UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Ivana Lučin

THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN WATER TRANSPORTATION SYSTEMS

DOCTORAL DISSERTATION

Rijeka, 2022.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Ivana Lučin

THE APPLICATION OF ARTIFICIAL INTELLIGENCE IN WATER TRANSPORTATION SYSTEMS

DOCTORAL DISSERTATION

Thesis Supervisor: Prof. D. Sc. Zoran Čarija
Thesis Co-supervisor: Prof. D. Sc. Siniša Družeta

Rijeka, 2022.

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET

Ivana Lučin

PRIMJENA UMJETNE INTELIGENCIJE U SUSTAVIMA TRANSPORTA VODE

DOKTORSKA DISERTACIJA

Mentor: prof. dr. sc. Zoran Čarija
Komentor: prof. dr. sc. Siniša Družeta

Rijeka, 2022.

Thesis Supervisor: Prof. D. Sc. Zoran Čarija, University of Rijeka, Faculty of Engineering
Thesis Co-supervisor: Prof. D. Sc. Siniša Družeta, University of Rijeka, Faculty of Engineering

This doctoral dissertation was discussed on ________________ at the University of Rijeka, Croatia, Faculty of Engineering in front of the following Evaluation Committee:

1.

2.

3.

Abstract

Water distribution systems are designed to assure safe water transportation to the end-users. Since the water needs to have required quality and hydraulic characteristics, these systems are regularly monitored, controlled, and improved. In this doctoral dissertation, an investigation of different applications of artificial intelligence methods for the purpose of improving water distribution systems was conducted. Firstly, the optimization procedure coupled with numerical simulations is used for improving the design of the parts of the water system intake structure. In the further investigation of optimization applications, pollution detection strategy is developed, where novel optimization approach based on search space reduction method and independent optimizations conducted for each possible source node is proposed. Machine learning has been applied in the prediction of a number of pollution sources, based on a wide range of pollution scenarios with a various number of pollution sources. Additionally, machine learning has been used for leak localization, based on a wide range of leak scenarios. As a further development of leak localization methodology, pipe segmentation approach was proposed in which additional divisions of pipes were introduced to simulate a more realistic scenario where leaks can occur not only at pipe junctions but at any point of pipe. The conducted research showed several new possible utilizations of artificial intelligence methods which were previously not considered mainly due to their considerable computational demand. These applications need to be further explored since with the rapid increase of computational power these methods could provide valuable insight into water system behavior and improve water transportation system operation.

Keywords: water distribution systems, shape optimization, leak localization, pollution localization, Random Forest

Sažetak

Sustavi transporta vode služe za opskrbu različitih korisnika pri čemu je glavna funkcija sustava osiguranje želje kvalitete vode i njenih hidrauličkih karakteristika. Problemi u sustavu mogu uzrokovati značajne gubitke, trajna oštećenja, a u konačnosti mogu predstavljati opasnost za ljudske živote, te se zbog toga sustavi transporta vode redovito prate i reguliraju. S povećanjem količine dostupnih mjerenja kao i s povećanjem računalnih resursa, primjena umjetne inteligencije prilikom dizajniranja i kontrole sustava transporta vode postala je sve zastupljenija. U ovoj doktorskoj disertaciji predloženo je nekoliko novih smjerova primjene umjetne inteligencije u svrhu poboljšanja sustava transporta vode. Prvi od istraženih smjerova je optimizacijski pristup koji je primjenjen za poboljšanje dizajna djelova ulazne strukture sustava transporta vode, konkretno zaštitne rešetke. Primjenom optimizacijskih metoda moguće je prilagoditi geometriju poprečnog presjeka kako bi se minimizirali hidraulički gubitci uz zadovoljenje ekoloških i inženjerskih zahtjeva. Optimizacijski pristup je primjenjen i na problem detekcije mjesta unosa onečišćenja u sustav transporta vode. U slučaju pojave onečišćenja u sustavu potrebno je brzo odrediti lokaciju i parametre onečišćenja, u cilju upozorenja korisnika i poduzimanja potrebnih zaštitnih radnji. Primjenom nove metode koja smanjuje broj potencijalnih čvorova unosa onečišćenja, za svaki preostali potencijalni čvor proveden je zaseban optimizacijski postupak, čime je smanjena dimenzionalnost problema što pojednostavljuje i ubrzava optimizacijski postupak. Nadalje, strojno učenje primjenjeno je za predviđanje nekoliko mogućih lokacija unosa onečišćenja na temelju ograničenih senzorskih mjerenja. Predikcijski model je istreniran na sintetičkim mjerenjima dobivenim iz većeg broja numeričkih simulacija provedenih za varijabilni broj lokacija onečišćenja i za varijabilne parametre unosa onečišćenja. Slična metodologija provedena je i za određivanje mjesta oštećenja cjevovoda, gdje je predikcijski model istreniran na sintetičkim podatcima o izmjerenim tlakovima, koji su dobiveni

iz većeg broja simulacija sa varijabilnim mjestom i veličinom oštećenja. Za razliku od standardne metodologije, u kojoj se pretpostavlja da se oštećenje dogodilo u nekom od čvorova vodovodne mreže, u ovom radu predlaže se novi pristup u kojem se nakon preliminarne lokalizacije oštećenja provodi dodatna segmentacija cijevi kako bi se lokacija oštećenja mogla točnije odrediti. Provedeno istraživanje pokazalo je da se metode umjetne inteligencije danas mogu uspješno primjeniti na probleme koji se prethodno nisu rješavali na ovaj način, ponajviše zbog prevelikih računalnih zahtjeva. Metode predložene u ovom radu pokazuju da se povećanjem računalnih resursa i korištenjem poboljšanih tehnika umjetne inteligencije može poboljšati rad i kontrola sustava transporta vode.

Ključne riječi: sustavi transporta vode, optimizacija oblika, lokalizacija oštećenja, lokalizacija onečišćenja, slučajna šuma

Contents

Part I

Introduction

Chapter 1

Introduction

Water transportation systems are designed for the safe distribution of water from the catchment area to the end-users, e.g. industry or households. Depending on the designated purpose (technical water, drinking water etc.), water needs to have the required quality and hydraulic characteristics. To obtain these attributes, water transportation systems are regularly monitored and controlled to provide optimal system operation [46, 85, 93].

The first step for achieving this goal is the appropriate design of water intake, where trash racks or screens are installed so as to prevent the entrance of debris or fish in the water distribution system [11, 89, 37]. Trash racks or screens can cause additional clogging due to debris accumulation or can reduce system efficiency due to losses caused by flow disturbance, hence they should be optimally designed [73, 33, 98]. Subsequent water treatment can be conducted using filters and water purification systems if water is to be used as drinking water [10, 23]. Although precautions are being taken, unexpected events such as accidental or intentional pollution intrusion in water distribution network can occur [84, 72]. Intrusions of pollution can also occur in pipe leak locations where contaminated soil may enter the pipe under certain conditions [53, 9]. These intrusions can cause serious health problems to the end-users, thus sensors are installed in water distribution networks for water quality monitoring [32, 65]. In the case of an accident event, various mathematical and statistical techniques based on the sensor measurements can be used to identify conditions under which the accident occurred.

With growing technological trends such as Smart Cities and the Internet of Things, the complexity of water distribution systems is continually increasing and consequently available amount of data which can provide valuable insight in system operation. Additionally, improvements to the existing infrastructure are constantly being implemented. These trends indicate a strong need for computational techniques that can process and analyze obtained data and consequently provide engineers with useful knowledge which can enhance water transportation systems operation.

1.1 Accident events

1.1.1 Structural failures

Structural failures in water transport systems can be caused by various factors, e.g. vibrations due to fluid-structure interaction [44, 80, 86], impact of large debris collision with trash-racks, or smaller debris entering water distribution systems and causing damage to system parts [17, 21]. For this reason, the geometry of intake structures needs to be carefully considered to provide good protection and sustain debris load while at the same time produce minimal disturbance of fluid flow. This presents an optimization problem since opposing goals need to be satisfied [14].

Additionally, structural failures can occur due to material deterioration [42, 67, 15]. Various factors can influence pipe material deterioration, such as temperature variation, soil influence, age, stress due to pressure changes, etc. Due to expensive installation, water distribution network pipes are not regularly substituted, which can cause corrosion and cracks in the material over the long term. The problem is that when a leak occurs, it can precipitate material deterioration, which can ultimately lead to pipe bursts. Pipe bursts are the greatest problem since they cause serious damage and losses, thus a number of papers have considered methods for predicting pipe failures [78, 22, 7, 87, 19]. A greater pipe burst can cause flooding of a populated or industrialized area, which can cause considerable material losses, while the consequent water supply outage can last for an extended period of time. Therefore, different methods for detecting and localizing leak and burst locations have been explored [16, 88, 45, 35].

1.1.2 Water quality failure

Water quality failures can occur because of deliberate contamination injection in water distribution networks or due to accident events. Pollution intrusion through leak location is serious issue, since the soil at the location of the leak may be contaminated with harmful micro-organisms and pathogens [41, 39, 9, 24]. Additionally, pipe corrosion can lead to reduced water quality, i.e. occurrence of ”red water” [54]. In the case of these incidents, it is of utmost importance to rapidly determine the location of intrusion, starting time, duration of intrusion, and contamination concentration. With these parameters identified, simulations of pollution spreading through water distribution networks can be conducted. Simulation of contamination spreading identifies which parts of the water distribution network, and in what amount, have been contaminated. On the basis of this knowledge, required actions can be taken, such as prevention of further contamination and warning of users in the contaminated area.

The presented problem is difficult to solve since it is an inverse problem, where based on sensor measurements, different analytical techniques need to be employed to determine causal factors of the event. For solving this problem optimization methods coupled with numerical simulations are predominantly used, which is known as the simulation-optimization approach [62, 63, 2, 91, 92, 81, 90]. In this approach, optimization methods try to find contamination parameters for the numerical simulation of pollution scenario which will produce results most similar to the sensor measurements obtained from the real event. In recent years machine learning approach is also often employed, where machine learning algorithms for prediction of pollution parameters are trained on a wide range of simulated contamination scenarios [51, 28, 29, 30].

1.2 Numerical simulations in water transportation systems

Due to large scale and great complexity of water systems, model testing is usually not the feasible approach for analyzing existing systems or investigating possible improvements. In-field measurements usually provide only limited information, thus these measurements are typically used for calibration of the numerical model, which is then used for obtaining detailed information regarding fluid flow. Therefore, numerical simulations are increasingly being used to enhance the existing design, to provide better insight into system behavior under different conditions, or for designing a new infrastructure. Numerical simulations can be used for one-dimensional, two-dimensional, and three-dimensional fluid flow analysis, depending on the considered problem. One-dimensional simulation represent flows through long pipes, preferably circular pipes, which are mostly used for water distribution systems. The most widely used software for this purpose is EPANET, a public domain software developed by US Environmental Protection Agency [71]. This software enables fast simulation of system behavior even for most complex networks. However, this speed is due to considerable simplifications of its flow model, such as the assumption of complete mixing at junctions, the assumption of constant pressure and velocity values along a pipe, etc. However, wrong results can be obtained as a result of these simplifications. Thus, two-dimensional and three-dimensional fluid flow analysis is conducted when more detailed and more precise information regarding fluid flow is needed. For example, pressure and velocity distribution at intake structures, identification of recirculation zones that occur due to trash-rack and screen installation, mixing at junctions, etc.

Numerical simulations are the basis for using artificial intelligence methods such as optimization algorithm or machine learning algorithms. Optimization methods can be used to enhance existing system design if system design variations are evaluated by numerical simulations of system behavior. Optimal sensor placements can be determined if synthetic measurements obtained from the simulations are compared for various sensor layouts. Fault events can be detected by optimizing simulation parameters of accident events by comparing simulation results to observed values. With the increasing amount of sensor data, machine learning algorithms, which are based on finding patterns and underlying correlations in the data, can be used to extract meaningful information from sensor measurements, which can then provide better insight into existing systems and enhance its monitoring. With increasing computational resources, the importance and applications of these methods are growing, especially machine learning methods. In this thesis, multiple novel areas of applications of proposed methods in water systems are presented, and further research possibilities are proposed.

1.3 Organization of the thesis

In the presented thesis, several applications of optimization and machine learning techniques are presented for known problems in water transportation systems, namely reduction of hydraulic losses, pollution localization, and leak localization. For each of these directions, limitations of conducted research are discussed and future work is proposed. The thesis is organized as follows.

In the second chapter, an overview of optimization methods used for water transportation systems analysis is presented. Numerical and optimization approaches are applied for the evaluation of novel trash-rack geometry designs. Previous research of improved designs was mostly relying on experimental testing, whereby the proposed artificial intelligence based methods provide an investigation of novel designs which were previously not considered in experimental testing. The future application of the proposed approach is discussed. Additionally, in the same chapter, the application of optimization techniques for the detection of the pollution source is presented. It is known that the pollution localization problem is very complex due to the categorical variable which represents the source node. Additionally, it is a multi-modal problem, since multiple equally good solutions can exist. Therefore, a novel approach that reduces the search space of the considered problem is presented, with the addition of the novel optimization approach. In this optimization approach, separate optimizations are conducted for each suspect source node that survived the search space reduction. This reduces the dimensionality and complexity of the considered problem, since it eliminates the categorical variable which is also the most problematic from the optimization standpoint. It also enables obtaining the best solution for each potential source node, i.e. deals with the multi-modality of the problem. Limitations of the proposed method are mentioned, and future work is proposed.

In the third chapter, an overview of machine learning methods used in water transportation systems is presented. A novel machine learning approach based on a Random Forest classifier employed for the prediction of a limited number of pollution sources in water distribution network in the case of accident event is presented. Since the efficiency of search space reduction methods and optimization methods depends on the number of pollution injection locations, it is greatly beneficial to have information about the number of pollution sources. Assumptions used for the proposed method are explained, and further areas of investigation are explored. Machine learning application was also used for the identification of possible leak locations. Random Forest classifier is employed to detect source location based on large number of synthetic pressure data obtained from the leak simulations conducted for various node demands and leak sizes. As the further improvement of the proposed approach pipe segmentation is introduced. Since the leaks can occur anywhere in the pipe segment, the proposed approach simulates a more realistic case, where prediction model accuracy is estimated for leaks that can occur anywhere in pipe segment, not only in network nodes. Further development of the proposed idea is also briefly mentioned. In the fourth chapter, a conclusion is provided where a summary of main contributions is presented and a proposal of future research is given.

Chapter 2

Optimization methods in water transportation systems

Optimization methods are aimed at finding the best solution for a defined goal function, which can be formulated as a maximization or minimization problem. During the iterative process of the optimization method, values of input variables are varied, typically within a predefined range, until stopping criteria are satisfied. The most straightforward optimization problem in water transportation systems is finding optimal geometry design for defined criteria. For example, the design of trash-racks or screens within provided limits to assure needed blockage while producing minimial hydraulic losses. In this case, an optimization algorithm is used to iteratively provide values for chosen geometry parameters, e.g. bar width, bar spacing, and bar length, and numerical simulations are conducted to evaluate fitness function value, i.e. define hydraulic losses for the chosen design. An increase in the number of considered geometry parameters increases geometry flexibility and enables finding improved designs; however, a greater number of optimization parameters considerably widen the search space. Therefore, enhanced optimization methods need to be used to reduce the probability of obtaining only local optima and to increase convergence speed, since numerical simulations for evaluation of structure design usually require 2D or 3D numerical simulations which are often computationally quite expensive.

Another type of optimization problem in water transport systems are inverse problems. In case of an incident event, it is of main importance to find conditions that caused the incident. Optimization techniques and numerical simulations are jointly used where incident event parameters used for simulation are changed by optimization algorithm, with results obtained by simulation being compared to the true sensor measurements. Parameters that provide the best agreement between computed ”measurements” from the simulated scenario and real sensor measurements are the optimal solution of the optimization problem solved. This is considered as an inverse problem since based on the recorded output (sensor measurements) the algorithm tries to find optimal input (incident event parameters). Detection of pollution event parameters is considered as an inverse type of problem, where based on sensor measurements pollution source location and pollution parameters need to be obtained. The main problem is the node variable which is the categorical variable that makes considered optimization problem of the most complex type since the mixture of continuous and categorical values is present in the optimization problem. Additional problem is that in case of pollution event rapid reaction time is needed to minimize the harm for end-users, thus fast optimization methods are needed.

Additional optimization problems in water distribution networks include optimal sensor placement, calibration of numerical models based on in-field measurements, optimization of pipe diameters and lengths for new infrastructure, etc. However, these optimization problems will not be covered in this thesis.

2.1 Optimization of hydraulic element design

To determine the geometry of various infrastructure segments engineering practice and model testings are mostly used. For example protection racks or trash-racks can be chosen in accordance with known guidelines regarding hydraulic losses [36]. However, with growing ecological concern, it is observed that classic designs can cause injuries and increase mortality of fish species [5, 37]. Therefore, novel designs are being considered where turbulence zones are being induced, which fish naturally avoid. However, these novel designs, which usually consist of angled trash racks and angled bars, increase hydraulic losses, thus a compromise between engineering and ecological concern needs to be made. In recent years, model testing is being conducted to investigate different trash-rack and screen designs, bar spacing, length and inclination [95, 97, 68, 69, 4, 96]. However, cross-sections are usually kept rectangular or with a streamlined shape. Only recently novel cross-section designs are being considered [8, 55]. The main limitation of model testing is that designs used in investigation need to be defined before testing, which makes it difficult to pinpoint the true optimal design. Additionally, in model testing intake design is defined by testing facility equipment, which is usually a straight channel and unique specifics of each intake cannot be considered. This is a considerable simplification, since it is reasonable to believe that each intake has a unique optimal design that ideally needs to be defined. Papers dealing with numerical simulations of intake structures are rather sparse [70, 59, 3, 40]. In all of these papers, a limited number of pre-defined designs was investigated, thus coupling numerical simulations with optimization techniques would enable exploration of these new, innovative cross-sections specifically designed for the considered intake.

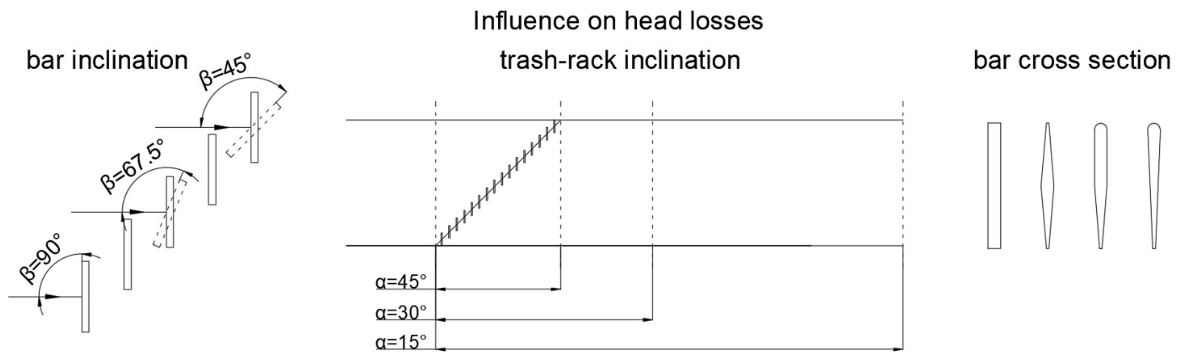

Figure 2.1: Numerically investigated bar inclinations, trash-rack inclinations, and bar cross sections [48].

In this thesis, numerical analysis was conducted on four different types of bar cross-sections: rhombus, rectangular, rounded front edge with inclined back in the lower half, and rounded front edge with inclination starting after rounded edge, for different bar and trash rack inclinations (Figure 2.1). Numerical analysis was conducted in the ANSYS Fluent using 2D $k-\epsilon$ turbulent flow model. The validation of numerical setup was made by comparing with experimental results obtained from [4] for different bar inclinations (45°, 67.5°, and 90°) and trash-rack inclinations (15°, 30°, and 45°). It was observed that depending on thrash rack and bar inclination angles different cross-sections produce an optimal solution, i.e. yield smallest head losses. This indicated that optimization of bar cross-section could be beneficial for specific intake geometry. Therefore, an optimization procedure using Particle Swarm Optimization (PSO) was conducted for a real intake structure of Hydroelectric power plant Senj, Croatia, where three different bar cross-sections were considered: with all rounded edges, with all inclined edges, and design with rounded front edges and inclined back edges. In the considered case, fluid flow was adjacent to the bars, so the optimization for all considered cases converged in cross-section with the lowest cross-section area since it produced the smallest disturbance in fluid flow (Figure 2.2).

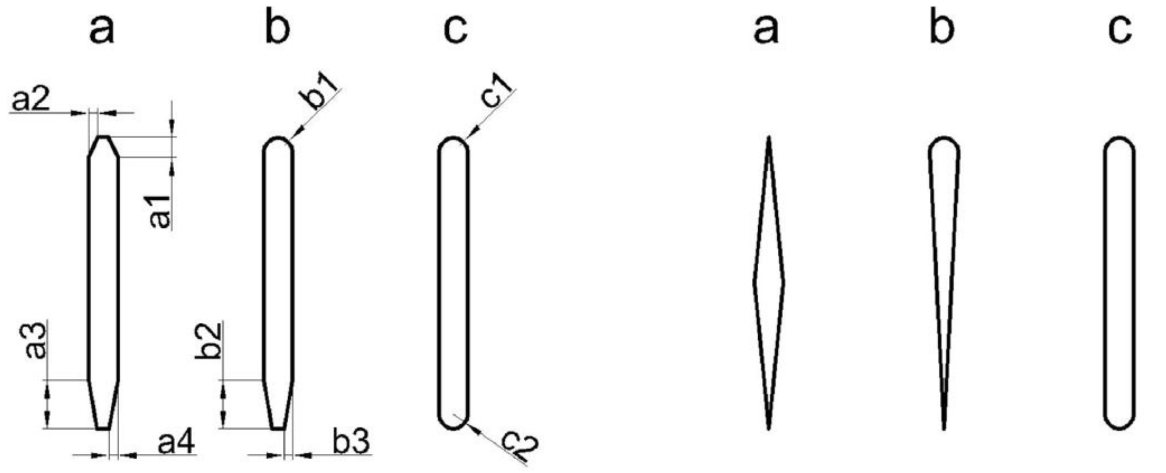

Figure 2.2: Investigated cross sections for optimization approach (left) with final optimal design (right) [48].

The obtained results give a strong indication that further investigation of numerical and optimization methods needs to be conducted. It must be noted that in the current optimization procedure parametric shape was considered which is often the case in optimization methods since it provides a smaller search space due to the smaller number of optimization variables. However, it limits the optimization possibilities since innovative designs, such as curved bars [8] cannot be obtained. However, with growing computational power, optimization procedures in which cross-sections are described with the set of points where coordinate values are optimization variables could enable obtaining more innovative designs such as curved bars with irregular cross-sections. Although such shapes could be hydraulically more efficient, construction constraints must be also taken into consideration such as the strength of the material, manufacturing demands, method of cleaning of the trash-rack structures, etc. These types of problems need to satisfy multiple constraints and have multiple goals which are often opposed, therefore coupling of computational fluid dynamics (CFD) and structural

(FEM) numerical analysis with optimization methods can provide improved solutions while considering all present limitations.

2.2 Pollution localization with optimization - simulation approach

Monitoring of complex systems such as water transportation systems can be a challenging task due to various accident events that can occur. Accidental or intentional pollution intrusions need a rapid reaction, where identification of pollution source, contamination concentration, starting time, and duration need to be identified as soon as possible. Based on this information, simulation of pollution spreading can be conducted so polluted water distribution network areas can be identified, warnings can be given to the affected users, and the source of pollution can be eliminated. Since the reaction time is of main importance, different methods and techniques are being used to simplify the considered problem and narrow down the search space to provide faster response time [20, 43, 64]. When using stochastic optimization methods, such as PSO, multiple optimization runs are needed. Keeping in mind the necessity of rapid intervention, multiple runs of optimization cycles can be time extensive and still do not assure obtaining the optimal solution. This is especially important when discrepancy in sensor measurements [63] and water demand uncertainties [81, 90] are included. Additionally, it is known that the considered problem is a multimodal, i.e. multiple solutions exist. As a solution to this, niching algorithm has been proposed in work by [34, 91] where during the optimization run the best solution for each network node is stored in its niche. This approach produces multiple solutions from a single optimization run.

In this thesis, a search space reduction technique has been proposed in which, prior to conducting the optimization process, preliminary evaluation of possible source nodes is conducted. For each network node, an unrealistically high contamination value was injected during the entire simulation time. Simulations were conducted using EPANET2 software. If sensors did not register contamination for this extreme case it is concluded that they would not be able to detect contamination for any other, less severe, contamination scenario parameters (e.g. smaller pollution concentration value, shorter injection time, etc.), thus these nodes are excluded from the search space. In this way, a considerable percentage of network nodes can be eliminated before commencing the optimization procedure. The main benefit of this forward approach is that number of needed simulations is equal to the number of network nodes and the execution of these simulations can be run in parallel. The proposed approach is not computationally demanding but can provide considerable search space reduction. An extensive investigation of the proposed method was conducted for 5 different sized benchmark water distribution networks which are shown in Figure 2.3. Details of considered networks and sensor layouts can be found in Table 2.1.

Table 2.1: Overview of investigated networks and sensor layouts in [52].

Network	No. of network nodes	Simulation time	Sensor placement
Anytown	19	24 h	70, 160
			90, 110, 140
Net3	92	24 h	117, 143, 181, 213
			115, 119, 187, 209
			113, 120, 147, 211
			117, 149, 167, 213, 253
			117, 173
Richmond	865	72 h	123, 219, 305, 393, 589
			93, 352, 428, 600, 672
BWSN Network 1	126	96 h	10, 31, 45, 83, 118
			10, 83
BWSN Network 2	12523	48 h	871, 1334, ..., 11519

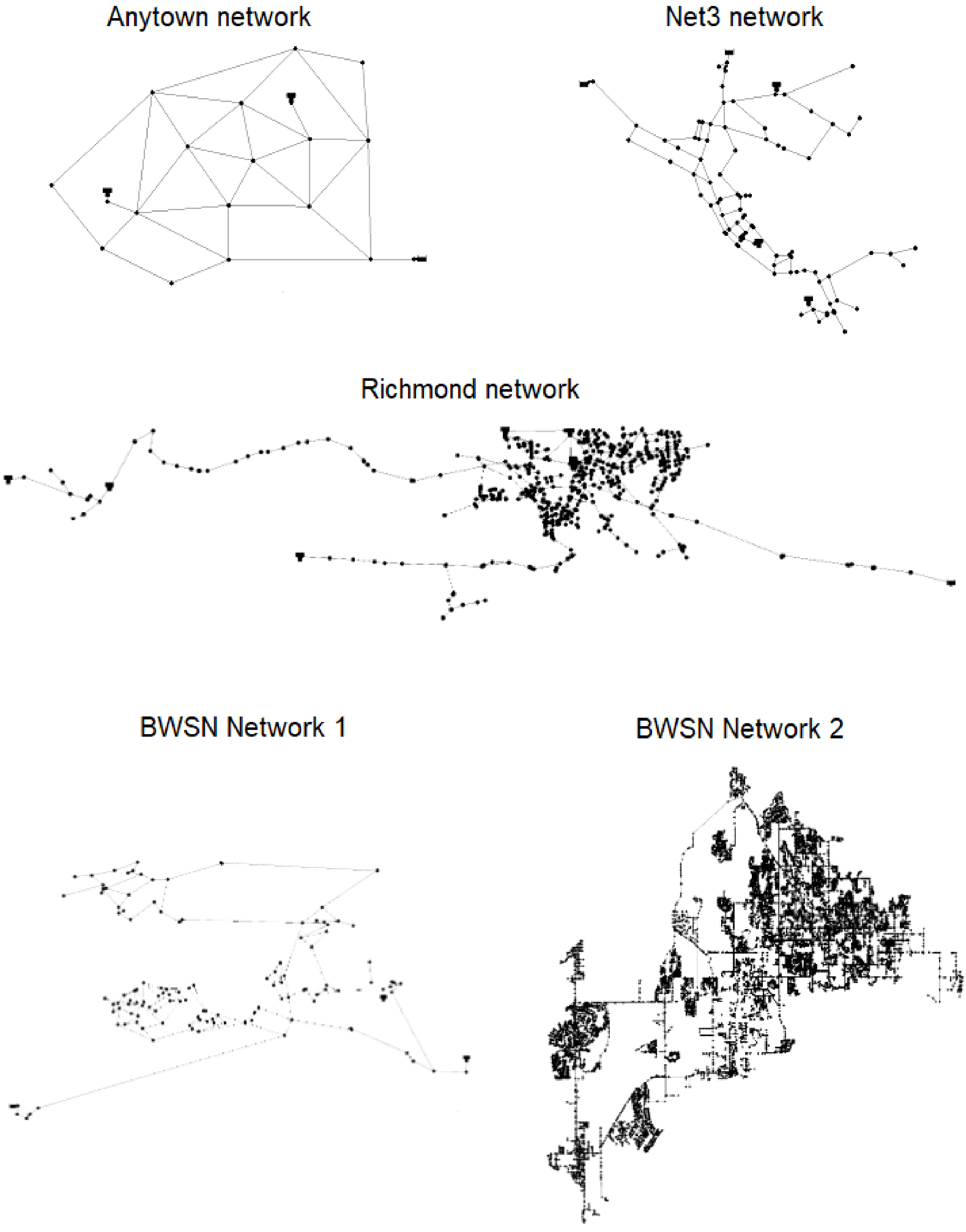

Figure 2.3: Networks investigated in [52] for search space reduction method.

Additionally, two different approaches were considered, one in which multiple injection locations are possible and one in which only a single pollution source is considered. If multiple sources of pollution are considered, network nodes were eliminated only if all sensors did not record pollution in the extreme scenario. If only a single source of pollution is considered, the condition was that exactly all sensors that detected pollution in the real pollution event must detect pollution in the case of an extreme scenario. Greater reduction of network nodes was obtained for the approach for a single source of pollution, however, the assumption of only one pollution source can cause wrong results if multiple sources are present.

Based on these findings, it was observed that if a reasonable number of suspect nodes remained, independent optimization procedures for each suspect node can be conducted in reasonable time. In this way, for each optimization the remaining optimization variables are pollution injection starting time, injection duration, and concentration value. This considerably reduces optimization complexity since the most problematic categorical variable (pollution source location) is eliminated. Additionally, the proposed approach eliminates the problem of obtaining only local optima. The niching algorithm [34, 91] also provides multiple solutions, but the advantage of the approach proposed in this thesis is that for each injection the optimum fitness is obtained, since independent optimization runs performed for each injection node. In the case of the niching algorithm, it is expected that the optimization algorithm investigates the most in the vicinity of the optimal solution, where it is not given that obtained solutions from other niches are the best solutions that can be obtained for these source nodes.

The performance of the proposed search space reduction method should be investigated under demand uncertainties and sensor measurement imperfection, which is a more realistic case. The research conducted in this thesis, and many other pollution detection techniques and methods found in literature, are tested with the measurements from an extended period. However, in a real case scenario, methods for pollution detection will be utilized from the first positive sensor detection. Therefore, the proposed approach should be evaluated for more realistic conditions. The independent optimization approach was further explored in work by [30] where based on machine learning prediction several most suspect nodes are detected and are then used in independent optimization procedures. However, these independent optimizations assume a single injection location, where multiple injection locations should be also explored.

Chapter 3

Machine learning in water transportation systems

Machine learning algorithms are designed to find underlying patterns and correlations in the data. Their best performance is achieved when a big amount of data is available but at the expense of increased model complexity. With recent advances in technology, the competence of machine learning models is increasing for various problems which were previously not considered for machine learning application. Machine learning models can be used as a substitute for extensive experimental or numerical testing when a considerable number of parameters is investigated. For example, when a large number of parameters need to be considered, such as pipe diameters, pipe lengths, temperature, water quality parameters, etc., a substantial number of experiments or simulations would be required to obtain valid conclusions about the investigated phenomena. However, if a limited but significant number of experiments or simulations are conducted, a machine learning model can be used to predict the desired output variable for various input parameter combinations which were not evaluated through the experiment. Additionally, machine learning methods are also often used for finding anomalies in sensor readings which can indicate accident events such as contamination occurrence or pipe bursts. Machine learning was previously used for evaluation of mixing in double pipe junctions [27], water quality monitoring [56, 26], prediction of possible sources of pollution intrusion [83, 29, 28, 30], anomaly detection [82], prediction of pipe failures [74, 25], leakage localization [47, 13], detection of cyber-attacks [58, 1] etc.

In case of accident events in which reaction time is the most important, the main advantage of machine learning methods over optimization methods is considerable reduction in needed computational time. In the case of machine learning applications, the majority of computational time is used for data preparation and model training. This can be conducted prior to the accident event, so later the prepared prediction model can be used with only a small computational effort needed. However, the main disadvantage of the machine learning approach is that real conditions of water distribution networks in the case of the accident event cannot be known, and can be considerably different from those used for the construction of the machine learning model. This is especially important for water network demands which can vary considerably on a daily or hourly basis, where for the machine learning approach an estimate of system behavior is used. Therefore, the advantage of the optimization approach is that it is prepared after the accident event is observed, thus calibrated water distribution network model, based on observed sensor measurements during the accident event, can be used.

It should be noted that the recorded accident events in water distribution networks are rather rare, therefore the data for prediction model training is limited. However, simulations of the wide range of different conditions can be utilized to consider various uncertainties that can occur to gather a considerable amount of synthetic data. Ultimately, multiple prediction models can be prepared for utilization in the case of an accident event. In this thesis, machine learning algorithms have been used to predict the number of pollution locations and to determine possible leak locations based on sensor measurements.

3.1 Machine learning application for determination of number of pollution sources

Based on a larger number of pollution simulations with various pollution parameters, machine learning models can be used to provide the most probable pollution source with the prediction of injection time, injection duration and contamination concentration [28, 29]. After a machine learning based source localization, further finer determination of pollution parameters can be conducted with optimization techniques [30]. The main problem is that the majority of both optimization and machine learning techniques in the literature only consider a single pollution location. In [83] Bayesian approach coupled with Support Vector Regression was used for the probability distribution of possible contaminant sources with the assumption of a single injection location. In [75] the efficiencies of the Bayesian probability-based method, backtracking method, and optimization-based method were evaluated, where it was noted that the Bayesian method was designed only for a single contamination location. Machine learning predictions of pollution scenario parameters based on Random Forest algorithm [28, 29, 30] all assume a single pollution location. As mentioned in the previous chapter, the pollution localization technique proposed in [52] showed better pollution localization in the case of a single pollution source, although, it can be extended to multiple injection sources. If several pollution locations are assumed, optimization variables need to be assured for additional source locations which considerably increases the search space. If these variables are ultimately not needed, the reaction time is prolonged due to an unnecessary increase in problem complexity.

In this thesis, the machine learning approach is presented where the number of contamination sources is predicted based on a large number of simulations for various pollution scenario parameters. A considerable number of pollution scenarios were generated with a randomly chosen number of pollution sources, concentration value, injection time, and duration time. The number of injection locations varied from 1 to 4, and investigated networks were Net3 and Richmond network. When multiple injection locations are chosen, simplification was made and the same pollution parameters (injection time, duration, and concentration value) were used for all locations. Similarly as in [52], it was observed that random pollution parameters can cause pollution scenario which is undetected by all sensors in the water distribution network, thus the prior check of prepared data is conducted. Network nodes for which pollution is not detected are eliminated, and simulation with multiple locations was conducted only for source nodes that contribute to pollution readings. The example of conducted data preparation can be observed in Figure 3.1 where 3 different source nodes were randomly chosen; however, ultimately only 2 network nodes were considered since for the one source node (node 151) sensors did not detect contamination.

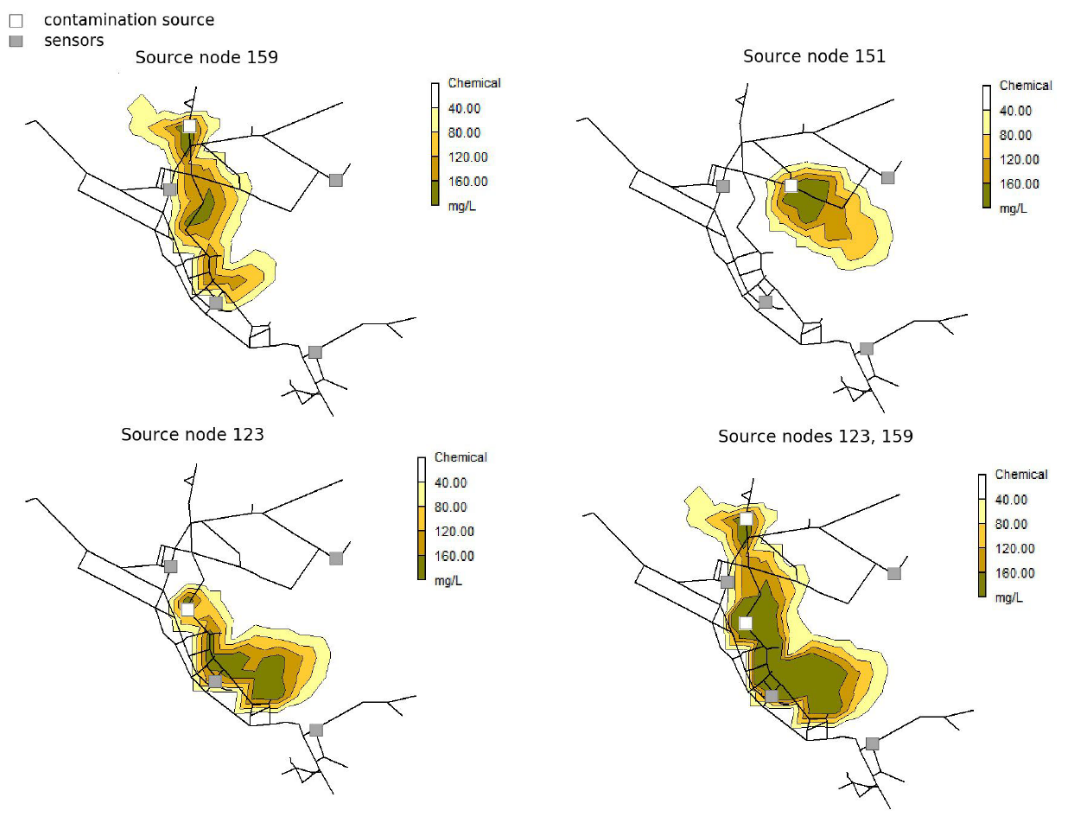

Figure 3.1: Example of single and multiple sources pollution scenarios of Net3 considered in [51].

Random Forest classifier implementation in the Python library Scikit-learn [60] is used to determine the number of pollution sources. This is important since the number of optimization variables needs to be defined before the optimization process and some space reduction techniques have better performance in the case of the single pollution source. Based on the right prediction of the number of contamination sources only needed optimization variables can be used for the optimization problem. It was observed that good accuracy can be obtained when an exact number of sources is predicted, however considerable improvement is obtained when it is predicted if single or multiple numbers of sources are present. This information is also important since some pollution localization techniques can be used only in the case of a single source or are more efficient when that is assumed, as was shown in [52].

In the case of the proposed machine learning approach, a considerable assumption was made where all sources of pollution had the same parameters. This could be realistic in the case of simultaneous intentional intrusions; however, even then exactly equal parameters are hard to expect. Therefore, further investigation should be conducted for the evaluation of the number of pollution sources but with various pollution scenario parameters. Additionally, due to a large amount of considered data, especially when uncertainties are incorporated, data reduction techniques should be explored and other machine learning algorithms which could provide better model accuracy. It must be noted that the reaction time is most important in this case, thus all possible improvements that could reduce the computational time should be explored.

3.2 Machine learning application for leak localization

The presence of leak locations in water distribution networks can be a considerable problem due to substantial water losses. The main problems are small leaks that can be often hard to detect, and over time can cause considerable cumulative water losses. Additionally, for some types of terrain, water is absorbed in the soil, thus leak presence is not evident on the surface and even greater leaks can remain undetected for longer periods. Leak locations can also be hazardous due to possible contamination intrusion from surrounding contaminated soil. Therefore, the detection, localization, and repair of even small leaks is important. Usual methods for leak localization consider hardware-based methods and software-based methods. Hardware-based methods use in-situ measurements, for example, infrared thermography, acoustic methods, or ground-penetrating radar. The main problem is that these methods require an experienced operator and are time and money-consuming. Software-based methods use various software for simulation of water systems or analysis of measured data from sensor measurements. The main problem is that many software based methods rely on residual-based analysis where pressure sensor measurements are compared to expected (predicted) values, which can considerably differ from true pressure measurements.Thus, these methods can have a considerable number of false positive predictions since an unexpected surge in water demand can be interpreted as leak occurrence. If numerical simulations are being conducted for the evaluation of expected system behavior, the numerical model needs to be a good representative of the real network, which is often the problem, due to various uncertainties. For example, the accumulation of corrosion byproducts and suspended particles with time can cause a reduction in pipe diameter and can change pipe roughness. These values can be calibrated with in-field measurements; however, considerable estimations still remain present such as unknown valve opening status. An overview of some of leak detection and localization methods, with their advantages and limitations can be found in [18, 31, 6, 94, 38, 88, 12, 94].

Recently different machine learning algorithms have been used for leak detection and localization such as principal component analysis (PCA) [66], convolutional neural network (CNN) [99], artificial neural network (ANN) [61, 57], k-nearest neighbours [76], Bayesian classifier [77], deep learning [99], linear discriminant analysis (LDA) and neural network classifiers [79]. However, the main problem is a sparse number of data for actual leak and burst events, which is the requirement for high prediction model accuracy. However, a considerable amount of data can be obtained if simulations are conducted with the variation of leak location, leak size, and node demands. The idea is similar as in [29] and [51] where a large number of simulations for pollution scenarios were conducted. Additionally, uncertainties can also be addressed by simulating leak scenarios under various conditions, such as node demands, pipe diameters, etc.

For the leak localization using big data Random Forest algorithm implementation in the Python library Scikit-learn [60] was employed. Two different-sized networks were considered, Hanoi (Figure 3.2) and Net3 network. An overview of considered networks, sensor placements, and simulation parameters is given in Table 3.1. The prediction model accuracy was assessed for different base node variations, leak sizes, and sensor layouts. It was observed that better prediction accuracy can be obtained for greater leaks, which was expected since they cause greater disturbance in the pressure measurements. Additionally, with the increase of network complexity and a wider range of demand uncertainties, a possible number of leak scenarios rapidly increases, thus prediction model accuracy for the same number of training data is reduced. Although for these cases true leak node is often not detected, if several top nodes with the greatest prediction model certainty are considered, a significant leak localization can be obtained. Additionally, the prediction model can be prepared for a specific range, such as nighttime when there are smaller water demands and with that smaller demand variations. It must be noted that considerable simplification is made, since the proposed classification method is trained, and thus can only predict leak locations that are network nodes. In reality, this is not the case since the leak can occur anywhere in the pipe segment. These issues are further investigated.

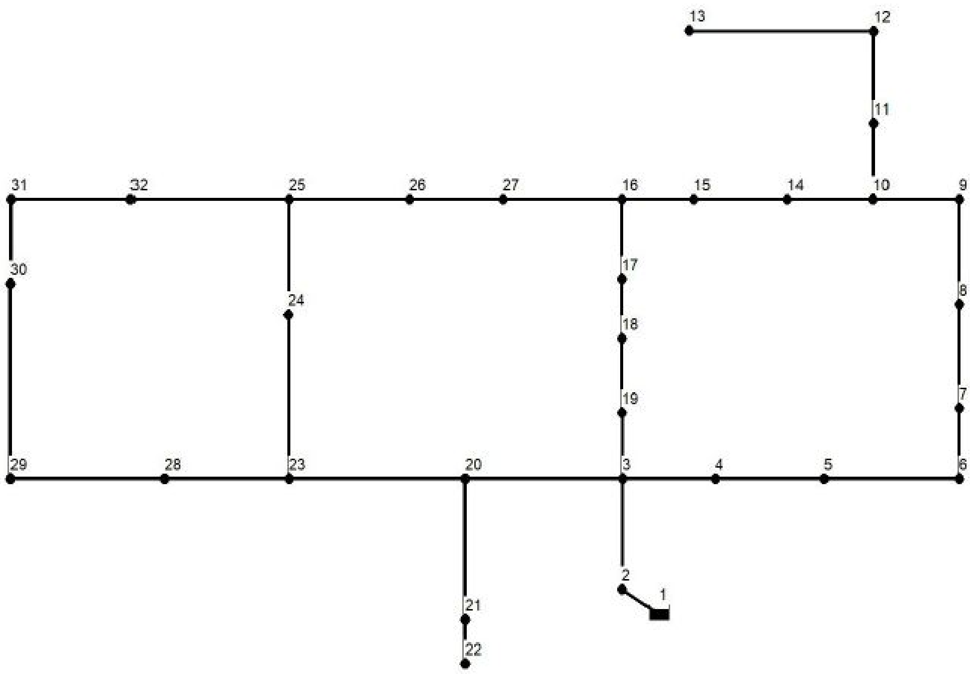

Figure 3.2: Hanoi network.

Table 3.1: Overview of investigated networks and sensor layouts in [49].

Network	No. of network nodes	Sensor placement	Simulation time
Hanoi	31	14, 30	24 h
Net3	92	117, 143, 181, 213	24 h
		115, 119, 187, 209	
		117,181	
		119,209	

Further investigation was conducted for the pipe segmentation approach. Each pipe is divided into additional segments, where three different machine learning frameworks are considered and compared. The first model was trained and tested on leak scenarios with leak locations in original network nodes, the same approach as in [49]. In the second approach, the model was trained on all original and segmentation nodes and the prediction of leak node is original or segmentation node. The third model was trained on original network nodes, and prediction is made for scenarios with leak locations in both original and segmentation nodes. Since the number of classes, in this case, is the same as the number of network nodes, segmentation nodes are associated with their nearest original network nodes. If a prediction of that nearest original network node is made, it is considered as the correct prediction for the segmentation node. It was observed that the second approach considerably increases computational demands and as such is currently not a feasible approach for larger networks. However, the last approach simulates the most realistic case and as such can successfully localize the area of leak location if several top nodes with the greatest prediction model certainty are considered.

To further localize the leak location, sequential prediction models are used. After the initial leak localization is made, pipe segmentation around most suspect nodes is conducted (Figure 3.3). Sequential prediction models were trained with simulations conducted with possible leak locations in the original most suspect network nodes and segmentation nodes. However, it was observed that machine learning models have a problem with detecting fine differences in pressure sensor measurements for different leak scenarios, and although true leak location is always in several top nodes, it is not always the prediction with the greatest certainty. Optimization methods could provide finer parameter tuning, therefore, coupling of machine learning approach for general localization with optimization approach for finer detection of leak location should be explored.

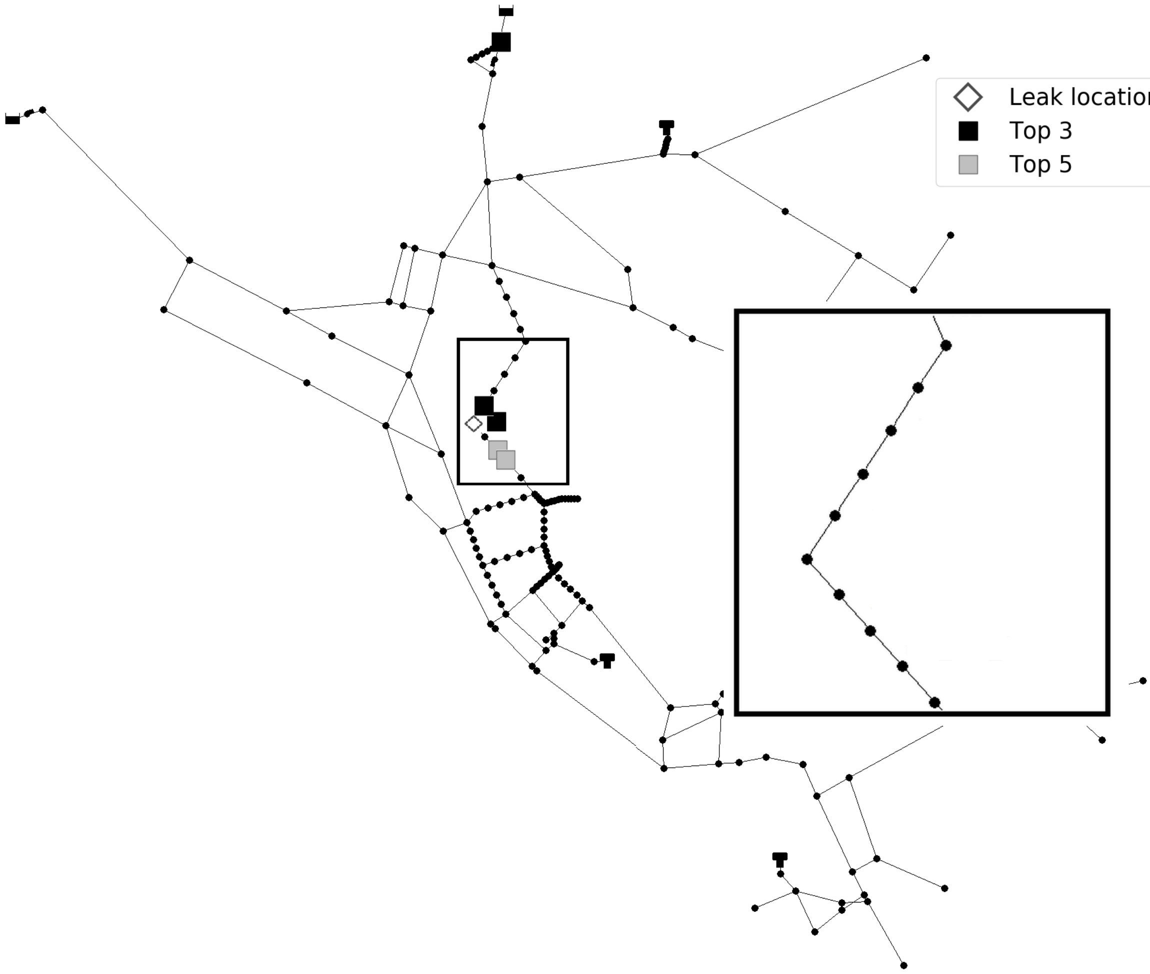

Figure 3.3: Pipe segmentation after initial machine learning localization as conducted in [50].

Chapter 4

Conclusion

4.1 Main contributions

In the presented thesis, multiple novel applications of artificial intelligence in water transportation systems are presented and discussed. The enclosed papers show the in-depth methodology of proposed applications, where multiple directions of possible further utilization are discussed in the present thesis. It was shown that with growing computational resources utilization of novel approaches is greatly beneficial, where previously used methods were less efficient.

The main contributions from the presented research are as follows:

- Numerical analysis and optimization methods have been used to determine optimal cross-section shape for the defined bar and trash-rack inclinations of the water intake structure. Conducted research showed that optimal cross-section shape varied for different trash-rack configurations, which showed the importance of optimization methods that can freely adjust geometry-shape for specific intake structures.

- A novel search space reduction method in pollution detection was presented, which considerably reduced the number of potential pollution sources. Based on the reduced number of solutions, a novel pollution localization technique was presented, in which for each remaining suspect node an independent optimization procedure was conducted to obtain pollution starting time, duration of injection, and concentration value. This approach significantly reduced the complexity of the optimization problem since the categorical variable was removed. Additionally, since the considered problem is multi-modal, multiple solutions are simultaneously obtained which is overall computationaly more efficient.

- A novel machine learning approach that identifies the number of pollution sources in the water distribution network is proposed. The presented method show good accuracy when the exact number of pollution locations are predicted, which is important information for reducing the number of unnecessary optimization variables. Additionally, when it is predicted only whether single or multiple pollution sources are present, the accuracy of the technique increases. This is important since some space reduction methods are more efficient and others are specialized for a single pollution location.

- Random Forest prediction model trained on synthetic pressure data obtained by simulated leak scenarios was utilized for the leak localization problem. It was shown that the proposed approach can incorporate various uncertainties regarding water distribution network behavior and provide considerable leak localization. That represents a strong benefit since software-based methods that use residuals as the criteria for anomalies detection use an estimated water distribution network model and thus can cause wrong results.

- Further investigation of machine learning application in leak localization was conducted for the pipe segmentation approach. Since the usage of 1D numerical simulations enables detecting only network nodes as leak locations, and leaks can occur anywhere in pipe segment, additional network nodes were created to further localize leak location. Sequential machine learning models were used, where the first Random Forest model was used to identify leak area for pipe segmentation, and the second model was used to try to identify exact leak location. It was observed that the proposed approach narrows down the leak area for leaks occurring both in network nodes and in pipe segments with great accuracy; however, the exact location cannot be determined since multiple leak locations with various leak sizes can produce similar pressure sensor measurements.

4.2 Future work

As a continuation of conducted research possible future research areas are:

- The design optimization of the trash-rack cross-section should be conducted with more parameters where profiles would be described with a large number of points that would be able to converge into any shape, such as curved cross-sections. Additionally, the expansion of fitness function for design optimization with ecological goals should be explored, such as including fish avoidance ability considering the proposed design.
- Search space reduction technique coupled with independent optimization methods proposed for the detection of pollution scenario parameters should be explored in the context of rapid reaction time, where the search for the pollution parameters will start immediately after pollution detection, not after a longer measurement time which was the case in this study. Additionally, time-varying pollution injection should be considered since the assumption of constant injection concentration is assumed.
- Other machine learning algorithms should be investigated to achieve improved accuracy for the identification of a number of pollution sources and leak location. Additionally, dimensionality reduction methods should be also explored to reduce prediction model complexity.
- The machine learning application that identifies the number of pollution sources should be further explored with various pollution scenario parameters for each injection node.
- Coupling of machine learning algorithm that would identify the leak area and optimization methods which would find exact leak location should be explored.
- The segmentation approach should be applied to pollution location detection since leak locations are also possible pollution intrusions locations. Thus, the current assumption of pollution locations only in network nodes is also a considerable simplification of the problem, which should be avoided if possible.

Chapter 5

Summary of papers

A Assessment of Head Loss Coefficients for Water Turbine Intake Trash-Racks by Numerical Modeling

In this work, numerical simulations of fluid flow around trash-rack for different bar cross sections are conducted to investigate cross section influence on head losses. Comparison with experimental data is conducted to validate the usage of numerical simulations which enable investigation of great number of trash-rack configurations. In previous experimental studies researchers mostly focused on trash-rack parameters (bar spacing, bar length, inclinations etc.) where bar cross section was mainly rectangular or streamlined shape. Therefore, 2D simulations for different cross sections are carried out for a range of trash-rack configurations in order to provide better insight how it affects energy losses. It is shown that head loss reduction due to change in cross section is greatly dependent on trash-rack configuration, therefore optimization of simplified real water turbine trash-rack is also conducted to produce the cross section that generates smallest head losses for given configuration.

Lučin, I. , Čarija, Z., Grbčić, L., Kranjčević, L., 2020. Assessment of Head Loss Coefficients for Water Turbine Intake Trash-Racks by Numerical Modeling. Journal of Advanced Research, 21, pp. 109-119.; https://doi.org/10.1016/j.jare.2019.10.010

B Source Contamination Detection Using Novel Search Space Reduction Coupled with Optimization Technique

Contaminant intrusion in a water distribution network is an important concern because it can have hazardous consequences for the population. Reacting immediately is crucial to prevent or reduce the further propagation of contamination. In terms of contamination scenario characteristics, optimization is researched extensively as a valuable methodology to provide information. This work presented a procedure preceding the optimization which considerably reduces the search space for a potential contaminant source location. For each suspect node, a simulation is conducted with unrealistically high contaminant concentration injected throughout the whole simulation. If the sensors do not register contamination in a subsequent scenario, then that node can be eliminated as a possible contaminant source. The methodology is applicable for both single and multiple contaminant injection nodes. This approach was investigated in multiple benchmark networks and for different sensor placements in the literature. By coupling the proposed search space reduction method with an optimization approach, a novel efficient methodology for contamination source detection was presented.

Lučin, I., Grbčić, L., Družeta, S., Čarija, Z., 2021. Source Contamination Detection Using Novel Search Space Reduction Coupled with Optimization Technique. Journal of Water Resources Planning and Management, 147 (2), p. 04020100.; https://doi.org/10.1061/(ASCE)WR.1943-5452.0001308

C Machine-Learning Classification of a Number of Contaminant Sources in an Urban Water Network

In the case of a contamination event in water distribution networks, several studies have considered different methods to determine contamination scenario information. It would be greatly beneficial to know the exact number of contaminant injection locations since some methods can only be applied in the case of a single injection location and others have greater efficiency. In this work, the Neural Network and Random Forest classifying algorithms are used to predict the number of contaminant injection locations. The prediction model is trained with data obtained from simulated contamination event scenarios with random injection starting time, duration, concentration value, and the number of injection locations which varies from 1 to 4. Classification is made to determine if single or multiple injection locations occurred, and to predict the exact number of injection locations. Data was obtained for two different benchmark networks, medium-sized network Net3 and large-sized Richmond network. Additionally, an investigation of sensor layouts, demand uncertainty, and fuzzy sensors on model accuracy is conducted. The proposed approach shows excellent accuracy in predicting if single or multiple contaminant injections in a water supply network occurred and good accuracy for the exact number of injection locations.

Lučin, I., Grbčić, L., Čarija, Z., Kranjčević, L., 2021. Machine-Learning Classification of a Number of Contaminant Sources in an Urban Water Network. Sensors, 21 (1), p. 245.; https://doi.org/10.3390/s21010245

D Data-Driven Leak Localization in Urban Water Distribution Networks Using Big Data for Random Forest Classifier

In the present paper, a Random Forest classifier is used to detect leak locations on two different sized water distribution networks with sparse sensor placement. A great number of leak scenarios were simulated with Monte Carlo determined leak parameters (leak location and emitter coefficient). In order to account for demand variations that occur on a daily basis and to obtain a larger dataset, scenarios were simulated with random base demand increments or reductions for each network node. Classifier accuracy was assessed for different sensor layouts and numbers of sensors. Multiple prediction models were constructed for differently sized leakage and demand range variations in order to investigate model accuracy under various conditions. Results indicate that the prediction model provides the greatest accuracy for the largest leaks, with the smallest variation in base demand (62% accuracy for greater- and 82% for smaller-sized networks, for the largest considered leak size and a base demand variation of $\pm 2.5\%$). However, even for small leaks and the greatest base demand variations, the prediction model provided considerable accuracy, especially when localizing the sources of leaks when the true leak node and neighbor nodes were considered (for a smaller-sized network and a base demand of variation $\pm 20\%$ the model accuracy increased from 44% to 89% when top five nodes with greatest probability were considered, and for a greater-sized network with a base demand variation of $\pm 10\%$ the accuracy increased from 36% to 77%).

Lučin, I., Lučin, B., Čarija, Z., Sikirica, A., 2021. Data-Driven Leak Localization in Urban Water Distribution Networks Using Big Data for Random Forest Classifier. Mathematics, 9 (6), p. 672.; https://doi.org/10.3390/math9060672

E Detailed Leak Localization in Water Distribution Networks Using Random Forest Classifier and Pipe Segmentation

In this paper, a Random Forest classifier was used to predict leak locations for two differently sized water distribution networks based on pressure sensor measurements. The prediction model is trained on simulated leak scenarios with randomly chosen parameters - leak location, leak size, and base node demand uncertainty. Leak localization methods found in literature that rely on numerical simulations can only predict network nodes as leak nodes; however, since a leak can occur at any point along a pipe segment, additional spatial discretization of suspect pipe is proposed in this paper. It was observed that pipe segmentation of the whole network is a non-feasible approach since it rapidly increases the number of potential leak locations, consequently increasing the complexity of the prediction model. Therefore, a novel approach is proposed, in which a prediction model is trained on scenarios with leaks occurring in original network nodes only, but with its accuracy assessed against pressure sensor measurements from scenarios in which leaks occur in points between network nodes. It was observed that this approach can successfully narrow down the suspect leak area and, followed by additional segmentation of that network area and subsequent prediction, a precise leak localization can be achieved. The proposed approach enables incorporation of various uncertainties by simulating leak scenarios under different conditions. Investigation of leak size uncertainty and base demand variation showed that several different scenarios can produce similar sensor measurements which makes it difficult to unambiguously determine leak location using the prediction model. Therefore, future approaches of coupling prediction modeling with optimization methods are proposed.

Lučin, I., Čarija, Z., Lučin, B., Družeta, S., 2021. Detailed Leak Localization in Water Distribution Networks Using Random Forest Classifier and Pipe Segmentation. IEEE Access, 9, pp. 155113-155122.; https://doi.org/10.1109/ACCESS.2021.3129703

Bibliography

[1] A. A. Abokifa et al. Real-time identification of cyber-physical attacks on water distribution systems via machine learning–based anomaly detection techniques. *Journal of Water Resources Planning and Management*, 145(1):04018089, 2019.

[2] O. S. Adedoja et al. Towards development of an optimization model to identify contamination source in a water distribution network. *Water*, 10(5):579, 2018.

[3] H. O. Åkerstedt et al. Numerical investigation of turbulent flow through rectangular and biconvex shaped trash racks. *Engineering*, 9(05):412, 2017.

[4] I. Albayrak et al. An experimental investigation on louvres and angled bar racks. *Journal of Hydraulic Research*, 56(1):59–75, 2018.

[5] S. V. Amaral et al. Survival of fish passing downstream at a small hydropower facility. *Marine and Freshwater Research*, 69(12):1870–1881, 2018.

[6] U. Baroudi et al. Pipeline leak detection systems and data fusion: A survey. *IEEE Access*, 7:97426–97439, 2019.

[7] N. A. Barton et al. Improving pipe failure predictions: Factors affecting pipe failure in drinking water networks. *Water Research*, 164:114926, 2019.

[8] C. Beck et al. Improved hydraulic performance of fish guidance structures with innovative bar design. In *12th International symposium on ecohydraulics (ISE 2018)*, 2018.

[9] M.-C. Besner et al. Assessing the public health risk of microbial intrusion events in distribution systems: conceptual model, available data, and challenges. *Water Research*, 45(3):961–979, 2011.

[10] C. Binnie et al. *Basic water treatment*. Royal society of chemistry, 2002.

[11] M. G. Carleton and J. S. Nielsen. A study of trash and trash interception devices. *Water Science and Technology*, 22(10-11):287–290, 1990.

[12] T. K. Chan et al. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*, 6:78846–78867, 2018.

[13] J. Chen et al. An iterative method for leakage zone identification in water distribution networks based on machine learning. *Structural Health Monitoring*, 20(4):1938–1956, 2021.

[14] S. P. Clark et al. Experimental study of energy loss through submerged trashracks. *Journal of Hydraulic Research*, 48(1):113–118, 2010.

[15] I. S. Cole and D. Marney. The science of pipe corrosion: A review of the literature on the corrosion of ferrous metals in soils. *Corrosion Science*, 56:5–16, 2012.

[16] A. F. Colombo et al. A selective literature review of transient-based leak detection methods. *Journal of Hydro-environment Research*, 2(4):212–227, 2009.

[17] S. H. Crandall et al. Destructive vibration of trashracks due to fluid-structure interaction. *Journal of Engineering for Industry*, 97(4):1359–1365, 1975.

[18] S. Datta and S. Sarkar. A review on different pipeline fault detection methods. *Journal of Loss Prevention in the Process Industries*, 41:97–106, 2016.

[19] T. Dawood et al. Water pipe failure prediction and risk models: state-of-the-art review. *Canadian Journal of Civil Engineering*, 47(10):1117–1127, 2020.

[20] A. E. De Sanctis et al. Real-time identification of possible contamination sources using network backtracking methods. *Journal of Water Resources Planning and Management*, 136(4):444–453, 2010.

[21] E. Egusquiza et al. Failures due to ingested bodies in hydraulic turbines. *Engineering Failure Analysis*, 18(1):464–473, 2011.

[22] R. Farmani et al. Pipe failure prediction in water distribution systems considering static and dynamic factors. *Procedia Engineering*, 186:117–126, 2017.

[23] S. D. Faust and O. M. Aly. *Chemistry of water treatment*. CRC press, 2018.

[24] J. Gibson et al. Predicting health risks from intrusion into drinking water pipes over time. *Journal of Water Resources Planning and Management*, 145(3):04019001, 2019.

[25] M. M. Giraldo-González and J. P. Rodríguez. Comparison of statistical and machine learning models for pipe failure modeling in water distribution networks. *Water*, 12(4):1153, 2020.

[26] V. Gomez-Alvarez and R. P. Revetta. Monitoring of nitrification in chloraminated drinking water distribution systems with microbiome bioindicators using supervised machine learning. *Frontiers in Microbiology*, 11:2254, 2020.

[27] L. Grbčić et al. Efficient double-tee junction mixing assessment by machine learning. *Water*, 12(1):238, 2020.

[28] L. Grbčić et al. A machine learning-based algorithm for water network contamination source localization. *Sensors*, 20(9):2613, 2020.

[29] L. Grbčić et al. Water supply network pollution source identification by random forest algorithm. *Journal of Hydroinformatics*, 22(6):1521–1535, 2020.

[30] L. Grbčić et al. Machine learning and simulation-optimization coupling for water distribution network contamination source detection. *Sensors*, 21(4):1157, 2021.

[31] A. Gupta and K. D. Kulat. A selective literature review on leak management techniques for water distribution system. *Water Resources Management*, 32(10):3247–3269, 2018.

[32] J. Hall et al. On–line water quality parameters as indicators of distribution system contamination. *Journal-American Water Works Association*, 99(1):66–77, 2007.

[33] D. Honingh et al. Urban river water level increase through plastic waste accumulation at a rack structure. *Frontiers in Earth Science*, 8:28, 2020.

[34] C. Hu et al. A mapreduce based parallel niche genetic algorithm for contaminant source identification in water distribution network. *Ad Hoc Networks*, 35:116–126, 2015.

[35] Z. Hu et al. A comprehensive review of acoustic based leak localization method in pressurized pipelines. *Mechanical Systems and Signal Processing*, 161:107994, 2021.

[36] I. E. Idelchik. Handbook of hydraulic resistance. *Washington, DC, Hemisphere Publishing Corp.*, 1986.

[37] M. L. Inglis et al. Testing the effectiveness of fish screens for hydropower intakes. Report Project SC120079. https://www.gov.uk/government/publications/testing-the-effectiveness-of-fish-screens-for-hydropower-intakes, 2016. Accessed: 16-July-2019.

[38] M. I. M. Ismail et al. A review of vibration detection methods using accelerometer sensors for water pipeline leakage. *IEEE Access*, 7:51965–51981, 2019.

[39] M. R. Karim et al. Potential for pathogen intrusion during pressure transients. *Journal-American Water Works Association*, 95(5):134–146, 2003.

[40] L. A. Khan et al. Computational fluid dynamics modeling of turbine intake hydraulics at a hydropower plant. *Journal of Hydraulic Research*, 42(1):61–69, 2004.

[41] G. J. Kirmeyer and K. Martel. *Pathogen intrusion into the distribution system*. American Water Works Association, 2001.

[42] Y. Kleiner and B. Rajani. Comprehensive review of structural deterioration of water mains: statistical models. *Urban Water*, 3(3):131–150, 2001.

[43] K. A. Klise et al. Two-tiered sensor placement for large water distribution network models. *Journal of Infrastructure Systems*, 19(4):465–473, 2013.

[44] C. S. W. Lavooij and A. S. Tusseling. Fluid-structure interaction in liquid-filled piping systems. *Journal of Fluids and Structures*, 5(5):573–595, 1991.

[45] R. Li et al. A review of methods for burst/leakage detection and location in water distribution systems. *Water Science and Technology: Water Supply*, 15(3):429–441, 2015.

[46] M. Lin et al. Wireless sensor network: Water distribution monitoring system. In *2008 IEEE radio and wireless symposium*, pages 775–778. IEEE, 2008.

[47] Y. Liu et al. Water pipeline leakage detection based on machine learning and wireless sensor networks. *Sensors*, 19(23):5086, 2019.

[48] I. Lučin et al. Assessment of head loss coefficients for water turbine intake trash-racks by numerical modeling. *Journal of advanced research*, 21:109–119, 2020.

[49] I. Lučin et al. Data-driven leak localization in urban water distribution networks using big data for random forest classifier. *Mathematics*, 9(6):672, 2021.

[50] I. Lučin et al. Detailed leak localization in water distribution networks using random forest classifier and pipe segmentation. *IEEE Access*, 9:155113–155122, 2021.

[51] I. Lučin et al. Machine-learning classification of a number of contaminant sources in an urban water network. *Sensors*, 21(1):245, 2021.

[52] I. Lučin et al. Source contamination detection using novel search space reduction coupled with optimization technique. *Journal of Water Resources Planning and Management*, 147(2):04020100, 2021.

[53] D. McInnis. A relative-risk framework for evaluating transient pathogen intrusion in distribution systems. *Urban Water Journal*, 1(2):113–127, 2004.

[54] L. S. McNeill and M. Edwards. Iron pipe corrosion in distribution systems. *Journal-American Water Works Association*, 93(7):88–100, 2001.

[55] J. Meister et al. Hydraulics of horizontal bar racks for fish downstream migration. In *38th International Association for Hydro-Environmental Engineering and Research World Congress (IAHR 2019)*, 2019.

[56] H. Mohammed et al. Machine learning: based detection of water contamination in water distribution systems. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, pages 1664–1671, 2018.

[57] S. R. Mounce et al. Development and verification of an online artificial intelligence system for detection of bursts and other abnormal flows. *Journal of Water Resources Planning and Management*, 136(3):309–318, 2010.

[58] P. Nader et al. Detection of cyberattacks in a water distribution system using machine learning techniques. In *2016 Sixth International Conference on Digital Information Processing and Communications (ICDIPC)*, pages 25–30. IEEE, 2016.

[59] S. S. Paul and M. S. Adaramola. Analysis of turbulent flow past bar-racks. In *ASME international mechanical engineering congress and exposition*, volume 46545, page V007T09A029. American Society of Mechanical Engineers, 2014.

[60] F. Pedregosa et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[61] E. J. Pérez-Pérez et al. Leak diagnosis in pipelines using a combined artificial neural network approach. *Control Engineering Practice*, 107:104677, 2021.

[62] A. Preis and A. Ostfeld. A contamination source identification model for water distribution system security. *Engineering Optimization*, 39(8):941–947, 2007.

[63] A. Preis and A. Ostfeld. Genetic algorithm for contaminant source characterization using imperfect sensors. *Civil Engineering and Environmental Systems*, 25(1):29–39, 2008.

[64] T. Qin and D. L. Boccelli. Grouping water-demand nodes by similarity among flow paths in water-distribution systems. *Journal of Water Resources Planning and Management*, 143(8):04017033, 2017.

[65] Y. Qin et al. Microfabricated electrochemical pH and free chlorine sensors for water quality monitoring: recent advances and research challenges. *RSC Advances*, 5(85):69086–69109, 2015.

[66] M. Quiñones-Grueiro et al. An unsupervised approach to leak detection and location in water distribution networks. *International Journal of Applied Mathematics and Computer Science*, 28(2):283–295, 2018.

[67] B. Rajani and Y. Kleiner. Comprehensive review of structural deterioration of water mains: physically based models. *Urban Water*, 3(3):151–164, 2001.

[68] S. Raynal et al. An experimental study on fish-friendly trashracks–part 1. inclined trashracks. *Journal of Hydraulic Research*, 51(1):56–66, 2013.

[69] S. Raynal et al. An experimental study on fish-friendly trashracks–part 2. angled trashracks. *Journal of Hydraulic Research*, 51(1):67–75, 2013.

[70] S. Raynal et al. Numerical simulations of fish-friendly angled trashracks at model and real scale. In *35th World Congress of the International Association for Hydro-Environment Engineering and Research*, 2013.

[71] L. A. Rossman. Epanet 2: Users manual. Tech. Rep. EPA/600/R-00/057. 2000.

[72] R. Sadiq et al. Water quality failures in distribution networks—risk analysis using fuzzy logic and evidential reasoning. *Risk Analysis: An International Journal*, 27(5):1381–1394, 2007.

[73] M. Salah Abd Elmoaty. An experimental investigation of the impact of aquatic weeds trash racks on water surface profile in open channels. *Water Science*, 33(1):84–92, 2019.

[74] A. M. A. Sattar et al. Extreme learning machine model for water network management. *Neural Computing and Applications*, 31(1):157–169, 2019.

[75] A. Seth et al. Testing contamination source identification methods for water distribution networks. *Journal of Water Resources Planning and Management*, 142(4):04016001, 2016.

[76] A. Soldevila et al. Leak localization in water distribution networks using a mixed model-based/data-driven approach. *Control Engineering Practice*, 55:162–173, 2016.

[77] A. Soldevila et al. Leak localization in water distribution networks using bayesian classifiers. *Journal of Process Control*, 55:1–9, 2017.

[78] A. M. St. Clair and S. Sinha. State-of-the-technology review on water pipe condition, deterioration and failure rate prediction models! *Urban Water Journal*, 9(2):85–112, 2012.

[79] C. Sun et al. Leak localization in water distribution networks using pressure and data-driven classifier approach. *Water*, 12(1):54, 2020.

[80] A. S. Tijsseling. *Fluid-structure interaction in the case of waterhammer with cavitation.* PhD thesis, Technische Universiteit Delft (Netherlands), 1995.

[81] P. Vankayala et al. Contaminant source identification in water distribution networks under conditions of demand uncertainty. *Environmental Forensics*, 10(3):253–263, 2009.

[82] D. Vries et al. Application of machine learning techniques to predict anomalies in water supply networks. *Water Science and Technology: Water Supply*, 16(6):1528–1535, 2016.

[83] H. Wang and K. W. Harrison. Improving efficiency of the bayesian approach to water distribution contaminant source characterization with support vector regression. *Journal of Water Resources Planning and Management*, 140(1):3–11, 2014.

[84] T. Westrell et al. A theoretical approach to assess microbial risks due to failures in drinking water systems. *International Journal of Environmental Health Research*, 13(2):181–197, 2003.

[85] A. J. Whittle et al. Sensor networks for monitoring and control of water distribution systems. In *6th International Conference on Structural Health Monitoring of Intelligent Infrastructure (SHMII 2013)*. International Society for Structural Health Monitoring of Intelligent Infrastructure, 2013.

[86] D. C. Wiggert and A. S. Tijsseling. Fluid transients and fluid-structure interaction in flexible liquid-filled piping. *Applied Mechanics Reviews*, 54(5):455–481, 2001.

[87] D. Wilson et al. State-of-the-art review of water pipe failure prediction models and applicability to large-diameter mains. *Urban Water Journal*, 14(2):173–184, 2017.

[88] Y. Wu and S. Liu. A review of data-driven approaches for burst detection in water distribution systems. *Urban Water Journal*, 14(9):972–983, 2017.

[89] F. Xiang et al. Experimental study of debris capture efficiency of trash racks. *Journal of Hydro-environment Research*, 3(3):138–147, 2009.

[90] X. Yan et al. Research on contaminant sources identification of uncertainty water demand using genetic algorithm. *Cluster Computing*, 20(2):1007–1016, 2017.

[91] X. Yan et al. Multimodal optimization problem in contamination source determination of water supply networks. *Swarm and Evolutionary Computation*, 47:66–71, 2019.

[92] X. Yan et al. Pollution source localization in an urban water supply network based on dynamic water demand. *Environmental Science and Pollution Research*, 26(18):17901–17910, 2019.

[93] A. Yazdani and P. Jeffrey. Complex network analysis of water distribution systems. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 21(1):016111, 2011.

[94] D. Zaman et al. A review of leakage detection strategies for pressurised pipeline in steady-state. *Engineering Failure Analysis*, 109:104264, 2020.

[95] M. Zayed et al. An experimental investigation of head loss through a triangular “v-shaped” screen. *Journal of Advanced Research*, 10:69–76, 2018.

[96] M. Zayed et al. An experimental study on angled trash screen in open channels. *Alexandria Engineering Journal*, 57(4):3067–3074, 2018.

[97] M. Zayed et al. Experimental investigation of curved trash screens. *Journal of Irrigation and Drainage Engineering*, 146(6):06020003, 2020.

[98] M. Zayed and E. Farouk. Effect of blocked trash rack on open channel infrastructure. *Water Practice and Technology*, 16(1):247–262, 2021.

[99] X. Zhou et al. Deep learning identifies accurate burst locations in water distribution networks. *Water Research*, 166:115058, 2019.

List of Figures

List of Tables

Curriculum Vitae

Ivana Lučin was born in Rijeka, Croatia in 1991. She received mag. ing. mech. title in 2015. at Faculty of Engineering, University of Rijeka. In 2016. she enrolled in Postgraduate University Doctoral Study in Computational Mechanics at Faculty of Engineering, University of Rijeka under the supervision of Prof. D. Sc. Zoran Čarija. Since 2016. she is working as research and teaching assistant at Faculty of Engineering, University of Rijeka at Department of Fluid Mechanics and Computational Engineering. She is teaching assistant on the courses Computer Applications in Engineering, Computational Methods, Hydraulic Machines, Programming: Scripting Languages, Computational Fluid Dynamics, and Visualisation and Setup of Computer Simulations at the undergraduate and graduate university study of mechanical engineering and computer science. Her research areas are application of optimization and machine learning methods in engineering. She is a member of the Center for Advanced Computing and Modeling at the University of Rijeka.

List of Publications

Scientific papers in peer-reviewed journals:

- Lučin, I., Čarija, Z., Lučin, B. and Družeta, S., 2021. Detailed Leak Localization in Water Distribution Networks Using Random Forest Classifier and Pipe Segmentation. IEEE Access , 9, pp. 155113-155122.
- Grbčić, L., Kranjčević, L., Lučin, I. and Sikirica, A., 2021. Large Eddy Simulation of turbulent fluid mixing in double-tee junctions. Ain Shams Engineering Journal, 12(1), pp.789-797.
- Lučin, I., Grbčić, L., Družeta, S. and Čarija, Z., 2021. Source Contamination Detection Using Novel Search Space Reduction Coupled with Optimization Technique. Journal of Water Resources Planning and Management, 147 (2), p. 04020100.
- Lučin, I., Lučin, B., Čarija, Z. and Sikirica, A., 2021. Data-Driven Leak Localization in Urban Water Distribution Networks Using Big Data for Random Forest Classifier. Mathematics, 9 (6), p. 672.
- Lučin, I., Grbčić, L., Čarija, Z. and Kranjčević, L., 2021. Machine-Learning Classification of a Number of Contaminant Sources in an Urban Water Network. Sensors, 21 (1), p. 245.
- Sikirica, A., Čarija, Z., Lučin, I., Grbčić, L. and Kranjčević, L., 2020. Cavitation Model Calibration Using Machine Learning Assisted Workflow. Mathematics, 8(12), p.2107.
- Grbčić, L., Lučin, I., Kranjčević, L. and Družeta, S., 2020. Water supply network pollution source identification by random forest algorithm. Journal of Hydroinformatics, 22(6), pp.1521-1535.

- Družeta, S., Ivić, S., Grbčić, L. and Lučin, I., 2020. Introducing languid particle dynamics to a selection of PSO variants. Egyptian Informatics Journal, 21(2), pp.119-129.

- Grbčić, L., Lučin, I., Kranjčević, L. and Družeta, S., 2020. A machine learning-based algorithm for water network contamination source localization. Sensors, 20(9), p.2613.

- Grbčić, L., Kranjčević, L., Družeta, S. and Lučin, I., 2020. Efficient double-Tee junction mixing assessment by machine learning. Water, 12(1), p.238.

- Lučin, I. , Čarija, Z., Grbčić, L., Kranjčević, L., 2020. Assessment of Head Loss Coefficients for Water Turbine Intake Trash-Racks by Numerical Modeling. Journal of Advanced Research, 21, pp. 109-119.

- Sikirica, A., Čarija, Z., Kranjčević, L. and Lučin, I., 2019. Grid type and turbulence model influence on propeller characteristics prediction. Journal of Marine Science and Engineering, 7(10), p.374.

- Grbčić, L., Kranjčević, L., Lučin, I. and Čarija, Z., 2019. Experimental and numerical investigation of mixing phenomena in double-Tee junctions. Water, 11(6), p.1198.

- Ivić, S., Družeta, S., Hreljac, I., 2017. S-Lay pipe laying optimization using specialized PSO method. Structural and Multidisciplinary Optimization, 56(2), pp.297-313.

Scientific papers in conferences:

- Sikirica, A., Lučin, I., Čarija, Z. and Lučin, B., 2020. CFD Analysis of Marine Propeller Configurations in Cavitating Conditions. Pomorski zbornik, (3), pp.251-264.

- Lučin, I., Kranjčević, L., Čarija, Z. and Mogorović, A., 2018. Experimental Setup of Fluid Mixing in Double Tee-Junctions. In Proceedings of the 29th DAAAM international symposium (Vol. 29, pp. 1046-52).

- Čarija, Z., Lučin, I., Lučin, B. and Grbčić, L., 2018, January. Investigation of numerical simulation parameters on fluid flow around trash-racks. In Proceedings of the 29th DAAAM international symposium (Vol. 29, pp. 1046-52).

- Zeng, H., Grbčić, L., Lučin, I. and Kranjčević, L., 2018. Mesh creation for realistic terrain cases for shallowfoam-2D OpenFoam solver. In Proceedings of the 29th DAAAM international symposium (Vol. 29, pp. 1046-52).

Part II

Included Publications