

# Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant

---

Volf, Goran; Sušanjan Čule, Ivana; Žic, Elvis; Zorko, Sonja

Source / Izvornik: **Sustainability**, 2022, 14

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.3390/su141811481>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:157:358175>

Rights / Prava: [Attribution 4.0 International](#) / [Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-11-20**



image not found or type unknown

Repository / Repozitorij:




[Repository of the University of Rijeka, Faculty of Civil Engineering - FCERI Repository](#)



image not found or type unknown

## Article

# Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant

Goran Volf <sup>1,\*</sup> , Ivana Sušanĳ Čule <sup>1</sup> , Elvis Žic <sup>1</sup>  and Sonja Zorko <sup>2</sup>

<sup>1</sup> Department of Hydraulic Engineering, Faculty of Civil Engineering, University of Rijeka, 51000 Rijeka, Croatia

<sup>2</sup> Istarski Vodovod d.o.o., 52420 Buzet, Croatia

\* Correspondence: goran.volf@uniri.hr

**Abstract:** In order to improve the treatment processes of the drinking water treatment plant (DWTP) located near the Butoniga reservoir in Istria (Croatia), a prediction of the water quality index (WQI) was done. Based on parameters such as temperature, pH, turbidity, KMnO<sub>4</sub>, NH<sub>4</sub>, Mn, Al and Fe, the calculation of WQI was conducted, while for the WQI prediction models, along with the mentioned parameters, O<sub>2</sub>, TOC and UV254 were additionally used. Four models were built to predict WQI with a time step of one, five, ten, and fifteen days in advance, in order to improve treatment processes of the DWTP regarding the changes in raw water quality in the Butoniga reservoir. Therefore, obtained models can help in the optimization of treatment processes, which depend on the quality of raw water, and overall, in the sustainability of the treatment plant. Results showed that the obtained correlation coefficients for all models are relatively high and, as expected, decrease as the number of prediction days increases; conversely, the number of rules, and related linear equations, depends on the parameters set in the WEKA modelling software, which are set to default settings which give the highest values of correlation coefficient (R) for each model and the optimal number of rules. In addition, all models have high accuracy compared to the measured data, with a good prediction of the peak values. Therefore, the obtained models, through the prediction of WQI, can help to manage the treatment processes of the DWTP, which depend on the quality of raw water in the Butoniga reservoir.

**Keywords:** water quality index; prediction models; machine learning; water quality; treatment processes improvement; drinking water treatment plant; Butoniga reservoir



**Citation:** Volf, G.; Sušanĳ Čule, I.; Žic, E.; Zorko, S. Water Quality Index Prediction for Improvement of Treatment Processes on Drinking Water Treatment Plant. *Sustainability* **2022**, *14*, 11481. <https://doi.org/10.3390/su141811481>

Academic Editors: Elena Cristina Rada and Agostina Chiavola

Received: 6 July 2022

Accepted: 10 September 2022

Published: 13 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Nowadays, because of increasing demand for potable water, problems concerning water quantity and quality are very important, and, indeed, a crucial issue for urban sustainable development and ecological environment [1]. During significant climate changes and other (mainly human) impacts which currently influence on the quality, quantity and availability of water and water resources, their proper management is needed. As such, an adequate supply of safe potable water is vital to human health and survival [2–4].

Water quality is mainly assessed using traditional water quality assessment methods that measure chemical parameters against established standards [5]. The water quality index (WQI) is used worldwide for water quality assessment, and to cope with big and complex water quality data. Moreover, the WQI method is most often used because this method's main objective is to turn complex water quality data into understandable and usable information [6]. Today, various versions of WQI methods are widespread and are receiving more and more attention all over the world for representing and modeling water quality data [7]. The WQI method is also a useful tool to manage and/or reveal the performance and sustainability of the drinking water treatment plant (DWTP) [7,8]. To provide the best treatment efficiency so that the treatment processes are determined

by prioritizing water quality in a more prominent way in the preferred water source, DWTPs are constantly changing both in terms of easy operation and functionality. The most preferred DWTP, which can treat all types of water with a certain quality, regardless of the water source, is called the conventional DWTP [7]. With these plants, the chemical and microbiological parameters in water can be treated with sufficient quality. As such, according to this, treatment plant strategies should be developed considering the reduction of human health risks and the improvement of water quality after water purification on DWTP. Moreover, it is important to determine and assess the parameters affecting water quality [7].

In this research, the WQI method was used to reveal the performance of the Butoniga DWTP located in Istria (Croatia), together with use of machine learning (ML) tools, if necessary, to help in the optimization of certain treatment processes which are dependent on the quality of raw water, and overall, for the sustainability of the DWTP. Calculation of WQI is based on a number of physical, chemical, and bacteriological parameters [6]. Detailed calculation of WQI is presented in Section 3.2.

A previous modelling attempt regarding a similar problem was done in [7], where the effluent water quality of a DWTP in Turkey according to WQI and health risk assessment (HRA) was investigated using principal component analysis (PCA). Other authors, such as [9–16], used in their research various ML algorithms and artificial intelligence (AI) methods in predicting WQI. Authors in [9] developed a model for predicting WQI using artificial neural networks (ANNs), namely, the nonlinear autoregressive neural network (NARNET) and the long short-term memory (LSTM) deep learning algorithm. Improvement of WQI prediction was done in research conducted in [10], where prediction was done using ML algorithms such as multilayer perceptron (MLP), convolutional neural network (CNN) and LSTM. In other research [11], prediction of WQI was done with ML techniques such as random forest (RF), neural networks (NN), multinomial logistic regression (MLR), support vector machine (SVM), and bagged tree models (BTM). In other research [12], eight artificial intelligence algorithms were used, e.g., MLR, RF, M5P tree, random subspace (RSS), additive regression (AR), ANNs, support vector regression (SVR), and locally weighted linear regression (LWLR) to generate WQI prediction. In another study [13], WQI prediction was done using supervised ML, e.g., MLP, Gaussian naïve Bayes (GNB), MLR, stochastic gradient descent (SGD), k-nearest neighbors (KNN), decision tree (DT), RF, SVM, gradient boosting classifier (GBC) and bagging classifier (BC). Moreover, various ML models such as boosting-based algorithms, DT-based algorithms and ANN-based algorithms were used for predicting WQI in research [14]. ANNs were also used for the WQI model in research [15], while in research [16], simple prediction of WQI was done through supervised ML, e.g., simple linear regression (SLR), MLR, and ridge and lasso regression (RLR). This confirms the fact that the WQI method is used worldwide for water quality assessment.

Previous studies regarding functioning and problems on Butoniga DWTP were made in studies [17,18]. The first study, [17], describes the first experiences of the use of Butoniga DWTP for drinking water supply. In addition, this study analyzed raw water quality data from the Butoniga reservoir, and some management guidelines regarding the Butoniga reservoir and the related DWTP were given. In the second study, [18], the impact of the Butoniga reservoir raw water quality on water treatment was considered. This study provided some interesting conclusions, mentioning that the main problem with the Butoniga reservoir, and thus the related DWTP, appears in the summer months when water temperature is the most critical parameter, because, in order to be suitable for use and for treatment processes, water must not exceed the maximum allowable concentration (MAC) of 25 °C, according to Croatian regulations for drinking water [19]. In that time period, water is captured from the lowest water intake, which captures water from the lowest water layer in the Butoniga reservoir, which have increased concentrations of manganese (Mn), iron (Fe), ammonium (NH<sub>4</sub>) and lower pH values. Increased concentration of Mn, Fe, NH<sub>4</sub> and lower pH values of water from the lowest water intake requires enhanced continuous process control and higher consumption of chemicals for the treatment process on DWTP;

however, the process is also stable in these conditions, and all samples of the effluent are in accordance with the Croatian regulations for drinking water [19], while the exceeded temperature values due to the heating of untreated water from the reservoir cannot be influenced [18]. This is also the main reason why this research was conducted.

To cope with these problems, the ML technique was applied on data measured at the intake of raw water from the Butoniga reservoir, which is after the purification process of the DWTP used for drinking water purposes. The first step was to calculate WQI from a long-term dataset (2011–2020) consisting of physical and chemical parameters measured once a day, to obtain usable and understandable information about the water quality of the Butoniga DWTP. Second, based on calculated WQI, the rule-based ML method in the form of rule-based models was applied to predict WQI with a time step of one, five, ten and fifteen days in advance, in order to improve specific treatment processes on the DWTP regarding changes in raw water quality in the Butoniga reservoir. The rule-based models for numeric prediction use regression equations in the terminal nodes which allow a more accurate prediction of the target attribute. The models are interpreted as a set of IF THEN rules, where each rule is associated with a multivariate linear model (equation). Unlike other AI-based methods, which provide very good predictions, but sometimes are limited in terms of interpretability (black box models), the rule-based models tend to be more descriptive and interpretable (white box models) [20]. Therefore, the specific objective of this study is to obtain WQI prediction models that can be used in the performance and optimization of treatment processes which are dependent on the quality of raw water, and overall, for the sustainability of the Butoniga DWTP.

The paper is organized as follows: Section 2 describes the study area and measured data used in the modelling exercise. Section 3 describes the modelling method, calculation of WQI and description of the modelling experiment for the modelling task. In Section 4, the results are presented, i.e., constructed prediction models. Discussion of the results is presented in Section 5, and finally, Section 6 contains the conclusions of this research.

## 2. Study Area and Data Description

For the process of obtaining drinking water on the DWTP, raw water was taken from the Butoniga reservoir (Figure 1). The Butoniga reservoir is an artificial lake created in 1987 with two main objectives: I) protection from adverse water impacts, and II) drinking water supply. It has a watershed area of about 73 km<sup>2</sup>, ranging in elevation from 40 to 500 masl. The volume of the reservoir is 19.5 million m<sup>3</sup>, the surface area about 2.5 km<sup>2</sup>, with an average depth of 7.8 m and a maximum depth of 17.5 m.

As a small and relatively shallow reservoir, it is very sensitive to eutrophication and degradation processes caused by climate change and human activities. Known and conventional pressures in the surrounding watershed include erosion and leaching of nutrients from agricultural lands, as well as from untreated wastewater from settlements that are drained to the reservoir through black pits or open sewers [17].

The Butoniga DWTP is located about 600 m downstream from the dam of the Butoniga reservoir, on an area of 80,000 m<sup>2</sup> (Figure 1). The first phase of the Butoniga DWTP is designed to process 1000 L/s or 3600 m<sup>3</sup>/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-h 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 plant was completed and put into operation in June 2002, while it has been in continuous operation since the spring of 2004 [18].

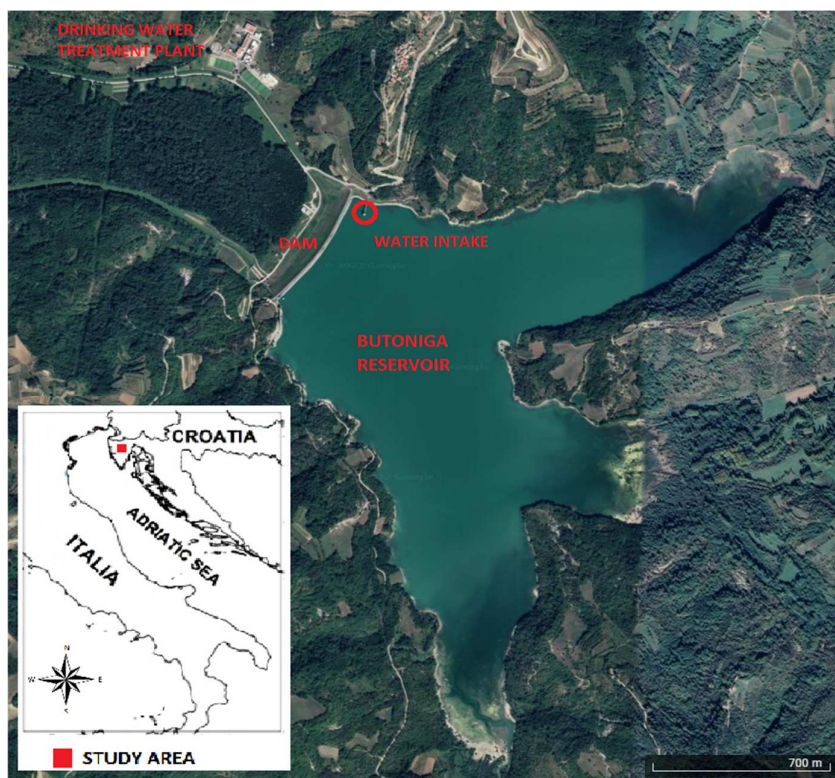


Figure 1. Study area; location of the Butoniga reservoir and the drinking water treatment plant.

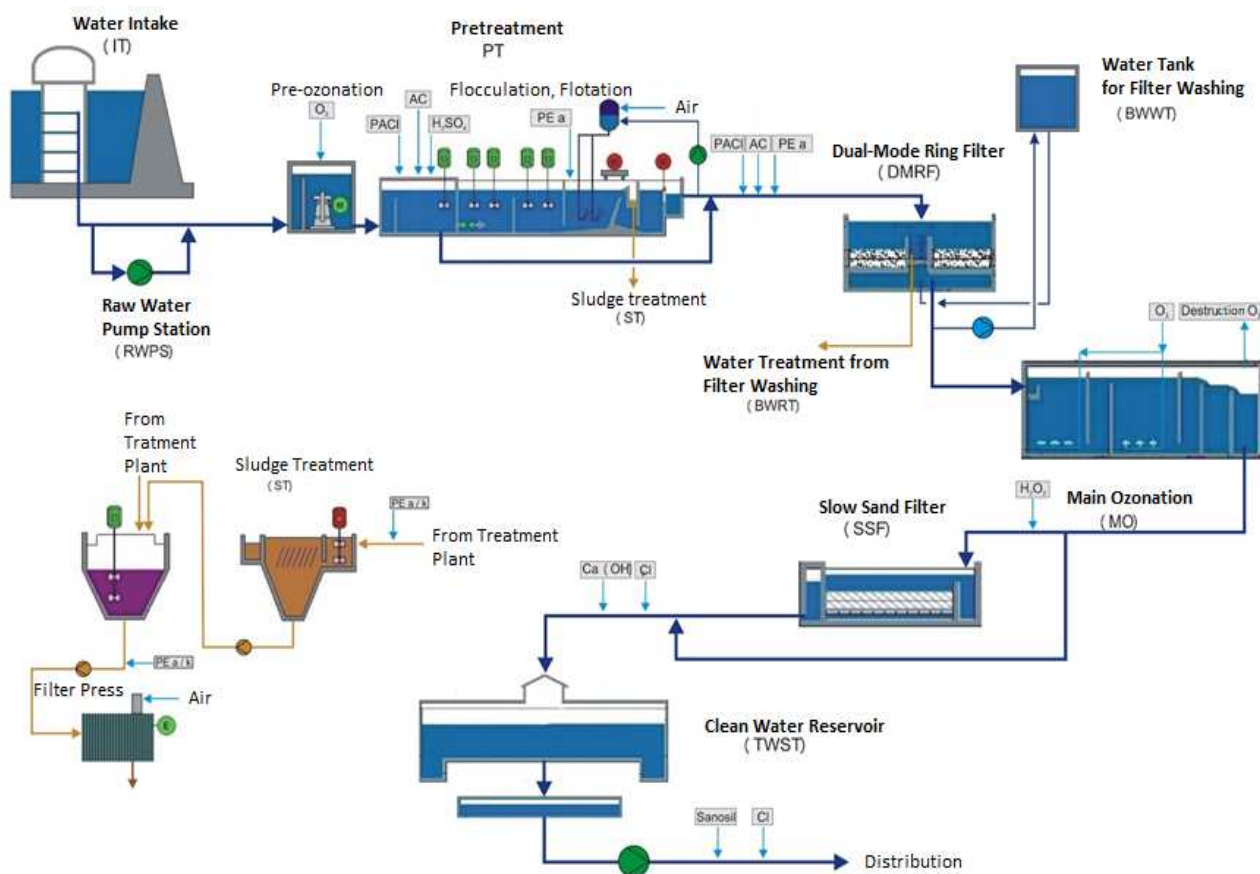


Figure 2. Butoniga drinking water treatment plant processes scheme [18].

The operation of the DWTP is mainly related to the tourist season, and, of the 5,000,000 m<sup>3</sup> of produced and distributed water annually, 3,000,000 m<sup>3</sup> is produced and distributed during 15 June to 15 September, when the water quality in the Butoniga reservoir is the worst [18].

As mentioned in the introduction, the main problem with the Butoniga reservoir, and thus the DWTP, appears in the summer months when the 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<sub>4</sub> and lower pH values, and thus influence the treatment processes [18].

The data set for the calculation and prediction of WQI consists of physical and chemical parameters measured once a day at the input of raw water to the DWTP, from 2011 to 2020. Physiochemical parameters include temperature (Temp), pH, turbidity (Tur), oxygen concentration (O<sub>2</sub>), total organic carbon (TOC), potassium permanganate (KMnO<sub>4</sub>), ammonia (NH<sub>4</sub>), manganese (Mn), aluminum (Al), iron (Fe) and the amount of organic substances (UV254), whose concentrations are determined in the internal laboratory of the Butoniga DWTP by standard analytical methods according to ISO standards [21] and standard methods for the examination of water and wastewater [22].

For the calculation of WQI, the following parameters were used: temperature, pH, turbidity, KMnO<sub>4</sub>, NH<sub>4</sub>, Mn, Al and Fe; while for the prediction of WQI, along with the other parameters, oxygen concentration, total organic carbon and UV254 were additionally used.

The data were pre-processed regarding to the modelling and research goal. For the WQI prediction models, the entire span of the measured data was used, from 2011 to 2020, while missing data were managed with a cubic spline interpolation.

### 3. Materials and Methods

#### 3.1. Modelling Methods; Rule-Based Models

Rule-based regression models for numeric prediction are yet another model representation, where the models are interpreted as a set of IF THEN rules, where each rule is associated with a multivariate linear model. A rule indicates that, whenever a case satisfies all the conditions, the linear model is appropriate for predicting the value of the target attribute. The algorithms for rule induction mostly represent different variations of the M5 algorithm. The algorithm implemented in a software package WEKA [23] was applied for modelling, in which the variation of the M5 algorithm (M5P) was enhanced by combining model-based and instance-based learning [24].

After the rule-based models are 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 a cross-validation method. The given (training) dataset is partitioned into a chosen number of folds (n). In turn, each fold is used for testing, while the remainder (n−1 folds) is used for training. The final error is the averaged 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, mean absolute error, root relative squared error, relative absolute error, and correlation coefficient (R) [23].

In this research, for the experiments, the accuracy of the prediction models was evaluated through the correlation coefficient, and the quality of the models was done using the cross-validation method in the software package WEKA [23].

#### 3.2. Water Quality Index

To calculate WQI, Croatian drinking water quality standards and recommendations were used [19], so that the resulting WQI represents water appropriateness for use in the drinking water supply. The WQI was conducted in three steps.

In the first step, each of the 8 parameters is assigned weight ( $w_i$ ) according to its relative importance in the total quality of the drinking water (Table 1). A maximum weight

of 5 is assigned to Mn, Al and Fe as a parameter, which has great impact in water quality assessment for the Butoniga reservoir and for drinking water production [18]. Furthermore, temperature, oxygen concentration and UV254 were assigned a minimum weight of 1, because these parameters are not harmful or toxic at the expected concentrations for drinking water production [18].

**Table 1.** Relative weight of each parameter.

Parameter	Limit Values [19]	Water Quality Index (WQI)	
		$w_i$	$W_i$
Temperature	25 °C	1	0.034
pH	8	3	0.103
Turbidity	4 NTU	2	0.069
KMnO <sub>4</sub>	5 mg/L	4	0.138
NH <sub>4</sub>	0.5 mg/L	4	0.138
Mn	0.05 mg/L	5	0.172
Al	0.2 mg/L	5	0.172
Fe	0.2 mg/L	5	0.172

In the second step, relative weight ( $W_i$ ) is calculated from the formula:

$$W_i = \frac{w_i}{\sum_i^n w_i}, \quad (1)$$

where  $W_i$  is relative weight;  $w_i$  is the weight of each parameter; and  $n$  is the number of parameters.

Calculated values of the relative weights ( $W_i$ ) of each parameter are shown in Table 1.

In the third step, the quality based on concentration of the  $i$ th parameter ( $q_i$ ) is determined by dividing its concentration in each water sample and its corresponding limit values according to the Croatian guidelines for drinking water quality standards and recommendations [19], and the result is multiplied by 100.

$$q_i = \frac{C_i}{S_i} \times 100, \quad (2)$$

where  $q_i$  is the rating based on concentration of  $i$ th parameter;  $C_i$  is the concentration of each analyzed parameter expressed in mg/L; and  $S_i$  is the recommendation of Croatian legislation for each analyzed parameter expressed in mg/L.

To calculate WQI, first, the sub-index for each analyzed parameter ( $SI_i$ ) used for determination of WQI is determined.

$$SI_i = W_i \times q_i, \quad (3)$$

$$WQI = \sum_{i=1}^n SI_i, \quad (4)$$

where  $SI_i$  is the sub-index of each analyzed parameter;  $q_i$  is the rating based on concentration of  $i$ th parameter; and  $n$  is the number of parameters.

If the WQI is calculated with a smaller number of indicators (less than 12), as in this case, then the sum of sub-indices of all parameters are divided by the sum weighting factors of available analyzed parameters.

Based on the calculated WQI values, tested water samples are classified into five categories of “excellent water” to “unfit for drinking” (Table 2) [25].

**Table 2.** Water quality classification based on WQI [25].

WQI	Water Quality
<50	Excellent
50–100	Good
100–200	Bad
200–300	Very bad
>300	Unfit for drinking

### 3.3. Description of the Modelling Experiment

The data used for building WQI prediction models are depicted in Table 3. All data were pre-processed with respect to modelling and research goals based on the knowledge of modelling experts on the analyzed DWTP. Four models were built to predict WQI one, five, ten and fifteen days in advance, with the purpose to improve treatment processes on DWTP regarding changes of raw water quality in the Butoniga reservoir. Therefore, these models would help to manage treatment processes which are dependent on the quality of raw water in the Butoniga reservoir.

**Table 3.** Physico-chemical data used for building models.

Symbol	Interpretation	Unit
Temp	Water temperature	-
pH	pH	-
Turb	Turbidity	NTU
KMnO <sub>4</sub>	Potassium permanganate	mg/L
NH <sub>4</sub>	Ammonium	mg/L
Mn	Manganese	mg/L
Al	Aluminum	mg/L
Fe	Iron	mg/L
O <sub>2</sub>	Oxygen concentration	mg/L
O <sub>2</sub> p	Oxygen saturation	%
TOC	Total organic carbon	mg/L
UV254	Organic matter in water	1/cm
WQI	Calculated water quality index	-
WQI_1_day_pred	WQI 1 day prediction	-
WQI_5_day_pred	WQI 5 days prediction	-
WQI_10_day_pred	WQI 10 days prediction	-
WQI_15_day_pred	WQI 15 days prediction	-

For the experiment, the ML algorithm M5P for rule-based models integrated in the WEKA modelling software [23] was used. Predicted WQI (one, five, ten and fifteen days) was set as a target (dependent) variable, whereas water temperature, pH, turbidity, KMnO<sub>4</sub>, NH<sub>4</sub>, Mn, Al, Fe, O<sub>2</sub>, TOC and UV254 (Table 3) were set as independent variables (descriptors) from which the predicted values of WQI were modelled. The above parameters were mainly used because they best represent the parts of the system (DWTP) on top of which the target variable relays.

The aim of all prediction models is to be as much as possible applicable and valid for the prediction of WQI, meaning that they should perform as accurately as possible. To achieve this, the most commonly used procedures of building and testing models was applied; the entire data set was taken for training while validating with 10-fold cross-validation (see Section 3.1). 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 [23].

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 correlation coefficient (R).



#### 4. Results

The models for predicting WQI with a time step of one, five, ten and fifteen days in advance were built with the M5P algorithm for induction of rule-based models, which was implemented in the software package WEKA [23]. For the prediction models, the 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 to select the best model, the model selection procedure described in Section 3.1 was used.

First, the built prediction model with a time step of one day is presented in Table 4. The model has only one rule/equation with a very high correlation coefficient (R) of 0.93 when tested with 10-fold cross validation. To predict WQI, the following parameters are needed at present: O<sub>2</sub>, pH, turbidity, UV254, NH<sub>4</sub>, Mn, Al, Fe and value of WQI calculated from present data. Using this data in the prediction model, the WQI with a time step of one day can be obtained.

**Table 4.** Prediction model for WQI with a time step of one day (units for the parameters used are reported in Table 3).

Rule No.	Rule	Equation
1.	-	$\text{WQI}_{1\_day\_pred} = -2.1437 \times \text{O}_2 - 6.613 \times \text{pH} - 0.6083 \times \text{Tur} + 78.459 \times \text{UV254} + 20.5083 \times \text{NH}_4 + 72.6239 \times \text{Mn} + 168.8286 \times \text{Al} + 74.7368 \times \text{Fe} + 0.5695 \times \text{WQI} + 76.5001$

The second prediction model with time step of five days has twelve rules/equations (Table 5), with also a very high correlation coefficient (R) of 0.81 when tested with 10-fold cross validation. To predict WQI, all parameters presented in Table 3 are needed at present. The rule selection depends on the values of the variables in the rule (second column in Table 5). When a rule is selected, a corresponding equation is applied to obtain the WQI five days in advance.

**Table 5.** Prediction model for WQI with a time step of five days (units for the parameters used are reported in Table 3).

Rule No.	Rule	Equation
1.	Mn ≤ 0.178 UV254 > 0.05 O <sub>2</sub> p ≤ 91.263 WQI ≤ 67.5 Temp ≤ 14.647	$\text{WQI}_{5\_day\_pred} = 0.0492 \times \text{Temp} - 0.1442 \times \text{O}_2 + 0.0116 \times \text{O}_2\text{p} + 18.4678 \times \text{pH} + 0.1147 \times \text{Tur} + 0.1954 \times \text{TOC} - 0.0833 \times \text{KMnO}_4 + 1.0771 \times \text{UV254} + 0.359 \times \text{NH}_4 + 3.4449 \times \text{Mn} + 1.4357 \times \text{Al} + 194.1906 \times \text{Fe} + 0.0154 \times \text{WQI} - 117.4061$
2.	Mn ≤ 0.178 UV254 > 0.05 NH <sub>4</sub> > 0.116 NH <sub>4</sub> ≤ 0.217	$\text{WQI}_{5\_day\_pred} = -0.0314 \times \text{Temp} - 4.0616 \times \text{O}_2 + 0.4277 \times \text{O}_2\text{p} - 30.9666 \times \text{pH} + 0.0146 \times \text{Tur} + 0.409 \times \text{TOC} - 0.1268 \times \text{KMnO}_4 + 1.9415 \times \text{UV254} + 0.4537 \times \text{NH}_4 + 5.6159 \times \text{Mn} + 8.273 \times \text{Al} + 218.0231 \times \text{Fe} + 0.2294 \times \text{WQI} + 271.4172$
3.	Mn ≤ 0.238 UV254 > 0.05	$\text{WQI}_{5\_day\_pred} = -1.3998 \times \text{Temp} - 3.7955 \times \text{O}_2 + 0.2478 \times \text{O}_2\text{p} - 73.5681 \times \text{pH} + 0.0134 \times \text{Tur} - 0.0776 \times \text{TOC} - 13.3644 \times \text{KMnO}_4 - 234.667 \times \text{UV254} + 0.4444 \times \text{NH}_4 + 1.6204 \times \text{Mn} + 1.9645 \times \text{Al} + 132.2723 \times \text{Fe} - 0.2347 \times \text{WQI} + 717.3281$
4.	Mn > 0.154 Mn ≤ 0.556 pH > 7.514 Mn > 0.273 NH <sub>4</sub> ≤ 0.444 NH <sub>4</sub> > 0.168	$\text{WQI}_{5\_day\_pred} = 2.3592 \times \text{Temp} - 0.344 \times \text{O}_2 - 0.0068 \times \text{O}_2\text{p} - 1.3764 \times \text{pH} + 0.2721 \times \text{Tur} - 1.6152 \times \text{TOC} - 19.6394 \times \text{KMnO}_4 - 13.095 \times \text{UV254} + 6.8627 \times \text{NH}_4 + 4.9796 \times \text{Mn} + 45.1687 \times \text{Al} + 69.3252 \times \text{Fe} + 0.0377 \times \text{WQI} + 137.004$

Table 5. Cont.

Rule No.	Rule	Equation
5.	Mn > 0.154 Mn ≤ 0.625 WQI > 145.5 O <sub>2</sub> > 4.29 TOC > 2.144	$\text{WQI}_{5\_day\_pred} = -2.2601 \times \text{Temp} + 1.0762 \times \text{O}_2 - 0.5878 \times \text{O}_2\text{p} - 51.0854 \times \text{pH} + 4.2073 \times \text{Tur} + 0.4781 \times \text{TOC} - 442.8518 \times \text{UV254} + 53.2808 \times \text{NH}_4 + 83.5576 \times \text{Mn} - 462.669 \times \text{Al} - 14.6656 \times \text{Fe} + 0.0126 \times \text{WQI} + 598.894$
6.	Mn > 0.154 WQI ≤ 145.5	$\text{WQI}_{5\_day\_pred} = -0.1431 \times \text{Temp} - 15.6738 \times \text{O}_2 + 0.9714 \times \text{O}_2\text{p} - 2.2167 \times \text{pH} + 1.5839 \times \text{Tur} + 29.7531 \times \text{TOC} + 3.9299 \times \text{NH}_4 + 119.515 \times \text{Mn} + 41.3706 \times \text{Al} + 146.8431 \times \text{Fe} + 0.4928 \times \text{WQI} - 18.3667$
7.	Mn > 0.196 NH <sub>4</sub> ≤ 0.656 Temp > 14.521	$\text{WQI}_{5\_day\_pred} = -0.2457 \times \text{Temp} + 16.8424 \times \text{O}_2 - 2.17 \times \text{O}_2\text{p} - 1.9383 \times \text{pH} + 4.5045 \times \text{Tur} + 1.5087 \times \text{TOC} + 36.4246 \times \text{UV254} + 4.2032 \times \text{NH}_4 + 466.3063 \times \text{Mn} + 68.4722 \times \text{Al} - 11.5351 \times \text{Fe} - 1.1276 \times \text{WQI} + 250.7915$
8.	UV254 ≤ 0.051 Tur > 1.77 O <sub>2</sub> > 6.671	$\text{WQI}_{5\_day\_pred} = -0.1311 \times \text{Temp} - 0.8676 \times \text{O}_2 + 0.0209 \times \text{O}_2\text{p} - 4.7746 \times \text{pH} + 0.2683 \times \text{Tur} - 11.8676 \times \text{TOC} + 12.0203 \times \text{KMnO}_4 - 1084.1325 \times \text{UV254} + 0.9434 \times \text{NH}_4 - 197.3272 \times \text{Mn} + 28.2832 \times \text{Al} - 11.6473 \times \text{Fe} + 0.4091 \times \text{WQI} + 149.8425$
9.	UV254 ≤ 0.051 O <sub>2</sub> > 6.685	$\text{WQI}_{5\_day\_pred} = -2.1766 \times \text{Temp} - 12.4999 \times \text{O}_2 + 1.0932 \times \text{O}_2\text{p} - 30.7421 \times \text{pH} + 6.042 \times \text{Tur} - 9.7481 \times \text{TOC} + 0.5713 \times \text{KMnO}_4 + 60.3924 \times \text{UV254} - 44.7147 \times \text{NH}_4 + 7.7437 \times \text{Mn} + 55.508 \times \text{Al} + 69.3385 \times \text{Fe} + 345.6699$
10.	UV254 > 0.051 O <sub>2</sub> ≤ 1.646	$\text{WQI}_{5\_day\_pred} = -5.4435 \times \text{Temp} + 1.8677 \times \text{O}_2 - 0.6007 \times \text{O}_2\text{p} - 5.9718 \times \text{pH} - 0.8719 \times \text{Tur} + 37.473 \times \text{TOC} + 12.2012 \times \text{KMnO}_4 - 2209.8031 \times \text{UV254} - 45.2861 \times \text{NH}_4 + 171.9741 \times \text{Mn} - 12.49 \times \text{Fe} + 416.9164$
11.	UV254 > 0.051 Temp ≤ 14.146 NH <sub>4</sub> > 0.243	$\text{WQI}_{5\_day\_pred} = 7.2088 \times \text{O}_2 - 0.3865 \times \text{O}_2\text{p} - 10.7903 \times \text{pH} + 6.4896 \times \text{Tur} + 4.3084 \times \text{TOC} + 7.6783 \times \text{KMnO}_4 + 68.6909 \times \text{UV254} - 22.0766 \times \text{NH}_4 - 78.7705 \times \text{Mn} - 16.5265 \times \text{Fe} + 173.4252$
12.	-	$\text{WQI}_{5\_day\_pred} = -12.6524 \times \text{Temp} - 51.1841 \times \text{O}_2 + 4.544 \times \text{O}_2\text{p} - 169.1456 \times \text{pH} + 43.3802 \times \text{KMnO}_4 + 723.3245 \times \text{UV254} + 36.264 \times \text{NH}_4 + 99.2001 \times \text{Mn} + 1428.6758 \times \text{Al} + 1506.6314$

In Table 6, the third prediction model with a time step of ten days is presented, which has fourteen rules/equations with a slightly lower correlation coefficient (R) than the second model of 0.79, but also very high when tested with 10-fold cross-validation. To predict WQI, all parameters presented in Table 3 are needed at present. The rule selection is the same as for the second model, which depends on the values of the variables in the rule, and when a rule is selected, a corresponding equation is applied to calculate the WQI ten days in advance.

**Table 6.** Prediction model for WQI with a time step of ten days (units for the parameters used are reported in Table 3).

Rule No.	Rule	Equation
1.	Mn ≤ 0.161 UV254 > 0.049 Fe > 0.08 pH > 7.903 UV254 ≤ 0.07	$\text{WQI}_{10\_day\_pred} = 1.5616 \times \text{Temp} - 0.1592 \times \text{O}_2 + 0.2154 \times \text{O}_2\text{p} + 31.8486 \times \text{pH} - 0.1474 \times \text{Tur} + 16.4278 \times \text{TOC} + 0.0481 \times \text{KMnO}_4 - 6.5069 \times \text{UV254} + 0.3496 \times \text{NH}_4 - 0.7361 \times \text{Mn} + 2.4 \times \text{Al} + 246.8828 \times \text{Fe} + 0.0231 \times \text{WQI} - 292.6473$
2.	Mn ≤ 0.228 UV254 > 0.049 O <sub>2</sub> > 6.037 TOC > 2.068	$\text{WQI}_{10\_day\_pred} = -1.3609 \times \text{Temp} - 10.3742 \times \text{O}_2 + 1.3011 \times \text{O}_2\text{p} - 33.9897 \times \text{pH} - 3.2088 \times \text{Tur} + 0.0438 \times \text{TOC} - 565.14 \times \text{UV254} + 0.3834 \times \text{NH}_4 + 65.7253 \times \text{Mn} - 221.5725 \times \text{Al} + 188.8055 \times \text{Fe} + 0.1485 \times \text{WQI} + 360.0167$

Table 6. Cont.

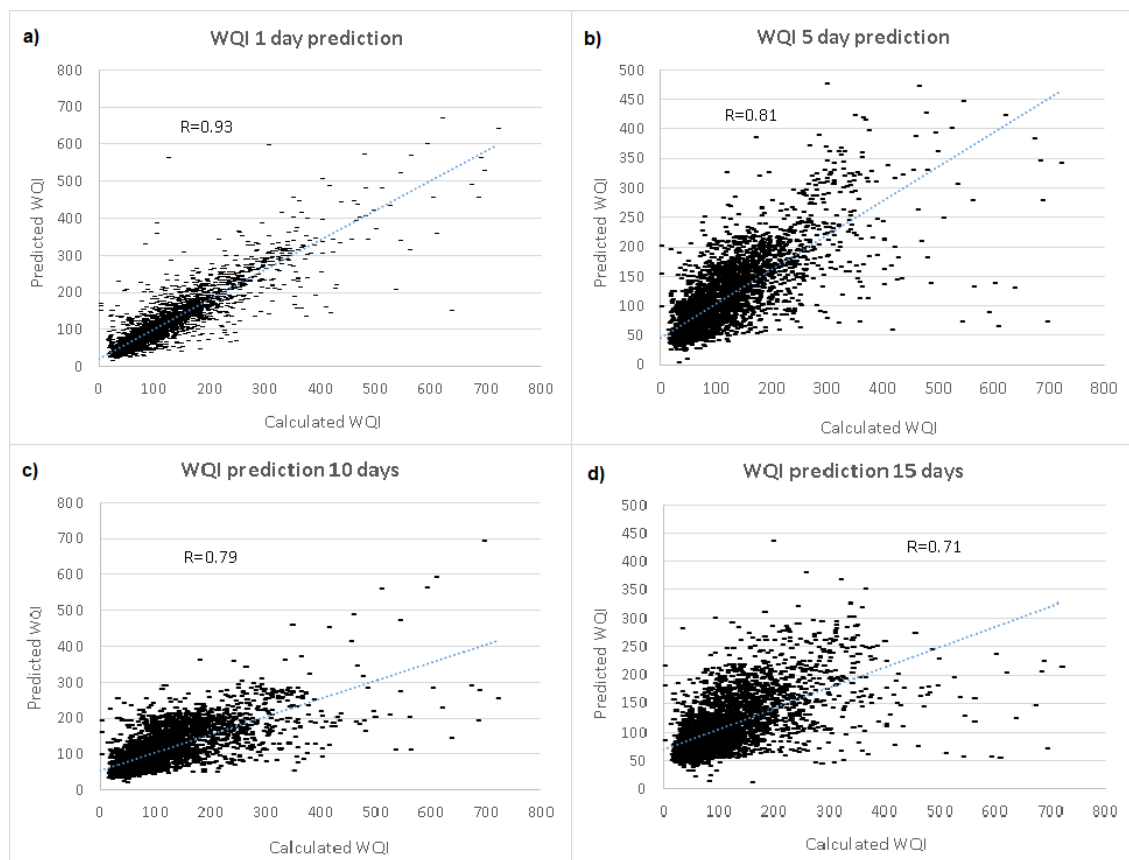
Rule No.	Rule	Equation
3.	Mn > 0.131 pH > 7.63 Mn ≤ 0.42	$\text{WQI}_{10\_day\_pred} = 1.5174 \times \text{Temp} - 0.4902 \times \text{O}_2 + 0.0296 \times \text{O}_2\text{p} - 1.0916 \times \text{pH} - 2.1281 \times \text{Tur} + 0.0978 \times \text{TOC} - 362.8943 \times \text{UV254} + 58.198 \times \text{NH}_4 + 11.7774 \times \text{Mn} + 436.819 \times \text{Al} + 110.5269 \times \text{Fe} + 0.1953 \times \text{WQI} + 70.6604$
4.	Mn > 0.147	$\text{WQI}_{10\_day\_pred} = -4.7909 \times \text{Temp} - 29.3983 \times \text{O}_2 + 1.9398 \times \text{O}_2\text{p} + 27.2021 \times \text{pH} - 3.6554 \times \text{Tur} + 0.4596 \times \text{TOC} + 36.9032 \times \text{NH}_4 + 176.6099 \times \text{Mn} + 556.3115 \times \text{Al} + 56.7678 \times \text{Fe} - 0.3055 \times \text{WQI} + 79.0885$
5.	UV254 ≤ 0.05 Tur ≤ 1.349 Temp > 13.085	$\text{WQI}_{10\_day\_pred} = -1.1396 \times \text{Temp} - 6.1336 \times \text{O}_2 + 0.4967 \times \text{O}_2\text{p} - 31.8387 \times \text{pH} + 0.4681 \times \text{Tur} - 13.8161 \times \text{TOC} + 5.2705 \times \text{KMnO}_4 + 90.5952 \times \text{UV254} - 4.0772 \times \text{NH}_4 - 234.7263 \times \text{Mn} - 22.4896 \times \text{Al} + 8.6443 \times \text{Fe} + 0.041 \times \text{WQI} + 349.673$
6.	UV254 > 0.05 Al ≤ 0.002 Temp ≤ 19.35	$\text{WQI}_{10\_day\_pred} = 0.0935 \times \text{Temp} + 12.3922 \times \text{O}_2 - 1.4555 \times \text{O}_2\text{p} - 21.8854 \times \text{pH} + 4.3067 \times \text{Tur} + 31.5541 \times \text{TOC} - 28.5229 \times \text{KMnO}_4 - 107.7914 \times \text{UV254} + 15.9927 \times \text{NH}_4 + 21.5347 \times \text{Al} + 131.9561 \times \text{Fe} - 0.7713 \times \text{WQI} + 263.7697$
7.	UV254 > 0.05 UV254 > 0.052	$\text{WQI}_{10\_day\_pred} = 9.6063 \times \text{Temp} + 36.7844 \times \text{O}_2 - 3.7629 \times \text{O}_2\text{p} - 29.6123 \times \text{pH} + 14.0689 \times \text{Tur} + 7.3178 \times \text{TOC} - 66.8263 \times \text{UV254} + 439.762 \times \text{Mn} - 692.8787 \times \text{Al} - 313.2304 \times \text{Fe} + 0.0563 \times \text{WQI} + 146.4862$
8.	UV254 ≤ 0.048 O <sub>2</sub> ≤ 7.389 pH > 7.755	$\text{WQI}_{10\_day\_pred} = -0.3282 \times \text{Temp} - 8.0303 \times \text{O}_2 + 0.5504 \times \text{O}_2\text{p} + 37.8754 \times \text{pH} + 18.7798 \times \text{Tur} + 3.8309 \times \text{TOC} + 339.0088 \times \text{UV254} - 62.2836 \times \text{NH}_4 - 173.1733 \times \text{Mn} - 83.2919 \times \text{Al} + 21.4339 \times \text{Fe} + 0.0471 \times \text{WQI} - 263.5586$
9.	UV254 ≤ 0.049 pH > 7.785 Tur ≤ 1.468	$\text{WQI}_{10\_day\_pred} = -0.2551 \times \text{Temp} - 0.1407 \times \text{O}_2 - 11.9135 \times \text{pH} + 9.0883 \times \text{Tur} - 5.6761 \times \text{TOC} + 7.0392 \times \text{KMnO}_4 + 406.0863 \times \text{UV254} - 29.4935 \times \text{Mn} - 60.929 \times \text{Al} + 69.954 \times \text{Fe} + 109.2525$
10.	Tur > 1.504 pH > 7.762 pH ≤ 8.157	$\text{WQI}_{10\_day\_pred} = 0.6663 \times \text{Temp} + 0.2275 \times \text{O}_2 - 11.1661 \times \text{pH} + 4.3653 \times \text{Tur} - 8.2495 \times \text{TOC} + 5.6081 \times \text{KMnO}_4 + 564.6441 \times \text{UV254} - 10.1607 \times \text{NH}_4 - 92.4626 \times \text{Mn} + 216.0347 \times \text{Fe} - 0.4035 \times \text{WQI} + 109.4049$
11.	WQI > 38.5 O <sub>2</sub> > 6.874 Tur ≤ 2.111 O <sub>2</sub> > 8.921	$\text{WQI}_{10\_day\_pred} = -3.2172 \times \text{Temp} - 0.7486 \times \text{O}_2 - 59.4558 \times \text{pH} - 0.6238 \times \text{Tur} + 6.9855 \times \text{TOC} + 8.644 \times \text{KMnO}_4 + 1134.6654 \times \text{UV254} - 174.1622 \times \text{Al} - 55.1062 \times \text{Fe} + 0.2754 \times \text{WQI} + 496.4482$
12.	UV254 ≤ 0.049 pH > 7.673	$\text{WQI}_{10\_day\_pred} = -1.5223 \times \text{Temp} + 9.0823 \times \text{O}_2 - 0.84 \times \text{O}_2\text{p} - 43.5423 \times \text{pH} + 8.2979 \times \text{Tur} + 13.9948 \times \text{KMnO}_4 + 2670.0583 \times \text{UV254} + 272.8255$
13.	Temp > 9.654	$\text{WQI}_{10\_day\_pred} = -7.5086 \times \text{Temp} - 112.8856 \times \text{pH} + 69.6063 \times \text{TOC} + 967.9153$
14.	-	$\text{WQI}_{10\_day\_pred} = -61.5182 \times \text{O}_2 - 751.5462 \times \text{Tur} + 1847.615$

The fourth and final model with a prediction of fifteen days' time step is presented in Table 7. The prediction model has only five rules/equations, but with a slightly low correlation coefficient (R) of 0.71 when tested with 10-fold cross-validation. To predict WQI, all parameters presented in Table 3 are needed at present. The rule selection is the same as for other models, from which the WQI can be calculated fifteen days in advance. The number of rules was expected to be higher as the time step increases and with given default values of the parameters in WEKA to obtain a higher correlation coefficient and the optimal number of rules, as explained in Section 3.3; however, surprisingly, the number of rules is less than for the second and third model.

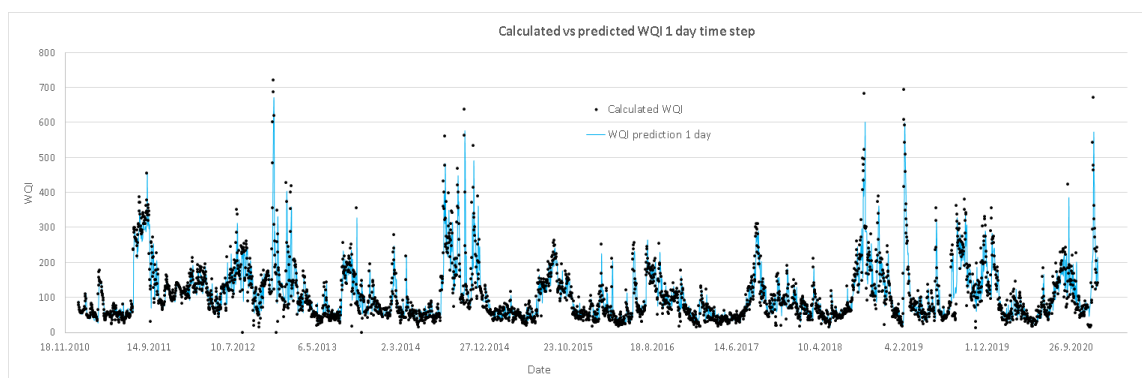
The correlation between the calculated and the modelled (predicted) values of WQI is presented on Figure 3, as the correlation coefficient (R). From Figure 3, it can be seen that as the time step increases, the accuracy of the model slowly decreases as expected: for the one day prediction model, the correlation coefficient is 0.93; for the five days prediction model, 0.81; for the ten days prediction model, 0.79; and finally, for the fifteen days prediction model, 0.71.

**Table 7.** Prediction model for WQI with a time step of fifteen days (units for the parameters used are reported in Table 3).

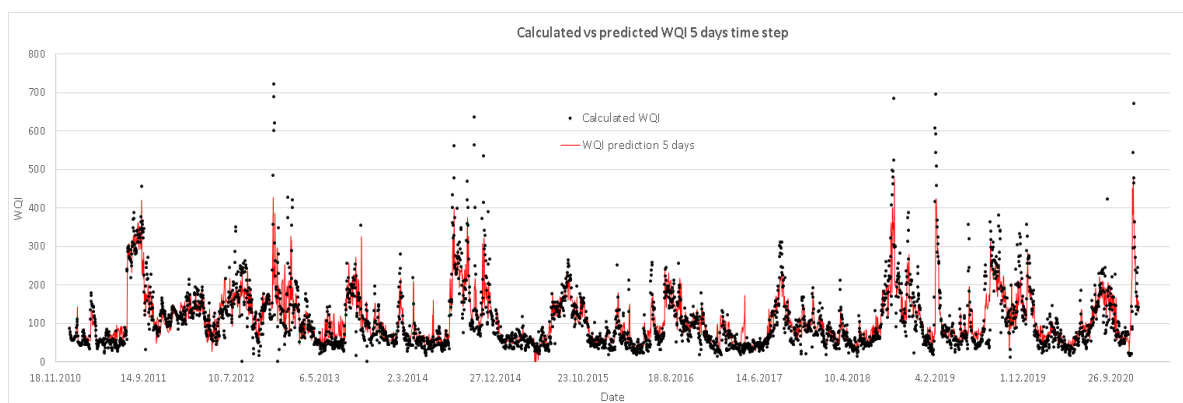
Rule No.	Rule	Equation
1.	pH > 7.722 Mn ≤ 0.182 Tur > 1.675 Temp ≤ 15.32	$\text{WQI}_{15\_day\_pred} = -2.308 \times \text{Temp} - 7.9764 \times \text{O}_2 + 1.0935 \times \text{O}_2\text{p} + 28.5839 \times \text{pH} + 3.8658 \times \text{Tur} + 0.2701 \times \text{TOC} + 0.183 \times \text{KMnO}_4 - 12.663 \times \text{UV254} + 1.2919 \times \text{NH}_4 + 2.8251 \times \text{Mn} - 224.3635 \times \text{Al} + 119.207 \times \text{Fe} + 0.0008 \times \text{WQI} - 184.4439$
2.	pH > 7.703 Mn > 0.134 pH > 7.948	$\text{WQI}_{15\_day\_pred} = -1.1171 \times \text{Temp} - 9.6787 \times \text{O}_2 + 0.7265 \times \text{O}_2\text{p} - 47.8948 \times \text{pH} - 1.7395 \times \text{Tur} + 10.9147 \times \text{TOC} - 16.9889 \times \text{KMnO}_4 - 571.2309 \times \text{UV254} + 2.2305 \times \text{NH}_4 + 6.0254 \times \text{Mn} + 431.2766 \times \text{Al} + 209.4744 \times \text{Fe} + 0.0623 \times \text{WQI} + 518.4762$
3.	Mn > 0.134 Mn ≤ 0.52 Temp ≤ 21.01 pH > 7.63 WQI > 114.5	$\text{WQI}_{15\_day\_pred} = -2.3239 \times \text{Temp} - 4.9107 \times \text{O}_2 + 0.1752 \times \text{O}_2\text{p} - 5.5331 \times \text{pH} - 5.4288 \times \text{Tur} - 1.1149 \times \text{TOC} + 7.5097 \times \text{NH}_4 + 145.6845 \times \text{Mn} + 676.501 \times \text{Al} + 5.0832 \times \text{Fe} - 0.0928 \times \text{WQI} + 201.3688$
4.	Mn > 0.134 Temp ≤ 21.065 Tur > 3.084 UV254 ≤ 0.071	$\text{WQI}_{15\_day\_pred} = -0.2501 \times \text{Temp} + 5.9114 \times \text{O}_2 + 0.2627 \times \text{O}_2\text{p} - 73.2176 \times \text{pH} - 6.2534 \times \text{Tur} + 1.7585 \times \text{TOC} - 1.2392 \times \text{KMnO}_4 + 1308.1435 \times \text{UV254} + 117.9952 \times \text{NH}_4 + 11.2328 \times \text{Mn} + 1364.4201 \times \text{Al} + 1.451 \times \text{Fe} - 0.0057 \times \text{WQI} + 584.6712$
5.	-	$\text{WQI}_{15\_day\_pred} = -3.1871 \times \text{Temp} - 11.7564 \times \text{O}_2 + 0.7338 \times \text{O}_2\text{p} - 101.7279 \times \text{pH} - 1628 \times \text{Tur} + 24.9861 \times \text{TOC} - 245.935 \times \text{UV254} + 48.1042 \times \text{NH}_4 + 151.9509 \times \text{Mn} + 417.6806 \times \text{Al} + 108.3222 \times \text{Fe} - 0.1082 \times \text{WQI} + 918.6201$

**Figure 3.** Correlation coefficient between calculated and predicted values for WQI: (a) one day time step, (b) five days' time step, (c) ten days' time step, and (d) fifteen days' time step.

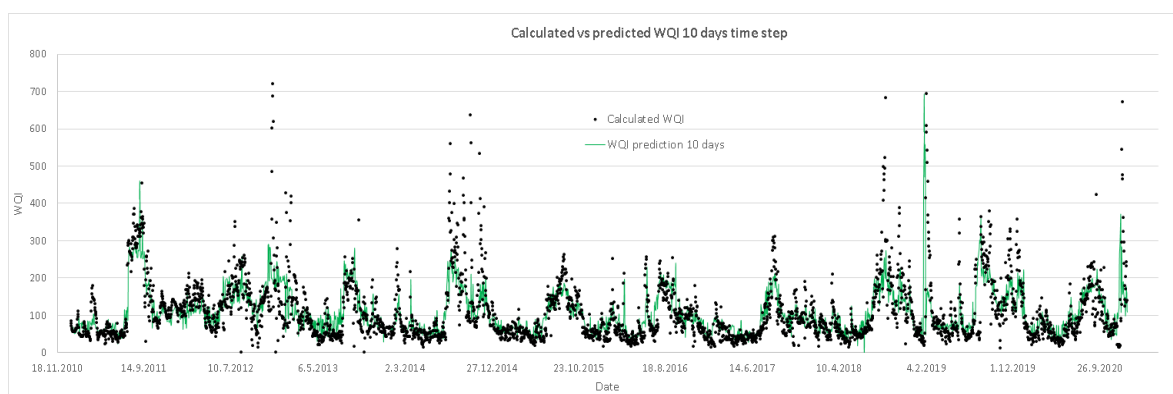
The performance of the prediction models is presented in Figures 4–7. Each figure represents the modelled (predicted) vs. calculated values of the WQI for the given time period of one, five, ten and fifteen days in advance. Figure 4 indicates very high accuracy of the first prediction model with a one day time step compared to the calculated data of the WQI, with a good prediction of the peak values included. Increasing the time step to five (Figure 5), ten (Figure 6) and fifteen days (Figure 7), the accuracy of the models slowly decreases, as can also be seen from Figure 3; however, regardless of that, the good prediction of the peak values remains, and as such, models can be used in prediction purposes.



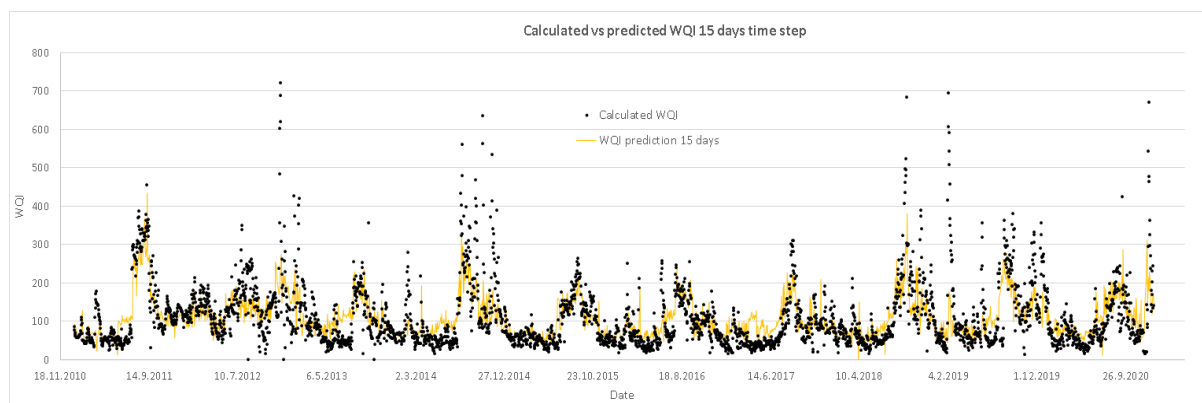
**Figure 4.** Performance of the prediction models; calculated vs. predicted values of the WQI for the one day time step.



**Figure 5.** Performance of the prediction models; calculated vs. predicted values of the WQI for the five days' time step.



**Figure 6.** Performance of the prediction models; calculated vs. predicted values of the WQI for the ten days' time step.



**Figure 7.** Performance of the prediction models; calculated vs. predicted values of the WQI for the fifteen days' time step.

Overall, obtained results, i.e., prediction models, are at an acceptable level when considering the correlation coefficients and the prediction of the peak values. As such, they can help in optimization of the treatment processes which are dependent on the quality of raw water in the Butoniga reservoir, and overall, in the sustainability of the DWTP.

## 5. Discussion

Within this research, simple and yet efficient WQI prediction models were developed for the Butoniga DWTP using ML tools. The models predict the values of WQI one, five, ten and fifteen days in advance, given the present observed values of thirteen variables, and can thus be used as a warning tool and to manage treatment processes of the DWTP which are dependent on the quality of raw water in the Butoniga reservoir. The WQI method was used to determine the water quality of influent at the DWTP.

Simple statistical analysis of WQI showed that only 23% of WQI falls under the “Excellent” water rank according to Table 2; 36% of the WQI falls under the “Good” water rank, 29% falls under the “Bad” water rank, only 8% under the “Very bad” and 4% under the “Unfit for drinking” water rank. Generally, it can be said that most of the water (52%; “Excellent” plus “Good” water rank) is suitable for drinking according to the WQI. As such, much more attention should be addressed to these samples with “Excellent” and “Good” water quality because there is no need to worry about the quality of water and these waters are marginally suitable for drinking and do not need to go through all treatment processes, while the rest of the samples must go through all treatment processes on the Butoniga DWTP. Regardless of that, the DWTP provides water that is safe for drinking according to all examined parameters prescribed by the Croatian regulations [19].

Prediction models for the WQI with a time step of one, five, ten and fifteen days in advance were built with the use of rule-based models integrated in WEKA modelling software [23]. Prediction models are interpreted as a set of IF THEN rules, where each rule is associated with a multivariate linear model. A rule indicates that, whenever a case satisfies all the conditions, the linear model (equation) is appropriate for predicting the value of the target attribute, in this case, WQI. In its structure (rules and equations), the models contain variables (parameters) from Table 3 measured at present to predict WQI one, five, ten and fifteen days in advance in order to improve the treatment processes of the DWTP regarding changes in raw water quality in the Butoniga reservoir.

Previous modelling attempts regarding a similar problem was done in [7], where the effluent water quality of a treatment plant in Turkey, according to WQI and HRA, was investigated using PCA. Generally, the main conclusion of this study is that the water was suitable for drinking purposes, and the treatment plant provides excellent stable water quality; as a result of HRA, parallel with WQI results, it has been found that the risk for people is “negligible”.

In other research, such as [7,9–16], various ML algorithms and methods were used, such as PCA, ANN, NARNET, LSTM, MLP, CNN, RF, NN, MLR, SVM, BTM, RSS, AR, SVR, LWLR, GNB, SGD, KNN, DT, GBC, BC, SLR and RLR in predicting WQI for various purposes, i.e., efficient and better prediction of WQI [9,11,13], water quality management [10], health risk assessment and water supply quality for drinking purposes [7], water quality classification of rivers, lakes, reservoirs [14–16], etc.

As mentioned in the introduction, the main problem with the Butoniga reservoir, and thus affecting treatment processes on DWTP, appears in the summer months when the water temperature is the most critical parameter, and water for the treatment process must be captured from the lowest layer of the reservoir which has increased concentrations of Mn, Fe, NH<sub>4</sub> and lower pH values, and thus has influence on the treatment processes of the DWTP [18]. To help to manage treatment processes in mentioned situations during the summer months, the obtained prediction models can be used in the prediction of the WQI, i.e., the quality of raw water in the Butoniga reservoir, and thus prepare and define water treatment steps and methods to be used.

Finally, the aim of these models was the prediction of the WQI, and given the accuracy of the models' performance on unseen data, they can help in the optimization and management of treatment processes on the DWTP which are dependent on the quality of raw water in the Butoniga reservoir, and overall, in the sustainability of the DWTP.

## 6. Conclusions

In times of increasing demand for potable water, proper optimization and management of DWTPs is needed, both in terms of obtaining the required water quality, and also in terms of the adequate treatment processes used and the consumption of the electric energy. In light of this, in this research, prediction of the WQI was done for the improvement of specific treatment processes of the Butoniga DWTP.

ML methods in the form of rule-based models were applied on the data measured at the intake of raw water from the Butoniga reservoir to predict the WQI one, five, ten and fifteen days in advance. Predictions of the WQI are done according to current values of measured parameters at the intake of raw water. Obtained models have high correlation coefficients and provide accurate predictions of the WQI, correctly predicting the peak values when compared to calculated data of the WQI. As expected, the highest correlation coefficient was the one day prediction, followed by five and ten days; the lowest correlation coefficient was with the fifteen-day prediction.

As such, obtained prediction models can help with the optimization and management of treatment processes of the DWTP, especially during the summer months (tourist season) when the quality of raw water in the Butoniga reservoir is the worst, and where changes in the raw water quality can result in direct action and optimization of the operation of the DWTP. Overall, in this way, obtained models will also help in the sustainability of the DWTP and will improve the operation of the DWTP, which will result in a better water quality, thus benefiting overall public safety.

Future work of this research is focused on defining boundaries, i.e., limit values of the WQI obtained through the prediction models to manage and control specific treatment processes of the Butoniga DWTP. For this, an analysis of the output water quality, the treatment processes included, and the consumed electric energy on the DWTP will be conducted.

**Author Contributions:** Conceptualization, G.V. and I.S.Č.; methodology, G.V.; software, G.V.; validation, G.V., I.S.Č. and E.Ž.; formal analysis, G.V. and S.Z.; investigation, G.V.; resources, S.Z.; data curation, S.Z.; writing—original draft preparation, G.V.; writing—review and editing, G.V., I.S.Č., E.Ž. and S.Z.; visualization, G.V.; supervision, G.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from Istarski vodovod d.o.o. Buzet and are available [[www.ivb.hr](http://www.ivb.hr)] with the permission of Istarski vodovod d.o.o. Buzet.

**Acknowledgments:** This research has been fully supported by the University of Rijeka under projects: “Sustainable River Basin Management by Implementation of Innovative Methodologies, Approaches and Tools” (uniri-tehnic-18-129) and “Hydrology of Water Resources and Identification of Flood and Mudflow Risk in Karst” (uniri-tehnic-18-54). Also this research has been supported by project “Influence of Summer Fire on Soil and Water Quality”, founded by the Croatian Science Foundation (IP-2018-01-1645). Authors greatly appreciate the comments of the three anonymous reviewers who helped substantially in improving the paper.

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

### Abbreviations

AI	Artificial Intelligence
ANN	Artificial Neural Network
AR	Additive Regression
BC	Bagging Classifier
BTM	Bagged Tree Models
CNN	Convolutional Neural Network
DT	Decision Tree
DWTP	Drinking Water Treatment Plant
GBC	Gradient Boosting Classifier
GNB	Gaussian Naïve Bayes
HRA	Health Risk Assessment
KNN	k-Nearest Neighbors
LSTM	Long Short-Term Memory
LWLR	Locally Weighted Linear Regression
MAC	Maximum Allowable Concentration
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multinomial Logistic Regression
NARNET	Nonlinear Autoregressive Neural Network
NN	Neural Network
PCA	Principal Component Analysis
R	Correlation Coefficient
RF	Random Forest
RLLR	Ridge and Lasso Regression
RSS	Random Subspace
SGD	Stochastic Gradient Descent
SLR	Simple Linear Regression
SVM	Support Vector Machine
SVR	Support Vector Regression
WQI	Water Quality Index

### References

1. Aalipour Erdi, M.; Gasempour Niari, H.; Meshkini, S.R.M.; Foroug, S. Surveying drinking water quality (Balikhlou River, Ardabil Province, Iran). *Appl. Water Sci.* **2018**, *8*, 34. [[CrossRef](#)]
2. Kumar, A.; Sharma, M.P.; Rai, S.P. A novel approach for river health assessment of Chambal using fuzzy modeling, India. *Desalination Water Treat.* **2017**, *58*, 72–79. [[CrossRef](#)]
3. Kumar, A.; Taxak, A.K.; Mishra, S.; Pandey, R. Long term trend analysis and suitability of water quality of River Ganga at Himalayan hills of Uttarakhand, India. *Environ. Technol. Inno.* **2021**, *22*, 101405. [[CrossRef](#)]
4. Tiri, A.; Belkhiri, L.; Mouni, L. Evaluation of surface water quality for drinking purposes using fuzzy inference system. *Groundw. Sustain. Dev.* **2018**, *6*, 235–244. [[CrossRef](#)]



5. Braga, F.H.R.; Dutra, M.L.S.; Lima, N.S.; Silva, G.M.; Miranda, R.C.M.; Firmo, W.C.A.; Moura, A.R.L.; Monteiro, A.S.; Silva, L.C.N.; Silva, D.F.; et al. Study of the Influence of Physicochemical Parameters on the Water Quality Index (WQI) in the Maranhão Amazon, Brazil. *Water* **2022**, *14*, 1546. [CrossRef]
6. Xiao, J.; Wang, L.; Deng, L.; Jin, Z. Characteristics, sources, water quality and health risk assessment of trace elements in river water and well water in the Chinese Loess Plateau. *Sci. Total Environ.* **2019**, *650*, 2004–2012. [CrossRef] [PubMed]
7. Alver, A. Evaluation of conventional drinking water treatment plant efficiency according to water quality index and health risk assessment. *Environ. Sci. Pollut. Res.* **2019**, *26*, 27225–27238. [CrossRef] [PubMed]
8. Mbavarira, T.M.; Grimm, C. A Systemic View on Circular Economy in the Water Industry: Learnings from a Belgian and Dutch Case. *Sustainability* **2021**, *13*, 3313. [CrossRef]
9. Aldhyani, T.H.H.; Al-Yaari, M.; Alkahtani, H.; Maashi, M. Water Quality Prediction Using Artificial Intelligence Algorithms. *Appl. Bionics Biomech.* **2020**, *2020*, 6659314. [CrossRef] [PubMed]
10. Sheikh Khozani, Z.; Iranmehr, M.; Wan Mohtar, W.H.M. Improving Water Quality Index prediction for water resources management plans in Malaysia: Application of machine learning techniques. *Geocarto Int.* **2022**, 1–18. [CrossRef]
11. Hassan, M.; Hassan, M.; Akter, L.; Rahman, M.; Zaman, S.; Hasib, K.; Jahan, N.; Smrity, R.N.; Farhana, J.; Raihan, M.; et al. Efficient Prediction of Water Quality Index (WQI) Using Machine Learning Algorithms. *Hum. Cent. Intell. Syst.* **2021**, *1*, 86–97. [CrossRef]
12. Kouadri, S.; Elbeltagi, A.; Islam, A.R.M.T.; Kateb, S. Performance of machine learning methods in predicting water quality index based on irregular data set: Application on Illizi region (Algerian southeast). *Appl. Water. Sci.* **2021**, *11*, 190. [CrossRef]
13. Ahmed, U.; Mumatz, R.; Anwar, H.; Shah, A.A.; Irfan, R.; Garcia-Nieto, J. Efficient Water Quality Prediction Using Supervised Machine Learning. *Water* **2019**, *11*, 2210. [CrossRef]
14. Khoi, D.N.; Quan, N.T.; Linh, D.Q.; Nhi, P.T.T.; Thuy, N.T.D. Using Machine Learning Models for Predicting the Water Quality Index in the La Buong River, Vietnam. *Water* **2022**, *14*, 1552. [CrossRef]
15. Aalipour, M.; Št'astný, B.; Horký, F.; Jabbarian Amiri, B. Scaling an Artificial Neural Network-Based Water Quality Index Model from Small to Large Catchments. *Water* **2022**, *14*, 920. [CrossRef]
16. Fernández del Castillo, A.; Yebra-Montes, C.; Verduzco Garibay, M.; de Anda, J.; Garcia-Gonzalez, A.; Gradilla-Hernández, M.S. Simple Prediction of an Ecosystem-Specific Water Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning. *Water* **2022**, *14*, 1235. [CrossRef]
17. Hajduk Černeha, B. Akumulacija Butoniga u Istri—Prva iskustva u korištenju za vodoopskrbu. In Proceedings of the Vodni dnevi 2021, Rimske Toplice, Slovenia, 7–8 October 2021.
18. Zorko, S. Akumulacija Butoniga—pritisci u slijevu i zaštita voda. In Proceedings of the Zbornik radova—Upravljanje jezerima i akumulacijama u Hrvatskoj i Okrugli stol o aktualnoj problematici Vranskog jezera kod Biograda na Moru, Biograd na Moru, Croatia, 4–6 May 2017.
19. Republic of Croatia regulations, Act on water intended for human consumption. 2020. In *Cro., Zakon o Vodi za Ljudsku Potrošnju*; NN 56/13, 64/15, 104/17, 115/18, 16/20; Effective from 20 February 2020. Available online: [https://narodne-novine.nn.hr/clanci/sluzbeni/2013\\_05\\_56\\_1138.html](https://narodne-novine.nn.hr/clanci/sluzbeni/2013_05_56_1138.html) (accessed on 9 March 2022).
20. Volf, G.; Atanasova, N.; Kompare, B.; Precali, R.; Ožanić, N. Descriptive and prediction models of phytoplankton in the northern Adriatic. *Ecol. Model.* **2011**, *222*, 2502–2511. [CrossRef]
21. Technical Committees, ISO/TC 147, Water Quality. Available online: <https://www.iso.org/committee/52834/x/catalogue> (accessed on 4 February 2022).
22. Baird, R.B.; Eaton, A.D.; Rice, E.W. *Standard Methods for the Examination of Water and Wastewater*, 23rd ed.; AWWA: Washington, DC, USA, 2017.
23. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Elsevier: Amsterdam, The Netherlands, 2005.
24. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Francisco, CA, USA, 1993.
25. Yisa, J.; Jimoh, T. Analytical Studies on Water Quality Index of River Landzu. *Am. J. Appl. Sci.* **2010**, *7*, 453–458. [CrossRef]