PREDICTION MODELS FOR MANGANESE, IRON AND AMMONIUM IN RAW WATER FOR A DRINKING WATER TREATMENT PLANT BUTONIGA (CROATIA)

Goran Volf[1*] – Morana Krbavčić[1] – Ivana Sušanj Čule[1] – Sonja Zorko[2]

[1]Department for Hydrotehnics, Faculty of Civil Engineering, University of Rijeka, Radmile Matejčić 3, 51000 Rijeka, Croatia
[2]Istarski Vodovod d.o.o., 52420 Buzet, Croatia

ARTICLE INFO

***Article history*:**
Received: 21.04.2023.
Received in revised form: 10.08.2023.
Accepted: 03.10.2023.

Keywords:
Drinking Water Treatment Plant
Prediction models
Manganese
Iron
Ammonium
Butoniga reservoir
Machine learning
Model trees
Performance and optimization
Treatment processes

DOI: https://doi.org/10.30765/er.2232

Abstract:

Drinking water treatment plant Butoniga is one of the main water supply facilities for potable water in Istria (Croatia). Water for treatment process is captured from the Butoniga reservoir which is a small and relatively shallow reservoir. As such, the reservoir is very sensitive to eutrophication and degradation processes caused by climate change and human activities in the watershed. In summer months during tourist season, when at highest water demand and lowest water level at the reservoir, the water temperature is the most critical parameter during treatment process. To capture colder water, raw water for treatment is taken from the lowest water intake, i.e. from the deepest layer in the Butoniga reservoir. This layer has another problem, namely increased concentrations of manganese, iron and ammonium under lower pH values. This study provides prediction models for manganese, iron and ammonium for seven days in advance, which are some of the most critical parameters during summer months and have significant influence on treatment process of raw water. For modelling purposes, machine learning software Weka was used to build models in form of model trees. Obtained prediction models for manganese, iron and ammonium have high accuracy compared to the measured data with a good prediction of the peak values. Therefore, obtained models can help in optimization of the treatment processes at the treatment plant, which are depending on the quality of raw water in Butoniga reservoir.

1 Introduction

Drinking water treatment plants (DWPT) are facilities which are used to produce clean water for human consumption with high quality and low cost which can adapt to standards and norms of public health. So, this makes DWTP a vital necessary facility to supply sufficient and quality potable water to the public [1]. Many DWTP use only mathematical models or empirical formulas as predictive models for their control systems which mainly lack a macro understanding of the overall dynamics, and the nonlinear relationships that are widely present in drinking water treatment. Accurate prediction models for drinking water treatment and production processes must be established to guide water treatment processes. [2]. The amount of data generated along DWTP allows development of data-based models like machine learning (ML) methods which are subfield of artificial intelligence (AI) and are able to predict operational parameters which can be incorporated into environmental decision support systems (DSS) [3]. Unlike mathematical models or statistical methods, ML methods have ability to handle complex nonlinear relationships and an accurate understanding of the overall dynamics of water treatment processes.

* Corresponding author
E-mail address: goran.volf@uniri.hr

Therefore, ML methods have the ability to monitor the evolution of water quality, analyse and predict water quality, and also reveal the process of pollutant migration and transformation, thereby shifting the focus from solving existing problems to identifying risks in advance and dynamically optimizing the facilities [1]. Also, models based on ML methods have a certain degree of interpretability, and appropriate analysis methods are capable of mining the hidden physical meaning and chemical or some other information to deepen the understanding of water treatment processes methods [4], [5]. Some applications of ML can be seen in research made by Wang et al. [6] where effluent turbidity in the drinking water flocculation process was estimated with an improved random forest model. In research made by Kim and Parnichkun [7] prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant was done using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system. Use of multiple linear regression (MLR) and artificial neural network (ANN) models were used by Ayanshola et al. [8] to predict treated water turbidity in a water treatment plant, while Alsaeed et al. [9] have done prediction of turbidity and aluminium in drinking water treatment plants using hybrid network algorithm (GA-ANN) and gene expression method (GEP). ANN were used also by Godo-Pla et al. [1] for predicting the oxidant demand in full-scale drinking water treatment plant. Recent advances in artificial intelligence and ML for nonlinear relationship analysis and process control in drinking water treatment can be found in research made by Li et al. [3]. Previous studies regarding functioning and problems of the Butoniga DWTP were done by Hajduk-Černeha [10], Zorko [11] and Volf et al. [12]. First experiences in use of the Butoniga DWTP for drinking water supply are described by Hajduk-Černeha [10]. Also, in this study [10] is analysed raw water quality from Butoniga reservoir and there are given some management guidelines regarding Butoniga reservoir and related DWTP. In the study made by Zorko [11] the impact of Butoniga reservoir raw water quality on water treatment processes is considered.

This study also gives some interesting conclusions such as, the main problem with Butoniga reservoir and thus related DWTP appear in summer months when water temperature is the most critical parameter, so in order to be suitable for use and also for treatment processes water temperature must not exceed the maximum allowable concentration (MAC) of 25 ^{o}C according to Croatian regulations for drinking water [13]. During this time period water is captured from the lowest water intake which captures water from the deepest water layer in the Butoniga reservoir. This layer has another problem, namely increased concentrations of manganese (Mn), iron (Fe) and ammonium (NH_4) under lower pH values. Increased concentration of Mn, Fe and NH_4 under lower pH values of raw water from the lowest water intake requires enhanced continuous process control and higher consumption of chemicals for the treatment process on DWTP. During these conditions the process is also stable and all water samples at the effluent are in accordance with the Croatian regulations for drinking water [13], where on exceeding values of the water temperature due to the heating of the Butoniga reservoir cannot be influenced [11]. Volf et al. [12] in his study presents and describes water quality index (WQI) for the Butoniga DWTP and associated WQI prediction models which are used for optimization of treatment processes on the DWTP. This study is an extension of the work done by Volf et al. [12] and deals with the investigations made by Zorko [11] regarding increased concentrations of Mn, Fe and NH_4 during summer months. In order to address the issue of increased concentrations of Mn, Fe, and NH_4 during the summer months, when higher chemical consumption is required for treatment processes, this study developed prediction models for Mn, Fe, and NH_4 concentrations with a seven-day forecast. Additionally, the research includes relevant findings regarding the operational aspects of the Butoniga DWTP, utilizing ML techniques in the form of model trees integrated into the Weka modelling software [14].

The model trees used for numeric prediction use linear equation in the terminal nodes (leaves) which allow a more accurate prediction of the target attribute (Please see section 2.2). Unlike other ML based methods which provide very good predictions, but sometimes are limited in terms of interpretability (black box models), the model trees tend to be more descriptive and interpretable (white box models) [15]. Therefore, specific objective of this study is to develop prediction models for Mn, Fe and NH_4 for seven days in advance that can be used for optimizing of the treatment processes of the Butoniga DWTP. The paper is organized as follows: Section 2 describes study area, measured data and modelling methods used in modelling exercise. Section 3 gives results, i.e. constructed prediction models for Mn, Fe and NH_4 with related discussion of the results, and finally Section 4 contains the conclusions of this research.

2 Experimental investigation

2.1 Study area and data description

DWTP Butoniga produces potable water from the Butoniga reservoir, which is located upstream from the DWTP (Figure 1). The Butoniga reservoir is an artificial lake created in year 1987 with two main objectives: i) protection from adverse water impacts, and ii) drinking water supply. Butoniga reservoir has a watershed area of about 73 km^2, ranging in elevation from 40 to 500 masl. The volume of the reservoir is 19.5 million m^3, while the surface area is around 2.5 km^2 with an average depth of 7.8 m and maximum depth of 17.5 m [10].

According to the above characteristics of the reservoir, Butoniga is a small and relatively shallow reservoir which is very sensitive to eutrophication and degradation processes caused by climate change and anthropogenic activities (human activities in the watershed). Known pressures in the surrounding watershed include erosion and leaching of nutrients (mainly nitrogen and phosphorus) from agricultural lands, as well as from untreated wastewater from settlements that are drained to the reservoir through black pits or open sewers [10].

Figure 1. Location of the Butoniga DWTP and Butoniga reservoir

Butoniga DWTP is located about 600 m downstream from the dam of the Butoniga reservoir on an area of 80.000 m^2 (Figure 1). First phase of Butoniga DWTP is designed to process 1000 l/s or 3600 m^3/h. Parts of the process are designed for a final capacity of 2000 l/s, which is planned in the second phase. All process units are designed for 24-hour full capacity with a hydraulic reserve of 25 %. The plant can operate flexibly by changing the capacity from 20 to 100 % of the nominal capacity. The main drinking water treatment process (Figure 2) consists of the following units: raw water intake, pre-ozonation, coagulation-flocculation, flotation, rapid filtration, main ozonation, slow sand filter, disinfection, final pH correction, pressure pumping and chlorination. The auxiliary process (Figure 2) of drinking water treatment consists of the following units: station for cleaning sand from slow sand filters, treatment of water from washing filters, sludge treatment and neutralization of wastewater from chemicals. The construction of the plant was finished and it became operational in June 2002. Since the spring of 2004, it has remained in continuous operation [11].

The operation of the DWTP is primarily related to the tourist season. Out of the total annual production and distribution of 5.000.000 m^3 of water, 3.000.000 m^3 is generated and distributed between June 15 and September 15. During this period, the water quality in the Butoniga reservoir is the worst [11]. As mentioned in introduction section main problem with Butoniga reservoir and thus DWTP appear in summer months when water temperature is the most critical parameter and water for the treatment process must be captured from the lowest layer which have increased concentrations of Mn, Fe, NH_4 and lower pH values and thus influence on treatment processes [11].

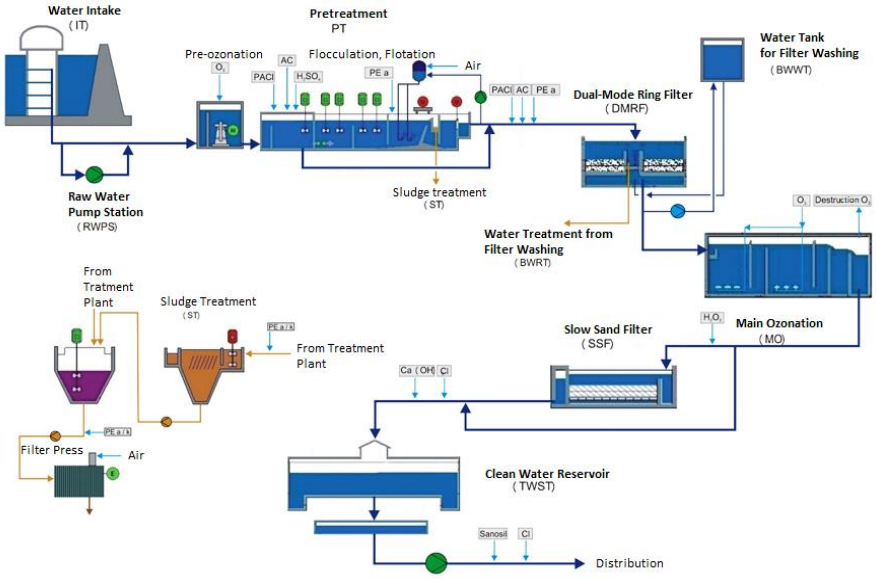

Figure 2. Treatment processes scheme for drinking water treatment plant Butoniga

The data set for building the prediction model of manganese, iron and ammonium (see Table 1) consists of physical and chemical parameters measured once a day at the inflow of raw water to the DWTP, from 2011 to 2020. Physiochemical parameters include water temperature in the reservoir (Temp), pH, turbidity (Tur), oxygen concentration (O_2), total organic carbon (TOC), potassium permanganate ($KMnO_4$), ammonia (NH_4), manganese (Mn), aluminium (Al), iron (Fe) and amount of organic substances (UV254) whose concentration were determinate in internal laboratory of the Butoniga DWTP by standard analytical methods according to ISO standards [16] and Standard methods for the examination of water and wastewater [17]. In addition to the data measured on the DWTP, for prediction models were also used reservoir water level data (Lake level) and data from nearby meteorological station which are obtained from Croatian Hydro-Meteorological Service (CHMS).

This data contain daily precipitations (Prec) and air temperatures (Air temp). Also, for better prediction of the models, sum of 30, 25, 20, 15, 10 and 5 days precipitations were used for modelling purposes which are significant from a hydrological aspect due to the precipitation concentration runoff. All the data were pre-processed regarding to modelling and research goals. For the prediction models the entire span of the measured data was used; from 2011 to 2020. Missing data were managed with a cubic spline interpolation due to best fit.

Table 1. Data used for prediction models

Symbol	Description	Unit
Temp	Water temperature	°C
O_2	Oxygen concentration	mg/l
pH	pH	-
Tur	Turbidity	NTU
TOC	Total organic carbon	mg/l
$KMnO_4$	Potassium permanganate	mg/l
UV254	Organic matter in water	1/cm
Al	Aluminium	mg/l
NH_4	Ammonium	mg/l
Mn	Manganese	mg/l
Fe	Iron	mg/l
Prec 30 days	30 days sum of precipitations	mm
Prec 25 days	25 days sum of precipitations	mm
Prec 20 days	20 days sum of precipitations	mm
Prec 15 days	15 days sum of precipitations	mm
Prec 10 days	10 days sum of precipitations	mm
Prec 5 days	5 days sum of precipitations	mm
Prec	Precipitations	mm
Air temp	Air temperature	°C
Lake level	Lake level	m
NH_4_train	**Ammonium values shifted 7 days in advance**	**mg/l**
Mn_train	**Manganese values shifted 7 days in advance**	**mg/l**
Fe_train	**Iron values shifted 7 days in advance**	**mg/l**

2.2 Modelling methods; model trees

Model trees are hierarchical structures composed of nodes and branches. Internal nodes contain tests on the input attributes while each branch of an internal test corresponds to an outcome of the test and the predictions for the values of the target variable (i.e. the class) are stored in the leaves that are the terminal nodes in the tree. If the leaf contain a single value for the class prediction, then it is talked about simple regression trees, while if a linear equation is used for prediction in the leaf, then it is talked about model trees [14], [18]. One of the most used algorithm for induction of model trees is the M5 algorithm [18], based on the TDIDT top-down induction of decision trees (TDIDT) algorithm [19]. For the experiments conducted in this research a variation of the M5 algorithm was used, called M5P, implemented in the software package Weka [14].

The M5P algorithm uses a decision tree as the base model, with the addition of linear regression models at the leaf nodes of the tree. This allows for more complex models that can better handle continuous target variables. The M5P algorithm works by first growing a decision tree based on the input data, and then fitting a linear regression model to the data at each leaf node. The final model is a combination of these decision tree and linear regression models, which can be used to make predictions on new data [14]. After the tree is constructed from the training (learning) set of data, it is necessary to assess the model quality, i.e., the accuracy of prediction. This can be done by simulating the model on a testing set of data and comparing the predicted values of the target with the actual values. Another option is to employ cross-validation. The given (training) data set is partitioned on a chosen number of folds (n), usually 10. In turn, each fold is used for testing, while the remainder (n-1 folds) is used for training. The final error is the average error of all the models throughout the procedure.

The size of the error between the actual and the predicted values can be calculated by several measures to evaluate the model accuracy: root mean-squared error (RMSE), mean absolute error (MAE), root relative squared error (RRSE), relative absolute error (RAE), and correlation coefficient (R) [14]. In the performed experiments the accuracy of the models is evaluated through all the measures of accuracy.

2.3 Design of the modelling experiments

The data used for building prediction models of manganese, iron and ammonium are depicted in Table 1. As mentioned, all data were pre-processed with respect to modelling and research goals based on the knowledge of modelling experts on the analysed DWTP. Models are built to predict concentrations of manganese, iron and ammonium seven days in advance with purpose to improve treatment processes on DWTP regarding to changes of raw water quality in the Butoniga reservoir. Therefore, these models would help to manage treatment processes which are depending on the quality of raw water in the Butoniga reservoir. For the experiments the machine learning algorithm M5P for induction of model trees integrated in the Weka modelling software [14] was used. Predicted concentrations of Mn, Fe and NH_4 seven days in advance were set as a target (dependent) variable for each build model, whereas water temperature, pH, turbidity, $KMnO_4$, NH_4, Mn, Al, Fe, O_2, TOC, UV254, Prec 30 days, Prec 25 days, Prec 20 days, Prec 15 days, Prec 10 days, Prec 5 days, Prec, Air temp and Lake level (Table 1) were set as independent variables (descriptors) from which the predicted values of Mn, Fe and NH_4 were modelled in separate models. The above parameters were mainly used because they best represent the parts of the system (DWTP and Butoniga reservoir) on top of which the target variable relays.

The aim of obtained prediction models is to be as much as possible applicable and valid for the prediction of Mn, Fe and NH_4, meaning that they should perform as accurately as possible. To achieve this, the most commonly used procedures of building and testing models were applied; the entire data set was taken for training while validating with 10-fold cross-validation (see Section 2.2). To achieve the highest correlation coefficient (R) and the optimal number of rules default values of parameters for building models were used in Weka modelling software [14]. The model performing most accurately according to the validation method was selected as a representative model for the prediction purposes. The accuracy of the models is evaluated through the RMSE, MAE, RRSE, RAE and R.

3 Results and discussion

As stated in the introduction section, the main problem concerning the Butoniga reservoir and the associated DWTP arises during the summer months when water temperature becomes the most critical parameter. During this time, water for the treatment process must be captured from the deepest layer of the reservoir, which exhibits increased concentrations of Mn, Fe, NH_4, lower pH values, and reduced oxygen concentrations. These factors significantly impact the treatment processes on the DWTP [11]. Figure 3 illustrates the measured concentrations of Mn, Fe, NH_4, along with pH values and oxygen concentrations, throughout the observed period from 2011 to 2020. It is evident from the data that all parameters exceed the Maximum Allowable Concentrations (MAC) for drinking water according to Croatian regulations [13]. The exceedance of limit values, such as 0.5 mg/l for NH_4, 0.05 mg/l for Mn, and 0.2 mg/l for Fe, is recorded throughout the whole year, with a more pronounced occurrence during the summer months, particularly during the tourist season. During this period, the exceedances of MAC are significantly higher, coinciding with lower pH values and oxygen concentrations, while the water temperature is higher than the rest of the year, as depicted in Figure 3.

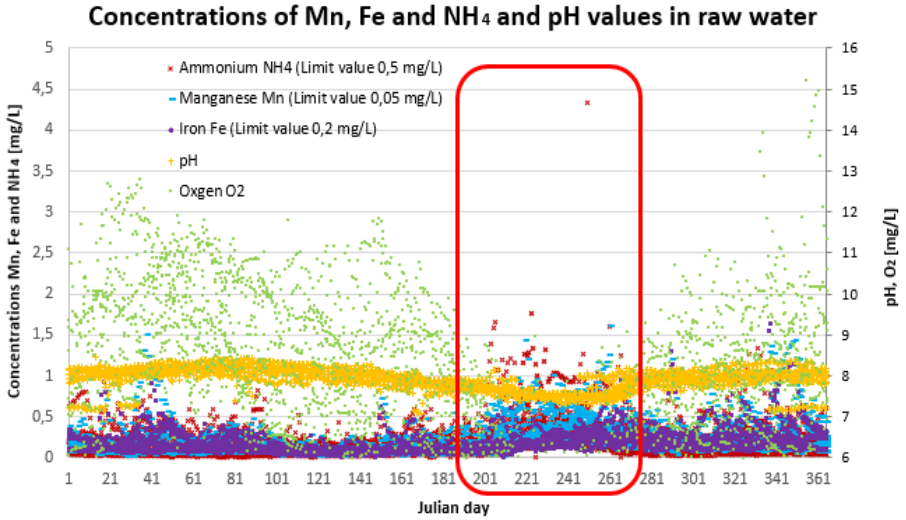

Figure 3. Concentrations of manganese, iron and ammonium in raw water along with pH values and oxygen concentrations

Prediction models of manganese, iron and ammonium concentrations for seven days in advance were built with Weka modelling software using M5P algorithm for induction of model trees [14]. As mentioned in section 2.3 for building prediction models entire span of the measured data was used, from 2011 to 2020 with daily sampling frequency. Missing data were managed by using a cubic spline interpolation, and selection of the best model was done according to the procedure mentioned in section 2.3. Prediction model for manganese is presented on Figure 4, while related model equations are given in Table 2. The model is composed of nine leaves, i.e. equations, were each equation is used to predict manganese concentrations seven days in advance using parameters given in model tree nodes (Figure 4). The model tree nodes in Figure 4 demonstrate that the most prevalent parameters are the current concentration of manganese (as anticipated), pH values, the cumulative sum of five days' precipitation, water temperature, and iron concentration. Table 2 contains the equations associated with the model tree leaves, including parameters such as temperature, pH values, manganese, iron, and the sum of five day precipitation.

The selection of equations in the model tree leaves is contingent upon the variable values in the tree nodes. Once the selection based on the variable values in the tree nodes is made, the corresponding equation is applied to predict the manganese concentration seven days ahead. For example, if manganese concentration (top node) is lower or equal than 0.232 (left side of the tree) and again manganese concentration is higher than 0.134 and pH value is higher than 7.873 than equation 6 is applied for manganese prediction; otherwise if pH value is lower or equal than 7.873 than equation 5 is applied.

The model has very high correlation coefficient (R) of 0.92, while MAE is 0.049, RMSE is 0.084, RAE of 33.71 % and RRSE of 41.12 %.

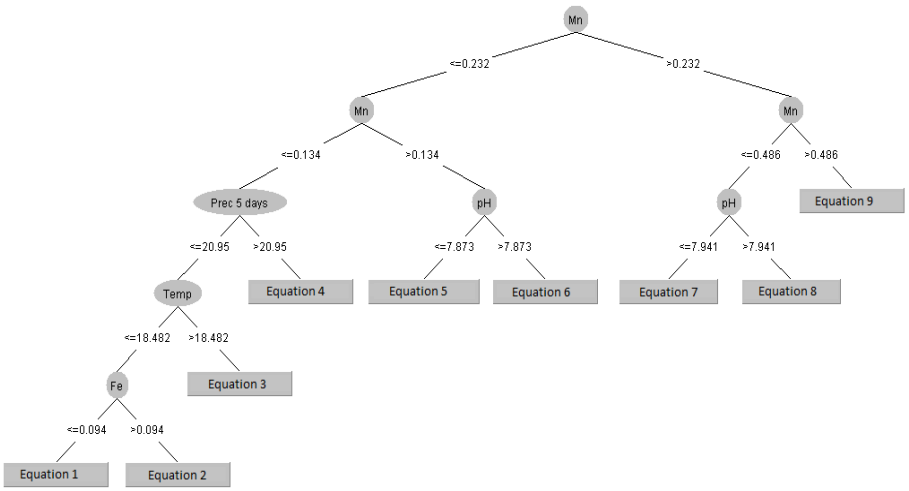

Figure 4. Model tree for manganese prediction 7 days in advance

Table 2. Equations for model tree presented on Figure 6 (manganese prediction, i.e. Mn pred)

Equation number	Equations
Equation 1	Mn_pred = 0.0001*Temp - 0.0018*pH + 0.0065*Mn + 0.0276*Fe + 0*Prec 5 days + 0.0742
Equation 2	Mn_pred = 0.0001*Temp - 0.0018*pH + 0.0065*Mn + 0.0273*Fe + 0*Prec 5 days + 0.1071
Equation 3	Mn_pred = 0.0002*Temp - 0.0018*pH + 0.0065*Mn + 0.0345*Fe + 0*Prec 5 days + 0.1365
Equation 4	Mn_pred = 0*Temp - 0.0018*pH + 0.0065*Mn + 0.022*Fe + 0.0001*Prec 5 days + 0.1447
Equation 5	Mn_pred = −0*Temp - 0.0157*pH + 0.0094*Mn + 0.0077*Fe + 0*Prec 5 days + 0.378
Equation 6	Mn_pred = −0*Temp - 0.007*pH + 0.0094*Mn + 0.0077*Fe + 0*Prec 5 days + 0.2279
Equation 7	Mn_pred = −0*Temp - 0.0108*pH + 0.0189*Mn - 0.0011*Fe + 0*Prec 5 days + 0.43
Equation 8	Mn_pred = −0*Temp - 0.015*pH + 0.0189*Mn - 0.0011*Fe + 0*Prec 5 days + 0.3625
Equation 9	Mn_pred = −0*Temp - 0.008*pH + 0.0298*Mn - 0.0011*Fe + 0*Prec 5 days + 0.6041

The prediction model for iron is presented on Figure 5, with related model equations in Table 3. Similar as the model for manganese, the iron prediction model is consisted of nine leaves, i.e. model equations were each equation is used to predict the iron concentrations for seven days in advance using parameters given in the model tree (Figure 5). Figure 5 reveals that the model tree's prominent parameters in the tree nodes are the current concentration of iron (as anticipated), pH values, the cumulative sum of five days' precipitation, turbidity, lake level, and oxygen concentration. Table 3 includes the equations associated with the model tree leaves, which involve parameters such as oxygen concentration, pH values, turbidity, manganese, iron, the sum of five day precipitation, and lake level.

The equation selection process remains the same as in the previous model. This model has also very high correlation coefficient (R) of 0.91, while MAE is 0.034, RMSE is 0.060, RAE of 34.82 % and RRSE of 42.52 %.

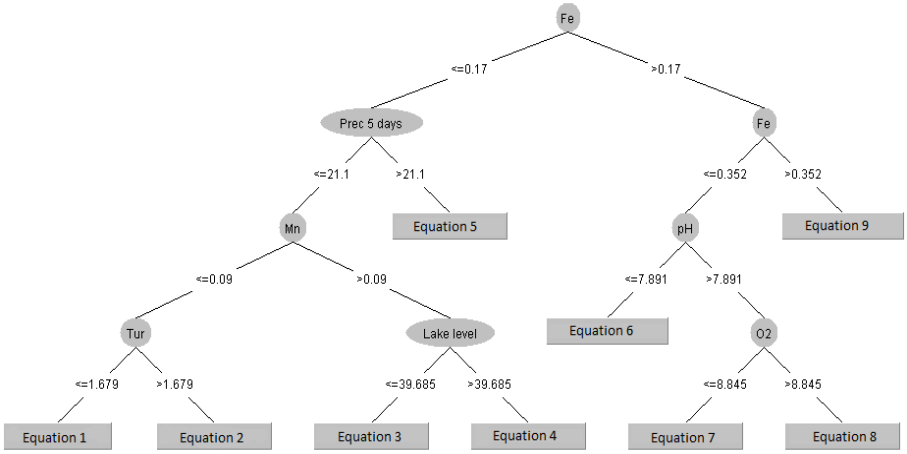

Figure 5. Model tree for iron prediction 7 days in advance

Table 3. Equations for model tree presented on Figure 7 (iron prediction, i.e. Fe_pred))

Equation number	Equations
Equation 1	Fe_pred = 0.0001*O_2 - 0.0006*pH + 0.001*Tur + 0.0022*Mn + 0.0037*Fe + 0 Prec 5 days - 0.0003*Lake level + 0.0853
Equation 2	Fe_pred = 0.0001*O_2 - 0.0006*pH + 0.0008*Tur + 0.0022*Mn + 0.0037*Fe + 0*Prec 5 days - 0.0003*Lake level + 0.1148
Equation 3	Fe_pred = 0.0001*O_2 - 0.0006*pH + 0.0005*Tur + 0.0025*Mn + 0.0037*Fe + 0*Prec 5 days - 0.0005*Lake level + 0.1703
Equation 4	Fe_pred = 0.0001*O_2 - 0.0006*pH + 0.0005*Tur + 0.0025*Mn + 0.0037*Fe + 0*Prec 5 days - 0.0008*Lake level + 0.1423
Equation 5	Fe_pred = 0.0001*O_2 - 0.0006*pH + 0.0005*Tur + 0.0028*Mn + 0.0037*Fe + 0*Prec 5 days - 0.0003*Lake level + 0.1624
Equation 6	Fe_pred = 0.0004*O_2 - 0.0054*pH - 0*Tur + 0.0004*Mn + 0.0104*Fe + 0*Prec 5 days - 0.0002*Lake level + 0.3082
Equation 7	Fe_pred = 0.0004*O_2 - 0.0043*pH - 0*Tur + 0.0004*Mn + 0.0104*Fe + 0*Prec 5 days - 0.0002*Lake level + 0.2567
Equation 8	Fe_pred = 0.0006*O_2 - 0.0043*pH - 0*Tur + 0.0004*Mn + 0.0104*Fe + 0*Prec 5 days - 0.0002*Lake level + 0.2712
Equation 9	Fe_pred = 0.0005*O_2 0.0052*pH - 0*Tur + 0.0004*Mn + 0.0246*Fe + 0*Prec 5 days - 0.0002*Lake level + 0.43

The prediction model for ammonium is given in Figure 6, while Table 4 contains related model equations for prediction of the ammonium concentration for seven days in advance. Similar as the prediction models for manganese and iron, the prediction model for the ammonium is consisted of nine leaves, i.e. model equations were each equation is used to predict the ammonium concentrations for seven days in advance using parameters given in model tree nodes (Figure 6). Figure 6 demonstrates that the model tree primarily features are the current concentration of ammonium (as anticipated), pH values, the cumulative sum of thirty days' precipitation, and manganese concentration as the most appearing parameters.

The related equations in Table 4 involve parameters such as pH values, ammonium, the sum of thirty day precipitation, and manganese concentration. The equation selection process remains the same as in the previous model. Model for NH_4 has very high correlation coefficient of (R) 0.92, while MAE is 0.047, RMSE is 0.087, RAE of 32.79 % and RRSE of 39.53 %.

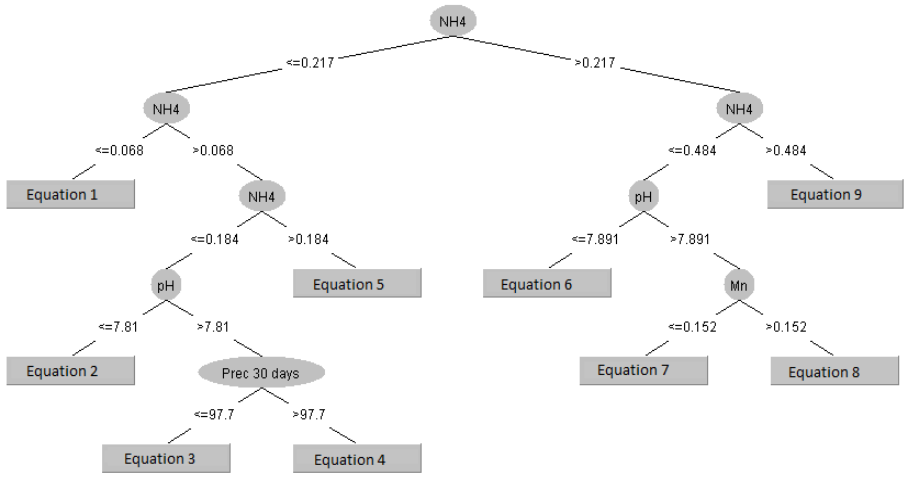

Figure 6. Model tree for ammonium prediction 7 days in advance

Table 4. Equations for model tree presented on Figure 7 (ammonium prediction, i.e. NH_4_pred)

Equation number	Equations
Equation 1	NH_4_pred = −0.0015*pH + 0.0285*NH_4 - 0*Prec 30 days + 0.0588
Equation 2	NH_4_pred = −0.0027*pH + 0.0258*NH_4 + 0*Prec 30 days + 0.1922
Equation 3	NH_4_pred = −0.0018*pH + 0.0258*NH_4 + 0*Prec 30 days + 0.1587
Equation 4	NH_4_pred = −0.0018*pH + 0.0258*NH_4 + 0*Prec 30 days + 0.1969
Equation 5	NH_4_pred = −0.0024*pH + 0.0429*NH_4 + 0*Prec 30 days + 0.2365
Equation 6	NH_4_pred = −0.0042*pH + 0.0119*NH_4 + 0.01*Mn - 0*Prec 30 days + 0.3968
Equation 7	NH_4_pred = −0.0033*pH + 0.0119*NH_4 + 0.032*Mn - 0*Prec 30 days + 0.2405
Equation 8	NH_4_pred = −0.0033*pH + 0.0119*NH_4 + 0.0194*Mn - 0*Prec 30 days + 0.3301
Equation 9	NH_4_pred = −0.0037*pH + 0.0271*NH_4 + 0.0088*Mn - 0*Prec 30 days + 0.609

The performance of the prediction models is presented in Figures 7 to 9. Each figure represents the modelled (predicted) vs. measured values of manganese (Figure 7), iron (Figure 8) and ammonium (Figure 9) for the given time period from 2011 to 2020 for seven days in advance. Figures 7 to 9 indicate very high accuracy of the prediction models for all predicted parameters with a relatively good prediction of the peak values included. Overall, obtained results i.e., prediction models are acceptable for optimizing the management of the Butoniga DWTP, looking at the correlation coefficients and prediction of the peak values. As such, the models can be used for prediction purposes and therefore for optimization of the treatment processes on the Butoniga DWTP.

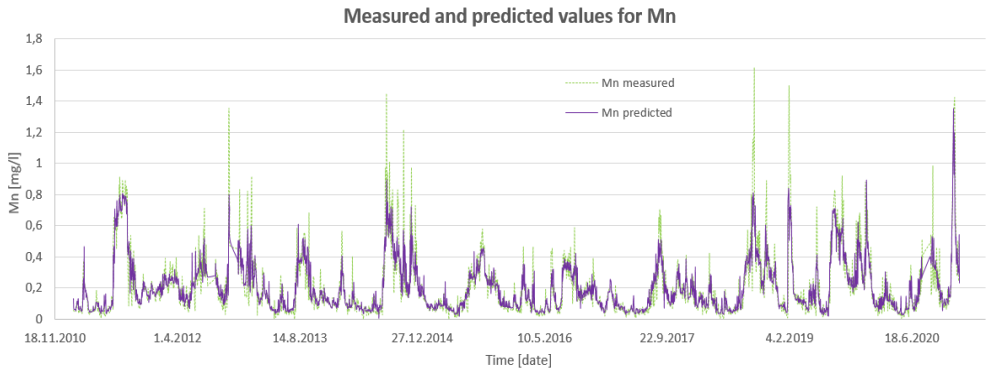

Figure 7. Comparison of measured and predicted values of manganese concentration for the modelled period

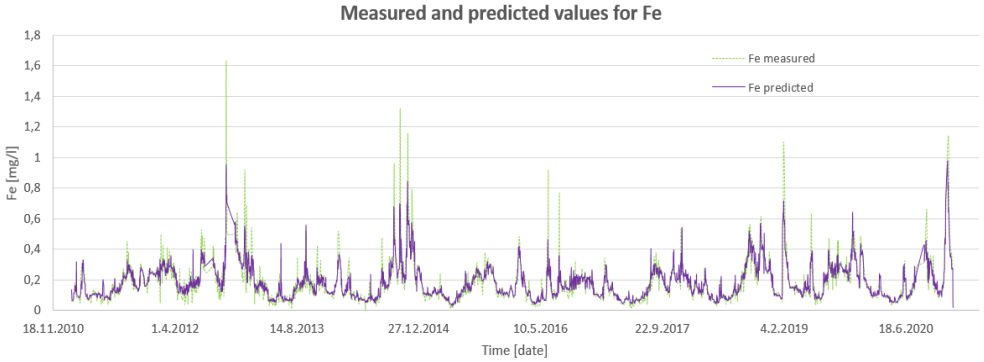

Figure 8. Comparison of measured and predicted values of iron concentration for the modelled period

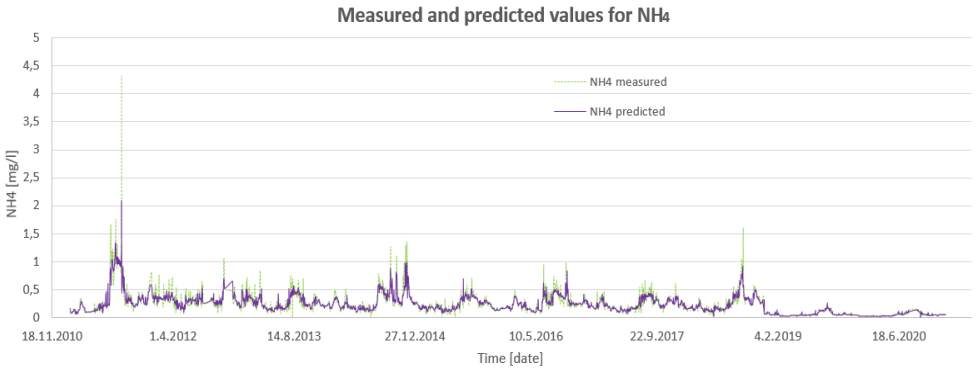

Figure 9. Comparison of measured and predicted values of ammonium concentration for the modelled period

In recent times, the rising complexity of water quality and more stringent drinking water standards have created challenges for the economic efficiency and operation of DWTPs. This has necessitated the development of more complex modelling and optimization techniques for treatment processes, as well as an overall proper management approach for DWTPs as a whole [1]. As mentioned earlier, traditional mathematical models and statistical methods have been commonly used for modelling DWTPs. However, these approaches often lack a comprehensive understanding of the overall dynamics and nonlinear relationships that are prevalent in DWTP systems. On the other hand, ML methods have the capability to handle complex nonlinear relationships and provide a more accurate understanding of the overall dynamics in drinking water treatment processes. ML methods enable monitoring of water quality evolution, analysis, and prediction of water quality, as well as revealing the migration and transformation processes of pollutants.

This allows for a shift in focus from merely solving existing problems to proactively identifying risks in advance and dynamically optimizing the facilities [1]. Today, the utilization of ML methods is becoming increasingly widespread in various aspects of DWTP modelling, including prediction, optimization, and facility management. Below are some examples showing the application of ML methods. Wang et al. [6] conducted research on the estimation of effluent turbidity in the drinking water flocculation process. They used an improved random forest (IRF) model that consisted of both long-term and short-term components.

The study demonstrated that the IRF model exhibited high flexibility and adaptability. In a separate study, Kim and Parnichkun [7] used a hybrid model combining k-means clustering and adaptive neuro-fuzzy inference system to predict settled water turbidity and determine optimal coagulant dosage in a drinking water treatment plant. Their approach showed improved prediction capability. Ayanshola et al. [8] utilized MLR and ANN models to predict treated water turbidity in a water treatment plant, achieving reasonable prediction accuracy with both methods. Alsaeed et al. [9] developed prediction models for turbidity and aluminium in drinking water treatment plants using GA-ANN and GEP which can be used as early warning systems to provide information about water treatment plants. Godo-Pla et al. [1] employed ANN to predict the oxidant demand in a full-scale drinking water treatment plant. This predictive capability is integrated into an environmental decision support system, which can be fed with real-time data to enhance and expedite the

decision-making process for plant managers and operators. Furthermore, recent research by Li et al. [3] has explored the use of AI and ML for analysing nonlinear relationships and process control in drinking water treatment. In summary, ML methods have demonstrated remarkable performance in capturing complex relationships and making predictions in DWTPs. These methods excel in processing data with nonlinear relationships that are challenging to fit using single mathematical models or traditional statistical methods [3].

In this study, ML methods in the form of model trees were used to develop prediction models for manganese, iron, and ammonium. Model trees employ linear equations within their terminal nodes, enabling numerical predictions. This method is straightforward to use and can be very simple integrated into a DSS for the DWTP. The model trees' "decision" nodes incorporate parameters that are significant for the predicted variable, providing insights into the dependencies of the target variable on specific parameters. Unlike other ML based methods that offer highly accurate predictions but may lack interpretability (black box models", model trees aim to be more descriptive and interpretable (white box models) [15]. The prediction models for manganese, iron, and ammonium developed in this study are an extension of the research conducted by Volf et al. [12].

The previous research introduced and described the WQI for the Butoniga DWTP, along with prediction models for the WQI, which were used to optimize treatment processes at the DWTP. As the Butoniga DWTP faces challenges primarily related to increased concentrations of manganese, iron, and ammonium throughout the year, with a particular focus on the summer months during the tourist season, as described by Zorko [11], this current research is primarily focused on predicting manganese, iron, and ammonium concentrations. The aim of this prediction models is to be able to respond promptly and effectively manage the Butoniga DWTP. During this critical period characterized by increased concentrations of manganese, iron, and ammonium, enhanced and continuous process control is required, along with higher chemical consumption, to maintain stability in the treatment process and to ensure that all water samples in the effluent remain below the MAC [13].

4 Conclusion

The importance of DWTP which produce potable water for human consumption is nowadays very significant. Therefore, a model that represents functioning and dynamics of the DWTP becomes very useful and important tool for developing strategies and management that improve its behaviour and use of resources through proper optimization of the treatment processes and also with minimizing the risks of an inadequate actions. In this research prediction models for Mn, Fe and NH_4 concentrations in raw water for seven days in advance were built with use of ML method in form of the model trees, which was applied on measured data for the Butoniga DWTP. Models were built to cope with problem of high concentrations of Mn, Fe and NH_4, which occurs during summer months, when water is captured from the lowest water intake, i.e. the lowest water layer in the Butoniga reservoir. Predictions of the Mn, Fe and NH_4 are done according to current values of measured parameters at the intake of the raw water. Obtained models have high correlation coefficients and thus provide accurate predictions of the Mn, Fe and NH_4, including the predictions of the peak values when compared to measured data. As such, the models can help with the optimization and management of treatment processes at the DWTP, especially during the summer months (tourist season), when the quality of raw water in the Butoniga reservoir is poorest, and where changes in the raw water quality can result in direct action and optimization of the operation of the Butoniga DWTP.

Acknowledgements

This study received support from several projects. The first project, titled "Sustainable river basin management by implementation of innovative methodologies, approaches, and tools" (uniri-tehnic-18-129), and the second project, titled "Hydrology of water resources and identification of flood and mudflow risk in karst" (uniri-tehnic-18-54), were funded by the University of Rijeka. Additionally, the Croatian Science Foundation provided funding for the project "Influence of summer fire on soil and water quality" (IP-2018-01-1645). The research also falls under the ZIP UNIRI project line of the University of Rijeka, specifically ZIP-UNIRI-1500-3-22 and ZIP-UNIRI-1500-2-22. Furthermore, this work is part of Morana Krbavčić's (second author) undergraduate thesis, titled "Drinking water treatment and quality - an example of a drinking water treatment plant Butoniga". Finally, a portion of the research conducted in this manuscript was presented at the XVI International Conference on Durability of Building Materials and Components-DBMC 2023.

References

[1] L. Godo-Pla, P. Emiliano, F. Valero, G. Sin and H. Monclus, "Predicting the oxidant demand in full-scale drinking water treatment using an artificial neural network: Uncertainty and sensitivity analysis," *Process Saf. Environ. Prot.*, vol. 125, pp. 317-327, 2019, https://doi.org/10.1016/j.psep.2019.03.017.

[2] H. Maier, "Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters," *Environ. Modell. Software*, vol. 19. Pp. 485-494, 2004, https://doi.org/10.1016/S1364-8152(03)00163-4.

[3] L. Li, S. Rong, R. Wang and S. Yu, "Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: A review," *J. Chem. Eng.*, vol. 405, 2021, https://doi.org/10.1016/j.cej.2020.126673.

[4] J. D. Olden and D. A. Jackson, "Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks," *Ecol. Model*., vol. 154, pp. 135-150, 2002, https://doi.org/10.1016/S0304-3800(02)00064-9.

[5] S. Park, S. S. Baek, J. Pyo, Y. Pachepsky, J. Park and K. H. Cho, "Deep neural networks for modeling fouling growth and flux decline during NF/RO membrane filtration," *J. Membr. Sci*., vol. 587, 2019, https://doi.org/10.1016/j.memsci.2019.06.004.

[6] D, Wang, X. Chang, K. Ma, Z. Li and L. Deng, "Estimating effluent turbidity in the drinking water flocculation process with an improved random forest model," *Water Supply*, vol . 22, pp. 1107-1119, 2022, https://doi.org/10.2166/ws.2021.213.

[7] C. M. Kim and M. Parnichkum, "Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of k-means clustering and adaptive neuro-fuzzy inference system," *Appl Water Sci*., vol. 7, pp 3885-3902, 2017, DOI 10.1007/s13201-017-0541-5.

[8] A. M. Ayanshola, A. A. Alao, A. W. Salami, S. O. Bilewu, A. A. Mohammed, O. O. Adeleke and O. O. Olofintoye, "Modelling of turbidity variation in a water treatment plant," *Acta Tech Corvin., Bull. Eng*., vol. 4, pp. 41-44, 2021.

[9] R. Alsaeed, B. Alaji and M. Ebrahim, "Predicting turbidity and Aluminum in drinking water treatment plants using Hybrid Network (GA- ANN) and GEP," *Drink. Water Eng. Sci*., https://doi.org/10.5194/dwes-2021-8.

[10] B. Hajduk Černeha, "Akumulacija Butoniga u Istri - Prva iskustva u korištenju za vodoopskrbu," *Proceedings of Vodni dnevi*, 2021, Rimske Toplice, Slovenia, 7.-8. October 2021.

[11] S. Zorko, "Akumulacija Butoniga - pritisci u slijevu i zaštita voda," *Zbornik radova-Upravljanje jezerima i akumulacijama u Hrvatskoj i Okrugli stol o aktualnoj problematici Vranskog jezera kod Biograda na Moru*, 2017, Biograd na Moru, Croatia, 4.-6. May 2017.

[12] G. Volf, I. Sušanj Čule, E. Žic, and S. Zorko, "Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant," *Sustainability*, vol. 14, 2022, https://doi.org/10.3390/su141811481.

[13] Croatian regulations for drinking water, "Pravilnik o parametrima sukladnosti, metodama analize, monitoringu i planovima sigurnosti vode za ljudsku potrošnju te načinu vođenja registra pravnih osoba koje obavljaju djelatnost javne vodoopskrbe," available online https://narodne-novine.nn.hr/clanci/sluzbeni/2017_12_125_2848.html. [Accessed: 21 March 2023].

[14] I. H. Witten and E. Frank, „Data Mining: Practical Machine Learning Tools and Techniques," 2nd ed., *Elsevier*, 2005.

[15] G. Volf, N. Atanasova, B. Kompare, R. Precali and N. Ožanić, „Descriptive and prediction models of phytoplankton in the northern Adriatic," *Ecol. Model*., vol. 222, pp 2502-2511, 2011.

[16] Technical Committees, ISO/TC, *Water quality*. Available online: https://www.iso.org/committee/52834/x/catalogue F. [Accessed 4 Feb. 2022].

[17] R. B. Baird, A. D. Eaton and E. W. Rice, „Standard Methods for the Examination of Water and Wastewater," 23rd ed., *AWWA*, USA, 2017.

[18] J. R. Quinlan, „C4.5: Programs for Machine Learning," *Morgan Kaufmann*, 1993.

[19] J. R. Quinlan, „Induction of decision trees," *Machine Learning*, vol 1, pp. 81-106, 1986.