number of connections (degree), while the latter one contains "hubs" where edges tend to accumulate, making them vulnerable in terms of connectivity [4]. Traditionally, real networks were assumed to be random graphs; however, in recent years it was revealed that certain real-life networks cannot be accurately modelled as random graph (see [4]). The

goal of the chapter is to determine whether there is *a special graph type that is able to describe the segment graphs* accurately.

3.2 Degree Distribution

In the field of graph theory, the degree of a node is the number of edges connecting trough that node. The degree distribution p_k provides the probability that a randomly selected node in the network has degree k. Since p_k is a probability, it must be a normalized quantity, i.e. $\sum p_k = 1$, see [4]. For the purposes of this study, I am using 27 real-life WDNs from Western-Hungary. The degree distributions of segment graphs of real-life WDNs can be seen in Figure 4 at the left-side, while degree distributions of special graphs at the right side. As it shows the real-life WDNs are more similar to the traditional random graphs rather than scale-free networks. The WDNs do not contain segments with high degree i.e. high isolation valve number that could be critical from a purely topological viewpoint.

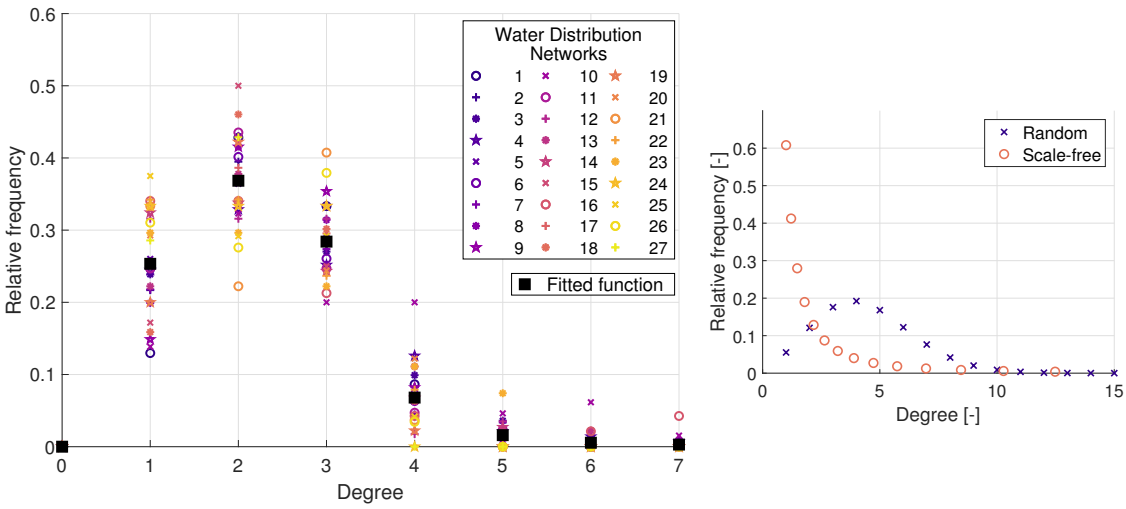

Figure 4. Degree distributions of the segment graphs of 27 real-life WDNs at the left-side, and special graphs at the right side.

3.3 Diameter and clustering coefficient

Beside the degree distribution, the last decades brought numerous topological indicators for quantifying different behaviours of networks (again, I refer to [4]) which are already utilized in the field of WDNs, e.g. [30]. In this section, I am focusing on two of them, namely the *diameter* and the *clustering coefficient*. For the efficient calculation of these quantities, the `igraph` toolbox ([5]) is applied. A path between two distinct nodes in a graph is a sequence of edges. Since typically for two nodes, there are numerous different paths, the shortest path is considered, where the number of edges along the path is the

smallest. In the case of a network, between every pair of nodes, the shortest paths can be determined, and *the length of the longest shortest path is called the diameter of the graph.* For the special graphs, the diameter can be estimated with analytical formulas [31, 4]. The left side of Figure 5 depicts these correlations, and also the data points of real-life WDNs. As it shows, even the prediction of the random graph is slightly underestimates the diameter of real-life WDNs, however the function from scale-free networks is not even in the same magnitude.

The second wide-spread parameter is the *clustering coefficient* of a graph, which characterises the average density of the graph, i.e. closeness to the complete graph (where every distinct pair of nodes is directly connected with a unique edge). This quantity can be defined as *the average number of the connections between the neighbours of an arbitrarily selected node in the network*, see [32]. The clustering coefficient is a number between zero and one, if it equals to one, it means that every neighbouring node is connected directly, if it equals to zero, none of them are connected. For a scale-free network, the clustering coefficient is expected to be high (due to the hubs), while for random networks it is expected to be close to zero. This quantity can also be approximated for special cases [31]. The right-side of Figure 5 presents the results for the approximations with the results from the real-life WDNs. The scale-free estimation is clearly incorrect, while the prediction from the random graph is significantly closer.

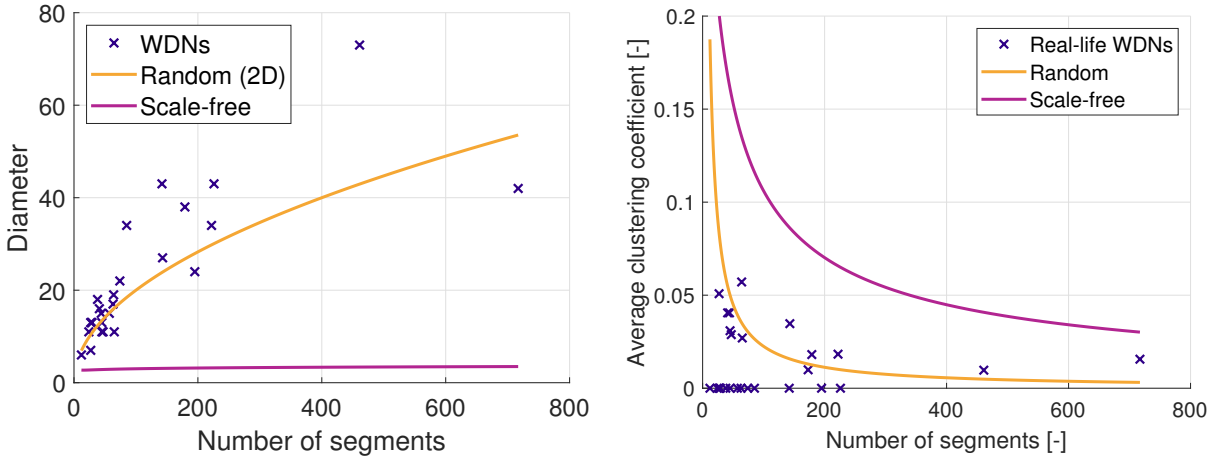

Figure 5. Diameter (left side) and average clustering coefficients (right side) of segment graphs of real-life WDNs compared to random and scale-free network approximations.

3.4 Universal degree distribution function for real-life WDNs

The purpose of the current section is to create a universal function, which is able to approximate the degree distribution of the segment graph of a general real-life WDN. Based on the degree distribution and the structural properties, it seems that the segment graphs of WDNs behave as random graphs; thus, one might fit the Poisson distribution

for describing the degree distribution in general. However, the fact that the topology of a WDN is planar by nature and every node is necessarily connected; thus there are no nodes with zero edges, i.e. there are no segments without ISO valves; these properties make this approach unsuitable. For example, if one would like to describe a WDN using a Poisson distribution with average degree of 2.25, more than 10% of the nodes would have zero degrees.

Due to these difficulties, I have created a general function for the approximation of the degree distributions of segment graphs of WDNs. The function was defined by taking the average of the relative frequencies at each degree, for numerical values, see Thesis 4. The fitted function also satisfies the criterion of the distribution function, that is, the sum of the values is equal to one. This fitted function can be observed in Figure 4 with the real-life WDNs.

Thesis 3

The segment graph of a real-life water distribution network consists of the isolation valves as edges and segregated islands (segments) as nodes. Such graphs show the nature of a planar, connected random graph.

Related publications: [J1], [J3], [J4].

Thesis 4

The segment graph of a real-life water distribution network consists of the isolation valves as edges and segregated islands (segments) as nodes. The discrete distribution function in the Table below describes the degree distribution of the segment graph of common Hungarian water distribution networks with 85% of success rate with a significance level of 95%.

Degree	Rel. freq.
0	0
1	0.2538
2	0.3690
3	0.2846
4	0.0682
5	0.0160
6	0.0055
7	0.0029

Related publications: [J1], [J3], [J4].

4 Vulnerability of water distribution networks

4.1 Definition of local vulnerability

[33] reviewed a few different definitions of vulnerability from the literature, and most of them are based on topological approaches. The goal of the vulnerability, according to the definition in this paper, is to catch the general hydraulic behaviour of WDNs in the case of a single pipe break, thus it is based on the hydraulic model. Most of the vulnerability definitions from the literature belongs either to pipelines or nodes, while this paper is focusing on segments, similar to [34]. As a result, every element from *a segment has the same vulnerability*, since every failure is causing the loss of the whole segment.

The proposed vulnerability is *dimensionless*, thus it can be applied for comparing WDNs with different sizes. First, a *failure rate* (α_i) needs to be determined for each segment, this represents the probability of having a pipe failure in that segment. This quantity must be normalized i.e. $\sum \alpha_i = 1$. Numerous different issues can increase the chance of pipe bursts, e.g. pipe material, pipe age, pH value of the surrounding soil, increased traffic on the surface or construction works around the pipelines [35, 36]. In this dissertation I analyse two different approaches. The first assumption is based on the relative pipeline length found in a segment, i.e. it is assumed that each meter of pipeline has an equal chance to break. During the second estimation, the material of the pipelines are considered with also using the pipe burst statistics from the last 28 years. Second, the amount of drinking water, that cannot be provided in the case of the segregation of the i-th segment, is calculated by the 1D hydraulic solver, i.e. $b_i = \sum d_i - \sum c_i$, where d_i indicates the nominal demand and c_i represents the actual amount of served water according to the model. From the b_i values, the dimensionless β_i values can be introduced, i.e. $\beta_i = b_i / \sum d_i$.

Using the previously defined quantities, the local vulnerability (γ_i) of the network with respect to the i-th segment is

$$\gamma_i = \alpha_i \beta_i. \tag{8}$$

This parameter is dimensionless and segment-specific, moreover it is a product of the failure rate (α_i) and the relative loss (β_i) caused by the i-th segment failure. According to this definition, *the vulnerability is high if a segment has a high probability of a pipe burst, and causes a considerable amount of loss in the provided drinking water in the case of its isolation.*

4.2 Distribution of local vulnerability

The vulnerability analysis was performed for all real-life WDNs from the Sopron Waterworks with both failure rate approximations, here the results with the pipeline material are presented. Although every network was analysed, only 27 of them were large enough for analysing the vulnerability distribution. Figure 6 depicts the sampled probability density function of local vulnerabilities (γ_i, see Equation 8) for 27 real-life WDNs using log-log scale. The points are located along a linear line that implies *power law probability distribution*, which indicates similar behaviour to the *scale-free networks*. Due to the nature

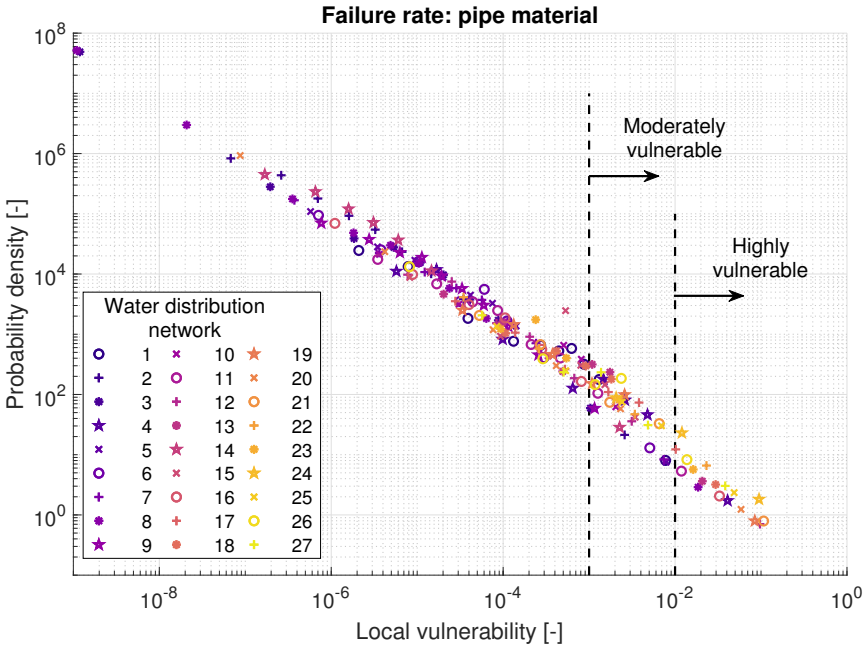

Figure 6. Probability density functions (PDF) of the local vulnerability values of 27 real-life WDNs if the failure rate is based on the material. Systems including segments with higher than 1% vulnerability (e.g. 10% of failure probability and 10% of lost demand) are deemed to be highly vulnerable.

of a scale-free network, it contains critical nodes with high degree (hub), which might be significantly higher than the average, and the loss of a hub could lead to the detachment of the network. The linear trend (that is, power law probability distribution) also highlights that a large part of the network is slightly (or not at all) vulnerable, some regions are moderately vulnerable, but, most importantly, there will be *a few segments that are highly vulnerable*. For example, in the case of SN-8, the probability of a randomly picked segment having larger vulnerability than 10^{-2} (i.e. it is highly vulnerable) is more than 5%, meaning that more than 5% of the segments are highly exposed.

4.3 Network vulnerability

Besides the importance of the local distribution of vulnerabilities inside a network, it is also useful to evaluate the overall quality of a network from the viewpoint of vulnerability for comparison with other WDNs. Thereby, the decision making of the utility company can be supported in the optimal allocation of maintenance and development resources. Therefore, the network vulnerability is introduced, which is the weighted average of the relative demand losses (β)

$$\Gamma = \frac{\sum_i \beta_i \alpha_i}{\sum_i \alpha_i}. \tag{9}$$

This is the weighted average of the local consumption outages, i.e. Γ is *the expected value of the amount of water loss in the case of a single, accidental pipe break* according to the hydraulic model. Since the sum of the failure rate is a normalized quantity, the formula simplifies to

$$\Gamma = \sum_i \beta_i \alpha_i = \sum_i \gamma_i. \tag{10}$$

This means that the *sum of the local vulnerabilities is equal to the network vulnerability.*

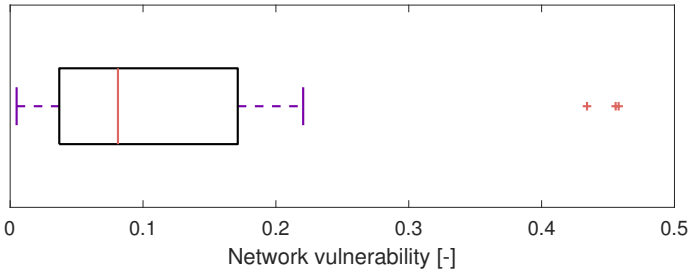

Figure 7. Box plot of network vulnerability (Γ) of the 27 real-life WDNs, revealing 3 highly exposed (outlier) systems.

Thesis 5

The local vulnerability γ_i of a WDN with respect to a segment is defined as the product of the failure probability α_i and the relative loss in the unserved demands β_i during the isolation of the segment: $\gamma_i = \alpha_i \beta_i$. The probability density function of this quantity shows a power law trend.

Related publications: [J1], [J3], [J4].

Thesis 6

The network vulnerability Γ is the weighted average of the relative loss in the unserved demands β_i with the weights being the failure rate α_i. This parameter indicates the expected value in the loss of unfulfilled demands in the case of a random pipe burst in the network. The network vulnerability is the sum of the local vulnerabilities.

Related publications: [J1], [J3], [J4].

Published Papers

Journal articles

[J1] **Richárd Wéber**, Tamás Huzsvár, Dr. Csaba Hős, Vulnerability analysis of water distribution networks to accidental pipe burst, *Water Research*, 184(1), 2020, doi: 10.1016/j.watres.2020.116178, IF(2019): 9.13

[J2] **Richárd Wéber**, Dr. Csaba Hős: Efficient Technique for Pipe Roughness Calibration and Sensor Placement for Water Distribution Systems, *Journal of Water Resources Planning and Management*, 146(1), 2020, doi: 10.1061/(ASCE)WR.1943-5452.0001150, IF(2018): 3.404

[J3] Tamás Huzsvár, **Richárd Wéber**, Dr. Csaba Hős: Analysis of the Segment Graph of Water Distribution Networks. *Periodica Polytechnica Mechanical Engineering*, 63(4):295-300, 2019, doi: 10.3311/PPme.13739

[J4] **Wéber Richárd**, Huzsvár Tamás, Dr. Hős Csaba: Mire lehet még használni egy ivóvízhálózat hidraulikai modelljét?, *MASZESZ HÍRCSATORNA*, (2):35-43, 2020

Conference papers

[C1] **Richárd Wéber**, Dr. Csaba Hős: Comparison of sensor placement strategies on water distribution systems in the aspect of roughness calibration, *11 th Eastern European Young Water Professionals Conference*, Prague, 2019.10.01.-2019.10.05., Water for All, Water for Nature, Reliable Water Supply, Wastewater, Treatment and Reuse, (2019)

[C2] **Richárd Wéber**, Dr. Csaba Hős: Opimal measurement layout on water distribution system, *10 th Eastern European Young Water Professionals Conference*, Zagreb, 2018.05.07.-2018.05.13., New Technologies in Water Sector, (2018)

[C3] **Wéber Richárd**, Dr. Hős Csaba: Optimális mérési elrendezés vízhálózaton, *XXVI. Nemzetközi Gépészeti Találkozó OGÉT*, Marosvásárhely, 2018.04.26.-2018.04.29., (2018)

[C4] **Wéber Richárd**, Dr. Hős Csaba: Hidraulikus alaphálózatok érzékenység vizsgálata, *XXV. Nemzetközi Gépészeti Találkozó OGÉT*, Déva, 2017.04.27.-2017.04.30., Déva, (2017)

References

[1] L. A. Rossman, "EPANET 2: users manual," *Cincinnati US Environmental Protection Agency National Risk Management Research Laboratory*, vol. 38, no. September, p. 200, 2000.

[2] G. Halász, K. Gergely, and L. Kullmann, *Áramlás a csőhálózatokban.* Budapest: Műegyetem Kiadó, 2002.

[3] R. C. Mittal and A. Al-Kurdi, "LU-decomposition and numerical structure for solving large sparse nonsymmetric linear systems," *Computers and Mathematics with Applications*, vol. 43, no. 1-2, pp. 131–155, 2002.

[4] A. L. Barabási, *Network Science.* Cambridge University Press, 2016.

[5] G. Csardi and T. Nepusz, "igraph Reference Manual," p. 812, 2012.

[6] T. Koppel and A. Vassiljev, "Calibration of a model of an operational water distribution system containing pipes of different age," *Advances in Engineering Software*, vol. 40, no. 8, pp. 659–664, 2009.

[7] G. Meirelles, D. Manzi, B. Brentan, T. Goulart, and E. Luvizotto, "Calibration Model for Water Distribution Network Using Pressures Estimated by Artificial Neural Networks," *Water Resources Management*, vol. 31, no. 13, pp. 4339–4351, 2017.

[8] X. Xie, H. Zhang, and D. Hou, "Bayesian Approach for Joint Estimation of Demand and Roughness in Water Distribution Systems," *Journal of Water Resources Planning and Management*, vol. 143, no. 8, p. 04017034, 2017.

[9] A. Khedr, B. Tolson, and S. Ziemann, "Water distribution system calibration: Manual versus optimization-based approach," *Procedia Engineering*, vol. 119, no. 1, pp. 725–733, 2015.

[10] K. Du, R.-y. Ding, Z.-h. Wang, Z.-g. Song, B.-f. Xu, M. Zhou, Y. Bai, and J. Zhang, "Direct Inversion Algorithm for Pipe Resistance Coefficient Calibration of Water Distribution Systems," *Journal of Water Resources Planning and Management*, vol. 144, no. 7, p. 04018027, 2018.

[11] G. Sanz and R. Pérez, "Comparison of demand calibration in water distribution networks using pressure and flow sensors," *Procedia Engineering*, vol. 119, no. 1, pp. 771–780, 2015.

[12] N. C. Do, A. R. Simpson, M. Asce, J. W. Deuerlein, and O. Piller, "Calibration of water demand multipliers in water distribution systems using genetic algorithms," *Journal of Water Resources Planning and Management*, vol. 142, no. 11, p. 4016044, 2016.

[13] Q. Zhang, F. Zheng, H.-f. Duan, Y. Jia, T. Zhang, and X. Guo, "Efficient Numerical Approach for Simultaneous Calibration of Pipe Roughness Coefficients and Nodal Demands for Water Distribution Systems," *Journal of Water Resources Planning and Management*, vol. 144, no. 10, pp. 1–12, 2018.

[14] T. Gao, "Roughness and Demand Estimation in Water Distribution Networks Using Head Loss Adjustment," *Journal of Water Resources Planning and Management*, vol. 143, no. 12, p. 04017070, 2017.

[15] M. Dini and M. Tabesh, "A New Method for Simultaneous Calibration of Demand Pattern and Hazen-Williams Coefficients in Water Distribution Systems," *Water Resources Management*, vol. 28, no. 7, pp. 2021–2034, 2014.

[16] K. N. Mallick, I. Ahmed, K. S. Tickle, and K. Lansey, "Determining Pipe Groupings for Water Distribution Networks," *Journal of Water Resources Planning and Management*, vol. 128, no. April, pp. 130–139, 2002.

[17] J. Francés-Chust, B. M. Brentan, S. Carpitella, J. Izquierdo, and I. Montalvo, "Optimal placement of pressure sensors using fuzzy DEMATEL-based sensor influence," *Water (Switzerland)*, vol. 12, no. 2, 2020.

[18] X. Xie, B. Zeng, and M. Nachabe, "Sampling design for water distribution network chlorine decay calibration," *Urban Water Journal*, vol. 12, no. 3, pp. 190–199, 2015.

[19] Q. Zhang, F. Zheng, Z. Kapelan, D. Savic, G. He, and Y. Ma, "Assessing the global resilience of water quality sensor placement strategies within water distribution systems," *Water Research*, vol. 172, p. 115527, 2020.

[20] O. Giustolisi, A. Simone, and D. B. Laucelli, "A proposal of optimal sampling design using a modularity strategy," *Water Resources Research*, vol. 52, no. 8, pp. 6171–6185, 2016.

[21] F. Soroush and M. J. Abedini, "Optimal selection of number and location of pressure sensors in water distribution systems using geostatistical tools coupled with genetic algorithm," *Journal of Hydroinformatics*, vol. 21, no. 6, pp. 1030–1047, 2019.

[22] C. A. Bush and J. G. Uber, "Sampling Design Methods for Water Distribution Model Calibration," *Journal of Water Resources Planning and Management*, vol. 124, no. 6, pp. 334–344, 1998.

[23] W. de Schaetzen, G. Walters, and D. Savic, "Optimal sampling design for model calibration using shortest path, genetic and entropy algorithms," *Urban Water*, vol. 2, no. 2, pp. 141–152, 2000.

[24] K. Klapcsik, R. Varga, and C. Hos, "Optimal Pressure Measurement Layout Planning in Real-Life Water Distribution Systems," *Perodica Polytechnica Mechanical Engineering*, vol. 62, no. 1, pp. 51–64, 2018.

[25] M. D. Jolly, A. D. Lothes, L. S. Bryson, and L. Ormsbee, "Research database of water distribution system models," *Journal of Water Resources Planning and Management*, vol. 140, no. 4, pp. 410–416, 2014.

[26] H. Hwang and K. Lansey, "Water Distribution System Classification Using System Characteristics and Graph-Theory Metrics," *Journal of Water Resources Planning and Management*, vol. 143, no. 12, p. 04017071, 2017.

[27] T. Walski, "Water distribution valve topology for reliability analysis.pdf," *Reliability Engineering and System Safety*, vol. 42, pp. 21–27, 1993.

[28] P. Erdős and A. Rényi, "On random graphs I.," *Publicationes Mathematicae*, vol. 6, pp. 290–297, 1959.

[29] A. L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," vol. 286, no. October, pp. 509–512, 1999.

[30] A. Yazdani, R. A. Otoo, and P. Jeffrey, "Resilience enhancing expansion strategies for water distribution systems: A network theory approach," *Environmental Modelling and Software*, vol. 26, no. 12, pp. 1574–1582, 2011.

[31] B. Bollobás and O. Riordan, "The diameter of a scale-free random graph," *Combinatorica*, vol. 24, no. 1, pp. 5–34, 2004.

[32] F. Meng, G. Fu, R. Farmani, C. Sweetapple, and D. Butler, "Topological attributes of network resilience: A study in water distribution systems," *Water Research*, vol. 143, pp. 376–386, 2018.

[33] M. Maiolo, D. Pantusa, M. Carini, G. Capano, F. Chiaravalloti, and A. Procopio, "A new vulnerability measure for water distribution network," *Water (Switzerland)*, vol. 10, no. 8, 2018.

[34] N. Abdel-Mottaleb and T. Walski, "Identifying Vulnerable and Critical Water Distribution Segments," pp. 329–339, 2020.

[35] F. Roland, *Ivóvízhálózatok rekonstrukciós stratégiájának kiválasztása térbeli és időbeli meghibásodás modellezéssel*. 2012.

[36] I. Bogárdi and R. Fülöp, "A spatial probabilistic model of pipeline failures," *Periodica Polytechnica Civil Engineering*, vol. 55, no. 2, pp. 161–168, 2011.