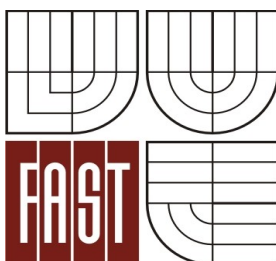


**VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ**  
BRNO UNIVERSITY OF TECHNOLOGY



**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**

# Acknowledgments

I would have never been able to finish my PhD Thesis without the advice of my supervisor, help from friends and support from family and friends. I would like to express my gratitude to my supervisor, Ladislav Tuhovčák (VUT Brno), for his excellent guidance, patience, and providing me with a good atmosphere for doing my research.

I would also like to thank the Institute of Municipal Water Management in Brno University of Technology, Civil Engineering Faculty, in general for providing the tools and helping me to create the framework for this research. Special thanks go to the people here. Also I would like to thank the partners, Brněnské vodárny a kanalizace a.s. (BVK) and Vodárenská Akciová společnost a.s. (VAS) for allowing and helping me to develop and test my research in a real water distribution network.

Last but not least, I thank my parents Aracelis Cordoba and Rito Cuesta, my sisters and friends. They were always supporting and encouraging me with their best wishes.

# Abstract

A water distribution system (WDS) is based in a network of interconnected hydraulic components to transport the water directly to the customers. Water must be treated in a Water Treatment Plant (WTP) to provide safe drinking water to consumers, free from pathogenic and other undesirable organisms. The disinfection is an important aspect in achieving safe drinking water and preventing the spread of waterborne diseases. Chlorine is the most commonly used disinfectant in conventional water treatment processes because of its low cost, its capacity to deactivate bacteria, and because it ensures residual concentrations in WDS to prevent microbiological contamination. Chlorine residual concentration is affected by a phenomenon known as chlorine decay, which means that chlorine reacts with other components along the system and its concentration decrease. Chlorine is measured at the output of the WTP and also in several considered points within the WDS to control the water quality in the system. Simulation and modeling methods help to predict in an effective way the chlorine concentration in the WDS. The purpose of the thesis is to assess chlorine concentration in some strategic points within the WDS by using the historical measured data of some water quality parameters that influence chlorine decay. Recent investigations of the water quality have shown the need of the use of non-linear modeling for chlorine decay prediction. Chlorine decay in a pipeline is a complex phenomenon so it requires techniques that can provide reliable and efficient representation of the complexity of this behavior. Statistical models based on Artificial Neural Networks (ANN) have been found appropriated for the investigation and solution of problems related with non-linearity in the chlorine decay prediction offering advantages over more conventional modeling techniques. In this sense, this thesis uses a specific neural network application to solve the problem of forecasting the residual chlorine in a water distribution system at selected points inside the network. Hydraulic and water quality parameters such as flow, pH, temperature, turbidity etc. will be used to forecast chlorine concentration in few selected points in the network distribution system and then the results obtained from the forecasting will be used to check and compare the simulation in the whole system in investigation.

# 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.

# List of Figures

Figure 1: Example of an Artificial Neural Network .....	14
Figure 2: Basic elements of an artificial neuron.....	15
Figure 3: Examples of activation functions.....	18
Figure 4: Modeling methodology.....	23
Figure 5: Program Structure.....	38
Figure 6: Scheme water supply system Našiměřice.....	41
Figure 7: Našiměřice, Czech Republic, Water supply system.....	42
Figure 8: Water tank (Hydroglobus) in Našiměřice .....	43
Figure 9: Instruments installation for measurement in Našiměřice.....	45
Figure 10: Predictions for Multiple Linear Regression Model 3.....	48
Figure 11: Training, Test and validation Predictions for MLP Model 1.....	49
Figure 12: Training, Test and validation Predictions for MLP Model 2.....	49
Figure 13: Water supply scheme.....	52
Figure 14: Hydraulic model run in EPANET 2.0.....	54
Figure 15: Time-Series plot Calibration for Tank level in Kohoutovice.....	56
Figure 16: Calibration Plot from EPANET for Head values at Tank Kohoutovice (Red line represents simulation and Green line the observed values).....	56
Figure 17: Calibration Plot from EPANET for Flow values at Link 327409 (Red line represents simulation and Green line the observed values).....	57
Figure 18: Calibration Plot in Excel for Flow values at Link 327409.....	57
Figure 19: Free Chlorine measurement using Instrument Spectrofotometer HACH LANGE DR 2800.....	59
Figure 20: Relationship between bulk decay coefficient and time.....	60
Figure 21: ANN simulation of residual chlorine in nodes 1, 2 and 3.....	63
Figure 22: ANN Structure for residual chlorine modeling.....	64
Figure 23: Simulation in node 1 – ANN and EPANET.....	71
Figure 24: Simulation in node 2 – ANN and EPANET.....	72
Figure 25: Simulation in node 3 – ANN and EPANET.....	72

A. Figure 2.1: Calibration report of the Correlation Plot for Head values at Tank Kohoutovice.....	87
A. Figure 2.2: Calibration report of the Correlation Plot for Flow values at Link 327409.....	87
A. Figure 3.1: Legend.....	89
A. Figure 3.2: Chlorine concentration in network nodes, $K_b = -0.14 \text{ day}^{-1}$ $K_w = -0.12 \text{ mm/day}$ at 10:00 (left) and at 14:00 (right).....	89
A. Figure 3.3: Chlorine concentration in network nodes, $K_b = -0.14 \text{ day}^{-1}$ $K_w = -0.12 \text{ mm/day}$ at 18:00 (left) and at 22:00 (right).....	89
A. Figure 3.4: Chlorine concentration in network nodes, $K_b = -0.14 \text{ day}^{-1}$ $K_w = -0.12 \text{ mm/day}$ at 26:00 (left) and at 30:00 (right).....	90

A. Figure 3.5: Chlorine concentration in network nodes, $K_b = -0.14 \text{ day}^{-1}$ $K_w = -0.12 \text{ mm/day}$ at 34:00 (left) and at 38:00 (right).....	90
A. Figure 3.6: Chlorine concentration in network nodes, $K_b = -0.14 \text{ day}^{-1}$ $K_w = -0.12 \text{ mm/day}$ at 42:00 (left) and at 46:00 (right).....	90
A. Figure 4.1: Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 1.....	94
A. Figure 4.2: Histogram Free Chlorine Node 1 (Target) for Model Subset 1...94	
A. Figure 4.3: Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 1.....	95
A. Figure 4.4: Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 2.....	95
A. Figure 4.5: Histogram Free Chlorine Node 1 (Target) for Model Subset 2...96	
A. Figure 4.6: Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 2.....	96
A. Figure 4.7: Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 3.....	97
A. Figure 4.8: Histogram Free Chlorine Node 1 (Target) for Model Subset 3...97	
A. Figure 4.9: Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 2.....	98
A. Figure 4.10: Free Chlorine Node 2 (Target) vs. Free Chlorine Node 2 (Output) For Model Subset 4.....	98
A. Figure 4.11: Histogram Free Chlorine Node 2 (Target) for Model Subset 4.....	99
A. Figure 4.12: Free Chlorine Node 2 (Target), Free Chlorine Node 2 (Output), Free Chlorine Node 2 (Square residual) for Model Subset 4.....	99
A. Figure 4.13: Free Chlorine Node 2 (Target) vs. Free Chlorine Node 2 (Output) For Model Subset 5.....	100
A. Figure 4.14: Histogram Free Chlorine Node 2 (Target) for Model Subset 5.....	100
A. Figure 4.15: Free Chlorine Node 2 (Target), Free Chlorine Node 2 (Output), Free Chlorine Node 2 (Square residual) for Model Subset 5.....	101
A. Figure 4.16: Free Chlorine Node 3 (Target) vs. Free Chlorine Node 3 (Output) For Model Subset 6.....	101
A. Figure 4.17: Histogram Free Chlorine Node 3 (Target) for Model Subset 6.....	102
A. Figure 4.18: Free Chlorine Node 3 (Target), Free Chlorine Node 3 (Output), Free Chlorine Node 3 (Square residual) for Model Subset 6.....	102
A. Figure 4.19: Free Chlorine Node 3 (Target) vs. Free Chlorine Node 3 (Output) For Model Subset 7.....	103
A. Figure 4.20: Histogram Free Chlorine Node 3 (Target) for Model Subset 7.....	103
A. Figure 4.21: Free Chlorine Node 3 (Target), Free Chlorine Node 3 (Output), Free Chlorine Node 3 (Square residual) for Model Subset 7.....	104

# List of Tables

Table 1: Overview of water demand year 2010.....	43
Table 2: Našiměřice Water supply system – Overview and demand in nodes..	44
Table 3: Details of input subset and variable used for the Models.....	46
Table 4: Descriptive statistics of the database (2007 -2010).....	46
Table 5: Matrix of correlation between all variables.....	47
Table 6: Model Performance .....	48
Table 7: Roughness Coefficient (mm) for the pressure zone studied.....	54
Table 8: Statistics of the parameters, used for generation of random numbers .....	67
Table 9: Statistics of the parameters calculated by MC method.....	68
Table 10: Details of the seven input subsets (in grey color) selected for the ANN model.....	69
Table 11: Model Performance of each phase of the seven selected parameters	70
A. Table 1.1: Raw Data and Data simulated using MC Method (MC simulated in Bold).....	86
A. Table 2.1: Calibration Statistics for Head at Tank Kohoutovice and Flow at Link 327409.....	88
A. Table 4.1: Summary of Network Results.....	91
A. Table 4.2: Predictions Statistics for each Model Subset.....	91
A. Table 4.3: Sensitivity Analysis for Model subsets 1, 2 and 3.....	91
A. Table 4.4: Sensitivity Analysis for Model subsets 4 and 5.....	92
A. Table 4.5: Sensitivity Analysis for Model subsets 6 and 7.....	92
A. Table 4.6: Predictions results for Nodes 1, 2 and 3 in each Model subset....	93
A. Table 5.1: Four-years database of parameters measured in Našiměřice....	105

# Principal Symbols and Abbreviations

ADALINE	Adaptive Linear Neuron
$ai(t-1)$	Activation function in previous state
$ai(t)$	Activation function at a time t
ANN	Artificial Neural Network
BNN	Bayesian Neural Networks
BOS	Bosonohy
BUT	Brno University of Technology
BVK	Brněnské vodárny a kanalizace a.s.
C(t)	Chlorine concentration at time t (mg/l)
C <sub>0</sub>	Initial chlorine concentration (mg/l)
Ca(ClO) <sub>2</sub>	Calcium hypochlorite
Cl <sub>2</sub>	Chlorine
DBP	Disinfection by Products
DN	Diametre Nominal/Nominal Diameter
GIS	Geographic Information Systems
GRNN	General Regression Neural Network
GUI	Graphical User Interface
H <sub>2</sub> O	Water
$hi$	Net Input
$kb$	Bulk water decay coefficient
$kf$	Mass-transfer coefficient
KOH	Kohoutovice
$kw$	Wall decay coefficient
MADALINE	Multiple Adaptive Linear Neuron
MASL	Meters Above Sea Level
MC	Monte-Carlo
MLP	Multi-Layer Perceptron



MLR	Multi-Linear Regression
MSE	Means Squared Error
NaClO	Sodium hypochlorite
NOM	Natural Organic Matter
NPS	Nominal Pipe Size
OCR	Optical Character Recognition
PNN	Probabilistic Neural Networks
PVC	Polyvinylchlorid
Q <sub>h</sub>	Hourly Flow rate
Q <sub>max</sub>	Maximum flow rate
RBF	Radial Bass Function
RMSE	Root Mean Squared Error
SCADA	Supervisor Control and Data Acquisition
SOFM	Self-organizing Feature Map
USEPA	US Environmental Protection Agency
UV	Ultra-violet treatment
VAS	Vodárenská Akciová společnost a.s.
VOV	Vir Regional Water Main system (Vírský oblastní vodovod)
WDS	Water Distribution System
W <sub>ij</sub>	ANN Weights
WTP	Water Treatment Plant
X <sub>j</sub>	ANN Inputs
$\lambda$	Coefficient of specific lethality - Deactivated constant
$\theta_j$	Threshold value

# Contents

1	Introduction.....	1
1.1.	Goals of the Thesis.....	3
1.2.	Thesis Contributions.....	4
1.3.	Thesis Outline.....	4
2	State of the Art.....	6
2.1.	A standard approach.....	7
2.1.1	Bulk and Wall Reactions.....	9
2.2.	Where are the problems?.....	10
2.3.	Artificial Neural Network for Water Quality in Water Distribution.....	11
2.4.	Artificial Neural Network: An Overview.....	12
2.4.1.	Structure of an ANN.....	14
2.4.2.	Components of ANN.....	15
2.4.3.	Neural Network Architectures.....	18
2.4.4	Use of Artificial Neural Network in Water Management.....	20
3	Theoretical Methodology for ANN Models Design.....	22
3.1.	Data preparation.....	24
3.1.1.	Hydraulic assessment of the WDS.....	24
3.2.	Selection of inputs and outputs of the ANN model.....	25
3.2.1.	Factors affecting chlorine decay in a Water Distribution System (WDS).....	26
3.3.	Construction of the input database using the Monte-Carlo method.....	31
3.3.1.	Simulation of factors affecting chlorine decay.....	33
3.3.2.	Test of Significance.....	33
3.4	Creation of the Neural Network Model.....	34
3.4.1	Division of models into subsets of parameters.....	36
3.4.2	Training, Testing and Validation.....	37
3.4.3	Performance Evaluation.....	37
4	Case Studies .....	39
4.1	ANN for Prediction of Chlorine in a WDS - Case Study: Našiměřice, Czech Republic.....	40
4.1.1	Current state of the drinking water supply system.....	41
4.1.2	Hydraulic assessment of water supply system Našiměřice.....	43
4.1.3	Construction of ANN Models.....	45

4.1.4 Outcomes of the research.....	47
4.1.5 Results and further work.....	50
4.2 Development of neural network model for forecasting chlorine concentrations in a pressure zone of a WDS - Case Study: Brno-Kohoutovice, Czech Republic.....	50
4.2.1 Hydraulic model for the selected pressure zone.....	51
4.2.2 Water quality model (chlorine decay) for the selected pressure zone...58	
4.2.3 ANN model for simulation of residual chlorine concentration.....	62
4.2.4 Construction of the input database using the Monte-Carlo method....	64
4.2.5 Creation of the Neural Network Model.....	67
4.2.6 Outcomes and results.....	73
5 Conclusions and discussion.....	74
5.1 Fulfillment of Thesis Goals.....	75
5.2 Main Findings of the Thesis.....	76
5.3 Known Limitations of the Research.....	78
5.4 Future Research Directions and Practical Use.....	79
Bibliography.....	80
APPENDIX 1: Database of 18 parameters used for ANN simulation and statistics in Brno-Kohoutovice, Case study.....	86
APPENDIX 2: Calibration Reports and Statistics.....	87
APPENDIX 3: Chlorine concentration in the network nodes.....	89
APPENDIX 4: ANN Models simulation and results.....	91
APPENDIX 5: Database of parameters used for ANN simulation, Našiměřice, case study.....	105

# 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 consumption is one of the most challenging task faced by the water utility industries, taking in consideration the components of the WDS, such as pipe materials, tanks, valves etc. and other risks related to water distribution. Basically water for human consumption is treated in a Water Treatment Plant (WTP) to make it safe and maintain the water quality as expressed before. Inside the WTP there is a combination of processes for drinking water treatment. Principal processes of a conventional WTP include; aeration, coagulation, sedimentation, filtration and disinfection. The purpose of treating the water is to provide a product that is micro-biologically and chemically safe for consumption. In all WDS applying disinfection, a disinfectant residual should be maintained throughout the distribution system at all times. Maintenance and monitoring of a residual disinfectant offer two benefits. First, a disinfectant residual will limit the growth of organisms within the system and may afford some protection against re-contamination by pathogenic or micro-organism, which can be originated in the biofilm formed inside the system, as well as in negative pressure areas (created by pipe cracks, fissures, etc.); second, the disappearance of the residual provides an immediate indication of the entry of oxidized matter into the system or of a malfunction of the treatment process.

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. Dosing too much chlorine has a number of negative effects, as it increases water treatment costs and has a deleterious effect on the taste and smell properties of the water. High chlorine levels are frequently related to consumer complaints and are commonly the largest source of customer concern for water utilities. Increased chlorine levels also raise the risk of forming disinfection by-products (DBP), which may be harmful to human health. In the other hand low levels of chlorine could cause diseases. Therefore, it is important to achieve a balance between the objectives of ensuring an adequate chlorine residual for microbiological quality and preventing high chlorine residuals that impact on the aesthetic qualities of the drinking water and may also pose health problems. Utilities can maintain the satisfaction and safety of their customers by strictly controlling residual chlorine throughout the WDS. In this sense, mathematical modeling of the decay is essential to correctly project new systems or make changes in existing ones [4]. The creation of predictive models for water quality in WDS can be divided into two categories, physically based model and stochastic model, the first is based on a deterministic idea where is necessary the use of hydraulic WDS model to predict the water quality parameters (e.g. EPANET 2.0), the second is known as a black box system, because of the use of historical data (inputs) to make a prediction (output) according to the relationship between all the data itself through mathematical and statistical concepts (e.g. regression, transfer functions, neural networks). Artificial Neural Networks (ANN) models are used frequently in various fields such as pattern recognition or data classification. The use of this technique in water management is increasing due the accurate prediction of parameters.

In this thesis, Artificial Neural Network (ANN) models are used to predict the free chlorine concentration of a given WDS. Findings of the study will be implemented in a special zone of water distribution systems, where the designed program is used for forecasting free chlorine residual at several water sections. Two case studies, which are real distribution networks are located in the Czech Republic. Those are small distribution networks with a simple architecture and the data of pipe age, material dimensions and others water quality parameters are available for at least 4 years, also the hydraulic behavior has been first run and calibrated for each WDS and then through the ANN models it was proved that free chlorine residual investigated can be predicted. The historical data that will be used as inputs for the models is

supposed to be collected in some nodes inside the WDS. There should be measured the chlorine and also all the variables or parameters that are likely to influence on the chlorine decay. The outputs will be characterized by a high degree of correspondence with the results of actual measurements recorded in the real water distribution systems. In addition to predicting chlorine, the program developed here may also be used to specify hydraulic parameters that must be maintained in the water system in order to obtain preset residual chlorine at specific nodes of the system. These may be obtained by selecting in EPANET an individual operating zone connected to a reservoir or tank in the water distribution system. The approach described in this thesis, will employ Artificial Neural Networks (ANN) for forecasting free chlorine residual on the basis of both historical data and simulation data, enabling an analysis of water distribution systems of any complexity.

### ***1.1. Goals of the Thesis***

The primary objective of the thesis is to develop an Artificial Neural Network (ANN) model that can simulate residual chlorine concentrations 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. An important part of the thesis is related to the development of a proposal method followed by required steps to reach the principal objective.

To predict residual chlorine decay in a WDS using ANN it was proposed to achieve the following main goals:

#### **[Goal 1]**

Construction of a database using the historical data of water quality parameters that can influence chlorine decay. ANN works with the analysis of historical data and it is required a large amount of data to be success in the application. Monte-Carlo method can be implemented to simulate some parameters affecting chlorine decay when not enough data will be available.

#### **[Goal 2]**

Methodology for construction and calibration of the hydraulic and water quality models: The hydraulic model will be run using EPANET 2.0 follow by ANN approach, which will be used to run the quality model.

### [Goal 3]

Analysis of results in the network: it is necessary to check the model efficiency and it will be achieved by taking our predictive output (from ANN) and using it as input of a previously calibrated water model (EPANET 2.0).

## 1.2. 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.

## 1.3. 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

---

This section gives a comprehensive overview of the state of art in predictive models of drinking water quality in Water Distribution Systems (WDS) and also outlines the general use of the Artificial Neural Networks (ANN) technique, which is a method that examines the data, learn from them and define the nonlinear relationship between the input and output variables.

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 this regard, may be evaluated, in the field of drinking water quality, suitable parameters such as: Disinfectant decay (free chlorine and chloramines) Disinfection by products (DBP), microbial water quality, Discoloration (production, accumulation, disturbance and transport of particles or sediments). It should, however, take into account other factors affecting water quality 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. A standard approach

The classical approach to model free chlorine residual is given by the simulation of the water quality reactions. The rate of most chemical reactions is proportional to the concentration reactants, that is to say, the rate decreases as the concentration of the reactants decreases and is typically approximated using the Chick (1908) equation; kinetic of organism disinfection by the first order as the following form:

$$\frac{dN}{dt} = -kN \quad (1)$$

Where:

$N$  = Number of microorganisms present at time  $t$

$t$  = Time (min)

$k$  = Rate constant which is characteristic of the type of disinfectant, microorganism or water quality

Integrating with respect to time, and replacing limits ( $N = N_0$  at  $t = 0$ ) yields:

$$\ln\left(\frac{N}{N_0}\right) = -kt \quad (2)$$

Or alternatively expressed:

$$N = N_0 e^{-kt} \quad (3)$$

Where:

$N_0$  = Number of microorganisms in early stage

Watson (1908) studied the coefficient of disinfection and proposed:

$$K = \lambda C^n \quad (4)$$

Where

$C$  = Disinfectant concentration (mg/l)

$\lambda$  = Coefficient of specific lethality - Deactivated constant (l min/mg)

$n$  = Coefficient of dilution - Constant of other solutions

Combining the two equations (2) and (4) becomes the *Chick-Watson* model:

$$\ln\left(\frac{N}{N_0}\right) = -\lambda C^n t \quad (5)$$

The Chick-Watson model is well accepted to represent the disinfection kinetic in water systems. The rate expression was applied to water distribution systems by Vasconcelos (1995) [13]. The first-order kinetic model for the disappearance of residual chlorine due to reactions with materials in the aqueous phase at different residence times in the network was expressed as:

$$\frac{dC}{dt} = -kC \quad (6)$$

Where:

$C$  = Chlorine concentration (mg/l)

$k$  = First-order decay constant ( $\text{min}^{-1}$ )

$t$  = Time (min)

Integrating equation (6) gives:

$$C(t) = C_0 e^{-kt} \quad (7)$$

Where:

$C(t)$  = Chlorine concentration at time  $t$  (mg/l)

$C_0$  = Initial chlorine concentration (mg/l)

$t$  = Residence time in the pipe (min)

The second-order rate process of a single compound was expressed as:

$$\frac{dC}{dt} = -kC^2 \quad (8)$$

or

$$C = \frac{C_0}{(1 + C_0 kt)} \quad (9)$$

Chlorine decay constant  $k$  is site specific and must be verified by field measurements. The  $k$  value may vary with water quality, water temperature, flow velocity, pipe material and area of contact with the pipe [13]. Many chemical reactions are complex, involving more than one compound; this means that the overall reaction kinetics will be complex and difficult to define.

### 2.1.1 Bulk and Wall Reactions

Rossman (1994) [14] developed a mass-transfer based model of chlorine decay in pipe networks that applies for turbulent and laminar conditions. The model considers that first-order reactions of chlorine occur both in the bulk flow and at the pipe wall. The model contains two set of rate coefficients that must be estimated. One represents a rate constant for the reaction in the bulk flow, and the other is a pipe wall reaction rate constant. The overall reaction coefficient can be defined as the sum of a bulk and wall coefficient.

$$k = k_b + k_w \quad (10)$$

Where:

$k_b$  = Bulk water decay coefficient

$k_w$  = Wall decay coefficient

In this model the rate of radial mass-transfer is defined by the coefficients  $k_f$ , which is itself a function of turbulence and molecular diffusivity.

$$k_w = \frac{4k_b k_f}{d(k_b + k_f)} \quad (11)$$

Where:

$k_f$  = mass-transfer coefficient

$d$  = Pipe diameter (mm)

US Environmental Protection Agency (USEPA) implemented the mass-transfer model within a network analysis software called EPANET. EPANET

is a network model that attempts to define the hydraulic behavior of distribution networks by mathematical equations to be solved by the computer. It incorporates a graphical user interface (GUI), giving the scheme of the network which also displays output data. The GUI can be used to edit the properties of the nodes, links or any components of the network. EPANET provides water quality modeling capabilities such as: non-reactive tracer material movement through the network, reactive material as it grows (e.g. a disinfection by-product) or decay (e.g., chlorine residual), age of water, reactions in bulk and pipe walls etc. This model has been widely used to predict free chlorine decay.

### ***2.2. Where are the problems?***

The standard approach described above suffers from several problems which are dealt with in this thesis:

- The use of first order kinetic for modeling chlorine decay is the most frequent model used in water distribution, because of its analytical solution. The limitation of this model technique is that it uses only one or two parameters (chlorine concentration and time) to be determined, ignoring the role of other important influencing factors on chlorine decay as the role of temperature, flow, age of pipes, and other chemical compounds.
- In  $n^{th}$  order models have been proved to have more accurate results modeling chlorine decay as it uses other parameters such as natural organic matter (NOM) and chlorine concentration but still others physico-chemical parameters are still missing in the prediction.
- Modeling bulk and wall decay separately requires additional sampling to estimate bulk decay coefficient for each water source and in addition it is hard to estimate wall decay as it must be measure on each site and it is inappropriate to use literature values as it depends on the site specification, pipe age and material.
- Many modelers consider the influence of pipe wall but they do not take into account the variation of pipe properties as pipe material, condition and diameter. Also variation in flow, velocity and water age can affect the chlorine decay.
- Statistical techniques such as Multi-Linear Regression (MLR) have been used to predict chlorine in water distribution using the historical data but MLR only takes into account a linear relationship between the variables. Chlorine decay is a phenomenon not yet completely understood and the use of non-linearity to find a relationship between the variables can lead to have better results in model performance.

### ***2.3. Artificial Neural Network for Water Quality in Water Distribution***

The power of ANNs in learning and generalizing complex phenomena lead to successful applications in several areas including science, engineering, technology and business. Some of the popular applications include, speech/handwriting, recognition, optical character recognition (OCR), automated classification of spam/no-spam email messages, prediction of stock prices, credit scoring, time series forecasting of stream flows, forecasting number of sun-spots, the list goes on. There are score of commercial and public domain/shareware general purpose ANN software packages available in the market [3].

Several successful applications in environmental engineering have been reported in literature as well. Maier and Dandy (2000) provide an extensive review of literature on the use of ANNs in water resources modeling and water quality modeling in terms of chlorine decay; they used the general regression neural network architecture to forecast chlorine residuals in Adelaide Australia. Also used data driven methods as Multi-Layer Perceptron to investigate the relationships between chlorine decay and water distribution parameters in the same case of study Adelaide Australia, the results in both researches ANNs architecture were found useful tools for forecasting chlorine residual and suitable for estimating disinfectant concentrations in WDS respectively. The same authors presented in 2006 chlorine decay forecasting using general regression neural network in a town in the south of Australia called Myponga, which distribute from the water treatment plant drinking water to a population of 45000 approximately. This study is based in the comparison between two used statistical models; Multiple Linear Regression (MLR) and Artificial Neural Networks (ANN). They describe exactly the way the information was taken, through experiments, to get the inputs data making tests in some real points in the trunk main (conduction line pipe) to get the parameters or variables (Chlorine in the filtered water tank outlet, chlorine in Cactus Canyon, chlorine in Aldinga, flow filtered water tank out let, flow in Sellicks hill pump station off-take, Temperature in Cactus Canyon, turbidity in WTP, and pH in WTP). In this test were used 8 variables with 384 inputs, but through the ANN architecture called self-organizing map, they reduced the inputs until 140, taking the non-significant inputs out and then again they used ANN General Regression Neural Network (GRNN) an architecture developed by D. Specht [5], to take the most statistical significant inputs for the experiment. The modeling by MLR was obtained using the R statistical package. The main challenge in this study was to obtain the forecasting horizon length, and it reached 72 hrs in advance with ANN and 24 hrs in advance with MLR.

Rodriguez and Serodes (1999) [9] used empirical linear and non-linear modeling of residual chlorine in urban drinking water system, case of study in the WDS of Sainte-Foy and the main water pipeline in the city of Quebec (Canada), the non-linear model used was an ANN and they referenced because of its recognized capacity for establishing complex relationships between input and output sets of data, conclusions for this research demonstrated that there exists an interesting potential for empirical-base modeling (linear and non-linear) in identifying the patterns of evolution of residual chlorine in drinking water systems.

In Czech Republic several researches have been implemented 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 Nacházel (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.

### ***2.4. Artificial Neural Network: An Overview***

Artificial Neural Network (ANN) is a simulation of the real nervous system, in other words, is a mathematical model based in biological neural networks [10]. It is a system that contains a collection of units (neurons) communicating with each other in a network that works to produce an output stimulus. ANNs are inspired by the activity of human brain. The key is the creation of neural networks (The system structure), which is composed of a large number of highly interconnected basic units (neurons) in layers that work together to solve specific problems. ANNs can be configured for specific applications, such as pattern recognition or data classification using learning process. This learning process as well as the biological system provides adjustments of developed models.

The use of Artificial intelligence technologies, specifically Artificial Neural Networks (ANN), is increasing in the drinking water treatment industry as they allow for the development of robust nonlinear models of complex units processes, improving the drinking water quality, and at the same time, reducing operating costs with the advanced process control. Despite widespread use of ANN technology in several areas of civil engineering most water supply engineering professionals appear to be highly skeptical of this

powerful technology. However, a few groups of researches who employed this technology for water supply engineering applications reported findings that were beyond the reach of traditional statistical/mathematical modeling tools [3]. ANNs are capable of modeling data whose functional relationships are not known in advance. By choosing appropriate architecture and activations functions, neural networks can be trained to capture knowledge from the data available within acceptable performance level [4]. Although an ANN captures and stores the knowledge as a well defined functional relationship, the relationship remains transparent to most users. For this reason many modelers treat ANNs as black-box models.

Artificial neural networks emerged after the introduction of simplified neurons by McCulloch and Pitts in 1943 (McCulloch & Pitts, 1943). These neurons were presented as models of biological neurons and as conceptual components for circuits that could perform computational tasks. The basic model of the neuron is founded upon the functionality of a biological neuron. "Neurons are the basic signaling units of the nervous system" and "each neuron is a discrete cell whose several processes arise from its cell body".

In engineering, neural networks serve two important functions: as pattern classifiers and as nonlinear adaptive filters. It will be provided a brief overview of the theory, learning rules, and applications of the most important neural network models. Definitions and style of computation an ANNs is an adaptive, most often nonlinear system that learns to perform a function (an input/output map) from data. Adaptive means that the system parameters are changed during operation, normally called the training phase. After the training phase the Artificial Neural Network parameters are fixed and the system is deployed to solve the problem at hand (the testing phase). The ANN is built with a systematic step-by-step procedure to optimize a performance criteria or to follow some implicit internal constraint, which is commonly referred to as the learning rule. The input/output training data are fundamental in neural network technology, because they convey the necessary information to *discover* the optimal operating point. The nonlinear nature of the neural network processing elements provides the system with lots of flexibility to achieve practically any desired input/output map, i.e., some Artificial Neural Networks are universal mappers.

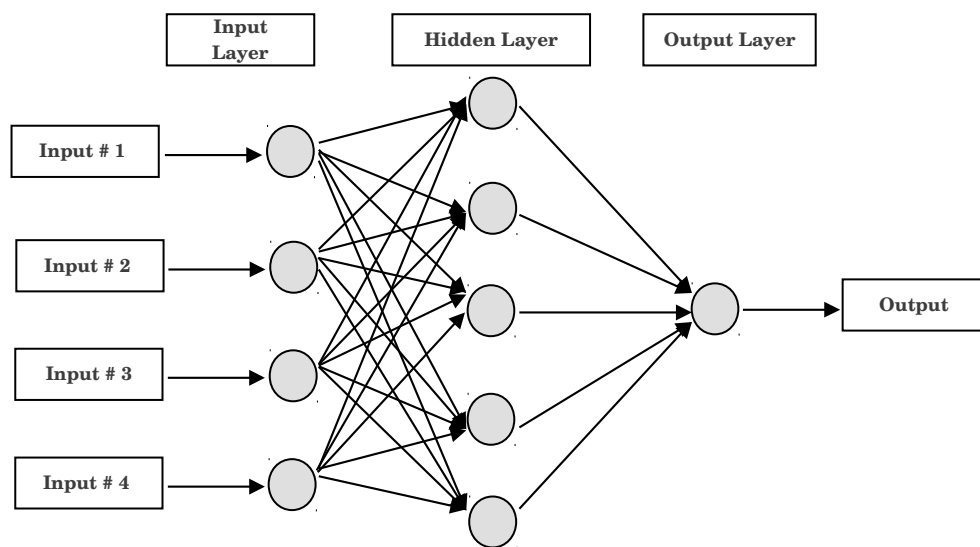
An input is presented to the neural network and a corresponding desired or target response set at the output (when this is the case the training is called supervised). An error is composed from the difference between the desired response and the system output. This error information is fed back to the system and adjusts the system parameters in a systematic fashion (the learning rule). The process is repeated until the performance is acceptable.



### 2.4.1. Structure of an ANN

The basic units of a biological neural system are neurons, which are grouped into sets, consisting of millions of them organized in layers and constitute a system with own functionality [10]. A set of these subsystems create a global system.

In **Figure 1** we see an ANN as a collection of parallel processors connected in the form of a guided or directed graph, organized such as the network structure itself leading us to consider it as a feature to keep in mind when creating an ANN. We can represent in a systematic way, each item (unit) that process the information of the network, as a node with connections between units represented by arrows, these arrows also indicate the direction in which information flows. Basically, neural networks are built from simple units, sometimes called neurons or cells by analogy with the real thing. These units are linked by a set of weighted connections. Learning is usually accomplished by modification of the connection weights. Each unit codes or corresponds to a feature or a characteristic of a pattern that we want to analyze or that we want to use as a predictor [5].



**Figure 1:** Example of an Artificial Neural Network

Neural networks usually organize their units into several layers. The first layer is called the input layer, the last one the output layer. The intermediate layers (if any) are called the hidden layers.

The information to be analyzed is fed to the neurons of the first layer and then propagated to the neurons of the second layer for further processing. The result of this processing is then propagated to the next layer and so on until the last layer. Each unit receives some information from other units (or from the external world through some devices) and processes this information, which will be converted into the output of the unit [5].

### 2.4.2. Components of ANN

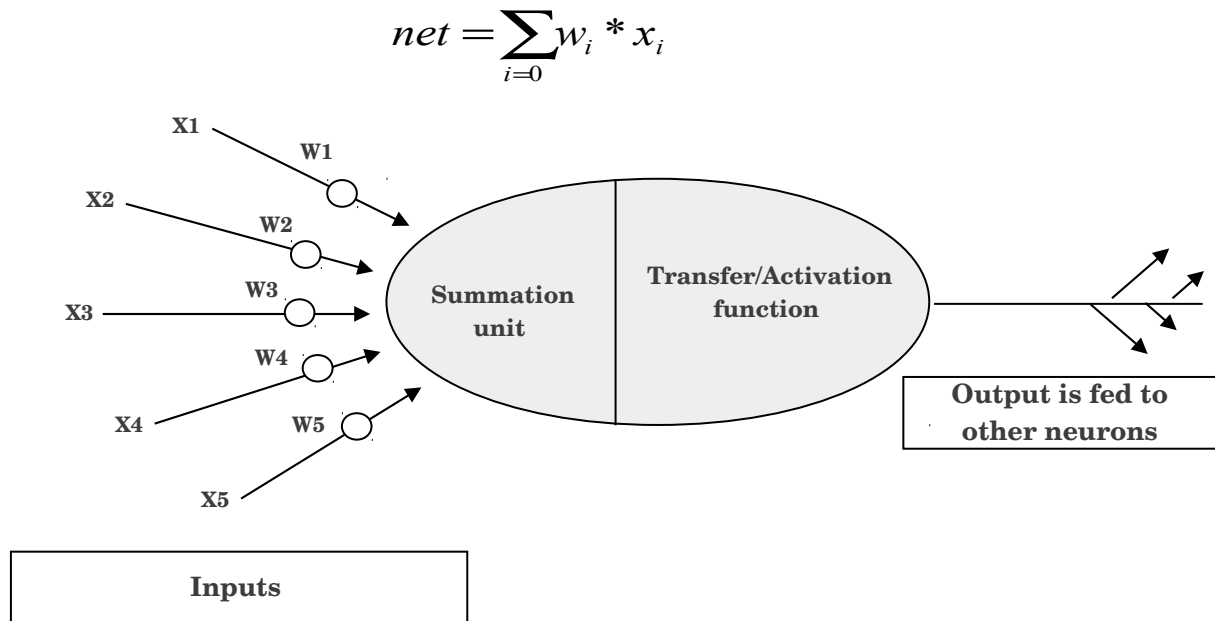
The first thing to understand is that the components of an ANN are an attempt to recreate the interacting process of the human brain. Next it is going to be describe most of the components contains in a neural network. These components are valid even if the neuron is used like input, output or hidden layer.

- **Inputs ( $X_j$ )**

As shown in **Figure 2**,  $X_j$  are inputs to the neuron and the appropriate selection of these variables or inputs in a group of potential measurement, to the system under investigation, is a vital step in model development. This is particularly important in data driven techniques, such as artificial neural networks and fuzzy systems, as the performance of the final model is heavily dependent on the input variables used to develop the model

For example in the treatment of drinking water, the application of ANN has demonstrated great effectiveness and efficiency. This is because in such networks can be calculated very quickly the hydraulic of the water (or the quality), by using a few specific inputs. However, in the distribution of drinking water (over 1,000 pipes) and wide dynamic models (e.g. every 5 minutes for a week), the algorithm is extremely slow.

#### *A Single Neuron*



**Figure 2:** Basic elements of an artificial neuron

- **Weight ( $W_{ij}$ )**

Typically a neuron receives many simultaneous and multiple inputs. Each input has its own relative weight which gives the importance of the input within the activation function of the neuron. These weights do the same role performed by the biological neurons in synaptic phase. In both cases, some inputs are more important than others so they have more effect on the processing of the neuron combined to produce a neuronal answer.

The weights are coefficients that can be adapted within the network to determine the intensity of the input signal, received by the artificial neuron. They are the measure of the strength of an input connection. These forces can be modified in response to the training examples according to the specific topology or because of the training rules.

- **Summing part**

This rule provides, from the inputs and weights the potential post-synaptic value  $h_i$  of the neuron:

$$h_i = \sum_i w_{ij} * x_i \quad (12)$$

Where:

$h_i$  = Net input

The most common function is the sum of all weights and inputs, by grouping the inputs and weights in two vectors  $(x_1, x_2, \dots, x_n)$  and  $(w_{1j}, w_{2j}, \dots, w_{nj})$  and then calculate this amount making the scalar product of two vectors.

The role of this summing part may be more complex than a simply sum of products. The inputs and weights can be combined in different ways before passing the value to the activation function. For example we can use as the summing part, the minimum, maximum, most, products, or different normalization algorithms. The specific algorithm for the propagation of neural inputs is determined by the choice of architecture.

- **Activation or transfer function**

The result of the summing part, in most cases is a weighted sum, which is transformed into the actual output of the neuron through an algorithmic process known as activation function. The purpose of utilizing an activation function is to allow the summation output to vary with respect to time. Activation functions currently are pretty much confined to research. Most of

the current network implementations use an *identity* activation function, which is equivalent to not having one.

$$a_i(t) = f_i(a_i(t-1), h_i) \quad (13)$$

$a_i(t)$  = Activation function at a time t

$a_i(t-1)$  = Activation function in previous state

The other parameters are defined above

In equation **13** the activation function depends on the post synaptic potential  $h_i$  And its previous state of activation. However, in many models of ANN is considered that the current state of the neuron does not depend on its previous state  $a_i(t-1)$ , but only the current as shown in equation **14**.

$$a_i(t) = f_i(h_i) \quad (14)$$

In the Activation function, the value of the output combination can be compared with a threshold value for determining the output of the neuron. If the sum is greater than the threshold value it generates a neuron signal. If the sum is less than the threshold, no signal is generated.

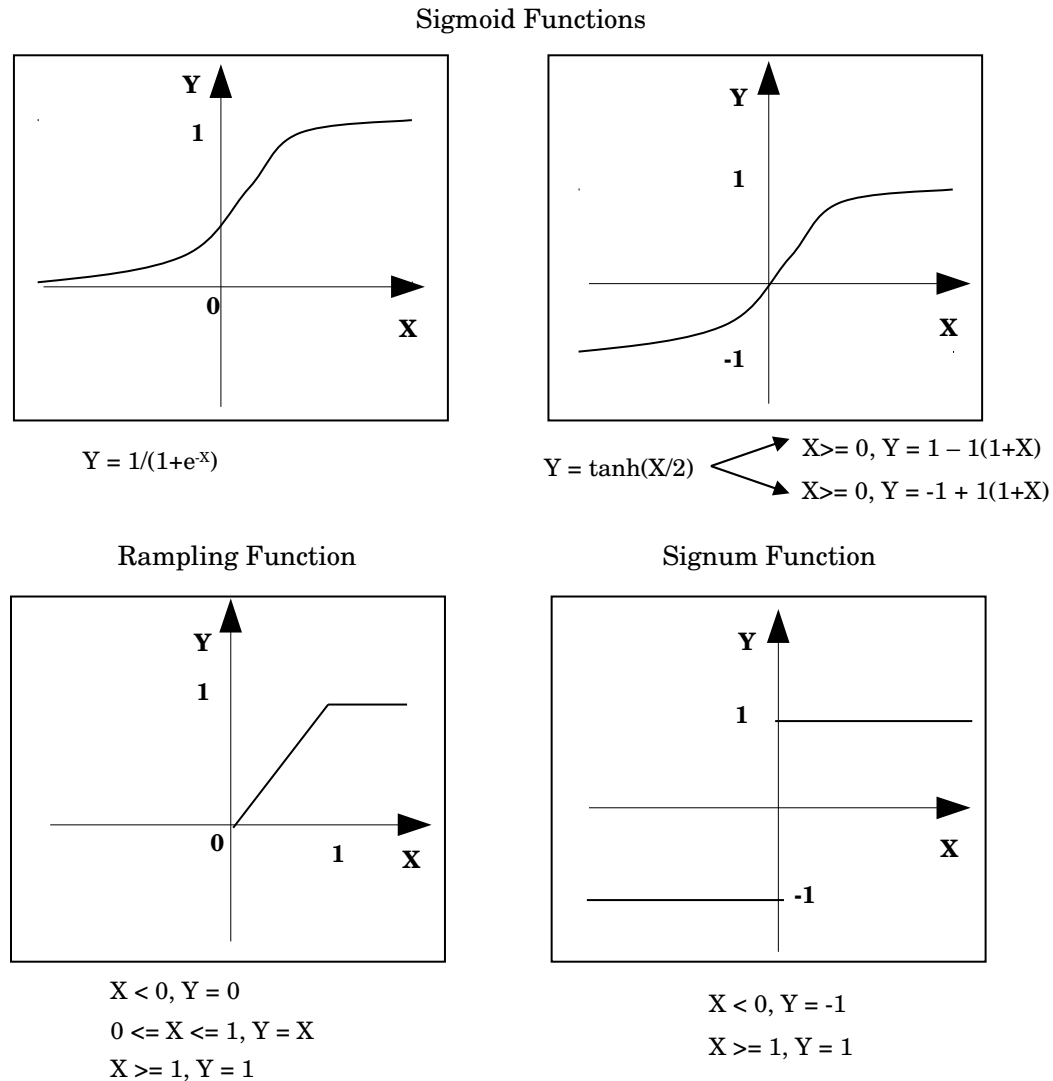
Usually the threshold value, or transfer function value is typically nonlinear. The use of linear functions is limited since the value of the output is proportional to the input; in fact this was one of the problems in the early models of artificial neural networks in Perceptrons.

The activation function could be something as simple as it only depends on whether the result of the combination function is positive or negative. Some transfer or activation function can be seen in **Figure 3**.

From the functions presented in **Figure 3**, the sigmoid function stands out. From a mathematical point of view, the usefulness of these functions is that the function and its derivative are continuous. These functions work quite well and are usually elected. There are other activation functions that are specific to some architectures.

Before applying the activation function, we can add some noise to the inputs. The source and amount of this noise is determined by the training of a

particular network. This noise is commonly known as temperature of the neuron. In fact by adding different noise levels to the result of the combination or summing, leads to create a model more similar to the brain. The use of the noise by temperature is still under investigation and is not usually applied in the praxis.



**Figure 3:** Examples of activation functions

### 2.4.3. Neural Network Architectures

Architecture is called to the topology, structure or connection pattern of a neural network. In an ANN nodes are connected by synapses, this structure of synaptic connections determines the behavior of the network. In general, neurons are usually grouped into structural units that are called layers and finally, the set of one or more layers is the neural network. There are three types of layers:

- Input: an input layer or sensory layer consists of neurons that receive data or signals from the environment.
- Hidden: is the one that has no direct connection with the contour, i.e. is not directly connected to body sensors nor effectors.
- Output: is the layer in which the neurons provide the response of the neural network.

The connections between neurons can be excitatory or inhibitory: a synaptic weight defines a negative inhibitory connection, while a positive determines an excitatory connection.

Intra-layer connections, also called side connections, take place between neurons in one single layer, while the inter-layer connections occur between neurons in different layers. There are also feedback connections that have an opposite way input - output. In some cases can exist feedback even in a neuron itself. Based on these concepts, we can set different neural architectures:

- Single- Layer Network: are those composed by just one layer of neurons.
- Multi-Layer Networks: (layered networks) are those whose neurons are organized in several layers.

In response to the data flow in a neural network, we can talk about:

- Feed-forward Networks: information circulates in one direction from the input neurons to the output.
- Feedback Networks: the information circulates between the layers in any direction.

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, as shown in **Figure 1**. 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. The strength of each connection, referred to as its connection

weight, can be adjusted. The connection weight from the  $i_{th}$  node to the  $j_{th}$  node is denoted by  $w_{ij}$  [8].

Input data are presented to the network through the input layer, the values of which are denoted by  $x_i$ . Data are passed from the input layer to the hidden layer. Each node in the hidden layer receives the weighted outputs ( $w_{ij}x_i$ ) of the nodes in the preceding layer. These outputs are then summed and added to a threshold value ( $\theta_j$ ) to produce the node input ( $I_j$ ) as shown in equation 15

$$I_j = \sum_i w_{ij}x_i + \theta_j \quad (15)$$

Where

$\theta_j$  = Threshold value

$I_j$  = Node Input

The node input ( $I_j$ ) is then passed through an activation function ( $f(I_j)$ ) to produce the node output,  $y_j$ . This node output is then used to compute the inputs for nodes in the following layer, until the final output is calculated [8].

#### **2.4.4 Use of Artificial Neural Network in Water Management**

An Artificial Neural Network (ANN) is nowadays recognized as a very promising tool for relating input data to output data. It is said that the possibilities of artificial neural networks are unlimited. The applications 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.

### **Advantages and disadvantages of ANN**

Scientific and engineering community has acquired already an extensive experience in developing and using data-driven techniques. Not all sectors of water industry, however, have used advantages of these methods. Applications of two mostly widely used particular types of data-driven models, namely artificial neural networks (ANN) and fuzzy logic-based models, to modeling in the water resources management field must be considered. The advantage of such integrated tool is that it connects the numerical modeling technique with the Artificial Intelligence technique to include the advantages of both.

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

Every method has its advantages and disadvantages and is well suited to some specific problems.

Generally it is presented the following list of advantages and disadvantages:

Advantages:

- A neural network can perform tasks that a linear program can not.
- When an element of the neural network fails, it can continue without any problem by their parallel nature.
- A neural network learns and does not need to be reprogrammed.
- It can be implemented in any application.
- It can be implemented without any problem.

Disadvantages:

- The neural network needs training to operate.
- Another aspect of the artificial neural networks is that there are different architectures, which consequently requires different types of algorithms
- Requires high processing time for large neural networks.

Despite to be an apparently complex system, a neural network is relatively simple.



# 3

## Theoretical Methodology for ANN Models Design

---

Some different topologies of feed forward artificial neural networks (ANNs) using the backpropagation learning algorithm can be studied to approach the behavior of chlorine decay for varying levels of chlorine residual in several nodes inside a water distribution system, in addition, some physic-chemical input parameters (e.g. pH, Temperature, turbidity and flow) can be assessed as they can affect chlorine decay.

The systematic modeling procedure proposed to be implemented in this thesis can be seen in **Figure 4**. The main steps for modeling chlorine residual in a WDS using ANN involve:

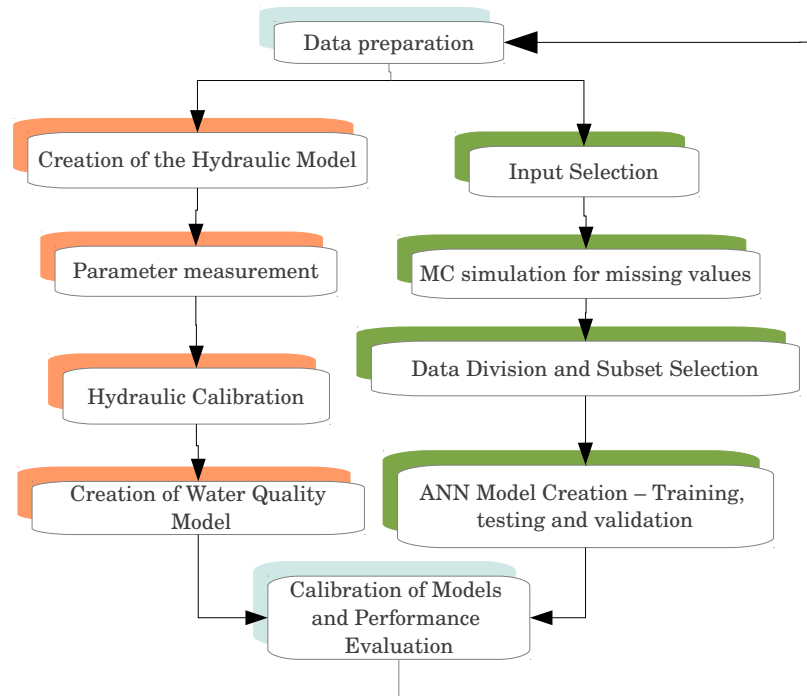
- Data preparation
- Input selection
- Monte Carlo simulation for missing values
- Data division and selection of subsets
- Model creation
- Model calibration
- Performance evaluation

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 for the creation of ANN models for chlorine decay prediction in a WDS is the creation of the hydraulic model. The hydraulic model must be also calibrated in the best possible way since the 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. 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).



**Figure 4:** Modeling methodology

The objective is to create a big database for each input and output parameter. The results obtained from MC method can be again analyzed with descriptive statistics (average, standard deviation and confidence interval). This analysis can be done again for each input and output parameter of the model. As an example for the procedure implemented it can be taken the initial chlorine data measured. It is supposed that the values of chlorine are variables that follow a normal distribution. The objective will be the creation of a large database for this parameter and next the data will be analyzed to obtain the best fitting value for the missing value in the original database, the results obtained from the random generator will then be analyzed using the function of mean, standard deviation and confident interval. With these results from the MC method and the data of the original database it must be determine the

expected chlorine in the network. The same will be done for each parameter estimated in this study, affecting chlorine decay.

### ***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 and they depend on the water source of the location. Depending on the utility the data resource is stored in databases applications such as Supervisor Control and Data Acquisition (SCADA) systems, metadata in excel or even in Geographic Information Systems (GIS) databases. Care must be taken in how the data was collected, the format given, the supplier of the data and it must be verified whether the information is current and accurate. Seasons also influence in the value data, so the approach can be taken in two ways. First divide the data for each season and run the study for different seasons as in each season, the variables such as temperature and pH have great changes or the second way, create a big database and run the analysis for all the data available. ANN will find the relationship between the changes, but a big amount of data is required to get better results.

An important part of the data collection also goes to preparation of the data for the creation of the hydraulic model. 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 created by the commercial company DHI 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.

#### ***3.1.1. Hydraulic assessment of the WDS***

The hydraulic assessment of the water distribution system will be run using the computational model EPANET 2.0 implemented in a special zone of the water distribution system. To calibrate the model it is needed to put into action a measuring campaign of the hydraulic parameters and also the water quality parameters in the water distribution system. It requires the installation of various devices in the distribution system for the measure of

some parameters and scheduling regular visits to check the level chlorine in the investigation zone.

An essential part in the hydraulic model is the identification of the intended use and this thesis focuses on water-quality in terms of chlorine concentration in the system. Water-quality and operational studies required an extended-period analysis, whereas some planning or design studies can be performed using a study state analysis [10]. In general water-quality analysis required a higher level of model calibration.

Several authors have describe the calibration of hydraulic network models in different way. e.g. S. Lingeriddy and Ormsbee [25] suggested a seven-step network calibration methodology based in two levels. Macro-level calibration of the model describe the general performance and global adjustments of each hydraulic parameter, for example, reported to use this type of calibration to experiment the impact of each parameter in the model. The second level is called micro-level calibration and it is a detailed description of the model performance and modifying the parameters to find a rational agreement between the observed values and the values provided by the hydraulic model. Normally the attempt for micro-level calibration is done by the traditional trial-and-error approach, but in recent years some researchers have proposed different algorithms such as Genetic Algorithms to adjust some of the model results and match with the measured parameters.

### ***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
- Availability and quality of the required data in the training set

Initially the model will be run with all possible input variables that had an impact in the output. Also it has to be run a linear regression to estimate the correlation of each input and then choose the possible inputs or generates an optimized ANN model consisting in 3 steps:

- Build the model with all the possible inputs

- Re-build the model with reduce inputs in different combinations
- Verify the effect of the combination and reduction and remove the insignificant

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. The objective of the model is the prediction of residual chlorine decay, so we need to specify the variables affecting this water quality parameter.

Recent studies have shown physic-chemistry parameters such as temperature, pH, turbidity, and natural organic matter (NOM) between other that affect the chlorine decay, also hydraulic parameters as flow, pressure and properties of the pipe line as pipe material, diameter and age of pipes should also take into account as an influence to the chlorine decay.

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

In most WDS the water is disinfected using chlorine and to ensure the water is adept for human consumption, it must be guaranteed that chlorine should stay in the pipes for a specific time until is consumed by the customer. Chlorine decay is a phenomenon in which chlorine disappears in the network due several reasons as chemical reactions with the pipe wall and other compounds (organic or inorganic) in the water. Chlorine also reacts differently depending on the temperature of the water and other physical, physic-chemical and hydraulic parameters. Some relevant parameters affecting chlorine decay are following:

Physical Parameters:

- Pipe roughness – Influenced by pipe age, pipe material and water quality.
- Diameter

Hydraulic Parameters:

- Flow
- Pressure
- Velocity

Physic-chemical Parameters:

- Chlorine added to water (initial chlorine) - At the beginning of the network or in the treatment plant
- pH
- Turbidity
- 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. The following sections explain in detail the information of each parameter that will be studied.

#### – Physical Parameters

The following parameters describe the physical properties of the distribution network, which become an influence of the hydraulic parameters:

#### **Pipe roughness**

The internal roughness of a pipe is an important factor when considering the friction losses of a fluid moving through the pipe. For each pipe material either a single pipe roughness value or a range of roughness values is normally provided by the manufacturer if the pipe is new, but in most cases depending of the age of pipe construction the pipe roughness varies. The roughness value, usually denoted as  $k$ , is used in the calculating the relative roughness of a pipe against the size of its diameter. For water-quality case study it has to be checked the historical data to compare the pipe age and material. The pipe roughness has to be a variable depending on the two respectively parameters.

#### **Diameter**

Commercial pipes come in many different materials and many different sizes. Water pipes are pipes or tubes, frequently made of polyvinyl chloride (PVC/uPVC), ductile iron, steel, cast iron, polypropylene, polyethylene, or copper, that carry pressurized and treated fresh water to buildings (as part of a municipal water system), as well as inside the building. In America the standard size for pipes used for high or low pressure and temperatures is called Nominal Pipe Size (NPS), The European designation equivalent to NPS is  $DN$  (diamètre nominal/nominal diameter), in which sizes are measured in millimeters. Although the diameter should not significantly influence the

distribution flow or head-loss it may be affected in a high level the pipe velocity, which in turn could influence the results of the water-quality analysis [10].

#### **– Hydraulic Parameters**

These parameters define the behavior of the distribution system in terms of quantity; an insight of each parameter is described below:

##### **Flow**

The first step in evaluating a water distribution system is to determine design flows the representative built-upon region or area of study. The fundamentals of calculating water system design flow are as follows:

- Determine the average daily demand
- Determine the maximum daily demand or estimate it from the average daily demand.
- Determine the maximum and minimum hourly demand from consumption records

The assessment of water-quality is always related with the determination of water flow. It is not possible to assess water quality without the evaluation of the water quantity.

##### **Pressure**

Pressure is defined as a force per unit area when the force acts at right angles to a surface. A water distribution system should be designed to maintain a specific minimum pressure at all water taps including fire hydrant locations under all conditions of design flow. Pressure is one of the principal hydraulic parameters to maintain stability in the distribution system and high and low levels of pressure can influence the degree of microorganism in the water and at the same time the chlorine decay.

##### **Velocity**

There is a strong influence of flows and velocities on transport, mixing, production and decay of substances in the network which impose a different approach to water-quality modeling. The hydraulic simulation of velocities is made to create a clear picture of water movement together with calibration and validation of the model. The flow velocity in the water distribution networks is very different depending on the part of the networks considered.

Larger diameter pipes also result in lower water flow velocities in the water system that lead, in turn, to the deposition of sediments.

#### **– Physic-chemical Parameters**

Water quality is a complex subject, therefore, people have found differences regarding the state of the water, such as temperature, color, taste, odor, etc. these parameters now can be expressed as measurements and indicators of water quality, using a simple measure that can be done on-side, but the quality of drinking water is a health consideration and must have not excessive concentration of minerals and chemical compounds, it must be free of toxins and must not contain disease organisms. To make it possible, it is necessary to collect a sample and analyze it in another location, like a laboratory. As an overview of the factors affecting water quality in terms of chlorine we have chosen four physic-chemical parameters to be analyzed. A brief summary describing the parameters is express as follows:

#### **Chlorine added to water or Initial chlorine**

Chlorine ( $\text{Cl}_2$ ) is one of the most widely used disinfectants. It is very applicable and very effective for the deactivation of pathogenic microorganisms. Chlorine can be easily applied, measures and controlled. It is fairly persistent and relatively cheap. Chlorine has been used for applications, such as the deactivation of pathogens in drinking water, swimming pool water and waste water, for the disinfection of household areas and for textile bleaching, for more than two hundred years. When chlorine was discovered we did not now that disease was caused by microorganisms. However, we only started using disinfectants on a wider scale in the nineteenth century, after Louis Pasteur discovered that microorganisms spread certain diseases. Chlorine has played an important role in lengthening the life-expectancy of humans.

Public water systems use chlorine in the gaseous form, which is consider too dangerous and expensive for home use. Other systems uses also liquid chlorine (sodium hypochlorite  $\text{NaClO}$ ) or dry chlorine (calcium hypochlorite  $\text{Ca}(\text{ClO})_2$ ). While any of these forms of chlorine can effectively disinfect drinking water, each has distinct advantages and limitations for particular applications.

The importance of knowing the quantity of the chlorine added to water, falls exactly on the need of knowing the amount of chlorine that has disappeared, in other words, with this parameter it is possible to calculate chlorine decay. Chlorine added to water has a high influence on chlorine decay as it has been demonstrated by some researchers including, Dandy et al. – Forecasting chlorine residuals in a water distribution system using a general regression neural network. Chlorine added to water has a significant relationship with the output variable that will be modeled.



## **pH**

pH is the measure of the alkalinity in a solution and it is defined as a measure of the activity of the hydrogen ion ( $H^+$ ) and is reported as the reciprocal of the logarithm of the hydrogen ion activity.

$$pH = -\log a_{H^+} \quad (16)$$

In water, pH is one of the most sensitive indicator in balance state. In general, a water with a  $pH < 7$  is considered acidic and with a  $pH > 7$  is considered basic. The normal range for pH in surface water systems is 6.5 to 8.5 and for groundwater systems 6 to 8.5. Alkalinity is a measure of the capacity of the water to resist a change in pH that would tend to make the water more acidic. The measurement of alkalinity and pH is needed to determine the corrosive of the water. The pH of pure water ( $H_2O$ ) is 7 at 25 °C and is neutral, but when exposed to the carbon dioxide in the atmosphere this equilibrium results in a pH of approximately 5.2. pH is a suitable input to predict chlorine decay, even though previous studies consider it as a medium and low influence in chlorine prediction.

## **Turbidity**

Turbidity is a measure of the degree to which light is scattered by suspended particulate material and soluble colored compounds in the water. It provides an estimate of the muddiness or cloudiness of the water due to clay, silt, finely divided organic and inorganic matter, soluble colored organic compounds, plankton, and microscopic organisms. The measurement of turbidity is a key test of water quality.

Fluids can contain suspended solid matter consisting of particles of many different sizes. While some suspended material will be large enough and heavy enough to settle rapidly to the bottom container if a liquid sample is left to stand (the settle-able solids), very small particles will settle only very slowly or not at all if the sample is regularly agitated or the particles are colloidal. These small solid particles cause the liquid to appear turbid. In water treatment plants the sedimentation and filtration help to settle the major part of materials and suspended solids.

Turbidity is an optical property of a suspension and its measurement can indicate that water carry materials such as pesticide, heavy metals and bacteria. As one of the most important cause of chlorine disappearance in water is the matters (organic and inorganic) in it, turbidity is part of the physic-chemical indicators affecting or influencing chlorine decay.

## **Temperature**

The temperature of water is a measure of its internal, thermal energy content. It is a property that can be sensed and measured directly with a thermometer. Heat content is a capacity property that must be calculated. Heat content usually is considered as the amount of energy above that held by liquid water at 0°C it is a function of temperature and volume. Temperature affects not only the solubility and toxicity in water but also it represents a reference to define many other parameters. Drinking water optimal temperature rates between 8°C to 12°C. As the temperature increases the rate of reaction increases as well. As a rough approximation, for many reactions happening in a pipe line, the rate of reaction doubles for every 10°C rise in temperature, in other words, particles can only react when they collide. If the temperature rise in the water, the particles move faster and so collide more frequently. That will speed up the rate of reaction.

### ***3.3. Construction of the input database using the Monte-Carlo method***

The Monte Carlo method consists of the performance of a simulation using random numbers to determine the future behavior of a random variable. The application of the Monte Carlo method combines the random numbers obtained, with the function that represents the frequency distribution of the historical variations of the variable. The simulation will be run only in few selected points inside the network to obtain a large database of variables affecting the residual chlorine decay with the following methodology:

- Specify the variables and objectives of the model
- Estimate the probability distribution that explains the behavior of random variables not-controlled in the model
- Calculate the cumulative probabilities of each of the variables
- Generate a random number
- Link the random number with variables which cumulative probability is less than or equal to the random number obtained
- Repeat the process a large number of times, until obtain the required or desired number of sample values
- Performed with the variables obtained the operations specified in the model
- Analyze the distribution functions of target variables obtained with the indicated operations as a tool for decision making

Monte Carlo simulation technique will be used as it provides flexibility and can include number of sources of uncertainty. Next the data of each parameter are laid out. Initial Chlorine, pH, pipe roughness, Turbidity between others factors of the initial conditions of the water distribution system will be generated by Monte Carlo simulation technique that requires the use of random number generator. Random number generator used for simulating factors affecting chlorine decay in WDS generates the numbers that follow a normal distribution and uniform distribution depending of the parameter simulated. Generated factors are then to be compared with the actual factor to check the significance of the simulated factor. Factors affecting chlorine decay as chlorine added to water, pH, flow, turbidity, temperature, etc generated by using MC Method will be used in the next chapters for calculation of chlorine decay using Artificial Neural Networks and also in the rest of the point in the WDS with EPANET 2.0.

Monte Carlo method was invented by Stanislaw Ulam, a Polish born mathematician, in 1946 while he was determining the probabilities of winning in a card game of solitaire. The Monte Carlo method provides approximate solutions for many mathematical problems by generating random numbers and calculating what fraction of the numbers obey some property or properties. Monte Carlo method is useful for examining numerical solutions to problems which are too complex to solve.

Nicholas Metropolis and Stanislaw Ulam [16] presented motivation and a general description of Monte Carlo method dealing with a class of problems in mathematical physics. Monte Carlo method is, essentially, a statistical approach to the study of differential equations, or more generally, of integro-differential equations that occur in various branches of the natural sciences. In general Monte Carlo method can be performed by carrying out the following steps:

- Define a domain of possible inputs.
- Generate inputs randomly from the domain and perform deterministic computation on them.
- Aggregate the results.

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 followed 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. For calculation of each factor affecting chlorine decay, all the parameters can be simulated using MS Excel. Microsoft Office Excel uses a function called Data Analysis this statistics add-in provides new functionality to the spreadsheet. Among them, we emphasize the Generation of random numbers. As per rule, the numbers can be generated from the minimum number that can be achieve to the maximum number, after many replications, the stored results will mimic the sampling distribution of the statistic. Monte Carlo techniques can provide information about sampling distributions when exact theory for the sampling distribution is not available.

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.

### **3.3.2. Test of Significance**

In order to confirm that simulated hydraulic and quality parameters agree to the actual values of each measured parameter, a test of significance is required. Chi-square ( $\chi^2$ ) test will be performed on each observed measures. The primary purpose of Chi-square test is to compare some observed values to the expected values. Chi-square test is performed to test whether observed differences between expectation values and measured values are acceptable.

The hypothesis is to use a measure of goodness of fit to describe how well a simulation fits a set of observation, in other words whether the two values, simulated values and actual values of the hydraulic and quality parameters affecting chlorine decay, follow a normal distribution or not. Equation 17 gives the formula for calculating chi-square values.

$$\chi^2 = \sum_{i=1} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} \quad (17)$$

### 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:

- Multilayer Perceptron Neural Network – MLP

It is located between the most known and used ANN. Its basic element (Neuron) is Perceptron. MLP is a supervised network, that is to say, the training sample must have besides the input values also the appropriate outputs values. This Network is especially suitable for regression problems and prediction, but it can also be used for classification. The MLP consists of a feed-forward network of neurons which map input vectors to output vectors. Each artificial neuron consists of a linear combination of weighted inputs which is passed through a non-linear activation function to produce the neuron's output. Backpropagation algorithm can be used very generally to train neural networks, but it is most famous for applications to layered feedforward networks, or multilayer Perceptrons.

- Radial Basis Function (RBF)

RBF was proposed as an independent neural network. It has one input, one hidden and one output layer. All the outputs are completely connected with all the neuron in the next hidden layer. The neuron in the hidden layer is so called Radial type, that is to say, instead of a summing part it is calculated by the distance of the input vector from the threshold. RBF is a simple network

that has a fast learning process and is special for classification as well for regression.

- Self-Organizing Feature Map – SOFM

Unsupervised learning is the name to the process performed by the SOFM which only uses the input values (independent variables). The neural network only does the analysis of these data. SOFM have only one layer named competitive layer. The inputs to the network are completely connected, that is to say, all neuron have the information of all the inputs.

- Adaptive Linear Neuron - ADALINE, MADALINE

ADALINE is a single layer neural network developed by Professor Bernard Widrow and Ted Hoff in 1960. It consists in a weight, a bias and summation function. MADALINE is an extension of ADELIN of two layer neural network. For problems with multiple input variables and outputs variables, each input is applied to one ADALINE and MADALINE can be used in parallel. It is usually involved in prediction problems with several inputs.

- Bayesian Networks (Probabilistic Neural Networks - PNN and General Regression Neural Network - GRNN)

Is a probabilistic graphical model that represents a set of random variables via a directed acyclic graph. Generalizations of Bayesian networks can be represent and solve decision problems under uncertainty. Probabilistic Neural Networks (PNNs) and Generalized Regression Neural Networks (GRNNs) were introduced by Specht in 1990 and 1991, respectively. While PNNs are generally used for classification problems and, therefore, distinguish between different categories of patterns, GRNNs estimate the most probable value for continuous dependent values. Both network types compute the probability density functions of the given patterns and finally attribute them to the class or value to which they most likely belong. BNNs are feed-forward networks which do not use back-propagation. Their architecture is comprised of four layers [4].

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. After the selection of the ANN type the next step is the creation of several models based

on the input parameters, that is to say, divide the input parameters and create models to compare the performance and the influence of each input parameter.

#### **3.4.1 Division of models into subsets of parameters**

To obtain better results in the performance of the ANN model is necessary to have a clear image of the parameters that can influence chlorine decay. To create a database of parameter that can influence chlorine decay the best way is to divide the data into subset and then build several ANN models combining the parameters until the better results are found. First it will be necessary to create one model with all the parameters available and the next step is to divide those parameters into subset to try to find the best combination of parameter which produces the best model performance. The performance should be measured with the value of the coefficient of correlation and the least squared error. Sensitivity analysis should be also performed to know the influence of each parameter and from the first point intent to create the combination of parameters in each subset depending on the influence of the factors or parameters studied.

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.

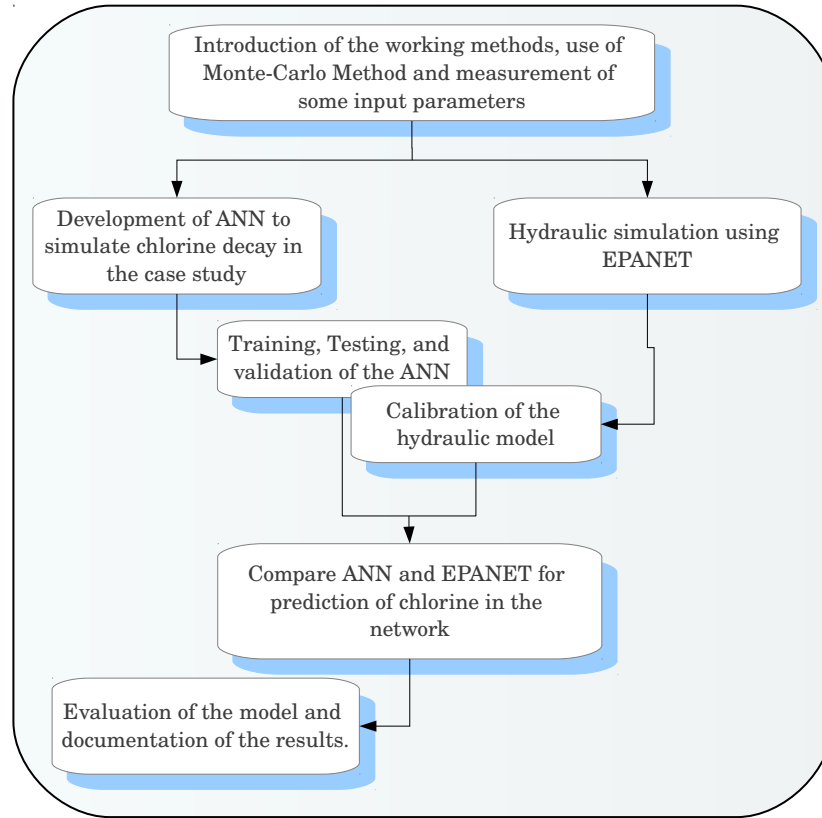
### **3.4.3 Performance Evaluation**

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 5**.

The next step should be the performance of the quality model to simulate chlorine decay in EPANET, as explained in section 3.1.1, 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.

A Sensitivity analysis should then be performed to explore the reaction of the system caused by changes in the inputs. Several investigators have advocated the use of sensitivity analysis for the confirmation and evaluation of simulation models. It is clear that a measure of the importance and error levels of model terms can be helpful during model development and can also provide some insight into the confirmation process [17].





**Figure 5:** Program Structure

Sensitivity analysis, then, does not yield a measure of model confirmation, but it can provide information that is extremely useful in model testing and development. Several studies have examined issues in the development, calibration, confirmation, and prediction of water quality simulation models. Sensitivity analysis and Monte Carlo simulation have been used to examine the impact of error terms on the prediction error. Among the strong conclusions apparent in several articles and projects about the sensitivity analysis is that data are often inadequate for effective calibration and confirmation of mathematics models. This situation clearly limits the degree to which we may apply statistical confirmatory criteria and ultimately affects the reliability of planning models and methods [13].

# 4

## Case Studies

---

In this chapter will be described the experiments and tests conducted to evaluate the methodology proposed in the thesis 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 under the support of the grant-competition for specific projects in Brno University of Technology, BUT (Juniorský grantový projekt 2010). The outcomes of the project are aimed to compare the forecasting accuracy between two data-driven models (MLR and MLP) and the effect of the physical and chemical parameters that directly influence chlorine decay. Complex chemical and physical processes within a certain node of a water supply network can exist and one of the outcomes of this project was to compare exactly all the inputs that affect the output parameter, that is to say, the historical data collected and all the information can be adapted and analyzed to choose the input parameter that have a strong influence in chlorine decay. In a general view, the tests done in this distribution system are aimed to compare the results obtained from the traditional linear regression statistical method and the ANN models.

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. The project was partly supported by the grant *Juniorský grantový projekt 2011* at BUT and the second part of the project was supported by Innovation vouchers 2011, which is a tool for promoting cooperation between companies and Brno research institutions. The cooperation was done between the water utility BVK and BUT. 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. A comparison between the water quality prediction using EPANET and ANN is also carried out in this part of the chapter, making the ANN model useful as a tool for comparison and calibration of water-quality model results in EPANET.

#### ***4.1 ANN for Prediction of Chlorine in a WDS - Case Study: Našiměřice, Czech Republic***

The project is focused 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 data-driven methods to identify and predict the evolution of selected water quality parameters. The wide open used data-driven methods in water management are 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 continuous variable. The performance of MLP and MLR are evaluated using 4-years old database (set of inputs) collected in the city of Našiměřice Czech Republic. The first part of the paper shows a summary of the state of the knowledge in modeling using ANN and the second part describes the collection of data and construction of the models.

The main expected contributions of the project are:

- (1) prediction of the given water quality parameter or coefficient using both stochastic methods MLR and MLP through the evaluation of the historical data of some physic-chemical parameters measured and added at several point of the water distribution system (WDS).
- (2) Combination of Data-driven model and hydraulic model analysis, to predict water quality parameters within a WDS under consideration. The way these models are going to be mixed is creating a stochastic model to calculate a water quality parameter in a WDS. Data-driven method will be designed to predict the evolution of some water quality parameters using historical data.

This approach will be tested on a concrete site using a hydraulic model built on the software EPANET 2.0.

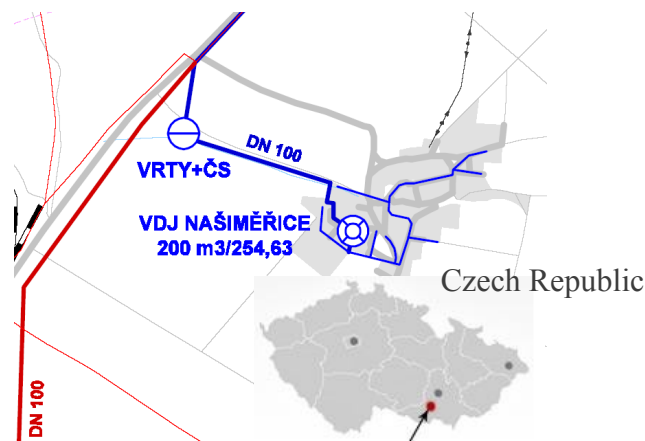
##### 4.1.1 Current state of the drinking water supply system

###### Available Data

Našiměřice village is situated about 10 km west of Pohořelice (Czech Republic), the altitude ranges between 210 to 230 m. 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 HV 3 to Našiměřice town over the power of Iron pipe lines DN100 and PVC 90 to a hydroglobe tank of 200 m<sup>3</sup>. See **Figure 6**.

Drinking water is distributed to town by gravity through pipes that ranges from steel, PVC and cast iron with diameters from 25, 80 and 100 mm. The average age of the drinking water network system is 27.7 years.



**Figure 6:** Scheme water supply system Našiměřice

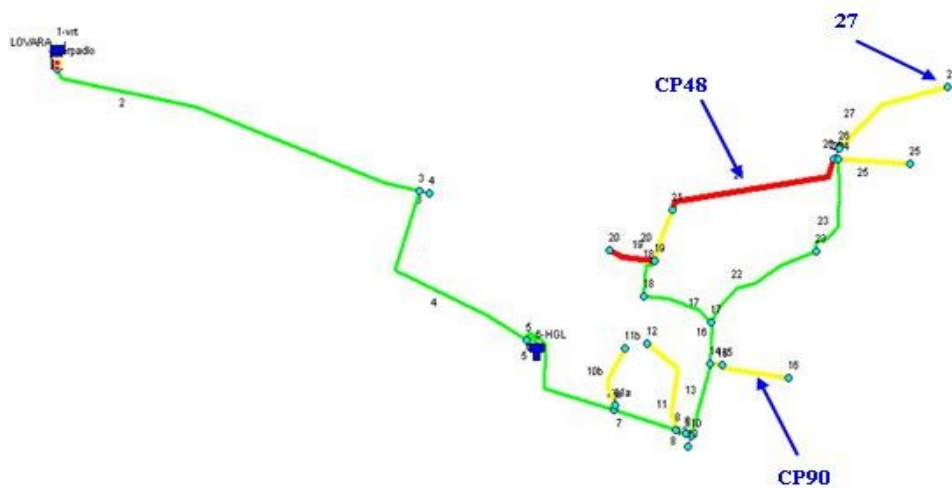
The main source of drinking water for the village Našiměřice is HV 3 well effluent, where it is permitted to take monthly an amount of water allowed by  $Q_{m,max} = 1667 \text{ m}^3/\text{month}$  and the annual allowable amount of water  $Q_{r,max}$

= 20,000 m<sup>3</sup>. Water from the well effluent is healthy and securely chlorinated, more recently, there were problems with water quality in terms of the increased presence of uranium, it is being investigated the possibility of re-bonding to another source of drinking water. Water is pumped from a well effluent to a water tank (Hydroglobus) in the village Našiměřice with a volume of 200 m<sup>3</sup> with a pump Lovara L4GS11T.

The water is pumped from a well, HV3 (put into operation in 1962) for Našiměřice to yield 1.7 liters/sec, the discharge into Našiměřice height is 49 m. The water treatment is only with chlorine but high levels of uranium has been found lately, it is expected that in the past when the country was still sea, heavy metals were settled (hence uranium) and were kept in the basement today. The total length of pipe is about 16,183 km.

There is a water tank (Hydroglobus) in Našiměřice with the following properties:

- Volume 200m<sup>3</sup>
- Max. water level. 254, 63 m n. m.
- Water temperature in the range of consumer from 3 to 16° C
- Put into operation in 1975



**Figure 7:** Našiměřice, Czech Republic, Water supply system



**Figure 8:** Water tank (Hydroglobus) in Našiměřice

#### **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. included in the project Našiměřice – Analysis of water leakage risks in water distribution systems. (*Našiměřice - Analýza rizik úniků vody ve vodovodní síti*) it was built 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.

**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	<b>0,29</b>	<b>0,44</b>	<b><u>0,80</u></b>

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, HGL 200m3. 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.

### Data used in the construction of the hydraulic model

Tests 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	1994	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

The summary in **Table 2** shows the material of the pipes for each section, the diameter, the number of connections to each node and the demands which was calculated based on the global reference obtained for year 2010 in Našiměřice

provided by the water utility and then distributed spatially taking into account the number of connections, houses and areas. The **Table 2** also shows the ground elevation of each node and the age of pipes for each section.

### 4.1.3 Construction of ANN Models

#### Selection of the Input Indicators

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. In this case of study 693 data records were collected and the 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** where the categorical values were transformed to numerical values 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 9**.



**Figure 9:** Instruments installation for measurement in Našiměřice



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.

**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 measure in the each location listed)
		1	2	3	
Inputs		x	x	x	ZD WC
	Temperature	x	x	x	č.p. 117,kohout na zahradu
	pH	x	x	x	č.p. 83, OÚ, kuchyňka
	Flow	x	x	x	č.p. 44,dvůr - venkovní kohout
	Pipe material	x	x	x	VODOJEM, Našiměřice – odtok
	Diameter	x	x	x	odkalení naproti č. 113
	Age of pipes	x	x	x	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, Uřad - 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

A preliminary study of the correlation using the Statsoft STATISTICA 9.0 Software Package, allowed us to analyze the influence of the input indicators on the free chlorine determination on the water network. This analysis showed that Temperature and Pipe material have major influence on the predicted

free chlorine; also the Diameter represents a significant influence in the forecasting. Age of pipe has a moderate influence and finally Flow, Age of pipe, and pH have a low impact on the free chlorine forecasting. The **Table 5** summarizes the matrix of correlation between the indicators. On the basis of these results 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 transform into numerical and Model 3 was constructed using MLR analysis. A summary of data used in each model is also shown in **Table 3**.

**Table 5:** Matrix of correlation between all variables

	Temperature	pH	Flow	Pipe material	Diameter	Age of pipes	Free Chlorine
Temperature	1						
pH	0.301318	1					
Flow	0.128414	0.176775	1				
Pipe material	-0.063939	0.057395	0.152551	1			
Diameter	-0.178952	-0.152685	-0.309294	-0.247625	1		
Age of pipes	0.36235	0.190642	0.154843	0.691905	-0.030569	1	
Free Chlorine	-0.543125	-0.269448	0.152221	-0.31402	0.441737	-0.34212	1

#### 4.1.4 Outcomes of the research

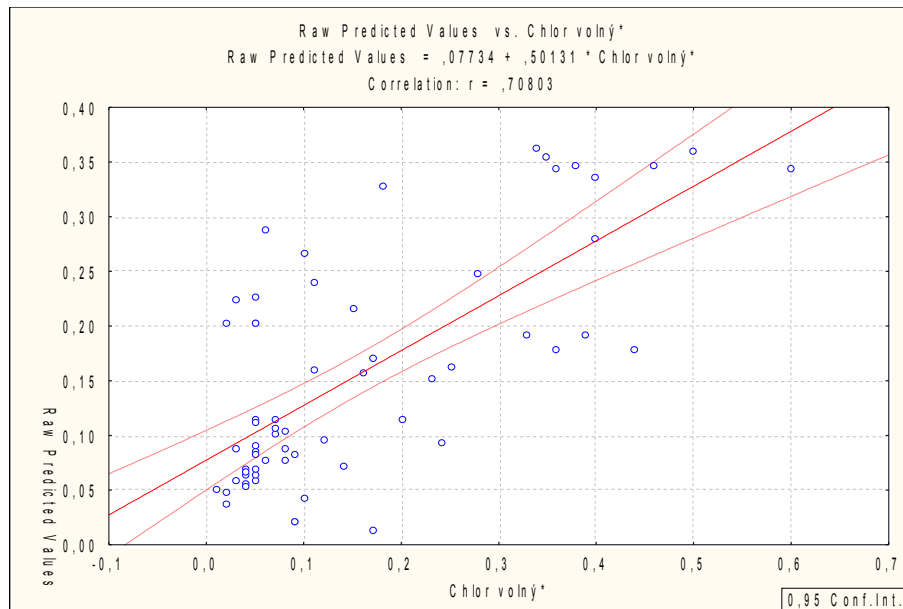
The type of ANN model used was the well known Multilayer Perceptron (MLP). The MLP is a feed forward ANN model that maps the sets of input data onto a set of appropriate output. A MLP consist of multiple layers of nodes in a directed graph, which is fully connected from one layer to the next. Each node in the hidden layer and the output layer uses a nonlinear activation function as mentioned in Chapter 2, Section 2.4.3. 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).

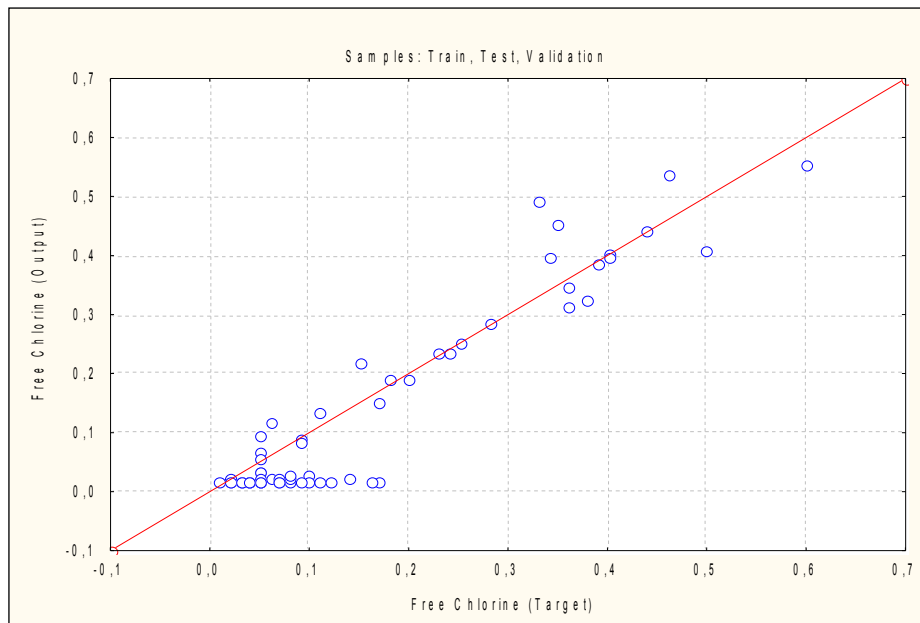
The training, testing and validation results for each of these models are given in **Table 6** (Models 1-3). By the virtue of the MLP architecture, the training set can be predicted to a high level of accuracy in comparison with MLR the MLP achieved a significantly lower error for the training, testing and validation sets and this proves 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 10** The performance obtained by MLR was 0.70803 which is significantly lower than the achieved by MLP in **Figure 11** and **12**.

**Table 6:** Model Performance

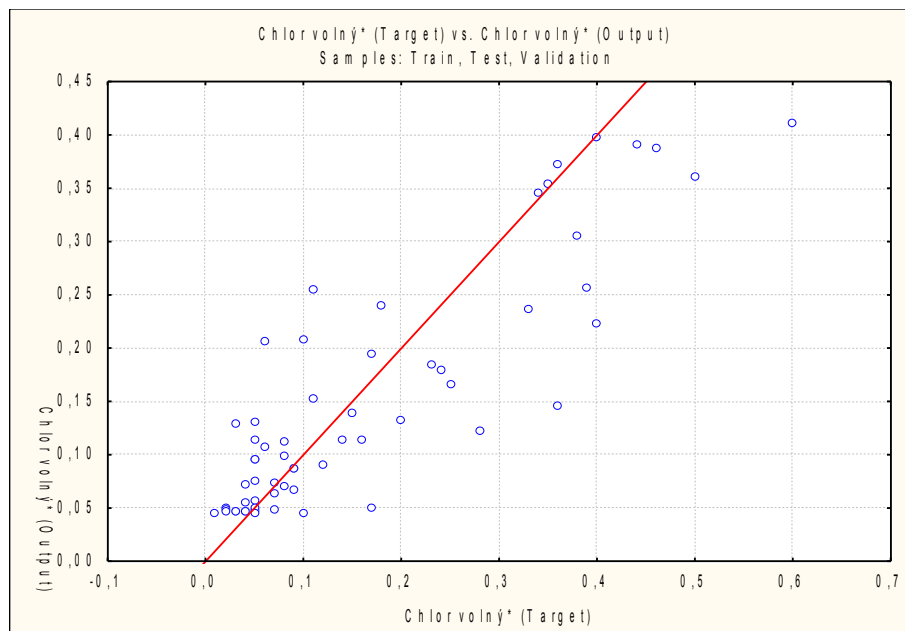
Model No.	Model Type	Training perf.	Testing pef.	Validation perf.
1	MLP	0,97838	0,87358	0,89696
2	MLP	0,90734	0,73009	0,94906
3	MLR	0,70803	0,70803	0,70803



**Figure 10:** Predictions for Multiple Linear Regression Model 3



**Figure 11:** Training, Test and validation Predictions for MLP Model 1



**Figure 12:** Training, Test and validation Predictions for MLP Model 2

#### ***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. The descriptive statistic and correlation matrix allowed creating a sensitivity analysis of the input selected. 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. The construction of a large database in a specific WDS with the historical data can be used to create a software based in ANN approach and calibrate the water quality parameters of the model. 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***

There are several methods of disinfection that can be used to adjust the flow of larger quantities. They can be divided into three main groups: UV radiation, Ozone and chlorination. Disinfection with chlorine has the advantage of efficiency and durability. The concept of residual disinfectant concentration is associated with durability of disinfection, to ensure the disinfection of drinking water is essential to keep residual disinfectant concentration in the water supply system. This can prevent contamination of the water supply system by means of pathogens or microorganisms. In this study the goal to determine the factors influencing the chlorine decay in the pressure zone of Brno, Kohoutovice. These factors are based on local measurements of residual chlorine.

For determining factors affecting chlorine decay in water distribution systems (WDS), under different parameter conditions, values of historical data are required. Initial Chlorine, pH, Flow, Temperature and turbidity between other

factors are used in this study as Input for forecasting chlorine decay in Brno, Kohoutovice - pressure zone, Czech Republic Case Study. The methodology for simulation of factors affecting chlorine decay in WDS using historical data obtained from the water utility is laid out in this chapter. Initial Chlorine, pH, temperature, between others factors of the initial conditions of the water distribution system will be generated by Monte Carlo simulation technique that requires the use of random number generator. Random number generator used for simulating factors affecting chlorine decay in WDS generates the numbers that follow a normal distribution and uniform distribution depending of the parameter simulated. Generated factors are than to be compared with the actual factor to check the significance of the simulated factor. Factors affecting chlorine decay as chlorine added to water, pH, flow, temperature, etc generated by using MC Method will be used next for calculation of chlorine decay in the rest of the point in the WDS using Artificial Neural Networks.

### ***4.2.1 Hydraulic model for the selected pressure zone***

#### **Selection of the pressure zone**

Some specification should be taken into account before choosing the pressure zone. The following is a description of the zone needed:

- Small pressure zone (less than 500 nodes and sections) to confirm easily the objective of the project.
- Measurement of chlorination at the beginning of the system
- Accessibility for measurement campaign and installation of equipment for testing and taking samples.

Based on the considerations, it was proposed by *Brněnské vodárny a kanalizace a.s.* (BVK) three pressure zones:

#### ***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 13** 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.

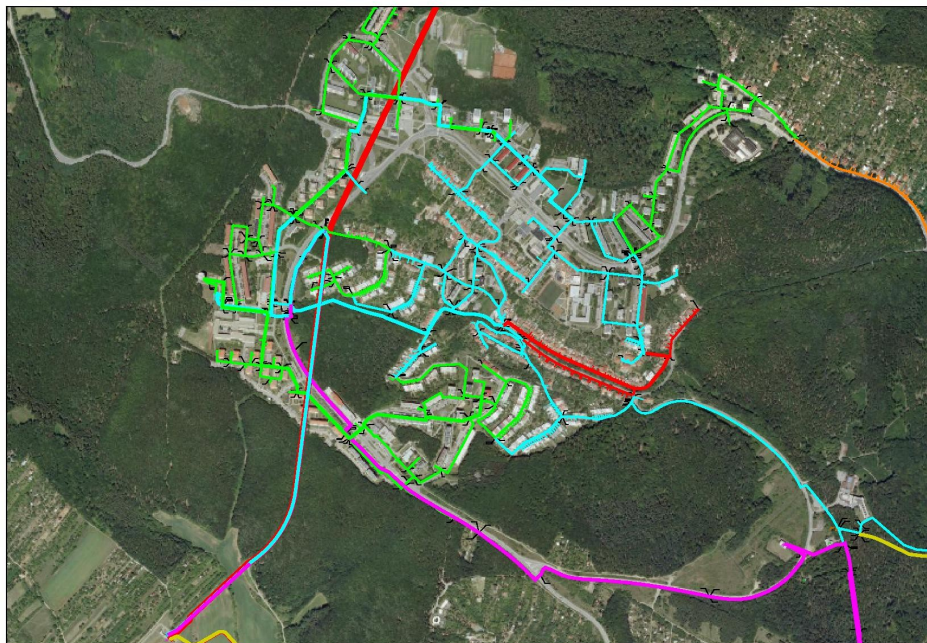
### Description of the current situation in the drinking water supply system

City District Brno-Kohoutovice lies approximately 7.3 km west of Brno-center, the altitude ranges between 410 to 285 MSL. 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 Myslívna. 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  $6550 \text{ m}^3$  with maximum water level of 320 MSL, the water column height is of 6.5 meters. Kohoutovice tank has two chambers and volume is  $3000 \text{ m}^3$  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 13:** Water supply scheme

### Creation of the hydraulic model

Creation of the hydraulic model for the studied zone inside the water distribution system involved the following steps:

1. The base model was taken originally from the General Plan of the 2008 water supply system, Brno. The model does not contain coordinates.
2. Using the software Mike Urban with the export function from the original file GDB, it was created a new file INP and NET that has the possibility to be read by the computational software EPANET 2.0.
3. Spatially was allocated customer demands in the network nodes and also it was incorporated two new pumps to pump the water from Bosonohy to Kohoutovice.
4. Characteristics such as the elevation in nodes, tanks and some valves were already in the base model.
5. In the pressure zone *1.3.2 VDJ Zemní Kohoutovice* there are connected via pressure regulators these pressure reduction valves (included in the model):

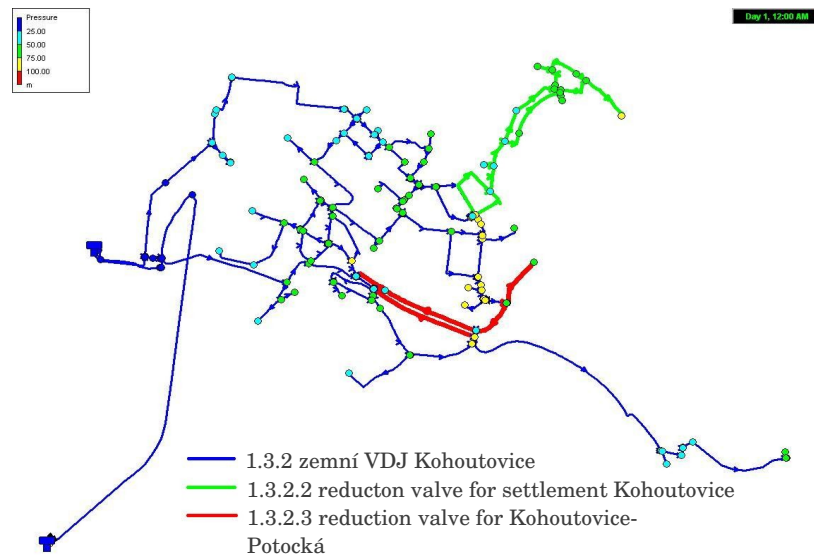
- 1.3.2.3 Pressure reduction valve in Kohoutovice-Potočka
- 1.3.2.2 Pressure reduction valve for the down town zone. Settlement Kohoutovice

The base model was created only for a maximum hourly need  $Q_h = 22.2$  l/s (one point in time). There are no large customers, that is to say, customers over 10000 m<sup>3</sup>/year. Water demand was divided into nodes according to the actual annual consumption of all customers connected to that section.

The base model had not been properly calibrated but in this study, there was performed calibration and quality control review of the resulting model. This model covered only 1 hour in time but the roughness coefficients and demands were already calculated. 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 (48 hrs) using EPANET 2.0. See **Figure 14**.

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. Appropriate pipe roughness coefficients were assigned as per **Table 7**. The authors applied suggested safety factors and it is represented variability in the pipe absolute roughness values.





**Figure 14:** Hydraulic model run in EPANET 2.0

**Table 7:** Roughness Coefficient (mm) for the pressure zone studied

Mat.	Abb.	Description	2000 2005	1990 1999	1975 1989	1960 1974	1949 1959	Before 1949
-1	~	Not Assigned						
0	-	Not Assigned						
8	PP	Prolypropylene						
9	PB	Lead						
1	GI	Grey Iron	0,4	1	2	3	5	5
2	DICL	Ductile Iron	0,1	0,4	1			
3	STL	Steel	0,3	1,5	2,5	4	5	5
4	PE	Polithylene	0,01	0,03	0,1	0,2		
5	GFRP	Fiberglass	0,01	0,03				
6	ACP	Eternit			0,4	1	2	3
7	PVC	Polyvinyl Chloride	0,03	0,05	0,1	0,2		

### Measurement of the water quality and hydraulic parameters

After the creation of the hydraulic model an important next step is to calibrate the model using measure parameters on-site and compare it with the results values of the model in EPANET 2.0. The most common parameters to analyze and measure for hydraulic model calibration are; pressure, flow and water level in tanks. 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.

### **Calibration of the hydraulic model**

Observable model outputs are pressures, flows, tank water levels, and water-quality predictions. There were performed two levels of calibration. One level serves as a reality check that the model is producing reasonable, but not necessarily highly accurate results. It was then checked for the following problematic behavior:

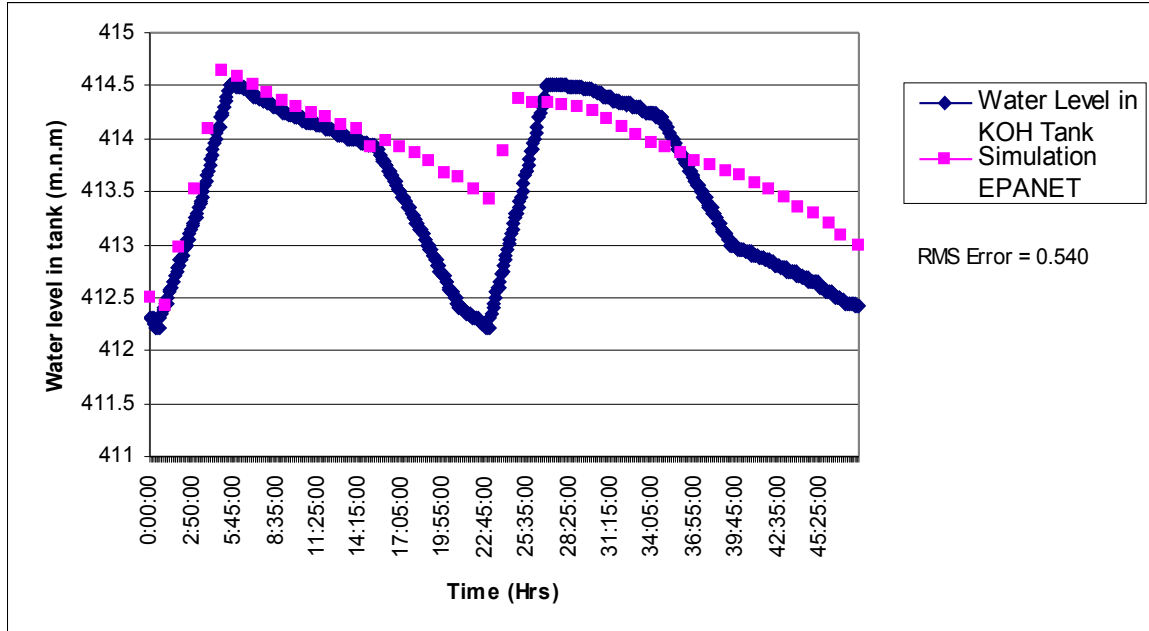
- Unreasonably low (e.g., negative) or high pressures.
- Pumps operating outside of their allowable range or being shut down for this reason.
- Pumps cycling on/off in an unreasonable fashion.
- Tanks that continuously keep filling or emptying.
- Nodes disconnected from any source because of closed pipes, pumps, or valves.

All of these conditions were checked resulting there was no problem in representing some aspect of the system to the computer.

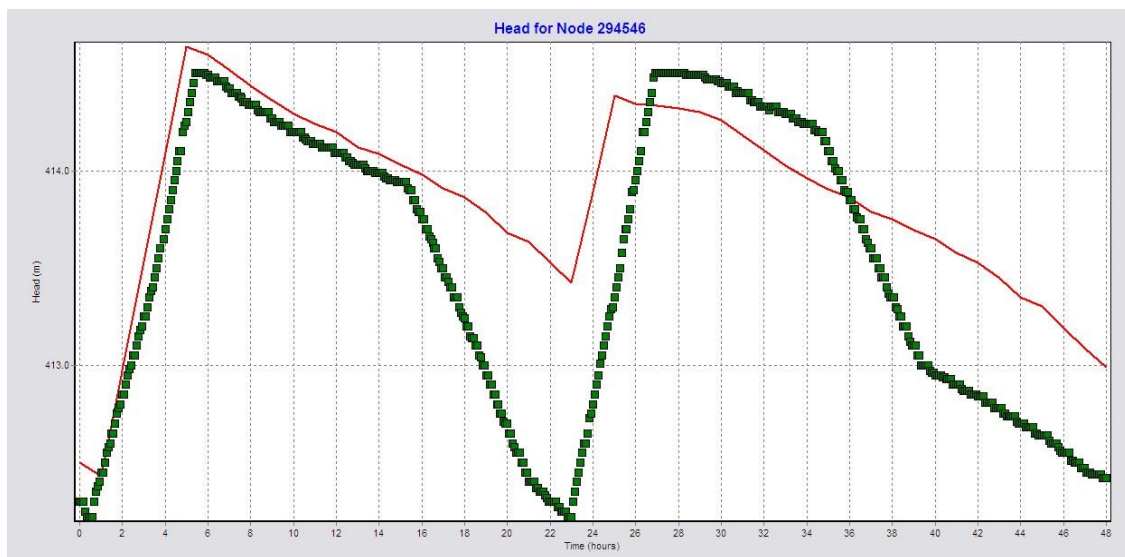
The second level of calibration involved adjustments to model input parameters that match best with field observations. This required the collection of field data, under more than one operating condition. This included flow rates and pressures at supply points of the pressure zone and water levels in storage tanks. 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.

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 15** 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 magenta line represents the water level simulation from the computational software EPANET 2.0 and the blue line shows the water level in Kohoutovice

tank measured by BVK using the real time measurement capability of the SCADA system.

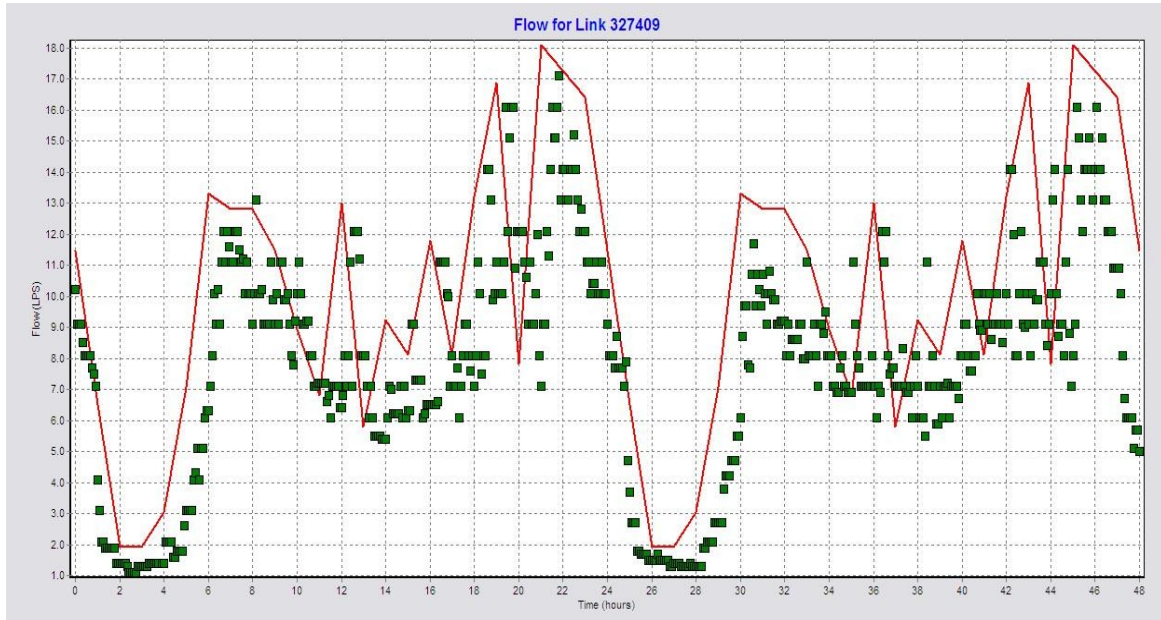


**Figure 15:** Time-Series plot Calibration for Tank level in Kohoutovice

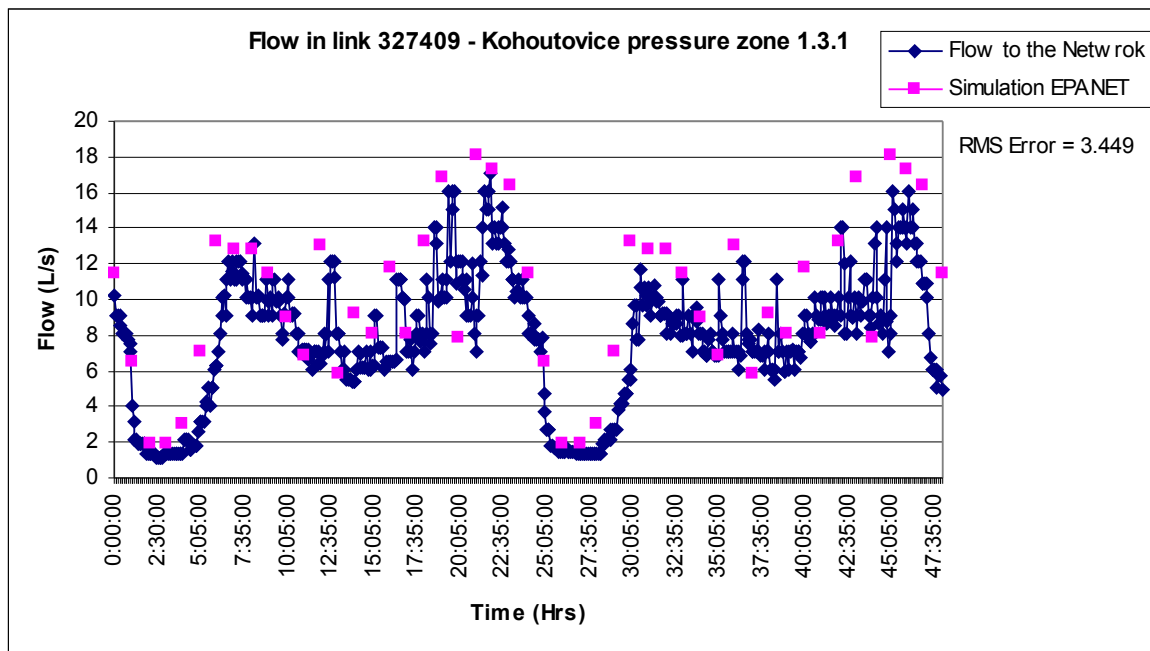


**Figure 16:** Calibration Plot from EPANET for Head values at Tank Kohoutovice (Red line represents simulation and Green line the observed values)

**Figure 16** shows the same time-series plot results as in **Figure 15** but exported from the computational software EPANET 2.0, where the calibration was done, including the data obtain from BVK about the water level in Kohoutovice tank and the results from the EPANET simulation itself. The Root Mean Squared Error was 0.54 for a 48 hours of this simulation.



**Figure 17:** Calibration Plot from EPANET for Flow values at Link 327409 (Red line represents simulation and Green line the observed values)



**Figure 18:** Calibration Plot in Excel for Flow values at Link 327409

Other results from the hydraulic calibration can be seen in the following **Figures 17** and **18**. Where the quantity of flow was calibrated using the data measured from the flow-meter installed at link 327409. It was then compared using the calibration inside EPANET 2.0 as shown in **Figure 17** and MS Excel in **Figure 18**. The Root Mean Squared Error from the calibration was 3.449, for the 48 hours of simulation.

#### ***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

$K_w$  - Wall reaction coefficient

Ideal calibration for  $K_b$  – Bulk reaction coefficient must be conducted by taking samples of water in a series of non-reacting bottles and analyzing the contents of each bottle at different points in time. For  $K_w$  - Wall reaction coefficient, samples may be measured under ideal conditions (i.e. long isolated pipes, no connections, controlled flow, inline chlorine measure) which in real world these conditions are very infeasible. The value of  $K_b$  was determined based on local free residual chlorine measurements. The EPANET software models the chlorine decay through a first order kinetic law.

To calculate the model for predicting disinfectant rates were determined by collecting field samples as well as conducting laboratory bottle tests using the instrument Spectrophotometer HACH LANGE DR 2800. The data were then analyzed to understand those reactions occurring in the main portion of the stream flow (bulk reaction) as well as those occurring on the near the pipe wall (wall reactions).



**Figure 19:** Free Chlorine measurement using Instrument Spectrofotometer HACH LANGE DR 2800

### **$K_b$ - Bulk Reaction coefficient**

Chlorine decay is usually represented as first order, with the decay coefficients  $K_b$  typically being between 0.05 and 15 d<sup>-1</sup>[20]. Coefficients for decay reactions can be reported as negative values. To determine decay coefficient  $K_b$  in Kohoutovice pressure zone 1.3.1 There was used several bottle tests where chlorine decay in a particular volume of water is 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 20**, from which  $K_b$  is extracted using the following relationship [21]:

$$K_b = \frac{L_n \left( \frac{C_1}{C_2} \right)}{(t_1 - t_2)} \quad (18)$$

Where:

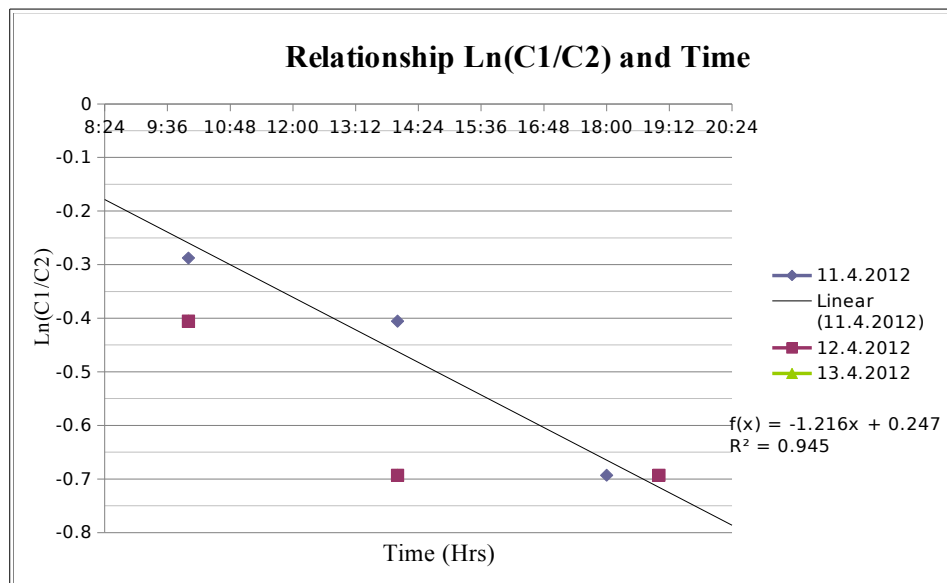
$K_b$  = Bulk reaction rate coefficient ( $\text{day}^{-1}$ )

$t_1$  = 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 20:** Relationship between bulk decay coefficient and time

The global average bulk decay coefficient (e.g.  $T = 12^\circ\text{C}$  :  $-0.14 \text{ d}^{-1}$ ) for samples in Kohoutovice Water System was incorporated into the EPANET 2.0 model using a first-order reaction rate. Initial chlorine concentrations were specified at Bosonohy tank according to field concentration data as it can increase model accuracy. The  $K_b$  for first-order reactions was estimated by placing a sample of water in a series of non-reacting glass bottles and analyzing the contents of each bottle at different points in time. If the reaction is first-order, then plotting the natural  $\log (C_1/C_2)$  against time should result in a straight line, where  $C_1$  is concentration at time  $t_1$  and  $C_2$  is concentration at  $t_2$ . See **Figure 20**.  $K_b$  was then estimated as the slope of this line. Bulk reaction coefficients usually increase with increasing temperature. Running multiple bottle tests at different temperatures will provide more accurate assessment of how the rate coefficient varies with temperature.

### **$K_w$ - 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 equations (19) through (20).

### **Hazen – Williams:**

$$K_w = \frac{F}{C} \quad (19)$$

### **Darcy – Weisbach:**

$$K_w = \frac{-F}{\log\left(\frac{k}{d}\right)} \quad (20)$$

Where:

F = Correlation coefficients of wall reaction and pipe roughness

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

d = Pipe diameter (mm)

C = H-W C factor

Wall decay coefficients were assigned to this model using the Darcy–Weisbach Equation (4). The value  $F = -0.2$  for all the links and the roughness coefficient  $k$  for each pipe link were determined depending on the pipe material and pipe age as indicated in **Table 7**. The value  $F = -0.2$  was selected with the knowledge that on average it would result in a  $K_w = -0.12$  mm/d. This  $K_w$  is in



the mid-range of typical values that have been estimated for the types of pipes found in Kohoutovice pressure zone.

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***

Some different topologies of feed forward artificial neural networks (ANNs) using the backpropagation learning algorithm can be studied to approach the behavior of chlorine decay for varying levels of chlorine residual in the three nodes inside Kohoutovice pressure zone, in addition, some physic-chemical input parameters (e.g. pH, Temperature, turbidity and flow) will be assessed as they can affect chlorine decay.

#### **Selection of inputs and outputs of the ANN model**

Input selection is based on the existence of a known or suspected relationship with the output parameter, relevant literature, and data availability. The selected output parameter to be studied in this model is residual chlorine. For the given case, 18 available parameters were used to construct the raw database. In the model for chlorine decay simulation in the pressure zone 1.3.1 Kohoutovice it was choose 15 input parameters and 3 outputs parameters.

The model inputs were selected from the available parameters.

- **Input parameters:**

Initial condition parameters

- (1) Initial chlorine in Bosonohy tank
- (2) Free chlorine in Kohoutovice tank
- (3) pH in Kohoutovice tank
- (4) Temperature in Kohoutovice tank
- (5) Flow measured at the out flow of Kohoutovice tank
- (6) Turbidity in Kohoutovice tank

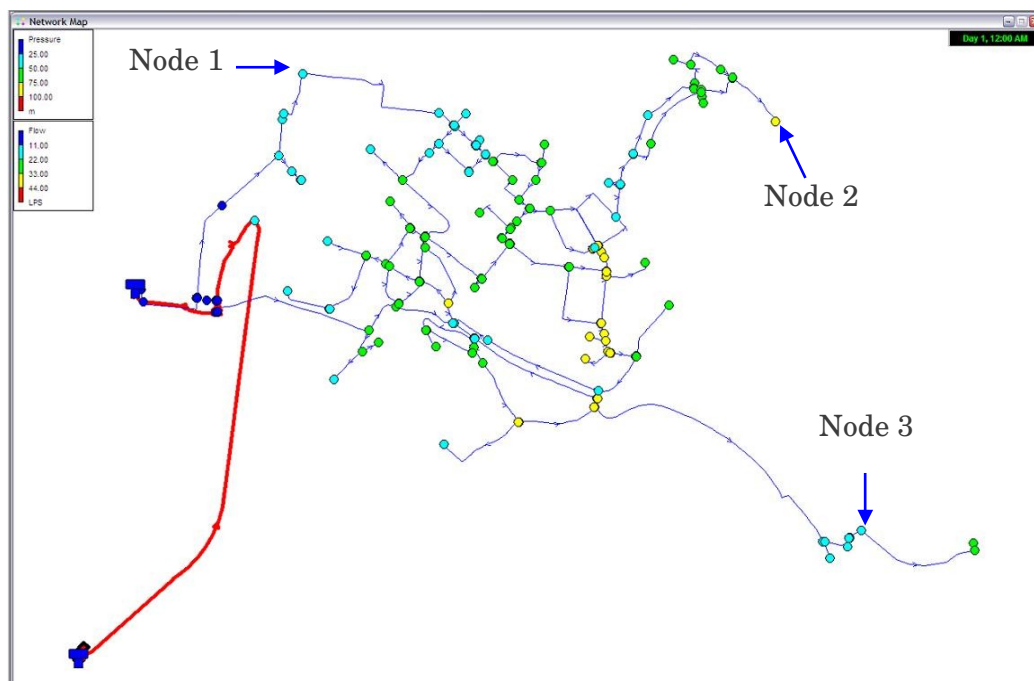
Local condition parameters

- (7) pH node 1
- (8) Temperature node 1
- (9) Turbidity node 1
- (10) pH node 2
- (11) Temperature node 2
- (12) Turbidity node 2
- (13) pH node 3
- (14) Temperature node 3
- (15) Turbidity node 3

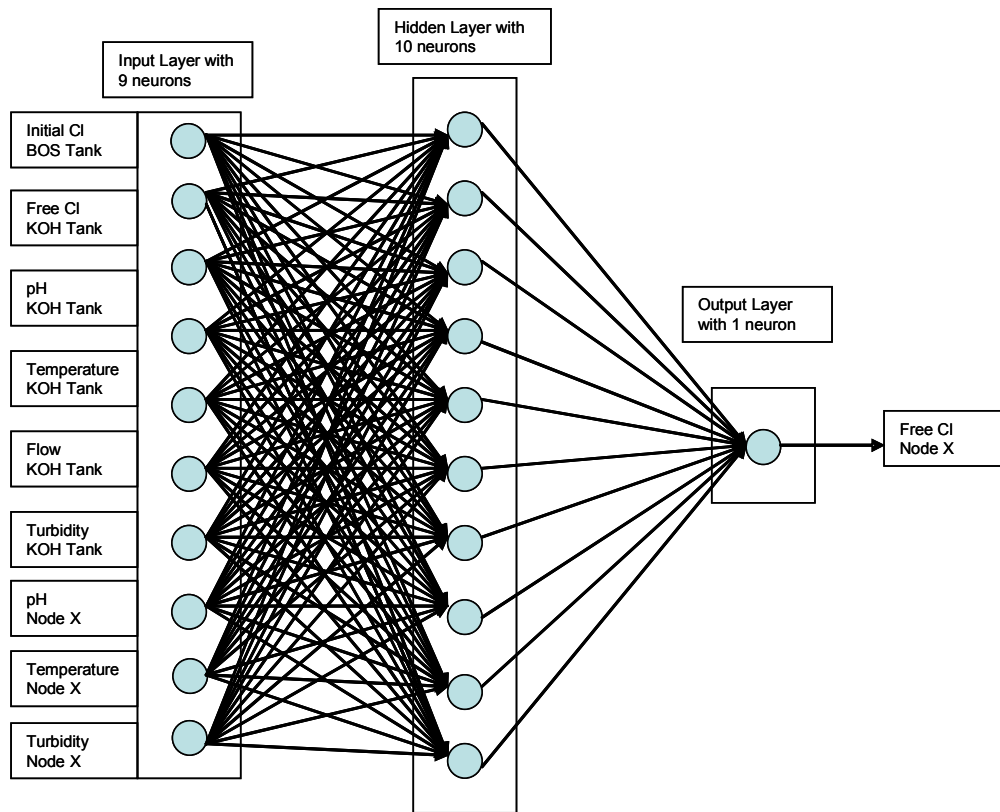
• **Output Parameters**

- (16) Free Chlorine node 1
- (17) Free Chlorine node 2
- (18) Free Chlorine node 3

Chlorine residual was proposed to be simulated in three nodes inside the pressure zone Kohoutovice. See **Figure 21**.



**Figure 21:** ANN simulation of residual chlorine in nodes 1, 2 and 3



**Figure 22:** ANN Structure for residual chlorine modeling

- 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 22** to check the architecture of the ANN used in modeling chlorine decay for the pressure zone in Kohoutovice.

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

The main steps proposed for ANN modeling involves:

- Data preparation,
- Input selection
- Monte Carlo simulation for missing values

- Data division and selection of subsets
- Model creation
- Model calibration
- Performance evaluation

The analytical techniques used to help with the input determination process were the coefficients of correlation and sensitivity analysis. Performance evaluation was then used to test the accuracy of each calibrated model, which included the Root Mean Squared Error (RMSE).

The model methodology is being tested in the Kohoutovice pressure zone and it can only be used for prediction of chlorine residual specifically in this distribution network. Some historical data of parameters affecting chlorine decay was available for Kohoutovice pressure zone and the measurements can be performed on-site easily. 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.

For determining factors affecting chlorine decay in Kohoutovice pressure zone, under different parameter conditions, values of historical data of Initial

Chlorine, pH, Flow, Turbidity and Temperature are required. In this section is laid out the methodology for simulation of factors affecting chlorine decay in Kohoutovice pressure zone using historical data obtained from the water utility BVK. Monte Carlo (MC) method was used as it provides flexibility. Initial Chlorine, pH, Turbidity between others factors of the initial conditions of the water distribution system will be generated by Monte Carlo method that requires the use of random number generator. Generated factors are than to be compared with the actual factor to check the significance of the simulated factor. Factors affecting chlorine decay as chlorine added to water, pH, flow, turbidity, and temperature, generated using MC Method will be used next for calibration of chlorine decay in the rest of the point in Kohoutovice pressure zone using the computational software EPANET 2.0 and Artificial Neural Networks technique.

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 will be 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. MC method can provide information about sampling distributions when exact theory for the sampling distribution is not available.

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. Physicochemical parameters lead a continuous probability distribution and normal distribution can be used for each of the parameters to generate the database. **Table 8** shows basic statistics of the parameters, used to generate the random numbers in a given probability distribution for each input and output parameter.

The objective was to create a big 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. As an example for the procedure implemented 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; it must be determined 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. The same was done for each parameter estimated in this study, affecting chlorine decay.

**Table 8:** 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

#### 4.2.5 Creation of the Neural Network Model

The type of ANN model used was the well known Multilayer Perceptron (MLP). The MLP is a feed forward ANN model that maps the sets of input data onto a set of appropriate output. A MLP consist of multiple layers of nodes in a directed graph, which is fully connected from one layer to the next. Each node in the hidden layer and the output layer uses a nonlinear activation function. 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. Values

resulting from the MC simulation of each parameter are gathered in **Appendix 1** marked in bold. The Monte Carlo calculation uncertainty was kept below 2%. **Table 9** shows the statistics of the MC calculation for each parameter simulated.

**Table 9:** 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 to 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 10** 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 10**.

**Table 10:** 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							
<b>Total</b>	<b>540</b>	<b>300</b>	<b>300</b>	<b>540</b>	<b>360</b>	<b>540</b>	<b>360</b>

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 11** (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.

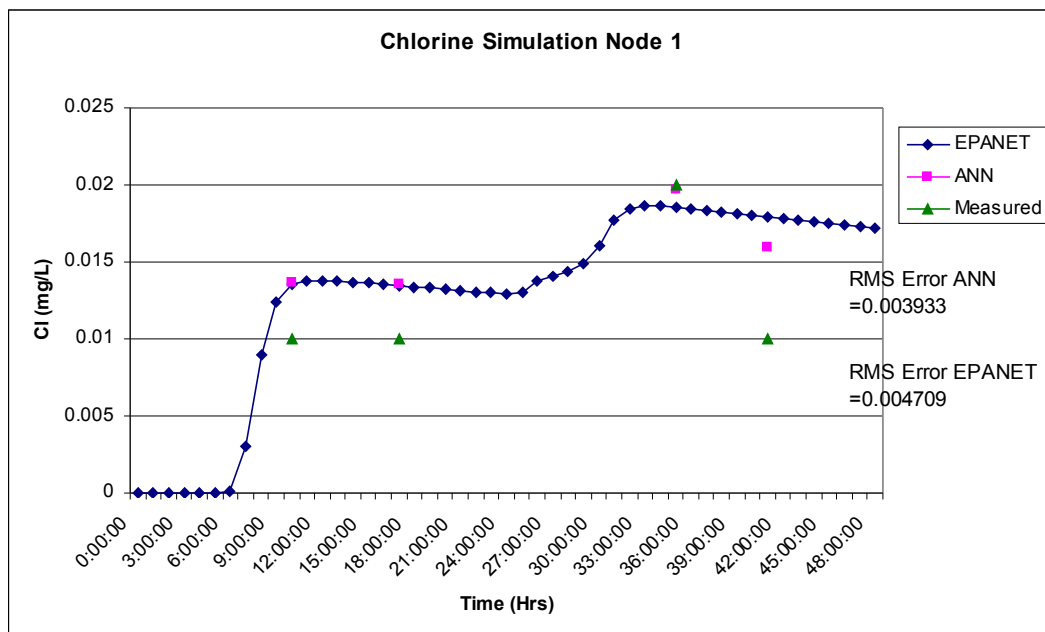
**Table 11:** Model Performance of each phase of the seven selected parameters

	Subset	Model Type	Training perf.	Testing pef.	Validation perf.
<b>Node 1</b>	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
<b>Node 2</b>	4	MLP 9-8-1	0.9778	0.4072	0.7034
	5	MLP 6-11-1	0.9876	0.4219	0.6672
<b>Node 3</b>	6	MLP 9-9-1	0.9990	0.5338	0.5387
	7	MLP 6-10-1	0.9841	0.4026	0.7811

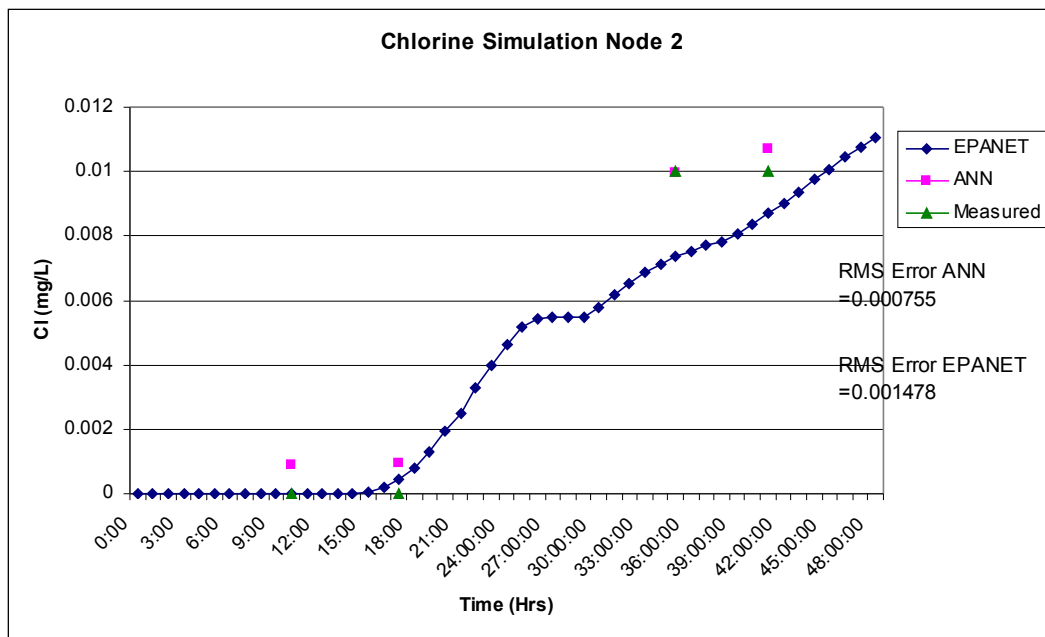
The data presented in **Table 11** shows that the first three networks had attained values for Training, Testing and validation performance close to 1. The networks with architectures listed as subset 1, 2 and 3 are those that had, moreover, attained the lowest values for the mean squared error. A combination 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 10** 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.

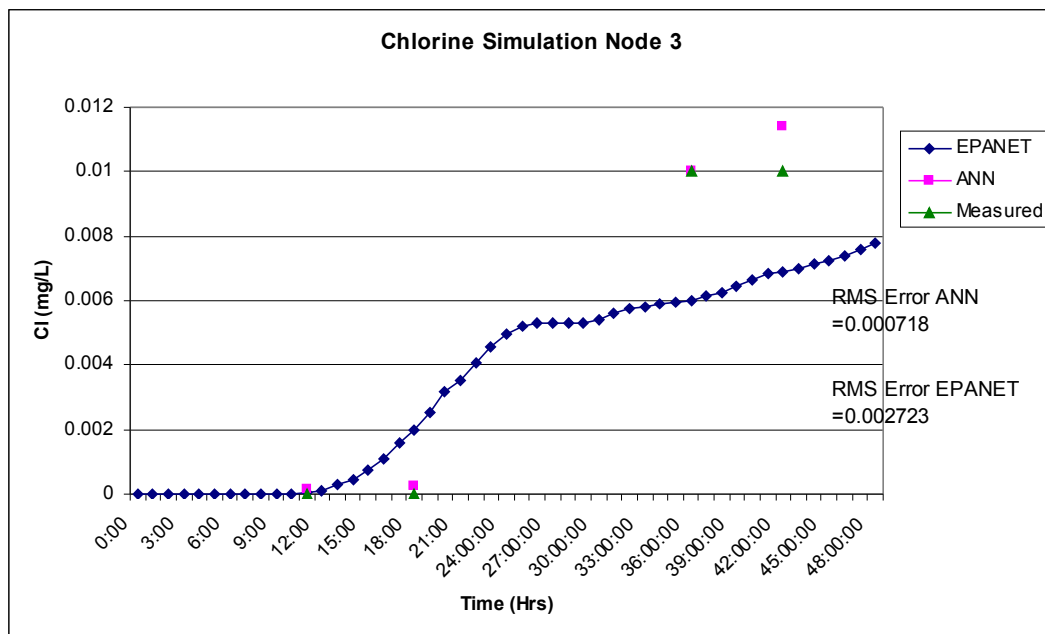
Particularly using the *Custom Predictions* tool in Statistica 10 from Statsoft it is possible to predict chlorine in Nodes 1, 2 and 3 with the known conditions of input parameters introducing the values to the ANN model proposed. The data obtained from the ANN can be compared with the results from the EPANET and used as calibration in the three nodes and at the same time providing more reliable results of the final model. **Figures 23, 24 and 25** 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 23:** Simulation in node 1 – ANN and EPANET



**Figure 24:** Simulation in node 2 – ANN and EPANET



**Figure 25:** 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.
- The proposed methodology for the creation of ANN models can be used for predicting chlorine decay in different pressure zones of any water distribution system.
- 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.
- The model showed that some nodes in Kohoutovice pressure zone stayed without residual chlorine concentration, several hours, during the simulated period of time of 48 hrs. See **Appendix 3**.
- The sensitivity analysis (See **Appendix 4**) showed that the water temperature entering the system and in the referenced node have a high influence in chlorine decay in the water supply system under consideration. To achieve the ideal calibration the model must take into account the evolution of residual free chlorine concentration for different temperatures at different times.

# 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 different sources and the 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 where 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 of the 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 were 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*.

### **5.2 Main Findings of the Thesis**

The ANN architecture used in this thesis was the Multi-Layer Perceptron (MLP) using backpropagation algorithm and has been evaluated to check the model performance and the least squared error to be compared with other water-quality techniques.

Main findings of this thesis:

#### **[Finding 1]**

Continual distributions are used to quantify, evaluate and fit these parameters. Basic statistics such as mean, standard deviation and confidence intervals were used for evaluation of these parameters calculated with Monte

Carlo method. 3000 readings were generated for Brno, Kohoutovice pressure zone case study. The use of the Monte Carlo calculations in combination with artificial neural network have been proven to be a powerful tool to perform chlorine residual prediction in three nodes inside the Kohoutovice pressure zone.

#### **[Finding 2]**

Several ANN topologies have responded accurately to the values of the training data set and also to testing and validation data set of value calculated using the MC method. Network topologies for Brno, Kohoutovice pressure zone (Subset 1) can be fully used to predict the values of chlorine residual concentration in Node 1. Results show that the problem dealt within this paper is indeed very complex because chlorine decay is depending on several parameters such as temperature, pH, turbidity between others.

#### **[Finding 3]**

For training the ANN in Brno, Kohoutovice pressure zone, the performance was very accurate, even though it could not be achieved the same results in Subsets 4 to 7 in testing and validation phase, as already written in the evaluation of the data obtained section. Therefore, the training data set should be chosen to best represent the expected chlorine residual in the selected Nodes inside the pressure zone Kohoutovice. The present findings suggest that:

- ANNs is capable to predict free chlorine at Kohoutovice pressure zone using historical data and data generated by MC method.
- The key parameters Initial chlorine, flow and temperature have the most influence in the chlorine decay prediction in the pressure zone 1.3.1 Kohoutovice.
- Some nodes inside Kohoutovice pressure zone stay without residual chlorine for few hours during the day. See *Appendix 3*.

Recommendations for the use of this model are following:

1. The present model can only be used for chlorine decay prediction in Kohoutovice pressure zone.
2. Try to create a complete database based on the measurements of the parameters affecting chlorine decay in the same places inside the pressure zone.



3. Create a large database which include parameters from all the season and study different model subset for each season.
4. It is shown in the sensitivity analysis in appendix 3 that temperature from both sides (Initial and local conditions) have a high influence in chlorine decay. Ideal calibration should take into account the behavior of chlorine at different temperature values in different season.
5. This methodology can be use to assess chlorine decay behavior in different pressure zones of any water distribution system, following the procedure shown in **Figure 4**.
6. Monte Carlo simulation can be used to generate missing values of the original data set. This is a new approach proposed in this research that can be used as it provide flexibility, manage the uncertainty and even provide more accurate results that simple descriptive statistics.

### 5.3 Known Limitations of the Research

The following assumptions and simplifications should be noted because they can limit usability of the present research:

- A withdraw of ANN in general is regarding the slow performance when working in a very large system (referring to big distribution networks) but one way to avoid this situation is dividing the network into small zones (pressure zones) and analyze one by one.
- One of the limitation of the research regards to the lack of data or information from the water utilities. Although it is collected data in some nodes inside the network and in the main components of the system such as reservoirs, tanks etc. it exist a need to have more laboratory data such a pH, turbidity and more chemical parameters that could influence chlorine decay.
- In *Chapter 3* there was mention several types of ANN architectures. In the experimental section in *Chapter 4* the only architecture used was Multi-Layer Perceptron (MLP) but other type of architectures such as GRNN or SOFM were not studied to check the performance. SOFM give the possibility to predict when the objective parameter (chlorine concentration) is not measure very often
- As a recommendation for water utilities the collection of measure data and storage in a database is important for development of such models or the continual improvement of the current one.
- Depending on the amount of data, ANN such as MLP using backpropagation (gradient-descendant) algorithm can be very slow in the performance. Another annotation for future work can be the creation

of ANN using another type of algorithm, can be a hybrid algorithm or a Genetic Algorithm to train the network, as it provides better results.

### **5.4 Future Research Directions and Practical Use**

This thesis represents the research I have conducted and I believe that the research brings new ideas to Artificial Neural Network modeling in water quality for water distribution systems as well for the standard approach currently used by the water utilities in the Czech Republic and world wide. On the other hand, the chosen research topic is very complex, and spans several research fields. This work, shall be therefore, considered only as a tiny aspect of the water-quality modeling. I assume that presented methodology can be used in applications that are mainly focus on the control of chlorine concentration in water supply systems and can be improve each time using the new collected data measured in the specific network.

I propose the following subjects for further research:

- Divide the network distribution system into specific pressure zones for a better and quickly performance of the ANN. In the same way divide the measurement for chlorine and other physic-chemical parameters in four season, as temperature is one of the parameters that influence the most the chlorine concentration in the system.
- An incrementation of the original database based in measurements in specific nodes inside the water supply system studied. The water utilities should focus in strategic nodes in the system and keep regular control on them and in the same meaning update the database to gather sufficient data to run the model again and maintain the model calibrated.
- When training, testing and validating the model, it can be proposed a different approach such as the use of new data measured for testing or specific data collected in a given season for validating. Also the separation of data can be done using another type of algorithm such as Genetic algorithms or even SOFM. Other than the approach used in the thesis that was based in dividing the data into percentage of 50, 25 and 25 of random values of the parameters for training, testing and validating respectively.
- As a final phase the creation of a software capable of predicting chlorine concentration using ANN techniques in a very simple form and with a graphical user interface in which only the initial parameters of the WDS will be needed and even can be managed to connect the new software with a SCADA system for a real time prediction.

## Bibliography

1. James Powell, Jonathan Clement Malcolm Brandt. Predictive Models for Water Quality in Distribution Systems, Black & Veatch, 69 London Road, Redhill, Surrey RH1 1LQ, UK: Awwa research Foundation, 2004. 105. ISBN 1 84339 913 X.
2. Gavin J. Bowden, John B. Nixon, Graeme C. Dandy, Holger R. Maier, Mike Holme, Forecasting chlorine residuals in a water distribution system using a general regression neural network, Adelaide, Australia: Mathematical and Computer Modeling, 11 February 2005. 16. ISBN 44 (2006) 469-484.
3. Srinivasa Lingireddy, Ph.D., P.E., Gail M. Brion, Ph.D., Artificial Neural Networks in Water Supply Engineering, Reston, Virginia, 20191-4400, American Society of Civil Engineers, 2005. ISBN 0-7844-0765-7.
4. Zhengfu Rao, Fernando Alvarruiz. Use of an artificial neural network to capture the domain knowledge of a conventional hydraulic simulation model, IWA Publishing, Journal of Hydroinformatics, 09.01.2007.
5. Herve Abdi, Neural Networks, The University of Texas at Dallas, 1998, ISBN 9780761914402.
6. James C. Powell, Nicholas B. Hallam, Factors which control bulk chlorine decay rates, Water Research Vol. 34 No 1, pp. 117-126, 1999 Elsevier Science Ltd. PII: S0043-1354(99)00097-4.
7. N.B. Hallam, J.R. West, C.F. Forster, The decay of chlorine associated with the pipe wall in water distribution systems, Water Research 36 (2002) 3479-3488, 2002 Elsevier Science Ltd. PII: S0043-1354(02)00056-8.
8. M.S. Gibbs, N. Morgan, H.R. Maier, G.C. Dandy, Investigation into the relationship between chlorine decay and water distribution parameters using data driven methods, Mathematical and Computer Modelling 44 (2006) 485-498.
9. Manuel J. Rodriguez, Jean-B Serodes, Assessing empirical linear and non-linear modelling of residual chlorine in urban drinking water systems, Environmental modelling & Software 14 (1999) 93-102.
10. Michael Damien Kazantzis, Strengths and Weakness of EPANET for Modelling Water Quality in Distribution Systems, Centre of Applied Modelling in Water Engineering, School of Civil and Environmental Engineering, The University of Adelaide, February 2002.

11. Raed Jafar, Isam Shahrour, Ilan Juran, Application of Artificial Neural Networks (ANN) to model the failure of urban water mains, Mathematical and Computer Modelling, Brooklyn, NY 11201, United States, 51 (2010) 1170-1180.
12. Robert J. May, Graem C. Dandy, Holger R. Maier, John B. Nixon, Application of partial mutual information variable selection to ANN forecasting of water quality in water distribution systems. Adelaide, SA 5005, Australia, Environmental Modelling & Software, 23 (2003) 10289-1299.
13. John J. Vasconcelos, Characterization and Modeling of Chlorine Decay in Distribution Systems, AWWA Research Foundation, 1996. 105. ISBN 0-89867-870-6
14. Lewis A. Rossman, Computer Models/Epanet, Cincinnati, OH, United states, U.S. Environmental Protection Agency.
15. Pedro Castro, Mário Neves, Chlorine Decay in Water Distribution Systems Case Study – Lousada Network, University of Porto, Portugal, Electron. J. Environ. Agric. Food Chem, ISSN 1579-4377.
16. J. Izquierdo, R. Perez And P. L. Iglesias, Mathematical Models and Methods in the Water Industry, Mathematical and Computer Modelling 39 (2004) 1353-1374.
17. Manuel Herrera, Luis Torgo, Joaquin Izquierdo, Rafael Perez-Garcia, Predictive models for forecasting hourly urban water demand, 46022 Valencia, Spain, Journal of Hydrology 387 (2010) 141–150
18. Jinhui Jeanne Huang, Edward Arthur McBean, Using Bayesian statistics to estimate the coefficients of a two- component second-order chlorine bulk decay model for a water distribution system, Ontario, Canada N1G 2W1, Water Research 41 (2007) 287 – 294.
19. Lihui Ma, Kunlun Xin, and Suiqing Liu, Using Radial Basis Function Neural Networks to Calibrate Water Quality Model, World Academy of Science, Engineering and Technology 38 2008
20. Baggett C, Guohua Li, Rosario R. From Start to Finish: Calibrating Pinellas County's 2,000-Mile Hydraulic Water Distribution System Model. Florida Water Resources Journal. 2008. 44 – 54.
21. Clark R, Grayman W. Modeling water quality in drinking water distribution systems. American Water Works Association (Denver, CO). 1998. ISBN 0898679729
22. Slavičková K. Změny kvality vody při dopravě a jejich modelování v programu Epanet 2. Voda Zlín 2002. ISBN 80-238-8326-7.

23. Vojtová I. Vírský oblastní vodovod a jeho funkce v Brněnské vodárenské soustavě (BVS). Masarykova univerzita, 2010.
24. Starý, Nacházel, Zezulák. Využití metod umělé inteligence ve vodním hospodářství. Academia. ISBN 80-200-0229-4. 2004
25. Larry W. Mays. Water distribution System Handbook. McGraw-Hill. ISBN 9780071342131. 2000.
26. Bull, R. J.; Birnbaum, L. S.; Cantor, K. P.; Rose, J. B.; Butterworth, B. E, Pegram, R.; Tuomisto, J. Water chlorination: essential process or cancer hazard? Fund. Appl. Toxicol. 1995, 28, 155–166.
27. Ahmad Jabari Kohpaei, Arumugam Sathasivan, Chlorine decay prediction in bulk water using the parallel second order model: An analytical solution development, Department of Civil Engineering and Construction, Curtin University, Chemical Engineering Journal 171 (2011) 232–241
28. Nieuwenhuijsen, M. J.; Toledano, M. B.; Eaton, N. E.; Fawell, J.; Elliott, P. Chlorination disinfection byproducts in water and their association with adverse reproductive outcomes: a review. Occup. Environ. Med. 2000, 57, 73–85.
29. R.M. Clark, M. Sivaganesan, Predicting chlorine residuals in drinking water: second order model, Journal of Water Resources Planning and Management 128 (2002) 152–161.
30. Boccelli, D. L.; Tryby, M. E.; Uber, J. G.; Rossman, L. A.; Zierolf, M. L.; Polycarpou, M. M. Optimal Scheduling of Booster Disinfection in Water Distribution Systems. J. Water Res. Pl. 1998, 124, 99–111.
31. I. Fisher, A. Jabari Kohpaei, A. Sathasivan, Comment on “Using Bayesian statistics to estimate the coefficients of a two-component second-order chlorine bulk decay model for a water distribution system by Huang J.J., McBean E.A. Water Res., 2007.”, Water Research 44 (2010) 3309–3310.
32. Tryby, M. E.; Boccelli, D. L.; Uber, J. G.; Rossman, L. A. Facility Location Model for Booster Disinfection of Water Supply Networks. J. Water Res. Pl. 2002, 128, 322–333.
33. D. Gang, T.E. Clevenger, S.K. Banerji, Modeling chlorine decay in surface water, Journal of Environmental Informatics 1 (2003) 21–27.
34. P. Jonkergouw, S.T. Khu, D. Savic, D. Zhong, X. Hou, H.B. Zhao, A variable rate coefficient chlorine decay model, Environmental Science and Technology 43 (2009) 408–414.
35. V. Jegatheesan, S.H. Kim, C.K. Joo, Evaluating the drinking water quality through an efficient chlorine decay model, Water Science and Technology, Water Supply 6 (2006) 1–7.

36. G. Kastl, I. Fisher, V. Jegatheesan, Evaluation of chlorine decay kinetics expressions for drinking water distribution systems modeling, *Journal of Water Supply: Research and Technology-AQUA* 48 (1999) 219–226.
37. Online: <http://www.statsoft.com/textbook>, Website of StatSoft, Inc. the STATISTICA Data Analysis Software and Services. Electronic Statistics Textbook.
38. Prasad, T. D.; Walters, G. A.; Savic, D. A. Booster Disinfection of Water Supply Networks: Multiobjective Approach. *J. Water Res. Pl.* 2004, 130, 367–376.
39. Ian Fisher a\*, George Kastl b, Arumugam Sathasivan, Evaluation of suitable chlorine bulk-decay models for water distribution systems, Department of Civil Engineering and Construction, Curtin University of Technology, *water research* 45 (2011) 4896- 4908
40. Propato, M.; Uber, J. G. Linear least-squares formulation for operation of booster disinfection systems. *J. Water Res. Pl.* 2004, 130, 53–62.
41. Shang, F., Uber, J., Rossman, L., 2008. Modeling reaction and transport of multiple species in water distribution systems. *Environmental Science and Technology* 42 (3), 808-814.
42. Cozzolino, L.; Pianese, D.; Pirozzi, F. Control of DBPs in water distribution systems through optimal chlorine dosage and disinfection station allocation. *Desalination* 2005, 176, 113–125.
43. Biswas, P.; Lu, C.; Clark, R. M. A model for chlorine concentration decay in pipes. *Wat. Res.* 1993, 27, 1715–1724.
44. Clark, R. M.; Haught, R. C. Characterizing pipe wall demand: Implications for water quality modeling. *J. Water Res. Pl.* 2005, 131, 208–217.
45. Zhang, G. R.; Kiene, L.; Wable, O.; Chan, U. S.; Duget, J. P. Modeling of chlorine residual in the water distribution system network of Macao. *Environ. Technol.* 1992, 13, 937–946.
46. Dharmarajah, H.; Patania, N. L.; Jacangelo, J. G.; Aieta, E. M.; Montgomery, J. M. Empirical modeling of chlorine and chloramine residual decay. In *Proceedings AWWA Annual Conference*. 1991, Philadelphia, USA.
47. Hallam, N. B.; Hua, F.; West, J. R.; Forster, C. F.; Simms, J. Bulk decay of chlorine in water distribution systems. *J. Water Res. Pl.* 2003, 129, 78–81.
48. Boccelli, D. L.; Tryby, M. E.; Uber, J. G.; Summers, R. S. A reactive species model for chlorine decay and THM formation under rechlorination conditions. *Wat. Res.* 2003, 37, 2654–2666.
49. Hua, F.; West, J. R.; Barker, R. A.; Forster, C. F. Modelling of chlorine decay in municipal water supplies. *Wat. Res.* 1999, 33, 2735–2746.

50. Vasconcelos, J. J.; Rossman, L. A.; Grayman, W. M.; Boulos, P. F.; Clark, R. M. Kinetics of chlorine decay. *J. Am. Water Works Ass.* 1997, 89, 54–65.
51. Munavalli, G. R.; Mohan Kumar, M. S. Water quality parameter estimation in a distribution system under dynamic state. *Wat. Res.* 2005, 39, 4287–4298.
52. Jonkergouw, P. M. R.; Khu, S.-T.; Kapelan, Z. S.; Savić, D. A. Water quality model calibration under unknown demands. *J. Water Res. Pl.* 2008, in press.
53. Ostfeld Avi, Enhancing Water-Distribution System Security through Modeling, American Society of Civil Engineers, Technion. ISSN: 1943-5452.
54. Opher T., Rom M., Kronaveter L., Friedler E., and Ostfeld A. (2011), Comparison of model trees, artificial neural networks, and statistical diagnostic models for pipeline drinking water biofouling prediction." *Urban Water*, DOI:10.1080/1573062X.2011.633611
55. Preis A. and Ostfeld A. (2008). "Multiobjective contaminant response modeling for water distribution systems security." *Journal of Hydroinformatics*, Vol. 10, No. 4, pp. 267-274.
56. Perelman L. and Ostfeld A. (2008). "Water distribution system aggregation for water quality analysis", *Journal of Water Resources Planning and Management Division*, ASCE, Vol. 134, No. 3, pp. 303-309.
57. Ručka, J.; Tuhovčák, L., Využití neuronových sítí pro predikci poruchovosti vodovodních sítí, *Juniorstav 2005*, ISBN 80-214-2830-9, VUT v Brně, Brno, 2005
58. Tuhovčák, L.; Ručka, J., Hazard identification and risk analysis of water supply systems, kapitola v *Strategic Asset Management of Water Supply and Wastewater Infrastructures*, ISBN 1-84339-186-4, IWA Publishing, London, UK, 2009
59. Tuhovčák, L., Ručka, J., Svoboda, M., Riziková analýza systémů zásobování pitnou vodou, 4. vodohospodářská konference 2004, ISBN 80-7204-360-9, Ústav vodních staveb, FAST, VUT v Brně, Brno, 2004
60. 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.
61. 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.
62. 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.
63. Ian Flood, Nabil Kartam, Artificial Neural Networks for Civil Engineers: Advanced Features and Applications. American Society of Civil Engineers and Artificial Intelligence Committee. ISBN: 0-7844-0344-9. 1998.
64. Huang J, McBean E. Using Bayesian Statistics to Estimate Chlorine Wall Decay Coefficients for Water Supply System. Journal of Water Resources, Planning and Management. ASCE. March 2008. DOI: 0733-9496(2008) 134.

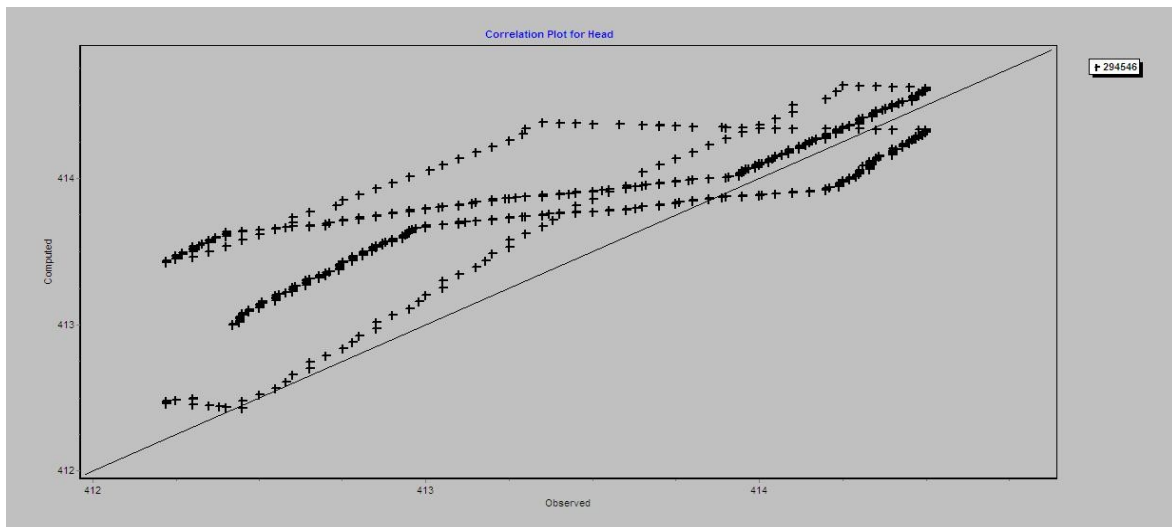


## APPENDIX 1: Database of 18 parameters used for ANN simulation and statistics in Brno-Kohoutovice, Case study

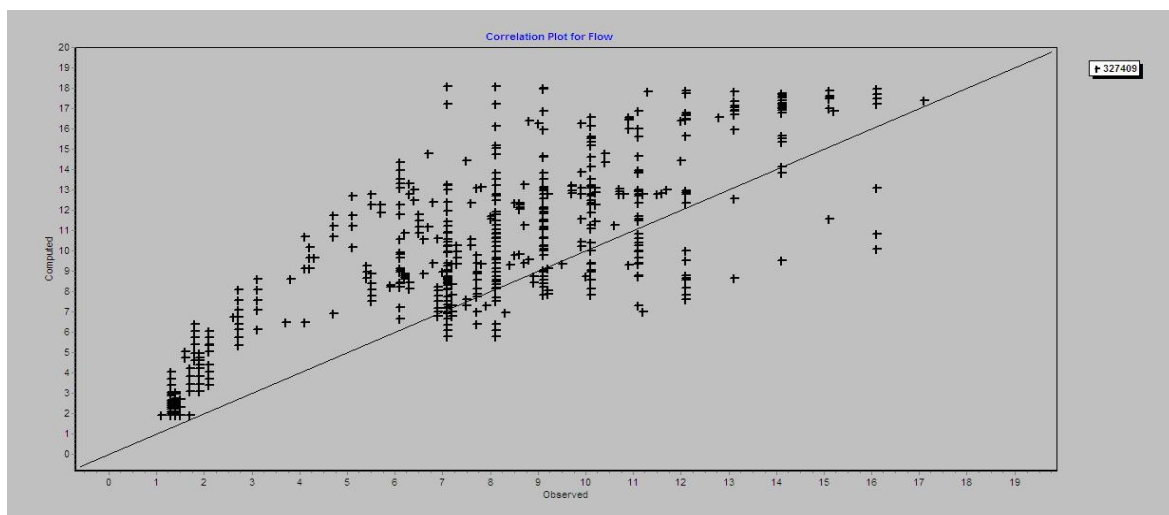
**A. Table 1.1:** Raw Data and Data simulated using MC Method (MC simulated in Bold)

Date	Initial Chlorine Bosonohy Tank	Free Chlorine Kohoutovice Tank	pH	Temperature	Flow Kohoutovice Tank	Turbidity	Free Chlorine Junction 290172 (Libusina tr.4)	pH Junction 290172 (Libusina tr.4)	Temperature Junction 290172 (Libusina tr.4)	Turbidity Junction 290172 (Libusina tr.4)	Free Chlorine Junction 291750 Libusino udoli 66 - 58 - 160	pH Junction 291750 Libusino udoli 66 - 58 - 160	Temperature Junction 291750 Libusino udoli 66 - 58 - 160	Turbidity Junction 291750 Libusino udoli 66 - 58 - 160	Free Chlorine Junction 294526 Nad Pisárkami 2	pH Junction 294526 Nad Pisárkami 2	Temperature Junction 294526 Nad Pisárkami 2	Turbidity Junction 294526 Nad Pisárkami 2
9.1.2007	0.04	0.01	7.86	10.1	10.20	1.19	0.03	7.84	17.00	1.21	0.01	7.91	14.00	2.56	<b>0.01</b>	<b>7.65</b>	<b>11.23</b>	<b>0.69</b>
26.2.2007	0.03	0.01	7.23	10.1	9.10	0.41	0.02	<b>7.74</b>	<b>18.46</b>	<b>1.38</b>	0.02	7.59	17.20	4.28	0.01	7.55	12.60	0.76
6.3.2007	0.03	0.02	7.28	10.5	9.10	0.67	0.02	7.37	14.90	0.38	0.01	7.31	9.50	0.46	<b>0.01</b>	<b>7.65</b>	<b>11.25</b>	<b>0.69</b>
2.4.2007	0.04	0.01	7.64	11.3	9.10	0.67	0.04	<b>7.73</b>	<b>18.33</b>	<b>1.39</b>	<b>0.01</b>	<b>7.60</b>	<b>18.34</b>	<b>0.97</b>	<b>0.01</b>	<b>7.65</b>	<b>11.27</b>	<b>0.70</b>
2.5.2007	0.02	0.01	7.50	12.4	8.50	2.39	0.01	7.75	22.00	4.57	0.01	7.49	15.00	1.87	<b>0.01</b>	<b>7.65</b>	<b>11.26</b>	<b>0.68</b>
5.6.2007	0.02	0.02	7.58	11.6	8.10	2.11	0.01	<b>7.73</b>	<b>18.34</b>	<b>1.41</b>	<b>0.01</b>	<b>7.60</b>	<b>18.45</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.19</b>	<b>0.69</b>
2.7.2007	0.02	0.02	7.63	14.7	8.10	1.99	0.01	7.62	18.50	2.63	0.01	7.56	18.30	1.04	<b>0.01</b>	<b>7.65</b>	<b>11.26</b>	<b>0.69</b>
1.8.2007	0.02	0.02	7.48	14.7	8.10	0.46	0.01	<b>7.73</b>	<b>18.33</b>	<b>1.34</b>	<b>0.01</b>	<b>7.60</b>	<b>18.40</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.20</b>	<b>0.68</b>
17.9.2007	0.02	0.01	7.54	12.4	8.10	0.68	0.01	7.67	22.10	1.62	0.01	7.64	17.70	2.59	<b>0.01</b>	<b>7.65</b>	<b>11.29</b>	<b>0.68</b>
15.10.2007	0.03	0.02	7.54	11.5	6.30	2.58	0.01	<b>7.73</b>	<b>18.32</b>	<b>1.40</b>	<b>0.01</b>	<b>7.60</b>	<b>18.45</b>	<b>0.96</b>	<b>0.01</b>	<b>7.65</b>	<b>11.23</b>	<b>0.69</b>
6.11.2007	0.03	0.01	7.49	10.2	7.10	1.69	0.02	7.57	13.50	0.53	0.02	7.49	13.70	0.78	<b>0.01</b>	<b>7.65</b>	<b>11.33</b>	<b>0.69</b>
7.1.2008	0.02	0.02	7.48	7.3	8.10	0.60	0.01	7.47	11.40	2.98	0.01	7.47	10.40	0.90	<b>0.01</b>	<b>7.65</b>	<b>11.27</b>	<b>0.69</b>
5.2.2008	0.04	0.03	7.50	8.8	10.10	1.36	0.01	<b>7.72</b>	<b>18.39</b>	<b>1.39</b>	<b>0.01</b>	<b>7.60</b>	<b>18.38</b>	<b>0.98</b>	<b>0.01</b>	<b>7.65</b>	<b>11.21</b>	<b>0.69</b>
5.3.2008	0.03	0.02	7.45	8.4	9.10	1.36	0.01	7.60	17.30	2.40	0.01	7.45	10.30	1.27	<b>0.01</b>	<b>7.65</b>	<b>11.30</b>	<b>0.69</b>
2.4.2008	0.02	0.01	7.68	9.8	10.20	4.74	0.01	<b>7.74</b>	<b>18.37</b>	<b>1.41</b>	0.01	7.74	9.80	0.85	0.01	7.67	7.70	0.59
6.5.2008	0.03	0.02	7.40	11.5	9.10	1.07	0.02	7.51	19.50	1.51	0.02	7.46	15.60	0.65	<b>0.01</b>	<b>7.65</b>	<b>11.19</b>	<b>0.69</b>
2.6.2008	0.02	0.01	7.41	13.3	11.10	0.19	0.01	<b>7.73</b>	<b>18.44</b>	<b>1.34</b>	<b>0.01</b>	<b>7.60</b>	<b>18.34</b>	<b>0.97</b>	<b>0.01</b>	<b>7.65</b>	<b>11.24</b>	<b>0.69</b>
8.7.2008	0.02	0.01	7.44	14.1	12.10	1.5	0.01	7.67	19.10	0.90	0.01	7.41	20.10	1.10	<b>0.01</b>	<b>7.65</b>	<b>11.24</b>	<b>0.68</b>
4.8.2008	0.02	0.01	7.45	14.3	11.10	0.5	0.01	<b>7.73</b>	<b>18.29</b>	<b>1.36</b>	<b>0.01</b>	<b>7.60</b>	<b>18.30</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.68</b>
1.9.2008	0.02	0.01	7.50	14.4	12.10	0.5	0.01	7.75	23.30	0.80	0.01	7.52	19.70	1.40	<b>0.01</b>	<b>7.65</b>	<b>11.30</b>	<b>0.70</b>
6.10.2008	0.03	0.01	7.69	12.4	11.60	0.9	0.02	<b>7.73</b>	<b>18.36</b>	<b>1.36</b>	<b>0.01</b>	<b>7.60</b>	<b>18.29</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.69</b>
24.11.2008	0.03	0.02	7.58	9.3	11.10	3.9	0.02	7.47	15.70	1.80	0.01	7.48	16.00	0.90	<b>0.01</b>	<b>7.65</b>	<b>11.21</b>	<b>0.70</b>
8.12.2008	0.03	0.03	7.40	9.5	11.10	2.4	0.01	7.57	13.20	1.90	0.01	7.63	12.40	0.80	<b>0.01</b>	<b>7.65</b>	<b>11.26</b>	<b>0.68</b>
6.1.2009	0.03	0.02	7.55	7.4	12.10	0.7	0.01	7.63	18.20	1.10	0.01	7.67	15.30	1.30	<b>0.01</b>	<b>7.65</b>	<b>11.24</b>	<b>0.68</b>
4.2.2009	0.03	0.02	7.62	8.9	12.10	1.3	0.01	<b>7.74</b>	<b>18.33</b>	<b>1.34</b>	<b>0.01</b>	<b>7.60</b>	<b>18.32</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.19</b>	<b>0.70</b>
3.3.2009	0.03	0.02	7.51	9.4	10.10	0.9	0.01	7.79	18.30	4.70	<b>0.01</b>	<b>7.60</b>	<b>18.30</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.31</b>	<b>0.69</b>
14.4.2009	0.03	0.02	7.65	9.2	9.10	0.4	0.01	<b>7.73</b>	<b>18.32</b>	<b>1.40</b>	<b>0.01</b>	<b>7.60</b>	<b>18.33</b>	<b>0.97</b>	<b>0.01</b>	<b>7.65</b>	<b>11.27</b>	<b>0.69</b>
4.5.2009	0.02	0.01	7.54	13.0	8.10	0.5	0.01	7.77	21.80	1.50	<b>0.01</b>	<b>7.60</b>	<b>18.26</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.23</b>	<b>0.70</b>
9.6.2009	0.02	0.02	7.50	14.0	7.80	0.6	0.01	<b>7.73</b>	<b>18.46</b>	<b>1.34</b>	0.01	7.44	16.70	0.50	0.01	7.65	13.90	0.60
13.7.2009	0.02	0.01	7.61	13.7	9.20	2.6	0.01	7.78	19.90	1.70	0.01	7.50	18.80	1.10	<b>0.01</b>	<b>7.65</b>	<b>11.21</b>	<b>0.69</b>
18.8.2009	0.02	0.01	7.64	14.2	10.10	1.0	0.01	<b>7.73</b>	<b>18.29</b>	<b>1.41</b>	<b>0.01</b>	<b>7.60</b>	<b>18.32</b>	<b>0.96</b>	<b>0.01</b>	<b>7.65</b>	<b>11.32</b>	<b>0.69</b>
1.9.2009	0.03	0.02	7.65	14.9	11.10	0.5	0.01	7.86	23.00	1.30	0.01	7.57	20.10	0.50	<b>0.01</b>	<b>7.65</b>	<b>11.26</b>	<b>0.68</b>
14.10.2009	0.02	0.01	7.55	11.0	10.10	2.3	0.01	<b>7.74</b>	<b>18.36</b>	<b>1.36</b>	<b>0.01</b>	<b>7.59</b>	<b>18.50</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.31</b>	<b>0.70</b>
2.11.2009	0.02	0.01	7.84	10.3	9.10	1.5	0.01	8.06	21.20	1.50	0.01	7.84	14.10	0.80	<b>0.01</b>	<b>7.65</b>	<b>11.18</b>	<b>0.70</b>
2.12.2009	0.02	0.01	7.80	11.4	9.10	0.6	0.01	7.94	19.60	1.00	0.01	7.84	12.40	0.40	<b>0.01</b>	<b>7.65</b>	<b>11.32</b>	<b>0.69</b>
5.1.2010	0.03	0.03	7.71	9.1	9.20	0.2	0.01	7.72	20.20	0.40	0.01	7.61	14.50	0.20	<b>0.01</b>	<b>7.65</b>	<b>11.29</b>	<b>0.68</b>
8.2.2010	0.02	0.01	7.55	8.4	9.20	1.1	0.01	<b>7.74</b>	<b>18.33</b>	<b>1.38</b>	0.01	7.57	7.30	0.30	0.01	7.61	11.50	0.50
22.3.2010	0.03	0.02	7.73	11.6	8.10	1.2	0.02	7.56	19.00	0.70	<b>0.01</b>	<b>7.60</b>	<b>18.24</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.26</b>	<b>0.68</b>
6.4.2010	0.03	0.01	7.53	9.5	8.10	0.3	0.02	<b>7.73</b>	<b>18.28</b>	<b>1.38</b>	<b>0.01</b>	<b>7.60</b>	<b>18.39</b>	<b>0.94</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.69</b>
3.5.2010	0.02	0.01	7.58	9.4	7.10	0.1	0.01	7.72	21.50	0.90	0.01	7.68	17.50	0.40	<b>0.01</b>	<b>7.65</b>	<b>11.23</b>	<b>0.70</b>
2.6.2010	0.02	0.01	7.53	10.6	14.10	2.5	0.01	<b>7.73</b>	<b>18.39</b>	<b>1.36</b>	<b>0.01</b>	<b>7.60</b>	<b>18.38</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.27</b>	<b>0.68</b>
19.7.2010	0.03	0.02	7.77	12.3	14.10	0.1	0.01	7.80	19.60	0.80	0.01	7.81	21.20	0.20	<b>0.01</b>	<b>7.65</b>	<b>11.19</b>	<b>0.69</b>
31.8.2010	0.02	0.01	7.57	14.6	13.10	0.8	0.01	<b>7.73</b>	<b>18.40</b>	<b>1.35</b>	<b>0.01</b>	<b>7.60</b>	<b>18.40</b>	<b>0.95</b>	<b>0.01</b>	<b>7.65</b>	<b>11.20</b>	<b>0.69</b>
15.9.2010	0.02	0.01	7.40	13.3	9.90	0.2	0.01	7.66	22.00	0.30	0.01	7.49	18.60	0.30	<b>0.01</b>	<b>7.65</b>	<b>11.28</b>	<b>0.69</b>
6.10.2010	0.03	0.02	7.55	15.0	10.10	0.6	0.01	7.69	15.60	0.40	0.01	7.62	20.50	0.50	<b>0.01</b>	<b>7.65</b>	<b>11.24</b>	<b>0.69</b>
2.11.2010	0.02	0.01	7.63	12.3	11.10	0.7	0.01	7.81	19.00	0.80	0.01	7.76	12.80	0.30	<b>0.01</b>	<b>7.65</b>	<b>11.21</b>	<b>0.69</b>
7.12.2010	0.03	0.02	7.53	10.2	10.10	0.3	0.01	7.80	19.60	0.80	<b>0.01</b>	<b>7.59</b>	<b>18.28</b>	<b>0.96</b>	<b>0.01</b>	<b>7.65</b>	<b>11.21</b>	<b>0.69</b>
12.1.2011	0.04	0.03	7.57	9.7	10.10	4.3	0.03	7.71	15.20	4.70	<b>0.01</b>	<b>7.60</b>	<b>18.41</b>	<b>0.96</b>	0.03	7.78	10.70	1.00
2.3.2011	0.04	0.03	7.43	9.0	10.10	0.4	0.03	7.52	12.60	0.30	<b>0.01</b>	<b>7.60</b>	<b>18.34</b>	<b>0.92</b>	<b>0.01</b>	<b>7.65</b>	<b>11.29</b>	<b>0.68</b>
6.4.2011	0.04	0.03	7.81	12.1	11.10	0.7	0.03	<b>7.73</b>	<b>18.41</b>	<b>1.34</b>	<b>0.01</b>	<b>7.60</b>	<b>18.36</b>	<b>0.96</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.70</b>
11.5.2011	0.04	0.03	7.82	11.2	16.10	2.5	0.03	7.95	19.30	0.50	<b>0.01</b>	<b>7.60</b>	<b>18.51</b>	<b>0.98</b>	<b>0.01</b>	<b>7.65</b>	<b>11.28</b>	<b>0.68</b>
13.6.2011	0.04	0.03	7.49	13.6	12.10	2.1	0.03	<b>7.73</b>	<b>18.29</b>	<b>1.40</b>	<b>0.01</b>	<b>7.60</b>	<b>18.37</b>	<b>0.98</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.69</b>
13.7.2011	0.04	0.03	7.80	13.4	15.10	1.1	0.03	7.94	20.30	0.80	<b>0.01</b>	<b>7.60</b>	<b>18.36</b>	<b>0.98</b>	<b>0.01</b>	<b>7.65</b>	<b>11.22</b>	<b>0.69</b>
29.8.2011	0.04	0.03	7.85	14.3	16.10	0.2	0.03	<b>7.73</b>	<b>18.40</b>	<b>1.39</b>	<b>0.01</b>	<b>7.59</b>	<b>18.32</b>	<b>0.97</b>	<b>0.01</b>	<b>7.65</b>	<b>11.24</b>	<b>0.69</b>
13.9.2011	0.04	0.03	7.92	13.5	16.10	0.2	0.03	7.94	20.20	0.30	<b>0.01</b>	<b>7.59</b>	<b>18.42</b>	<b>0.96</b>	<b>0.01</b>	<b>7.65</b>	<b>11.27</b>	<b>0.69</b>
25.10.2011	0.04	0.03	8.13	12.0	10.90	1.5	0.03	<b>7.73</b>	<b>18.44</b>	<b>1.35</b>	<b>0.01</b>	<b>7.60</b>	<b>18.42</b>	<b>0.97</b>	<b>0.01</b>	<b>7.65</b>	<b>11.23</b>	<b>0.69</b>
7.11.2011	0.05	0.04	7.59	12.0	12.10	0.3												

## APPENDIX 2: Calibration Reports and Statistics



**A. Figure 2.1:** Calibration report of the Correlation Plot for Head values at Tank Kohoutovice

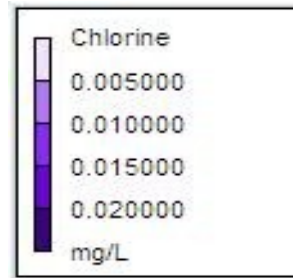


**A. Figure 2.2:** Calibration report of the Correlation Plot for Flow values at Link 327409

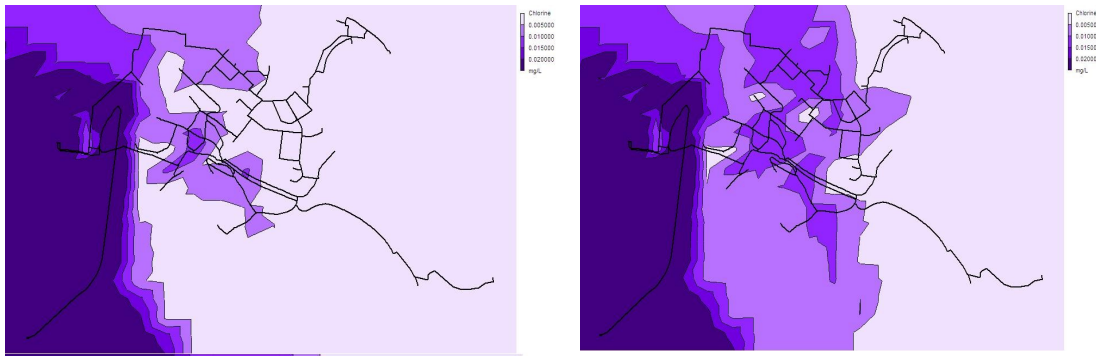
**A. Table 2.1:** Calibration Statistics for Head at Tank Kohoutovice and Flow at Link 327409

<b>Location</b>	<b>Num Obs</b>	<b>Observed Mean</b>	<b>Computed Mean</b>	<b>Mean Error</b>	<b>RMS Error</b>
Tank 294546	575	413.52	413.86	0.417	0.54
Link 327409	575	7.79	10.17	2.8	3.449
Correlation Between Means:			-410.942		
Correlation Between Means:			0		

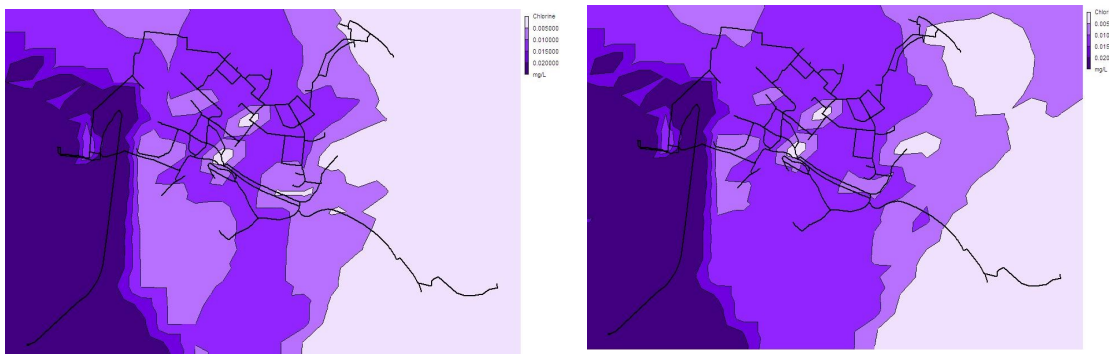
### APPENDIX 3: Chlorine concentration in the network nodes



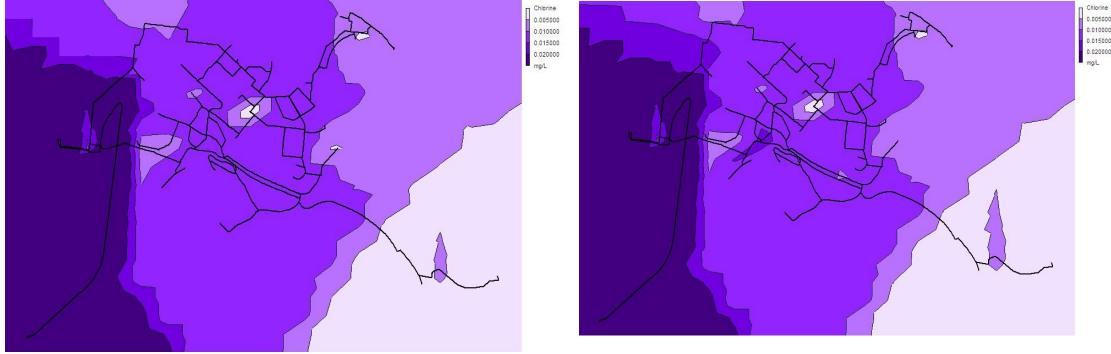
**A. Figure 3.1: Legend**



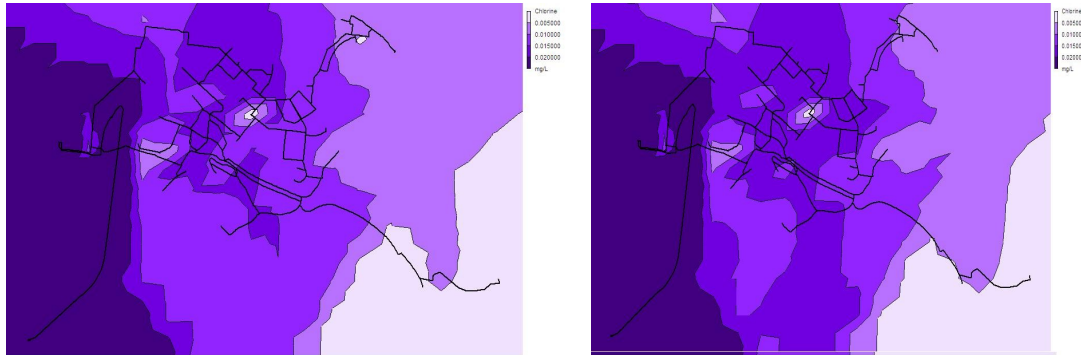
**A. Figure 3.2: Chlorine concentration in network nodes,  $K_b = -0.14 \text{ day}^{-1}$   $K_w = -0.12 \text{ mm/day}$  at 10:00 (left) and at 14:00 (right)**



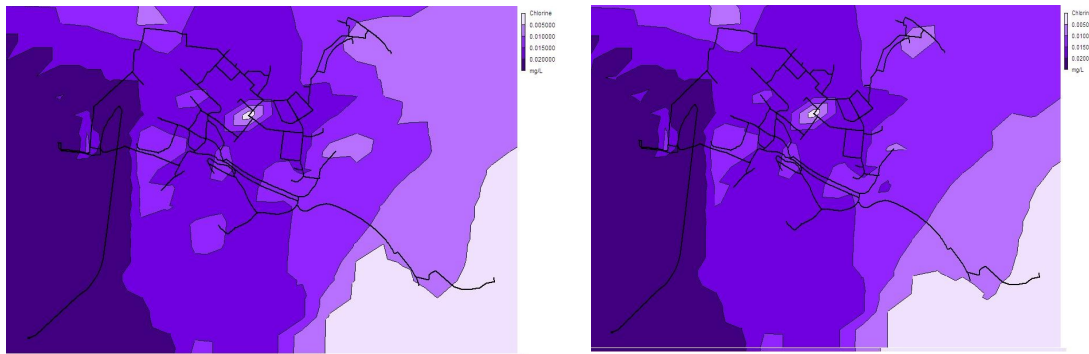
**A. Figure 3.3: Chlorine concentration in network nodes,  $K_b = -0.14 \text{ day}^{-1}$   $K_w = -0.12 \text{ mm/day}$  at 18:00 (left) and at 22:00 (right)**



**A. Figure 3.4:** Chlorine concentration in network nodes,  $K_b = -0.14 \text{ day}^{-1}$   $K_w = -0.12 \text{ mm/day}$  at 26:00 (left) and at 30:00 (right)



**A. Figure 3.5:** Chlorine concentration in network nodes,  $K_b = -0.14 \text{ day}^{-1}$   $K_w = -0.12 \text{ mm/day}$  at 34:00 (left) and at 38:00 (right)



**A. Figure 3.6:** Chlorine concentration in network nodes,  $K_b = -0.14 \text{ day}^{-1}$   $K_w = -0.12 \text{ mm/day}$  at 42:00 (left) and at 46:00 (right)

## APPENDIX 4: ANN Models simulation and results

**A. Table 4.1:** Summay of Network Results

Subset	Net. name	Training perf.	Test perf.	Validation perf.	Training error	Test error	Validation error	Training algorithm	Error function	Hidden activation	Output activation
1	MLP 9-8-1	0.934331	0.915416	0.907718	0.000008	0.000007	0.000006	BFGS 41	SOS	Identity	Logistic
2	MLP 5-5-1	0.898331	0.908285	0.944157	0.000009	0.000008	0.000003	BFGS 15	SOS	Exponential	Identity
3	MLP 5-5-1	0.946651	0.907248	0.928412	0.000005	0.000007	0.000007	BFGS 23	SOS	Tanh	Identity
4	MLP 9-8-1	0.977804	0.407221	0.703433	0.000000	0.000003	0.000004	BFGS 15	SOS	Exponential	Identity
5	MLP 6-11-1	0.987626	0.421866	0.667195	0.000000	0.000003	0.000003	BFGS 303	SOS	Tanh	Identity
6	MLP 9-9-1	0.998974	0.533841	0.538665	0.000000	0.000000	0.000001	BFGS 44	SOS	Exponential	Identity
7	MLP 6-10-1	0.984109	0.402605	0.781112	0.000000	0.000000	0.000001	BFGS 57	SOS	Identity	Exponential

**A. Table 4.2:** Predictions Statistics for each Model Subset

Subset #	1	2	3	4	5	6	7
	1.MLP 9-8-1	1.MLP 5-5-1	3.MLP 5-5-1	2.MLP 9-8-1	3.MLP 6-11-1	2.MLP 9-9-1	3.MLP 6-10-1
Minimum prediction (Train)	0.01069	0.00823	0.00416	0.00920	0.00947	0.01001	0.01141
Maximum prediction (Train)	0.03371	0.03291	0.03408	0.03002	0.02927	0.03006	0.03095
Minimum prediction (Test)	0.01051	0.00550	0.00450	0.00955	0.00947	0.01343	0.01334
Maximum prediction (Test)	0.03099	0.03388	0.03065	0.01356	0.01354	0.01406	0.01385
Minimum prediction (Validation)	0.01039	0.00695	0.00466	0.00920	0.00882	0.01207	0.01250
Maximum prediction (Validation)	0.02748	0.03327	0.02792	0.01408	0.01610	0.01435	0.01387
Minimum prediction (Missing)							
Maximum prediction (Missing)							
Minimum residual (Train)	-0.01024	-0.00960	-0.00941	-0.00146	-0.00122	-0.00029	-0.00238
Maximum residual (Train)	0.01149	0.00980	0.00611	0.00130	0.00100	0.00031	0.00045
Minimum residual (Test)	-0.00559	-0.00388	-0.00737	-0.00199	-0.00269	-0.00026	-0.00008
Maximum residual (Test)	0.00632	0.01032	0.00624	0.00755	0.00689	0.00034	0.00067
Minimum residual (Validation)	-0.00328	-0.00371	-0.00189	-0.00171	-0.00268	-0.00375	-0.00324
Maximum residual (Validation)	0.00730	0.00519	0.00739	0.00802	0.00742	0.00070	0.00055
Minimum standard residual (Train)	-3.69551	-3.12985	-4.22423	-2.74567	-2.96357	-2.88105	-5.88961
Maximum standard residual (Train)	4.14669	3.19292	2.74284	2.44055	2.41937	3.06991	1.12217
Minimum standard residual (Test)	-2.12238	-1.37736	-2.72790	-1.23018	-1.69938	-2.11501	-0.38732
Maximum standard residual (Test)	2.39616	3.66529	2.30731	4.65770	4.35255	2.80910	3.43825
Minimum standard residual (Validation)	-1.36060	-1.99406	-0.71980	-0.84973	-1.48017	-4.54618	-4.21464
Maximum standard residual (Validation)	3.03278	2.78933	2.82041	3.98882	4.09271	0.84857	0.71946

**A. Table 4.3:** Sensitivity Analysis for Model subsets 1, 2 and 3

Subset	Initial Chlorine Bosonohy Tank	pH	Free Chlorine Kohoutovice Tank	Temperature Junction 290172 (Libusina tr.4)	Temperature	Flow Kohoutovice Tank	pH Junction 290172 (Libusina tr.4)	Turbidity	Turbidity Junction 290172 (Libusina tr.4)
1.MLP 9-8-1	6.032489	1.260633	1.199060	1.181358	1.121093	1.040400	1.010297	0.968116	0.940948
2.MLP 5-5-1	7.418023	1.395260	2.107620		1.042762	1.096422			
3.MLP 5-5-1	10.52243		2.240832	1.487396	1.133877	1.482115			

**A. Table 4.4: Sensitivity Analysis for Model subsets 4 and 5**

Subset	pH	Initial Chlorine Bosonohy Tank	Temperature Junction 291750 Libusino udoli 66 - 58 - 160	Free Chlorine Kohoutovice Tank	pH Junction 291750 Libusino udoli 66 - 58 - 160	Turbidity Junction 291750 Libusino udoli 66 - 58 - 160	Temperature	Flow Kohoutovice Tank	Turbidity
4.MLP 9-8-1	2.514563	1.967296	1.731978	1.472971	1.124257	1.072625	1.052683	1.042012	1.011154
5.MLP 6-11-1	3.096307	1.224592	2.287724	1.471534			1.218488	1.113323	

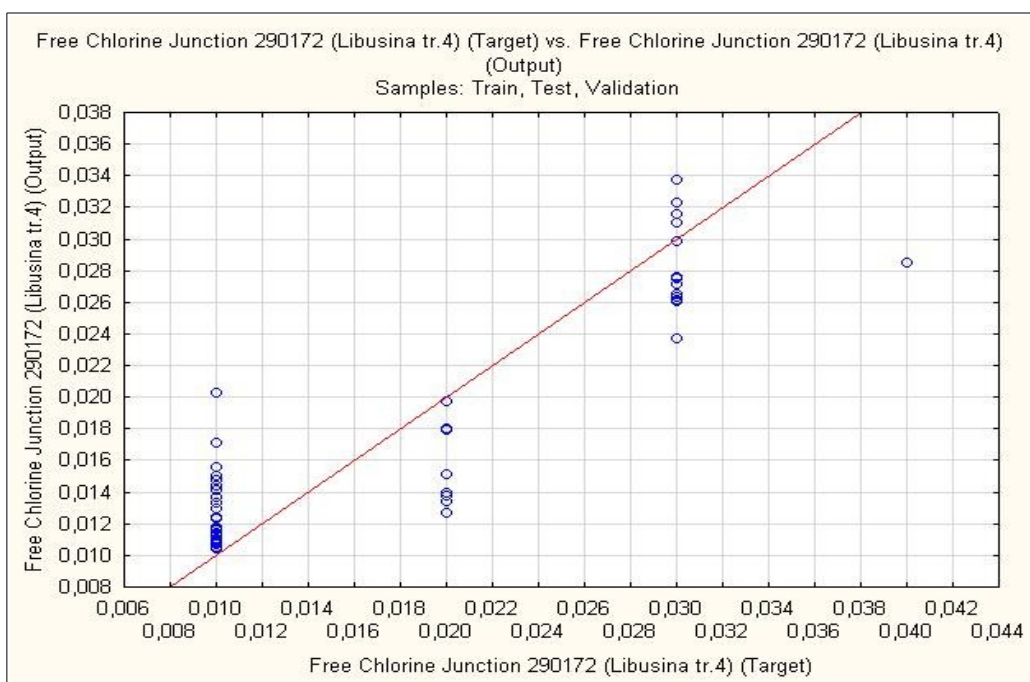
**A. Table 4.5: Sensitivity Analysis for Model subsets 6 and 7**

Subset	Turbidity Junction 294526 Nad Pisárkami 2	pH Junction 294526 Nad Pisárkami 2	Turbidity	Temperature	Free Chlorine Kohoutovic e Tank	pH	Flow Kohoutovice Tank	Temperature Junction 294526 Nad Pisárkami 2	Initial Chlorine Bosonohy Tank
6.MLP 9-9-1	6.918514	4.696061	1.674389	1.133401	1.038547	0.977815	0.969035	0.960977	0.903444
7.MLP 6-10-1	8.085072	3.693667	1.089613		1.113357	1.001689			1.061986

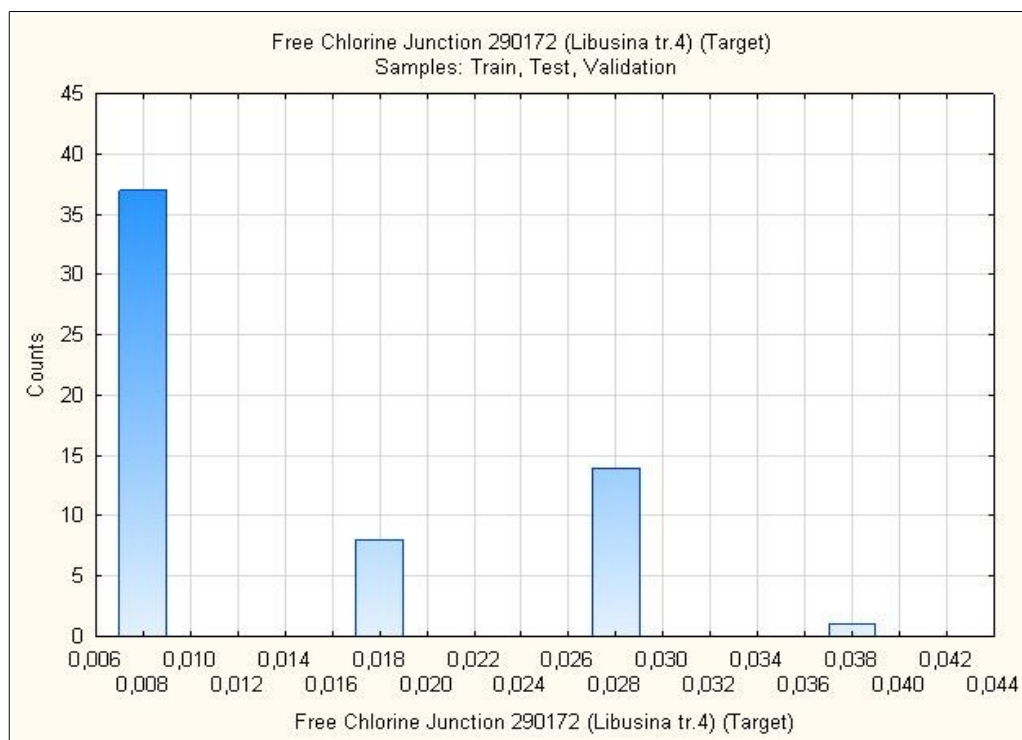
**A. Table 4.6:** Predictions results for Nodes 1, 2 and 3 in each Model subset

		Node 1				Node 2			Node 3		
	Sample	Target	1	2	3	Target	4	5	Target	6	7
		Free Chlorine Junction 290172 (Libusina tr.4)	Free Chlorine Junction 290172 (Libusina tr.4) - Output	Free Chlorine Junction 290172 (Libusina tr.4) - Output	Free Chlorine Junction 290172 (Libusina tr.4) - Output	Free Chlorine Junction 291750 Libusino udoli 66 - 58 - 160	Free Chlorine Junction 291750 Libusino udoli 66 - 58 - 160 - Output	Free Chlorine Junction 291750 Libusino udoli 66 - 58 - 160 - Output	Free Chlorine Junction 294526 Nad Pisárkami 2	Free Chlorine Junction 294526 Nad Pisárkami 2 - Output	Free Chlorine Junction 294526 Nad Pisárkami 2 - Output
1	Train	0.030000	0.031528	0.032908	0.034080	0.010000	0.010285	0.010889	0.013796	0.013790	0.014135
2	Validation	0.020000	0.012698	0.018690	0.018644	0.020000	0.014081	0.016102	0.010000	0.012069	0.013236
3	Test	0.020000	0.013919	0.015365	0.015105	0.010000	0.011783	0.010815	0.013608	0.013864	0.013684
4	Train	0.040000	0.028510	0.032454	0.033888	0.012375	0.013155	0.012183	0.013966	0.013924	0.014181
5	Test	0.010000	0.011254	0.010102	0.010343	0.010000	0.010373	0.012070	0.013569	0.013577	0.013414
6	Validation	0.010000	0.010900	0.006952	0.007497	0.012247	0.010208	0.010891	0.013509	0.013820	0.013374
7	Validation	0.010000	0.011697	0.008042	0.008888	0.010000	0.009200	0.008818	0.013519	0.013851	0.013394
8	Test	0.010000	0.010783	0.008546	0.008752	0.012471	0.010246	0.010215	0.014019	0.013784	0.013345
9	Test	0.010000	0.010667	0.010060	0.010205	0.010000	0.010636	0.011685	0.013718	0.013740	0.013476
10	Train	0.010000	0.014757	0.015472	0.015136	0.012608	0.011681	0.011920	0.013761	0.013877	0.013609
11	Test	0.020000	0.017931	0.018303	0.021577	0.020000	0.012448	0.013113	0.014037	0.013791	0.013814
12	Test	0.010000	0.011171	0.005501	0.004504	0.010000	0.009555	0.009472	0.013776	0.013844	0.013339
13	Train	0.010000	0.020240	0.019603	0.019413	0.012602	0.011948	0.012168	0.013538	0.013534	0.013717
14	Train	0.010000	0.013692	0.013624	0.011891	0.010000	0.010703	0.009862	0.013688	0.013516	0.013550
15	Validation	0.010000	0.012335	0.008690	0.009331	0.010000	0.009936	0.009767	0.010000	0.013746	0.012502
16	Validation	0.020000	0.013414	0.015367	0.012610	0.020000	0.011983	0.012577	0.013457	0.013853	0.013650
17	Train	0.010000	0.010937	0.010393	0.009282	0.012635	0.011788	0.012558	0.013781	0.013852	0.013632
18	Validation	0.010000	0.011362	0.010546	0.008610	0.010000	0.010566	0.011661	0.013999	0.013505	0.013458
19	Train	0.010000	0.011227	0.010724	0.009145	0.012209	0.011198	0.011518	0.013650	0.013630	0.013504
20	Train	0.010000	0.010688	0.010547	0.010905	0.010000	0.010525	0.010850	0.013889	0.013719	0.013679
21	Validation	0.020000	0.018015	0.018718	0.016006	0.012386	0.011401	0.010729	0.013684	0.013774	0.013872
22	Train	0.020000	0.019671	0.013620	0.014789	0.010000	0.011463	0.011219	0.013754	0.013670	0.013607
23	Test	0.010000	0.015594	0.009028	0.017372	0.010000	0.010508	0.009572	0.013792	0.013452	0.013344
24	Train	0.010000	0.012949	0.012809	0.010487	0.010000	0.010804	0.010384	0.013588	0.013542	0.013601
25	Train	0.010000	0.014081	0.013508	0.011080	0.012592	0.011292	0.011597	0.013445	0.013640	0.013686
26	Test	0.010000	0.014355	0.013808	0.011487	0.012265	0.011866	0.012443	0.013838	0.013740	0.013695
27	Validation	0.010000	0.013276	0.013713	0.011886	0.012313	0.010883	0.011620	0.013660	0.013832	0.013638
28	Validation	0.010000	0.010656	0.010315	0.010879	0.012223	0.011177	0.011303	0.013904	0.014011	0.013662
29	Train	0.010000	0.010738	0.008231	0.009026	0.010000	0.009991	0.010477	0.010000	0.010016	0.012384
30	Train	0.010000	0.011697	0.010354	0.011693	0.010000	0.010139	0.010030	0.013816	0.013664	0.013452
31	Validation	0.010000	0.011739	0.010423	0.010427	0.012217	0.010112	0.009476	0.013765	0.013564	0.013515
32	Train	0.010000	0.013622	0.015985	0.011541	0.010000	0.010751	0.009888	0.013513	0.013682	0.013585
33	Train	0.010000	0.011320	0.009277	0.009929	0.012276	0.011470	0.012191	0.013516	0.013650	0.013573
34	Train	0.010000	0.010866	0.008917	0.009097	0.010000	0.009195	0.009655	0.013735	0.013862	0.013612
35	Train	0.010000	0.010968	0.009290	0.010775	0.010000	0.009265	0.009476	0.013721	0.013762	0.013542
36	Train	0.010000	0.011701	0.009018	0.004158	0.010000	0.010148	0.011103	0.013868	0.013736	0.013414
37	Train	0.010000	0.010719	0.008300	0.009146	0.010000	0.010160	0.009471	0.010000	0.010012	0.011415
38	Validation	0.020000	0.015079	0.014806	0.013577	0.012301	0.010346	0.010433	0.013679	0.013756	0.013555
39	Test	0.020000	0.013769	0.017698	0.019103	0.012210	0.013030	0.013543	0.014013	0.014058	0.013839
40	Validation	0.010000	0.010392	0.008818	0.007259	0.010000	0.011708	0.012685	0.013733	0.014355	0.013696
41	Validation	0.010000	0.011551	0.008895	0.004659	0.012331	0.011427	0.012357	0.013937	0.013250	0.013385
42	Train	0.010000	0.014983	0.015579	0.009944	0.010000	0.011109	0.011005	0.013490	0.013781	0.013713
43	Test	0.010000	0.011782	0.010419	0.006083	0.012357	0.010236	0.010190	0.013526	0.013595	0.013557
44	Test	0.010000	0.010509	0.010559	0.011004	0.010000	0.011995	0.012689	0.013740	0.013849	0.013616
45	Train	0.010000	0.017054	0.016390	0.011735	0.010000	0.011392	0.010349	0.013718	0.013733	0.013583
46	Train	0.010000	0.011078	0.009617	0.009582	0.010000	0.010118	0.010326	0.013833	0.013683	0.013526
47	Train	0.010000	0.012401	0.014143	0.011306	0.012452	0.011537	0.012215	0.013654	0.013850	0.013660
48	Train	0.030000	0.032293	0.020203	0.026657	0.012363	0.012314	0.012172	0.030000	0.030064	0.030950
49	Test	0.030000	0.023684	0.019682	0.030648	0.012370	0.012086	0.012397	0.013699	0.013426	0.013591
50	Test	0.030000	0.026295	0.023716	0.024203	0.012363	0.011175	0.011800	0.014065	0.014002	0.013846
51	Train	0.030000	0.027469	0.029439	0.028462	0.012613	0.012241	0.012345	0.013440	0.013651	0.013598
52	Train	0.030000	0.026518	0.021715	0.026850	0.012413	0.011744	0.012505	0.013987	0.014000	0.013693
53	Test	0.030000	0.025997	0.027036	0.026241	0.012395	0.011442	0.012252	0.013837	0.013713	0.013646
54	Train	0.030000	0.029866	0.029979	0.032186	0.012339	0.011359	0.012343	0.013806	0.013678	0.013685
55	Validation	0.030000	0.026134	0.033270	0.027921	0.012643	0.011841	0.012407	0.013825	0.013765	0.013790
56	Test	0.030000	0.030995	0.033877	0.023764	0.012566	0.013563	0.013077	0.013781	0.013904	0.013700
57	Train	0.030000	0.033707	0.031718	0.031160	0.012650	0.012839	0.013438	0.013552	0.013649	0.013701
58	Train	0.030000	0.027117	0.030001	0.030456	0.012318	0.012298	0.012644	0.014061	0.013751	0.013733
59	Validation	0.030000	0.027480	0.026884	0.023873	0.012632	0.012498	0.012651	0.014005	0.013306	0.013672
60	Train	0.030000	0.027525	0.029959	0.025808	0.030000	0.030018	0.029265	0.013609	0.013861	0.013783

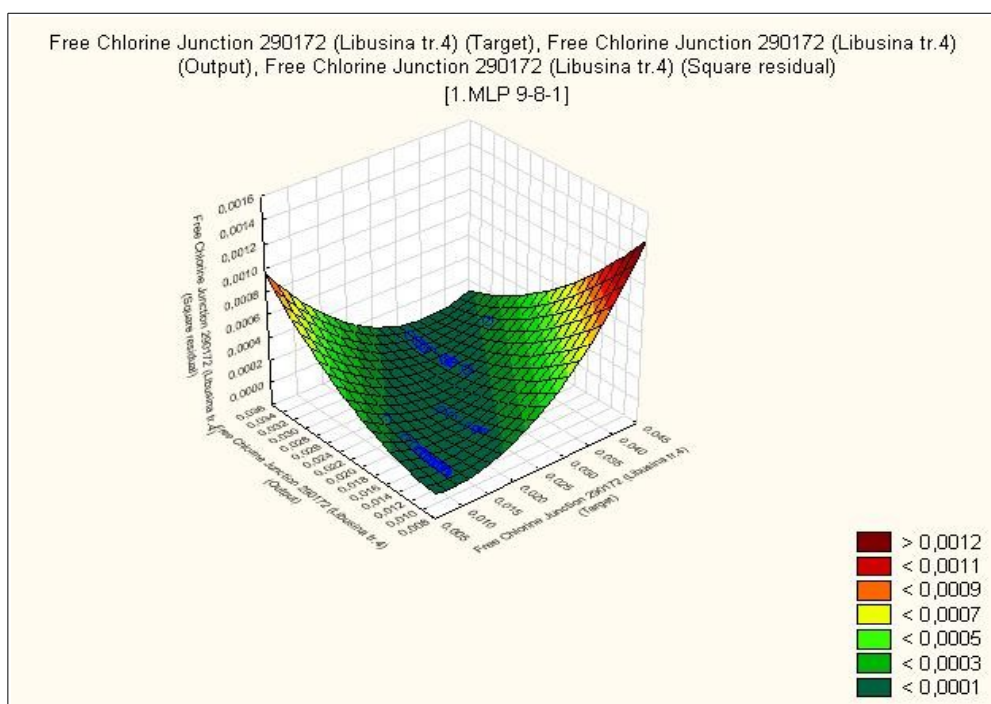




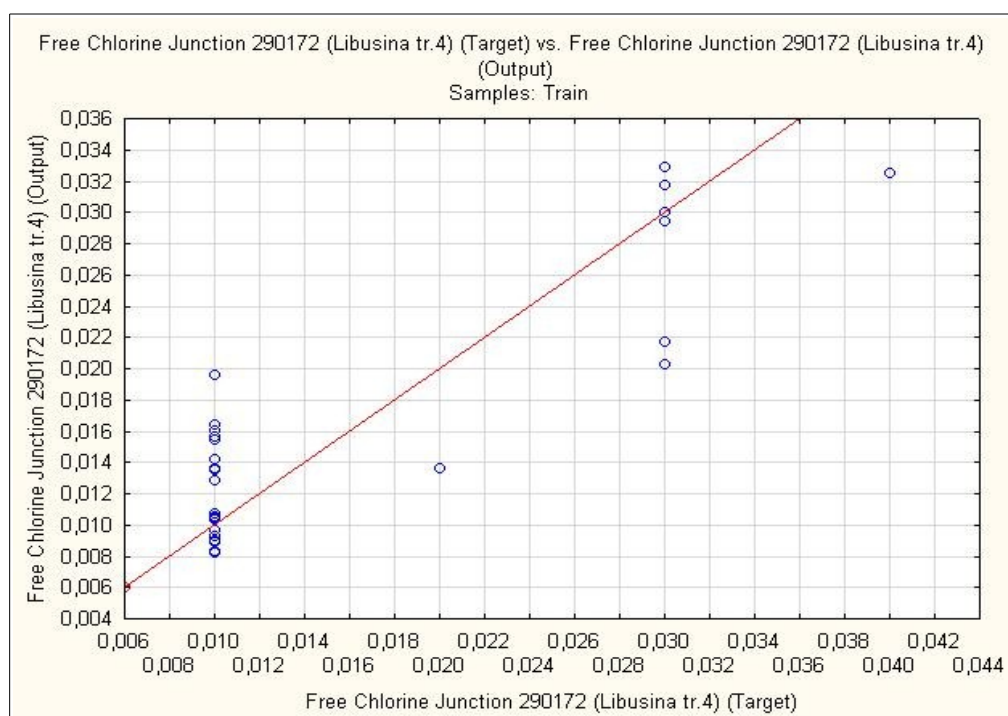
**A. Figure 4.1:** Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 1



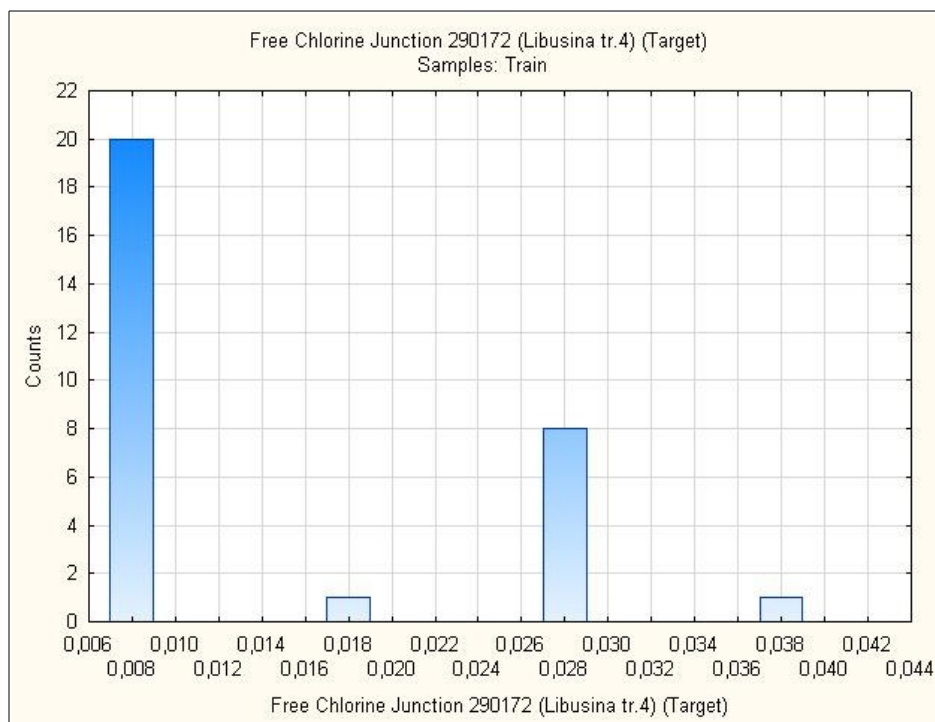
**A. Figure 4.2:** Histogram Free Chlorine Node 1 (Target) for Model Subset 1



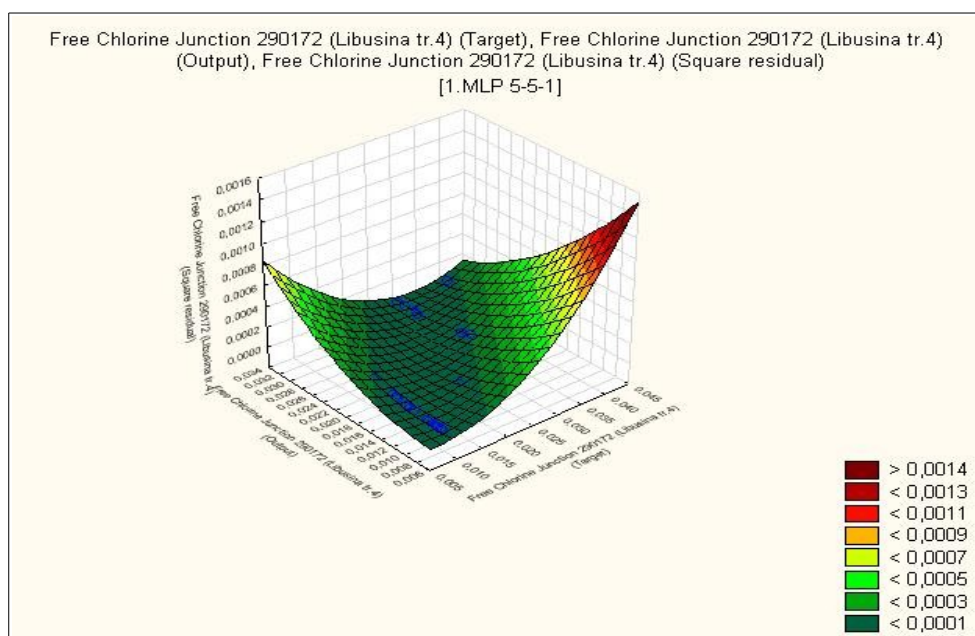
**A. Figure 4.3:** Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 1



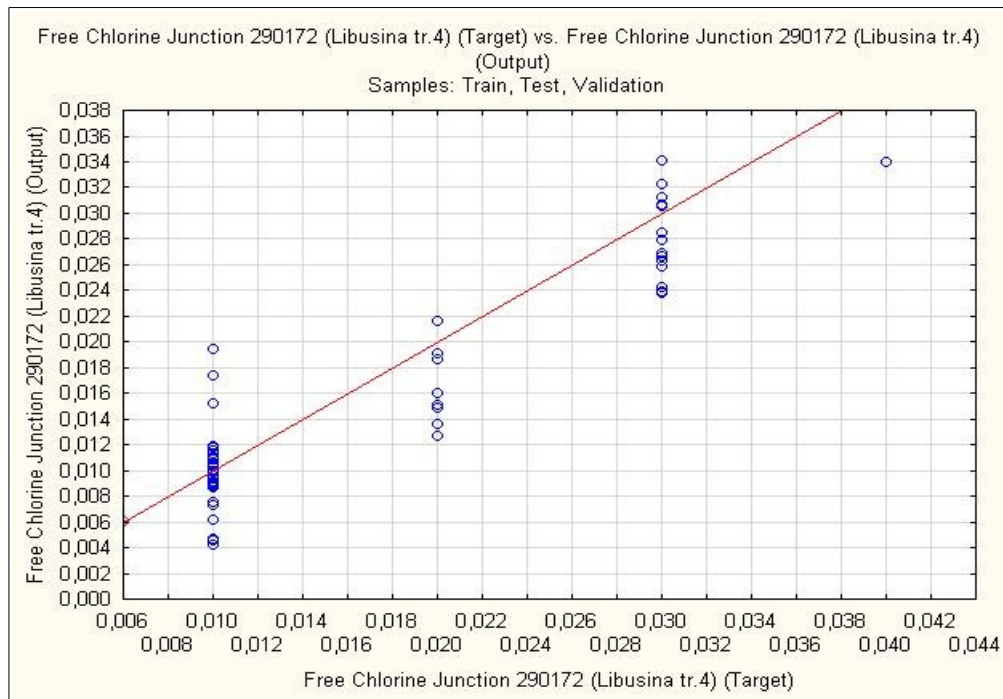
**A. Figure 4.4:** Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 2



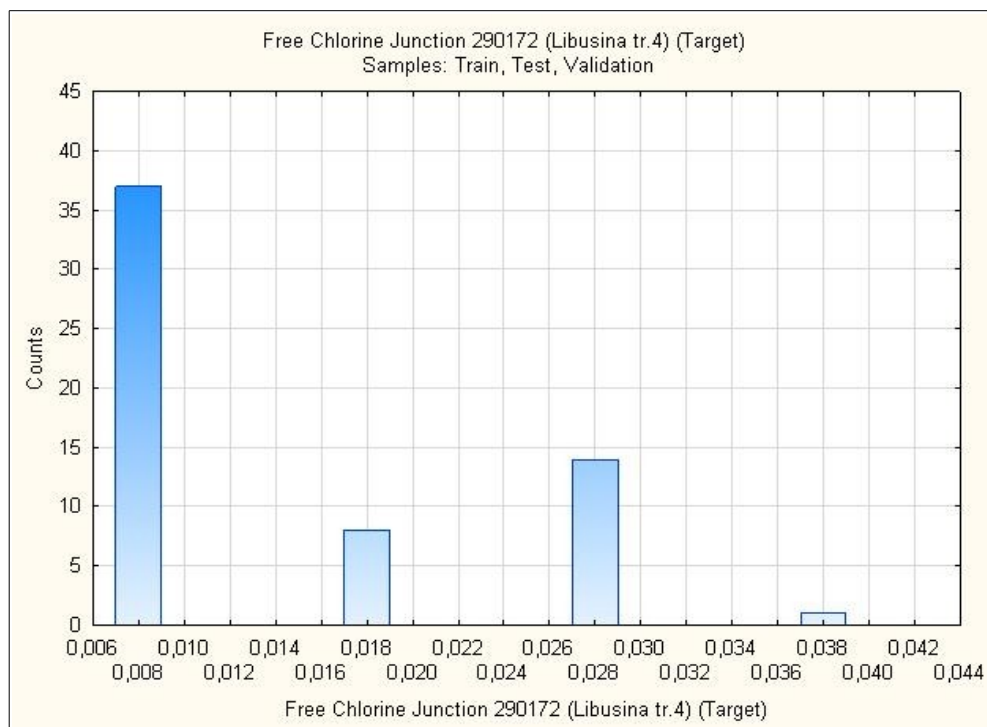
**A. Figure 4.5:** Histogram Free Chlorine Node 1 (Target) for Model Subset 2



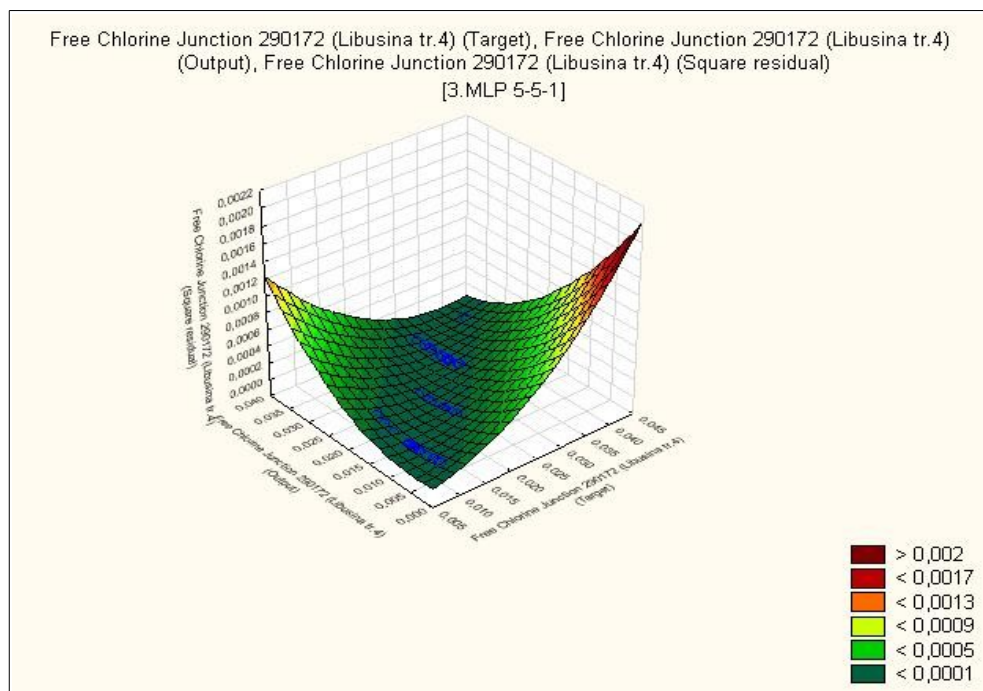
**A. Figure 4.6:** Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 2



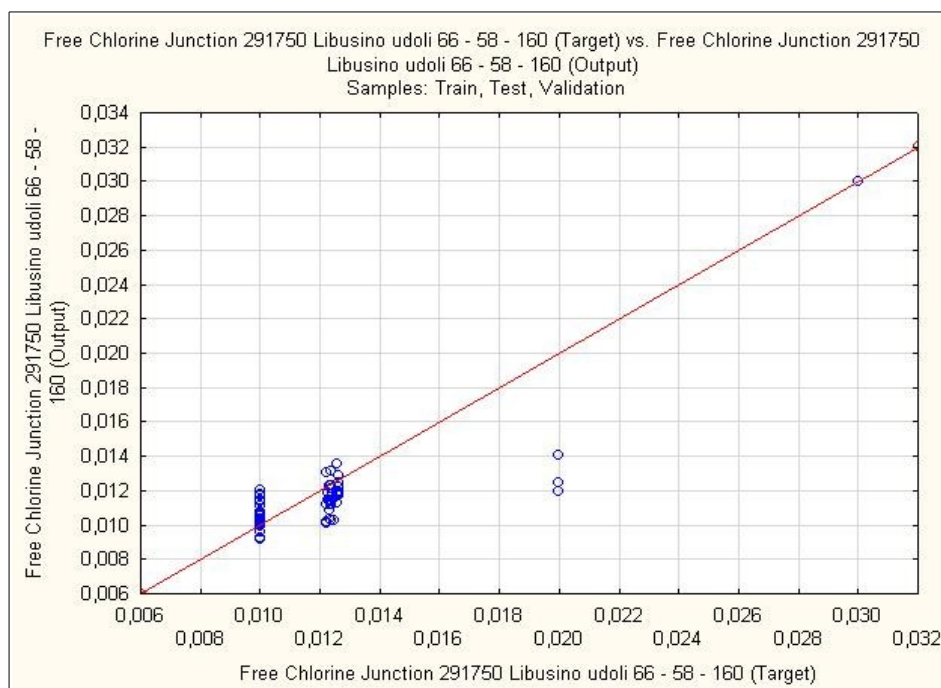
**A. Figure 4.7:** Free Chlorine Node 1 (Target) vs. Free Chlorine Node 1 (Output) For Model Subset 3



**A. Figure 4.8:** Histogram Free Chlorine Node 1 (Target) for Model Subset 3

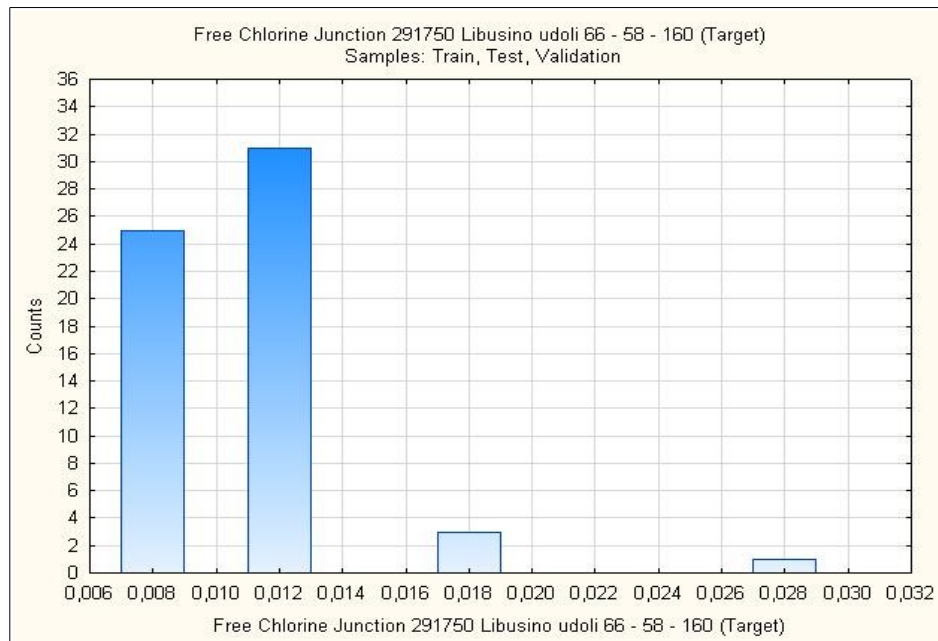


**A. Figure 4.9:** Free Chlorine Node 1 (Target), Free Chlorine Node 1 (Output), Free Chlorine Node 1 (Square residual) for Model Subset 2

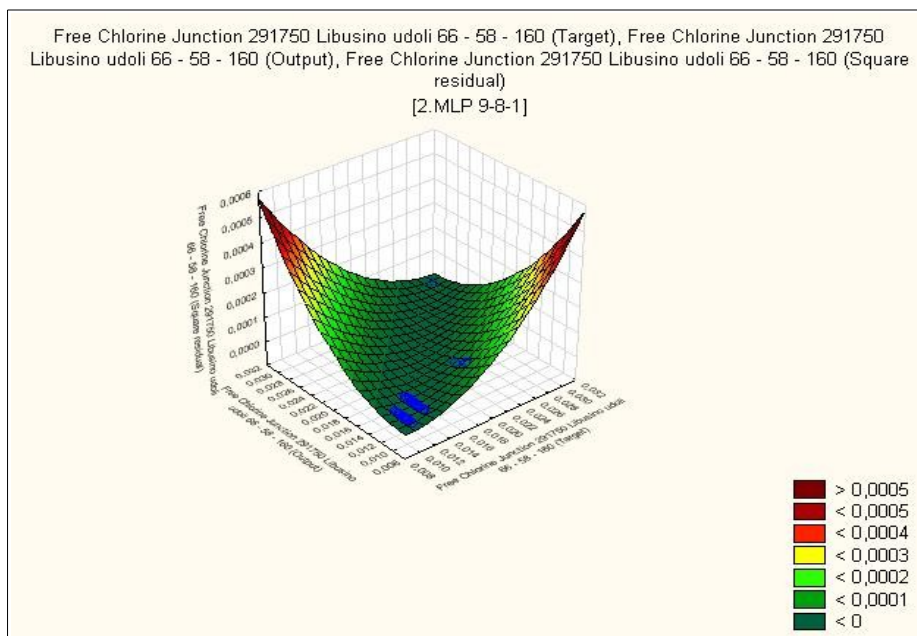


**A. Figure 4.10:** Free Chlorine Node 2 (Target) vs. Free Chlorine Node 2 (Output) For Model Subset 4

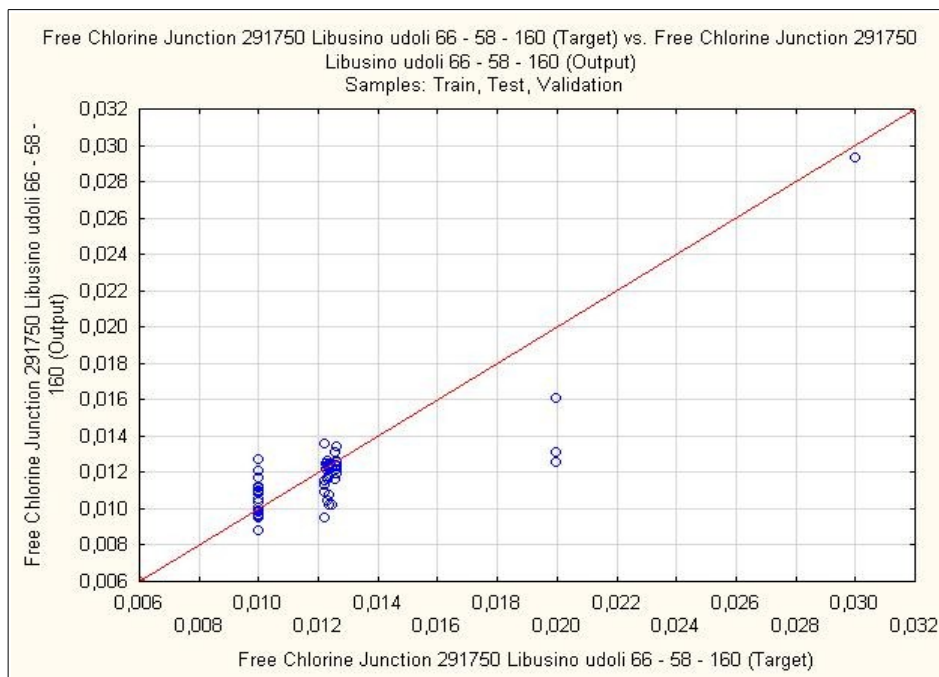




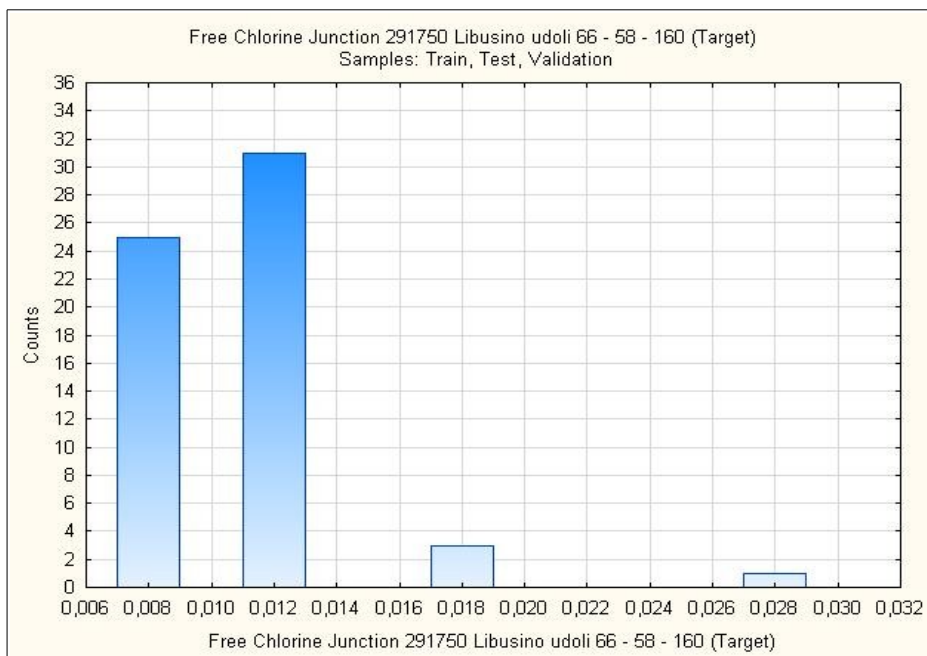
**A. Figure 4.11:** Histogram Free Chlorine Node 2 (Target) for Model Subset 4



**A. Figure 4.12:** Free Chlorine Node 2 (Target), Free Chlorine Node 2 (Output), Free Chlorine Node 2 (Square residual) for Model Subset 4



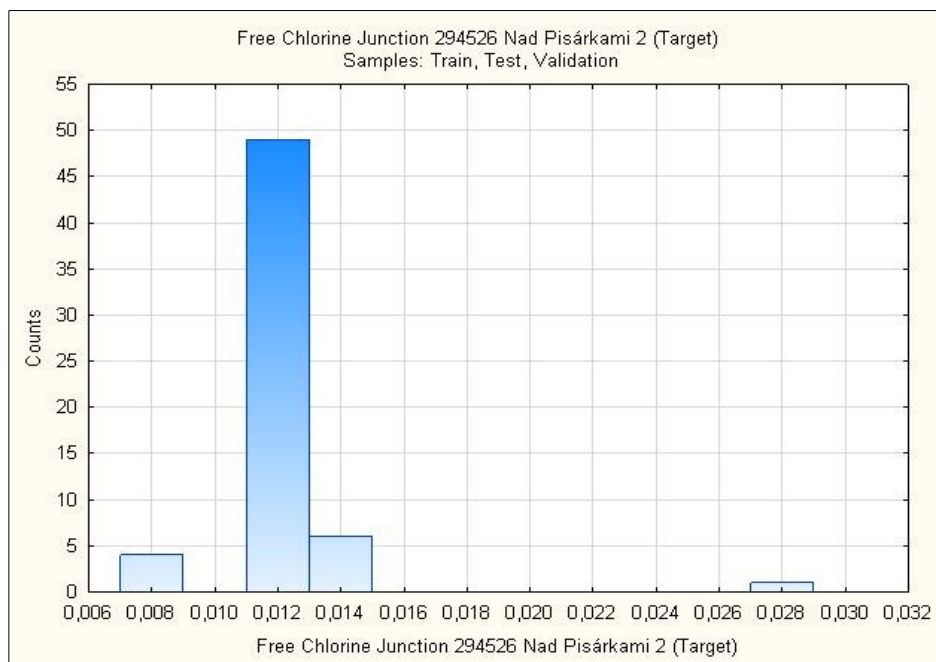
**A. Figure 4.13:** Free Chlorine Node 2 (Target) vs. Free Chlorine Node 2 (Output) For Model Subset 5



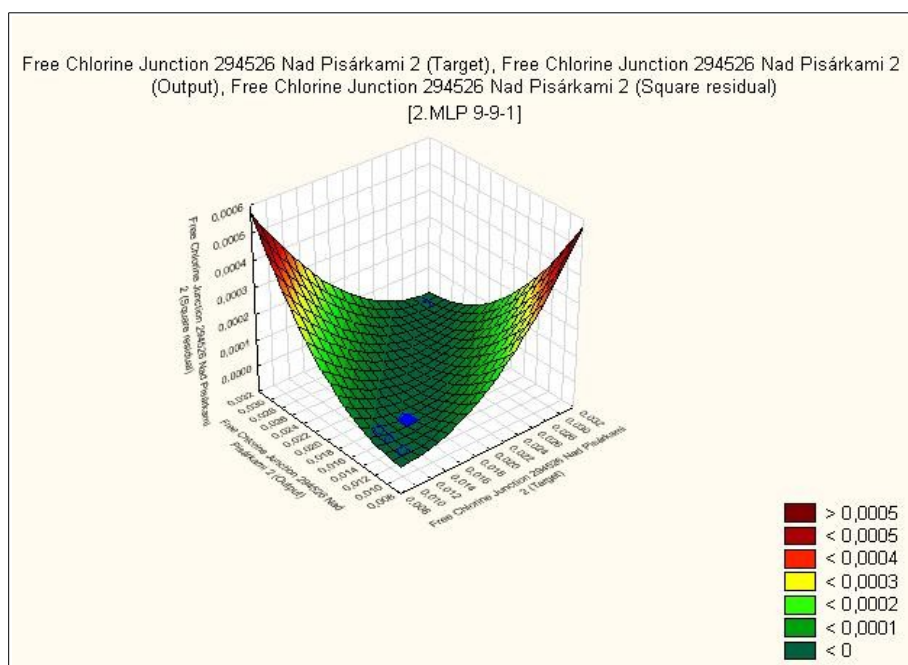
**A. Figure 4.14:** Histogram Free Chlorine Node 2 (Target) for Model Subset 5



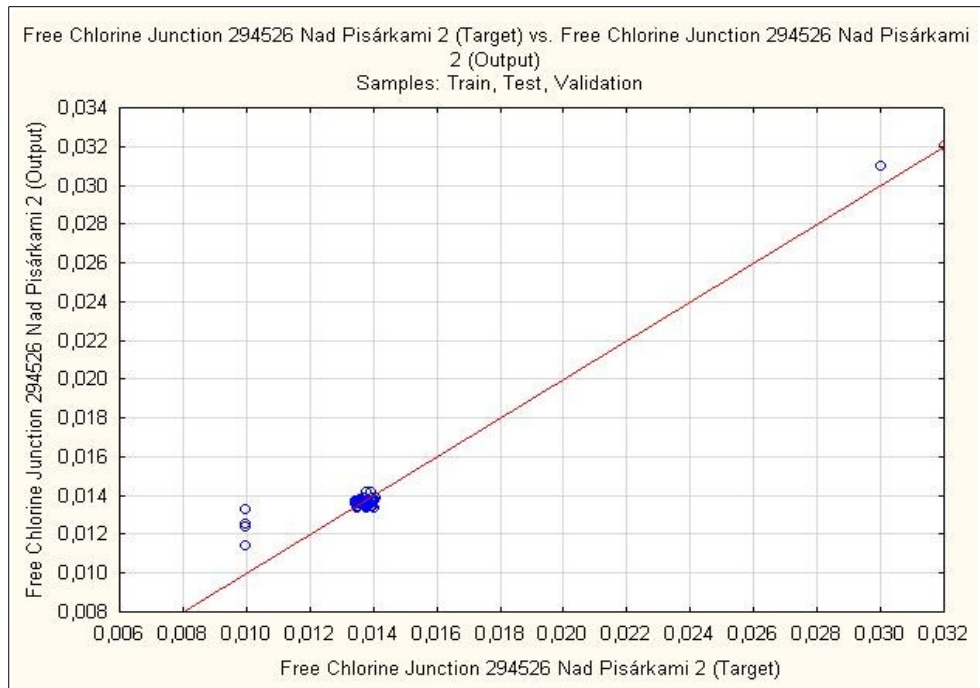




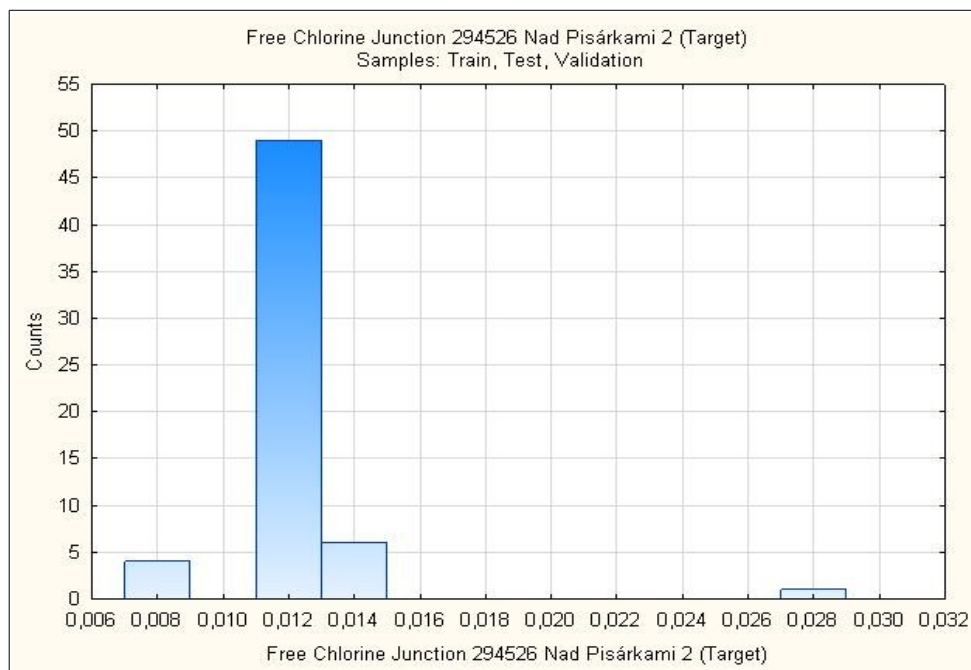
**A. Figure 4.17:** Histogram Free Chlorine Node 3 (Target) for Model Subset 6



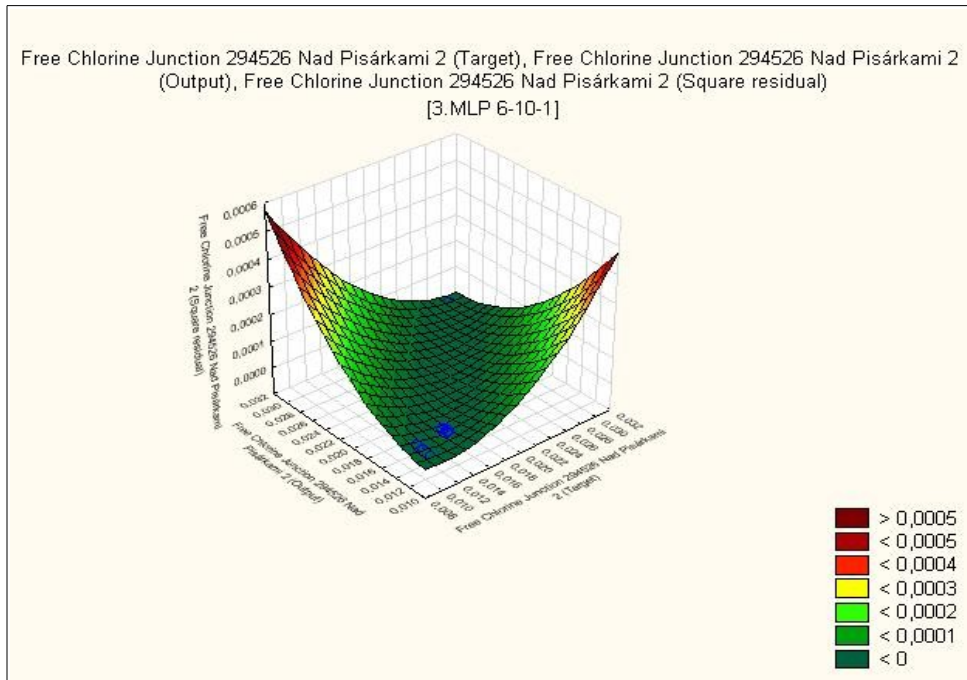
**A. Figure 4.18:** Free Chlorine Node 3 (Target), Free Chlorine Node 3 (Output), Free Chlorine Node 3 (Square residual) for Model Subset 6



**A. Figure 4.19:** Free Chlorine Node 3 (Target) vs. Free Chlorine Node 3 (Output) For Model Subset 7



**A. Figure 4.20:** Histogram Free Chlorine Node 3 (Target) for Model Subset 7



**A. Figure 4.21:** Free Chlorine Node 3 (Target), Free Chlorine Node 3 (Output), Free Chlorine Node 3 (Square residual) for Model Subset 7

## APPENDIX 5: Database of parameters used for ANN simulation, Našiměřice, case study.

**A. Table 5.1:** Four-years database of parameters measured in Našiměřice

Location	Temperature	pH	Free Chlorine	Flow	Pipe material	Diameter	Age of pipe
Našiměřice c.p. 117,kohout na zahradu	10.2	7.46	0.09	0.52	OC	DN80	A3
Našiměřice c.p. 117,kohout na zahradu	10.2	7.47	0.09	0.52	OC	DN80	A3
Našiměřice, c.p. 83, OÚ, kuchynka	14.1	7.38	0.05	0.12	LT	DN100	A1
Našiměřice, c.p. 83, OÚ, kuchynka	16.7	7.41	0.05	0.35	LT	DN100	A1
Našiměřice c.p. 44,dvur - venkovní kohout	18.4	7.28	0.05	0.5	LT	DN100	A4
NAŠIMERICE, VODOJEM, Našiměřice - odtok	10	7.49	0.11	0.25	LT	DN100	A3
Našiměřice - odkalení naproti c. 113	10.6	7.42	0.05	0.269	LT	DN100	A1
Našiměřice, c.p. 83, OÚ, kuchynka	7.4	7.3	0.06	0.27	LT	DN100	A1
Našiměřice, c.p. 83, OÚ, kuchynka	9.2	7.25	0.28	0.25	LT	DN100	A1
NAŠIMERICE, VODOJEM, Našiměřice - přítok	9.5	7.35	0.1	0.28	LT	DN100	A3
Našiměřice c.114 venkovní kohout	17.8	7.22	0.05	0.5	PVC	DN110	A1
Našiměřice, c.p. 83, OÚ, kuchynka	16.9	7.3	0.05	0.36	LT	DN100	A1
Našiměřice, c.p. 83, OÚ, kuchynka	8.4	7.31	0.4	0.6	LT	DN100	A1
Našiměřice - c.p.20 - kuchyne, umyvadlo	17.9	7.9	0.02	0.4	LT	DN100	A4
Našiměřice - c.16 kuchyn	17.3	7.59	0.08	0.6	LT	DN100	A4
Našiměřice - Obecní úřad - kuchynka - umyvad	18.8	7.5	0.1	0.2	LT	DN100	A4
Našiměřice - c.16 kuchyn	20.4	7.42	0.17	0.29	LT	DN100	A4
Našiměřice - Obecní úřad - kuchynka - umyvad	8.4	7.31	0.4	0.8	LT	DN100	A4
Našiměřice - VDĽ - odtok - kohout	10.8	7.42	0.03	0.2	LT	DN100	A3
NAŠIMERICE, VODOJEM, Našiměřice - odtok	13	7.23	0.39	0.285	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	13	7.25	0.15	0.29	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	13	7.28	0.02	0.27	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	13	7.3	0.05	0.28	LT	DN100	A4
Našiměřice c.p. 90,kohout na zahradu	16.5	7.41	0.2	0.24	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.7	7.42	0.01	0.21	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	18.1	7.38	0.04	0.22	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	11.3	7.36	0.5	0.26	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	13.7	7.25	0.44	0.36	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	18.3	7.42	0.14	0.25	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.7	7.49	0.02	0.26	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	17.9	7.59	0.03	0.27	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	11.3	7.41	0.35	0.22	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	13.9	7.29	0.25	0.22	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	18.4	7.48	0.08	0.54	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.5	7.5	0.04	0.36	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	17	7.56	0.05	0.25	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	11.2	7.43	0.6	0.01	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	14.3	7.26	0.11	0.29	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	18.2	7.48	0.05	0.24	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.4	7.5	0.04	0.26	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	18	7.5	0.06	0.57	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	11.4	7.66	0.36	0.51	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	12.9	7.26	0.33	0.28	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	17.5	7.5	0.07	0.65	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.6	7.42	0.04	0.45	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	17.4	7.22	0.05	0.49	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	11.3	7.22	0.34	0.05	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	13.8	7.25	0.17	0.24	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	17.4	7.56	0.05	0.48	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	17.3	7.42	0.03	0.26	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	17.6	7.49	0.24	0.67	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	12.1	7.3	0.38	0.30	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	13.2	7.28	0.36	0.19	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	17.2	7.4	0.08	0.35	PVC	DN90	A1
Našiměřice c.p. 118,kohout na zahradu	18.1	7.43	0.04	0.24	LT	DN100	A4
Našiměřice c.p. 48, Úrad - venkovní kohout	17.5	7.51	0.09	0.45	LT	DN100	A4
Našiměřice - vrt HV3 - kohout	13.2	7.23	0.18	0.33	PVC	DN110	A1
NAŠIMERICE, VODOJEM, Našiměřice - odtok	14.3	7.3	0.23	0.24	LT	DN100	A3
Našiměřice c.p. 90,kohout na zahradu	17.4	7.22	0.07	0.35	PVC	DN90	A1