Empirical modeling of turbidity removal in a dissolved air flotation system: application of artificial neural networks
Ana Cláudia Oliveira e Souza, Newton Libanio Ferreira and Flávio Vasconcelos da Silva

ABSTRACT
Dissolved air flotation (DAF) is a physical separation process that uses air microbubbles to remove suspended material dispersed in a liquid phase. Even though DAF is considered a well-established unit operation, modeling it is difficult due to the complexity of the phenomena involved, resulting in conceptual models with no practical application. Thereby, the objective of this work was to evaluate empirical modeling efficiency in predicting the turbidity removal dynamic using artificial neural networks applied to a DAF prototype. For the study of the neural network input variables, a two-level, full-factorial design was utilized to verify the statistical significance of the saturation pressure and the saturated water flow rate in relation to the turbidity removal. Using a time-delay recurrent neural network architecture, two empirical models were proposed to simulate the dynamic behavior of the turbidity removal promoted by the DAF prototype. The real-time model provided good predictions with $R^2 = 0.9717$ and $MSE = 1.0482$, and the simulation model was also able to predict the process behavior presenting performance criteria equal to $R^2 = 0.9475$ and $MSE = 1.8640$.

Key words | artificial neural networks, dissolved air flotation, process modeling, water treatment

HIGHLIGHTS
- Development of empirical models capable of efficiently predicting the turbidity removal dynamic behavior of a DAF prototype using artificial neural networks.
- The simulation model can predict the $TBD_R$ without the need to perform an experimental run with $R^2 = 0.9475$.
- The real-time model can be used in applications that demand an online model, and it provided good predictions with $R^2 = 0.9717$. 

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (http://creativecommons.org/licenses/by/4.0/).
doi: 10.2166/ws.2021.152
INTRODUCTION

Dissolved air flotation (DAF) is a unit operation capable of removing solid or liquid contaminant particles present in a liquid phase. At the beginning of its use (the early 1900s), DAF was widely used in the ore processing industry as a method of separating mineral ores (Edzwald 1995). Sixty years later, water and wastewater treatment plants began applying DAF to remove color, natural organic matter, and suspended particles using small air bubbles produced in a saturation vessel.

When the DAF process is applied to treating water for public supply purposes, the technique is used in the clarification stage of the raw water treatment to remove turbidity. In this case, the flotation process usually takes place in rectangular tanks divided into a contact zone and a separation zone. In the contact zone, air microbubbles collide with coagulated impurity flocs to form particle–bubble agglomerates. These agglomerates have a lower density than water, so they rise to the separator surface where they form a floating layer that can be collected and removed in the separation zone (Edzwald 2010). The microbubbles are formed from the depressurization of the water flow that is saturated with compressed air at pressures ranging from four to six bar (Edzwald & Haarhoff 2011).

Therefore, the raw water feed flow rate, the saturation vessel pressure, and the air bubble flow rate injected into the flotation tank are the physical parameters related to the process efficiency. These parameters must be considered not only in the project phase but also during DAF unit operation. The coagulant and flocculant doses, the pH values, and the ambient temperature are another group of important parameters to guarantee the success of water treatment. However, these parameters are mainly related to the preliminary treatment stages of raw water. This paper focuses on modeling and simulating the relation between turbidity removal and the physical parameters directly related to the flotation stage efficiency.

Over the years, process simulation has been used for describing the behavior of industrial processes using adequate mathematical models capable of representing the real process operation. Once implemented, the simulation is used as an auxiliary tool for making decisions about operational changes in the process. It allows the prediction of the system reaction in different case scenarios without direct disturbances being applied to the real process. Many papers have been published about modeling and simulation of wastewater treatment plants (Alex et al. 1999; Miyata et al. 2004; Man et al. 2017; Seco et al. 2020); however, there is a lack of work concerning the simulation of water treatment plants using dissolved air flotation in the clarification stage.

The absence of papers in this area can be explained by the fact that simulating the processes involved in DAF using first-principles models derived from mass, energy, and momentum balances is nevertheless an arduous task (Haarhoff 2008). This case scenario is due to the complexity of the phenomena involved, especially in the coagulation/flocculation steps and in the collision and attachment processes among flocs and microbubbles. Besides that, although DAF is considered a well-established unit operation, in some countries (like Brazil) sedimentation is the
most-used clarification process in water treatment stations due to its simplicity and low power consumption. However, despite being the conventionally applied clarification process, sedimentation presents several disadvantages when compared with DAF (e.g., the sludge produced is less concentrated than the float layer formed in DAF tanks, larger areas are required for the installation of the sedimentation tanks, and the residence time of the raw water being treated is higher; Zabel 1985).

Over the years, several studies have presented models capable of mathematically replicating the phenomena observed during DAF operation (Edzwald 1995; Fukushi et al. 1995; Shawwa & Smith 2000; Edzwald & Haarhoff 2011; Ksenofontov & Ivanov 2015). Although they have contributed to a better understanding of the theory involved, the existing models are complex and difficult to implement in simulation software or to solve mathematically. There is lots of research that has been done regarding the simulation, control and optimization of the flotation column process (Bergh & Yianatos 1995; Carvalho & Durão 2002; Hacifazlioglu & Sutcu 2007; Bouchard et al. 2014), which is normally used for ore processing. However, the theory and principles applied are restricted to the phenomena observed in the flotation columns and they cannot be applied to dissolved air flotation in rectangular tanks.

On the other hand, relevant and interesting studies have been recently carried out considering different DAF process aspects for water and wastewater treatment. Rodrigues et al. (2019) combined computational fluid dynamics and population balance equations to study the microbubble distribution in a DAF tank contact and separation zones. They evaluated the effect of bubble breakage and coalescence events on the flow dynamics of the DAF tank. Fanaie & Khiaadani (2020) investigated the influence of salinity on size distribution and hydrodynamics of microbubbles in a DAF experimental setup. Different degrees of salinity were tested using a mixture of sodium chloride and tap water. The authors noticed that the flow itself had greater impact in the flow patterns than the salinity. Despite the relevance of the studied subjects, none of the mentioned works address means of establishing an adequate mathematical relationship between turbidity removal and the physical variables of DAF (e.g. saturation pressure and saturated water flow rate).

The lack of valid models that could have practical application is a barrier to a more detailed study and simulation of the process dynamics. Thus, empirical models emerge as a feasible alternative to fulfill the demand for a model that represents the phenomena observed during DAF. In this regard, the prediction of important parameters related to water and wastewater treatment has been proven effective when working with artificial neural network (ANN) models (Maier et al. 2004; Labidi et al. 2007; Singh et al. 2009; Salari et al. 2018).

In the last five years, artificial intelligence (AI) has returned to be a reason for enthusiasm in the academic and industrial environments. According to Venkatasubramanian (2018), after going through two periods of great stagnation, also known as ‘AI winters’, the application of AI techniques, such as the use of artificial neural networks, has been very promising in solving several complex problems. ANNs are structures that mirror the functioning of the human brain and its basic processing units are known as neurons. Several ANN architectures have already been proposed and studied by AI researchers in recent decades (Leijnen & van Veen 2020). Among these architectures, the recurrent neural networks are especially useful when there is observed to be a temporal dependence between the process data being analyzed (e.g. industrial processes, like the dissolved air flotation unit operation).

Given that water crisis is an actual scenario in Brazil and other countries, it is extremely important to understand how to increase the efficiency of the available water treatment technologies, such as developing a better understanding of the dynamic behavior of the phenomena observed in a DAF unit. Therefore, it is equally useful for industrial process operators to have at their disposal a support tool that enables quick and safe testing of different operating conditions. This can be achieved by a reliable DAF simulation model. Thus, the aim of this work is to develop an empirical ANN model to simulate the dynamic behavior of a DAF prototype. Knowing that turbidity is one of the main water potability assessment parameters, and therefore its monitoring is an essential task in water treatment stations, turbidity removal was the chosen output variable and a time-delay recurrent neural network architecture was employed.
MATERIALS AND METHODS

DAF prototype

To collect the data needed to construct an ANN model, tests were carried out in a DAF prototype built and automated by Fonseca et al. (2017), located in the Control and Automation Processes Laboratory at the University of Campinas, Brazil. Figure 1 shows the DAF prototype used.

The unit consists of a coagulation and flocculation tank (1), subdivided into three sections, where the coagulant and flocculant agents are supplied, in addition to the raw synthetic water to be treated. After chemical pretreatment, the raw water is diverted to the flotation tank (3), consisting of contact and separation zones. In the contact zone, the raw coagulated water encounters a constant flow of air microbubbles. Subsequently, the microbubbles are generated in an unpacked saturator vessel (8), where clean water is saturated with compressed air. Then, the saturated water stream passes through a needle valve (9), which promotes depressurization and the consequent nucleation of the bubbles in this stream.

The clarified water flows out from the bottom of the separation zone, where a fraction is directed to a sand filter (4) that removes the non-floating particles. The other fraction of the clarified water is directed to an online turbidimeter (6) before it is purged. The filtered water is stored in a buffer tank (5) and part of it is used to feed the saturator vessel. Figure 2 shows the DAF pilot plant piping and instrumentation diagram (P&ID).

The prototype is automated and equipped with control loops for regulating the raw water flow rate, the needle valve opening, and the saturator vessel pressure (Fonseca et al. 2017). All sensors and actuators are monitored by a programmable logic controller (PLC), connected to a supervisory system, implemented in MATLAB (MathWorks, 2021).
Inc., Natwick, MA). Through SCADA (Supervisory Control and Data Acquisition), data were acquired at a sampling rate of one second, allowing the composition of an extensive database at the end of each test.

**Experimental runs and factorial design**

To simulate the characteristics of superficial water collected in rivers and lakes, a raw synthetic water was prepared using red clay, a typical soil found in the state of São Paulo, Brazil. The synthetic water was obtained by mixing tap water and the red clay; the amount of red clay required was proportional to the initial desired turbidity value. In this study, a constant value of 20 NTU was used for initial turbidity. Sodium aluminate 2% v/v (NaAlO₂) and Tanfloc SG® 5% v/v (natural cationic polymer primary composed of tannins) were the coagulant and flocculant agents used, respectively. To determine the coagulant and flocculant chemical dosages, jar tests were performed according to the methodology presented in Crittenden et al. (2012). Table 1 presents all the operational conditions constantly maintained during the experiments. These parameters were kept constant to discard their influence on the turbidity removal and to allow the investigation of the effects of saturation pressure and saturated water flow only.

The experimental procedure was divided into two stages. First, a two-level, full-factorial design was applied to investigate the influence of the physical parameters related to the DAF unit's operation on the amount of turbidity removed. Table 2 shows the factors and levels used in the design. The levels were chosen according to the prototype equipment's physical limitations, such as the maximum pressure supported by the saturation vessel (7.4 bar) and the maximum saturated water flow rate measured by the

---

**Table 1 | Operational conditions of the DAF experiments**

| Parameter                  | Value     |
|----------------------------|-----------|
| Raw water feed flow rate   | 2 L min⁻¹ |
| Coagulant addition dose    | 75 mg L⁻¹ |
| Flocculant addition dose   | 275 mg L⁻¹|
| Raw water turbidity        | 20 NTU    |

---

**Figure 2 | DAF pilot piping and instrumentation diagram.**
flow sensor used (0.50 L/min). Replicates were realized to allow pure error calculation within a 95% confidence interval. Statistica 10 (StatSoft Inc., USA) statistical software was used to treat the data and to determine the significant effects.

The saturated water flow rate can also be defined in terms of the recycle ratio. The recycle ratio, calculated as the ratio between the saturated water and the raw water feed flow rates, is an estimated measure of the microbubble quantity injected into the flotation tank contact zone and can be calculated using Equation (1). Turbidity removal (Equation (2)) was the response variable chosen for analysis since its dynamic behavior represents the flotation unit efficiency and is the output variable used at the empirical modeling stage.

\[ RR(\%) = \frac{Q_{SAT}}{Q_{F}} \times 100 \]  

(1)

where \( Q_{SAT} \) = saturated water flow rate (L min\(^{-1}\)); \( Q_{F} \) = raw water feed flow rate (L min\(^{-1}\)); and \( RR \) = recycle ratio (%).

\[ TBD_{R}(\%) = \frac{TBD_{F} - TBD_{SS}}{TBD_{F}} \times 100 \]  

(2)

where \( TBD_{F} \) = raw water turbidity (NTU); \( TBD_{SS} \) = final steady state turbidity (NTU); and \( TBD_{R} \) = turbidity removal (%).

The second part of the experimental procedure consisted of realizing tests where the physical variables related to the DAF (saturation pressure and recycle ratio) operation were perturbated. This strategy was adopted to ensure the observation of the turbidity removal dynamic behavior, thus allowing us to obtain a time series database. Four experiments were realized and the operational conditions applied are presented in Table 3. Step perturbations were used.

The authors are aware that the disturbance ranges applied are limited, but the minimum and maximum values for the recycle ratio and the saturation pressure are in agreement with the literature standard design and operational ranges (Edzwald & Haarhoff 2014). Besides that, it is not possible to achieve a good solid–liquid separation state if the DAF process operational conditions are drastically changed. To execute all the following analysis the basic premise adopted was that all the experimental runs should present an adequate flotation condition.

### Development of ANN models

Once all the experiments were performed, the four databases generated were unified to form a single database composed of the time series that represent the dynamic behavior of the chosen input and output signals.

A time-delay recurrent neural network (TDRNN) architecture was used to model the turbidity removal behavior over time. In the TDRNN architecture, the output prediction is not only based on the current value of the interest variable but also considers the past-calculated output. Therefore, the model’s calculated output values are fed back into the feed layer and used like input signals as well as the other exogenous input variables (Haykin 2009). This architecture was chosen due to its capability of modeling dynamic and nonlinear processes. The TDRNN architecture along with the variables studied in this work is represented in Figure 3.

In this work, two empirical models were proposed: a real-time model and a simulation model. The real-time
model receives the values of the input variables directly measured by the sensors present in the DAF unit during the realization of an experimental test. Therefore, this model is capable of providing real-time prediction of the turbidity removal behavior and it can be incorporated into the SCADA system. The second model, the simulation model, works totally independently of the realization of an experiment. The exogenous input data are given by the user and the turbidity removal input is the value predicted by the ANN model.

For both models, the exogenous inputs were the saturated water flow rate, the saturation pressure, and the raw water turbidity values collected during run time, and the predicted output was the turbidity removal time series. To determine the network hyperparameters, different tests were realized that changed the number of time delays applied, the number of hidden network layers, and the number of neurons in each layer.

Not only a training stage, but also a validation stage was realized using an algorithm built in MATLAB (MathWorks, Inc., Natwick, MA). Training and validation test sets corresponded to 80% and 20% of the total points collected, respectively. The split into training and validation sets respected the temporal dependency of the data, therefore shuffling was not applied in order to maintain the chronological connection between the data samples. A sequential split was adopted (i.e. the first 80% of data points were used to train the ANN and validation used the last 20%). The Levenberg–Marquardt backpropagation method was used to perform the supervised training step, the hidden layer(s) activation function was a hyperbolic tangent sigmoid transfer function, and the output layer activation function was a linear transfer function.

To avoid overfitting during the learning stage, the early stopping technique was applied using the validation set. Early stopping is a method that interrupts the training process when the validation error starts to increase. At each iteration performed by the training algorithm, the validation set is used to estimate the generalization error. So, it is a way of exiting training at a suitable point, preventing the model from memorizing data instead of learning it. The use of early stopping is quite useful, especially when the available dataset is not large (Sakizadeh et al. 2015). The training step was performed 20 times for each model tested and the best neural model was selected. The parameters used to evaluate ANN performance were the mean squared error (MSE) and the regression value (R).

Based on these criteria, an ANN with elevated

![Figure 3](image-url)
generalization capability must return low MSE values and high $R$ values.

**RESULTS AND DISCUSSION**

Two-level, full-factorial design

To analyze the effects of the input signals chosen to feed the empirical neural model, a two-level, full-factorial design was first carried out. The Pareto graph is presented in Figure 4, which indicates that both saturation pressure and saturated water flow rate are statistically significant within a 95% confidence interval. Since replicates were realized, a pure error of 0.97 was calculated. Table 4 shows the effect values and the significance of the factors analyzed.

Between the two factors, the saturated water flow rate had the highest influence on turbidity removal. The positive effect indicates that the $TBD_R$ rises on average $2.70\% \pm 0.69\%$ when the saturated water flow increases from 0.38 to 0.44 L/min. The saturation pressure also presented a positive effect, thus $TBD_R$ increases about $2.12\% \pm 0.69\%$ when the pressure is raised from six to seven bar.

These results are in qualitative agreement with theory described in the specialized literature (Edzwald 1995; Edzwald 2010; Edzwald & Haarhoff 2011). Therefore, it has been verified that these physical variables (saturation pressure and saturated water flow rate) affect the turbidity removal promoted by the DAF unit and are adequate to model its dynamic behavior.

According to Table 4, the interaction effect between the factors evaluated is negligible since it shows a $p$-value higher than 0.05. Therefore, when the pressure is increased or decreased this action does not influence the saturated water flow rate effect on the turbidity removal. Changes in the saturated water flow rate also do not influence the pressure on the turbidity removal.

Figure 5 presents the response surface for the turbidity removal. It indicates that higher pressures and saturated water flow rates contribute to greater turbidity removal. This was the expected result, since a good condition of formation and injection of microbubbles in the flotation tank’s contact zone is related to these two variables. In addition, Figure 5 confirms that the levels adopted in the factorial design were adequate. More aggressive operating conditions were not used due to the prototype’s physical limitations mentioned earlier.

It is important to emphasize that the two-level, full-factorial design was not applied to obtain a statistical model of the process, but only to validate the influence of the factors on the chosen response variable. Even though the influence of saturator pressure and saturator flow rate in DAF is already extensively described in the literature and very well known in industry, it was essential to confirm
the validity of it in the laboratory-scale system used. ANNs were used to obtain an empirical model representative of the DAF system as discussed in the next section.

ANN models

Once the statistical significance of the variables chosen as the ANNs’ input signals was verified, tests with disturbances in these variables were performed (Table 3). A database with 20,230 points was obtained. The dataset used is available online and can be found at Souza et al. (2020). To determine the best topology for each neural model (the real-time model and the simulation model), several tests were realized and the most relevant are shown in Table 5. The progressive increase of the complexity of the topology tested was the approach adopted to avoid overfitting in the training stage and to prevent the generation of unnecessary complex models, which would require excessive computational effort. This is a common procedure which has been used by other researchers (Hamed et al. 2004; Maier et al. 2004; Labidi et al. 2007; Singh et al. 2009; Nakhaei et al. 2012).

The best network for the real-time model presented three time-delays in the input layer and three neurons in the hidden layer. Figure 6 shows the regression plots for the training and validation stages. The high regression coefficient and the low MSE value calculated (0.9717 and 1.0482, respectively) indicate that the ANN was able to map the relationships between inputs and outputs provided during training. In addition, the regression plots show agreement exists between the targets and the predicted values. The comparison between the real dynamic behavior of TBD\textsubscript{R} and the one predicted by the ANN is shown in Figure 7(a). It is noticed that the real-time model’s prediction was very close to the expected TBD\textsubscript{R} values. This means that the real-time model correctly assimilated the influence that the input variables (i.e., the DAF unit’s physical parameters), the saturation pressure, the saturated water flow rate, and the raw water turbidity have on the calculated output variable (i.e., the turbidity removal at the flotation tank’s exit).

Figures 7(b) and 7(c) present the variations applied to the two physical independent variables disturbed during the experimental runs. It is perceived that the real-time model was capable of learning the dynamic changes of its input variables and correctly expressing the effect of these changes in the turbidity removal behavior predicted. Another important highlight is that the ANN model

| Table 4 | Effects and p-values of the factors evaluated |
| Factors | Effect | Effect standard error | p-value |
| Pressure (bar), P | 2.12 | 0.69 | 0.038 |
| Saturated water flow (L/min), Q\textsubscript{SAT} | 2.70 | 0.69 | 0.018 |
| P × Q\textsubscript{SAT} | −0.45 | 0.69 | 0.554 |

| Table 5 | Topologies tested for the real-time and the simulation models. |
| Input time delays | Hidden layer neurons | R\textsubscript{TRAINING} | MSE\textsubscript{TRAINING} | R\textsubscript{VALIDATION} | MSE\textsubscript{VALIDATION} |
| Real-time model | 1 | 1 | 0.7996 | 3.0316 | 0.6983 | 2.0134 |
| 3 | 3 | 0.9717 | 1.0482 | 0.7584 | 1.0049 |
| 3 | 5 | 0.9399 | 2.0319 | 0.7259 | 1.9157 |
| Simulation model | 1 | 5 | 0.9110 | 5.0012 | 0.6993 | 2.5127 |
| 3 | 5 × 3 | 0.9475 | 1.8640 | 0.7565 | 1.5346 |
| 3 | 15 | 0.9493 | 1.1125 | 0.7254 | 1.6574 |
| 5 | 20 | 0.9585 | 1.0487 | 0.7114 | 2.1473 |

Figure 5 | Response surface for turbidity removal as a function of pressure and saturated water flow rate.
accomplished good results predicting not only the new TBD<sub>R</sub>’s stationary states achieved by the DAF system after each operational perturbation, but also the transition periods (i.e. periods when the TBD<sub>R</sub> presents a transient behavior because it is still under the influence of the alterations in the system’s input conditions).

The prediction errors presented low amplitude, remaining in the range of ~2% and 2% during the entire simulation. Given the current configuration of the DAF prototype, the magnitude of the calculated errors is acceptable, since this variation is usually observed in all tests performed. The location of the in-line turbidimeter measurement outlet at...
the flotation tank allows unfloated flocs to eventually enter in this sensor’s measuring chamber, which causes the appearance of small peaks in the measured signal and makes it quite variable over time.

Although the presented model was trained with only three temporal delays in the input signals, the other tests (Table 5) showed that increasing the number of delays did not affect the ANN’s performance or increase the required computational effort. Therefore, it is possible to obtain models that provide greater prediction horizons, allowing follow-up of the DAF unit behavior in more detail. The real-time model can be used in applications that demand an online model (i.e. an online simulator which runs in real-time and tracks the process behavior at each time-stamp), like studies involving the control and optimization of the DAF process, aiming to increase the efficiency of this unit’s operation.

According to Table 5, the best ANN topology for the simulation model consisted of two hidden layers, with five and three neurons, respectively, and also three temporal delays in the input layer. Figure 8 presents the regression plots for the training and validation stages of the neural model when the recursive feeding of the predicted TBD$_R$ was applied.

The regressions between the predicted outputs and the provided targets presented high regression coefficients close to 1. It is perceived that greater dispersion occurred between the adjusted data (when compared with the results obtained with the real-time model), but a linear correlation is observed. It is important to point out that both regression validation plots for the real-time and the simulation models showed higher scattering when compared with the training ones. Despite this, they still confirm the positive relationship between the predicted TBD$_R$ values and the real targets.

Figure 9 shows the dynamic behavior of the TBD$_R$ predicted by the simulation model and the real TBD$_R$ behavior obtained in the experimental tests. A regression coefficient of 0.9475 and a mean squared error equal to 1.8640 were achieved. The simulation model was able to predict the macroscopic temporal evolution of TBD$_R$ in the tests in which disturbances were applied to saturation pressure and saturated water flow rate. However, the neural model did not capture the small variations inherent to the turbidity measurements provided by the in-line turbidimeter and which occur in short periods of time. This explains why the MSE values calculated in the recursive simulation increased, since the obtained empirical model was wrong in several punctual predictions.

On the other hand, the behavior presented by the simulation model indicates no overfitting occurred during the training step, as the neural network did not excessively memorize the information provided, but it was able to generalize and learn the global behavior of the temporal series that constituted the database used. This is an important outcome, since it is very common for recurrent neural network architectures just to repeat the pattern of the dependent variable (i.e. the target) displaced in time by the delay used, indicating a poor learning achievement. Figure 9 shows clearly that the simulation model is capable of generalization. The
prediction errors presented amplitudes around 5%, which are higher than those observed for the real-time model, but they did not significantly affect the quality of the calculated outputs.

As presented in Table 5, increasing the model complexity by adding neurons in the hidden layer did not improve the model’s response. Although a higher regression coefficient is obtained for the training stage, the $MSE_{VALIDATION}$ increase and the $R_{VALIDATION}$ decrease indicate the presence of overfitting, when the neural network stops learning from data and starts to decorate them.

Although the $R_{VALIDATION}$ was around 0.7, the validation regression plot (Figure 8(b)) shows that the predicted and the real TBD$_R$ values are aligned and well described by the regression line. It is important to emphasize that the validation dataset is used to allow the application of the early stopping technique during the learning step. Therefore, validation is a support stage used to enhance the training performance and to avoid model overfitting. Besides that, considering the process modeled, the deviations observed in the validation regression plot are acceptable since they do not represent errors bigger than 5% in the predicted output variable (i.e., the turbidity at the flotation tank’s exit). The same observations apply to the real-time model.

Therefore, the simulation model provides a prediction of the turbidity removal promoted by the DAF unit, once given the saturation pressure, the saturated water flow rate, and the raw water turbidity, without the need to perform an experimental run. This becomes quite useful in exploratory studies involving DAF and when a need exists to save the reagents used in the chemical pretreatment. In addition, the simulation model is extremely useful in investigating the process’s operational conditions on an industrial scale, since a plant does not need to be disturbed or placed in unsafe conditions to obtain the desired behaviors.

CONCLUSIONS

The novel application of ANNs to model the dynamic behavior of a dissolved air flotation prototype treating water for public supply purposes proved to be an effective and adequate methodology. Using a two-level full-factorial design, it was found that it is possible to use DAF physical variables (saturation pressure and saturated water flow rate) as input signs to the neural models, since these variables are statistically significant and influence the removal of turbidity promoted by the process in the range of operational conditions tested. The TBD$_R$ rises on average $2.70\% \pm 0.69\%$ and $2.12\% \pm 0.69\%$ when the saturated water flow rate and the saturation pressure increase, respectively.

Both models provided predictions consistent with the targets used in the training stage. The time-delay recurrent neural networks were able to deal with the nonlinearity of the process and the temporal dependence of the input and output variables. The real-time model, fed with data directly measured by the prototype sensors during an experimental run, presented a high regression coefficient of 0.9717 and MSE of 1.0482. The simulation model, which predicted
turbidity removal without the need to perform an experiment, provided a linear correlation of 0.9475 and MSE of 1.8640. Therefore, the neural empirical models developed are powerful tools to continue studying the phenomena involved in the DAF process, aiming to increase process efficiency and allowing the advancement of research that seeks the continuous improvement of DAF unit operations in water treatment.

Given the good initial results, complementary research is being developed to expand the scope of the models obtained for the dissolved air flotation process. Other variables will be tested as input signals of the neural network model, like the coagulation and flocculation pH, as well as the chemical dosages. Besides that, more experimental runs will be conducted to increase the information volume of the training database and to allow the study of other operational conditions when different initial turbidity values are considered.

ACKNOWLEDGEMENT

The authors appreciate the financial support provided by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (Grant number: 3.50.03017.03.4P8).

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: https://github.com/anasouza26/daf-ann.

REFERENCES

Alex, J., Jumar, U. & Bitter, U. 1999 On-line simulation of wastewater treatment plants. IFAC Proceedings 32 (2), 5794–5799. https://doi.org/10.1016/S1474-6670(17)56989-X.

Bergh, L. G. & Yianatos, J. B. 1995 Dynamic simulation of operating variables in flotation columns. Minerals Engineering 8 (6), 603–613. https://doi.org/10.1016/0892-6875(95)00024-K.

Bouchard, J., Desbiens, A. & del Villar, R. 2014 Column flotation simulation: a dynamic framework. Minerals Engineering 55, 30–41. https://doi.org/10.1016/j.mineng.2013.07.021.

Carvalho, M. T. & Durão, F. 2002 Control of a flotation column using fuzzy logic inference. Fuzzy Sets and Systems 125 (1), 121–133. https://doi.org/10.1016/S0165-0114(01)00048-3.

Crittenden, J. C., Trussell, R. R., Hand, D. W., Howe, K. J. & Tchobanoglous, G. 2012 MWH’s Water Treatment: Principles and Design, 3rd edn. John Wiley & Sons, Hoboken, NJ, USA.

Edzwald, J. K. 1995 Principles and applications of dissolved air flotation. Water Science and Technology 31 (3–4), 1–23. https://doi.org/10.1016/0273-1223(95)00200-7.

Edzwald, J. K. 2010 Dissolved air flotation and me. Water Research 44 (7), 2077–2106. https://doi.org/10.1016/j.watres.2009.12.040.

Edzwald, J. K. & Haarhoff, J. 2011 Dissolved Air Flotation for Water Clarification, 1st edn. McGraw-Hill, New York, USA.

Fanaie, V. R. & Khiadani, M. 2020 Effect of salinity on air dissolution, size distribution of microbubbles, and hydrodynamics of a dissolved air flotation (DAF) system. Colloids and Surfaces A: Physicochemical and Engineering Aspects 591, 124547. https://doi.org/10.1016/j.colsurfa.2020.124547.

Fonseca, R. R., Thompson, J. P., Franco, I. C. & da Silva, F. V. 2017 Automation and control of a dissolved air flotation pilot plant. IFAC-PapersOnLine 50 (1), 3911–3916. https://doi.org/10.1016/j.ifacol.2017.08.364.

Fukushi, K., Tambo, N. & Matsui, Y. 1995 A kinetic model for dissolved air flotation in water and wastewater treatment. Water Science and Technology 31 (3–4), 37–47. https://doi.org/10.1016/0273-1223(95)00202-X.

Haarhoff, J. 2008 Dissolved air flotation: progress and prospects for drinking water treatment. Journal of Water Supply: Research and Technology – AQUA 57 (8), 555–567. https://doi.org/10.1016/j.aqua.2008.04.046.

Hacifazlioglu, H. & Sutcu, H. 2007 Optimization of some parameters in column flotation and a comparison of conventional cell and column cell in terms of flotation performance. Journal of the Chinese Institute of Chemical Engineers 38 (3–4), 287–293. https://doi.org/10.1016/j.jcie.2007.03.006.

Hamed, M. M., Khalafallah, M. G. & Hassani, E. A. 2004 Prediction of wastewater treatment plant performance using artificial neural networks. Environmental Modelling & Software 19 (10), 919–928. https://doi.org/10.1016/j.envsoft.2003.10.005.

Haykin, S. 2009 Neural Networks and Learning Machines, 3rd edn. Prentice Hall, Upper Saddle River, NJ, USA.

Ksenofontov, B. S. & Ivanov, M. V. 2013 A novel multistage kinetic modeling of flotation for wastewater treatment. Water Science & Technology 68 (4), 807–812. https://doi.org/10.2166/wst.2013.303.

Labidi, J., Pélach, M. À., Turon, X. & Mutijé, P. 2007 Predicting flotation efficiency using neural networks. Chemical Engineering and Processing: Process Intensiﬁcation 46 (4), 314–322. https://doi.org/10.1016/j.cep.2006.06.011.
Leijnen, S. & van Veen, F. 2020 The neural network zoo. Proceedings 47 (1), 9. https://doi.org/10.3390/proceedings2020047009.

Maier, H. R., Morgan, N. & Chow, C. W. K. 2004 Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environmental Modelling & Software 19 (5), 485–494. https://doi.org/10.1016/S1364-8152(03)00163-4.

Man, Y., Shen, W., Chen, X., Long, Z. & Pons, M.-N. 2017 Modeling and simulation of the industrial sequencing batch reactor wastewater treatment process for cleaner production in pulp and paper mills. Journal of Cleaner Production 167, 643–652. https://doi.org/10.1016/j.jclepro.2017.08.236.

Miyata, J., Sudo, T. & Tsubone, T. 2004 Wastewater treatment processing simulation technology using ‘activated sludge model’. JFE Technical Report 3, 27–34.

Nakhaei, F., Mosavi, M. R., Sam, A. & Vaghei, Y. 2012 Recovery and grade accurate prediction of pilot plant flotation column concentrate: neural network and statistical techniques. International Journal of Mineral Processing 110–111, 140–154. https://doi.org/10.1016/j.minpro.2012.03.003.

Rodrigues, J. P., Batista, J. N. M. & Bettega, R. 2019 Application of population balance equations and interaction models in CFD simulation of the bubble distribution in dissolved air flotation. Colloids and Surfaces A: Physicochemical and Engineering Aspects 577, 723–732. https://doi.org/10.1016/j.colsurfa.2019.06.032.

Sakizadeh, M., Malian, A. & Ahmadpour, E. 2015 Groundwater quality modeling with a small data set. Groundwater 54 (1), 115–120. https://doi.org/10.1111/gwat.12317.

Salari, M., Shahid, E. S., Afzali, S. H., Ehteshami, M., Conti, G. O., Derakhshan, Z. & Sheibani, S. N. 2018 Quality assessment and artificial neural networks modeling for characterization of chemical and physical parameters of potable water. Food and Chemical Toxicology 118, 212–219. https://doi.org/10.1016/j.fct.2018.04.036.

Seco, A., Ruano, M. V., Ruiz-Martinez, A., Robles, A., Barat, R., Serralta, J. & Ferrer, J. 2020 Plant-wide modelling in wastewater treatment: showcasing experiences using the biological nutrient removal model. Water Science and Technology 81 (8), 1700–1714. https://doi.org/10.2166/wst.2020.056.

Shawwa, A. R. & Smith, D. W. 2000 Dissolved air flotation model for drinking water treatment. Canadian Journal of Civil Engineering 27 (2), 373–382. https://doi.org/10.1139/cjce-27-2-373.

Singh, K. P., Basant, A., Malik, A. & Jain, G. 2009 Artificial neural network modeling of the river water quality – a case study. Ecological Modelling 220 (6), 888–895. https://doi.org/10.1016/j.ecolmodel.2009.01.004.

Souza, A. C. O., Ferreira, N. L. & Vasconcelos, F. 2020 Dissolved Air Flotation Data for Turbidity Removal Empirical Modeling, v1. https://github.com/anasouza26/daf-ann.

Venkatasubramanian, V. 2018 The promise of artificial intelligence in chemical engineering: is it here, finally? AIChE Journal 65 (2), 466–478. https://doi.org/10.1002/aic.16489.

Zabel, T. 1985 The advantages of dissolved-air flotation for water treatment. Journal American Water Works Association 77 (5), 42–46. https://doi.org/10.1002/j.1551-8833.1985.tb0537.x.

First received 17 December 2020; accepted in revised form 8 May 2021. Available online 20 May 2021