

Article

Efficient Double-Tee Junction Mixing Assessment by Machine Learning

Luka Grbčić ^{1,2} , Lado Kranjčević ^{1,2,*} , Siniša Družeta ¹  and Ivana Lučin ¹ 

¹ Department of Fluid Mechanics and Computational Engineering, Faculty of Engineering, University of Rijeka, 51000 Rijeka, Croatia; lgrbcic@riteh.hr (L.G.); ilucin@riteh.hr (I.L.); sinisad@riteh.hr (S.D.)

² Center for Advanced Computing and Modeling, University of Rijeka, 51000 Rijeka, Croatia

* Correspondence: lado.kranjcevic@riteh.hr

Received: 8 November 2019; Accepted: 11 January 2020; Published: 15 January 2020



Abstract: 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.

Keywords: mixing phenomena; double-Tee junctions; machine learning; artificial neural networks; support vector regression; CFD model

1. Introduction

The mixing of fluids in water distribution networks is a complex phenomenon that has been extensively subjected to research as it is relevant to several specific areas of application such as water distribution quality and safety [1–3], pollution source detection systems (both large [4,5] and small networks [6]), and optimal pollution sensor placement in a water distribution network [7–10]. The elements that form water distribution networks are pipes and junctions. When modeling mixing in a complex system, a correct mixing model must be applied to accurately describe the contaminant transport through the network due to the fact that a wrong solution could present a hazard to a great number of network users.

Usually, mixing in a pipe network is modeled as either complete mixing or bulk mixing. Complete mixing can be described as an even split of contamination at a network junction. Complete mixing models such as the one developed in the hydraulic analysis software EPANET are implemented by calculating the flow-weighted concentrations at the inlet pipes of a junction and then assuming an even split in the outlet pipes. The complete mixing model can be assumed correct only if there is a single outlet at a junction and if the distance between two junctions is great enough. For certain distances [11,12], it does not describe the physical process of mixing correctly.

Bulk mixing is an idealized mixing in which the flow between two inlet pipes is not interacting and the diffusive behavior between the inlet pipe streams is ignored (the streams are only touching at an interface). To model the bulk mixing, the flow momentum in the inlet pipes is calculated, and it determines how the concentration splits between the outlets. The inlet pipe with the higher momentum will dominate in concentration transfer into the outlet pipe, which follows the same direction as the inlet pipe. The inlet pipe with lower flow momentum will only transfer the concentration into its neighboring outlet pipe. This kind of mixing behavior is usually specific to cross junctions. The bulk mixing model is implemented in the software EPANET-BAM [13] (which can be considered as an extension of EPANET). If the expected mixing behavior is neither complete nor bulk, an experimentally-calibrated mixing model parameter can be used, by which type of mixing is defined.

In research by McKenna et al. [14], it is shown that, for a cross junction (which is a special case of a double-Tee junction in which the distance between converging and diverging pipes is equal to zero), a strong deviation from complete mixing occurs when the Reynolds number at each inlet and outlet pipe is equal. On the other hand, stronger turbulence can also create a considerable unpredictability of mixing in cross junctions [15]. The mixing that occurs at two sequential Tee junctions is especially complex due to increased chaotic eddying of flow, which in turn increases the diffusion between the streams. It is proved in several studies [12,16,17] that the length between double-Tee junctions is important since it determines the behavior of mixing of fluids or rather the transport of contaminant into each diverging pipe. In addition, variation of flows at both inlet and outlet pipes causes the mixing behavior to vary [11,12]. Another property of a pipe system that can cause an incomplete mixing behavior are uneven pipe diameters, where a large difference between pipe diameters can produce a more complete mixing [18]. The orientation of pipes for a double-Tee junction also affects the mixing behavior, and it is shown that mixing for a case when the pipes are at a opposite side compared to the case where they are at the same side, behaves differently [11,12] (assuming planar configuration). It is also shown that transitional and laminar flows in pipes exhibit a specific mixing behavior [19].

Instead of an experimental approach to studying mixing behavior in double-Tee junctions, computational fluid dynamics (CFD) can also be used. A species transport model coupled with RANS turbulence models yields good results [12,16] that can be used to further enhance simpler 1D mixing models (such as EPANET-BAM). Additionally, a high fidelity LES turbulence model can also be used to study mixing with more detail [20]. The problem with CFD simulations and the contaminant transport model is that they need to be calibrated as they are dependant on the turbulent Schmidt number, the value of which is very much case-specific [21,22] and has been differently reported in several studies regarding CFD analysis of mixing in double-Tee junctions [12,23,24]. Furthermore, CFD simulations are computationally expensive and producing results for a range of combinations of mixing scenarios (varying inlet and outlet flow, distance between junctions, and even pipe diameters) is quite impractical. Alternatively, mixing behavior can be predicted by use of data analysis (i.e., interpolation on a fixed previously produced dataset), as was done in a previous study [25] where Kriging and Delaunay triangulation were combined to describe mixing for different Reynolds numbers.

Recently, machine learning (ML) algorithms used in hydrology and hydraulics include Artificial Neural Networks (ANN) and variations [26,27], Random Forest [28], and Support Vector Machine [29]. ML has been previously successfully implemented in pipe flow systems. A selection of ML algorithms such as Random Forest, Bagging Algorithm, and Regression Tree were used to predict various characteristics of a wastewater pipe system and it was found that Regression Trees are most successful [30]. Several ML algorithms were also used to predict pressure gradients, liquid holdup, and flow pattern identification of multiphase flow on a pipe segment (where data were generated by CFD) and, for some of the characteristics, the Gradient Boosting algorithm and Support Vector Machine yielded best results [31]. Deep learning (i.e., ANN) was successfully applied to produce the best valve scheduling scenario in the case of pipe network contamination instead of searching for a contamination

source [32]. Similarly, Support Vector Regression (SVR) was used to detect anomalies in a pipe network system [33] and a modified ANN approach was used for pipe network management [34].

A ML approach can be also used to evaluate the mixing behavior based on previously obtained data (either by experiment or by CFD). In this study, two ML algorithms—SVR and ANN—were tested in predicting mixing behavior for a combination of different inlet and outlet pipe flow ratios and different distances between double-Tee junctions, on one pipe configuration. Training data were generated by 3D CFD that was calibrated on an experiment presented in a previous study [12]. A supercomputer was used for CFD modeling due to its high computational demand and need to generate as many data as possible for better ML training.

2. Materials and Methods

2.1. Problem Description

Mixing prediction was performed for a previously reported experiment [12], in which the setup includes varying distances between double-Tee junctions and varying inflows in pipes.

Figure 1 shows the pipe configuration used. Branch pipes are on the opposite sides and the arrows represent the flow direction. Tap water (T) (which was used as a tracer) enters the main inlet pipe 1 while distilled water (D) enters the branch inlet pipe 2 causing the mixture (M) of tap and distilled water to exit both the main outlet pipe 3 and the branch outlet pipe 4. The electrical conductivity of tap water in the experiment was measured to be around $200 \mu\text{S}/\text{cm}^2$ while the distilled water conductivity was around $2 \mu\text{S}/\text{cm}^2$ at pipe inlets 1 and 2, respectively. This difference of electrical conductivity of the fluids was used to determine the mixing process in the double-Tee junctions since they differed by a factor of 100. The electrical conductivity of the mixture that leaves pipe outlets 3 and 4 was measured and was always constrained between the inlet tap water and inlet distilled water electrical conductivity. The rate of mixing or the electrical conductivity of the mixture at outlets depends on the flow conditions in the inlet and outlet pipes.

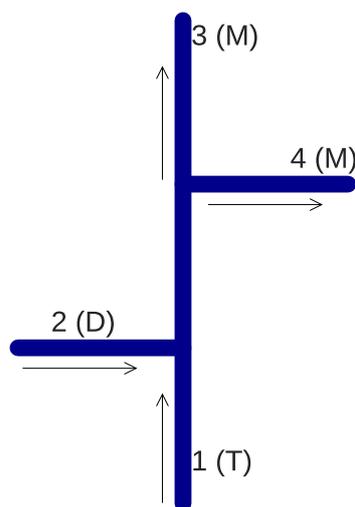


Figure 1. Scheme of the experimental configuration.

The measure of mixing taken from experimental data and used for further investigation of mixing was the branch outlet pipe 4 to main inlet pipe 1 (see Figure 1) conductivity ratio, which is defined as

$$R_c = \frac{c_4}{c_1} \quad (1)$$

where c_4 represents the conductivity at outlet pipe 4, while c_1 is the conductivity at inlet pipe 1.

This ratio was chosen due to its usage in previous studies [13,35] as it can be used in mixing prediction for simpler 1D models (EPANET-BAM). The inlet pipe flow ratio was also varied in the experiment and it is defined as

$$R_{q_{in}} = \frac{Q_1}{Q_2} \quad (2)$$

where Q_1 defines the flow at inlet pipe 1 and Q_2 the flow at inlet pipe 2. The values of Q_1 and Q_2 were controlled with valves in the experiment until a steady state behavior was achieved and then the measurement of the electrical conductivity at each outlet pipe was made.

The inflow pipe ratios $R_{q_{in}}$ used in the experiment were 0.333, 0.5, 1, 2, and 3, while the flow at both outlet pipes was the same. One of the purposes of this study was to examine the behavior of mixing when $R_{q_{in}}$ is not equal to 1 while the flow at both outlet pipes is also being varied since there is a lack of previous research regarding these combinations (which in a realistic case are most probable). Another important characteristic of the experiment is the distance between the double-Tee junction that was varied as $5.6D$, $10D$, and $15D$ (where D is the diameter of the pipes). All of the relevant specifications of the experiment are summarized in Table 1.

Table 1. Experimental data.

Parameter	Value	Unit
Internal pipe diameter (D)	18	mm
Inlet pipes length	$20D$	mm
Outlet pipes length	$40D$	mm
Tee distances	$5.6D, 10D, 15D$	mm
Flow in mixing zone	0.08–0.43	1/s
Reynolds number range	5658–30416	-
Inlet flow ratios	0.333, 0.5, 1, 2, 3	-
Kinematic viscosity of water	1×10^{-6}	m^2/s

2.2. CFD Simulations

2.2.1. CFD Model

A CFD approach was used as both a comparison of efficiency with the ML approach and a dataset generator for ML algorithms. The open-source CFD toolbox OpenFOAM [36] was used to model the mixing phenomena as it was previously proven to work well for the given problem. The isothermal 3D steady Reynolds-Averaged Navier–Stokes (RANS) equations with the passive scalar transport model and the k-Epsilon (k - ϵ) turbulence model were solved, where k represents the turbulence kinetic energy and ϵ is the turbulence energy dissipation. The passive scalar transport model is defined as:

$$\nabla \times (\mathbf{vc}) = \nabla \times (D_t \nabla c) \quad (3)$$

$$D_t = D_{mol} + \frac{\nu_t}{Sc_t} \quad (4)$$

$$\nu_t = C_\mu \frac{k^2}{\epsilon} \quad (5)$$

In Equation (3), parameter c represents a dimensionless scalar value, set as $c = 0$ for distilled water and $c = 1$ for tap water as it represents the tracer in pipes; D_t is the turbulent diffusivity defined by

Equation (4); and \mathbf{v} represents the advection velocity vector. Equation (3) can be physically interpreted as transport of a scalar quantity c by the fluid flow, which is resolved by the coupled RANS/(k - ϵ) model. The additional passive scalar transport model in Equation (3) was coupled with the steady-state turbulent OpenFOAM solver simpleFoam making it a new fluid flow solver referred to as tracerFoam.

Equation (4) is essential for modeling mixing phenomena as the turbulent Schmidt number S_{ct} needs to be calibrated. According to the results of the previous study [12], S_{ct} was defined as $S_{ct} = 0.5$. As for other parameters, D_{mol} is the molecular diffusivity of water (set as $D_{mol} = 1.7 \times 10^{-9} \text{ m}^2/\text{s}$ [37]), ν_t represents the kinematic eddy viscosity, and the constant C_μ was set to $C_\mu = 0.09$.

A bounded second-order upwind-biased scheme was used for the divergence terms of the model, while for the gradient terms the central difference scheme was used. The fluid used for all simulation runs was water with kinematic viscosity of $\nu = 1 \times 10^{-6} \text{ m}^2/\text{s}$.

A uniform value for velocity \mathbf{v} was defined at pipe inlets and outlets and a no-slip boundary condition was defined at the pipe walls. The values of \mathbf{v} were set at the pipe inlets and pipe outlets to form an exact ratio of both $R_{q_{in}}$ and $R_{q_{out}}$ for all cases (0.333, 0.5, 1, and 2). Pressure $p = 0$ was defined at outlet pipe 4 and zero pressure gradient in the normal direction was defined on the inlets, outlet pipe 3, and pipe walls. The concentration c was fixed at the inlets ($c = 1$ for inlet pipe 1 and $c = 0$ for inlet pipe 2), while, at the outlets and pipe walls, the derivative of concentration in the normal direction was set to zero. Additionally, the value of c was monitored at outlet pipe 3, and, to achieve a steady state regime, it had a convergence criterion of 10^{-5} . For the (k - ϵ) turbulence model, the values of k and ϵ at inlets and outlets were defined the same way as for c except at the pipe walls, where turbulence wall functions were used. Boundary conditions used in the numerical model are summarized in Table 2.

Table 2. Boundary conditions for the passive scalar model.

Variable	Inlets	Outlets	Pipe Walls
\mathbf{v}	\mathbf{v}	\mathbf{v}	$\mathbf{v} = 0$
p	$\partial p / \partial \mathbf{n} = 0$	p and $\partial p / \partial \mathbf{n} = 0$	$\partial p / \partial \mathbf{n} = 0$
c	c	$\partial c / \partial \mathbf{n} = 0$	$\partial c / \partial \mathbf{n} = 0$
k	k	$\partial k / \partial \mathbf{n} = 0$	f_{wall}
ϵ	ϵ	$\partial \epsilon / \partial \mathbf{n} = 0$	f_{wall}

2.2.2. Mesh Independence

Firstly, to use it as a dataset generator, the CFD model was validated with the previously obtained experimental data. A mesh independence study was done to find the optimal mesh sizes for data generation. Each distance between the double-Tee junctions yielded a differently sized mesh, with greater distances requiring more cells. The CFD approach mesh independence was validated with $R_{q_{in}}$ equal to 1 and with equal flow at outlet pipes ($R_{q_{out}}$ equal to 1) for all of the above-mentioned distances (Table 1). The grid convergence residual criterion was set to 10^{-5} for Equation (3). Meshes used in the mesh independence study can be seen in Table 3 where the average cell size ranged from 1.2 to 0.5 mm, while the summary of the mesh independence study is presented in Table 4 where R_c is the branch outlet pipe to main inlet pipe conductivity ratio. For ML model data generation, the fine mesh was chosen.

Table 3. Computational mesh sizes (number of cells in thousands) for different distances.

Distance	Coarse	Mid	Fine	Finest
5.6D	196	861	1819	4900
10D	401	1300	2950	5000
15D	538	1960	3000	5200

Table 4. Mesh independence study results (value of R_c) for different distances.

Distance	Coarse	Mid	Fine	Finest	Experiment
5.6D	0.602	0.607	0.613	0.613	0.610
10D	0.563	0.560	0.552	0.560	0.563
15D	0.541	0.541	0.530	0.530	0.539

2.2.3. CFD Data Generation

After the mesh independence study was completed and the numerical model was validated, CFD was used to generate data for ML. As it was defined in the experiment, the value of the inlet pipe flow ratio in Equation (2) was varied with the outlet pipe flow ratio, which is defined in Equation (6), where Q_3 is the main pipe 3 outlet flow and Q_4 is the branch pipe 4 outlet flow.

$$R_{q_{out}} = \frac{Q_3}{Q_4} \quad (6)$$

For an array of $R_{q_{in}}$ and $R_{q_{out}}$, the value of R_c was obtained ($R_{c_{CFD}}$). The ratios were all possible combinations of 0.333, 0.5, 1, 2, and 3 between the inlet pipes and the outlet pipes (e.g., $R_{q_{in}}$ set as 2 and $R_{q_{out}}$ as 0.333, $R_{q_{in}} = 1$ and $R_{q_{out}} = 2$, or $R_{q_{in}} = 0.333$ and $R_{q_{out}} = 0.5$).

This was conducted for all distances (5.6D, 10D, and 15D), comprising a total of 75 simulations on the fine mesh (25 simulations per distance).

2.3. ML Model

2.3.1. Artificial Neural Networks

Both an Artificial Neural Network (ANN) and Support Vector Regression (SVR) were used to train a model which predicts the behavior of mixing in a double-Tee junction case. For both methods, the input and output data were split as 70% for model training and 30% for model accuracy testing.

Generally, ANNs are a type of ML algorithms that are used to model complex phenomena, which are not as easy to model using conventional data analysis methods. The type of ANNs used in this study are Feedforward Multilayer Perceptron Neural Networks. This type of ANN is constructed using an input layer, which is usually data that cause the emergence of complex phenomena modeled, one or more hidden layers, and one output layer, which is usually the outcome of the complex phenomena. Layers consist of artificial neurons and a link exists between every layer of the ANN. Training an ANN is the process of fitting the weights of the artificial neuron links so that for each input datum a path of links is made to a certain output. Further information on ANNs can be found in the literature [38].

In this study, three hidden layers were added between the input and output layers and between every layer a dense connection was defined. The input layer consisted of three different data, the inlet pipe flow ratio $R_{q_{in}}$, the outlet pipe flow ratio $R_{q_{out}}$, and distance between the double-Tee junctions (5.6D, 10D, and 15D). The first hidden layer consisted of 50 neurons, both the second and third hidden layer consisted of 25 neurons, and all hidden layers used the rectifier activation function. The output layer was the value of Equation (1) or $R_{c_{ANN}}$. The sigmoid activation function was used at the output layer. The layer configuration is summarized in Table 5.

The ANN was trained using the ADAM optimizer [39] for 500 epochs with the binary cross-entropy loss function for error modeling. Initial weights of each layer link was randomized. The ANN was implemented in the Python Neural Network library Keras [40].

Table 5. ANN model definition with layers, activation function and data.

Layer	Neurons	Activation	Data
Input Layer	3	-	$R_{q_{in}}, R_{q_{out}}, \text{Distance}$
Hidden Layer 1	50	Rectifier	-
Hidden Layer 2	25	Rectifier	-
Hidden Layer 3	25	Rectifier	-
Output Layer	1	Sigmoid	$R_{c_{ANN}}$

2.3.2. Support Vector Regression

SVR can be considered as a special case of Support Vector Machine (SVM) models. SVMs can be used for both classification and regression problems and SVR would obviously belong to a family of algorithms used to create regression models. The SVR model is given training data in the form of inputs and outputs and it learns the relationship between the data. The inputs are n-dimensional vectors and the outputs are continuous values (unlike in classification problems). The input–output dependency or mapping is approximated using kernel functions and weights. Similar to in ANNs where the link weights are being fitted, the weights in SVR models are also fitted. In the process of learning, a hyperplane is created between the n-dimensional data along with error planes. The main goal is to find the data (or the support vectors) that are within the error margin. A further description of SVR can be found in the literature [41].

The data input and output values were the same as the ones for the ANN (defined in Section 2.3.1). A fifth-degree radial basis function (RBF) kernel was used for the model, a tolerance of 0.001 (which is the model fitting stopping criterion), an ϵ of 0.0001 (which specifies the error margin), and a penalty parameter equal to 3 (which is enforced for values outside the error margin). SVR was implemented in the Python ML library Scikit-learn [42].

3. Results and Discussion

3.1. CFD Results and Efficiency

The CFD analysis was done for 75 different scenarios by varying $R_{q_{in}}, R_{q_{out}}$, and three distances (5.6D, 10D, and 15D). The CFD simulation with $R_{q_{in}}$ equal to 0.333, 0.5, 2, and 3 and $R_{q_{out}}$ equal to 1 were compared with the experimental values to make sure that the procedure is valid and that a CFD model can be used as a data generator for ML (simulation results for $R_{q_{in}}$ equal to 1 are previously given in Section 2.2.2). The results are summarized in Table 6.

Table 6. Comparison of CFD and experimental results for different $R_{q_{in}}$ values.

Distance	$R_{q_{in}}$	$R_{c_{CFD}}$	R_c
5.6D	0.333	0.262	0.239
5.6D	0.5	0.377	0.342
5.6D	2	0.798	0.801
5.6D	3	0.880	0.857
10D	0.333	0.249	0.280
10D	0.5	0.356	0.370
10D	2	0.741	0.719
10D	3	0.825	0.790
15D	0.333	0.246	0.268
15D	0.5	0.343	0.359
15D	2	0.712	0.719
15D	3	0.800	0.809

For the distance of 5.6D between double-Tee junctions, the root-mean-square error (RMSE) is 0.0214 with the minimum relative error being +0.3% and the maximum relative error –10% between

the experimental and CFD values of R_c . The distance of 10D had the RMSE of 0.0246, the minimum relative error of experimental and CFD R_c as -3.05% , and the maximum relative error $+11\%$. For the 15D distance, the RMSE is 0.013 with minimum relative error as 0.9% and maximum relative error 8.9% . The error results are summarized in Table 7.

Table 7. CFD model error summary for all distances and $R_{q_{in}}$ from Table 6.

Distance	RMSE	Min. Error	Max. Error
5.6D	0.0214	+0.3%	-10%
10D	0.0246	-3.05%	+11%
15D	0.013	0.9%	8.9%

It can be seen that the error of the CFD model for the 15D distance is the lowest one. This can be attributed to a larger distance between double-Tees since the mixture is already averaged enough (and closer to complete mixing) across the pipe cross section before it is transported into the diverging pipes where, arguably, the most complex eddying of flow is happening. At smaller distances, that averaging of concentration before the diverging pipe is less visible. This can be seen in Figure 2 where Figure 2a is the concentration in the pipe cross section at $1.67D$ before the branch pipe 4 center-line for 5.6D distance, Figure 2b represents the same but for the 10D distance, and Figure 2c shows results for 15D. In Figure 2d, the modeled kinetic energy of turbulence is shown and it is the highest at the entrance of branch pipe 4, which could be the cause of a slightly higher error for smaller distances. Overall, the errors are acceptable and CFD proved to be a viable tool in modeling this phenomena.

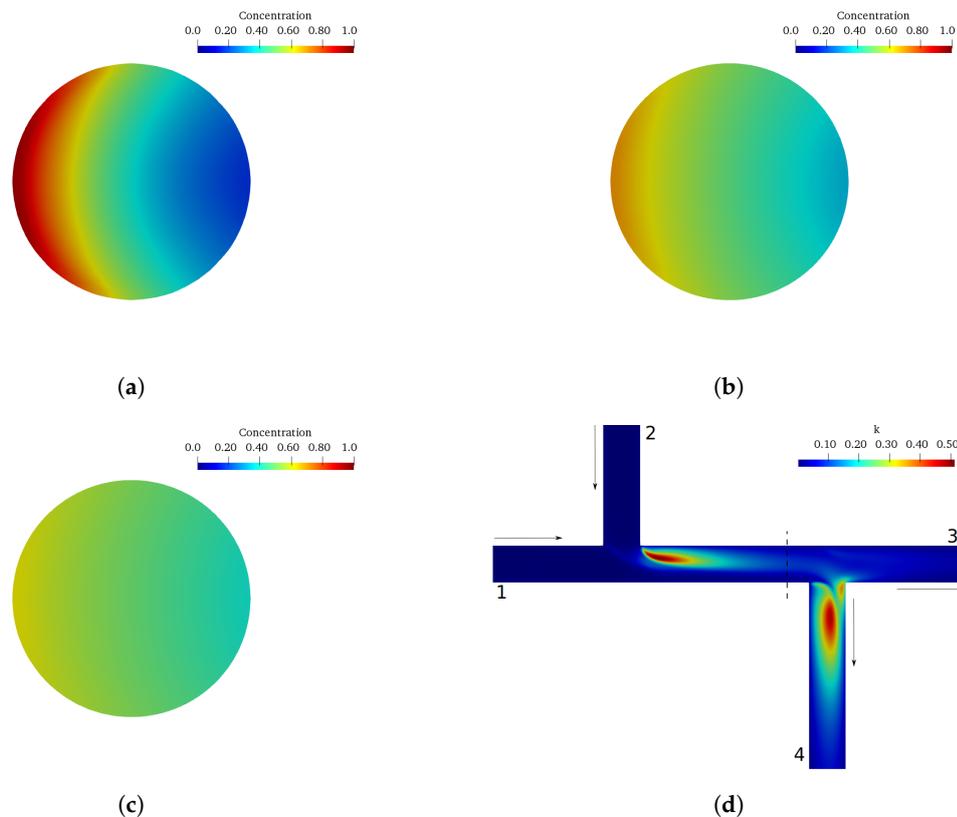


Figure 2. Concentration at pipe cross sections ($1.67D$ before the center-line of branch outlet pipe 4 and all cross sections are perpendicular to fluid flow) and turbulence kinetic energy with $R_{q_{in}}$ and $R_{q_{out}}$ equal to 1 for all cases: (a) cross section concentration for 5.6D; (b) cross section concentration for 10D; (c) cross section concentration for 15D; and (d) turbulence kinetic energy k (m^2/s^2) for 5.6D (plan view), where arrows represent the flow direction and a dashed line the position of the cross section of the 5.6D case.

Even though the CFD method used provides a good model for mixing phenomena in double-Tee junctions, there is still a problem of computational efficiency. All simulations (for the fine mesh that ranged from approximately 1.8 to 3 million cells) were done using the BURA supercomputer resources at the University of Rijeka. All runs were executed in parallel on 8 Intel E5 nodes with 20 threads per node resulting in 160 processes in total. The computational run times (until the residual of Equation (3) fell below 10^{-5}) for each case is summarized in Table 8.

Table 8. Computational run time for each distance on a fine mesh.

Distance	Cell Number	Time
5.6D	1.86 million	56 min
10D	2.95 million	82 min
15D	3.00 million	83 min

Taking into account that these results were obtained with a supercomputer on 160 Intel E5 cores, this methodology is very inefficient if employed on a personal computer, especially since a lot of computation is necessary for a large variety of possible combinations of inlet pipe flows, outlet pipe flows, and distances between double-Tee junctions.

3.2. ML Results

3.2.1. Models' Training and Testing

Of the 75 obtained simulation results, 70% were used for model training, and 30% were used for the purpose of model accuracy testing for both ANN and SVR since this train-to-test ratio provides that the training dataset should include all possible patterns used for defining the problem and should extend to the edge of the modeling domain. Both models were tested 10 times with randomly selected input data and the models with the minimum RMSE value of the test data were adopted and used further (both ML models RMSE varied for around 0.008 between the minimum and maximum value). The CFD obtained results were taken to be exact (since they exhibit a good agreement with the experimental results based on the average RMSE) and ANN and SVR model results were compared with them to examine their accuracy. For the same data inputs ($R_{q_{in}}$, $R_{q_{out}}$, and distance), the ANN model performed better than the SVR model regarding the accuracy of the testing data (predictions of R_c). The RMSE of ANN predictions was 0.0172 while for SVR it was 0.0361 with the maximum error of 0.0427 for ANN and 0.119 for SVR. The results of training data predictions are summarized in Table 9 while a detailed comparison of predicted data can be seen in Table 10.

Table 9. Accuracy metrics of ANN and SVR models.

Model	RMSE	Max Error
ANN	0.0172	0.0427
SVR	0.0361	0.119

The values of $R_{c_{SVR}}$ (SVR predicted R_c) and $R_{c_{ANN}}$ (ANN predicted R_c) in Table 10, which were obtained from the 30% of the input dataset, in bold better predict the CFD modeled values $R_{c_{CFD}}$, which serve as correct results of mixing behavior. ANN had a better prediction than SVR for 14 out of 23 input data values but for the other 9 predictions it was also quite close to the CFD modeled values. Overall, both models exhibit solid accuracy for predicting the input data even though ANN would obviously be a better choice for further examination.

Table 10. Comparison of the testing data (30% of the whole generated data) predictions for SVR and ANN with bold values representing a better prediction.

$R_{q_{in}}$	$R_{q_{out}}$	Distance	$R_{c_{CFD}}$	$R_{c_{SVR}}$	$R_{c_{ANN}}$
1	0.333	15D	0.519	0.516	0.520
0.5	0.333	15D	0.342	0.323	0.329
3	0.333	5.6D	0.812	0.810	0.833
3	3	15D	0.823	0.775	0.850
1	2	10D	0.565	0.577	0.578
1	0.5	10D	0.535	0.546	0.555
2	0.5	15D	0.702	0.699	0.710
1	3	15D	0.540	0.472	0.527
0.5	3	10D	0.360	0.327	0.333
0.5	0.333	5.6D	0.354	0.335	0.362
2	2	10D	0.769	0.807	0.756
0.5	3	15D	0.341	0.290	0.336
2	0.5	10D	0.718	0.714	0.707
2	1	10D	0.741	0.736	0.718
0.5	2	10D	0.360	0.334	0.357
3	1	15D	0.800	0.800	0.807
1	0.333	10D	0.529	0.545	0.546
3	3	10D	0.868	0.749	0.911
0.333	1	15D	0.246	0.267	0.257
0	0.5	15D	0.344	0.326	0.333
0	0.5	5.6D	0.360	0.343	0.361
1	1	15D	0.530	0.545	0.556
2	2	15D	0.730	0.751	0.729

3.2.2. R_c Prediction for Other Double-Tee Distances

Another useful test was done on the ANN and SVR models, which is the prediction of R_c for distances that were not in the input training data. The chosen distances were 7.5D and 12.5D since they are in between the training data values (5.6D, 10D, and 15D). The results of the 7.5D and 12.5D tests can be seen in Table 11 and again the results in bold are the ones that agree better with the CFD prediction.

Table 11. SVR and ANN model comparison for 7.5D and 12.5D (better prediction values are made bold).

$R_{q_{in}}$	$R_{q_{out}}$	Distance	$R_{c_{CFD}}$	$R_{c_{SVR}}$	$R_{c_{ANN}}$
2	1	7.5D	0.770	0.624	0.742
0.5	3	7.5D	0.362	0.456	0.371
2	1	12.5D	0.728	0.562	0.712
0.333	1	12.5D	0.245	0.447	0.245

After testing the prediction on different double-Tee junction distances (which were not included in the training data), it is obvious that ANN greatly outperforms SVR on this task. Henceforth, further double-Tee mixing ML analysis (of a whole range of $R_{q_{in}}$ and $R_{q_{out}}$ values) was only done using ANN.

3.2.3. R_c Prediction for a Range of $R_{q_{in}}$ and $R_{q_{out}}$ with ANN

A prediction of a range of possible combinations of inlet pipe flow and outflow ratios was done, albeit on a range that is still inside the minimum and maximum values of $R_{q_{in}}$, $R_{q_{out}}$ (from 0.333 to 3), and the distance of the training model, thus no extrapolation of data was done. The purpose of this test was to obtain the whole $R_{q_{in}}-R_{q_{out}}$ surface for a given distance between the double-Tee junction. Before extracting the surface data from the ANN model, which show mixing behavior for the whole array of $R_{q_{in}}$ and $R_{q_{out}}$, an additional check of validity of the ANN model was performed for a $R_{q_{in}} = 1.3$ and $R_{q_{out}} = 1$ since the testing data were done only for $R_{q_{in}}$ and $R_{q_{out}}$ of 0.333, 0.5, 1, 2, and

3. The CFD model result $R_{c_{CFD}}$ was 0.706 while the $R_{c_{ANN}}$ was 0.681, which corresponds to an error of +3.54% when $R_{c_{CFD}}$ is considered to be the exact value.

Figures 3–5 show the predicted values of R_c by ANN for the distances 5.6D, 10D, and 15D, respectively. Values of $R_{c_{ANN}}$ are represented by color and iso-lines. All three figures show continuous values of $R_{c_{ANN}}$ instead of the discrete ones obtained by CFD and experiments. It can be noticed by comparing the figures that the gradient of $R_{c_{ANN}}$ is smoother (the difference between the minimum and the maximum value is smaller) when the double-Tee distance is increased, which is logical since a much more complete mixing behavior is achieved with an increased distance. Another thing that can be noticed (for all distances) in the figures is that, when both $R_{q_{in}}$ and $R_{q_{out}}$ are increased, the value of $R_{c_{ANN}}$ tends towards 1, which can be interpreted as, when simultaneously both the flow in the main inlet pipe and the main outlet pipe are increasing, the concentration in the main outlet pipe is also increasing, and therefore deviating more from complete mixing. A similar thing happens when both ($R_{q_{in}}$ and $R_{q_{out}}$) are decreasing, making the concentration $R_{c_{ANN}}$ lower (it tends towards 0).

Additionally, in Figure 6, a 3D $R_{c_{ANN}}$ plot that contains the surfaces for 5.6D, 10D, and 15D distances is presented to better depict the gradient or the steepness of the surfaces.

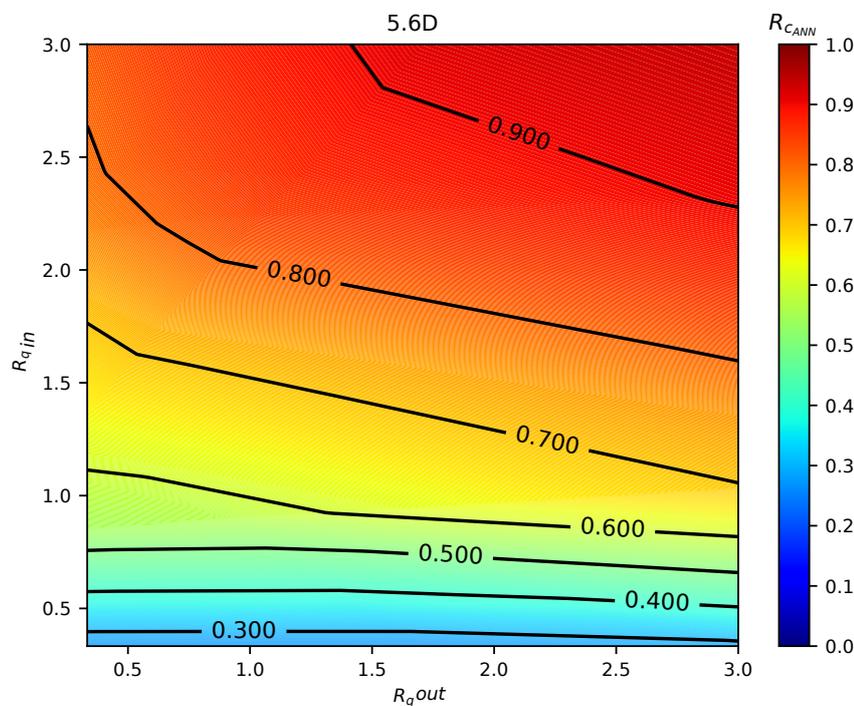


Figure 3. Predicted $R_{c_{ANN}}$ surface values for $R_{q_{in}}$, $R_{q_{out}}$, and 5.6D distance.

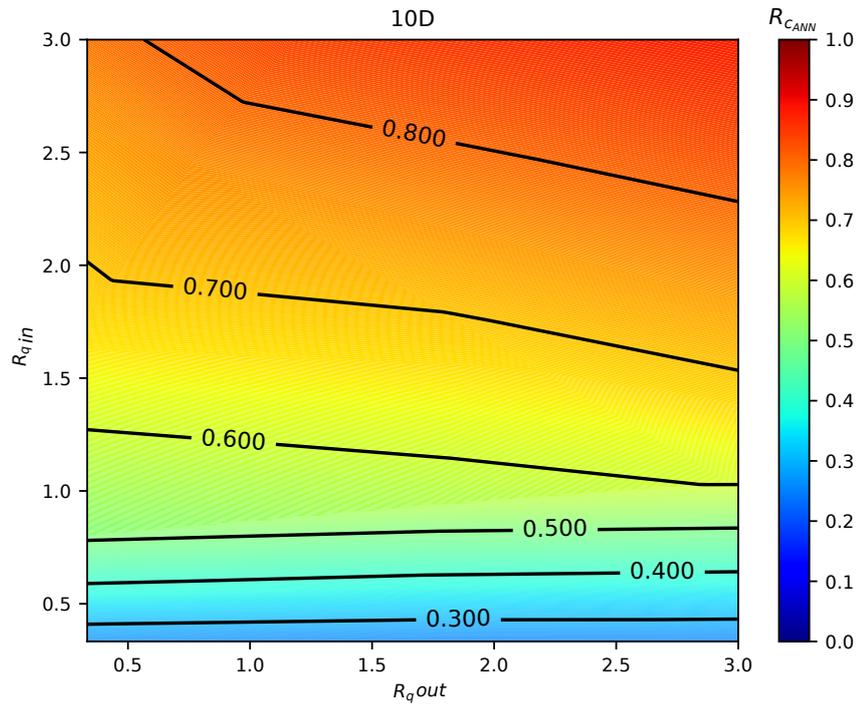


Figure 4. Predicted R_{CANN} surface values for R_{qin} , R_{qout} , and 10D distance.

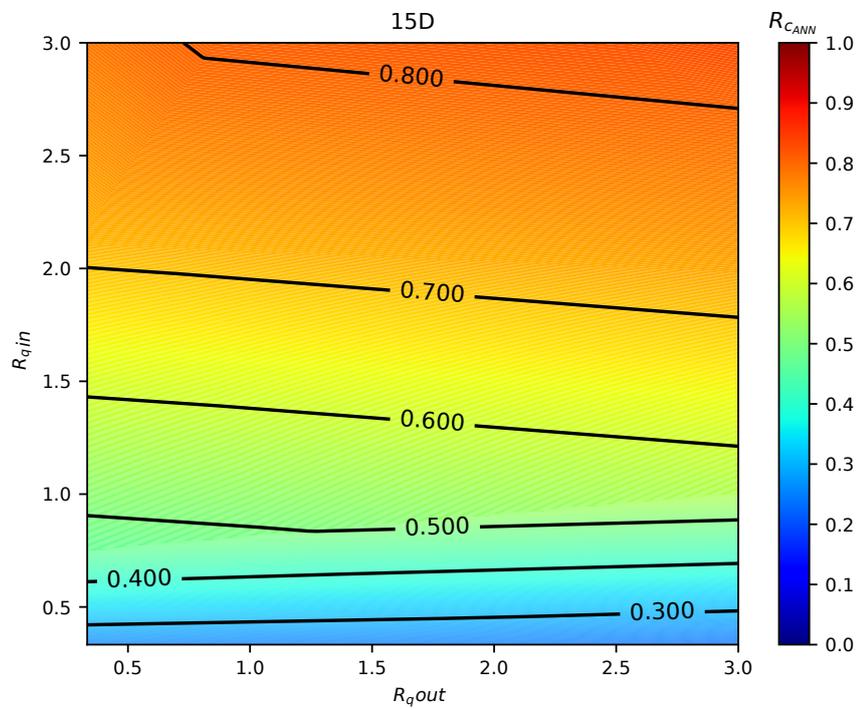


Figure 5. Predicted R_{CANN} surface values for R_{qin} , R_{qout} , and 15D distance.

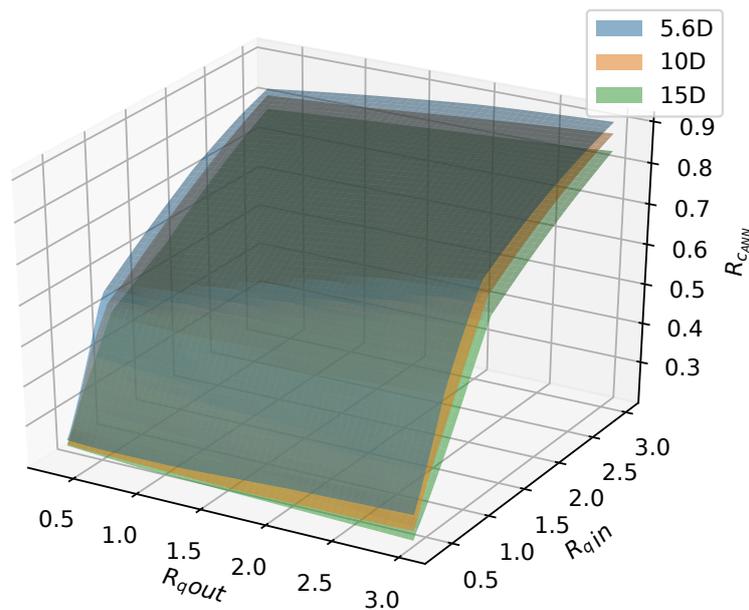


Figure 6. Predicted $R_{C_{ANN}}$ surfaces for 5.6D, 10D, and 15D.

3.2.4. On ANN Efficiency

It is hard to quantify the efficiency of the ANN model since it is dependant on previously generated data (either by CFD or experiment) but if a task were to generate, with a trained ANN, the whole range of new values of R_c from a range of input values, then it could be argued that ANN is extremely efficient since obtaining $R_{C_{ANN}}$ for 23 input data lasted 0.18 s.

4. Summary and Conclusions

In this paper, a new methodology for double-Tee junctions mixing modeling is presented based on an experiment from the previous literature. A CFD model was calibrated with the experimental data to serve both as a computational efficiency benchmark and a data-generator for a ML approach for the presented double-Tee junction mixing phenomena assessment problem.

In total, 75 simulations were done by the calibrated CFD RANS $k-\epsilon$ turbulence model and the data were used to train two different ML models. The input variables were defined as the inlet pipe flow ratio, the outlet pipe flow ratio, and distance between the double-Tee junction, while the output was the outlet branch pipe to main inlet pipe conductivity ratio, which is a relevant variable that describes the mixing behavior in a double-Tee junction.

The two tested ML algorithms were Support Vector Regression and Artificial Neural Network (specifically Feedforward Multilayer Perceptron), and, of the two, ANN outperformed SVR in accuracy. Empirical results obtained with the ANN model used show good approximation in predicting the mixing behavior for different distances between double-Tee junctions and inlet/outlet pipe flow ratios.

Besides the accuracy of ANN, it was determined that the computational efficiency of the ANN model is much greater than that of the CFD model, which makes it a great tool for modeling this kind of phenomena. Since many previous studies produced experimental data for a great number of different combinations (which include different Reynolds numbers ranges in pipes, varied pipe diameters, distances between double-Tees, pipe configuration, etc.), a further study could be done to create an ANN that takes into account all of these variables and produces all the relevant parameters to accurately model mixing. With its computational efficiency, a fast model such as ANN could be directly incorporated into complex water network supply pipe network models (such as EPANET and similar), which would greatly improve their accuracy and in turn make them a more powerful tool for predicting water network pollution dispersion and similar hazardous phenomena.

Author Contributions: Conceptualization, L.G., L.K., S.D., and I.L.; Data curation, L.G. and I.L.; Formal analysis, L.G.; Funding acquisition, L.K.; Investigation, L.G., L.K., and I.L.; Methodology, L.K., S.D., and I.L.; Project administration, L.K. and S.D.; Resources, L.K. and S.D.; Software, L.G.; Supervision, S.D.; Validation, L.G. and I.L.; Writing—original draft, L.G.; and Writing—review and editing, L.G., L.K., S.D., and I.L.

Funding: This research was funded by the Center for Advanced Computing and Modeling, University of Rijeka.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Di Nardo, A.; Di Natale, M.; Guida, M.; Musmarra, D. Water network protection from intentional contamination by sectorization. *Water Resour. Manag.* **2013**, *27*, 1837–1850. [[CrossRef](#)]
2. Grayman, W.M.; Murray, R.; Savic, D.A. Effects of redesign of water systems for security and water quality factors. In Proceedings of the World Environmental and Water Resources Congress 2009, Kansas City, MO, USA, 17–21 May 2009; pp. 1–11.
3. Bello, O.; Abu-Mahfouz, A.M.; Hamam, Y.; Page, P.R.; Adedeji, K.B.; Piller, O. Solving Management Problems in Water Distribution Networks: A Survey of Approaches and Mathematical Models. *Water* **2019**, *11*, 562. [[CrossRef](#)]
4. Kessler, A.; Ostfeld, A.; Sinai, G. Detecting accidental contaminations in municipal water networks. *J. Water Resour. Plan. Manag.* **1998**, *124*, 192–198. [[CrossRef](#)]
5. Adedija, O.; Hamam, Y.; Khalaf, B.; Sadiku, R. Towards development of an optimization model to identify contamination source in a water distribution network. *Water* **2018**, *10*, 579. [[CrossRef](#)]
6. Kranjčević, L.; Čavrak, M.; Šestan, M. Contamination source detection in water distribution networks. *Eng. Rev.* **2010**, *30*, 11–25.
7. Ciaponi, C.; Creaco, E.; Di Nardo, A.; Di Natale, M.; Giudicianni, C.; Musmarra, D.; 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* **2019**, *11*, 1315. [[CrossRef](#)]
8. Krause, A.; Leskovec, J.; Guestrin, C.; VanBriesen, J.; Faloutsos, C. Efficient sensor placement optimization for securing large water distribution networks. *J. Water Resour. Plan. Manag.* **2008**, *134*, 516–526. [[CrossRef](#)]
9. Berry, J.; Hart, W.E.; Phillips, C.A.; Uber, J.G.; Watson, J.P. Sensor placement in municipal water networks with temporal integer programming models. *J. Water Resour. Plan. Manag.* **2006**, *132*, 218–224. [[CrossRef](#)]
10. Hart, W.E.; Murray, R. Review of sensor placement strategies for contamination warning systems in drinking water distribution systems. *J. Water Resour. Plan. Manag.* **2010**, *136*, 611–619.
11. Song, I.; Romero-Gomez, P.; Andrade, M.A.; Mondaca, M.; Choi, C.Y. Mixing at junctions in water distribution systems: An experimental study. *Urban Water J.* **2018**, *15*, 32–38. [[CrossRef](#)]
12. Grbčić, L.; Kranjčević, L.; Lučin, I.; Čarija, Z. Experimental and Numerical Investigation of Mixing Phenomena in Double-Tee Junctions. *Water* **2019**, *11*, 1198. [[CrossRef](#)]
13. Ho, C.K.; O’Rear, L., Jr. Evaluation of solute mixing in water distribution pipe junctions. *J. Am. Water Work. Assoc.* **2009**, *101*, 116–127. [[CrossRef](#)]
14. McKenna, S.A.; Orear, L.; Wright, J. Experimental determination of solute mixing in pipe joints. In Proceedings of the World Environmental and Water Resources Congress 2007, Tampa, FL, USA, 15–19 May 2007; pp. 1–11.
15. Austin, R.; Waanders, B.v.B.; McKenna, S.; Choi, C. Mixing at cross junctions in water distribution systems. II: Experimental study. *J. Water Resour. Plan. Manag.* **2008**, *134*, 295–302. [[CrossRef](#)]
16. Shao, Y.; Yang, Y.J.; Jiang, L.; Yu, T.; Shen, C. Experimental testing and modeling analysis of solute mixing at water distribution pipe junctions. *Water Res.* **2014**, *56*, 133–147. [[CrossRef](#)] [[PubMed](#)]
17. Yu, T.; Tao, L.; Shao, Y.; Zhang, T. Experimental study of solute mixing at double-Tee junctions in water distribution systems. *Water Sci. Technol. Water Supply* **2014**, *15*, 474–482. [[CrossRef](#)]
18. Yu, T.; Qiu, H.; Yang, J.; Shao, Y.; Tao, L. Mixing at double-Tee junctions with unequal pipe sizes in water distribution systems. *Water Sci. Technol. Water Supply* **2016**, *16*, 1595–1602. [[CrossRef](#)]

19. Shao, Y.; Zhao, L.; Yang, Y.J.; Zhang, T.; Ye, M. Experimentally Determined Solute Mixing under Laminar and Transitional Flows at Junctions in Water Distribution Systems. *Adv. Civ. Eng.* **2019**, *2019*. [[CrossRef](#)]
20. Webb, S.W.; van Bloemen Waanders, B.G. High fidelity computational fluid dynamics for mixing in water distribution systems. In Proceedings of the Water Distribution Systems Analysis Symposium 2006, Cincinnati, OH, USA, 27–30 August 2006; pp. 1–15.
21. Gualtieri, C.; Angeloudis, A.; Bombardelli, F.; Jha, S.; Stoesser, T. On the values for the turbulent Schmidt number in environmental flows. *Fluids* **2017**, *2*, 17. [[CrossRef](#)]
22. Valero, D.; Bung, D.B. Sensitivity of turbulent Schmidt number and turbulence model to simulations of jets in crossflow. *Environ. Model. Softw.* **2016**, *82*, 218–228. [[CrossRef](#)]
23. Ho, C.K.; Wright, J.L.; McKenna, S.A.; Orear Jr, L. *Contaminant Mixing at Pipe Joints: Comparison between Laboratory Flow Experiments and Computational Fluid Dynamics Models*; Technical Report; Sandia National Lab. (SNL-NM): Albuquerque, NM, USA, 2006.
24. Romero-Gomez, P.; Choi, C.; van Bloemen Waanders, B.; McKenna, S. Transport phenomena at intersections of pressurized pipe systems. In Proceedings of the Water Distribution Systems Analysis Symposium 2006, Cincinnati, OH, USA, 27–30 August 2006; pp. 1–20.
25. Gilbert, D.; Mortazavi, I.; Piller, O.; Ung, H. Low dimensional modeling of Double T-junctions in water distribution networks using Kriging interpolation and Delaunay triangulation. *Pac. J. Math. Ind.* **2017**, *9*, 2. [[CrossRef](#)]
26. Le, X.H.; Ho, H.V.; Lee, G.; Jung, S. Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. *Water* **2019**, *11*, 1387. [[CrossRef](#)]
27. Tian, Y.; Xu, Y.P.; Yang, Z.; Wang, G.; Zhu, Q. Integration of a parsimonious hydrological model with recurrent neural networks for improved streamflow forecasting. *Water* **2018**, *10*, 1655. [[CrossRef](#)]
28. Saadi, M.; Oudin, L.; Ribstein, P. Random Forest Ability in Regionalizing Hourly Hydrological Model Parameters. *Water* **2019**, *11*, 1540. [[CrossRef](#)]
29. Gedik, N. Least Squares Support Vector Mechanics to Predict the Stability Number of Rubble-Mound Breakwaters. *Water* **2018**, *10*, 1452. [[CrossRef](#)]
30. Granata, F.; de Marinis, G. Machine learning methods for wastewater hydraulics. *Flow Meas. Instrum.* **2017**, *57*, 1–9. [[CrossRef](#)]
31. Kanin, E.; Vainshtein, A.; Osipov, A.; Burnaev, E. *The Method of Calculation the Pressure Gradient in Multiphase Flow in the Pipe Segment Based on the Machine Learning Algorithms*; IOP Conference Series: Earth and Environmental Science; IOP Publishing: Bristol, UK, 2018; Volume 193, p. 012028.
32. Hu, C.; Cai, J.; Zeng, D.; Yan, X.; Gong, W.; Wang, L. Deep reinforcement learning based valve scheduling for pollution isolation in water distribution network. *Math. Biosci. Eng.* **2019**, *17*, 105–121.
33. Vries, D.; van den Akker, B.; Vonk, E.; de Jong, W.; van Summeren, J. Application of machine learning techniques to predict anomalies in water supply networks. *Water Sci. Technol. Water Supply* **2016**, *16*, 1528–1535. [[CrossRef](#)]
34. Sattar, A.M.; Ertuğrul, Ö.F.; Gharabaghi, B.; McBean, E.A.; Cao, J. Extreme learning machine model for water network management. *Neural Comput. Appl.* **2019**, *31*, 157–169. [[CrossRef](#)]
35. Ho, C.K. Solute mixing models for water-distribution pipe networks. *J. Hydraul. Eng.* **2008**, *134*, 1236–1244. [[CrossRef](#)]
36. Jasak, H.; Jemcov, A.; Tukovic, Z. OpenFOAM: A C++ library for complex physics simulations. In *International Workshop on Coupled Methods in Numerical Dynamics*; IUC: Dubrovnik Croatia, 2007; Volume 1000, pp. 1–20.
37. Mills, R. Self-diffusion in normal and heavy water in the range 1–45. deg. *J. Phys. Chem.* **1973**, *77*, 685–688. [[CrossRef](#)]
38. Nielsen, M.A. *Neural Networks and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015; Volume 25.
39. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
40. Chollet, F. Keras: The Python Deep Learning Library. 2019. Available online: <https://keras.io/> (accessed on 24 October 2019).

41. Scholkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2001.
42. 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. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).