Numerical modeling of fluid mixing in pipe networks with machine learning applications

Grbčić, Luka

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Luka Grbčić

NUMERICAL MODELING OF FLUID MIXING IN PIPE NETWORKS WITH MACHINE LEARNING APPLICATIONS

DOCTORAL DISSERTATION

Rijeka, 2021.

UNIVERSITY OF RIJEKA
FACULTY OF ENGINEERING

Luka Grbčić

NUMERICAL MODELING OF FLUID MIXING IN PIPE NETWORKS WITH MACHINE LEARNING APPLICATIONS

DOCTORAL DISSERTATION

Thesis Supervisor: Prof. D. Sc. Lado Kranjčević

Rijeka, 2021.

Thesis Supervisor: Prof. D. sc. Lado Kranjčević

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

As water supply is one of the most recognizable and important public services contributing to the quality of life, the importance of a safe water supply system is a basic requirement for every urban water distribution network. Correct treatment of mixing phenomena in the network's hydraulic elements is of great importance for the accurate simulation and prediction of contamination events. In this doctoral thesis, the mixing of fluids in pipe networks and segments (double-Tee junctions) is investigated using an experimental approach, numerical models and machine learning methods. Firstly, experimental data is obtained for fluid mixing in different double-Tee configurations and a computational fluid dynamics transport model is calibrated and validated through setting the turbulent Schmidt number. A machine learning approach which uses Artificial Neural Networks is then proposed and trained with numerically generated data for the purpose of creating computationally efficient models for fluid mixing in double-Tee junctions. Additionally, a Large Eddy Simulation model coupled with a pure advection transport model is investigated and shown to produce accurate fluid mixing results in a double-Tee junction therefore showing a way to not include the turbulent Schmidt number. Secondly, three methodologies are presented which use machine learning algorithms for the purpose of contamination source detection in water supply pipe networks. The first methodology is based on Random Forest algorithm classification of the most probable contamination source in a water supply network. The RF classifier is trained with data obtained by Monte Carlo water quality simulations and it shows to be computationally efficient and can easily generate a list of most probable contamination sources in a water supply network. Furthermore, a novel machine learning based algorithm for water supply contamination source identification is presented and built for high performance parallel systems. The algorithm successfully utilizes the combination of Artificial Neural Networks for a parallel tournament style classification of the contamination source with Random Forests for regression analysis to determine significant variables of a contamination event. Lastly, two novel algorithmic frameworks are investigated which are based on coupling a machine learning algorithm for predicting the most probable contamination sources in a water distribution network with an optimization algorithm for determining the relevant parameters such as contamination start time, end time and contaminant concentration for each predicted node separately. Both algorithmic frameworks perform well in determining the true source node, start and end times and contaminant concentration.

Keywords: pipe fluid mixing; machine learning; contamination source detection; pipe networks; transport models; artificial neural networks; random forests; simulation-optimization

Prošireni sažetak

Vodoopskrba je jedna od najvažnijih javnih usluga koja doprinosi kvaliteti života, a važnost sigurnog vodoopskrbnog sustava osnovni je zahtjev svake gradske vodovodne mreže. Pravilno modeliranje miješanja u hidrauličkim elementima mreže od velike je važnosti za točnu simulaciju i predikciju širenja onečišćenja. U ovoj doktorskoj disertaciji istražuje se miješanje fluida u cjevovodnim mrežama i njenim segmentima (dvostruki T-spojevi) eksperimentalnim pristupom, numeričkim modelima i metodama strojnog učenja u svrhu točnijeg i efikasnijeg modeliranja širenja i detekcije onečišćenja u urbanim vodoopskrbnim mrežama. Dobiveni su novi eksperimentalni podaci za miješanje fluida u različitim konfiguracijama dvostrukih T-spojeva te je validiran numerički model kalibracijom turbulentnog Schmidtovog broja. Istražuje se i predlaže pristup baziran na strojnom učenju koji koristi umjetne neuronske mreže i koji se trenira podacima koji su generirani rezultatima dobivenim putem numeričkih analiza u svrhu stvaranja računski učinkovitih i robusnih modela za miješanje fluida u dvostrukim T-spojevima. S obzirom na računsku efikasnost, modeli koji su izrađeni strojnim učenjem mogu se integrirati u postojeće inženjerske sofvere koji su bazirani na jednostavnijim modelima, kako bi se povećala njihova točnost kod simulacije propagacije onečišćenja u cjevovodnim mrežama. Osim navedenog, istražuje se i modeliranje miješanja fluida u dvostrukom T-spoju putem modela turbulencije *Large Eddy Simulation* povezanim s modelom čistog advekcijskog transporta skalara. Predloženi način modeliranja miješanja fluida u dvostrukom T-spoju daje točne rezultate miješanja u dvostrukom T-spoju koji ne uključuje potrebu za kalibracijom turbulentnog Schmidtov broja te time stvara mogućnost za generiranje korekcijskih faktora za jednostavnije modele u svrhu točnijeg modeliranja onečišćenja u cjevovodima. Predstavljene su i istražene tri različite metodologije koje koriste algoritme strojnog učenja u svrhu otkrivanja izvora onečišćenja u vodovodnim mrežama. Prva od tri metodologija se temelji na klasifikaciji najvjerojatnijeg izvora onečišćenja u vodoopskrbnoj mreži te radi na temelju algoritma *Nasumičnih šuma (Random Forest)*. Algoritam *Nasumičnih šuma* istreniran je podacima koji su dobiveni *Monte Carlo* hidrauličkim simulacijama koje uključuju i modeliranje transporta onečišćenja. Algoritam se pokazuje računalno učinkovitim i lako generira popis najvjerojatnijih izvora onečišćenja u vodoopskrbnoj mreži te je testiran na dvije različite vodoopskrbne mreže različitih veličina. Nadalje, predstavljena je druga metodologija koja koristi algoritam zasnovan na strojnom učenju za klasifikaciju izvora onečišćenja u vodoopskrbnim mrežama koji je izrađen posebno za paralelne sustave visokih performansi. Algoritam koristi kombinaciju *Umjetnih neuronskih mreža*

(Artificial Neural Networks) za klasifikaciju izvora onečišćenja te algoritam *Nasumičnih šuma* za regresijsku analizu kako bi odredio značajne varijable događaja onečišćenja kao što su vrijeme početka onečišćenja, vrijeme završetka onečišćenja i koncentraciju onečišćenja. Uspješno sužava potencijalne izvore onečišćenja što dovodi do identifikacije izvora onečišćenja, vremena početka i vremena završetka događaja onečišćenja i koncentraciju onečišćenja te je vrlo računski efikasan kada se koristi u superračunalnom okruženju. Na kraju, trećom metodologijom predstavljena su i istražena dva nova algoritamska okvira koja se temelje na spajanju algoritma strojnog učenja za predviđanje najvjerojatnijih izvora onečišćenja u vodovodnoj mreži s optimizacijskim algoritmima (stohastičkih i determinističkih) za određivanje relevantnih parametara kao što su vrijeme početka, vrijeme završetka onečišćenja i koncentracija onečišćenja za svaki predviđeni potencijalni izvor zasebno. Oba algoritamska okvira imaju dobru izvedbu i pokazuju robusnost u određivanju istinskog izvora onečišćenja.

Ključne riječi: miješanje fluida u cijevima; strojno učenje; otkrivanje izvora onečišćenja; cjevovodne mreže; modeli transporta; umjetne neuronske mreže; nasumične šume; simulacija-optimizacija

Contents

Part I

Introduction

Chapter 1

Introduction

1.1 Water Supply Network Safety and Security

Water supply networks are one of the most important systems within an urban environment as most of the human population depends on it. The monitoring of water quality inside these systems is essential since a contamination event could inflict devastating consequences to water supply networks users [59].

Contamination events in water supply networks can have a whole variety of causes. Some of the causes would be pipe bio-film formation [86, 18], aged water in pipes [63] and corrosion formation on pipe surfaces [40]. There also exists a possibility of a terrorist attack through an intentional water supply network contamination [54, 21].

To enhance the control and safety of the water supply network systems from the aforementioned hazards, various methods to simulate contamination events are developed. These methods include water quality modeling and simulation of different scenarios using hydraulic simulation engineering software (e.g. EPANET2 [66]). In reality, the transport of contaminants through pipe networks is a complex process that involves fluid mixing and should be accurately modeled to enhance the safety of water supply networks through realistic, accurate prediction of contamination events.

1.2 The Double-Tee Junction Fluid Mixing Problem

Fluid mixing in water supply networks is a complex phenomenon that has been extensively researched because it is relevant to several specific areas of application such as quality and safety of water distribution [23, 7, 14], contamination detection systems (large [39, 2] and small water supply networks [42]) and optimal placement of contamination detection sensors in the water supply network [13, 27].

The elements that make up water networks are pipes and junctions. When modeling mixing in a complex system, the correct mixing model must be applied to accurately describe the trans-

port of contamination through the water supply network due to the fact that the wrong solution could pose a danger to a large number of water supply network users. Typically, mixing in a pipe network is modeled either as complete mixing or as bulk mixing. A complete mixing model can be assumed to be accurate only if there is one outlet at the junction or if the distance between the two junctions is large enough, while bulk mixing is an idealized mixing model in which the fluid flow from two inlet pipes do not mix, i.e. the diffusion processes are neglected. Both bulk and complete mixing models aren't accurate when mixing is occurring in a double-Tee junction, which is a common segment in a water supply network.

In this dissertation, the double-Tee junction fluid mixing problem is approached in three different ways. An experimental and numerical analysis of the fluid mixing process in double-Tee junctions is given. The research proposes an experimental procedure of studying the double-Tee fluid mixing and gives new experimental results for several double-Tee junction distances which are in turn used to validate the numerical model. Two different transport models are used to model the transport of contaminant and a turbulent Schmidt number calibration process is achieved.

The previously calibrated numerical model is used to generate data in order to use a machine learning approach to model the fluid mixing behavior in the double-Tee junction. Artificial Neural Networks and Support Vector Machine models are used to generate the fluid mixing parameters in double-Tee junctions in order to achieve an accurate and computationally efficient way of modeling these phenomena.

A novel method for modeling mixing in double-Tee junction is proposed. It involves using the LES turbulence model in conjunction with the pure advection transport model which doesn't incorporate the turbulent Schmidt number parameter. The numerical approach is compared with experimental results from previous studies.

1.3 Pipe Network Fluid Mixing and Contamination Localization

In addition to accurate modeling of mixing in a double-Tee junction system, determining the source of contamination in a water supply network is an important task which also benefits from an accurate model of fluid mixing, but for most pipe water supply networks, it is sufficient to use a complete mixing model within existing hydraulic simulators (i.e. EPANET) in order to exhibit the efficiency of the method or algorithm used for the complex task of contaminant source localization.

In this research, three novel methods for solving the contamination source detection and propagation problem are proposed. The Random Forests machine learning algorithm is used for contamination source detection. The classifier is trained by data obtained from Monte Carlo hydraulic and water quality simulations made with EPANET2. The input data for the RF classifier

were the water supply network water quality sensor measurements and the output was a list of most probable potential contamination sources in the network. The method was tested on two different benchmark networks with inclusion of hydraulic demand uncertainties and imperfect sensor measurements.

Secondly, an algorithm is constructed which searches for the source of contamination, start and end time of the contamination, and the injected chemical concentration in parallel. The algorithm is specifically build for high performance, massively parallel systems and utilizes an Artificial Neural Network for classification of the most probable source node and a Random Forest regression analysis to determine the other relevant contamination event variables. The robustness and the efficiency of algorithm is tested on two different water quality benchmark networks.

Finally, the Random Forests machine learning algorithm is then further explored by coupling the Random Forest classifier with a simulation-optimization algorithm. Two different algorithmic frameworks were created to investigate the coupling. One framework was constructed by coupling the classifier with a stochastic optimization algorithm, while the other incorporated an extra Random Forest regression analysis and then was coupled with a deterministic optimization algorithm. Both frameworks were tested on two benchmark networks.

Chapter 2

Double-Tee Junction Fluid Mixing Modeling

2.1 Experimental Analysis

Water supply networks can be considered as greatly complex systems whose improper operation could have a hazardous effect to the human population which uses it. To prevent events such as water supply networks accidental or deliberate contamination, hydraulic and water quality mixing models were developed to predict the behavior of such complex systems.

EPANET [66] is the most popular simulation software used for this purpose and it was developed by the Environmental Protection Agency of the United States. EPANET provides rapid results of water quality mixing in water supply networks and it can be used in several different applications such as algorithm efficiency and accuracy investigation for optimal sensor layouts in water supply networks [55, 84], detection of water supply network contamination sources [39, 34, 72, 42], characterization of contamination type [89], water supply network quality and safety [14, 23, 7] etc.

Generally, mass transport in water supply network junctions is complex since there is an emergence of secondary currents within the flow and there are enhanced flow instabilities which in turn further enhance turbulence and fluid mixing. For this reason, correct mixing models should be used to model and predict the contamination propagation in the water supply network which is mostly consisted of pipes and junctions.

EPANET uses a complete mixing model at every junction which means that the contamination concentration transported by the fluid is the same at all junction outlet pipes, and the value of the contamination concentration is determined by the junction inlet pipe flow-weighted concentrations. While the law of mass conservation is not violated by the complete mixing model formulation, the model works correctly only if applied to single-Tee junctions, i.e. a junction where two inlet pipes converge into a single outlet pipe.

Fluid mixing in a cross junction, which is a type of double-Tee junction, is not complete and

using a complete mixing model generates great inaccuracy. When the Reynolds (Re) numbers are equal at all inlet and outlet cross junction system pipes, it was observed that the greatest deviation from complete mixing emerges [51]. Water quality models in EPANET were created with data obtained from experimental analysis and computational fluid dynamics (CFD) for the purpose of improving cross junction mixing model accuracy [65]. When the fluid flow in a cross junction is turbulent, a considerable deviation from complete mixing is also apparent [4].

Besides the complete mixing model, the bulk-advective mixing model (BAM) was developed and it is used for modeling bulk mixing. With bulk fluid mixing at cross junctions, the instabilities and diffusion at mixing fluids interface are ignored, and bulk mixing model can be considered as the opposite of the complete mixing model [28]. With the bulk mixing model, the flow momentum at the cross junction inlet is used to calculate and determine how the contamination concentrations will diverge into the cross junction outlet pipes. The pipe which transports the fluid with the higher flow momentum is going to be the dominant in propagating the contamination concentration towards the cross junction outlet pipe which has the same direction, and the pipe which transports the lower momentum fluid will propagate the contamination concentration only to its adjacent outlet pipe.

An EPANET extension labeled as EPANET-BAM [30] was created and it uses the bulk mixing model to accurately model cross junction mixing. In the EPANET-BAM software, a mixing parameter value is set to model the mixing behavior. When the value is equal to zero, the mixing model will be bulk, and if it is equal to one, the complete mixing model will be used. Values in between can be set by the user in order to model incomplete mixing at the cross junction.

Double-Tee junction fluid mixing dynamics is investigated by varying the distances between the Tee junctions, flow values at the inlet and outlet pipes, and the configuration of pipes [74, 93]. Additionally, uneven inlet and outlet pipe diameters of the double-Tee junction configuration greatly influence the mixing behavior of the fluid and larger pipe diameter discrepancies generate a more complete mixing behavior [92]. The orientation of the inlet and outlet pipes of the double-Tee junction also has an enhanced influence on the mixing behavior [76], i.e. when the branch outlet pipe is located on the same or the opposite side of the branch inlet pipe.

In this dissertation, an experimental analysis of two double-Tee junction configurations is conducted, the influence of inlet flow values and the distance between the double-Tee junctions is investigated. The purpose of the experimental analysis is to investigate mixing dynamics and obtain data in order to validate numerical models. Additionally, experimental data is also used to obtain and interpolate mixing parameters for mixing models such as BAM.

Previous research has showed that that the distance between double-Tee junctions when should be around 7.5 and 10 pipe diameters in order to reach complete mixing [74, 94]. The complete mixing distance is also investigated within this dissertation.

The schematic of the experimental analysis and the two different double-Tee junction configurations can be observed in Figures 2.1 and 2.2. The experimental setup consists of two inlet

pipes and two outlet pipes that are connected with the double-Tee junction. To investigate the fluid mixing dynamics in the double-Tee junction, the electrical conductivity of the mixture is measured at the pipe outlets. Tap water is the fluid that flows through the main pipe inlet, while the branch inlet pipe transports distilled water. Tap and distilled water greatly differ in their electrical conductivity values and will produce a mixture with specific electrical conductivity values. Beside distilled and tap water mixtures, it is also possible to use salt and tap water at the inlet pipes in order to study the mixing behavior in double-Tee junctions [31].

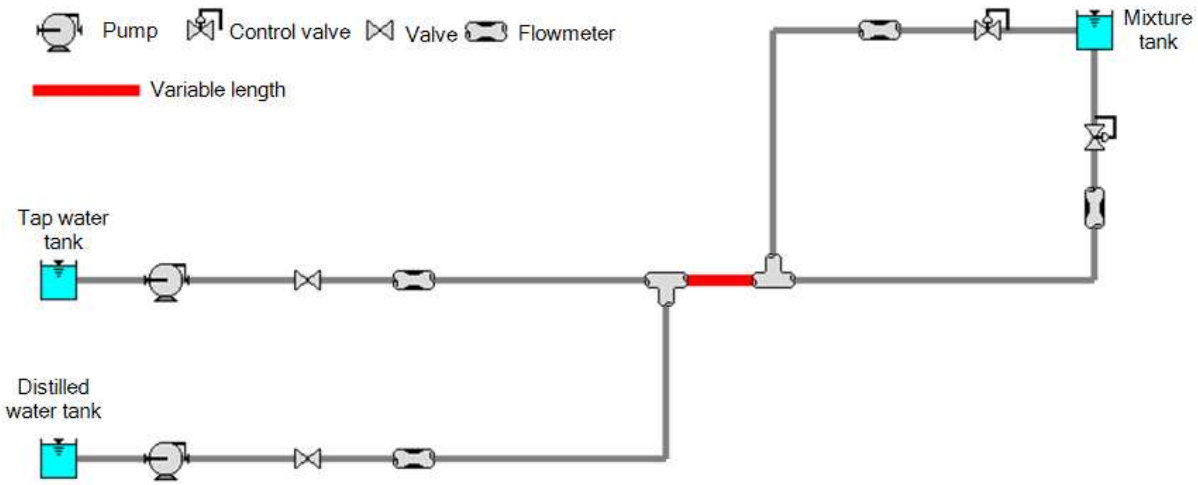

Figure 2.1: Experimental setup pipe system schematic with the double-Tee junction. [25]

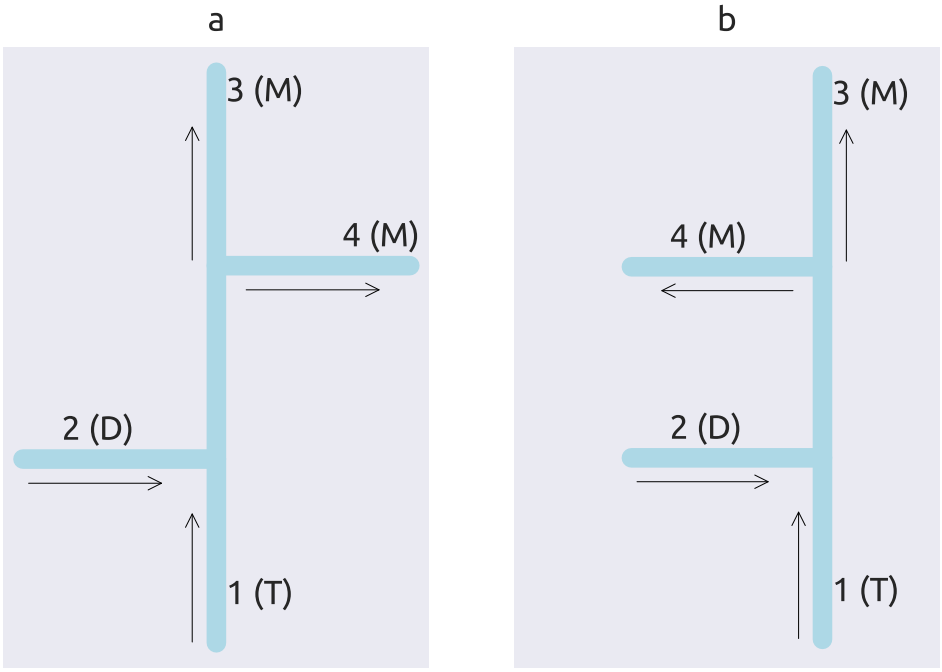

Figure 2.2: Double-Tee junction pipe configurations considered in the experimental analysis. [25]

2.2 Numerical Models

Along with an experimental approach, a CFD approach for double-Tee junction fluid mixing modeling is also possible and can provide accurate mixing results, however, CFD models need to be validated with experimentally obtained data. Double-Tee junction CFD mixing modeling of a real scale water supply network would be unpractical since it would require a massive amount of computational resources and time. CFD can be used to generate mixing parameters for simpler mixing models such as the bulk-advice mixing model which is implemented in EPANET-BAM.

When the species transport model is used, a calibration of the turbulent Schmidt number is needed and its value has been shown to be quite case-specific [26, 81]. Previous studies have reported different values of the turbulent Schmidt number when mixing in double-Tee junctions is modeled [31, 65].

A CFD approach using the Reynolds-averaged Navier-Stokes (RANS) turbulence k-Epsilon (k-ϵ) model coupled with the scalar transport model was investigated for a cross junction fluid mixing case and it was determined that adjusting the turbulent Schmidt number is a requirement to obtain values equal to those of an experimental analysis, ultimately a turbulent Schmidt number value of 0.135 was recommended [65].

A range of turbulent Schmidt number vales were also investigated for a cross junction mixing case [29] using the coupled RANS k-ϵ turbulence model with the scalar transport model.

The varied turbulent Schmidt number values were 0.001, 0.01, 0.1 and 1. A close relationship between Re and the turbulent Schmidt number was found, and that an adjustment of the turbulent Schmidt number is necessary when the double-Tee junction distance is changed.

The turbulent Schmidt number equal to 0.135 was additionally investigated for different double-Tee junction distances using the k-ϵ turbulence model and it was found that it produces a good fit with the experimental data [74].

The turbulent Schmidt number value of 0.2 was recommend for a double-Tee junction mixing model when the k-Omega (k-ω) sst turbulence model is used [11], further showing that the value is also dependant on the used turbulence model.

Finally, a high fidelity Large Eddy Simulation (LES) turbulence model coupled with the scalar transport model can also produce accurate fluid mixing results in double-Tee junctions [85].

In this dissertation, a CFD analysis is performed along with the experimental analysis. Both RANS and LES models were investigated using the open source finite volume-based CFD software OpenFOAM [37]. A multiphase and scalar transport model were both investigated with the RANS model, while the LES model was investigated coupled with a pure advection scalar transport model. The main contributions of this research is to validate the numerical models with the experimental data, investigate the efficiency of each numerical approach, calibrate and investigate the influence of the turbulent Schmidt number and generate mixing parameters for mixing models like BAM with CFD.

2.2.1 Transport Models

The double-Tee junction mixing was modeled with RANS k-ϵ turbulence model coupled with the scalar transport model and multiphase model, separately. Both models have the turbulent Schmidt number incorporated and it is defined in equation 2.1.

$$D_{\mathsf{t}} = D_{\mathsf{mol}} + {S_{\mathsf{ct}}}^{-1}\nu_{\mathsf{t}} \tag{2.1}$$

The D_{t} variable is the turbulent diffusion, D_{mol} denotes the molecular diffusion parameter, S_{ct} is the turbulent Schmidt number and ν_{t} is the kinematic turbulent viscosity which is calculated from the turbulent kinetic energy k and turbulent dissipation ϵ as seen in equation 2.2 [45].

$$\nu_{\mathsf{t}} = C_{\mu}\frac{k^2}{\epsilon} \tag{2.2}$$

The scalar transport model is defined in equation 2.3.

$$\nabla \cdot (\mathbf{u}c) = \nabla \cdot (D_{\mathsf{t}}\nabla c) \tag{2.3}$$

The c parameter denotes a dimensionless scalar which is 0 when the transported concentration is low, and 1 when it is high, and u denotes the velocity vector. The equation 2.3 is

physically interpreted as transport of a scalar quantity c by the fluid flow.

The second transport model investigated in this dissertation is the multiphase model and it is defined by equations 2.4, 2.5 and 2.6.

$$\frac{\partial \alpha_1}{\partial t} + \nabla \cdot (\mathbf{u}\alpha_1) = \nabla \cdot (D_{\mathsf{t}} \nabla \alpha_1) \tag{2.4}$$

$$\alpha_1 = 1 - \alpha_2 \tag{2.5}$$

$$\rho = \alpha_1 \rho_1 + \alpha_2 \rho_2 = \alpha_1 \rho_1 + (1 - \alpha_1)\rho_2 \tag{2.6}$$

The fluid density ρ_1 corresponds to the phase variable α_1, while ρ_2 is the second fluid's density which corresponds to the phase variable α_2. The density ρ in equation 2.6 is the density of the computed mixture of fluids or phases and it is further used as the density in the RANS k-$epsilon$ turbulence model. Similarly as the c value of the passive scalar model, α_1 is equal to 1 and it is the high transported concentration phase, while α_2 is 0 and the low concentration phase. If a phase has a value between 0 and 1 then it is a mixture.

Equation 2.4 is considered as an advection-diffusion equation for the α_1 phase, and it is transient which makes this approach more computationally inefficient than the passive scalar transport model.

In this dissertation, for the scalar transport model, a custom code was written within the OpenFOAM framework and coupled with the $simpleFoam$ solver which is RANS steady state turbulent fluid flow solver. The multiphase solver $twoLiquidMixingFoam$ was used for the multiphase transport model.

2.2.2 Large Eddy Simulation

LES turbulence model coupled with the pure advection scalar transport model was investigated for the double-Tee junction mixing problem. The pure advection scalar transport model does not incorporate the turbulent and molecular diffusion terms and it is defined in equation 2.7. The main benefit of this approach is that the model does not need any calibration since there is no S_{ct}.

$$\frac{\partial \bar{c}}{\partial t_{\mathsf{i}}} + \frac{\partial \bar{c}\bar{u}_{\mathsf{j}}}{\partial x_{\mathsf{j}}} = 0 \tag{2.7}$$

The variable $\bar{c}$ is filtered scalar quantity transported by the LES resolved flow, and $\bar{u}$ denotes the filtered velocity.

When the fluid flow is turbulent, there exists a process where the turbulent eddies of a larger length scale transfer kinetic energy on those eddies with a smaller length scale, and this process is referred to as the energy cascade. With the LES model there is a cut-off filter which directly

numerically solves the large scale eddies, while the small length scale eddies are modeled with a sub-grid scale (SGS) model. This cut-off filter is directly linked with the resolution of the numerical mesh. The Wall-Adapting Local Eddy-Viscosity (WALE) model introduced in [53] was used to model the double-Tee mixing process.

Unlike in the RANS approach, the equations solved with the LES model are the temporally and spatially filtered Navier-Stokes equations of fluid flow. In equation 2.8 the incompressible spatially and temporally filtered continuity equation is presented, while the incompressible momentum conservation equation is defined in equation 2.9.

$$\frac{\partial \bar{u}_{\mathsf{i}}}{\partial x_{\mathsf{i}}} = 0 \tag{2.8}$$

$$\frac{\partial \bar{u}_{\mathsf{i}}}{\partial t_{\mathsf{i}}} + \frac{\partial \bar{u}_{\mathsf{i}} \bar{u}_{\mathsf{j}}}{\partial x_{\mathsf{j}}} = \frac{\partial \sigma_{\mathsf{ij}}}{\partial x_{\mathsf{j}}} - \frac{\partial \bar{p}}{\partial x_{\mathsf{i}}} - \frac{\partial \tau_{\mathsf{ij}}}{\partial x_{\mathsf{j}}} \tag{2.9}$$

The variable $\bar{p}$ is the filtered pressure, while the stress tensor term σ_{ij} is present due to molecular viscosity, and τ_{ij} is the SGS stress which is modeled as defined in equation 2.11.

$$\tau_{\mathsf{ij}} = \overline{u_{\mathsf{i}} u_{\mathsf{j}}} - \bar{u}_{\mathsf{i}} \bar{u}_{\mathsf{j}} \tag{2.10}$$

Additionally, the further decomposition of τ_{ij} is decomposed defined in equation 2.11, where τ_{n} are the normal components of the SGS stress tensor, δ_{ij} denotes the Kronecker delta, ν_{t} the modeled turbulent viscosity and $\bar{S}_{\mathsf{ij}}$ the resolved scale rate of strain tensor which is defined in equation 2.12.

$$\tau_{\mathsf{ij}} - \frac{1}{3}\tau_{\mathsf{n}}\delta_{\mathsf{ij}} = -2\nu_{\mathsf{t}}\bar{S}_{\mathsf{ij}} \tag{2.11}$$

$$\bar{S}_{\mathsf{ij}} = \frac{1}{2}\left(\frac{\partial \bar{u}_{\mathsf{i}}}{\partial x_{\mathsf{j}}} + \frac{\partial \bar{u}_{\mathsf{j}}}{\partial x_{\mathsf{i}}}\right) \tag{2.12}$$

Furthermore, specifically in the LES WALE model, the variable ν_{t} is modeled with the terms defined in equation 2.13, and C_{w} is the constant present in the WALE model, Δ_{w} denotes the characteristic length scale which is dependant on the numerical mesh resolution and the term $S^{\mathsf{d}}_{\mathsf{ij}}$ is the deviation from $\bar{S}_{\mathsf{ij}}$.

$$\nu_{\mathsf{t}} = (C_{\mathsf{w}}\Delta_{\mathsf{w}})^2 \frac{(S^{\mathsf{d}}_{\mathsf{ij}} S^{\mathsf{d}}_{\mathsf{ij}})^{3/2}}{(\bar{S}_{\mathsf{ij}}\bar{S}_{\mathsf{ij}})^{5/2} + (S^{\mathsf{d}}_{\mathsf{ij}} S^{\mathsf{d}}_{\mathsf{ij}})^{5/4}} \tag{2.13}$$

Previously, the WALE model was investigated on a scalar transport within a single-Tee junction case [5, 71]. The numerical mesh for the LES WALE model was generated with the cell size requirement estimation procedure which is dependant on the RANS generated Taylor microscale values [1]. Equation 2.14 represents the estimator and Δ is the maximum cut-off value which determines the size of a numerical cell. The λ_{R} variable is the Taylor microscale and

L_R is the turbulent energy length scale. This numerical mesh procedure was investigated on a single-Tee junction case in conjunction with the Vreman SGS LES model [44] and on a fluid oscillator case [56].

$$\Delta = max\left(\lambda_R, \frac{L_R}{10}\right) \tag{2.14}$$

Furthermore, equations 2.15 and 2.16 are used to calculate the values of λ_R and L_R, respectively.

$$\lambda_R = \sqrt{\frac{10k\nu}{\epsilon}} \tag{2.15}$$

$$L_R = \frac{k^{1.5}}{\epsilon} \tag{2.16}$$

The purpose of modeling the double-Tee junction mixing problem with the LES WALE model and the pure advection scalar transport model is to investigate whether it is possible to obtain accurate mixing results without S_{ct} calibration. In this dissertation, the numerical mesh RANS estimation and LES WALE modeling was done in OpenFOAM.

2.3 Machine Learning Approach

A machine learning approach to fluid mixing modeling in double-Tee junctions is also possible besides the experimental and numerical approaches.

Recently, machine learning (ML) algorithms have been applied in areas like hydrology and hydraulics, and those include ANN [46, 79], RF [68], and Support Vector Machines (SVM) [19]. ML was also implemented in pipe flow systems, where a selection of ML algorithms such as RF, Bagging Algorithm (BA), and Regression Trees (RT) were trained to predict characteristics of a wastewater pipe system [22].

ML algorithms were used to investigate pressure gradients, flow pattern identification and liquid holdup in a pipe segment. The ML algorithms were trained with data generated by CFD and it was found that the Gradient Boosting (GB) algorithm and SVM produce the best predictions [38]. Deep learning (i.e. ANN) was also applied to predict the best valve scheduling scenario in a water supply network during a contamination event [33]. Similarly, the Support Vector Regression (SVR) algorithm was trained to find anomalies in a pressurized pipe network system [83], while a modified ANN algorithm was trained for pipe network management [70].

A data analysis approach using Kriging and Delaunay triangulation in conjunction was used for double-Tee junction mixing modeling [20], where the mixing dynamics prediction was done for several double-Tee junction distances and for different flow regimes with data obtained by CFD simulations.

A ML approach is also possible for mixing behavior evaluation with data obtained by exper-

imental analysis of by CFD. In this dissertation, two different algorithms (SVR and ANN) were investigated for this purpose. Data for ML algorithm training was generated by a CFD model which was validated with experimental data. The prediction of both ML algorithms was tested on different double-Tee junction distances and different inlet and outlet pipe flow values. The SVR and ANN algorithm implementations in the Python machine learning module scikit-learn was used in this dissertation [57], and the CFD data was generated with the open source CFD solver OpenFOAM.

The main purpose of this research is to investigate whether a computationally efficient method like ML is accurate enough for predicting mixing behavior parameters. An efficient method like ML has the possibility to be directly implemented into hydraulic and water quality software like EPANET, and could be used in real time due to its rapid prediction capabilities.

Chapter 3

Water Supply Network Source Contamination Localization

3.1 The Challenge of Contamination Localization

Contamination events in water supply networks are a great concern as they pose a major threat on the health of the human population which uses the network, hence, it is essential that water supply networks are functional [9]. Contamination of water supply network can be accidental or deliberate and should be rapidly terminated by the authorities that govern water supply networks since contamination can quickly propagate through the water supply network. A number of recent studies investigated emergency reactions needed in case of such events ([73, 77]).

Water supply network contamination is usually caused by a great variety of events such as pipe biofilm formation, aging of water and pipe lining chemical contamination and corrosion [52, 36]. Localizing the contamination source of a contamination event in a water supply network is an important and challenging task as it requires a lot of information of the water supply network state such as hydraulic demands.

Additionally, a contamination event is a complex problem due to complex physics which are involved in the contamination transport through fluid flow. As it was previously mentioned, this complexity is greatly enhanced in water supply elements such as double-Tee junctions. Experimental analysis showed that in the case of a less turbulent regime of fluid flow, diffusion of contamination in the water supply network is greatly amplified and recently, water supply network mixing models try to incorporate these effects [58]. Another possibility to tackle the water supply network contamination event diffusion problem is to use a combination of CFD and statistical methods to incorporate the effects of incomplete mixing in problematic network elements such as double-Tee junctions [11].

Obtaining information from water quality sensors positioned in the water supply networks is essential for contamination source localization. An optimal placement of water quality sensors in a pipe network is essential for finding the source and determining the contamination propagation

dynamics in the water supply network, and a lot of previous research was focused on this issue [8, 55, 62, 43, 27, 80, 32, 69]. More specifically, an optimal sensor placement in a water supply network needs to maximize the detection rate and minimize the detection time of contamination sources.

Data that was obtained with water quality sensor measurements are used with optimization algorithms to determine the most probable source of contamination in the water supply network, the contamination event start and end time and the contaminant chemical concentration [60, 82, 95, 42, 87, 91]. It is also important to note that sensor measurements can be imperfect and that hydraulic demand uncertainties can be present in the water supply network, thus several studies have included this in the data used for source detection [82, 87, 91].

The methods which involve optimization algorithms are named usually denoted as simulation-optimization methods and are the most popular set of methods for water supply network contamination source localization problem. Generally, this algorithmic procedure involves coupling a stochastic or deterministic optimization algorithm with a hydraulic and water quality water supply network fluid flow simulator (such as EPANET). The objective function of the simulation-optimization algorithmic loop is to minimize the error between the measured water quality in the water supply network and the hydraulic simulator values in order to determine the contamination source, start and end times of the contamination event and the contaminant chemical concentration. Historically, Genetic Algorithms (GA) have been widely used as the optimization algorithm in the simulation-optimization loop [61, 90, 34, 88, 50].

The simulation-optimization approach is usually computationally expensive and requires a significant amount of time to detect the water supply network contamination source, starting and ending times of the event and the contaminant chemical concentration. This is even more apparent when a real sized water supply network is being considered, as it usually consists of a greater number of potential contamination source nodes. The increased size of the water supply network increases the complexity of the given problem and can affect the accuracy and the detection time, thus slowing the process of contamination event halting.

Data mining is another approach for contamination source identification. Usually, detecting a contamination source with data mining implies searching for a specific contamination event through a pre-made database of simulated results of contamination events. Statistical and optimization algorithms can be used in conjunction with data mining for the purpose of searching thorough a pre-made database for the most accurate contamination event variables which include the source, duration (start and end time) and contaminant chemical concentration.

Optimization and statistical algorithms can be applied with data mining in order to search through the precompiled database for the most compatible contamination event parameters (source, concentration, duration), while simulation-optimization methods simultaneously run contamination event scenarios in conjunction with optimization algorithms and repeatedly evaluate a fitness function in order to detect the previously mentioned parameters.

A data mining approach was developed in conjunction with a maximum likelihood method

for a water supply network with a sensor layout which consisted of five sensors and it was showed that it is possible to minimize the list of potential contamination sources in the network [35].

The data mining method requires a large database of simulated contamination event scenario results, and this can be considered a drawback, even though it was showed that an offline database which can be successfully used to correctly identify the water supply network contamination source [75].

Data-driven approach based on the logistic regression (LR) algorithm was investigated for the water supply network contamination detection problem. LR was coupled with an evolutionary algorithm in order to maximize computational efficiency and accuracy [48]. Similarly, a Monte Carlo (MC) based data-driven model was created for the same purpose [16]. Probabilistic methods based on Bayesian Belief Networks (BBN) have also been investigated and explored for contamination source detection in water supply networks.

ML methods have been increasingly successfully used for a great variety of environmental engineering related problems and phenomena. Long Short-Term Memory (LSTM) Neural Network was trained to forecast floods with rainfall and discharge data [46]. ANN and RF were used in conjunction to localize chemical leaks with data obtained from sensors [12]. Groundwater flow modeling has also been actively including machine learning methods as Convolutional Neural Network (CNN) coupled with a Markov Chain Monte Carlo (MCMC) method has been applied to determine contaminant sources in groundwater area flows [96]. RF was also used to create prediction models for contamination source detection in rivers with MCsimulation generated data [47]. Groundwater nitrate contamination areas was modeled with the RF algorithm which was trained with data obtained from field measurements [64]. Similarly, RF models were created with measured data for heave metal contamination from multiple sources [**?**].

Machine learning and simulation-optimization coupling has been also employed in the area of groundwater pollution source and pollution characteristics prediction. Coupling of non-dominated sorting genetic algorithm II (NSGA-II) and both Probabilistic Support Vector Machines (PSVM) and Probabilistic Neural Networks (PNN) has been done for characterizing an unknown pollution source in groundwater resources systems [6].

ML models are another possible approach which can provide an extremely fast and efficient way of prediction potential contamination sources in water supply networks. An ANN was trained and tested on data obtained from experimental analysis of a small-sized pipe network system with 5 sampling locations to detect the source E. Coli bacterial [41]. Even though the experimental system is small-sized, the complexity of the problem is apparent, and a detection accuracy of 87% was realized. Learning vector quantization Neural Network (LVQNN) have been used to detect zones in large water supply networks where contamination source nodes could be located [67].

More recently, a CNN was trained with water supply network user complaint data in order to detect contamination sources [78]. It was compared to a Multilayer Perceptron ANN (MLP-

ANN) and it was found that CNN performs better. Decision Trees (DT) were trained with water supply network contamination event data in order to isolate a problematic area [17]. Similarly, RF was used to determine the number of water supply network contamination sources [49].

Generally, the water supply network contamination source detection is a complex and multi-modal problem which requires a lot more research as a lot of proposed solutions still have drawbacks either due to inaccuracy or low computational efficiency. It is why in this research, the usage of ML methods for the water supply network contamination source detection is further explored and investigated on benchmark water supply networks. The RF and ANN algorithms are used to create ML prediction models. Additionally, a coupling between simulation-optimization (stochastic and deterministic) and ML algorithms is achieved and investigated. To examine the accuracy and robustness of proposed methods and frameworks, the EPANET2 water quality and hydraulic simulator is used.

3.2 Monte Carlo Simulations and Machine Learning

MC simulations using the hydraulic and water quality analysis software were made in order to obtain data which can be used to train and investigating ML algorithms for the purpose of water supply network contamination source detection. EPANET2, the water quality and hydraulic software which uses the complete mixing model at junctions was used to generate training and testing data. The accuracy of the ML approach does not depend on the mixing model used by the hydraulic simulator, however, in order for the approach to be employed in a real case, the contaminant propagation and mixing should be as accurate as possible.

Each MC simulation used a randomized set of water supply network contamination event parameters which include contamination source location, start and end times and the contaminant chemical concentration. Generally, the input values for the ML model are the simulated sensor water quality measurements, while the output values is the location of the contamination source and other relevant parameters that are needed in order to reconstruct the contamination event.

In this thesis the RF algorithm (implemented in the Python ML module sickit-learn [57]) was trained with MC simulated data and investigated for the purpose of predicting the contamination source in two water quality benchmark networks (seen in Figures 3.1 and 3.2) with different sensor layouts. Hydraulic demand uncertainty and imperfect sensor measurements were also investigated to test the accuracy and robustness of the approach.

Furthermore, besides the RF model used for prediction of contamination sources, in this dissertation, a hybrid algorithm is proposed which uses ANN for source location classification within the water supply network and RF for prediction of contamination start and end time, and contaminant chemical concentration. The algorithm is specifically made for high performance computing systems and uses a tournament-based selection of most probable contamination source nodes based on MC water quality simulations in parallel. The proposed hybrid

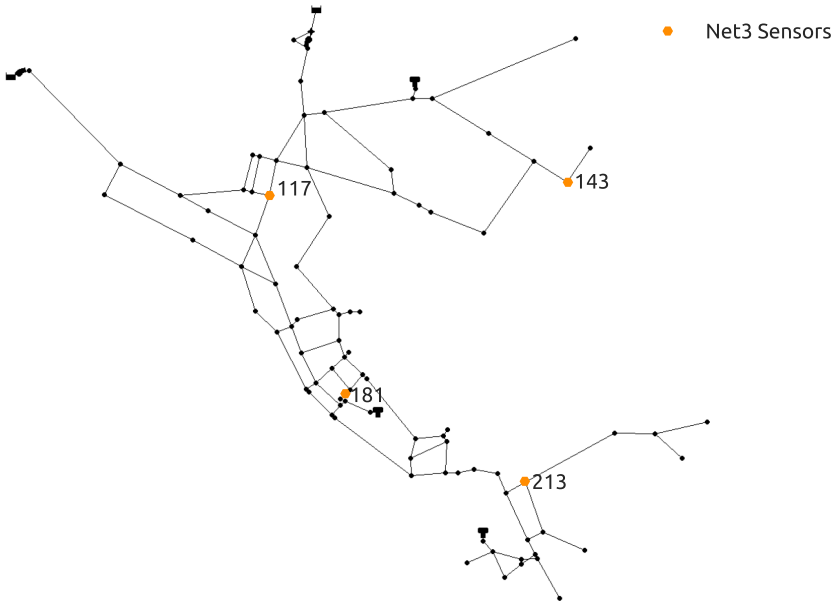

Figure 3.1: Water supply network benchmark Net3 (92 nodes) with a 4 sensor layout. [24]

algorithm is created using the Python ML module scikit-learn [57] and the Simple Linux Utility for Resource Management (SLURM) utility. The algorithm is also tested on the benchmark networks presented in Figures 3.1 and 3.2.

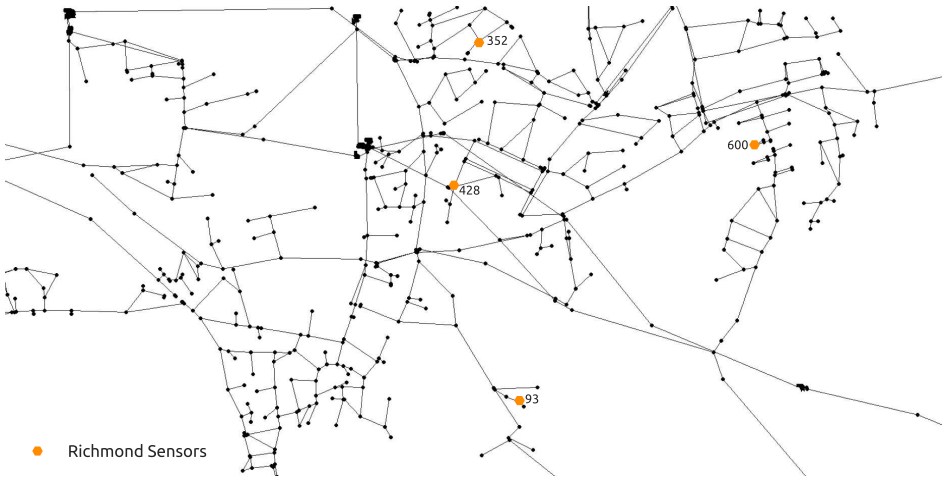

Figure 3.2: Water supply network benchmark Richmond (865 nodes) with a 5 sensor layout. [24]

3.3 Simulation-Optimization and Machine Learning

Finally, the coupling of the simulation-optimization approach and ML approach has been investigated in this dissertation. The goal is to investigate whether a coupled approach could eliminate the drawbacks of both methods when used separately. Two different novel coupling algorithmic frameworks are proposed and both use the MC simulation trained RF algorithm (scikit-learn [57] implementation) to reduce the space of potential contamination source nodes within the water supply network.

The first proposed algorithmic framework uses the simulation-optimization procedure loop with the RF classification in order to determine the contamination event parameters. Three different stochastic optimization algorithms (Particle Swarm Optimization (PSO), Fireworks Algorithm (FWA) and GA) are assessed within the first algorithmic framework in terms of efficiency and accuracy. PSO and FWA implenetations in the Python numerical optimization module indago [15], while the GA Python module for multiobjective optimization pymoo [10] implementation is used.

The second proposed algorithmic framework includes an additional RF model regression procedure for each RF classified potential contamination source location for the purpose of predicting each location's start time, end time and contaminant chemical concentration. After the

RF regression procedure, each contamination source node's predicted contamination start and end times, and concentration data is then used as starting values for the deterministic optimization algorithm Mesh Adaptive Direct Search (MADS) which is implemented in NOMAD 4.0 [3].

The EPANET2 hydraulic and water quality simulator is also used in the simulation-optimization loop of the algorithmic framework, as well as for generating the data for the RF classifier which is a part of both frameworks. The algorithmic frameworks are investigated on water supply benchmark networks seen in Figures 3.1 and 3.2.

Chapter 4

Conclusion

4.1 Main Contributions

This dissertation focuses on the mixing of fluids in pipe water supply networks. The main connection between all investigated topics is the modeling of fluid mixing. Mixing models are used to investigate the fluid mixing dynamics in water supply network elements such as double-Tee junctions and mixing models are used in conjunction with ML and optimization methods to create surrogate mixing models in large-scale water supply networks for the purpose of detecting contamination sources and other relevant contamination event parameters.

The main contributions of this research can be summarized into the following points:

- Through experimental analysis of fluid flow in a double-Tee junction system, new experimental data was obtained for different double-Tee distances, configurations, and inlet flow ratios. The obtained experimental results were compared with results obtained by previous research and it was shown that complete mixing is not occurring for the investigated cases. Additionally, experimental analysis was done for previously untested larger double-Tee distances. Also, new EPANET-BAM correction factors were recommended in this research for the investigated cases. Numerical models were also proposed as two different transport models were investigated (multiphase and passive scalar) and validated based on experimental data. It was found that both models are very accurate and the importance of calibrating the turbulent Schmidt number was emphasized. The value of the turbulent Schmidt number used in the study was 0.5 and it showed to be accurate for all of the numerically reproduced cases.

- A new approach to modeling fluid mixing in a double-Tee junction system based on machine learning was explored. The purpose of this research was to investigate a more computationally efficient approach (when compared to CFD or experimental analysis) to quickly obtain the parameters which determine the mixing dynamics in a double-Tee junction. The proposed machine learning model was trained with data obtained from

CFD analysis (the calibrated numerical model from the previous point was used). The input data for the machine learning model training were the inlet pipe flow ratio, the outlet pipe flow ratio and the distance between the double-Tee junction, while the output data was the relevant mixing ratio which could be used to determine a correction factor for simpler models. Two machine learning algorithms were investigated and compared (Artificial Neural Networks and Support Vector Machines) and it was found that Artificial Neural Networks perform better in terms of accuracy and problem generalization. It was found that the approach is computationally efficient and robust.

- The possibility of using the Large Eddy Simulation turbulence model with a pure advection scalar transport model on the double-Tee junction fluid mixing case was explored and compared with experimental results from previous research. The LES-WALE turbulent model was used with a grid size estimation method based on the RANS k-Epsilon model results through calculating turbulence scale values. When a pure advection transport model is used it is not necessary to calibrate the turbulent Schmidt number (since the equation does not incorporate the turbulent diffusion term) and it was shown that accurate mixing results were obtained with this model when used with LES-WALE. The numerical approach was explored and shown to be accurate for three different double-Tee junction configurations.

- A new approach to identify contamination sources in water supply networks based on the Random Forest algorithm was proposed and explored. The method was tested on two benchmark water distribution networks with different sizes and different sensor layouts. For each network, a large number of Monte Carlo contamination scenarios were simulated in the EPANET2 hydraulic simulator using a complete mixing model with random parameters such as contamination source location, start and end times of contamination and the concentration of the injected contamination. For each scenario, water quality sensor readings were saved and used as an input to the machine learning model, while the model output was the contamination source. The efficiency of the water network sensor layouts was investigated along with added hydraulic uncertainties and sensor imperfections. It was found that the RF classifier method is robust and efficient in creating a list of the most probable contamination sources for a given value of sensor measurements.

- A new algorithmic framework based on machine learning for the detection of contamination sources in water supply networks made especially for high-performance parallel systems was developed and investigated and presented. The algorithm uses a combination of Artificial Neural Networks to classify contamination sources and the Random Forest algorithm for regression analysis to determine significant contamination event variables such as start and end times and injected contamination concentration. The algorithm is based on parallel execution of Monte Carlo hydraulic simulations in EPANET2 of the water supply network and classifying the most probable source of contamination in a parallel

tournament selection with Artificial Neural Networks (potential sources are distributed to tournaments on CPU cores). After the tournament selection of the most probable sources, the Random Forest algorithm is used to predict other relevant variables, and all nodes are ranked according to the prediction error based on the sensor measurements data. The algorithm was tested on a smaller network of 92 potential sources and on a medium-sized network with 865 potential contamination sources. It was shown that the proposed algorithm is robust and computationally efficient when employed on high performance systems.

- Finally, a novel methodology is explored and presented for solving the problem of a water supply network contamination event, which includes determining the exact source of contamination, the contamination start and end times and the injected contaminant concentration. Two slightly different algorithmic frameworks were constructed based on the novel methodology. Both algorithmic frameworks utilize the Random Forest algorithm for classification of top source contamination node candidates, with one of the frameworks directly using the stochastic fireworks optimization algorithm to determine the contamination start time, end time and injected contaminant concentration for each predicted node separately. The second framework uses the Random Forest algorithm for an additional regression prediction of each top node's start time, end time and contaminant concentration and is then coupled with the deterministic global search optimization algorithm MADS. Both algorithmic frameworks perform well and show robustness for the true source node detection, start and end times and contaminant concentration, with the second framework being extremely efficient on the fuzzy sensor measurement benchmark network.

4.2 Future Work

Possible future work and expansions of the research presented in this dissertation includes:

- The turbulent Schmidt number calibrated RANS and passive scalar transport model is a useful tool which can be used to simulate other double-Tee junction distances, configurations and inflow values. It can used to generate more incomplete mixing parameters for hydraulic and water quality software such as EPANET-BAM without the need of additional experimental analysis if the flow regime is in the same range. Similarly, the LES WALE pure advection model in conjunction with the numerical mesh size estimation procedure, could be used for the same purpose but with more certainty as it shows that it doesn't require any turbulent Schmidt number calibration.

- The double-Tee junction ML model could be further trained and investigated with other ML algorithms and additional double-Tee junction configurations. The mixing surrogate model could be directly implemented into a water supply network mixing modeling

software as it could provide accurate mixing assessment for a wide array of double-Tee junction mixing combinations due to its efficiency and robustness.

- The water supply network contamination source detection algorithmic frameworks could be further explored with novel and cutting ML algorithms. Additionally, the methods should be investigated on real water supply networks which include hydraulic uncertainties of all pipe network elements in order to realistically assess their efficiency and accuracy.

Chapter 5

Summary of Papers

A Experimental and Numerical Investigation of Mixing Phenomena In Double-Tee Junctions

This work investigates mixing phenomena in a pressurized pipe system with two sequential Tee junctions and experiments are conducted for a range of different inlet flow ratios, varying distances between Tee junctions and two pipe branching configurations. Additionally, obtained experimental results are compared with results from previous studies by different authors and are used to validate the numerical model using the open source computational fluid dynamics toolbox OpenFOAM. Two different numerical approaches are used—Passive scalar model and Multiphase model. It is found that both numerical models produce similar results and that they are both greatly dependent on the turbulent Schmidt number. After the calibration procedure, both models provided good results for all investigated flow ratios, double-Tee junction distances, and pipe branching configurations, therefore both numerical models can be applied for a wide range of pipe networks configurations, but passive scalar model is the viable choice due to its much higher computational efficiency. Obtained results also describe the relationship between the double-Tee distances and complete mixing occurrence.

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.; https://doi.org/10.3390/w11061198

B Efficient Double-Tee Junction Mixing Assessment by Machine Learning

A new approach in modeling of mixing phenomena in double-Tee pipe junctions based on machine learning is presented in this paper. Machine learning represents a paradigm shift that can be efficiently used to calculate needed mixing parameters. Usually, these parameters are obtained either by experiment or by computational fluid dynamics (CFD) numerical modeling. A machine learning approach is used together with a CFD model. The CFD model was calibrated with experimental data from a previous study and it served as a generator of input data for the machine learning metamodels—Artificial Neural Network (ANN) and Support Vector Regression (SVR). Metamodel input variables are defined as inlet pipe flow ratio, outlet pipe flow ratio, and the distance between the pipe junctions, with the output parameter being the branch pipe outlet to main inlet pipe mixing ratio. A comparison of ANN and SVR models showed that ANN outperforms SVR in accuracy for a given problem. Consequently, ANN proved to be a viable way to model mixing phenomena in double-Tee junctions also because its mixing prediction time is extremely efficient (compared to CFD time). Because of its high computational efficiency, the machine learning metamodel can be directly incorporated into pipe network numerical models in future studies.

Grbčić, L., Kranjčević, L., Družeta, S. and Lučin, I., 2020. Efficient double-Tee junction mixing assessment by machine learning. Water, 12(1), p.238.; https://doi.org/10.3390/w12010238

C Large Eddy Simulation of Turbulent Fluid Mixing in Double-Tee Junctions

Double-Tee junctions serve as building blocks of pipe network systems and fluid mixing occurring in them is complex due to excessive eddying of flow. In this paper a LES model with pure advection was used on a double-Tee junction turbulent mixing phenomena problem. This kind of approach does not require the turbulent Schmidt number which is problem specific and needs to be calibrated with experimental data. When applying the LES-WALE turbulence model in conjunction with the pure advection transport of a scalar quantity on the double-Tee mixing problem it is achieved that the resolved turbulent flow field accurately transports the scalar when compared to experimental data from literature. This approach enables accurate mixing predictions in double-Tee junctions for the purpose of correcting simpler 1D numerical mixing models without the need of obtaining new experimental data since no turbulent Schmidt number calibration is needed. The approach can be applied to biological or chemical agent transport in a fluid and the model was tested and shown to be valid for a Reynolds number range of 20,000–400,000 on three different types of double-Tee junction configurations and five double-Tee distances with a constant inlet and outlet pipe diameter.

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.; https://doi.org/10.1016/j.asej.2020.06.004

D Water Supply Network Pollution Source Identification by Random Forest Algorithm

A novel approach for identifying the source of contamination in a water supply network based on the random forest classifying algorithm is presented in this paper. The proposed method is tested on two different water distribution benchmark networks with different sensor placements. For each considered network, a considerable amount of contamination scenarios with randomly selected contamination parameters were simulated and water quality time series of network sensors were obtained. Pollution scenarios were defined by randomly generated pollution source location, pollution starting time, duration of injection and the chemical intensity of the pollutant. Sensor layout's influence, demand uncertainty and imperfect sensor measurements were also investigated to verify the robustness of the method. The proposed approach shows high accuracy in localizing the potential sources of pollution, thus greatly reducing the complexity of the water supply network contamination detection problem.

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.; https://doi.org/10.2166/hydro.2020.042

E A Machine Learning-based Algorithm for Water Network Contamination Source Localization

In this paper, a novel machine learning based algorithm for water supply pollution source identification is presented built specifically for high performance parallel systems. The algorithm utilizes the combination of Artificial Neural Networks for classification of the pollution source with Random Forests for regression analysis to determine significant variables of a contamination event such as start time, end time and contaminant chemical concentration. The algorithm is based on performing Monte Carlo water quality and hydraulic simulations in parallel, recording data with sensors placed within a water supply network and selecting a most probable pollution source based on a tournament style selection between suspect nodes in a network with mentioned machine learning methods. The novel algorithmic framework is tested on a small (92 nodes) and medium sized (865 nodes) water supply sensor network benchmarks with a set contamination event start time, end time and chemical concentration. Out of the 30 runs, the true source node was the finalist of the algorithm's tournament style selection for 30/30 runs for the small network, and 29/30 runs for the medium sized network. For all the 30 runs on the small sensor network, the true contamination event scenario start time, end time and chemical concentration was set as 14:20, 20:20 and 813.7 mg/L, respectively. The root mean square errors for all 30 algorithm runs for the three variables were 48 min, 4.38 min and 18.06 mg/L. For the 29 successful medium sized network runs the start time was 06:50, end time 07:40 and chemical concentration of 837 mg/L and the root mean square errors were 6.06 min, 12.36 min and 299.84 mg/L. The algorithmic framework successfully narrows down the potential sources of contamination leading to a pollution source identification, start and ending time of the event and the contaminant chemical concentration.

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.; https://doi.org/10.3390/s20092613

F Machine Learning and Simulation-Optimization Coupling for Water Distribution Network Contamination Source Detection

This paper presents and explores a novel methodology for solving the problem of a water distribution network contamination event, which includes determining the exact source of contamination, the contamination start and end times and the injected contaminant concentration. The methodology is based on coupling a machine learning algorithm for predicting the most probable contamination sources in a water distribution network with an optimization algorithm for determining the values of contamination start time, end time and injected contaminant concentration for each predicted node separately. Two slightly different algorithmic frameworks were constructed which are based on the mentioned methodology. Both algorithmic frameworks utilize the Random Forest algorithm for classification of top source contamination node candidates, with one of the frameworks directly using the stochastic fireworks optimization algorithm to determine the contamination start time, end time and injected contaminant concentration for each predicted node separately. The second framework uses the Random Forest algorithm for an additional regression prediction of each top node's start time, end time and contaminant concentration and is then coupled with the deterministic global search optimization algorithm MADS. Both a small sized (92 potential sources) network with perfect sensor measurements and a medium sized (865 potential sources) benchmark network with fuzzy sensor measurements were used to explore the proposed frameworks. Both algorithmic frameworks perform well and show robustness in determining the true source node, start and end times and contaminant concentration, with the second framework being extremely efficient on the fuzzy sensor measurement benchmark network.

Grbčić, L., Kranjčević, L. and Družeta, S., 2021. Machine learning and simulation-optimization coupling for water distribution network contamination source detection. Sensors, 21(4), p.1157.; https://doi.org/10.3390/s21041157

Bibliography

[1] Addad, Y., Gaitonde, U., Laurence, D., and Rolfo, S. Optimal unstructured meshing for large eddy simulations. In *Quality and reliability of large-eddy simulations*. Springer, 2008, pp. 93–103.

[2] Adedoja, O., Hamam, Y., Khalaf, B., and Sadiku, R. Towards development of an optimization model to identify contamination source in a water distribution network. *Water 10*, 5 (2018), 579.

[3] Audet, C., Le Digabel, S., and Tribes, C. NOMAD user guide. Tech. Rep. G-2009-37, Les cahiers du GERAD, 2009.

[4] Austin, R., Waanders, B. v. B., McKenna, S., and Choi, C. Mixing at cross junctions in water distribution systems. ii: Experimental study. *Journal of Water Resources Planning and Management 134*, 3 (2008), 295–302.

[5] Ayhan, H., and Sökmen, C. N. Cfd modeling of thermal mixing in a t-junction geometry using les model. *Nuclear Engineering and Design 253* (2012), 183–191.

[6] Bashi-Azghadi, S. N., Kerachian, R., Bazargan-Lari, M. R., and Solouki, K. Characterizing an unknown pollution source in groundwater resources systems using psvm and pnn. *Expert Systems with Applications 37*, 10 (2010), 7154–7161.

[7] Bello, O., Abu-Mahfouz, A. M., Hamam, Y., Page, P. R., Adedeji, K. B., and Piller, O. Solving management problems in water distribution networks: A survey of approaches and mathematical models. *Water 11*, 3 (2019), 562.

[8] Berry, J., Hart, W. E., Phillips, C. A., Uber, J. G., and Watson, J.-P. Sensor placement in municipal water networks with temporal integer programming models. *Journal of water resources planning and management 132*, 4 (2006), 218–224.

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

[10] Blank, J., and Deb, K. pymoo: Multi-objective optimization in python. *IEEE Access 8* (2020), 89497–89509.

[11] Braun, M., Bernard, T., Ung, H., Piller, O., and Gilbert, D. Computational fluid dynamics modeling of contaminant mixing at junctions for an online security management toolkit in water distribution networks. *Journal of Water Supply: Research and Technology-Aqua 64*, 5 (2015), 504–515.

[12] Cho, J., Kim, H., Gebreselassie, A. L., and Shin, D. Deep neural network and random forest classifier for source tracking of chemical leaks using fence monitoring data. *Journal of Loss Prevention in the Process Industries 56* (2018), 548–558.

[13] Ciaponi, C., Creaco, E., Di Nardo, A., Di Natale, M., Giudicianni, C., Musmarra, D., and Santonastaso, G. F. Reducing impacts of contamination in water distribution networks: A combined strategy based on network partitioning and installation of water quality sensors. *Water 11*, 6 (2019), 1315.

[14] Di Nardo, A., Di Natale, M., Guida, M., and Musmarra, D. Water network protection from intentional contamination by sectorization. *Water Resources Management 27*, 6 (2013), 1837–1850.

[15] Družeta, S., and Ivić, S. Indago - python module for numerical optimization. `https://pypi.org/project/Indago/`, 2020. Accessed: September 10, 2020.

[16] Eliades, D., Lambrou, T., Panayiotou, C. G., and Polycarpou, M. M. Contamination event detection in water distribution systems using a model-based approach. *Procedia Engineering 89* (2014), 1089–1096.

[17] Eliades, D. G., and Polycarpou, M. M. Water contamination impact evaluation and source-area isolation using decision trees. *Journal of Water Resources Planning and Management 138*, 5 (2011), 562–570.

[18] Flemming, H.-C., Percival, S., and Walker, J. Contamination potential of biofilms in water distribution systems. *Water science and technology: water supply 2*, 1 (2002), 271–280.

[19] Gedik, N. Least squares support vector mechanics to predict the stability number of rubble-mound breakwaters. *Water 10*, 10 (2018), 1452.

[20] Gilbert, D., Mortazavi, I., Piller, O., and Ung, H. Low dimensional modeling of double t-junctions in water distribution networks using kriging interpolation and delaunay triangulation. *Pacific Journal of Mathematics for Industry 9*, 1 (2017), 2.

[21] Gleick, P. H. Water and terrorism. *Water policy 8*, 6 (2006), 481–503.

[22] Granata, F., and de Marinis, G. Machine learning methods for wastewater hydraulics. *Flow Measurement and Instrumentation 57* (2017), 1–9.

[23] Grayman, W. M., Murray, R., and Savic, D. A. Effects of redesign of water systems for security and water quality factors. In *World Environmental and Water Resources Congress 2009: Great Rivers* (2009), pp. 1–11.

[24] Grbčić, L., Kranjčević, L., and Družeta, S. Machine learning and simulation-optimization coupling for water distribution network contamination source detection. *Sensors 21*, 4 (2021), 1157.

[25] Grbčić, L., Kranjčević, L., Lučin, I., and Čarija, Z. Experimental and numerical investigation of mixing phenomena in double-tee junctions. *Water 11*, 6 (2019), 1198.

[26] Gualtieri, C., Angeloudis, A., Bombardelli, F., Jha, S., and Stoesser, T. On the values for the turbulent schmidt number in environmental flows. *Fluids 2*, 2 (2017), 17.

[27] Hart, W. E., and Murray, R. Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *Journal of Water Resources Planning and Management 136*, 6 (2010), 611–619.

[28] Ho, C. K. Solute mixing models for water-distribution pipe networks. *Journal of Hydraulic Engineering 134*, 9 (2008), 1236–1244.

[29] Ho, C. K., Orear, Jr, L., Wright, J. L., and McKenna, S. A. Contaminant mixing at pipe joints: Comparison between laboratory flow experiments and computational fluid dynamics models. In *Water Distribution Systems Analysis Symposium 2006* (2008), pp. 1–18.

[30] Ho, C. K., and O'Rear Jr, L. Evaluation of solute mixing in water distribution pipe junctions. *Journal-American Water Works Association 101*, 9 (2009), 116–127.

[31] Ho, C. K., Wright, J. L., McKenna, S. A., and Orear Jr, L. Contaminant mixing at pipe joints: Comparison between laboratory flow experiments and computational fluid dynamics models. Tech. rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2006.

[32] Hooshmand, F., Amerehi, F., and MirHassani, S. Risk-based models for optimal sensor location problems in water networks. *Journal of Water Resources Planning and Management 146*, 11 (2020), 04020086.

[33] Hu, C., Cai, J., Zeng, D., Yan, X., Gong, W., and Wang, L. Deep reinforcement learning based valve scheduling for pollution isolation in water distribution network. *Mathematical Biosciences and Engineering 17*, 1 (2019), 105.

[34] Hu, C., Zhao, J., Yan, X., Zeng, D., and Guo, S. A mapreduce based parallel niche genetic algorithm for contaminant source identification in water distribution network. *Ad Hoc Networks 35* (2015), 116–126.

[35] Huang, J. J., and McBean, E. A. Data mining to identify contaminant event locations in water distribution systems. *Journal of Water Resources Planning and Management 135*, 6 (2009), 466–474.

[36] Inkinen, J., Kaunisto, T., Pursiainen, A., Miettinen, I. T., Kusnetsov, J., Riihinen, K., and Keinänen-Toivola, M. M. Drinking water quality and formation of biofilms in an office building during its first year of operation, a full scale study. *Water research 49* (2014), 83–91.

[37] Jasak, H., Jemcov, A., Tukovic, Z., et al. Openfoam: A c++ library for complex physics simulations. In *International workshop on coupled methods in numerical dynamics* (2007), vol. 1000, IUC Dubrovnik Croatia, pp. 1–20.

[38] Kanin, E., Vainshtein, A., Osiptsov, A., and Burnaev, E. The method of calculation the pressure gradient in multiphase flow in the pipe segment based on the machine learning algorithms. In *IOP Conference Series: Earth and Environmental Science* (2018), vol. 193, IOP Publishing, p. 012028.

[39] Kessler, A., Ostfeld, A., and Sinai, G. Detecting accidental contaminations in municipal water networks. *Journal of Water Resources Planning and Management 124*, 4 (1998), 192–198.

[40] Kim, E. J., and Herrera, J. E. Characteristics of lead corrosion scales formed during drinking water distribution and their potential influence on the release of lead and other contaminants. *Environmental science & technology 44*, 16 (2010), 6054–6061.

[41] Kim, M., Choi, C. Y., and Gerba, C. P. Source tracking of microbial intrusion in water systems using artificial neural networks. *Water research 42*, 4-5 (2008), 1308–1314.

[42] Kranjčević, L., Čavrak, M., and Šestan, M. Contamination source detection in water distribution networks. *Engineering Review 30*, 2 (2010), 11–25.

[43] Krause, A., Leskovec, J., Guestrin, C., VanBriesen, J., and Faloutsos, C. Efficient sensor placement optimization for securing large water distribution networks. *Journal of Water Resources Planning and Management 134*, 6 (2008), 516–526.

[44] Kuczaj, A., Komen, E., and Loginov, M. Large-eddy simulation study of turbulent mixing in a t-junction. *Nuclear Engineering and Design 240*, 9 (2010), 2116–2122.

[45] Launder, B. E., and Spalding, D. B. The numerical computation of turbulent flows. In *Numerical prediction of flow, heat transfer, turbulence and combustion*. Elsevier, 1983, pp. 96–116.

[46] Le, X.-H., Ho, H. V., Lee, G., and Jung, S. Application of long short-term memory (lstm) neural network for flood forecasting. *Water 11*, 7 (2019), 1387.

[47] Lee, Y. J., Park, C., and Lee, M. L. Identification of a contaminant source location in a river system using random forest models. *Water 10*, 4 (2018), 391.

[48] Liu, L., Zechman, E. M., Mahinthakumar, G., and Ranjithan, S. R. Coupling of logistic regression analysis and local search methods for characterization of water distribution system contaminant source. *Engineering Applications of Artificial Intelligence 25*, 2 (2012), 309–316.

[49] Lučin, I., Grbčić, L., Čarija, Z., and Kranjčević, L. Machine-learning classification of a number of contaminant sources in an urban water network. *Sensors 21*, 1 (2021), 245.

[50] Lučin, I., Grbčić, L., Družeta, S., and Čarija, Z. Source contamination detection using novel search space reduction coupled with optimization technique. *Journal of Water Resources Planning and Management 147*, 2 (2020), 04020100.

[51] McKenna, S. A., Orear, L., and Wright, J. Experimental determination of solute mixing in pipe joints. In *World Environmental and Water Resources Congress 2007: Restoring Our Natural Habitat* (2007), pp. 1–11.

[52] Ng, D.-Q., Chen, C.-Y., and Lin, Y.-P. A new scenario of lead contamination in potable water distribution systems: galvanic corrosion between lead and stainless steel. *Science of the Total Environment 637* (2018), 1423–1431.

[53] Nicoud, F., and Ducros, F. Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow, turbulence and Combustion 62*, 3 (1999), 183–200.

[54] Ostfeld, A., and Salomons, E. Optimal early warning monitoring system layout for water networks security: inclusion of sensors sensitivities and response delays. *Civil Engineering and Environmental Systems 22*, 3 (2005), 151–169.

[55] Ostfeld, A., Uber, J. G., Salomons, E., Berry, J. W., Hart, W. E., Phillips, C. A., Watson, J.-P., Dorini, G., Jonkergouw, P., Kapelan, Z., et al. The battle of the water sensor networks (bwsn): A design challenge for engineers and algorithms. *Journal of Water Resources Planning and Management 134*, 6 (2008), 556–568.

[56] Pandey, R. J., and Kim, K.-Y. Comparative analysis of flow in a fluidic oscillator using large eddy simulation and unsteady reynolds-averaged navier–stokes analysis. *Fluid Dynamics Research 50*, 6 (2018), 065515.

[57] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. Scikit-learn: Machine learning in python. *Journal of machine learning research 12*, Oct (2011), 2825–2830.

[58] Piazza, S., Blokker, E. M., Freni, G., Puleo, V., and Sambito, M. Impact of diffusion and dispersion of contaminants in water distribution networks modelling and monitoring. *Water Supply 20*, 1 (2020), 46–58.

[59] Pieper, K. J., Tang, M., and Edwards, M. A. Flint water crisis caused by interrupted corrosion control: Investigating "ground zero" home. *Environmental Science & Technology 51*, 4 (2017), 2007–2014.

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

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

[62] Preis, A., and Ostfeld, A. Multiobjective sensor design for water distribution systems security. In *Water Distribution Systems Analysis Symposium 2006* (2008), pp. 1–17.

[63] Rajasärkkä, J., Pernica, M., Kuta, J., Lašňák, J., Šimek, Z., and Bláha, L. Drinking water contaminants from epoxy resin-coated pipes: A field study. *Water research 103* (2016), 133–140.

[64] Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., and Ribeiro, L. Predictive modeling of groundwater nitrate pollution using random forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (southern spain). *Science of the Total Environment 476* (2014), 189–206.

[65] Romero-Gomez, P., Choi, C., van Bloemen Waanders, B., and McKenna, S. Transport phenomena at intersections of pressurized pipe systems. In *Water Distribution Systems Analysis Symposium 2006* (2008), pp. 1–20.

[66] Rossman, L. A., et al. Epanet 2: users manual.

[67] Rutkowski, T., and Prokopiuk, F. Identification of the contamination source location in the drinking water distribution system based on the neural network classifier. *IFAC-PapersOnLine 51*, 24 (2018), 15–22.

[68] Saadi, M., Oudin, L., and Ribstein, P. Random forest ability in regionalizing hourly hydrological model parameters. *Water 11*, 8 (2019), 1540.

[69] Santonastaso, G. F., Di Nardo, A., Creaco, E., Musmarra, D., and Greco, R. Comparison of topological, empirical and optimization-based approaches for locating quality detection points in water distribution networks. *Environmental Science and Pollution Research* (2020), 1–10.

[70] Sattar, A. M., Ertuğrul, Ö. F., Gharabaghi, B., McBean, E. A., and Cao, J. Extreme learning machine model for water network management. *Neural Computing and Applications 31*, 1 (2019), 157–169.

[71] Selvam, P. K., Kulenovic, R., and Laurien, E. Large eddy simulation on thermal mixing of fluids in a t-junction with conjugate heat transfer. *Nuclear Engineering and Design 284* (2015), 238–246.

[72] Seth, A., Klise, K. A., Siirola, J. D., Haxton, T., and Laird, C. D. Testing contamination source identification methods for water distribution networks. *Journal of Water Resources Planning and Management 142*, 4 (2016), 04016001.

[73] Shafiee, M. E., Berglund, E. Z., and Lindell, M. K. An agent-based modeling framework for assessing the public health protection of water advisories. *Water resources management 32*, 6 (2018), 2033–2059.

[74] Shao, Y., Yang, Y. J., Jiang, L., Yu, T., and Shen, C. Experimental testing and modeling analysis of solute mixing at water distribution pipe junctions. *Water research 56* (2014), 133–147.

[75] Shen, H., and McBean, E. False negative/positive issues in contaminant source identification for water-distribution systems. *Journal of Water Resources Planning and Management 138*, 3 (2011), 230–236.

[76] Song, I., Romero-Gomez, P., Andrade, M. A., Mondaca, M., and Choi, C. Y. Mixing at junctions in water distribution systems: an experimental study. *Urban Water Journal 15*, 1 (2018), 32–38.

[77] Strickling, H., DiCarlo, M. F., Shafiee, M. E., and Berglund, E. Simulation of containment and wireless emergency alerts within targeted pressure zones for water contamination management. *Sustainable Cities and Society 52* (2020), 101820.

[78] Sun, L., Yan, H., Xin, K., and Tao, T. Contamination source identification in water distribution networks using convolutional neural network. *Environmental Science and Pollution Research 26*, 36 (2019), 36786–36797.

[79] Tian, Y., Xu, Y.-P., Yang, Z., Wang, G., and Zhu, Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water 10*, 11 (2018), 1655.

[80] Ung, H., Piller, O., Gilbert, D., and Mortazavi, I. Accurate and optimal sensor placement for source identification of water distribution networks. *Journal of Water Resources Planning and Management 143*, 8 (2017), 04017032.

[81] Valero, D., and Bung, D. B. Sensitivity of turbulent schmidt number and turbulence model to simulations of jets in crossflow. *Environmental modelling & software 82* (2016), 218–228.

[82] Vankayala, P., Sankarasubramanian, A., Ranjithan, S. R., and Mahinthakumar, G. Contaminant source identification in water distribution networks under conditions of demand uncertainty. *Environmental Forensics 10*, 3 (2009), 253–263.

[83] Vries, D., van den Akker, B., Vonk, E., de Jong, W., and van Summeren, J. Application of machine learning techniques to predict anomalies in water supply networks. *Water Science and Technology: Water Supply 16*, 6 (2016), 1528–1535.

[84] Wang, H., and Harrison, K. W. 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 (2012), 3–11.

[85] Webb, S. W., and van Bloemen Waanders, B. G. High fidelity computational fluid dynamics for mixing in water distribution systems. In *Water Distribution Systems Analysis Symposium 2006* (2008), pp. 1–15.

[86] Wingender, J., and Flemming, H.-C. Contamination potential of drinking water distribution network biofilms. *Water Science and Technology 49*, 11-12 (2004), 277–286.

[87] Xuesong, Y., Jie, S., and Chengyu, H. Research on contaminant sources identification of uncertainty water demand using genetic algorithm. *Cluster Computing 20*, 2 (2017), 1007–1016.

[88] Yan, X., Gong, J., and Wu, Q. Pollution source intelligent location algorithm in water quality sensor networks. *NEURAL COMPUTING & APPLICATIONS* (2020).

[89] Yan, X., Zhao, J., Hu, C., and Zeng, D. Multimodal optimization problem in contamination source determination of water supply networks. *Swarm and Evolutionary Computation* (2017).

[90] Yan, X., Zhao, J., Hu, C., and Zeng, D. Multimodal optimization problem in contamination source determination of water supply networks. *Swarm and Evolutionary Computation 47* (2019), 66–71.

[91] Yan, X., Zhu, Z., and Li, T. Pollution source localization in an urban water supply network based on dynamic water demand. *Environmental Science and Pollution Research 26*, 18 (2019), 17901–17910.

[92] Yu, T., Qiu, H., Yang, J., Shao, Y., and Tao, L. Mixing at double-tee junctions with unequal pipe sizes in water distribution systems. *Water Science and Technology: Water Supply 16*, 6 (2016), 1595–1602.

[93] Yu, T., Tao, L., Shao, Y., and Zhang, T. Experimental study of solute mixing at double-tee junctions in water distribution systems. *Water Science and Technology: Water Supply 15*, 3 (2014), 474–482.

[94] Yu, T., Tao, L., Shao, Y., and Zhang, T. Experimental study of solute mixing at double-tee junctions in water distribution systems. *Water Science and Technology: Water Supply 15*, 3 (2015), 474–482.

[95] Zechman, E. M., and Ranjithan, S. R. Evolutionary computation-based methods for characterizing contaminant sources in a water distribution system. *Journal of Water Resources Planning and Management 135*, 5 (2009), 334–343.

[96] Zhou, Z., and Tartakovsky, D. M. Markov chain monte carlo with neural network surrogates: application to contaminant source identification. *Stochastic Environmental Research and Risk Assessment* (2020), 1–13.

List of Figures

List of Publications

- 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. https://doi.org/10.3390/w11061198

- Grbčić, L., Kranjčević, L., Družeta, S. and Lučin, I., 2020. Efficient double-Tee junction mixing assessment by machine learning. Water, 12(1), p.238. https://doi.org/10.3390/w12010238

- 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. https://doi.org/10.1016/j.asej.2020.06.004

- 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. https://doi.org/10.2166/hydro.2020.042

- 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. https://doi.org/10.3390/s20092613

- Grbčić, L., Kranjčević, L. and Družeta, S., 2021. Machine Learning and Simulation-Optimization Coupling for Water Distribution Network Contamination Source Detection. Sensors, 21(4), p.1157. https://doi.org/10.3390/s21041157

- Lušić, D.V., Kranjčević, L., Maćešić, S., Lušić, D., Jozić, S., Linšak, Ž., Bilajac, L., Grbčić, L. and Bilajac, N., 2017. Temporal variations analyses and predictive modeling of microbiological seawater quality. Water research, 119, pp.160-170. https://doi.org/10.1016/j.watres.2017.04.046

- Lučin, I., Čarija, Z., Grbčić, L. and 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

- 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.
  https://doi.org/10.1016/j.eij.2019.11.005

- 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.
  https://doi.org/10.3390/math8122107

- 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.
  https://doi.org/10.3390/s21010245

- 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.
  https://doi.org/10.1061/(ASCE)WR.1943-5452.0001308

- 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.
  https://doi.org/10.1061/(ASCE)WR.1943-5452.0001308

- Ivic, S., Grbcic, L. and Družeta, S., 2015. Cooperative random walk for pipe network layout optimization. International Journal of Applied Engineering Research (0973-4562) 11 (2016), 4; 2839-2847

- Kranjčević, L., Grbčić, L., Mrazović, M. and Družeta, S., 2020. Rijeka Bay 3D VOF Costal Flow Model. Pomorski zbornik, (3), pp.125-132.

- Grbčić, L., Kranjčević, L., Filiplić, I. and Mavrić, K., 2020. Numerical Simulation of River Inflows in Rijeka Bay Coastal Area. Pomorski zbornik, (3), pp.117-124.

- Zeng, H., Grbčić, L., Lučin, I. and Kranjčević, L., 2018. MESH CREATION FOR REALISTIC TERRAIN CASES FOR SHALLOWFOAM-2D OPENFOAM SOLVER. Annals of DAAAM & Proceedings, 29.

- Kranjčević, L., Šrajbek, M., Grbčić, L. and Čarija, Z., 2018, January. Analysis of well field nitrates pollution distribution in agricultural area. In Proceedings of the 29th DAAAM International Symposium (pp. 1053-1058).
- 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).

Curriculum Vitae

Luka Grbčić was born in 1990. in Rijeka, Croatia. He finished primary and secondary school in Rijeka. In 2013, at the Faculty of Engineering of the University of Rijeka, he obtained a bachelor's degree in mechanical engineering and in 2015 he obtained a master's degree in Computational Engineering defending a thesis titled *Heuristic Algorithm for Pipeline Network Optimization*. In 2016, at the Faculty of Engineering he enrolled in a postgraduate doctoral study in Computational Mechanics under the supervision of prof. dr. sc. Lado Kranjčević.

Since 2016, he has been an employee of the Faculty of Engineering at the University of Rijeka, where he is employed at the Department of Fluid Mechanics and Comptational Engineering as an assistant. He is a member of the Chair of Computational Engineering and performs scientific research regarding the applications of artificial intelligence and machine learning in engineering. At the Department of Fluid Mechanics and Computational Engineering, he is a teaching assistant on the courses Fluid Mechanics, Computational Methods, Fluid Dynamics and Application of Parallel Computing at the undergraduate and graduate university study of mechanical engineering and computer science. Since 2019, he has been a member of the Center for Advanced Computing and Modeling at the University of Rijeka, where he participates in the use of the BURA supercomputer for scientific research purposes.

From 2020, he is a member of two projects financed from EU funds. In the KLIMOD project, he participates in the development of a numerical model of the flow, flooding and contamination dispersion in rivers and coastal marine areas, and also participates in the use of machine learning algorithms to predict and model microbiological sea water contamination. In the EUROCC project (EU HORIZON 2020) he is a member of a group from the Faculty of Engineering in Rijeka that participates in the formation of the Croatian National Center for High Performance Computing.

Areas of scientific interest and research include: fluid mechanics, computer modeling in fluid dynamics, numerical modeling of fluid mixing in pipe networks, supercomputing and application of supercomputers in computational fluid dynamics, application of machine learning algorithms in engineering. Since 2016, he has published 11 papers in internationally peer-reviewed scientific journals, of which he is the first author of 6 papers. He is also the first author of 1 scientific paper and co-author of 4 scientific papers that have been reviewed and published in the proceedings of an international scientific conference. He is fluent in English.

Part II

Included Publications