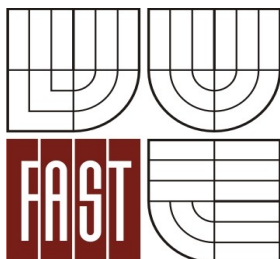


VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY



FAKULTA STAVEBNÍ
ÚSTAV VODNÍHO HOSPODÁŘSTVÍ OBCÍ

FACULTY OF CIVIL ENGINEERING
DEPARTMENT OF MUNICIPAL WATER MANAGEMENT

USING ARTIFICIAL NEURAL NETWORK MODELS TO ASSESS WATER QUALITY IN WATER DISTRIBUTION NETWORKS

VYUŽITÍ MODELŮ NEURONOVÝCH SÍTÍ PRO HODNOCENÍ KVALITY VODY VE VODOVODNÍCH SÍTÍCH

PhD THESIS
TÉZE DISERTAČNÍ PRÁCE

AUTHOR Ing. GUSTAVO ANDRES CUESTA CORDOBA
AUTOR PRÁCE

SUPERVISOR doc. Ing. LADISLAV TUHOVČÁK, Csc
VEDOUCÍ PRÁCE

BRNO 2012

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ
BRNO UNIVERSITY OF TECHNOLOGY
FAKULTA STAVEBNÍ
FACULTY OF CIVIL ENGINEERING
ÚSTAV VODNÍHO HOSPODÁŘSTVÍ OBCÍ
DEPARTMENT OF MUNICIPAL WATER MANAGEMENT

Ing. Gustavo Andres CUESTA CORDOBA

**USING ARTIFICIAL NEURAL NETWORK MODELS TO ASSESS
WATER QUALITY IN WATER DISTRIBUTION NETWORKS**

VYUŽITÍ MODELŮ NEURONOVÝCH SÍTÍ PRO HODNOCENÍ KVALITY
VODY VE VODOVODNÍCH SÍTÍCH

Short Version of PhD Thesis
Zkrácená verze PhD Thesis

Study Field:	Water Management and Water Structures
Supervisor:	doc. Ing. Ladislav Tuhovčák, Csc
Opponents:	

Presentation Date:

KEYWORDS

Water Distribution Systems, Chlorine decay, Artificial Neural Networks, Monte Carlo Method

KLÍČOVÁ SLOVA

Vodárenský distribuční systém, úbytek chloru, umělé neuronové sítě, metoda Monte Carlo

THE DISSERTATION THESIS IS AT ONE'S DISPOSAL:

Department of Municipal Water Management

Faculty of Civil Engineering

Žižkova 17

602 00 Brno

Gustavo Andres Cuesta, 2012

ISBN

ISSN

Contents

1	INTRODUCTION.....	1
1.1	THESIS CONTRIBUTIONS.....	1
1.2	THESIS OUTLINE.....	2
2	STATE OF THE ART.....	2
2.1	ARTIFICIAL NEURAL NETWORK FOR WATER QUALITY IN WATER DISTRIBUTION.....	3
2.1.1	Neural Network Architectures.....	3
3.1	DATA PREPARATION.....	5
3.2	SELECTION OF INPUTS AND OUTPUTS OF THE ANN MODEL.....	5
3.2.1	Factors affecting chlorine decay in a Water Distribution System (WDS).....	5
3.3	CONSTRUCTION OF THE INPUT DATABASE USING THE MONTE-CARLO METHOD.....	6
3.3.1	Simulation of factors affecting chlorine decay.....	6
3.4	CREATION OF THE NEURAL NETWORK MODEL.....	6
3.4.1	Division of models into subsets of parameters.....	7
3.4.2	Training, Testing and Validation.....	7
3.4.3	Performance Evaluation.....	8
4	CASE STUDIES	8
4.1	ANN FOR PREDICTION OF CHLORINE IN A WDS - CASE STUDY: NAŠIMĚŘICE, CZECH REPUBLIC.....	9
4.1.1	Current state of the drinking water supply system.....	9
4.1.2	Hydraulic assessment of water supply system Našiměřice.....	10
4.1.3	Construction of ANN Models.....	11
4.1.4	Outcomes of the research.....	13
4.1.5	Results and further work.....	14
4.2	DEVELOPMENT OF NEURAL NETWORK MODEL FOR FORECASTING CHLORINE CONCENTRATIONS IN A PRESSURE ZONE OF A WDS - CASE STUDY: BRNO-KOHOUTOVICE, CZECH REPUBLIC.....	15
4.2.2	Water quality model (chlorine decay) for the selected pressure zone.....	18
4.2.3	ANN model for simulation of residual chlorine concentration.....	20
4.2.4	Construction of the input database using the Monte-Carlo method.....	21
4.2.5	Creation of the Neural Network Model.....	22
4.2.6	Outcomes and results.....	27
5	CONCLUSIONS AND DISCUSSION.....	28
5.1	Fulfillment of Thesis Goals.....	29
	Bibliography.....	30

1 INTRODUCTION

The purpose of a Water Distribution System (WDS) is to make water available to customers with at least acceptable pressure, flow, continuity and water quality. Water quality can be measure in terms of the set of water quality parameters, for example, taste, odor, appearance and chlorine concentration between others parameters. Maintaining water quality through the WDS until the point of consume is one of the most challenging task faced by the utilities, taking in consideration the components of the WDS, such as pipe materials, tanks, valves etc. and other risks related to water distribution. Inside the Water Treatment Plant (WTP) there is a combination of processes for drinking water treatment. Principal processes of a conventional WTP include; aeration, coagulation, sedimentation, filtration and disinfection. A disinfectant residual should be maintained throughout the distribution system at all times.

Although it is recognized that excessive levels of disinfectant may result in taste and odor problems, it is therefore recommended that a disinfectant residual be maintained and monitored daily throughout the entire system. The most common disinfectant used in water management is chlorine. The concept of Residual Chlorine Concentration is associated with disinfection durability. There is, however, another problem regarding disinfection in a WDS. It is a phenomenon known as chlorine decay; chlorine reacts with other components along the system and its concentration decrease. Knowing the physic-chemical aspects behind chlorine decay is important in order to develop a strategy capable of disinfecting a WDS and, at the same time, preserving water quality until the point of use, without using more disinfectant than necessary.

The objective of this research is to develop an Artificial Neural Network (ANN) model that can simulate residual chlorine decay at selected nodes under a pressure zone of a water distribution system (WDS), with the advantages of a simple functional form and good accuracy. In addition, it can also be employed to estimate residual chlorine decay in the rest of points inside the network and remark the affected areas in which high or low levels of chlorine are presented in the system by using the computational model EPANET 2.0.

1.1 THESIS CONTRIBUTIONS

Main contributions of this thesis are seen in:

- Demonstrating that Artificial Neural Networks (ANN) are able to predict chlorine concentration in the distribution networks, case studies in Czech Republic with a simple functional form and good accuracy.
- Creating a specific database for each distribution network studied, with historical data of parameters affecting chlorine decay, which include; pH, temperature, turbidity, flow and initial chlorine.
- Presenting a sensitivity analysis of the input parameters to estimate which of them have the most influence in chlorine decay.

- Using Monte-Carlo method for simulation of input parameters affecting chlorine decay when not enough data will be available to run the ANN models.
- Showing that free chlorine concentration predicted with ANN technique, can be used in a physical based model (EPANET 2.0) to help with the calibration of the same in the remaining nodes within the WDS under consideration or for identification of areas affected by the maximum or minimum risk of significant changes of chlorine.

1.2 THESIS OUTLINE

The thesis is organized into the following main chapters:

- *State of the Art* outlines the research status in the fields of the standard methods used in water quality models, general Artificial Neural Network (ANN) modeling and ANN for water quality in water distribution systems (WDS). *Chapter 2.*
- *Theoretical Methodology for ANN Models Design* builds necessary theoretical foundation as a starting point to understand the thesis and research implementation presented in the next chapter. *Chapter 3.*
- *Case Studies* describes the implementation or test that I have conducted. Main ideas presented in previous chapter are verified with tests and experiments and then evaluation is made to the methods that have been used in the implementation or test. It also reviews the results providing detailed description of the outcomes. *Chapter 4.*
- *Conclusions and discussion* summarizes research observations. *Chapter 5.*

2 STATE OF THE ART

Predictive models are divided into three categories of models - Conceptual, Physical and Numerical [1]. Each of them represents a simplified system for understanding the behavior of a complex system. Models can usually address only a specific individual process, e.g. the user can control only part of the system and not the entire system. In water quality modeling it should be taken into account factors affecting in the entire distribution system, such as:

- The variation of the physical properties in the pipe (material, diameter, age of pipe in the distribution system, etc.).
- Variations in the hydraulic regime in the system under consideration (rate of water flow, velocity, time enhancements, etc)
- Temperature, pH, turbidity of water.

Numerical models that include empirical models are intended to describe the behavior of the system by mathematical equations. This thesis is based on a study of water quality parameters precisely with empirical model so called Artificial Neural Network (ANN) to predict chlorine concentration in a WDS.

2.1 ARTIFICIAL NEURAL NETWORK FOR WATER QUALITY IN WATER DISTRIBUTION

The applications of ANN in water management are widespread and vary from optimization of measuring networks, operational water management, prediction of drinking water consumption, on-line steering of waste water treatment plants and sewage systems, up to more specific applications such as establishing a relationship between the observed parameters such as pH, temperature, turbidity and chlorine or chloramine in a drinking water supply system. Especially where processes are complex, neural networks can open new possibilities for understanding and modeling these kinds of complex processes. ANNs are extensively applied for assessment purposes like rainfall-runoff modeling, water quality prediction in natural flows, approximating ecological relations. They have also been applied for optimal reservoir operation. A remarkable number of publications on application of fuzzy logic approach for process control in waste water treatment plants for deriving optimal control actions are available. Problem of real-time optimal operation of water related systems has been investigated by using neural networks, fuzzy logic approach and with neuro-fuzzy approach. The list goes on.

In Czech Republic several researches have been done using ANN in water management field. Grünwald et al. (2004 – 2008) [22], worked in a project called Innovation Process of Water Treatment Plant (WTP) and High Security of Water Quality in WDS (Inovace procesu úpravy vody a zabezpečení vysoké kvality pitné vody v distribučních sítích) in which used ANN for optimization of coagulant dosage in a WTP case of study - Plav (Czech Republic) and also for residual concentration of disinfectant (Chloramine) in the WTP. Starý and Nacházal (2004) [24] also published an interesting study called Using Artificial Intelligence in Water Management (Využití metod umělé inteligence ve vodním hospodářství) as an introduction of this qualitatively new method explaining the principles and showing examples of possible applications.

The advantages of an Artificial Neural Networks for quality water management can be summarized as follows:

- They can imitate the control actions of human operators through the description of the system behavior using historical data
- They are inherently non-linear and therefore, able to perform the control actions that are not possible purely with a traditional linear control

2.1.1 Neural Network Architectures

One of the most popular architectures in neural networks is the Multi-Layer Perceptron (MLP). Most of the networks with this architecture use the Widrow-Hoff (Delta) rule as their learning algorithm and the logistic function as the transfer function of the units of the hidden layer (the transfer function is in general non-linear for these neurons). These networks are very popular because they can approximate any multivariate function relating the input to the output [5]. The

Multi-Layer Perceptron (MLP) using the backpropagation training algorithm is the most widely used neural network for forecasting and prediction applications. MLPs generally consist of three layers: an input layer, a hidden layer and an output layer. However, MLPs may contain more than one hidden layer. Each layer consists of nodes or neurons, which are connected to nodes in the previous and following layers by connections.

3 THEORETICAL METHODOLOGY FOR ANN MODELS DESIGN

The systematic modeling procedure proposed to be implemented in this thesis can be seen in **Figure 1**. The main steps for modeling chlorine residual in a WDS using ANN involve: data preparation, input selection, Monte Carlo simulation for missing values, data division and selection of subsets, model creation, model calibration and performance evaluation.

Analytical techniques that can be used to help with the input determination process are the coefficients of correlation and sensitivity analysis. Performance evaluation can be then used to test the accuracy of each calibrated model, which included the Root Mean Squared Error (RMSE). Another important step is the creation of the hydraulic model. Water quality studies required high level calibration to avoid misleading data or error in the simulation. ANN uses historical data for prediction of parameters. When creating ANN models, some data may be missing from the original database.

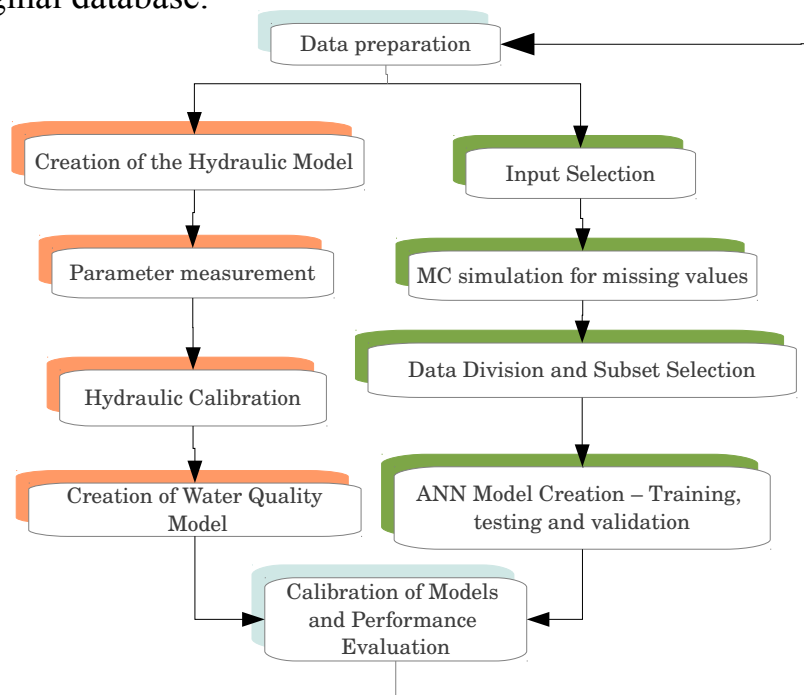


Figure 1: Modeling methodology for used in the Thesis

Modelers usually replace the missing data with the average of the sample or simply delete or ignore the complete row, causing the loss of important data. The

Monte Carlo (MC) method can be used to generate a database of each parameters affecting chlorine decay in the several nodes studied inside the WDS. Monte Carlo simulation can be performed to fulfill the missing values (if any) in the original database, as it provide flexibility, manage the uncertainty and even provide more accurate results than simple descriptive statistics (e.g. the average value).

3.1 DATA PREPARATION

Historical data of several parameters that are supposed to influence chlorine decay should be gathered to be successful in the application of ANN models. Usually utilities measure parameters such as pH, temperature, turbidity, color, manganese, iron, conductivity, e. coli, coliform bacteria etc. Those parameters are commonly used as indicator of sanitary quality of water. Utilities store the data in databases applications such as Supervisor Control and Data Acquisition (SCADA) systems, metadata in excel or even in Geographic Information Systems (GIS) databases. Several utilities already use GIS to store the data and it is easy to gather the information needed to create the hydraulic model, also the computational water modeling software MIKENET uses GIS function combined with EPANET based model and it can be easy to transfer the information between both applications. The creation of the hydraulic model is relevant in this thesis as it allows not only to compare the performance of ANN and first-order kinetic model used in EPANET but also hydraulic information such as flow, pressure etc, can be useful for the creation and evaluation of the ANN model.

3.2 SELECTION OF INPUTS AND OUTPUTS OF THE ANN MODEL

Successful application of artificial neural network model requires proper input data selection. The better way to choose the inputs for the ANN model is to minimize the size of the network and at the same time maintaining acceptable performance [7].

The better way to choose the inputs in practice is related to two primary considerations: Prior knowledge about the process and Availability and quality of the required data in the training set. The study will be run in a selected area of the WDS, which will be at least connected to a reservoir or a tank and will include the chlorine dosage for the zone selected.

3.2.1 Factors affecting chlorine decay in a Water Distribution System (WDS)

Some relevant parameters affecting chlorine decay are following:

- Physical Parameters: Pipe roughness (Influenced by pipe age, pipe material and water quality) and Diameter.
- Hydraulic Parameters: Flow, Pressure and Velocity
- Physic-chemical Parameters: Chlorine added to water (initial chlorine, at the begging of the network or in the treatment plant), pH, Turbidity and Temperature

These parameters will be taken for the prediction of chlorine decay in a special zone in a WDS and a database with several values of each parameter will be simulated using MC Method and the historical information provided by the water utility.

3.3 CONSTRUCTION OF THE INPUT DATABASE USING THE MONTE-CARLO METHOD

For analyzing the factors affecting chlorine decay the general Monte Carlo steps are modified as given below:

- Domain of possible inputs – Varies from minimum to maximum values of Chlorine added to water (initial chlorine), Flow, pH, Temperature, Turbidity, as per historical data received from water utility.
- Random Number Generator – The Software STATISTICA 10 will be used to generate random numbers within the domain.
- Artificial Neural Network – To aggregate the results Artificial Neural Network is plotted in order to calculate the chlorine decay.

3.3.1 Simulation of factors affecting chlorine decay

An initial simulation will be performed starting with the hydraulic parameters followed by the physic-chemical parameters. A normal distribution will be follow for the generation of the random number. For the analysis of the measured data, Statistica 10 from Statsoft uses a function called *Distribution Fitting* this option allows to verify whether the measure values follow a normal distribution and after the confirmation we can run a simulation using the Monte-Carlo Method proposed. It is planned to calculate a total of 3000 readings that will be generated for each factor or parameter. Physic-chemical parameters lead a continuous probability distribution and normal distribution can be used for each of the parameters to generate the database. As an example, we can take initial chlorine data provided by the water utility. We suppose the values of chlorine are variables that follow a normal distribution. Our objective will be the creation of a large database and after that, determine the expected chlorine in the network. The results obtained from the random generator will then be analyzed using the function of mean, standard deviation and confident interval of the sample.

3.4 CREATION OF THE NEURAL NETWORK MODEL

The selection of the ANN type depends on the type of problem to solve and the characteristic of data obtained. The selection is often difficult as it requires experience and sometimes is better to try with different types. The function of a single neuron is so simple that it can not solve by it self a complicated problem, that is the reason for the creation of a network of neurons, which are interconnected with each other. The following are different types of ANN:

- Mulit-Layer Perceptron Neural Network – MLP

- Radial Basis Function (RBF)
- Self-Organizing Feature Map – SOFM
- Adaptive Linear Neuron - ADALINE, MADALINE
- Bayesian Networks (Probabilistic Neural Networks-PNN and General Regression Neural Network - GRNN)

In several researches it has been proved that MLP is able to predict chlorine decay in WDS in a very accurate way. For this thesis MLP is used as the selected ANN type to predict chlorine concentration. The reason is because the data used is mainly numerical and our objective is to predict a given dependent variable (mostly a regression problem) and RBF for example have better performance for classification than regression or prediction.

3.4.1 Division of models into subsets of parameters

It is recommended to divide the data depending on the section where the data was gathered, that is to say, parameters that were measured at the initial part of the system can be called initial parameters e.g. Initial chlorine (chlorine added to the water at the beginning of the system), Flow at the output of tank, measure of pH and/or turbidity in the same point of chlorination. Parameters measured in specific nodes, can be called Local parameters, referring to measurements of pH, turbidity and chlorine concentrations in each specific node. The criteria to divide the parameters into subsets of models are the following:

- Run a preliminary ANN model which includes all the parameters suspected to influence chlorine decay.
- Run a Sensitivity Analysis of each parameter to obtain a better view of the influence weight.
- Based on the sensitivity analysis create a second model with only the parameters that have high influence on chlorine decay.
- Create a third model based on the Initial parameters with high influence in chlorine decay and only those local parameters.
- Continue with a combination of input parameters to compare the model performance.

There are several ways to combine the input parameters. Only the experience and the availability of data can suggest a good combination and provide good performance in the models.

3.4.2 Training, Testing and Validation

To calibrate the ANN models is needed to Test and Validate the Trained model. There are several approaches for testing and validation of ANN models. The most known approach is to divide the data into percentage for each phase, e.g. it can be used the 50% of the data for the training phase and the 25% and 25% of the data case be used for testing and validation phase respectively. When the data is divided into percentage, usually the values for each phase are taken randomly. Also using

Generic Algorithm the data can be successfully divided into the three phases depending on the criteria of the modeler. A different approach also can be used to test and validate the trained model. Depending on the type of data and availability, all the historical data can be used to train the model and the new measured data done in the calibration of hydraulic model for example, can be used in the testing and validation phase.

3.4.3 Performance Evaluation

The ANN model will predict chlorine decay concentration only in few selected points or nodes inside the WDS. The second step is to compare the results from ANN model with the entire system using the computational program EPANET 2.0. The objective of this analysis is to explore the system response to changes during a period of time and to check if there exist some zones affected with high or low among of chlorine. The way these two computational models (ANN and EPANET) will be weighed against is shown in the program structure of the **Figure 2**.

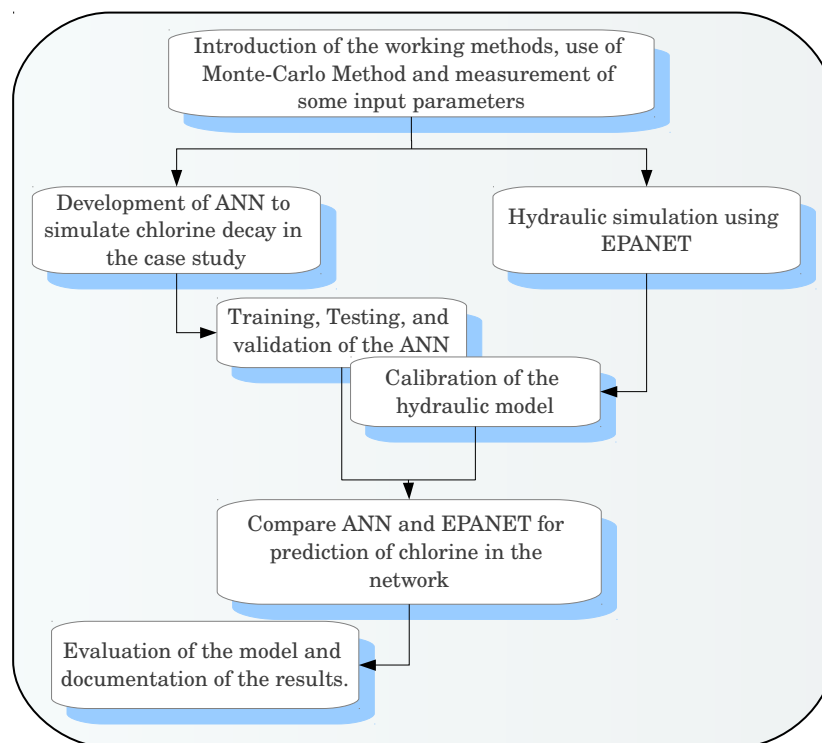


Figure 2: Program Structure

The next step should be the performance of the quality model to simulate chlorine decay in EPANET, the calibration of this quality model is going to be related with the data obtained from the ANN models. The comparison of both results (ANN and EPANET) and eventually the creation of charts showing errors, relationship and difference between the variables have to be necessary for the evaluation of the models.

4 CASE STUDIES

In this chapter will be described the experiments and tests conducted to evaluate the methodology proposed in chapter 3. The main ideas presented in the previous chapter are verified with tests and experiments. Also in the present chapter it will be reviewed the implementation of the methodology and will be provided an evaluation and detailed description of the results.

Mainly this chapter is divided in two parts. The first part refers to the study of the historical data collected by the water utility (Vodárenská Akciová společnost a.s.) in a four year database which also includes the on-site measured data during the time for the study of the project. The study was done in a real distribution system in a town called Našiměřice in the Czech Republic. The second part of the chapter is related only to the prediction of chlorine concentration using ANN models but at the same time, applying Monte-Carlo method to fulfill the cases where some values were missing from the raw data base. Historical data was gathered from the water utility (Brněnské vodárny a kanalizace a.s., BVK) using SCADA system and information stored in GIS application as well as on-line measurement conducted during the project.

4.1 ANN FOR PREDICTION OF CHLORINE IN A WDS - CASE STUDY: NAŠIMĚŘICE, CZECH REPUBLIC

The project is focus on the methods for evaluation the available historical data of water quality and the investigation of the impact for selected physical parameters of water quality and its development in a water distribution system. It will be solved by creating a model using statistical methods to identify and predict the evolution of selected water quality parameters. Multiple Linear Regression (MLR) based on the least square approach and Multi-Layer Perceptron (MLP), which is an Artificial Neural Network (ANN) architecture capable of predict any continues variable were used in this case of study. The performance of MLP and MLR are evaluated using 4-years old database set of inputs collected in the city of Našiměřice Czech Republic.

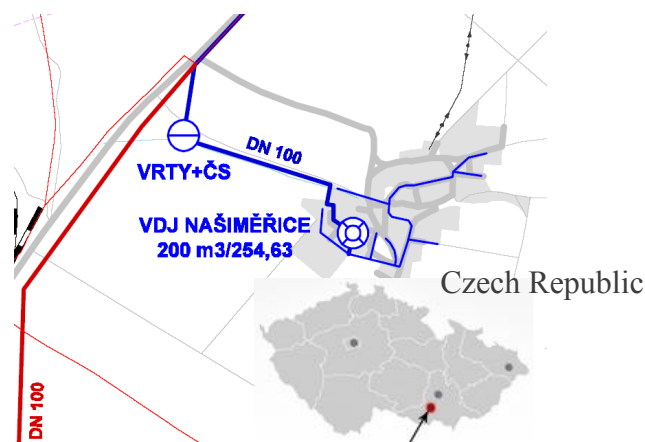


Figure 3: Scheme water supply system Našiměřice

4.1.1 Current state of the drinking water supply system

Našiměřice village is situated about 10 km west of Pohořelice (Czech Republic), In the village was built a water supply for public use, whose owner is an association of municipalities Znojmo Water and Sewage Systems and the operation is provided by Vodárenská akciová společnost a.s. division of Znojmo. The Population in the village is 207. The drinking water supply system is constructed by pumping the water from a well effluent named HV3 to Našiměřice town over the power of Iron pipe lines DN100 and PVC 90 to a hydroglobe tank of 200 m³. See **Figure 3**. The water is pumped from a well named HV3 to yield 1.7 liters/sec, the discharge into Našiměřice height is 49 m. The total length of pipe is about 16,183 km.

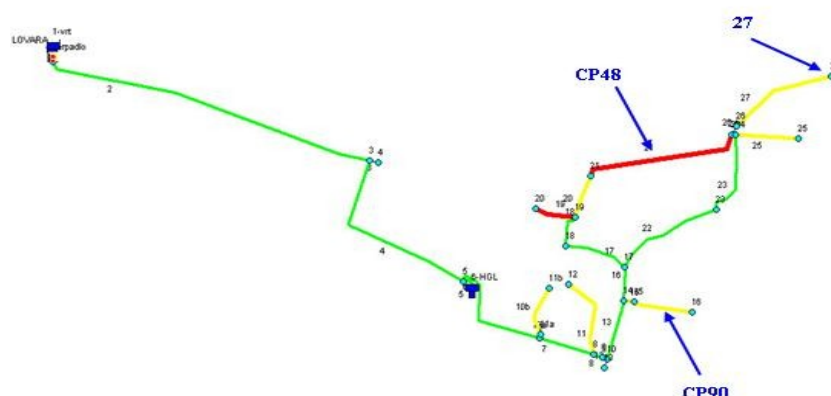


Figure 4: Našiměřice, Czech Republic, Water supply system

4.1.2 Hydraulic assessment of water supply system Našiměřice

Based on the hydraulic model provided by water system operator Vodárenská akciová společnost a.s. it was build a model within this project updated to the year 2010. The fundamental basis is the calculation of the demand of water in the system. It was used the operating data taken from real consumption over the past years. Hydraulic assessment was run using the computational model EPANET 2 and the hydraulic model includes 30 nodes, 28 sections, water source—well HV 3, pump Lovara L4GS11T and hydroglobe of 200m³ volume. The distribution network includes one independent circuit and a pressure zone. Predominant material is cast iron, followed by PVC and steel. Was then performed in a quasi-dynamic analysis, i.e. the simulation took place in the time step - in this case was 1 hour time step for two days.

Table 1: Overview of water demand year 2010

Town	Maximum of Service		
	Q _p (l/s)	Q _m (l/s)	Q _h (l/s)
Našiměřice	0,29	0,44	<u>0,80</u>

Data used in the construction of the hydraulic model

Test were taken in town Našiměřice during the month of September in 2010 to make the calibration of the hydraulic model in EPANET 2.0. Two pressure-meter Meinecke Cosmos CDL-2 were located in nodes CP90 and CP48, and a hydrant in node 27, a summary of the parameters used in the hydraulic model are shown in **Table 2**.

Table 2: Našiměřice Water supply system – Overview and demand in nodes

Section	N1	N2	L (m)	Material	DN (mm)	N. of connect.	Section demand	Nodes	Node demand (%)	Node demand	Ground elevation	Year of commissioning	Age of pipes (Years)	Lifespan from theoretical life (%)
1	1	2	15.6	PVC	110	0	0.0	1	0.0	0.0000	225.97	1964	16	32
2	2	3	550.0	LT	100	0	0.0	2	0.0	0.0000	224.49	1965	45	60
3	3	4	9.6	LT	100	0	0.0	3	0.0	0.0000	218.75	1965	45	60
4	3	5	324.9	PVC	90	0	0.0	4	0.0	0.0000	218.42	1994	16	32
5	5	6	8.0	LT	100	0	0.0	5	0.0	0.0000	226.02	1974	36	48
6	6	7	173.6	LT	100	0	0.0	6	0.0	0.0000	226.02	1974	36	48
7	7	8	93.1	LT	100	3	3.7	7	1.9	0.0147	225.11	1974	36	48
8	8	9	15.4	LT	100	0	0.0	8	3.7	0.0295	225.06	1974	36	48
9	9	10	5.7	PVC	90	0	0.0	9	0.0	0.0000	224.80	2000	10	20
10a	7	11a	5.3	PVC	90	0	0.0	10	4.3	0.0344	224.75	2000	10	20
10b	11a	11b	82.8	OC	80	5	6.2	11a	3.1	0.0246	225.25	1974	36	111
11	8	12	142.8	OC	25	3	3.7	11b	3.1	0.0246	223.51	1974	36	111
12	10	13	15.5	PVC	90	1	1.2	12	1.9	0.0147	223.22	2000	10	20
13	10	14	101.9	PVC	90	6	7.4	13	0.6	0.0049	224.80	2000	10	20
14	14	15	17.9	PVC	90	1	1.2	14	6.2	0.0492	223.70	2000	10	20
15	15	CP90	26.0	PVC	90	2	2.5	15	1.9	0.0147	223.50	2002	8	16
CP90	CP90	16	68.8	PVC	90	3	3.7	CP90	3.1	0.0246	223.70	2002	8	16
16	14	17	46.5	PVC	90	3	3.7	16	1.9	0.0147	220.00	2000	10	20
17	17	18	108.2	OC	80	1	1.2	17	4.3	0.0344	221.78	1970	40	123
18	18	19	53.2	OC	80	3	3.7	18	2.5	0.0197	215.38	1970	40	123
19	19	20	63.5	OC	80	2	2.5	19	8.6	0.0688	214.31	1965	45	138
20	19	21	72.0	OC	80	9	11.1	20	1.2	0.0098	214.83	1970	40	123
21	21	CP48	254.7	LT	100	5	6.2	21	8.6	0.0688	214.97	1970	40	53
CP48	CP48	22	77.0	LT	100	13	16.0	CP48	11.1	0.0885	216.32	1970	40	53
22	17	23	189.7	LT	100	3	3.7	22	8.0	0.0639	214.25	1991	19	25
23	23	24	145.7	LT	100	0	0.0	23	1.9	0.0147	213.39	1991	19	25
24	22	24	1.2	LT	100	0	0.0	24	5.6	0.0442	214.25	1974	36	48
25	24	25	109.6	LT	100	9	11.1	25	5.6	0.0442	213.01	1974	36	48
26	22	26	13.8	LT	100	0	0.0	26	5.6	0.0442	214.72	1970	40	53
27	26	27	183.0	LT	100	9	11.1	27	5.6	0.0442	219.32	1974	36	48

4.1.3 Construction of ANN Models

The database on the network considered in this study was constructed by collecting the available data provided by the water utility (Vodárenská Akciová společnost a.s.) of the historical data tests of free chlorine, pH, Temperature, flow among other water quality and hydraulic parameters. Details of input subsets used are shown in **Table 3**. The indicators were classified in 2 categories, according to the type of value (Temperature, pH and Flow were selected as continuous input and Pipe material, Diameter and Age of pipes as categorical inputs for Model 1). The descriptive statistics of the database are shown in **Table 4** were the categorical values were transform to numerical for statistical purpose and to be used in Model 2 and Model 3. It includes water quality test taken during the observation period

(September 2010), see **Figure 5**. The logs of Pipe material, diameter and age of pipes were used instead of their actual values to use descriptive statistics and Multi-Linear regression results. Categorical variables were converted into number of separate variables, a variable with different levels was transformed into a normalized number for each level, for instance, the age of pipes has 4 levels, the pipe material has 3 levels and each of them was represented in a numerical way.



Figure 5: Instruments installation for measurement in Našiměřice

Three models were constructed, they are classified according to the input indicators; two of them were done using ANN and one was done using Multi-Linear Regression Analysis (MLR) to compare results, potential and accuracy performance. In Model 1 all the data collected was used including categorical and numerical values, in Model 2 all the data collected was used as well but the categorical values were transformed into numerical and Model 3 was constructed using MLR analysis. A summary of data used in each model is also shown in **Table 4**.

Table 3: Details of input subset and variable used for the Models

Type	Variable	Subset of variable selected in each Model			Location (All variables were measured in the each location listed)
Inputs		1	2	3	
	Temperature	x	x	x	ZD WC
	pH	x	x	x	č.p. 117, kohout na zahradu
	Flow	x	x	x	č.p. 83, OÚ, kuchyňka
	Pipe material	x	x	x	č.p. 44, dvůr - venkovní kohout
	Diameter	x	x	x	VODOJEM, Našiměřice – odtok
	Age of pipes	x	x	x	odkalkení naproti č. 113
					VODOJEM, Našiměřice – přítok
					č.114 venkovní kohout
					č.p.20 - kuchyně, umyvadlo
					č.16 kuchyň
					Obecní úřad - kuchyňka – umyvadlo
					č.p. 90, kohout na zahradu
					č.p. 118, kohout na zahradu
					č.p. 48, Úřad - venkovní kohout

Table 4: Descriptive statistics of the database (2007 -2010)

Variables	Valid N	Mean	Minimum	Maximum	Std. Dev.
Temperature	63	14.84603	7,4	20,4	3,414949
pH	63	7.39079	7,22	7,9	0.128982
Flow	63	0.33143	0,0062	0,80	0.157556
Pipe material	63	1.74603	1	3	0.506992
Diameter	63	99.36508	80	110	6.444056
Age of pipes	63	2.5873	1	4	1.351641
Free Chlorine	63	0.15508	0,01	0,60	0.14677

4.1.4 Outcomes of the research

The type of ANN model used was Multi-Layer Perceptron (MLP). The MLP is a feed forward ANN model that maps the sets of input data onto a set of appropriate output. MLP utilizes the supervised learning technique “backpropagation” to train the network. For the model calibration, the data set was treated using the following analysis: the data set was divided into three subsets, The first subset (50% of the database) was used to train the network (Learning phase), the second part (25% of the database) was used to test the ANN models in order to determine when to stop the training stage (Testing phase) and the last part was used to validate the model data not involved in the training process (Validation phase). The training, testing and validation results for each of these models are given in **Table 5** (Models 1-3). By the virtue of the MLP architecture, the training set can be predicted to a high level of accuracy in comparison with MLR the MLP achieved a significantly lower error for the training, testing and validation sets and that is the prove that the MLP model was able to find nonlinear relationship between variables. Chlorine decay in a pipeline is a complex phenomenon, therefore it is not surprising that MLP was able to provide better predictions for this case of study when compared with linear regression model [2]. The set of predicted values produced by MLR are shown in **Figure 8** The performance obtained by MLR was 0.70803 which is significantly lower than the achieved by MLP in **Figure 6** and **7**.

Table 5: Model Performance

Model No.	Model Type	Training perf.	Testing pef.	Validation perf.
1	MLP	0,97838	0,87358	0,89696
2	MLP	0,90734	0,73009	0,94906
3	MLR	0,70803	0,70803	0,70803

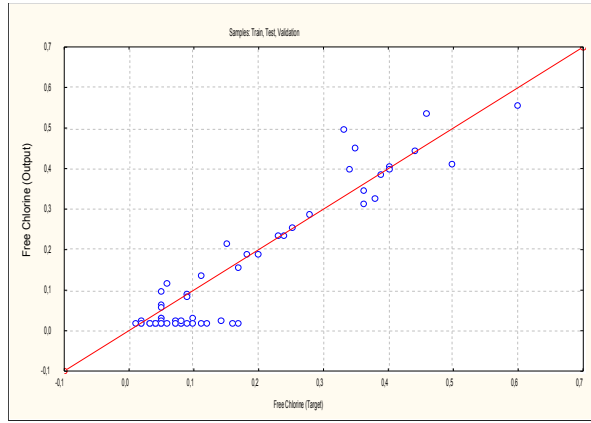


Figure 6: Training, Test and validation Predictions for MLP Model 1

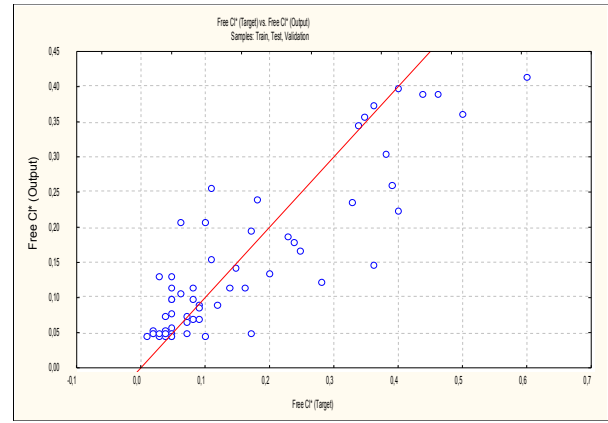


Figure 7: Training, Test and validation Predictions for MLP Model 2

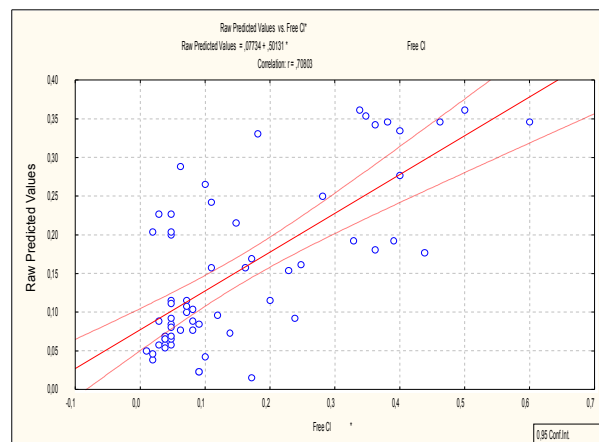


Figure 8: Predictions for Multiple Linear Regression Model 3

4.1.5 Results and further work

This research presented a study of the use of ANN approach for the evaluation and prediction of free chlorine within the water distribution network in Našiměřice town in Czech Republic. The performance of this approach was analyzed on a 4-year database of water quality and hydraulic parameters. Two ANN models were constructed and one model was created using statistical approach MLR. From the results obtained in this study MLP models were found to be useful tools for prediction of free chlorine in a WDS. The MLP models developed in this research were found to outperform MLR model significantly, suggesting that they are able to use non-linear relationship between the variables used in the input layer. In this study, ANNs is capable to predict free chlorine at Našiměřice town, the next level of complexity in this research is to repeat the construction of ANN models using a larger data set that also includes data for all season and the use of free chlorine within the network distribution system as an input indicator to predict chlorine residual in a specific node within the WDS. These predicted parameters can be used

in a physical based model (EPANET) for prediction of the same in the remaining nodes within the WDS under consideration or for identification of areas affected by the maximum or minimal risk of significant changes of this parameter.

4.2 DEVELOPMENT OF NEURAL NETWORK MODEL FOR FORECASTING CHLORINE CONCENTRATIONS IN A PRESSURE ZONE OF A WDS - CASE STUDY: BRNO-KOHOUTOVICE, CZECH REPUBLIC

Historical data was gathered from the water utility (Brněnské vodárny a kanalizace a.s., BVK) using SCADA system and information stored in GIS application as well as on-line measurement conducted during the project. The research was performed in the water pressure zone of Kohoutovice, which is a district within Brno city. This project deals with the use of MLP - Neural Network for solving the problem of predicting chlorine decay in the pressure zone of Kohoutovice. Hydraulic and water-quality parameters will be first introduced in few selected locations in the distribution system, then will be extended to the whole investigated system. Based on the considerations, it was proposed by *Brněnské vodárny a kanalizace a.s.* (BVK) the following pressure zones for the study:

1.3 VDJ Myslivna

1.3.2 Zemní VDJ Kohoutovice

1.3.2.1 Věžový VDJ Kohoutovice

The water comes from the VDJ Bosonohy to Zemní VDJ Kohoutovice and to VDJ Myslivna after that it is distributed to the system in separated pressure zones. Initial chlorine is measured at the VDJ Bosonohy. The chosen pressure zone is the *Zemní VDJ Kohoutovice*. In **Figure 9** it is shown as the pipe circuit color blue. This pressure zone was chosen because it filled all the specifications stated before and also there was already a hydraulic model and some hydraulic parameters as Pressure and Flow can be determined easily.

Description of the current situation in the drinking water supply system

City District Brno-Kohoutovice lies approximately 7.3 km west of Brno-center. In the district was built a water supply for public use, operated by Brněnské vodárny a kanalizace, a.s. The total population in the urban area is 13 338. The drinking water is obtained by pumping the water from the reservoir Čebín to Bosonohy tank. Pipeline “*Vírský oblastní vodovod*” (VOV – The Vir Regional water main system) under reservoir Čebín currently bring the water by gravitation to the Bosonohy tank. From the Bosonohy tank the water is pumped to the tanks Kohoutovice and Myslivna. The drinking water is distributed by gravity to the pressure zone ($Q_h = 22,2$ L/s) using pipes that ranges from steel, PVC, fiberglass, cast and ductile iron, in diameters as 80, 100 and 300 mm.

Bosonohy tank has two chambers and the volume is 6550 m³ with maximum water level of 320 MSL, the water column height is of 6.5 meters. Kohoutovice tank has two chambers and volume is 3000 m³ with maximum water level of 415 MSL,

the water column height is of 5 meters. Description of the main delivery pipe lines from tank Bosonohy to tank Kohoutovice:

- DN 300 fiberglass, year 2000, about 1372 m (in the tunnel VOV)
- DN 300 ductile iron, year 2000, about 290 m
- DN 300 ductile iron, year 2005, about 252 m



Figure 9: Water supply scheme

Model Creation, Measurement and calibration

The model consisted of approximately 287 nodes, 302 sections, 2 reservoir (Bosonohy and Kohoutovice) and two pumps. The distribution network includes one independent pressure zone together with the delivery pipe line from Bosonohy to Kohoutovice tank and two pressure reduction valves that are also connected to the pressure zone.

For the model calibration it was used the time series data in tank Bosonohy and the pump station from Bosonohy to Kohoutovice. The data was provided by BVK and it contains out-flow to Kohoutovice distribution network from the tank, water levels in tanks Bosonohy and Kohoutovice, in-flow to Bosonohy tank and pumped water from Bosonohy tank to Kohoutovice tank. For model calibration it was used the measured data obtained by BVK during September and October 2011.

The parameters chosen for calibration were flow and water level in Kohoutovice tank, as they are straightforward for measurement and calibration. For the water-quality model, it was collected constituent concentrations measured at these points as well. Selection of additional sampling points within the system was also taken.

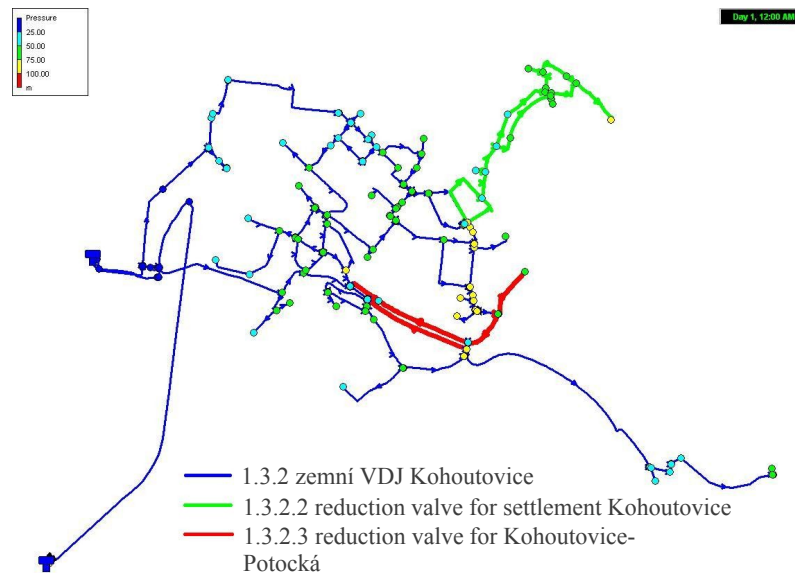


Figure 10: Hydraulic model run in EPANET 2.0

EPANET 2.0 allows the user to compare results of a simulation against measured field data. This can be done via Time Series plots for selected locations in the network or by special Calibration Reports that consider multiple locations. See **Figure 11** where the Kohoutovice tank level is compared with the observed data obtained by BVK measurements on-site, this measurement were compared using the software MS Office Excel. The Root Mean Squared Error was 0.54 for a 48 hours of this simulation.

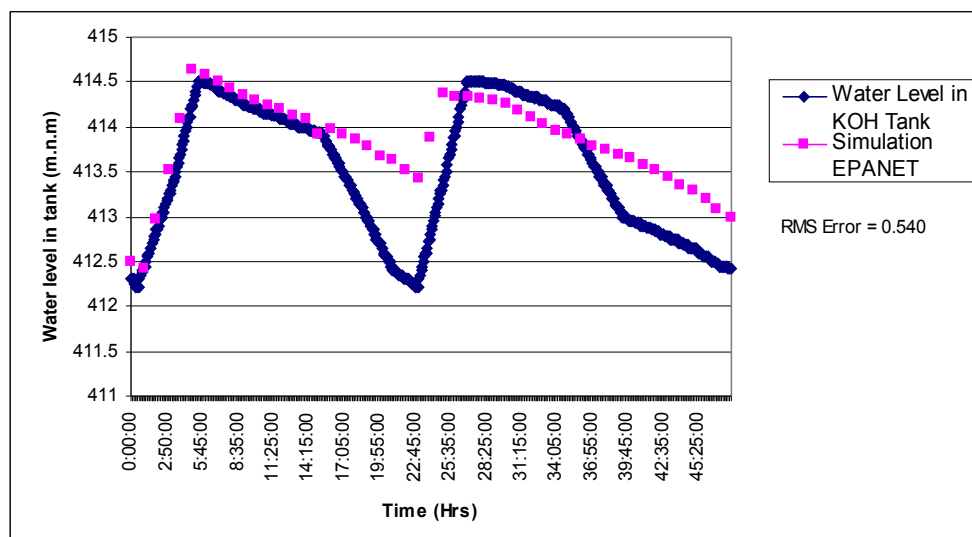


Figure 11: Time-Series plot Calibration for Tank level in Kohoutovice

Other results from the hydraulic calibration can be seen in the **Figures 12**. Where the quantity of flow was calibrated using the data measured from the flow-meter installed at link 327409. The Root Mean Squared Error from the calibration was 3.449, for the 48 hours of simulation.

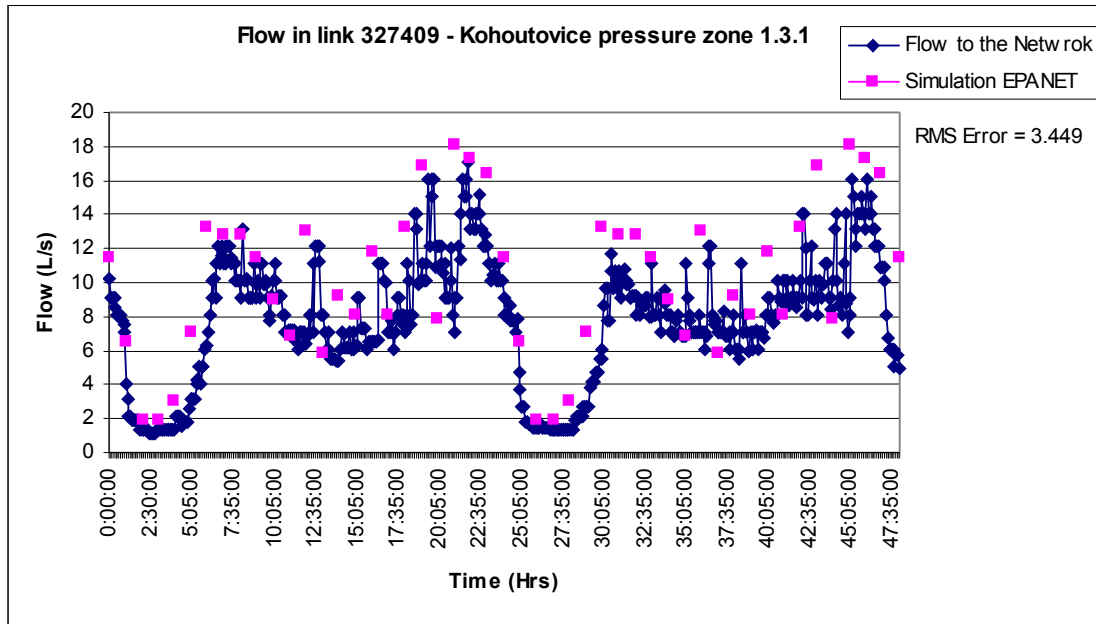


Figure 12: Calibration Plot in Excel for Flow values at Link 327409

4.2.2 Water quality model (chlorine decay) for the selected pressure zone

To assure the correct use of the water quality simulator, incorporated in the EPANET, one must conduct a model calibration process. This consists of attributing the correct values to K_b - Bulk reaction coefficient and K_w - Wall reaction coefficient

K_b - Bulk Reaction coefficient

To determine decay coefficient K_b in Kohoutovice pressure zone 1.3.1 it was used several bottle tests using the instrument Spectrophotometer HACH LANGE DR 2800. Chlorine decay in a particular volume of water was monitored during three days over the natural maximum water age of the system.

A plot of the relationship between bulk chlorine decay coefficients versus time was constructed. See **Figure 13**, from which K_b is extracted using the following relationship [21]:

$$K_b = \frac{L_n \left(\frac{C_1}{C_2} \right)}{(t_1 - t_2)} \quad (1)$$

Where:

K_b = Bulk reaction rate coefficient (day^{-1})

t_1 = Start time (hrs)

t_2 = End time (hrs)

C_1 =Chlorine concentration at t_1 (mg/l)

C_2 =Chlorine concentration at t_2 (mg/l)

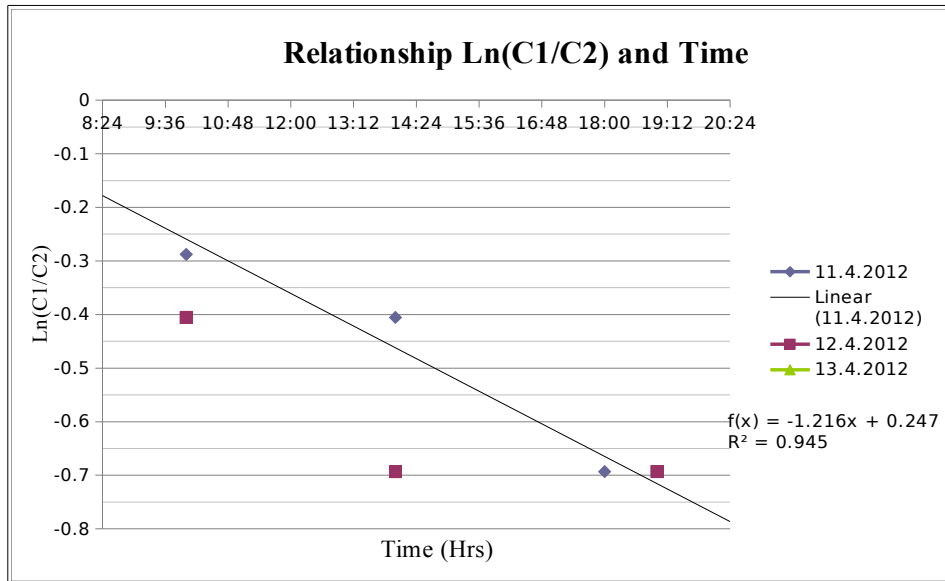


Figure 13: Relationship between bulk decay coefficient and time

K_w - Wall reaction decay coefficient

Wall reaction coefficient can be set as predicted concentration fit together observed concentration by trial and error. The dependency of K_w and the reaction order on pipe material and condition (i.e., age, encrustation, corrosion) make determining the coefficients difficult. Therefore, this model incorporates a calibrated K_w with initial estimates based upon pipe roughness coefficients, flow velocity, and pipe diameter. This approach is practiced widely in the industry because wall decay coefficients vary greatly due to pipe condition (material, roughness, corrosion, and biofilms) and can not be measured reasonably for large distribution systems. Determination of K_w coefficient is possible to be demonstrated in the model using equation (2).

Darcy – Weisbach:

$$K_w = \frac{-F}{\log\left(\frac{k}{d}\right)} \quad (2)$$

Where:

F = Correlation coefficients of wall reaction and pipe roughness

k = Roughness coefficient (Darcy – Weisbach) (mm)

d = Pipe diameter (mm)

After the bulk and wall decay coefficients were established, the model was run for 48 hours and the resulting free chlorine concentrations was compared to field data from three sites throughout the Kohoutovice pressure zone.

4.2.3 ANN model for simulation of residual chlorine concentration

For the given case, 18 available parameters were used to construct the raw database. It was choose 15 input parameters and 3 outputs parameters. The model inputs were selected from the available parameters.

- **Input parameters:**

Initial condition parameters (6 Inputs): Initial chlorine in Bosonohy tank, Free chlorine in Kohoutovice tank, pH in Kohoutovice tank, Temperature in Kohoutovice tank, Flow measured at the out flow of Kohoutovice tank and Turbidity in Kohoutovice tank

Local condition parameters (9 Inputs): pH node 1, Temperature node 1, Turbidity node 1, pH node 2, Temperature node 2, Turbidity node 2, pH node 3, Temperature node 3 and Turbidity node 3

- **Output Parameters:** Free Chlorine node 1, Free Chlorine node 2 and Free Chlorine node 3

Chlorine residual was proposed to be simulated in three nodes inside the pressure zone Kohoutovice. See **Figure 14**.

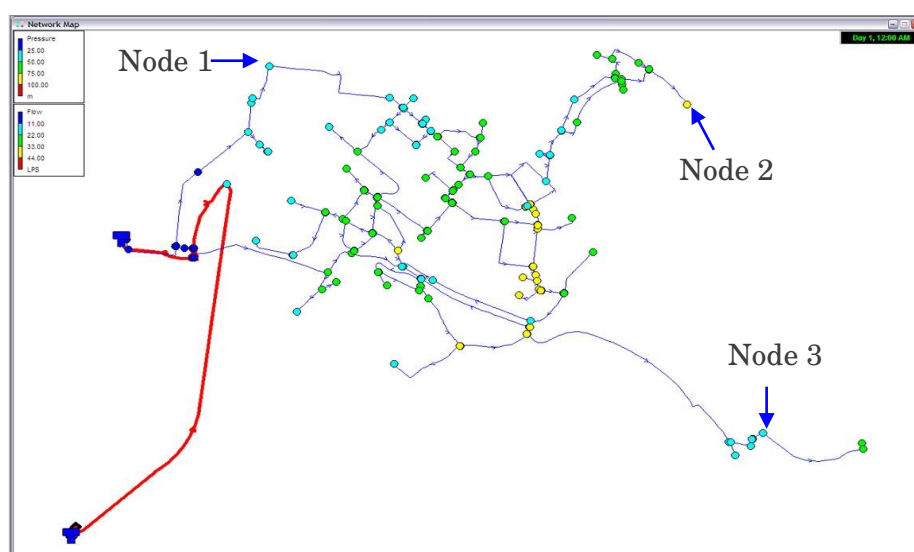


Figure 14: ANN simulation of residual chlorine in nodes 1, 2 a 3

- Node 1: Junction 290172 (Libusina tr.4)
- Node 2: Junction 291750 (Libusino udoli 66)
- Node 3: Junction 294526 (Nad Pisárkami 2)

Nine input parameters were selected and several subsets were constructed for each node by combining the inputs to obtain the best-fitting model and results. See **Figure 15** to check the architecture of the ANN used in modeling chlorine decay for the pressure zone in Kohoutovice.

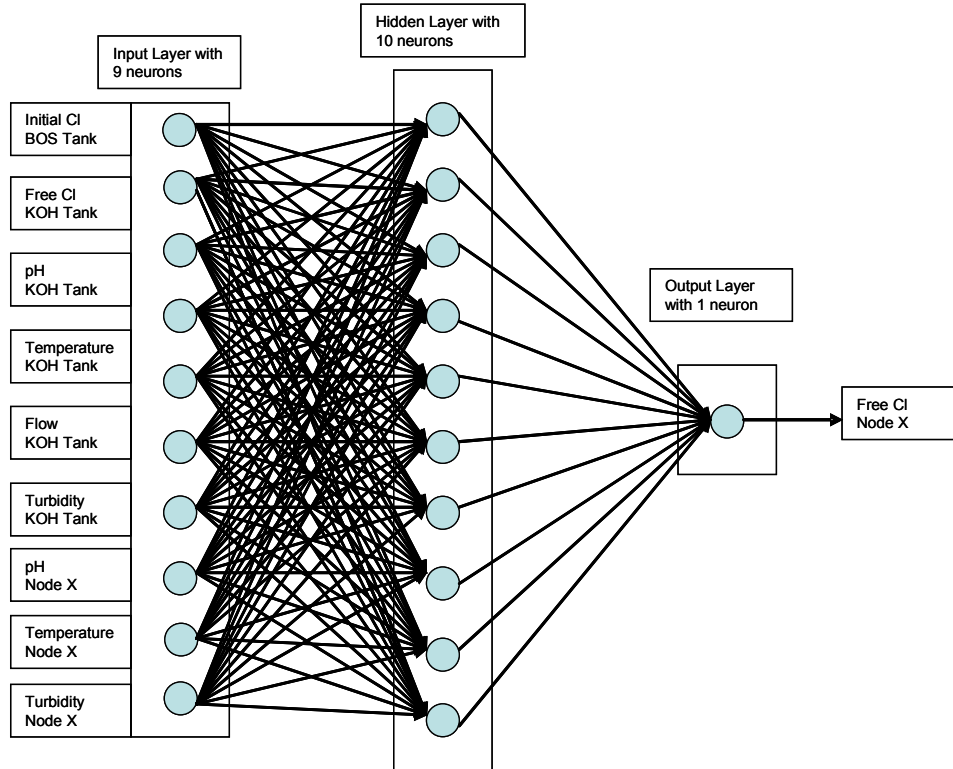


Figure 15: ANN Structure for residual chlorine modeling

4.2.4 Construction of the input database using the Monte-Carlo method

When creating ANN models, some data may be missing from the original database. Modelers usually replace the missing data with the average of the sample or simply delete or ignore the complete row, causing the loss of important data. The original database obtained from BVK for parameters in Kohoutovice Pressure zone 1.3.1 had 667 values for all the parameters including initial condition parameters, local parameters and output parameters. The ideal condition is to run the ANN model with all the historical data available. Few historical values were obtained from BVK in Kohoutovice pressure zone: 60 values for each initial condition parameter (input parameters), 37 values for each local parameter (input parameters) in Node 1 and 60 parameters for chlorine residual in Node 1 (Output parameter), 29 values for each local parameter in Node 2 including chlorine residual in Node 2, 5 values for each local parameter in Node 3 including chlorine residual for Node 3.

As the maximum given parameter was 60 it was needed to complete the database for 18 parameters to a number of 1080 values. From the measured parameters there were available only 667 values. Monte Carlo simulation was performed to fulfill the 413 missing values as it provide flexibility, manage the uncertainty and even provide more accurate results than simple descriptive statistics for example the average value. Note that this methodology has been implemented specifically for Kohoutovice pressure zone 1.3.1; although it can be used for other pressure zones.

An initial simulation was performed using the historical parameters measured in Kohoutovice pressure zone 1.3.1; Initial Chlorine, pH, Flow, Turbidity, Temperature and residual chlorine in three points inside the pressure zone. A normal distribution was followed to fit the distribution of the data measured and then a MC calculation was run. For the analysis of the measured data the computational software Statistica 10 from Statsoft was used. The software uses a function called *Distribution Fitting*, this option allows the verification whether the measure values follow a normal distribution and after the confirmation a calculation was run using the MC Method proposed in the same software package. All the parameters selected for the model were considered to follow a normal distribution. It was calculated a total of 3000 readings that generated random values for each factor or parameter affecting chlorine decay. **Table 6** shows basic statistics of the parameters, used to generate the random numbers in a given probability distribution for each input and output parameter.

Table 6: Statistics of the parameters, used for generation of random numbers

Parameter	Median	Standard Deviation	Min.	Max.
<i>Input Parameters:</i>				
pH Node 1	7.73	0.13	7.37	8.06
Temperature (°C) Node 1	18	2	11	23
Turbidity (ZF) Node 1	1.37	0.93	0.20	4.70
pH Node 2	7.60	0.10	7.31	7.91
Temperature (°C) Node 2	17	3	7	21
Turbidity (ZF) Node 2	0.97	0.61	0.20	4.28
pH Node 3	7.65	0.02	7.55	7.78
Temperature (°C) Node 3	11	1	8	14
Turbidity (ZF) Node 3	0.69	0.05	0.50	1.00
<i>Output Parameters:</i>				
Free chlorine (mg/l) Node 2	0.012	0.003	0.010	0.030
Free chlorine (mg/l) Node 3	0.014	0.002	0.010	0.030

The objective was to create a large database for each input and output parameter. The results obtained from MC method were again analyzed with descriptive statistics (average, standard deviation and confidence interval). This analysis was done again for each input and out parameter of the model.

4.2.5 Creation of the Neural Network Model

The type of ANN model used was Multilayer Perceptron (MLP). MLP utilizes the supervised learning technique “backpropagation” to train the network. For the model

calibration, the data set was treated using the following analysis: the data set was divided into three subsets, The first subset (50% of the database) was used to train the network (Learning phase), the second part (25% of the database) was used to test the ANN models in order to determine when to stop the training stage (Testing phase) and the last part was used to validate the model data not involved in the training process (Validation phase). The data set for each parameter were obtained from the data received by BVK and the missing parameters in a row were completed by the Monte Carlo calculations using the Software STATISTICA 10 from Statsoft. The Monte Carlo calculation uncertainty was kept below 2%. **Table 7** shows the statistics of the MC calculation for each parameter simulated.

Table 7: Statistics of the parameters calculated by MC method

Parameter	Mean	Standard Dev.	Max.	Min.	Conf. limits for means Interval 95%	
Input Parameter:						
pH Node 1	7.73	0.16	8.11	7.35	7.72	7.74
Temperature (°C) Node 1	18	3	26	9	18	18
Turbidity (NTU) Node 1	1.37	1.04	5.44	0.03	1.34	1.41
pH Node 2	7.60	0.15	7.97	7.29	7.59	7.60
Temperature (°C) Node 2	18	4	22	7	18	19
Turbidity (NTU) Node 2	0.95	0.85	15.66	0.06	0.92	0.98
pH Node 3	7.65	0.07	7.89	7.50	7.65	7.65
Temperature (°C) Node 3	11	2	17	4	11	11
Turbidity (NTU) Node 3	0.69	0.22	2.28	0.46	0.68	0.70
Output Parameter:						
Free Chlorine (mg/l) Node 2	0.0124	0.0063	0.0325	0.0011	0.0122	0.0127
Free Chlorine (mg/l) Node 3	0.0137	0.0089	0.0445	0.0163	0.0134	0.0141

It was then possible to generate ANN model combinations of different input parameters for the three nodes inside the pressure zone. The combination of these parameters for the three nodes inside the pressure zone 1.3.1, produced a data set of 900 input values. The data set composed by these input and output parameters are referred in this work as the training, testing and validation data set. This procedure allows the use of a suitable number of parameters for each model generated with acceptable uncertainties. As the objective of this work is the demonstration of the adequacy of using ANNs for the determination of the concentration of chlorine residual on different nodes inside Kohoutovice pressure zone, combining results from a historical measurement and using MC calculation for 413 missing values. The missing value correspond to the following parameters:

- 23 values of each local parameter in Node 1: pH, temperature and turbidity in Node 1
- 31 values of each local parameter in Node 2: pH, temperature, turbidity and residual chlorine in Node 2
- 55 values of each local parameter in Node 3: pH, temperature, turbidity and residual chlorine in Node 3

Based on the performance from each model combination it was selected 7 model results of ANN type MLP (Three models for node 1, two models for node 2 and two models for node 3), which performed the best results for each subset of input and output values. **Table 8** shows each input parameter, organized by columns for each node. Models are numbered from 1 to 7 and each model uses different combination of inputs shown in dark color in the cell. The total number of input used for each model is referenced in the last row of the **Table 8**.

Table 8: Details of the seven input subsets (in grey color) selected for the ANN model

Parameter / Input subset #	Node 1			Node 2		Node 3	
	1	2	3	4	5	6	7
Initial Chlorine BOS Tank							
Free Chlorine KOH Tank							
pH KOH Tank							
Temperature KOH Tank							
Flow KOH Tank							
Turbidity KOH Tank							
pH Node 1							
Temperature Node 1							
Turbidity Node 1							
pH Node 2							
Temperature Node 2							
Turbidity Node 2							
pH Node 3							
Temperature Node 3							
Turbidity Node 3							
Total	540	300	300	540	360	540	360

The 50% (training) of the data set was used to determine the best configuration of the ANNs and the rest 25% (test) and 25% (Validation) of data set was used to confirm the subsets chosen. The neurons at the input layer represent each parameter that influence chlorine decay, and the neuron at the output layer represents the chlorine residual in each of the three nodes inside the Kohoutovice pressure zone. The chosen values for the number of neurons for the studied ANN in the layers were the default values proposed by the *ANS tool* in Statistica 10 Software. Three subsets for node 1 and two subsets for nodes 2 and 3 were studied. After some initial observation it could be noticed that the default learning rate led to a slow reduction in the mean squared error (MSE). The training, testing and validation results for

each of the ANN subset are given in **Table 9** (Models 1-7). By the virtue of the MLP architecture, the training set can be predicted to a high level of accuracy. MLP achieved a significantly lower error for the training, testing and validation sets proving that the ANN models are able to find nonlinear relationship between variables.

Table 9: Model Performance of each phase of the seven selected parameters

	Subset	Model Type	Training perf.	Testing pef.	Validation perf.
Node 1	1	MLP 9-8-1	0.9343	0.9154	0.9077
	2	MLP 5-5-1	0.8983	0.9083	0.9442
	3	MLP 5-5-1	0.9467	0.9072	0.9284
Node 2	4	MLP 9-8-1	0.9778	0.4072	0.7034
	5	MLP 6-11-1	0.9876	0.4219	0.6672
Node 3	6	MLP 9-9-1	0.9990	0.5338	0.5387
	7	MLP 6-10-1	0.9841	0.4026	0.7811

The data presented in **Table 9** shows that the first three networks had attained values for Training, Testing and validation performance close to 1. The networks with architectures listed as subset 1, 2 and 3 are those that had, moreover, attained the lowest values for the mean squared error. A combination of input subsets was chosen to get an overview of the influence of the parameters in the chlorine decay. The three ANNs architecture performed in a good accurate way for the three subsets proposed in Node 1. Values of the free chlorine at Node 1 were completed only by the data obtained from BVK but some of the inputs as pH, temperature and turbidity at Node1 were calculated using MC method. It is also shown in **Table 9** the performance of the rest of the ANN architectures. Subset 4 and 5 correspond to Node 2 where is shown a poor performance in the testing phase (0,4072 and 0,4219) even though the training and validation have better results. The subset 4 was created using all the parameters available that could directly influence chlorine decay as shown in **Table 8**. These parameters included Initial condition parameters such as Initial chlorine at BOS Tank, Free Chlorine, pH, temperature and turbidity at KOH Tank in addition the local parameters such as pH, temperature and turbidity at Node 2. To check which parameters affected the most chlorine at Node 2 it was decided to create subset 5 with more influential parameters including only six inputs the initial condition parameters plus local Temperature at Node 2. The results only increased a bit more in the training and testing while the validation stayed in 0,6672. The results for the subsets 6 and 7 at Node 3 were taken in the same way as 4 and 5 and as well performed in a poor way varying the values from 0,4026 to 0,9990. This

behavior must be directly related with the low data availability and it could indicate that different ANNs should be tested for different input subset. Nevertheless the performance values for the every network subset for the training data set are very close to the unity, showing that the simulated and predicted data are strongly correlated. Particularly using the *Custom Predictions* tool in STATISTICA 10 from statsoft it is possible to predict chlorine in Nodes 1, 2 and 3 with the known conditions of input parameters introducing the values to the ANN model proposed. The data obtained from the ANN can be compared with the results from the EPANET and used as calibration in the three nodes and at the same time providing more reliable results of the final model. **Figures 16, 17 and 18** illustrate the typical correlation between ANN prediction and EPANET model data observed at each selected Node in Kohoutovice pressure zone. The simulated free chlorine concentrations at these select Nodes are in good agreement with the data obtained from ANN models. The models provide practical values by reasonably predicting locations of low chlorine residual and help to establish programs for measurement campaign in the network and residual chlorine assessment including additional parameters such as temperature, turbidity, pH and flow that were not taken into account with the first order simulation by EPANET.

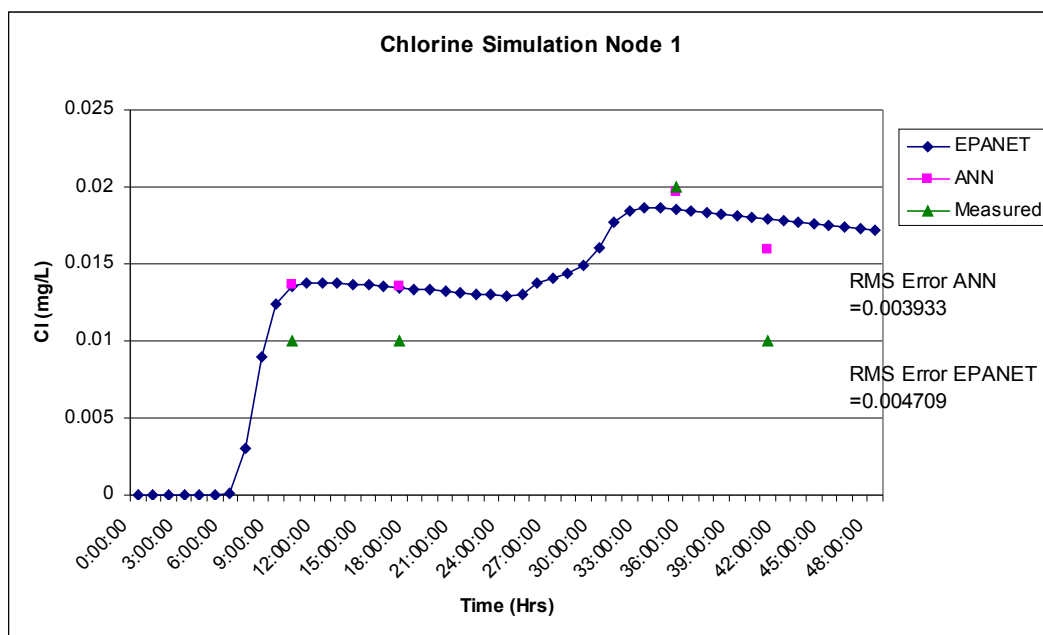


Figure 16: Simulation in node 1 – ANN and EPANET

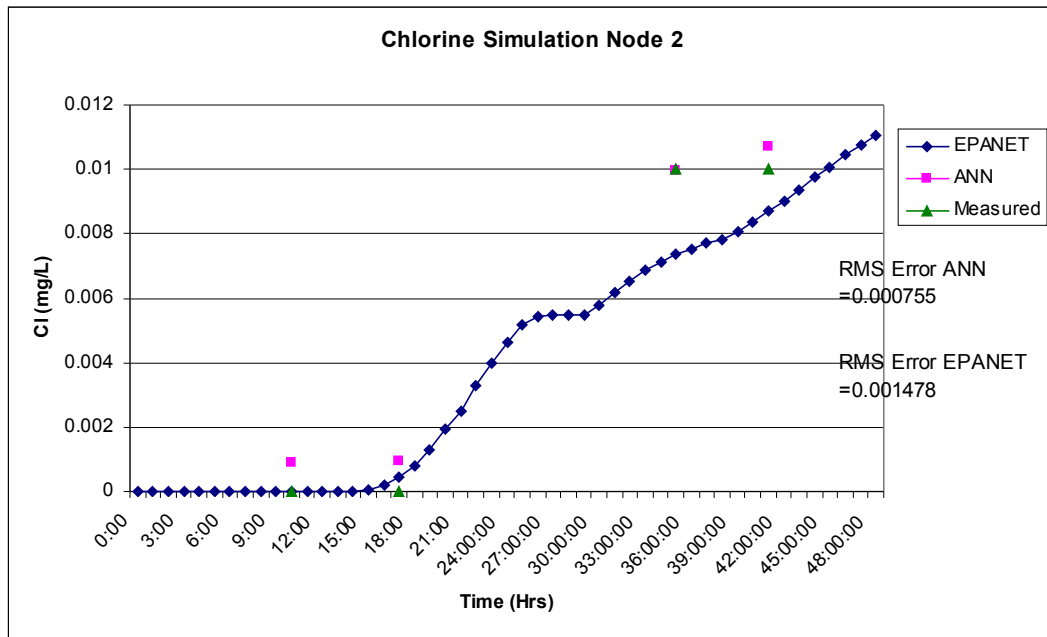


Figure 17: Simulation in node 2 – ANN and EPANET

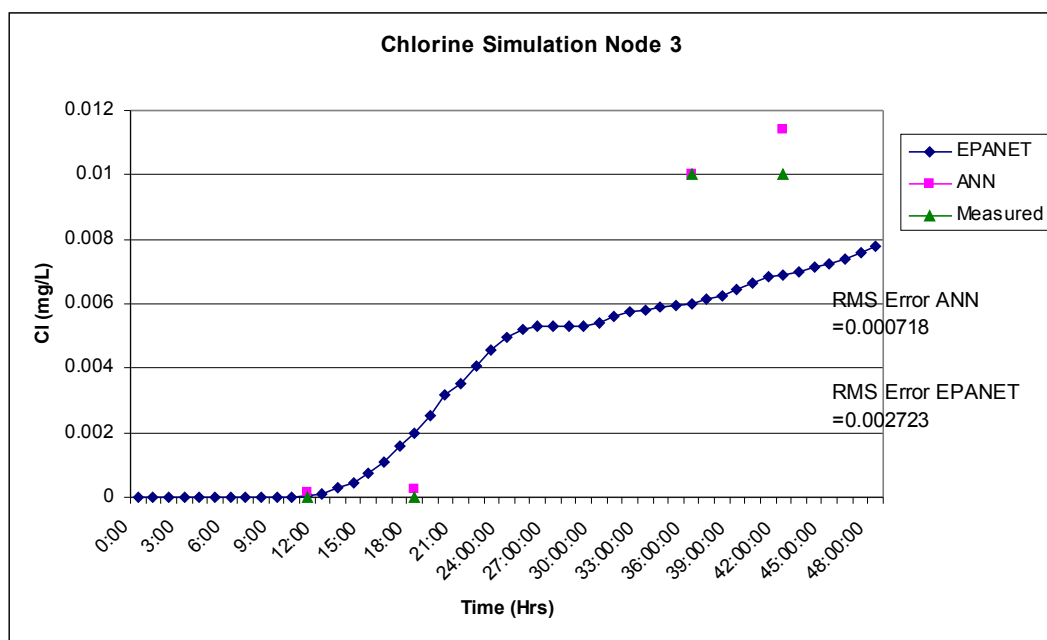


Figure 18: Simulation in node 3 – ANN and EPANET

4.2.6 Outcomes and results

Findings from the development and testing of ANN models and the obtained results allow to formulate the following conclusion and recommendations:

- The created ANN models using the tool ANS from the software Statistica 10 are able to predict the development of concentration of free chlorine in the pressure

zone Kohoutovice using the data measured and the data obtained by Monte-Carlo method.

- The ANN models created using the tool ANS from software Statistica 10 in this project can only be used for the prediction of chlorine decay in the pressure zone Kohoutovice.

- Monte-Carlo method can be used to generate the missing values of the original database of measured data. This new approach proposed in this project can be used thanks to the flexibility and uncertainty managed to achieve more accurate results than what could be achieved by using descriptive statistics (e.g. Average fulfillment)

- The key model input parameters are: Initial chlorine concentration, water temperature and flow rate. They have the greatest influence on the chlorine decay conditions at the selected nodes in Kohoutovice pressure zone.

5 CONCLUSIONS AND DISCUSSION

The first part of the thesis is dealing with the methodology proposed for the use of ANN for evaluation of historical data and chlorine decay prediction. The proposal was to estimate in chlorine decay in several points inside a network distribution system by using Artificial Neural Network techniques and some water quality parameters as inputs such as, Initial chlorine, pH, flow, Chlorine in several points inside the network, temperature, pipe roughness and pressure. Because of prior information is available about some or all the parameters, then the historical data was used to estimate the model parameters. In some cases, parameters are missing so it was also proposed to run a Monte Carlo simulation to fulfill the missing cases. Next distributions are used to quantify and evaluate these parameter value, summary statistics such as mean, median, standard deviation and confidence intervals are obtained for these parameters using Monte Carlo Method. Several readings have to be generated and tested using chi-squared test for each parameter to demonstrate that each of them can be simulated. An important part was also given to the hydraulic model as some values can be useful from it and a comparison of the modeling can be made. The second part of the thesis was more experimental as it was chosen two real network distribution systems as case studies to test the methodology proposed in the first part. The case studies were located in Czech Republic, first Našiměřice which is a small town and the distribution system was very suitable for the prediction of chlorine concentration in several nodes inside the WDS. Chlorine concentration was predicted in three nodes inside the WDS in Našiměřice using the following parameters as input: Initial chlorine, pH, flow, Chlorine in several points inside the network, temperature, pipe roughness and pressure. The finding shows, which parameters influenced the most the chlorine concentration and provide a high level of accuracy in the prediction compared to the Multi-Linear Regression method. The second case study was in a district of Brno City called Kohoutovice. The water for distribution comes from two difference sources and thee water is treated and

chlorinated in Čebín. Then the water is transported by means of pipes to Bosonohy tank and pumped to Kohoutovice tank, where finally is distributed to customers. This is also a small pressure zone suitable to the prediction of chlorine using ANN. The parameters used for forecasting chlorine concentration were: Initial chlorine, pH, turbidity, flow, Chlorine in several points inside the network and temperature. MC method was used as a technique to fulfill the missing values were not enough data were available. The results showed a good training performance and low Mean squared error. At the end a comparison was made with the EPANET water-quality modeling in the three nodes studied, which also helped identify some areas in which chlorine have a low or high concentration and remark the areas affected.

The main contributions of the thesis are perceived in:

- Demonstrating that Artificial Neural Networks (ANN) can be used to predict chlorine concentration in the distribution networks, case studies Našiměřice and Brno, Kohoutovice pressure zone, in Czech Republic.
- Creating a specific database for each distribution network studied (case studies), with historical data of parameters affecting chlorine decay obtained from the water utility, which includes; pH, temperature, turbidity, flow and initial chlorine.
- Using Monte-Carlo method in Brno, Kohoutovice case study, for simulation of some input and output parameters affecting chlorine decay when not enough data was available to run the ANN models.
- Showing that free chlorine concentration predicted with ANN technique, can be used in a physical based model (EPANET 2.0) to help with the calibration of the same in the remaining nodes within the WDS under consideration or for identification of areas affected by the maximum or minimum risk of significant changes of chlorine.

5.1 Fulfillment of Thesis Goals

All thesis objectives defined in *Chapter 1.1* have been achieved including the primary objective of the thesis which stated to develop an ANN model for chlorine concentration simulation.

[Goal 1]

Based on *Chapter 4*, case studies for Našiměřice and Brno, Kohoutovice pressure zone, a database for each water distribution system was created using the available parameters obtained from the water utility in each municipality. The database includes parameters such as: Initial chlorine, pH, turbidity, flow, Chlorine in several points inside the network and temperature.

Only for the case study Brno, Kohoutovice pressure zone, was implemented the Monte-Carlo method for simulation the values of parameters in which the cases

where not available or missing. Note that this databases can only be used for the specific distribution network although the methodology for the creation of the database is explained in *Chapter 3*.

[Goal 2]

Base on the tested case studies in *Chapter 4*, hydraulic models for each WDS was constructed and calibrated using the methodology technique proposed in Chapter 3. EPANET modeling network was used for both cases. For the calibration, a measurement campaign was done in each WDS including chlorine concentration measurements, flow and pressure instrument installation and using the available data provided by the water utility in each WDS.

[Goal 3]

Specifically in Brno, Kohoutovice pressure zone, case study the ANN model created was compared with EPANET water-quality model at the same time it was calibrated to obtain better results in terms of chlorine concentration prediction even in the rest of the nodes inside the WDS. More detailed information about the analysis can be seen in *Chapter 4* or for the general methodology used please refer to *Chapter 3*.

Bibliography

1. CLEMENT, J., J. POWELL and M. BRANDT. *Predictive models for water quality in distribution systems*. London: IWA Publishing, 2004, xxiii, 106 p. ISBN 18-433-9913-X.
2. BOWDEN, G., J. NIXON, G. DANDY, H. MAIER and M. HOLMES. Forecasting chlorine residuals in a water distribution system using a general regression neural network. *Mathematical and Computer Modelling*. 2006, vol. 44, no. 5-6, pp. 469-484. ISSN 08957177.
3. LINGIREDDY, S. and G. BRION. *Artificial neural networks in water supply engineering*. Reston, Va.: American Society of Civil Engineers, 2005, xvii, 173 p. ISBN 07-844-0765-7.
4. RAO, Z. and F. ALVARRUIZ. Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model. *Journal of Hydroinformatics*. vol. 9, no. 1, pp. 15-. ISSN 14647141.
5. ABDI, H., D. VALENTIN and B. EDELMAN. *Neural networks*. [Nachdr.]. Thousand Oaks, Calif [u.a.]: Sage, 1999. ISBN 978-076-1914-402.
6. POWELL, J., N. HALLAM, J. WEST, C. FORSTER and J. SIMMS. Factors which control bulk chlorine decay rates. *Water Research*. 2000, vol. 34, no. 1, pp. 117-126. ISSN 00431354.

7. HALLAM, N., B. J.R WEST, C.F FORSTER, J.C POWELL and I. SPENCER. The decay of chlorine associated with the pipe wall in water distribution systems. *Water Research*. 2002, vol. 36, no. 14, pp. 3479-3488. ISSN 00431354.
8. GIBBS, M.S., N. MORGAN, H.R. MAIER, G.C. DANDY, J.B. NIXON and M. HOLMES. Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods. *Mathematical and Computer Modelling*. 2006, vol. 44, no. 5-6, pp. 485-498. ISSN 08957177.
9. RODRIGUEZ, M. and J. SÉRODES. Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems. *Environmental Modelling*. 1998, vol. 14, no. 1, pp. 93-102. ISSN 13648152.
10. KAZANTZIS, D.M. Strengths and Weaknesses of EPANET for Modelling Water Quality in Distribution Systems. *Postgraduate Student Conference Distribution Systems Centre for Applied Modelling in Water Engineering*, School of Civil and Environmental Engineering, The University of Adelaide, 2004.
11. JAFAR, R., I. SHAHROUR and I. JURAN. Application of Artificial Neural Networks (ANN) to model the failure of urban water mains. *Mathematical and Computer Modelling*. 2010, vol. 51, no. 9-10, pp. 1170-1180. ISSN 08957177.
12. MAY, R., G. DANDY, H. MAIER and J. NIXON. Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. *Environmental Modelling*. 2008, vol. 23, no. 10-11, pp. 1289-1299. ISSN 13648152.
13. VASCONCELOS, J. and P. BOULOS. *Characterization and modeling of chlorine decay in distribution systems*. Denver, CO: American Water Works Association, c1996, xxxi, 402 p. ISBN 08-986-7870-6.
14. L. ROSSMAN, *Computer Models/Epanet*. Chapter 12, Larry W. Mays (ed.), *Water Distribution Systems Handbook*. McGraw-Hill Companies, New York, NY, , pp 12.1-12.23, (1999).
15. Castro, P., and Neves, M., *Chlorine decay in water distribution systems case study, Lousada Network*. *Electronic Journal of Environmental, Agricultural and Food Chemistry*, 2003, vol. 2, no. 2, pp. 261-266. ISSN: 1579-4377.
16. IZQUIERDO, J., R. PÉREZ and P.L. IGLESIAS. Mathematical models and methods in the water industry. *Mathematical and Computer Modelling*. 2004, vol. 39, no. 11-12, pp. 1353-1374. ISSN 08957177.

17. HERRERA, M., L. TORGO, J. IZQUIERDO and R. PÉREZ-GARCÍA. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*. 2010, vol. 387, no. 1-2, pp. 141-150. ISSN 00221694.
18. HUANG, J. and E. MCBEAN. Using Bayesian Statistics to Estimate Chlorine Wall Decay Coefficients for Water Supply System. *Journal of Water Resources Planning and Management*. 2008, vol. 134, no. 2, pp. 129-137. ISSN 0733-9496.
19. L. MA., X. KUNLUN and L. SUIQING, *Using Radial Basis Function Neural Networks to Calibrate Water Quality Model*, International Journal of Intelligent Systems and Technologies, 2008, vol. 3, no. 6, pp. 390-398, ISSN 1307-6884.
20. BAGGETT, C., G. LI and R. ROSARIO. From Start to Finish: Calibrating Pinellas County's 2,000-Mile Hydraulic Water Distribution System Model. *Florida Water Resources Journal*, 2008, vol. 60, no. 12, pp. 44-54.
21. CLARK, R., Water quality modeling in distribution systems. *Journal of Environmental Science and Health . Part A: Environmental Science and Engineering and Toxicology*. 1992, vol. 27, no. 5, pp. 1329-1366. ISSN 1077-1204.
22. SLAVÍČKOVÁ, K., Změny kvality vody při dopravě a jejich modelování v programu Epanet 2, In *proceedings Voda Zlín 2002*. Zlín: Vodovody a kanalizace Zlín, 2002, pp. 137-141. ISBN 80-238-8326-7.
23. VOJTOVÁ I., *Vírský oblastní vodovod a jeho funkce v Brněnské vodárenské soustavě (BVS)*. Brno, 2010. 47 p. Bakalářská práce na Ekonomicko-správní fakultě Masarykovy univerzity. Vedoucí diplomové práce doc. RNDr. Milan Viturka, Csc.
24. NACHÁZEL, K. and M. STARÝ. *Využití metod umělé inteligence ve vodním hospodářství*. Vyd. 1. Praha: Academia, 2004, 318 p. ISBN 80-200-0229-4.
25. MAYS, L., *Water distribution systems handbook*. New York: McGraw-Hill, c2000, 1 v. (various pagings). ISBN 00-713-4213-3.
26. Online: <http://www.statsoft.com/textbook>, Website of StatSoft, Inc. the STATISTICA Data Analysis Software and Services. Electronic Statistics Textbook.
27. HUA, F., J. WEST, R. BARKER and C. FORSTER. Modelling of chlorine decay in municipal water supplies. *Water Research*. 1999, vol. 33, no. 12, pp. 2735-2746. ISSN 00431354.
28. VASCONCELOS J., L. ROSSMAN, W. GRAYMAN, P. BOULOS and R. CLARK. Kinetics of chlorine decay. *Journal-American Water Works Association*. 1997, vol. 89, no. 7, pp. 54-65. ISSN 0003-150x.

29. JONKERGOUW, P. M., S.-T. KHU, Z. S. KAPELAN and D. A. SAVIC. Water Quality Model Calibration under Unknown Demands. *Journal of Water Resources Planning and Management*. 2008, vol. 134, no. 4, pp. 326-336. ISSN 0733-9496.
30. OSTFELD, A. Enhancing Water-Distribution System Security through Modeling. *Journal of Water Resources Planning and Management*. 2006, vol. 132, no. 4, pp. 209-210. ISSN 0733-9496.
31. RUČKA, J. and L. TUHOVČÁK. Využití neuronových sítí pro predikci poruchovosti vodovodních sítí. In *proceedings Juniorstav 2005: 7. odborná konference doktorského studia s mezinárodní účastí : sborník : Brno, 2.2.2005*. 1. vyd. Brno: Vysoké učení technické, 2005, 216 p. ISBN 80-214-2830-9.
32. TUHOVČÁK, L. and J. RUČKA. Hazard identification and risk analysis of water supply systems. *Strategic asset management of water supply and wastewater infrastructures: invited papers from the IWA leading Edge conference on Strategic Asset Management (LESAM), Lisbon, October 2007*. 1st pub. London: IWA, 2009, xii, pp. 287-298. ISBN 978-1-8433-9186-9.
33. TUHOVČÁK, L., J. RUČKA and M. SVOBODA. Riziková analýza systémů zásobování pitnou vodou. 4. *Vodohospodářská konference 2004 s mezinárodní účastí: sborník příspěvků*. Vyd. 1. Brno: Akademické nakladatelství CERM, 2004, 589 s. ISBN 80-720-4360-9.
34. CUESTA CORDOBA, G. ANN for Prediction of Chlorine in a Water Distribution System - Case of Study: Našiměřice, Czech Republic. In *Young Scientist 2011*. 1. Herľany - Slovakia: 3rd Phd. Student Conference of Civil Engineering-Technical University of Košice, 2011. ISBN: 978-80-553-0639-1.
35. CUESTA CORDOBA, G. Using of artificial neural network for evaluation and prediction of some drinking water quality parameters within a water distribution system. In *Juniorstav 2011*. 1. Brno, ČR: VUT v Brně, Fakulta stavební, 2011. pp. 243-253. ISBN: 978-80-214-4232-0.
36. CUESTA CORDOBA, G. Development of a neural model for forecasting residual chlorine decay - Našiměřice, Czech Republic case study. In *Juniorstav 2012*. 1. Brno, ČR: VUT v Brně, Fakulta stavební, 2012. pp. 313-320. ISBN: 978-80-214-4393-8.
37. FLOOD, I. and N. KARTAM. *Artificial neural networks for civil engineers: advanced features and applications*. Reston, VA: American Society of Civil Engineers, c1998, xv, 277 p. ISBN 07-844-0344-9.

CURRICULUM VITAE

Name Gustavo Andres Cuesta Cordoba
 Year of birth 1982
 Nationality Colombian
 Civil Status Single
 Telephone +420 774 958 045
 Email gustavoandres@seznam.cz
 Address Rousinovska 29
 Brno
 62700
 Czech Republic

EDUCATION

2008 – 2012 PhD Research Student at Brno University of Technology VUT (Vysoké učení technické v Brně), Faculty of Civil Engineering, Department of Water Management and Water Structures, Brno – Czech Republic
 1999 – 2004 Civil Engineer at University of Santander UDES (Universidad de Santander, UDES), University Campus “Lagos del cacique”, Bucaramanga – Santander, Colombia

EMPLOYMENT RECORD	
2012 – present	- SMDC Incident Coordinator, IBM Global Services, Brno – Czech Republic
2009 – 2011	- Customer Service Representative, IBM Global Services, Brno – Czech Republic
2005 – 2007	- Engineer Supervisor, Aguas del Sur de La Guajira, Fonseca – Colombia
2004 – 2005	- Resident Engineer, INCOL LTDA, Bucaramanga - Colombia

Experience in project assignments and education

VUT Projects	(Relevant and/or recent)
2012	Development of neural network model for forecasting chlorine concentrations in water distribution systems – Brno, Czech Republic. Case study.
2011	Development of a Neural Model for Forecasting Residual Chlorine Decay.
2010	Using of Artificial Neural Network for evaluation and prediction of some drinking water quality parameters within a Water Distribution System.

LANGUAGE PROFICIENCY

English Very Good
 Spanish Mother tongue
 Czech Basic

PUBLICATIONS

CUESTA CORDOBA, G. Development of a neural model for forecasting residual chlorine decay - Našiměřice, Czech Republic case study. In *JUNIORSTAV 2012*. 1. Brno, ČR: VUT v Brně, Fakulta stavební, 2012. s. 313-320. ISBN: 978-80-214-4393- 8.

CUESTA CORDOBA, G. Using of artificial neural network for evaluation and prediction of some drinking water quality parameters within a water distribution system. In *JUNIORSTAV 2011*. 1. Brno, ČR: VUT v Brně, Fakulta stavební, 2011. s. 243-253. ISBN: 978-80-214-4232- 0.

CUESTA CORDOBA, G. ANN for Prediction of Chlorine in a Water Distribution System - Case of Study: Našiměřice, Czech Republic. In *Young Scientist 2011*. 1. Herľany - Slovakia: 3rd Phd. Student Conference of Civil Engineering - TECHNICAL UNIVERSITY OF KOŠICE, 2011. ISBN: 978-80-553-0639- 1.

ABSTRAKT

Vodárenský distribuční systém je tvořen sítí dílčích prvků a subsystémů, které slouží k dopravě vody od zdroje až k odběratelům. Voda musí být upravena v úpravně vody pro zajištění bezpečné pitné vody pro spotřebitele, neobsahující patogenní a jiné nežádoucí organismy. Důležitým aspektem pro dosažení nezávadné pitné vody a prevencí před šířením chorob přenášených vodou je její hygienické zabezpečení. Chlor je běžným nejpoužívanějším dezinfekčním prostředkem v konvenčních procesech úpravy vody. Jeho rozšířené použití je dáno nízkou cenou a jeho vysokou schopností ničení bakterií. Proto se zajišťují jeho zbytkové koncentrace ve vodárenských distribučních systémech, aby se zabránilo mikrobiologické kontaminaci. Zbytková koncentrace chloru je ovlivněna fenoménem známým jako úbytek chloru, což znamená, že chlor reaguje uvnitř systému a jeho koncentrace se tak snižuje. Chlor je měřen na výstupu z úpravny vody a také v několika daných bodech ve vodárenském distribučním systému určeném pro kontrolu kvality vody. Metody simulace a modelování pomáhají efektivním způsobem předvídat koncentraci chloru ve vodárenských distribučních systémech. Účelem předložené disertační práce je hodnotit koncentraci chloru v některých strategických bodech v rámci vodárenského distribučního systému pomocí historických naměřených údajů některých parametrů kvality vody, které ovlivňují úbytek chloru. Nedávné výzkumy kvality vody prokázaly možnosti použití nelineárního modelování pro predikci úbytku chloru. Úbytek chloru v potrubí je složitý jev, proto vyžaduje techniky, které mohou zajistit spolehlivé a efektivní zastoupení složitosti tohoto chování. Statistické modely založené na umělých neuronových sítích byly shledány vhodnými pro zkoumání a řešení problémů spojených s nelinearitou v predikci úbytku chloru a nabízí výhodu na rozdíl od konvenčních modelovacích technik. V tomto ohledu používá tato disertační práce specifickou aplikaci neuronové sítě k vyřešení problému předpovídání zbytkového chloru ve vodárenském distribučním systému na vybraných bodech uvnitř sítě. Hydraulické parametry a parametry kvality vody, jako jsou průtok, pH, teplota, zákal atd, budou použity k předpovídání koncentrace chloru v několika vybraných bodech ve vodárenském distribučním systému. Získané výsledky z předpovídání koncentrace chloru pak budou použity ke kontrole a porovnání simulace celého zkoumaného systému.