

Article

Machine Learning Approach for Rapid Estimation of Five-Day Biochemical Oxygen Demand in Wastewater

Panagiotis G. Asteris¹, Dimitrios E. Alexakis^{2,*}, Markos Z. Tsoukalas¹, Dimitra E. Gamvroula¹
and Deniz Guney³

¹ Computational Mechanics Laboratory, School of Pedagogical and Technological Education, GR 15122 Athens, Greece

² Laboratory of Geoenvironmental Science and Environmental Quality Assurance, Department of Civil Engineering, School of Engineering, University of West Attica, GR 12241 Athens, Greece

³ Engineering Faculty, San Diego State University, San Diego, CA 92182, USA

* Correspondence: d.alexakis@uniwa.gr; Tel.: +30-210-538-1256

Abstract: Improperly managed wastewater effluent poses environmental and public health risks. BOD evaluation is complicated by wastewater treatment. Using key parameters to estimate BOD in wastewater can improve wastewater management and environmental monitoring. This study proposes a BOD determination method based on the Artificial Neural Networks (ANN) model to combine Chemical Oxygen Demand (COD), Suspended Solids (SS), Total Nitrogen (T-N), Ammonia Nitrogen (NH₄-N), and Total Phosphorous (T-P) concentrations in wastewater. Twelve different transfer functions are investigated, including the common Hyperbolic Tangent Sigmoid (HTS), Log-sigmoid (LS), and Linear (Li) functions. This research evaluated 576,000 ANN models while considering the variable random number generator due to the ten alternative ANN configuration parameters. This study proposes a new approach to assessing water resources and wastewater facility performance. It also demonstrates ANN's environmental and educational applications. Based on their RMSE index over the testing datasets and their configuration parameters, twenty ANN architectures are ranked. A BOD prediction equation written in Excel makes testing and applying in real-world applications easier. The developed and proposed ANN-LM 5-8-1 model depicting almost ideal performance metrics proved to be a reliable and helpful tool for scientists, researchers, engineers, and practitioners in water system monitoring and the design phase of wastewater treatment plants.

Keywords: artificial intelligence; artificial neural networks; biochemical oxygen demand (BOD); machine learning; wastewater



Citation: Asteris, P.G.; Alexakis, D.E.; Tsoukalas, M.Z.; Gamvroula, D.E.; Guney, D. Machine Learning Approach for Rapid Estimation of Five-Day Biochemical Oxygen Demand in Wastewater. *Water* **2023**, *15*, 103. <https://doi.org/10.3390/w15010103>

Academic Editor: Antonio Panico

Received: 2 December 2022

Revised: 21 December 2022

Accepted: 22 December 2022

Published: 28 December 2022



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

The U.K. Royal Commission on River Pollution suggested the biological measurement Biochemical Oxygen Demand (BOD) in 1908 to demonstrate the organic pollution of rivers [1]. BOD is defined as the quantity of oxygen taken up by the respiratory activity of microorganisms growing on organic substances present in the sample (e.g., sludge or water) while incubated at a specific temperature (typically 20 °C) for a fixed period (usually 5 days, BOD₅) [1] (Table 1). It is a measurement of the organic contaminants in water that can be broken down by biological processes [1]. The main downside of this measurement is the time (5 days) required to complete it [2].

Automated control solutions for wastewater treatment facilities and environmental monitoring applications need a reliable and accurate measurement of BOD in influent and effluent samples. Standard dilution is the classic technique for determining BOD [1]. This approach has been used to identify contaminants in most water bodies and assess BOD levels with reasonable precision. BOD monitoring necessitates the use of specialized equipment and procedures, which considerably increases the difficulty and expense of

the detection process. The difference in detection times for BOD (5 days) [3] and the other critical parameters COD (Chemical Oxygen Demand), SS (Suspended Solids), T-N (Total Nitrogen), NH₄-N (Ammonium Nitrogen) and T-P (Total Phosphorous) makes it challenging to match all the detection methods to simultaneously measure BOD, COD, SS, T-N, NH₄-N and T-P contents to evaluate the performance of wastewater treatment processes (Table 1). BOD, T-N and T-P, among others, are crucial parameters for estimating Water Quality Indices and assessing water bodies [4–6]. Moreover, the off-site detection and long detection times of the approaches mentioned above cannot meet the requirements of on-site, real-time monitoring of contaminants in automated water treatment operations [7]. According to Jouanneau et al. [1] the determination of BOD₅ is helpful in three ways: (a) it illustrates if the wastewater discharge and waste treatment technique comply with current objective values and legislation; (b) the COD/BOD₅ ratio demonstrates the biodegradable fraction of effluent; and (c) the ratio of BOD₅ to COD in wastewater treatment facilities represents the biodegradable portion of an effluent.

The high non-linear correlation between BOD and the five other parameters (COD, SS, T-N, NH₄-N and T-P) included in the analysis of this study reveals that it is hard to use classical computing techniques such as regression analysis to delineate related issues and extract significant results. Concerning water and wastewater treatment and assessment of water quality, many research studies and scientific publications have been conducted out applying Artificial Intelligence Techniques [8–10]. In the last two decades, soft computing techniques such as Machine Learning (ML) proved reliable and robust methods to model such topics with a strongly non-linear nature [11–15].

Because of the widespread usage of BOD techniques, alternatives ranging from static bioassays to online biosensors have been developed [1] (Tables 1 and 2). ML is a system that employs a series of programmed algorithms to anticipate the future patterns of any raw data by automatically analyzing the data's discreet relationships [16]. To identify the rules underlying the known data as precisely as feasible, it is often necessary to produce the dataset after treating it appropriately [17]. Then, a suitable ML method is selected based on the input data's features and output data's requirements. The chosen algorithm will then be trained using the meticulously gathered data and assessed to change its hyperparameters, producing the desired model. The suggested machine learning model is then qualified to predict new data [17].

Table 1. Groups of main methods for determining BOD, required time and measurement range for each method.

Method	Required Time (Median)	Measurement Range (mg L ⁻¹)	References
Chemical or Electrochemical measurement			
Standard reference method	5 days	0–6	ISO 5815-1:2003 [18]; Jouanneau et al. [1]
Modified reference method	5 days	0–6	McDonagh et al. [19]; McEvoy et al. [20]; Xiong et al. [21]; Xu et al. [22]
Photometric method	5 days	0–6	Jouanneau et al. [1]
Manometric method	5 days	0–700	Jouanneau et al. [1]
BOD prediction			
Biosensor based on bioluminescent bacteria	72 min	0–200	Sakaguchi et al. [23,24]
Microbial fuel cells	315 min	0–200	Jouanneau et al. [1]; Kim et al. [25]
Biosensor with entrapped bacteria	10 min	0–500	Karube [26]; Liu et al. [27]

Table 2. Approaches for estimating BOD, input, output variables and correlation coefficient (R^2) for each approach.

Approach	Number of Input Variables	Input Variables	Output Variables	R^2	Type of Water	References
ANN	4	TSS, TS, pH, T	BOD, COD	0.63–0.81	wastewater	Zare Abyaneh [28]
ANN	11	pH, TS, T-Alk, T-Hard, Cl, PO_4^{3-} , K, Na, NH_4-N , NO_3-N , COD	DO, BOD	0.77–0.85	river water	Singh et al. [29]
ANFIS	9	pH, alkalinity, T-Hard, TS, TDS, K, PO_4^{3-} , NO_3^- , DO	BOD	0.69–0.85	river water	Ahmed and Ali Shah [30]
ANN	11	pH, T-Alk, T-Hard, TS, COD, NH_4-N , NO_3-N , Cl, PO_4^{3-} , K, Na	DO, BOD	0.74–0.90	river water	Basant et al. [31]
ANN	8	T, turbidity, pH, CND, TDS, TSS, DO, COD	BOD	0.69	wastewater	Asami et al. [32]

Notes: ANN: Artificial neural network, ANFIS: Adaptive Neuro-Fuzzy Inference System, TSS: total suspended solids, TS: total solids, T-Alk: total alkalinity, T: temperature, T-Hard: total hardness, Cl: chloride, PO_4^{3-} : phosphate, K: potassium, Na: sodium, NH_4-N : ammonia nitrogen, NO_3-N : nitrate nitrogen, NO_3^- : nitrate, DO: dissolved oxygen, COD: chemical oxygen demand, CND: electrical conductivity, TDS: total dissolved solids.

The manuscript is organized into several sections. Section 2 presents the significance of this research. Section 3 presents material and Methods used for the development of the mathematical forecasting model for the BOD₅. Section 4 provides the presented results on the development of a closed-form equation for the estimation of BOD₅ in wastewater and the mapping of BOD₅, revealing its strongly nonlinear nature. In Section 5 the limitations of the proposed model are presented, followed by concluding remarks in Section 6.

2. Research Significance

The efficient operation and management of wastewater treatment plants (WWTPs) are gaining more consideration as environmental concerns receive increasing attention. The discharge of a WWTP's effluent into a receiving water body may cause or spread a variety of human health problems if it is improperly managed, hence posing significant environmental and public health risks. Better management of a WWTP may be attained by developing a robust mathematical method for estimating the BOD content in wastewater on a dataset of a minimum number of key parameters. Nevertheless, evaluating BOD content in wastewater is challenging owing to the intricacy of the treatment processes. The complex biological, chemical, and physical systems involved in the wastewater treatment process display nonlinear tendencies that are challenging to explain using linear mathematical models.

We rely on the advantages of the ANN model to explore the performance of the ANN model in determining BOD values in wastewater. Consequently, this study is a novel attempt at proposing a BOD determination method based on the ANN model to combine the COD, SS, T-N, NH_4-N and T-P concentrations in wastewater. This study may provide a new idea for monitoring water resources and the performance of the wastewater treatment plant.

3. Materials and Methods

3.1. Artificial Neural Networks

Artificial neural networks owe their name to the biological neural networks that they mimic significantly in structure and the basic principles that govern them. They are

mathematical simulants that, after suitable training, use capable and reliable databases that contain the existing knowledge for a particular problem. They aim to discover and expose the fundamental laws that govern the studied problem each time. They were first introduced by McCulloch and Pitts [33]. However, they were applied extensively from the decade 1990, mainly in medicine for the prediction of the disease of a patient according to a series of physicochemical parameters like age and hematological indices. Although that first application refers to medicine very quickly from the decade of the 1990's, the method of artificial neural networks was applied widely to the totality of sciences with a significant place to mechanical problems where the up to then classical deterministic mathematical methods were incapable of giving answers to multidimensional and complex problems with incredibly intense nonlinear behavior [34–40].

A classical feedforward ANN contains layers of nodes or neurons which have weighted connections with the nodes of the previous and preceding layers. Starting with the input layers in which each node presented for input, information or signals are propagated the subsequent layers until the information reaches the output layer. The output layer provided the predicted variable(s) and is compared with the labels of corresponding samples in the database. Figure 1 provides a typical example of the structure of ANN composed of one hidden layer.

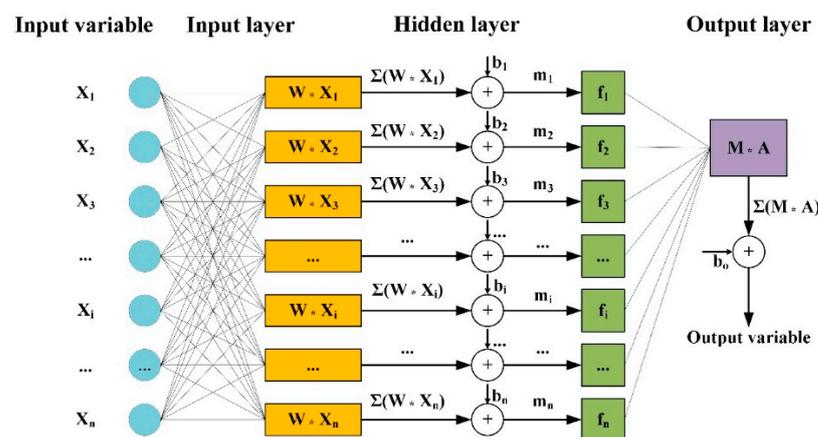


Figure 1. Architecture of an ANN model composed of one hidden layer.

The error of the predicted outputs and the labels is used to update the weights in the network. The weight adjustment is so-called the backpropagation and conducted each time a set of b samples (i.e., batch size) is “consumed”. This weight updating process is repeated until all samples of the train set are ingested, and an epoch is complete. The maximum number of epochs is set as the stop condition for the training process. In the end, a trained ANN model contains a set of optimized weights which provides the least error on the train set.

In work presented herein, the salient goal is developing a reliable and robust ANN model and deriving its closed-form equation to predict the 5-day biochemical oxygen demand (BOD_5). Specifically, for the estimation of BOD_5 in wastewater concerning COD, SS, TN, NH_4-N and TP, a plethora of different ANN architectures will be trained and developed. To this end, a detailed and in-depth investigation of the crucial parameters affecting the performance of ANN models, such as the number of neurons per hidden layers, activation functions, data normalization techniques and cost functions, has been conducted, and it is presented in the following sections.

3.2. Experimental Database

The primary target during a mathematical simulant’s training and development phase to predict the value of a parameter depending on several other parameters is the degree to which the proposed mathematical simulant is reliable and stable/robust. To this direction,

the majority of researchers give particular attention and diligence to the computational techniques and methods that shall employ for its development, while at the same time, they do not exercise the same attention and diligence concerning the database that shall be used for the training of the ANN. The authors of the present study consider that a mathematical simulant's reliability depends primarily on the reliability and effectiveness of the database that shall be used. We do not mean a database with a large amount of data with terms reliable and effective. Reliable and effective is considered a database with its data considered to be 'true' and covering statistically all the range of values capable of taking each of the parameters that infringe on each particular case studied problem.

The above has an even greater value when the database is comprised of experimental and not analytical results. In the case where the database is comprised of experimental data, its reliability is affected by a multitude of factors as follows: (a) the strict adherence to the international standards intended for the preparation of the specimens/samples and the conduct of the experimental tests-laboratory measurements; (b) the observance of the number of specimens with the same characteristics that must be checked; (c) the reliability of the experimental layout that was used; (d) the experience and specialization of the personnel that conducted the previous tests; and (e) the environmental conditions in which the aforementioned specimens were maintained as well the environmental conditions of the surroundings where the previous tests were conducted.

The observance requirement of the above rules is considered particularly imperative when the experimental database comprises experimental data produced from diverse laboratories and research groups.

According to the above principles for the training and development of a multitude of artificial neural networks and the selection among them of the best for the estimation of BOD₅ in wastewater, an experimental database was created, comprised of 387 datasets that correspond to 387 laboratory measurements that were conducted at the entrance of the sewage treatment plant located at Komotini region, Northern Greece. The samples were collected on a monthly basis from 2014–2021. Standard analytical methods were used to determine all parameters. All analytical methods were described in detail for water and wastewater experiments [41]. For each wastewater sample six water quality variables were laboratory measured. Precisely, for each sample were estimated the COD, SS, TN, NH₄-N, TP and BOD₅ concentrations. The measured values of the first five variables were used as input parameters, while the value of the sixth variable (BOD₅) as the output parameter during the training and development process of ANN models. The database is presented in Table S1 of Supplementary Materials.

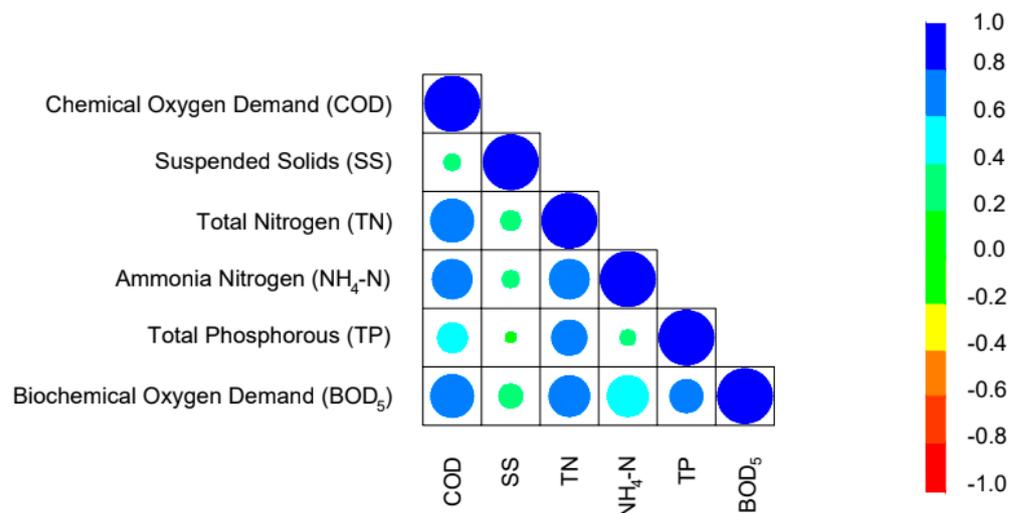
Table 3 presents for each parameter the minimum, average, and maximum value as well as the standard deviation (STD) and the coefficient of variation (CV). In Table 4 and Figure 2 the importance of the Pearson correlation factors between the six parameters are presented.

Table 3. Statistical analysis of the input and output parameters used in this research for the training and development of artificial neural networks.

Variable	Symbol	Units	Category	Data Used in NN Models				
				Min	Average	Max	STD	CV
Chemical Oxygen Demand	COD	mg L ⁻¹	Input	211.00	410.73	551.00	71.37	0.17
Suspended Solids	SS	mg L ⁻¹	Input	142.00	228.34	302.00	27.64	0.12
Total Nitrogen	TN	mg L ⁻¹	Input	44.20	66.44	84.25	9.17	0.14
Ammonia Nitrogen	NH ₄ -N	mg L ⁻¹	Input	39.30	52.52	70.10	7.67	0.15
Total Phosphorous	TP	mg L ⁻¹	Input	2.96	5.86	8.65	1.17	0.20
Biochemical Oxygen Demand	BOD ₅	mg L ⁻¹	Output	128.00	238.27	348.00	45.35	0.19

Table 4. Correlation matrix of the studied input and output variables.

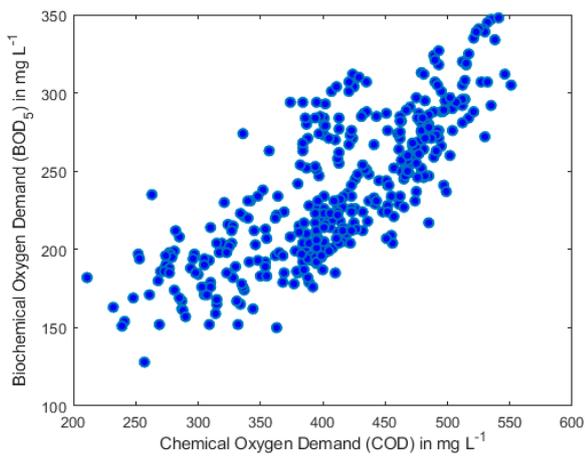
Variable	Symbol	COD	SS	TN	NH ₄ -N	TP	BOD ₅
Chemical Oxygen Demand	COD	1.00	0.29	0.78	0.72	0.54	0.78
Suspended Solids	SS	0.29	1.00	0.35	0.30	0.18	0.43
Total Nitrogen	TN	0.78	0.35	1.00	0.72	0.64	0.74
Ammonia Nitrogen	NH ₄ -N	0.72	0.30	0.72	1.00	0.27	0.58
Total Phosphorous	TP	0.54	0.18	0.64	0.27	1.00	0.60
Biochemical Oxygen Demand	BOD ₅	0.78	0.43	0.74	0.58	0.60	1.00

**Figure 2.** Pearson correlation coefficients between the examined variables.

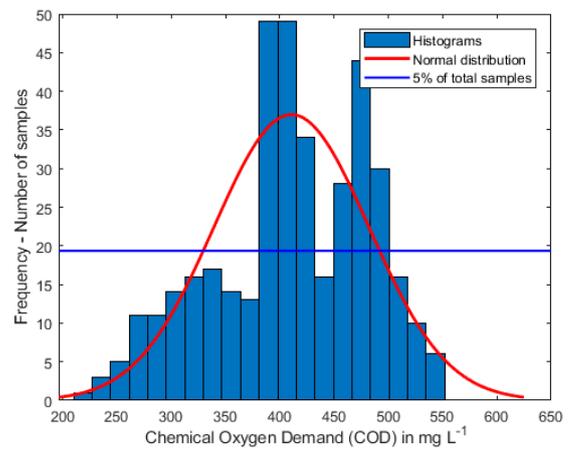
These values are beneficial since they indicate if there is a strong dependence of one parameter on the other. Additionally, the values of the last row of the table are indicative at the first level if there is dependence between each of the five input parameters with the output variable. It is observed that in rank, the parameters with the most significant correlation with the biochemical oxygen demand are the COD, the TN and the TP, with Pearson correlations factors 0.78, 0.74 and 0.60, respectively. In the following subsection, they will be presented thoroughly and in-depth the sensitivity analysis results of the output parameter of BOD₅ in relation with each one of the five input parameters using the Cosine Amplitude Method (CAM) [42] and the experimental database. Researchers have widely adopted CAM method to determine the effect of each input on the output [43–46].

In Figure 3 the histograms are presented for each of the six variables, and graphs showing the correlation between each of the input parameters with the BOD₅.

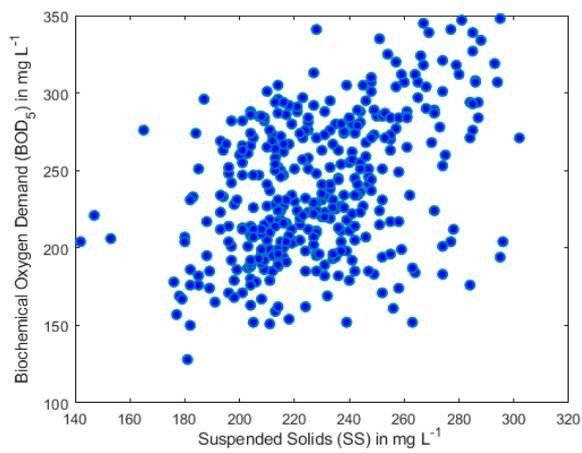
These graphs are particularly useful since they depict for each parameter the range of values it takes and their distribution. They are beneficial since, for the ranges where we have sufficient data, the reliability of the model to be proposed shall be exceptionally high. These ranges correspond to the values of parameters for which the number of samples is greater than 5% of the dataset and are defined by the blue horizontal line in each graph. For the areas that are under the blue line, it is required to be done enrichment of the database with more data.



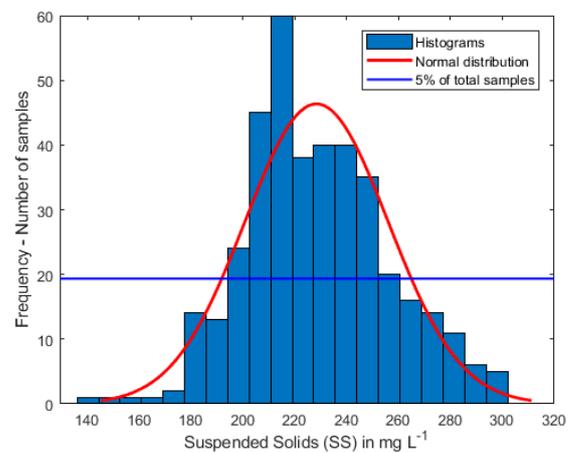
(a)



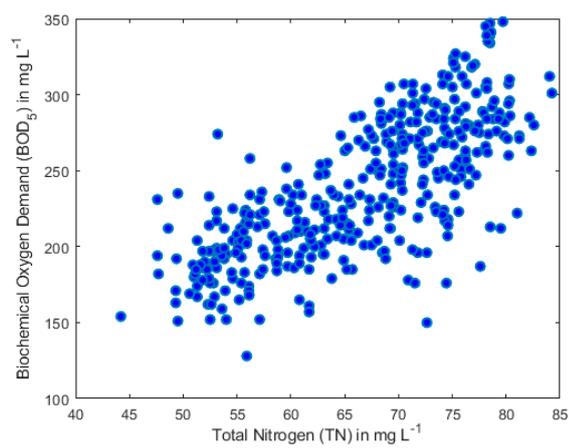
(b)



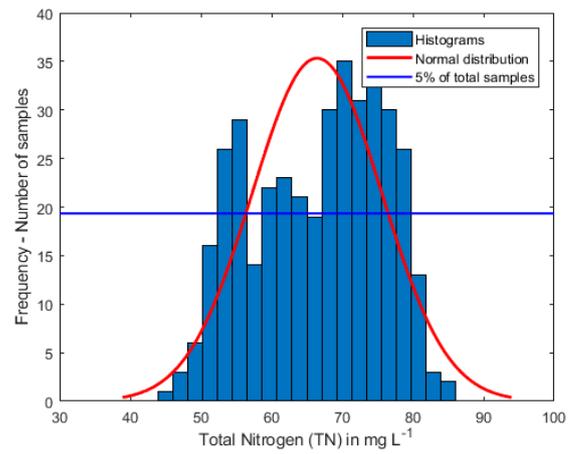
(c)



(d)



(e)



(f)

Figure 3. Cont.

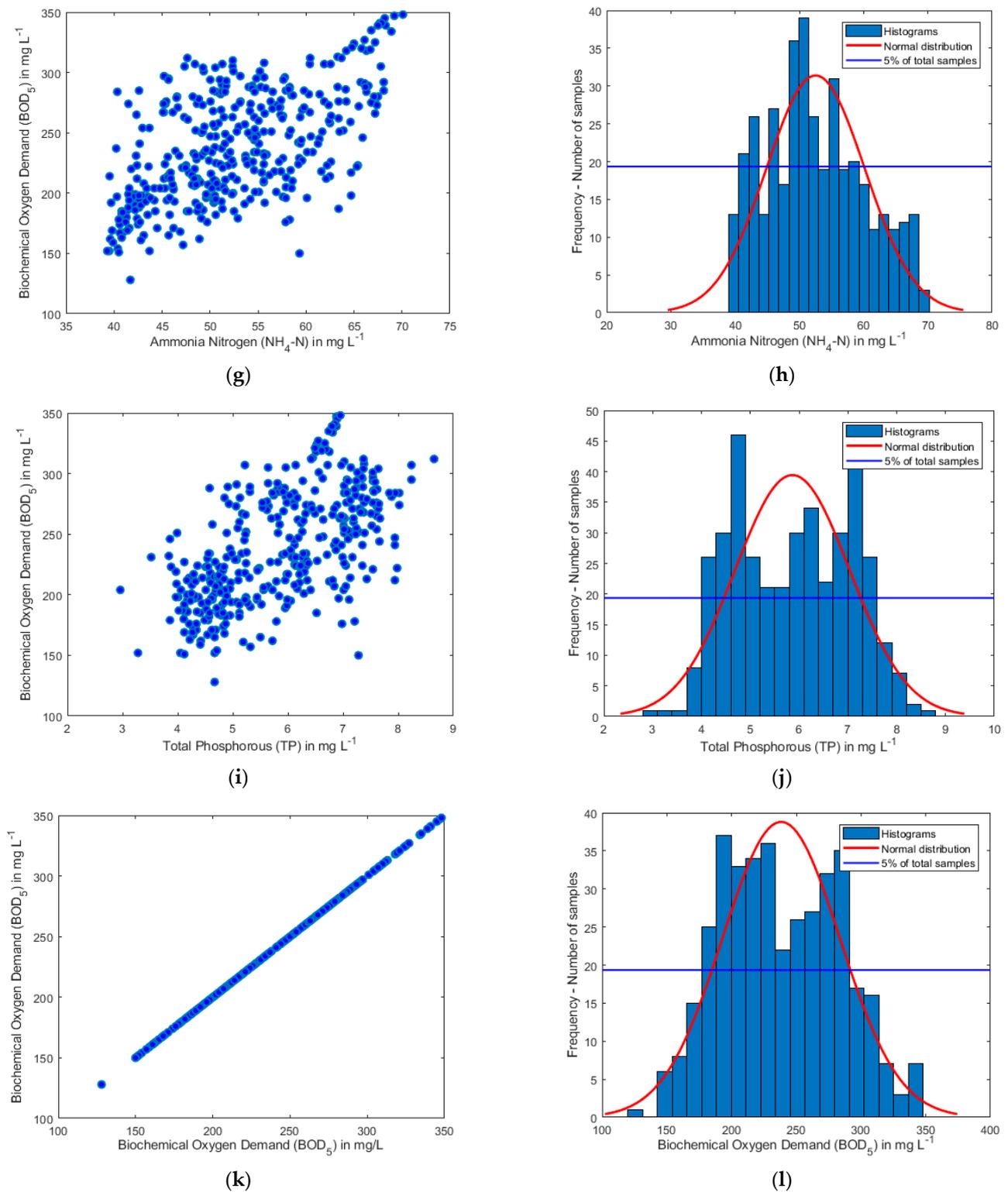


Figure 3. Scatter plots and histograms of the studied input and output parameters: (a,b) COD; (c,d) SS; (e,f) TN; (g,h) NH₄-N; (i,j) TP; and (k,l) BOD₅.

3.3. Sensitivity Analysis of the BOD₅ on the Input Parameters Based on the Experimental Database

During the training and development of the computational models for the prediction of the value of a parameter (output parameter) as a function of several other parameters (input parameters) that enter each case problem, it is interesting to explore the sensitivity that is

exhibited by the output parameter in terms of the input parameters. This is particularly useful for it enables us to exclude a series of parameters that do not affect the estimated parameter but, at the exact moment, guides to exhibit intense attention to the parameters that significantly affect the output parameter value. The exclusion of several parameters from the current studied problem has decreased the computing time and helped to discover the problem's nature and governing laws.

A first estimation of the dependence between the parameters and much more for the input parameters with the output parameter is given by the value of the Pearson correlation factor. Additionally, because of the subject's great importance, they have proposed a multitude of sensitivity analysis methods with a target the as much possible estimation of the sensitivity and dependence of the output parameter from the input parameters. Between them, there is the cosine amplitude method (CAM), which has been proposed by Jong and Lee [42] and has been accepted by a multitude of researchers [35,47–51].

The cosine amplitude method (CAM) was used to construct a data array, X , as follows:

$$X = \{x_1, x_2, x_3, \dots, x_i, \dots, x_n\} \quad (1)$$

where variable x_i in array, X is a length vector of m expressed by:

$$x_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}\} \quad (2)$$

The relationship between R_{ij} (strength of the relation) and datasets of x_i and x_j defined by:

$$R_{ij} = \frac{\sum_{k=1}^m x_{ik}x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2 \sum_{k=1}^m x_{jk}^2}} \quad (3)$$

The result of the sensitivity analysis, based on the datasets of the experimental database used in the present work, is presented in Figure 4. It depicted that from the five input parameters, the COD and TN had the strongest influence on the BOD₅ (Figure 4). This finding fully agrees with Pearson correlation factors presented in the preview's subsection. Furthermore, it is worth noting that all the input parameters can be characterized as crucial since they have also strongly related to BOD₅, achieving values greater than 0.98.

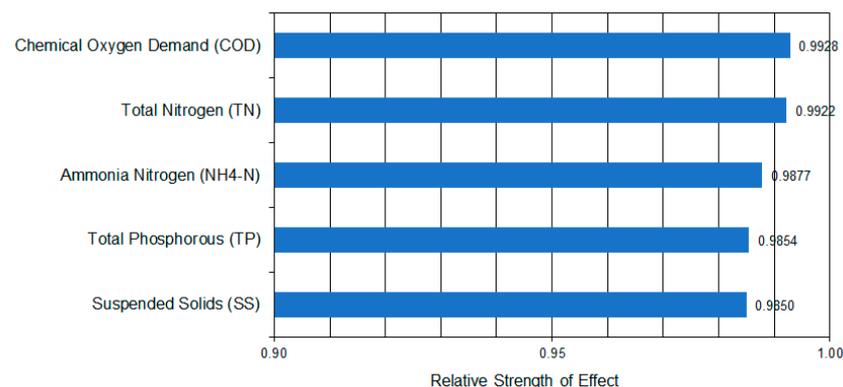


Figure 4. Sensitivity analysis of the Biochemical Oxygen Demand (BOD₅) on input parameters based on the experimental database.

3.4. Performance Indexes

The predictive accuracy of the developed neural network models was assessed using the root mean square error (RMSE), the mean absolute percentage error (MAPE), and the Correlation Coefficient (R^2) [49,51–55].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_i - y_i}{x_i} \right| \quad (5)$$

$$R^2 = 1 - \left(\frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right) \quad (6)$$

where n denotes the total number of datasets, and x_i and y_i represent the predicted and target values, respectively.

Recent research has highlighted the limitations of the root mean square error (RMSE), the mean absolute percentage error (MAPE), and the Correlation Coefficient (R^2) to assess the predictive accuracy of models [30,53]. The comparison of the performance of mathematical using the Pearson correlation factor is considered precarious given that except the comparison of the values of R or R^2 it is also required the comparison of the inclination angle of the line. Such a case is when a mathematical simulant always predicts a constant value regardless of the input parameters values. In this case, the value of $R = 1.00$ while the inclination angle is equal to zero.

To this end, the a20-index, has been recently proposed [43,56–62] for assessing the reliability of neural networks:

$$\text{a20-index} = \frac{m20}{M} \quad (7)$$

where M denotes the number of dataset samples and $m20$ denotes the number of samples with a ratio of the true value to the estimated-predicted value between 0.80 and 1.20. In an ideal forecasting model, the value of a20-index is equal to 1.00. The proposed a20-index is a simple statistical index having the advantage to include a physical engineering meaning, as it reveals the number of experiments that satisfy the predicted values with a 20% deviation from the 'true' values.

At this point it is worth stressing the very large significance of the database that shall be used for the training and development of the soft computing-based forecasting models. A comparison using different performance indices must be referring to an adequate number of data and indeed to be reliable must be based on the same database.

4. Results and Discussion

4.1. Development of ANN Models

Several hyperparameters and neural network structure/architecture must be determined ahead of time in the context of the training and development ANN models. This provides the benefit of developing an ANN model that is exceptionally optimized for the problem under investigation. However, unless a certain level of expertise is available, selecting appropriate values for these parameters and appropriate neuron layouts can be intimidating. In other words, the "human in the loop" concept is thought to be critical for tuning an efficient ANN model. Furthermore, special care about the overfitting problems should be paid. The selection of the optimum model among the plethora of training and developed soft computing models should be based not only on statistical indices but also on the derivation of curves which should be smooth revealing the nature of the problem under investigation.

However, in this study, the optimal ANN structure is not selected based on expertise or intuition. However, it is derived from an optimization procedure that trains and tests ANNs using a plethora of alternative hyper parameter combinations and ranks them according to predefined performance indices as well as it mentioned above taking care whether overfitting of the data taking place. Except for the fixed number of hidden layers, the optimization procedure combines the following parameters: (a) data normalization; (b) the number of neurons in the hidden layer; (c) cost function; and (d) activation function.

Table 5 shows the alternative options for these parameters, as well as some predetermined configuration options. Twelve different activation functions are investigated, including the common Hyperbolic Tangent Sigmoid (HTS), Log-sigmoid (LS), and Linear (Li) functions. If all other parameters are held constant, this results in 144 (12×12) alterna-

tive combinations to be trained and tested. Regarding the used cost functions, the MSE and SSE functions are investigated, whereas four data normalization techniques are applied on the input and output parameters, including no normalization at all. Considering the varying random number generator, all of these alternative ANN configuration parameters resulted in 576.000 different ANN models under evaluation (i.e., $50 \times 10 \times 42 \times 4 \times 10$) (10 alternatives). At this point, it should be pointed out that the Levenberg–Marquardt algorithm (LM) has been applied during training process of ANN mathematical models.

Table 5. Hyperparameters for the training and development of ANN models applied in this study.

Parameter	Value	Matlab Function(s)
Training Algorithm	Levenberg-Marquardt Algorithm	trainlm
Normalization	Without any normalization Minmax in the range [0.10–0.90], [0.00–1.00] and [–1.00–1.00] Zscore	Mapminmax zscore
Number of Hidden Layers	1	
Number of Neurons per Hidden Layer	1 to 50 by step 1	
Control random number generation	10 different random generation	rand(seed, generator), where the generator range from 1 to 10 by step 1
Training Goal	0	
Epochs	200	
Cost Function	Mean Square Error (MSE) Sum Square Error (SSE)	mse sse
Transfer Functions	Hyperbolic Tangent Sigmoid transfer function (HTS) Log-sigmoid transfer function (LS) Linear transfer function (Li) Positive linear transfer function (PLi) Symmetric saturating linear transfer function (SSL) Soft max transfer function (SM) Competitive transfer function (Co) Triangular basis transfer function (TB) Radial basis transfer function (RB) Normalized radial basis transfer function (NRB) Hard-limit transfer function (HL) Symmetric hard-limit transfer function (SHL)	tansig logsig purelin poslin satlins softmax compet tribas radbas radbasn hardlim hardlims

The training of the 576.000 alternative ANN architectures did not use the entire dataset of 387 samples. The dataset was divided into three sub-datasets to evaluate the generality of the developed ANN model: the first dataset included 66.7% of the entire database (258 specimens) and was used for training the ANN architectures, the second dataset included 16.8% of the entire database (65 specimens) and was used for testing the ANN architectures, and the third dataset included the remaining 16.5% of the entire sample (64 specimens). The three datasets are referred to as “training datasets”, “testing datasets”, and “validation datasets”, in that order. To eliminate potential bias, the sample was randomly divided into three datasets using a programmatic procedure.

Table 6 tabulates the top twenty ANN architectures based on their achieved values of RMSE index and for the case of testing datasets and configuration parameters. The first, which is preferred as the best and is regarded to as ANN LM 5-8-1 (the numbers refer to the five (5) input parameters, the eight (8) neurons in the hidden layer and the one (1) output parameter which is the BOD₅), achieves the best overall performance metrics, both in terms of RMSE (16.8563) and R (0.9443). The best ANN LM 5-8-1 model utilizes

the MinMax function for data normalization, which converts input and output values between $[-1.00, 1.00]$. It also applies as activation functions the Log-Sigmoid function (LS) for the input layer and the Symmetric saturating linear function (SSL) for the output layer, with the MSE function as its cost function. Figure 5 exhibits the neuron layout and overall architecture of the best ANN LM 5-8-1 model. Table 7 displays detailed and in-depth achieved performance indices for both training and testing datasets of the best ANN LM 5-8-1 model. Its performance on training datasets is expected to improve, particularly in the a20-index, where it matches over 98% of the samples within a 20 percent margin. The same index is 100.00% when compared to the testing datasets, which is an excellent value. At this point should be noted that the better achieved indices for testing datasets compared to training datasets clearly depict that not overfitting problems is taken place. To the authors best knowledge, the achieved performance is the better than any other performance reported in the related topic.

Table 6. Architectures and hyperparameters of the top twenty developed ANN LM models based RMSE index and testing phase.

Ranking	Normalization Technique	Cost Function	Transfer Function		Architecture	Datasets Performance Indices					
			Input Layer	Output Layer		Testing		Training		All	
						R	RMSE	R	RMSE	R	RMSE
1	Minmax $[-1.00, 1.00]$	MSE	logsig	satlins	5-8-1	0.9443	16.8563	0.9208	17.6044	0.9217	17.6065
2	Minmax $[0.10, 0.90]$	MSE	poslin	satlins	5-29-1	0.9421	17.8418	0.9057	19.1355	0.9110	18.7234
3	Minmax $[-1.00, 1.00]$	SSE	tribas	tansig	5-14-1	0.9407	17.7982	0.9319	16.3818	0.9287	16.8636
4	Minmax $[0.10, 0.90]$	MSE	tansig	purelin	5-13-1	0.9406	17.5336	0.9238	17.2764	0.9214	17.6425
5	Minmax $[0.10, 0.90]$	MSE	softmax	radbas	5-17-1	0.9400	17.7336	0.9156	18.1454	0.9164	18.2035
6	Minmax $[-1.00, 1.00]$	MSE	tansig	purelin	5-11-1	0.9400	17.5935	0.9298	16.6148	0.9224	17.5536
7	Minmax $[0.10, 0.90]$	MSE	logsig	satlins	5-8-1	0.9399	17.5855	0.9161	18.1001	0.9179	18.0053
8	Minmax $[0.10, 0.90]$	SSE	softmax	logsig	5-29-1	0.9396	17.7540	0.9311	16.4914	0.9275	17.0689
9	Minmax $[0.10, 0.90]$	SSE	satlins	purelin	5-15-1	0.9394	18.6357	0.9012	19.6172	0.9066	19.3523
10	Minmax $[0.00, 1.00]$	MSE	logsig	satlins	5-5-1	0.9393	18.7060	0.9136	18.4874	0.9150	18.5984
11	Minmax $[0.10, 0.90]$	SSE	softmax	satlins	5-26-1	0.9393	18.1376	0.9077	18.9646	0.9116	18.7815
12	Minmax $[0.00, 1.00]$	MSE	softmax	purelin	5-41-1	0.9388	17.8211	0.9217	17.5291	0.9215	17.7143
13	Minmax $[0.10, 0.90]$	MSE	tansig	purelin	5-8-1	0.9388	17.6942	0.9160	18.1425	0.9179	17.9936
14	Minmax $[0.10, 0.90]$	MSE	softmax	radbas	5-23-1	0.9388	19.0606	0.9135	18.4091	0.9155	18.4651
15	Minmax $[0.10, 0.90]$	MSE	logsig	logsig	5-12-1	0.9387	16.7798	0.9320	16.3995	0.9282	16.8627
16	Minmax $[0.00, 1.00]$	MSE	softmax	purelin	5-22-1	0.9387	18.2816	0.9105	18.6790	0.9125	18.6410
17	Minmax $[0.00, 1.00]$	MSE	softmax	poslin	5-22-1	0.9387	18.2816	0.9105	18.6790	0.9125	18.6410
18	Minmax $[0.00, 1.00]$	MSE	tansig	satlins	5-6-1	0.9386	17.5546	0.9164	18.0637	0.9166	18.1632
19	Zscore	MSE	poslin	purelin	5-9-1	0.9385	18.1809	0.9076	18.9472	0.9105	18.7648
20	Minmax $[0.00, 1.00]$	SSE	tansig	purelin	5-7-1	0.9384	17.1610	0.9100	18.7052	0.9110	18.7294

Table 7. Summary of prediction capability of the optimum ANN LM 5-8-1 model.

Model	Datasets	Performance Indices				
		a20-Index	R	RMSE	MAPE	VAF
ANN LM 5-8-1	Training	0.9806	0.9208	17.6044	0.0582	84.7803
	Test	1	0.9443	16.8563	0.0571	89.1499

Furthermore, values for the training and testing datasets are presented in Figure 6 as scatter plots of the ‘true’ vs. predicted by the best ANN LM 5-8-1 model. Except for the diagonal line (ideal prediction), two more dotted lines are drawn in these diagrams, indicating a $\pm 20\%$ error. Furthermore, Figure 6 is a more useful figure depicts the ratio of ‘true’ values to predicted values of BOD₅ in wastewater both for training and testing datasets.

4.2. Closed-Form Equation for the Estimation of BOD₅ in Wastewater

As presented in the previous section, the ANN LM 5-8-1 model is the best among the many different trained and developed architectures. That is the best is documented that this model has optimal values for the performance indices used for ranking other ANN models. It is worth noting that this is a standard procedure for the multitude of scientific publications.

At this point, the authors consider it necessary to state clearly that such a procedure is not safe and reliable because it is impossible to test the reliability-validity of these values.

Additionally, the results of such a research work are not immediately applicable to the scientists of this discipline and much more to the engineers in practice.

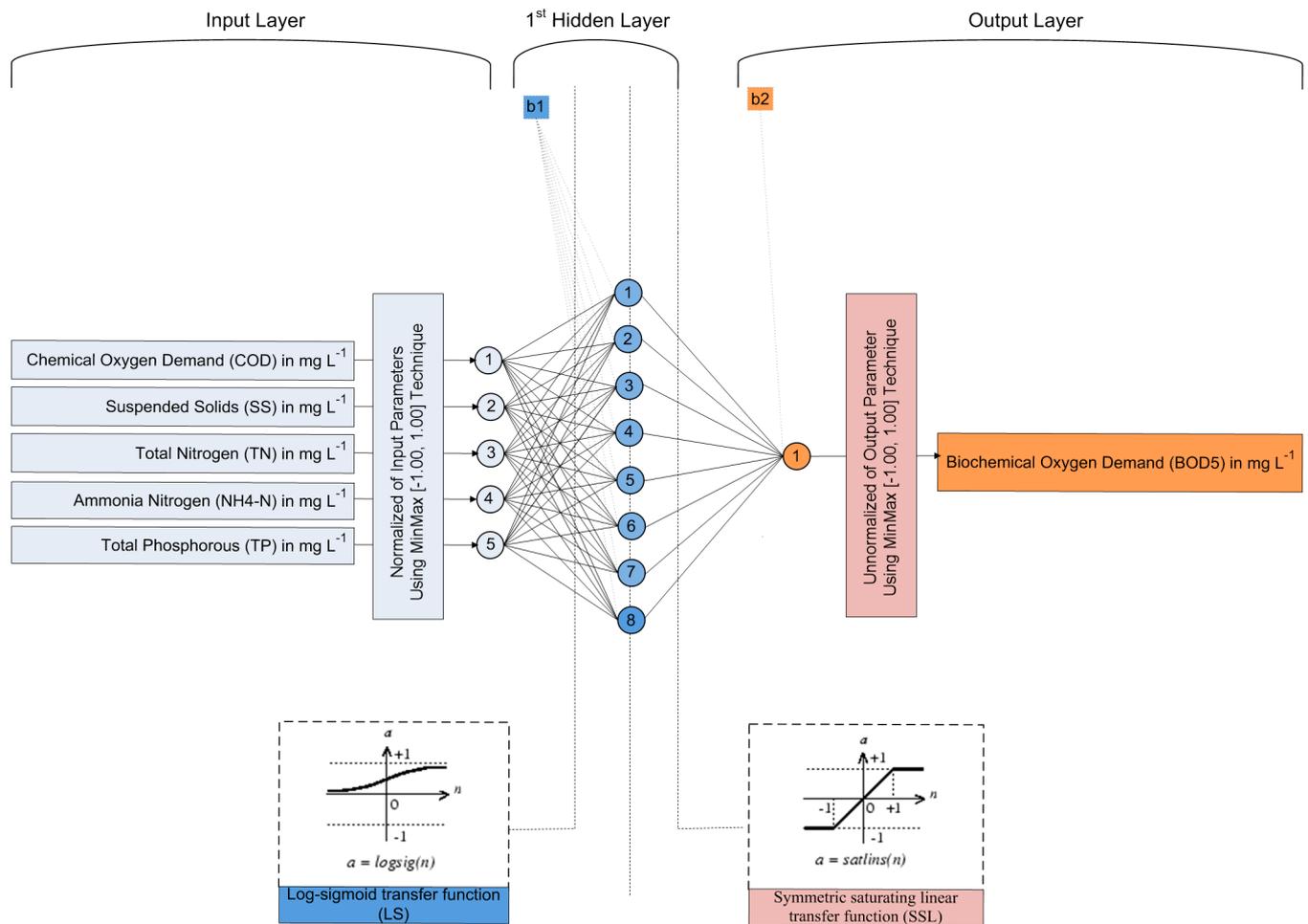


Figure 5. Architecture of the optimum ANN LM 5-8-1 model.

Intending to treat the above weaknesses, the authors consider it necessary to present the architecture corresponding to the best ANN model and the final values of weights and biases to make the design of the mathematical simulant possible. Giving the mathematical equation that describes the best mathematical simulant would be more beneficial. Under the prism of all the above in the present section, the derived equation for the prediction of both normalized and absolute values of BOD₅, using COD, SS, TN, NH₄-N and TP are expressed by the following equation for the optimum developed ANN LM 5-8-1 model:

$$BOD5_{norm} = \text{satlins}([LW\{2,1\}] \times [\text{logsig}([IW\{1,1\}] \times [IP] + [B\{1,1\}])] + [B\{2,1\}]) \quad (8)$$

$$BOD5_{real} = \frac{(BOD5_{norm} - a) \times (BOD5_{max} - BOD5_{min})}{b - a} + BOD5_{min} \quad (9)$$

where $a = -1.00$ and $b = 1.00$ are the lower and upper limits of the minmax normalization technique applied on the data, $BOD5_{max} = 348$ and $BOD5_{min} = 128$ are the maximum and minimum values of BOD₅ present in the database used for training and developing ANN models. The *satlins* and *logsig* are the symmetric saturating linear transfer function (SSL) and the Log-sigmoid transfer function (LS), respectively, as discussed. Their details (equations and graphs) are presented in detail in Table A1 of the Appendix A.

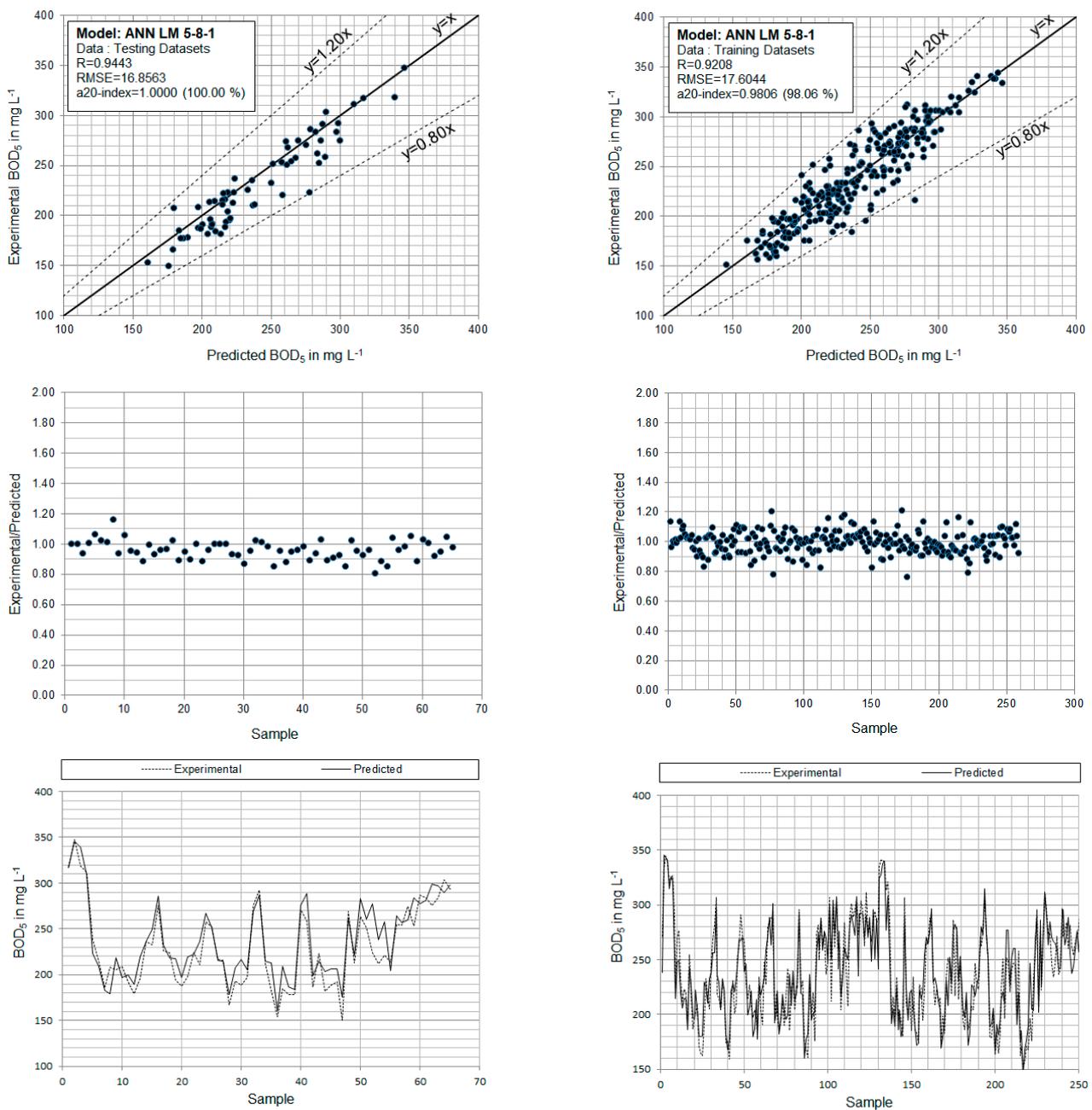


Figure 6. Experimental vs. predicted values of the five-day biochemical oxygen demand (BOD₅) in wastewater for training and testing datasets, using the developed ANN LM 5-8-1 model.

Equation (9) describes the developed ANN LM 5-8-1 model in a purely mathematical form so that its reproduction becomes straightforward. In Equation (8), $[IW\{1,1\}]$ is a 8×5 matrix that contains the weights of the hidden layer; $[LW\{2,1\}]$ is a 1×8 vector with the weights of the output layer; $[IP]$ is a 5×1 vector with the five (5) input variables; $[B\{1,1\}]$ is a 8×1 vector that contains the bias values of the hidden layer; and $[B\{2,1\}]$ is a 1×1 vector with the bias of the output layer. The $[IP]$ vector contains the five (5) normalized values of the input variables (COD, SS, TN, NH_4 -N and TP). It is expressed as:

$$[IP] = \begin{bmatrix} a + (b - a) \left(\frac{COD - \min(COD)}{\max(COD) - \min(COD)} \right) \\ a + (b - a) \left(\frac{SS - \min(SS)}{\max(SS) - \min(SS)} \right) \\ a + (b - a) \left(\frac{TN - \min(TN)}{\max(TN) - \min(TN)} \right) \\ a + (b - a) \left(\frac{NH_4 - N - \min(NH_4 - N)}{\max(NH_4 - N) - \min(NH_4 - N)} \right) \\ a + (b - a) \left(\frac{TP - \min(TP)}{\max(TP) - \min(TP)} \right) \end{bmatrix} \quad (10)$$

where the $\min(COD) = 211$, $\max(COD) = 551$; $\min(SS) = 142$, $\max(SS) = 302$; $\min(TN) = 44.20$, $\max(TN) = 84.25$; $\min(NH_4 - N) = 39.30$, $\max(NH_4 - N) = 70.10$ and $\min(TP) = 2.96$, $\max(TP) = 8.65$ are the input parameters minimum and maximum values (shown in Table 3). The values of final weights and biases that determine the matrices $[IW\{1, 1\}]$, $[LW\{2, 1\}]$, $[B\{1, 1\}]$ and $[B\{2, 1\}]$ are presented in Table 8.

Table 8. Finalized weights and bias of the optimum ANN-LM 5-8-1 model.

		$[IW\{1, 1\}]$				$[LW\{2, 1\}]^T$	$[B\{1, 1\}]$	$[B\{2, 1\}]$
		(8×3)				(1×8)	(8×1)	(1×1)
3.8747	-3.3026	0.8841	9.8281	-1.9581	2.9497	9.6809	-0.4917	
-3.8650	3.1701	0.9485	-1.1494	1.9099	-2.6746	2.1566		
1.2295	-2.3973	-1.2553	2.7258	-2.3405	-2.3231	1.0974		
-4.6036	3.7836	0.8690	-4.0565	7.1579	-1.1579	-1.3150		
-3.0096	3.4372	0.3911	-0.1472	2.3667	2.2639	2.2077		
-1.0593	-4.5452	1.0995	-4.5926	-1.9055	-0.6063	6.0657		
-4.7120	3.0951	-2.9974	-9.7771	2.9426	2.6242	-10.6922		
1.6124	-9.0346	0.1278	3.6006	8.0047	0.3364	7.7832		

Note: $[IW\{1, 1\}]$ is the matrix of weight values between the input layer and the first hidden Layer; $[LW\{2, 1\}]$ is the matrix of weight values between the 1st hidden layer and the output layer; $[B\{1, 1\}]$ is the matrix of bias values for hidden layer, and $[B\{2, 1\}]$ is the matrix of bias values for the output layer.

In this form of matrix multiplication, the prediction equation Equation (9) can be easily programmed in an Excel spreadsheet, and therefore it can be more easily evaluated and used in practice. It is worth noting that such an implementation can be used by various interested parties (i.e., researchers, students, engineers) without placing heavy requirements on effort and time.

4.3. Mapping of BOD₅

With the proposed optimal ANN-LM 5-8-1 model, a thorough analytical investigation was conducted of the parameters that affect the value of BOD₅. Based on the results of this analytical investigation they were derived a set of contour maps of the BOD₅ in relation of the input parameters (Figures 7–12). Based on these figures, it is shown in a robust manner that the proposed ANN-LM 5-8-1 model ensures that the known and widely encountered phenomenon of overfitting is not taking place. This is implied by the fact that all the derived charts and the derived curves are exceptionally smooth and do not display sudden variations having as a result to exhibit the laws that govern the variation of BOD₅ concerning COD, SS, TN, NH₄-N and TP.

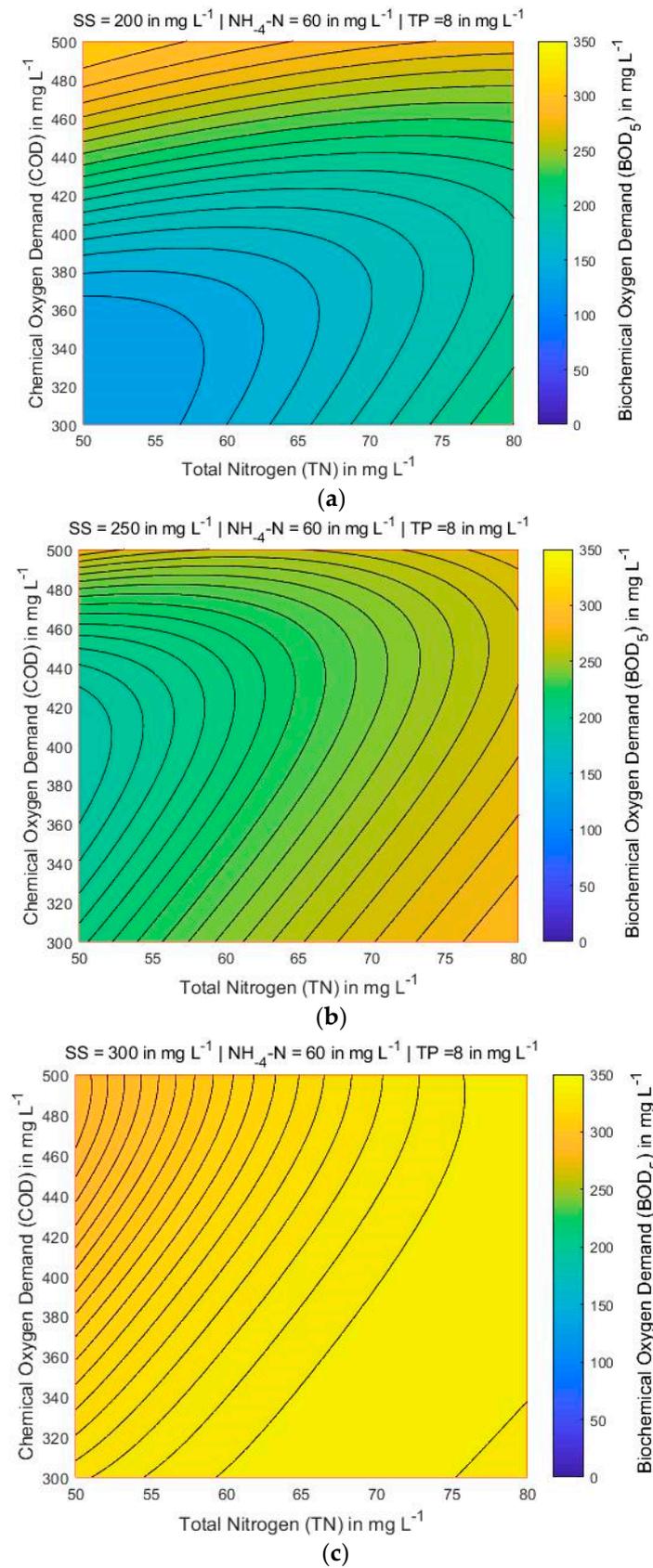


Figure 7. BOD_5 contour maps for three different values of SS (in mg L^{-1}) while $\text{NH}_4\text{-N} = 60$ mg L^{-1} and $\text{TP} = 8$ mg L^{-1} are constant: (a) $\text{SS} = 200$; (b) $\text{SS} = 250$; and (c) $\text{SS} = 300$.

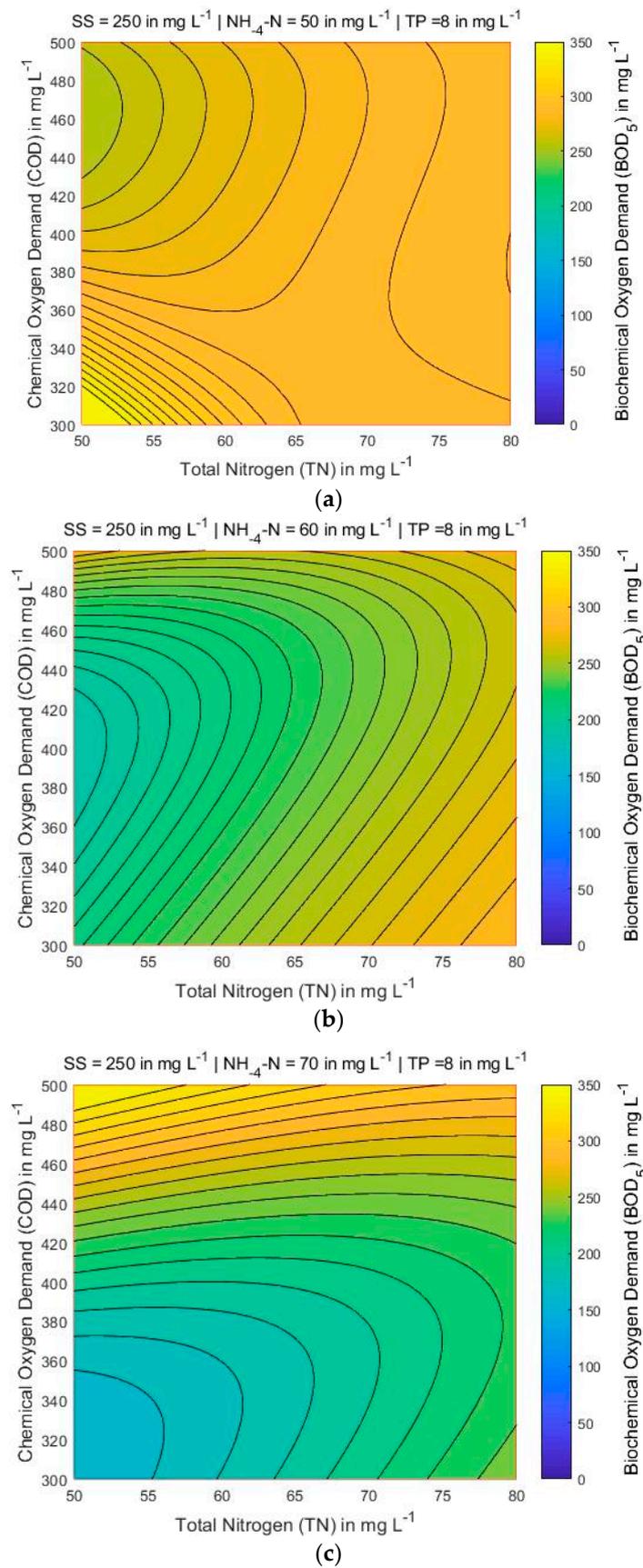


Figure 8. BOD₅ contour maps for three different values of NH₄-N (in mg L⁻¹) while SS = 250 mg L⁻¹ and TP = 8 mg L⁻¹ are constant: (a) NH₄-N = 50; (b) NH₄-N = 60; and (c) NH₄-N = 70.

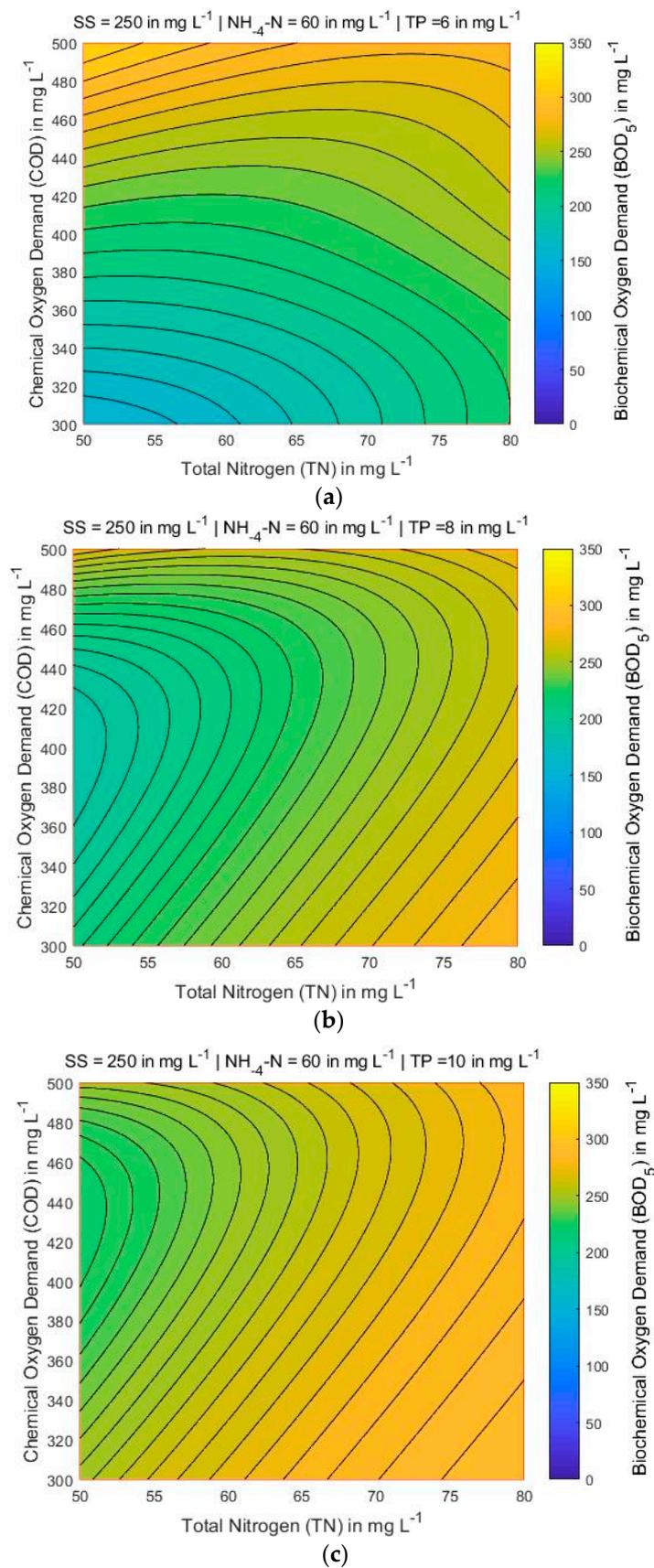
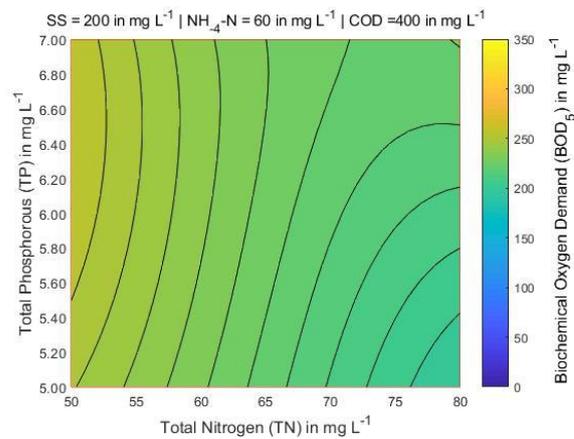
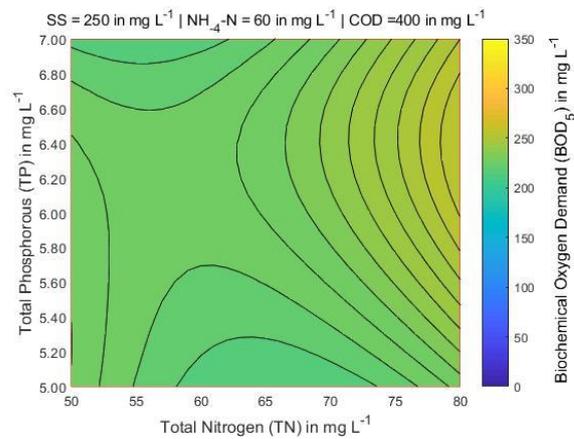


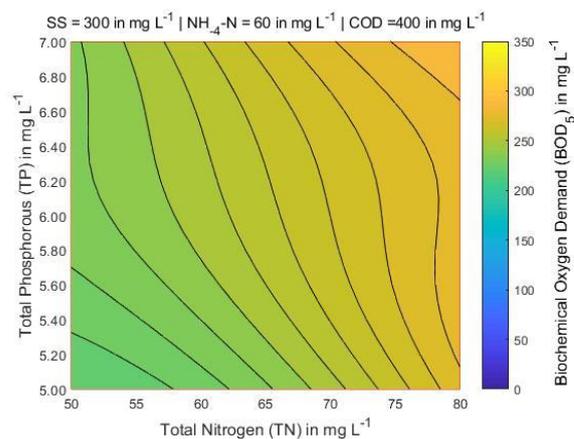
Figure 9. BOD_5 contour maps for three different values of TP (in mg L^{-1}) while $\text{SS} = 250 \text{ mg L}^{-1}$ and $\text{NH}_4\text{-N} = 60 \text{ mg L}^{-1}$ are constant: (a) TP = 6; (b) TP = 8; and (c) TP = 10.



(a)



(b)



(c)

Figure 10. BOD_5 contour maps for three different values of SS (in mg L^{-1}) while $\text{NH}_4\text{-N}$ = 60 mg L^{-1} and COD = 400 mg L^{-1} are constant: (a) SS = 200; (b) SS = 250; and (c) SS = 300.

Figure 7 shows that for the lowest COD and TN values, the BOD_5 also has the lowest value, searching all over the map area when $\text{NH}_4\text{-N}$ and TP are 60 and 8 mg L^{-1} , respectively, and SS varies between 200 and 300 mg L^{-1} . Figures 8 and 9, the variations of COD and TN present smooth curvature, searching all over the map area. A more detailed look at Figure 8 (left corner) shows that the lowest contents of COD and TN, the BOD_5 presents the highest value only in one part of the map. It is found that the COD and TN are sensitive to

BOD parameter. Figures 10–12 presents that for lowest contents of TP and TN, the BOD has moderate value when SS ranges from 250 and 300 mg L⁻¹, and NH₄-N and COD is 60 and 400 mg L⁻¹, respectively. A more detailed look at Figure 12, presents that the lowest concentrations of TP and TN, the BOD shows the highest value in the significant part of the map.

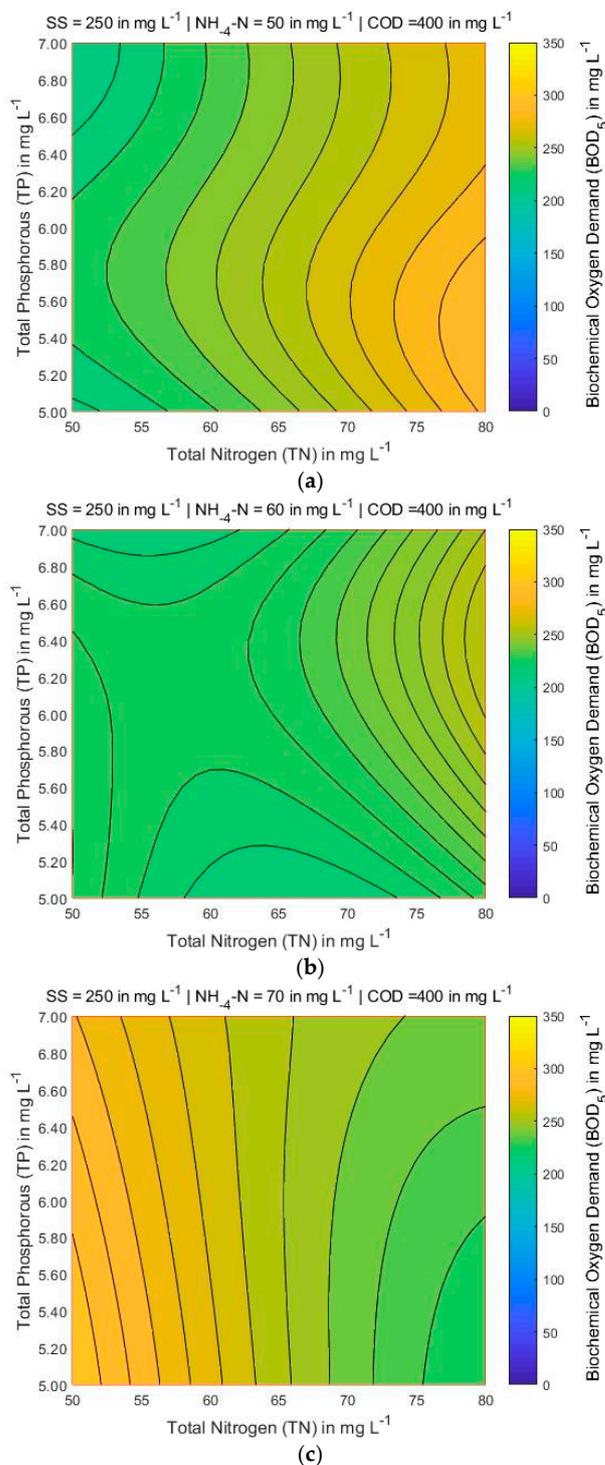


Figure 11. BOD₅ contour maps for three different values of NH₄-N (in mg L⁻¹) while SS = 250 mg L⁻¹ and COD = 400 mg L⁻¹ are constant: (a) NH₄-N = 50; (b) NH₄-N = 60; and (c) NH₄-N = 70.

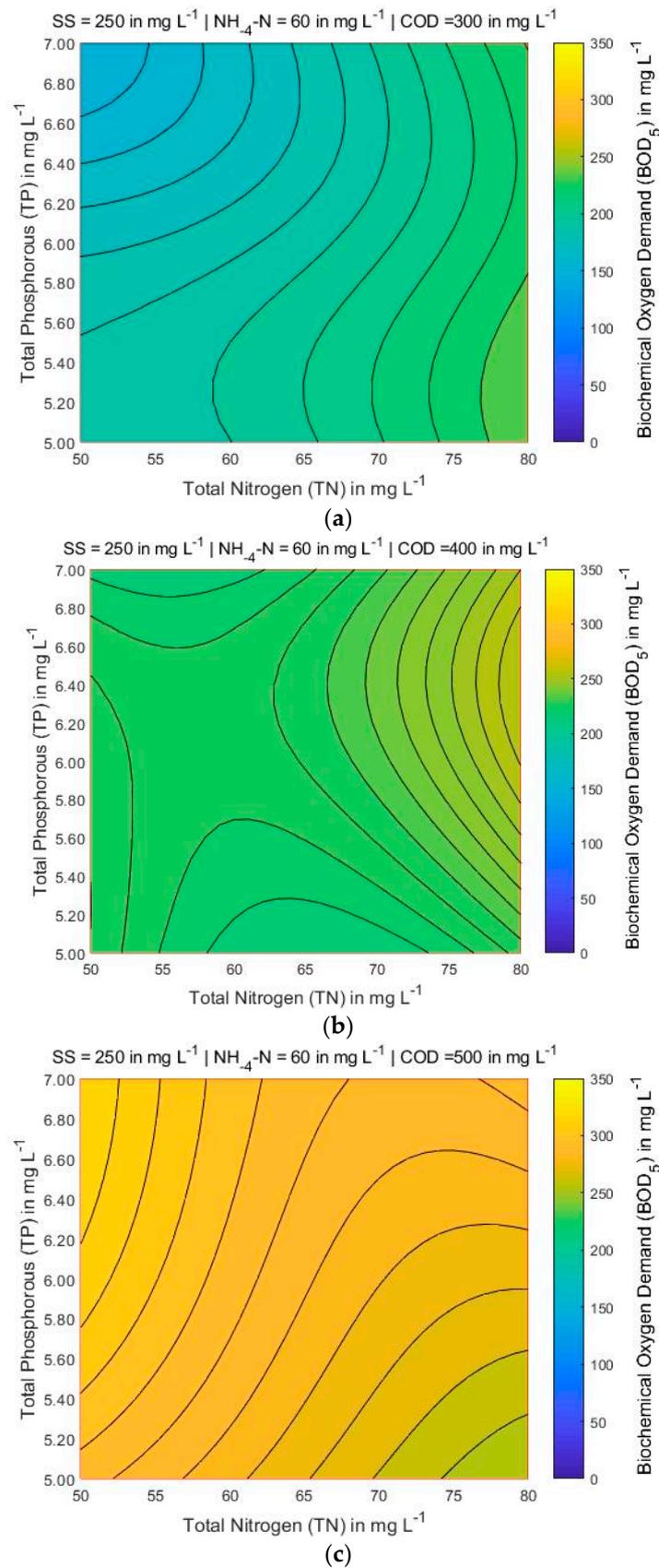


Figure 12. BOD₅ contour maps for three different values of COD (in mg L⁻¹) while SS = 250 mg L⁻¹ and NH₄-N = 60 mg L⁻¹ are constant: (a) COD = 300; (b) COD = 400; and (c) COD = 500.

5. Limitations and Future Works

The proposed optimal ANN-LM 5-8-1 model, like any other mathematical simulant, has validity for values of the input parameters between the minimum and maximum values of the database that was used for the training of ANN models (Table 3). Additionally, the reliability of the proposed model is exceptionally high for ranges of the values of the parameters where according to the histograms that were presented in the previous section (Figure 3), there exists sufficient data. For the regions where the data are not considered enough, we must update the database with further data that cover these areas satisfactorily. Based on those mentioned above, the authors' aims include updating the database and data from measurements from different sewage processing plants with the target of formulating one even more reliable model for estimating BOD₅ in wastewater.

6. Conclusions

The proposed ANN LM 5-8-1 approach can save costs and time for actual laboratory measurements. In other words, it is a practical need to illustrate a machine learning approach to conduct BOD estimation and receive accurate findings. The variation of COD and TN exhibit smooth curvature. COD and TN are found to be sensitive to the BOD parameter. The proposed optimal ANN model is valid for input parameter values between the minimum and maximum values of the database used for ANN model training. Furthermore, the proposed model's reliability is exceptionally high for parameter value ranges where there is sufficient data. The developed and proposed ANN model proved to be a robust and valuable tool for scientists, researchers, engineers and practitioners in monitoring water systems and the design phase of wastewater treatment plants. Moreover, it is an illustrative example of ANN methodology for environmental and educational applications.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/w15010103/s1>, Table S1: Experimental database used for the training, testing and development of ANN models.

Author Contributions: Conceptualization: P.G.A. and D.E.A.; methodology: P.G.A. and D.E.A.; software: P.G.A. and M.Z.T.; formal analysis and investigation: M.Z.T., D.E.G. and D.G.; writing—original draft preparation: all the authors; writing—review and editing: all the authors; supervision: P.G.A. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

Informed Consent Statement: Informed consent was obtained from all individual participants included in the study.

Data Availability Statement: The raw/processed data required to reproduce these findings will be made available on request.

Conflicts of Interest: The authors declare that there is no conflict of interest regarding the publication of this paper.

Notation

ANN(s)	Artificial Neural Network(s)
BOD	Biochemical Oxygen Demand
BPNN	Back Propagation Neural Network
COD	Chemical Oxygen Demand
CS	Compressive Strength
HL	Hard-limit transfer function
HTS	Hyperbolic Tangent Sigmoid transfer function
Li	Linear transfer function

LS	Log-Sigmoid transfer function
MAPE	Mean Absolute Percentage Error
MSE	Mean Square Error
n_e	Effective Porosity
N_{ip}	Number of input parameters
N_n	Number of hidden layers
N_{op}	Number of output parameters
N_{td}	Number of datasets
NRB	Normalized Radial Basis transfer function
NH_4-N	Ammonia Nitrogen
PLi	Positive Linear transfer function
R	Pearson correlation coefficient
RB	Radial Basis transfer function
R_n	Schmidt hammer rebound number
SHL	Symmetric hard-limit transfer function
SM	Soft Max transfer function
SS	Suspended Solids
SSE	Sum Square Error
SSL	Symmetric Saturating Linear transfer function
TB	Triangular Basis transfer function
TN	Total Nitrogen
TP	Total Phosphorous
UCS	Unconfined Compressive Strength
V_p	Ultrasonic Pulse Velocity
WWTP(s)	Wastewater Treatment Plant(s)

Appendix A

Table A1. Transfer functions.

SN	Transfer Function/Equation/ Matlab Function	Graph
1	<p>The symmetric saturating linear transfer function (SSL)</p> $a = f(n) = \begin{cases} -1, & n \leq -1 \\ n, & -1 < n < 1 \\ -1, & n \geq 1 \end{cases}$ <p>$a = f(n) = \text{satlins}(n)$</p>	
2	<p>The log-sigmoid transfer function (LS)</p> $a = \frac{1}{1 + \exp(-n)}$ <p>$a = f(n) = \text{logsig}(n)$</p>	

References

1. Jouanneau, S.; Recoules, L.; Durand, M.J.; Boukabache, A.; Picot, V.; Primault, Y.; Lakel, A.; Sengelin, M.; Barillon, B.; Thouand, G. Methods for Assessing Biochemical Oxygen Demand (BOD): A Review. *Water Res.* **2014**, *49*, 62–82. [[CrossRef](#)] [[PubMed](#)]
2. Riedel, K.; Kunze, G.; König, A. Microbial Sensors on a Respiratory Basis for Wastewater Monitoring. In *History and Trends in Bioprocessing and Biotransformation*; Dutta, N.N., Hammar, F., Haralampidis, K., Karanth, N.G., König, A., Krishna, S.H., Kunze, G., Nagy, E., Orlich, B., Osbourn, A.E., et al., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer: Berlin/Heidelberg, Germany, 2002; Volume 75, pp. 81–118, ISBN 978-3-540-42371-3.
3. Ngoc, L.T.B.; Tu, T.A.; Hien, L.T.T.; Linh, D.N.; Tri, N.; Duy, N.P.H.; Cuong, H.T.; Phuong, P.T.T. Simple Approach for the Rapid Estimation of BOD5 in Food Processing Wastewater. *Environ. Sci. Pollut. Res.* **2020**, *27*, 20554–20564. [[CrossRef](#)]
4. Alexakis, D.; Kagalou, I.; Tsakiris, G. Assessment of Pressures and Impacts on Surface Water Bodies of the Mediterranean. Case Study: Pamvotis Lake, Greece. *Environ. Earth Sci.* **2013**, *70*, 687–698. [[CrossRef](#)]
5. Alexakis, D.; Tsihrintzis, V.A.; Tsakiris, G.; Gikas, G.D. Suitability of Water Quality Indices for Application in Lakes in the Mediterranean. *Water Resour. Manag.* **2016**, *30*, 1621–1633. [[CrossRef](#)]
6. Gamvroula, D.E.; Alexakis, D.E. Evaluating the Performance of Water Quality Indices: Application in Surface Water of Lake Union, Washington State-USA. *Hydrology* **2022**, *9*, 116. [[CrossRef](#)]
7. Cheng, S.; Lin, Z.; Sun, Y.; Li, H.; Ren, X. Fast and Simultaneous Detection of Dissolved BOD and Nitrite in Wastewater by Using Bioelectrode with Bidirectional Extracellular Electron Transport. *Water Res.* **2022**, *213*, 118186. [[CrossRef](#)] [[PubMed](#)]
8. Zeinolabedini, M.; Najafzadeh, M. Comparative Study of Different Wavelet-Based Neural Network Models to Predict Sewage Sludge Quantity in Wastewater Treatment Plant. *Environ. Monit. Assess.* **2019**, *191*, 163. [[CrossRef](#)]
9. Najafzadeh, M.; Zeinolabedini, M. Derivation of Optimal Equations for Prediction of Sewage Sludge Quantity Using Wavelet Conjunction Models: An Environmental Assessment. *Environ. Sci. Pollut. Res.* **2018**, *25*, 22931–22943. [[CrossRef](#)]
10. Najafzadeh, M.; Zeinolabedini, M. Prognostication of Waste Water Treatment Plant Performance Using Efficient Soft Computing Models: An Environmental Evaluation. *Measurement* **2019**, *138*, 690–701. [[CrossRef](#)]
11. Gunjal, S.; Khobragade, M.; Chaware, C. Prediction of BOD from Wastewater Characteristics and Their Interactions Using Regression Neural Network: A Case Study of Naidu Wastewater Treatment Plant, Pune, India. In *Recent Trends in Construction Technology and Management*; Ranadive, M.S., Das, B.B., Mehta, Y.A., Gupta, R., Eds.; Lecture Notes in Civil Engineering; Springer Nature: Singapore, 2023; Volume 260, pp. 571–577, ISBN 978-981-19214-4-5.
12. Hu, Y.; Du, W.; Yang, C.; Wang, Y.; Huang, T.; Xu, X.; Li, W. Source Identification and Prediction of Nitrogen and Phosphorus Pollution of Lake Taihu by an Ensemble Machine Learning Technique. *Front. Environ. Sci. Eng.* **2023**, *17*, 55. [[CrossRef](#)]
13. Ismail, S.; Ahmed, M.F. Hydrogeochemical Characterization of the Groundwater of Lahore Region Using Supervised Machine Learning Technique. *Environ. Monit. Assess.* **2023**, *195*, 5. [[CrossRef](#)] [[PubMed](#)]
14. Zhou, P.; Li, Z.; El-Dakhkhni, W.; Smyth, S.A. Prediction of Bisphenol A Contamination in Canadian Municipal Wastewater. *J. Water Process Eng.* **2022**, *50*, 103304. [[CrossRef](#)]
15. Zhong, H.; Yuan, Y.; Luo, L.; Ye, J.; Chen, M.; Zhong, C. Water Quality Prediction of MBR Based on Machine Learning: A Novel Dataset Contribution Analysis Method. *J. Water Process Eng.* **2022**, *50*, 103296. [[CrossRef](#)]
16. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [[CrossRef](#)] [[PubMed](#)]
17. Huang, R.; Ma, C.; Ma, J.; Huangfu, X.; He, Q. Machine Learning in Natural and Engineered Water Systems. *Water Res.* **2021**, *205*, 117666. [[CrossRef](#)]
18. ISO 5815-1:2003; Water Quality-Determination of Biochemical Oxygen Demand after N Days (BOD_n)-Part 1: Dilution and Seeding Method with Allylthiourea Addition. International Organization for Standardization: Geneva, Switzerland, 2003.
19. McDonagh, C.; Kolle, C.; McEvoy, A.K.; Dowling, D.L.; Cafolla, A.A.; Cullen, S.J.; MacCraith, B.D. Phase Fluorometric Dissolved Oxygen Sensor. *Sens. Actuators B Chem.* **2001**, *74*, 124–130. [[CrossRef](#)]
20. McEvoy, A.K.; McDonagh, C.M.; MacCraith, B.D. Dissolved Oxygen Sensor Based on Fluorescence Quenching of Oxygen-Sensitive Ruthenium Complexes Immobilized in Sol-Gel-Derived Porous Silica Coatings. *Analyst* **1996**, *121*, 785–788. [[CrossRef](#)]
21. Xiong, X.; Xiao, D.; Choi, M.F. Dissolved Oxygen Sensor Based on Fluorescence Quenching of Oxygen-Sensitive Ruthenium Complex Immobilized on Silica-Ni-P Composite Coating. *Sens. Actuators B Chem.* **2006**, *117*, 172–176. [[CrossRef](#)]
22. Xu, W.; McDonough, R.C.; Langsdorf, B.; Demas, J.N.; DeGraff, B.A. Oxygen Sensors Based on Luminescence Quenching: Interactions of Metal Complexes with the Polymer Supports. *Anal. Chem.* **1994**, *66*, 4133–4141. [[CrossRef](#)]
23. Sakaguchi, T.; Kitagawa, K.; Ando, T.; Murakami, Y.; Morita, Y.; Yamamura, A.; Yokoyama, K.; Tamiya, E. A Rapid BOD Sensing System Using Luminescent Recombinants of Escherichia Coli. *Biosens. Bioelectron.* **2003**, *19*, 115–121. [[CrossRef](#)]
24. Sakaguchi, T.; Morioka, Y.; Yamasaki, M.; Iwanaga, J.; Beppu, K.; Maeda, H.; Morita, Y.; Tamiya, E. Rapid and Onsite BOD Sensing System Using Luminous Bacterial Cells-Immobilized Chip. *Biosens. Bioelectron.* **2007**, *22*, 1345–1350. [[CrossRef](#)] [[PubMed](#)]
25. Kim, B.H.; Chang, I.S.; Cheol Gil, G.; Park, H.S.; Kim, H.J. Novel BOD (Biological Oxygen Demand) Sensor Using Mediator-Less Microbial Fuel Cell. *Biotechnol. Lett.* **2003**, *25*, 541–545. [[CrossRef](#)] [[PubMed](#)]
26. Karube, I.; Matsunaga, T.; Mitsuda, S.; Suzuki, S. Microbial Electrode BOD Sensors. *Biotechnol. Bioeng.* **1977**, *19*, 1535–1547. [[CrossRef](#)] [[PubMed](#)]
27. Liu, C.; Zhao, H.; Zhong, L.; Liu, C.; Jia, J.; Xu, X.; Liu, L.; Dong, S. A Biofilm Reactor-Based Approach for Rapid on-Line Determination of Biodegradable Organic Pollutants. *Biosens. Bioelectron.* **2012**, *34*, 77–82. [[CrossRef](#)]

28. Zare Abyaneh, H. Evaluation of Multivariate Linear Regression and Artificial Neural Networks in Prediction of Water Quality Parameters. *J. Environ. Health Sci. Eng.* **2014**, *12*, 40. [[CrossRef](#)]
29. Singh, K.P.; Basant, A.; Malik, A.; Jain, G. Artificial Neural Network Modeling of the River Water Quality—A Case Study. *Ecol. Model.* **2009**, *220*, 888–895. [[CrossRef](#)]
30. Ahmed, A.A.M.; Shah, S.M.A. Application of Adaptive Neuro-Fuzzy Inference System (ANFIS) to Estimate the Biochemical Oxygen Demand (BOD) of Surma River. *J. King Saud Univ. Eng. Sci.* **2017**, *29*, 237–243. [[CrossRef](#)]
31. Basant, N.; Gupta, S.; Malik, A.; Singh, K.P. Linear and Nonlinear Modeling for Simultaneous Prediction of Dissolved Oxygen and Biochemical Oxygen Demand of the Surface Water—A Case Study. *Chemom. Intell. Lab. Syst.* **2010**, *104*, 172–180. [[CrossRef](#)]
32. Asami, H.; Golabi, M.; Albaji, M. Simulation of the Biochemical and Chemical Oxygen Demand and Total Suspended Solids in Wastewater Treatment Plants: Data-Mining Approach. *J. Clean. Prod.* **2021**, *296*, 126533. [[CrossRef](#)]
33. McCulloch, W.S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
34. Gavriilaki, E.; Asteris, P.G.; Touloumenidou, T.; Koravou, E.-E.; Koutra, M.; Papayanni, P.G.; Karali, V.; Papalexandri, A.; Varelas, C.; Chatzopoulou, F.; et al. Genetic Justification of Severe COVID-19 Using a Rigorous Algorithm. *Clin. Immunol.* **2021**, *226*, 108726. [[CrossRef](#)] [[PubMed](#)]
35. Asteris, P.G.; Gavriilaki, E.; Touloumenidou, T.; Koravou, E.; Koutra, M.; Papayanni, P.G.; Pouleres, A.; Karali, V.; Lemonis, M.E.; Mamou, A.; et al. Genetic Prediction of ICU Hospitalization and Mortality in COVID-19 Patients Using Artificial Neural Networks. *J. Cell. Mol. Med.* **2022**, *26*, 1445–1455. [[CrossRef](#)] [[PubMed](#)]
36. Upadhyay, A.K.; Shukla, S. Correlation Study to Identify the Factors Affecting COVID-19 Case Fatality Rates in India. *Diabetes Metab. Syndr. Clin. Res. Rev.* **2021**, *15*, 993–999. [[CrossRef](#)] [[PubMed](#)]
37. Niazkari, H.R.; Niazkari, M. Application of Artificial Neural Networks to Predict the COVID-19 Outbreak. *Glob. Health Res. Policy* **2020**, *5*, 50. [[CrossRef](#)] [[PubMed](#)]
38. Mahanty, C.; Kumar, R.; Asteris, P.G.; Gandomi, A.H. COVID-19 Patient Detection Based on Fusion of Transfer Learning and Fuzzy Ensemble Models Using CXR Images. *Appl. Sci.* **2021**, *11*, 11423. [[CrossRef](#)]
39. Rahimi, I.; Gandomi, A.H.; Asteris, P.G.; Chen, F. Analysis and Prediction of COVID-19 Using SIR, SEIQR, and Machine Learning Models: Australia, Italy, and UK Cases. *Information* **2021**, *12*, 109. [[CrossRef](#)]
40. Asteris, P.G.; Douvika, M.G.; Karamani, C.A.; Skentou, A.D.; Chlichlia, K.; Cavaleri, L.; Daras, T.; Armaghani, D.J.; Zaoutis, T.E. A Novel Heuristic Algorithm for the Modeling and Risk Assessment of the COVID-19 Pandemic Phenomenon. *Comput. Model. Eng. Sci.* **2020**, *125*, 815–828. [[CrossRef](#)]
41. APHA; AWWA; WPCF. *Standard Methods for the Examination of Water and Wastewater*, 20th ed.; American Public Health Association: Washington, DC, USA, 1999.
42. Jong, Y.-H.; Lee, C.-I. Influence of Geological Conditions on the Powder Factor for Tunnel Blasting. *Int. J. Rock Mech. Min. Sci.* **2004**, *41*, 533–538. [[CrossRef](#)]
43. Bardhan, A.; Biswas, R.; Kardani, N.; Iqbal, M.; Samui, P.; Singh, M.P.; Asteris, P.G. A Novel Integrated Approach of Augmented Grey Wolf Optimizer and ANN for Estimating Axial Load Carrying-Capacity of Concrete-Filled Steel Tube Columns. *Constr. Build. Mater.* **2022**, *337*, 127454. [[CrossRef](#)]
44. Mahmood, W.; Mohammed, A.S.; Asteris, P.G.; Kurda, R.; Armaghani, D.J. Modeling Flexural and Compressive Strengths Behaviour of Cement-Grouted Sands Modified with Water Reducer Polymer. *Appl. Sci.* **2022**, *12*, 1016. [[CrossRef](#)]
45. Mahmood, W.; Mohammed, A.S.; Sihag, P.; Asteris, P.G.; Ahmed, H. Interpreting the Experimental Results of Compressive Strength of Hand-Mixed Cement-Grouted Sands Using Various Mathematical Approaches. *Arch. Civ. Mech. Eng.* **2022**, *22*, 19. [[CrossRef](#)]
46. Emad, W.; Salih, A.; Kurda, R.; Asteris, P.G.; Hassan, A. Nonlinear Models to Predict Stress versus Strain of Early Age Strength of Flowable Ordinary Portland Cement. *Eur. J. Environ. Civ. Eng.* **2022**, *26*, 8433–8457. [[CrossRef](#)]
47. Asteris, P.G.; Apostolopoulou, M.; Armaghani, D.J.; Cavaleri, L.; Chountalas, A.T.; Guney, D.; Hajihassani, M.; Hasanipanah, M.; Khandelwal, M.; Karamani, C.; et al. On the Metaheuristic Models for the Prediction of Cement-Metakaolin Mortars Compressive Strength. *Metaheuristic Comp. Appl.* **2020**, *1*, 63–69. [[CrossRef](#)]
48. Asteris, P.G.; Argyropoulos, I.; Cavaleri, L.; Rodrigues, H.; Varum, H.; Thomas, J.; Lourenço, P.B. Masonry Compressive Strength Prediction Using Artificial Neural Networks. In *Transdisciplinary Multispectral Modeling and Cooperation for the Preservation of Cultural Heritage*; Moropoulou, A., Korres, M., Georgopoulos, A., Spyarakos, C., Mouzakis, C., Eds.; Communications in Computer and Information Science; Springer International Publishing: Cham, Switzerland, 2019; Volume 962, pp. 200–224, ISBN 978-3-030-12959-0.
49. Armaghani, D.J.; Asteris, P.G.; Fatemi, S.A.; Hasanipanah, M.; Tarinejad, R.; Rashid, A.S.A.; Huynh, V.V. On the Use of Neuro-Swarm System to Forecast the Pile Settlement. *Appl. Sci.* **2020**, *10*, 1904. [[CrossRef](#)]
50. Lu, S.; Koopialipour, M.; Asteris, P.G.; Bahri, M.; Armaghani, D.J. A Novel Feature Selection Approach Based on Tree Models for Evaluating the Punching Shear Capacity of Steel Fiber-Reinforced Concrete Flat Slabs. *Materials* **2020**, *13*, 3902. [[CrossRef](#)]
51. Le, T.-T. Practical Machine Learning-Based Prediction Model for Axial Capacity of Square CFST Columns. *Mech. Adv. Mater. Struct.* **2022**, *29*, 1782–1797. [[CrossRef](#)]
52. Le, T.-T.; Asteris, P.G.; Lemonis, M.E. Prediction of Axial Load Capacity of Rectangular Concrete-Filled Steel Tube Columns Using Machine Learning Techniques. *Eng. Comput.* **2022**, *38*, 3283–3316. [[CrossRef](#)]

53. Le, T.-T.; Le, M.V. Development of User-Friendly Kernel-Based Gaussian Process Regression Model for Prediction of Load-Bearing Capacity of Square Concrete-Filled Steel Tubular Members. *Mater. Struct.* **2021**, *54*, 59. [[CrossRef](#)]
54. Armaghani, D.J.; Asteris, P.G. A Comparative Study of ANN and ANFIS Models for the Prediction of Cement-Based Mortar Materials Compressive Strength. *Neural Comput. Appl.* **2021**, *33*, 4501–4532. [[CrossRef](#)]
55. Asteris, P.G.; Mokos, V.G. Concrete Compressive Strength Using Artificial Neural Networks. *Neural Comput. Appl.* **2020**, *32*, 11807–11826. [[CrossRef](#)]
56. Asteris, P.G.; Nguyen, T.-A. Prediction of Shear Strength of Corrosion Reinforced Concrete Beams Using Artificial Neural Network. *J. Sci. Transp. Tech.* **2022**, *2*, 1–12.
57. Li, N.; Asteris, P.G.; Tran, T.-T.; Pradhan, B.; Nguyen, H. Modelling the Deflection of Reinforced Concrete Beams Using the Improved Artificial Neural Network by Imperialist Competitive Optimization. *Steel Compos. Struct.* **2022**, *42*, 733–745. [[CrossRef](#)]
58. Lemonis, M.E.; Daramara, A.G.; Georgiadou, A.G.; Siorikis, V.G.; Tsavdaridis, K.D.; Asteris, P.G. Ultimate Axial Load of Rectangular Concrete-Filled Steel Tubes Using Multiple ANN Activation Functions. *Steel Compos. Struct.* **2022**, *42*, 459–475. [[CrossRef](#)]
59. Asteris, P.G.; Rizal, F.I.M.; Koopialipoor, M.; Roussis, P.C.; Ferentinou, M.; Armaghani, D.J.; Gordan, B. Slope Stability Classification under Seismic Conditions Using Several Tree-Based Intelligent Techniques. *Appl. Sci.* **2022**, *12*, 1753. [[CrossRef](#)]
60. Asteris, P.G.; Lourenço, P.B.; Hajihassani, M.; Adami, C.-E.N.; Lemonis, M.E.; Skentou, A.D.; Marques, R.; Nguyen, H.; Rodrigues, H.; Varum, H. Soft Computing-Based Models for the Prediction of Masonry Compressive Strength. *Eng. Struct.* **2021**, *248*, 113276. [[CrossRef](#)]
61. Asteris, P.G.; Skentou, A.D.; Bardhan, A.; Samui, P.; Lourenço, P.B. Soft Computing Techniques for the Prediction of Concrete Compressive Strength Using Non-Destructive Tests. *Constr. Build. Mater.* **2021**, *303*, 124450. [[CrossRef](#)]
62. Asteris, P.G.; Koopialipoor, M.; Armaghani, D.J.; Kotsonis, E.A.; Lourenço, P.B. Prediction of Cement-Based Mortars Compressive Strength Using Machine Learning Techniques. *Neural Comput. Appl.* **2021**, *33*, 13089–13121. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.