Prediction of the Bicarbonate Amount in Drinking Water in the Region of Médéa Using Artificial Neural Network Modelling

H. Tahraoui, a A. E. Belhadj,a* and A. E. Hamitouche b

a Laboratory of Biomaterials and Transport Phenomenon (LBMPT), Faculty of Technology, University Yahia Fares of Médéa, 26 000, Algeria
b Center for Scientific and Technical Research in Physico-chemical Analysis CRAPC, BP 384, Bou-Ismail, RP 42 004, Tipaza, Algeria

Abstract

The region of Médéa (Algeria) located in an agricultural site requires a large amount of drinking water. For this purpose, the water analyses in question are imperative. To examine the quality of the drinking water in this region, firstly, an experimental protocol was done in order to obtain a dataset by taking into account several physicochemical parameters. Secondly, the obtained data set was divided into two parts to form the artificial neural network, where 70 % of the data set was used for training, and the remaining 30 % was also divided into two equal parts: one for testing and the other for validation of the model. The intelligent model obtained was evaluated as a function of the correlation coefficient nearest to 1 and lowest mean square error (RMSE). A set of 64 data points were used in this study. Eighteen parameters in the input layer, five neurons in the hidden layer, and one parameter in the output layer were used for the ANN modelling. Levenberg Marquardt learning (LM) algorithm, logarithmic sigmoid, and linear transfer function were used, respectively, for the hidden and the output layers. The results obtained during the present study showed a correlation coefficient of $R = 0.99276$ with root mean square error $RMSE = 11.52613 \text{ mg dm}^{-3}$. These results show that obtained ANN model gave far better and more significant results. It is obviously more accurate since its relative error is small with a correlation coefficient close to unity. Finally, it can be concluded that obtained model can effectively predict the rate of soluble bicarbonate in drinking water in the Médéa region.

Keywords

Drinking water, physicochemical parameters, bicarbonate, modelling, artificial neural networks

1 Introduction

Industrial, agricultural, and urban development alters the quality of water and makes it unsafe. This is the case in Médéa region, which is subjected to diversification and increase in the quantity of pollutants released in the aquatic environment without treatment. Depending on the origin of the waste, the pollution may be of a chemical nature, especially heavy metals. To predict the amount of a pollutant, depending on many physicochemical parameters, we refer to methods based on mathematical and artificial intelligence models, such as linear regression (RLM) and artificial neural networks (ANN). Neural networks have been highly successful in modelling and predicting environmental parameters. They have been applied for various purposes, e.g.: prediction of groundwater remediation costs for drinking use based on quality of water resource; modelling for nitrate concentration in groundwater; prediction of groundwater suitability for irrigation; modelling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis; prediction of the quality of public water supply; modelling of TDS concentrations in river water; prediction of aluminium pitting in natural waters; prediction of fluoride concentration; prediction of bromate removal in drinking water; estimating rainfall relationship with river pollution; modelling of total dissolved solids; forecasting nitrate concentration in groundwater; effect of drought on pollution of river station by ANN; and prediction of heavy metal concentrations. All these works are linked to different physicochemical parameters of water.

In this work, we propose the use of ANN for the prediction of bicarbonate content of surface waters based on the physicochemical parameters.

2 Artificial neural network

Artificial neural network (ANN) represents a set of algorithms, the design of derives from and is schematically inspired by the functioning of biological neurons. ANNs are now used as a very powerful tool in modelling and analysing processes, as well as in predicting the behaviour of a given system. The ANN structure consists essentially of an input layer (independent variables), a number of hidden layers, and an output layer. Each of these layers consists of a number of interconnected processing units called neurons. These neurons interact by sending signals, and they are connected to all the neurons in the previous and next layers by links called weights and links. The architecture of the ANN model is shown in Fig. 1.

*Corresponding author: Dr Abd-Elmouneim Belhadj
Email: belhadj_1@yahoo.fr
3 Materials and methods

In this work, a procedure based on the development and optimisation of the architecture of an ANN using software MATLAB R2013a, is described by a flowchart presented in Fig. 2.20,21

3.1 Database

The data used for this study were obtained from the experimental analysis of several water samples carried out for 84 different days in all seasons in the period from 2018 to 2019 at Médéa region, Algeria. Analyses were done according to the Jean Rodier’s book of water analysis 9th Ed.22

The dependent variables are the contents of bicarbonate (HCO$_3^-$) soluble in water. The independent variables are the physicochemical parameters as follows: conductivity, turbidity, pH, hardness, calcium, magnesium, chlorides, total alkalinity titer (TAC), organic matter, nitrogen dioxide, nitrates, sodium, sulphates, potassium, heavy metals (Mn$^{2+}$, Fe$^{3+}$, and Al$^+$), and dry residues.

3.2 Normalisation and data pre-processing

The primary purpose of the data transformation is to modify the distribution of input variables so that they can better match outputs. Before training and validation, we scaled the inputs and targets using a normalised equation where the data always fall within a specified range:

$$x_{ni} = (y_{max} - y_{min}) \left( \frac{x - x_{min}}{x_{max} - x_{min}} \right) + y_{min} \tag{1}$$

The experimental data were normalised in the interval $[-1,1]$ using the mapminmax function,23 where $x_{ni}$ is the data value after normalisation, $x_{max}$ and $x_{min}$ denote the maximum and minimum of the data, respectively, $y_{max}$ and $y_{min}$ are taken as $-1$ and $1$; and $x$ represents the original values.

3.3 Data modelling techniques

Data modelling was carried out by ANNs that made it possible to justify the predictive quality of the models, by applying the same techniques on a set of data relating to 70% of the samples, chosen at random, among the totality of the samples, which constituted the group for learning a predictive model of the dependent variable. The remaining 30% of the samples, which did not participate in model learning, were divided in two (15% for the test, and 15% for the validation) to examine the validity and performance of the prediction of these models.24
3.4 Modelling performance

In the current study, we present the different statistical parameters used in order to evaluate the effectiveness of each network and its ability to make precise predictions.\textsuperscript{25–27}

The correlation coefficient ($R$), root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), standard preaching error (ESP), and error of prediction error model (EPM) were used to estimate the performance of the model. These statistical parameters were calculated using the following equations:\textsuperscript{28,29}

\begin{align}
R &= \frac{\sum_{i=1}^{N} (y_{\text{exp}} - \bar{y}_{\text{exp}}) (y_{\text{pred}} - \bar{y}_{\text{pred}})}{\sqrt{\sum_{i=1}^{N} (y_{\text{exp}} - \bar{y}_{\text{exp}})^2 \sum_{i=1}^{N} (y_{\text{pred}} - \bar{y}_{\text{pred}})^2}} \\
\text{RMSE} &= \sqrt{\frac{1}{N} \left( \sum_{i=1}^{N} (y_{\text{exp}} - y_{\text{pred}}) \right)^2} \\
\text{MSE} &= \frac{1}{N} \left( \sum_{i=1}^{N} (y_{\text{exp}} - y_{\text{pred}})^2 \right) \\
\text{MAE} &= \frac{1}{N} \sum_{i=1}^{N} |y_{\text{exp}} - y_{\text{pred}}| \\
\text{ESP}(\%) &= \frac{\text{RMSE}}{\bar{y}_{\text{exp}}} \cdot 100 \\
\text{EPM}(\%) &= \frac{100}{N} \sum_{i=1}^{N} \left( \frac{y_{\text{exp}} - y_{\text{pred}}}{y_{\text{exp}}} \right)
\end{align}

\begin{itemize}
\item $N$ is the number of data; $y_{\text{exp}}$ and $y_{\text{pred}}$ are the experimental and the predicted values, respectively, and $\bar{y}_{\text{exp}}$ and $\bar{y}_{\text{pred}}$ are the average values of the experimental and the predicted values, respectively.
\end{itemize}

4 Results and discussion

Preliminary tests have shown that, in order to improve the performance of a model established by ANN, it is necessary to modify the architecture of the network, by playing mainly on the number of hidden layers, or on the number of hidden layers, hidden neurons and/or the number of training cycles (number of iterations). For this, we successively changed the number of hidden neurons ($NH = 3–15$). The results of these tests are shown in Table 1. It was observed that the minimum of the MSE, the number of iterations, and the maximum correlation coefficient had been reached when $NH = 5$. The optimal acknowledgement architecture obtained from ANN has a topology of $18–5–1$ (Table 1), and its architecture is illustrated in Figs. 3 and 4. We could therefore choose five neurons of the hidden layer of the network in the present study, in order to predict the bicarbonate concentrations.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
Type of network & Learning algorithm & Input layer & Hidden layer & Output layer & \\
\hline
ANN & LM & 18 & 5 & Sigmoid & 1 & Linear \\
\hline
\end{tabular}
\caption{Structure of neural networks}
\end{table}

Fig. 5 describes the results of the ANN training phase. It showed that, after 236 iterations, the desired result had been achieved with 5 neurons in the hidden layer, the three curves related to the evolution of the error of the three phases converged correctly towards an acceptable MSE.

The results obtained (Fig. 6) enabled extraction of the different values relating to the learning parameters:

\begin{itemize}
\item maximum number of iterations (Epochs) = 236,
\item learning rate ($\eta$) = 0.001,
\item minimum gradient = $8.2405 \times 10^{-16}$.
\end{itemize}

Once the architecture of the neural network model is obtained, the model should be validated by comparing its results with the experimental data obtained in the laboratory. The results given in Fig. 7 show the performance and the efficiency of the developed ANN model. The relevance
Fig. 5 – Evolution of the MSE

Fig. 6 – Gradient variations of error, learning rate, and validation error as a function of number of iterations

Fig. 7 – Relation between observed bicarbonate levels and those estimated by ANN model
Results of the prediction performances in terms of all errors and in terms of the agreement vector values \([\alpha = \text{slope}, \beta (y \text{ intercept}), R (\text{correlation coefficient})]\) are summarised in Table 2. These results show a good correlation between the predicted ANN and the experimental values with a high correlation coefficient \((R = 0.99276)\).

Fig. 9 again shows the performance of the ANN configuration \([18–5–1]\) with a better superposition of the curves plotted experimentally and those predicted by the model deduced during our study. The results obtained by this study show a good agreement explained by a high coefficient of correlation, which has a value close to unity. In addition, the values of errors including \((\text{MSE, RMSE, ESP, EPM, MAE})\) are very low.
4.2 Residues study

Errors made by the artificial neuron networks model on the samples that were used in this study are named residues. Thus, the study of the relationship between the estimated contents of bicarbonate by the mathematical models and their residues \((y_{\text{exp}} - y_{\text{pred}})\) allowed us to ensure the performance of the model and verify them experimentally.

Table 2 – Prediction performance results in terms of all errors and agreement vector \([\alpha, \beta, R]\)

|        | RMSE / mg dm\(^{-3}\) | MSE / mg dm\(^{-3}\) | ESP / % | EPM / % | MAE / mg dm\(^{-3}\) | \(R\)   | \(\alpha\)   | \(\beta\)  |
|--------|------------------------|-----------------------|---------|---------|-----------------------|--------|-------------|------------|
| Training | 9.68913055             | 93.8792508            | 4.3826748 | 1.31881052 | 6.54764823             | 0.99523 | 0.9835336  | 3.6139936  |
| Test    | 15.734532              | 247.575496            | 8.28240564 | 0.09771214 | 10.9639556             | 0.9721 | 0.9592193  | −2.950842  |
| Validation | 14.0333691           | 196.935448            | 6.44782517 | 0.09058195 | 12.0114691             | 0.99106 | 0.9868746  | 0.2635232  |
| All     | 11.526138              | 132.85185             | 5.3433443 | 3.13764295 | 8.02930904             | 0.99276 | 0.9880565  | 0.5073722  |

Fig. 9 – Relationship between experimental data and the predicted data of samples using ANN modelling: (a) training, (b) test, (c) validation, and (d) all data.

Fig. 10 shows the relationships between the estimated levels of bicarbonate in the water of Médéa region and their residues obtained using neural networks (ANN), respectively. It shows that the residues obtained by neural networks are much less dispersed (closer to zero), and a clear improvement in the distribution of residues depending on the nature of the samples.
This distribution proves the predictive power of this model established by the neural networks in the prediction of the contents of bicarbonate from the environmental parameters. In general, the result obtained is very satisfactory and justifies the use of the neural network approach in the prediction of bicarbonate contents in Médéa region.

Fig. 10 – Residues relating to the models established by the artificial neural network depending on estimated values.

5 Conclusion

In the present study, the prediction of bicarbonate concentrations in drinking water in the Médéa region was obtained using artificial neural networks (ANN) with supervised learning, involving the “Levenberg-Marquardt” algorithm. This algorithm gives better results in terms of speed, convergence, and generation of performance. The results showed a high learning and predictive capacity for bicarbonate concentrations with a very high correlation coefficient of 0.99276 and a very low mean square error (11.52613 mg dm⁻³) for the whole database. In addition, they showed a better choice of learning algorithm, activation functions, and network architecture [18–5–1] obtained by applying statistical indicators of robustness. The prediction by artificial neural networks also shows a good correlation between the experimental and predicted values, meaning that the ANN model has better predictive power. This performance seems to be due to the fact that the concentrations of bicarbonates in drinking water in Médéa region are linked to the physicochemical characteristics of the environment by non-linear relationships.

This encourages us to consider in future the development of other aspects of this study related to the use of more parameters and prediction of other properties.

References

Literatura

1. J. S. Torrecilla, M. L. Mena, P. Yáñez-Sedeño, J. García, Application of artificial neural network to the determination of phenolic compounds in olive oil mill wastewater, J. Food Eng. 81 (3) (2007) 544–552, doi: https://doi.org/10.1016/j.jfoodeng.2006.12.003.
2. J. Rade, Développement d’un modèle statistique neuronal pour la description fine de la pollution atmosphérique par le dioxyde d’azote: application à la région Parisienne, Paris Est, 2008.
3. P. Perez, A. Trier, Prediction of NO and NO₂ concentrations near a street with heavy traffic in Santiago, Chile, Atmosph. Environ. 35 (10) (2001) 1783–1789, doi: https://doi.org/10.1016/S1352-2310(00)00288-0.
4. F. Qaderi, E. Babanezhad, Prediction of the groundwater remediation costs for drinking use based on quality of water resource, using artificial neural network, J. Clean. Prod. 161 (2017) 840–849, doi: https://doi.org/10.1016/j.jclepro.2017.05.187.
5. V. Wagh, D. Panaskar, A. Muley, S. Mukate, S. Gaikward, Neural-network modelling for nitrate concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India, Groundwater Sustain. Develop. 7 (2018) 436–445, doi: https://doi.org/10.1016/j.gsd.2017.12.012.
6. V. M. Wagh, D. B. Panaskar, A. A. Muley, S. V. Mukate, Y. P. Lolage, M. L. Aamalawar, Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India, Model. Earth Syst. Environ. 2 (2016) 1–10, doi: https://doi.org/10.1007/s40808-016-0250-3.
7. A. R. Khataee, M. B. Kasiri, Artificial neural network modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis, J. Mol. Catal. A-Chem. 331 (2010) 86–100, doi: https://doi.org/10.1016/j.molcata.2010.07.016.
8. H. Vicente, S. Dias, A. Fernandes, A. Abelha, J. Machado, J. Neves, Prediction of the quality of public water supply using artificial neural networks, J. Water Suppl. Res. Technol. – AQUA 61 (2012) 446–459, doi: https://doi.org/10.2166/aqua.2012.014.
9. P. Mageshkumar, T. Pradeep, M. R. John, S. A. Raj, S. Anandakumar, Neural network modeling of TDS concentrations in Cauvery River water, Tamilnadu, India, Int. J. Appl. Sci. Eng. Res. 1 (2012) 739–746, doi: https://doi.org/10.6088/ijaser.0020101074.
10. J. Leifer, J. I. Mickalonis, Prediction of aluminum pitting in natural waters via artificial neural network analysis, Corrosion 56 (2000) 563–571, doi: https://doi.org/10.5066/1.3280560.
11. A. A. Nadiri, E. Fijani, F. T.-C. Tsai, A. Asghari Moghadam, Supervised committee machine with artificial intelligence for prediction of fluoride concentration, J. Hydroinorm. 15 (2013) 1474–1490, doi: https://doi.org/10.2166/hydro.2013.008.
12. E. Karadurmus, N. Taşkan, E. Göz, M. Yüceer, Prediction of bromate removal in drinking water using artificial neural networks, Ozone: Sci. Eng. 41 (2019) 118–127, doi: https://doi.org/10.1080/01919512.2018.1510763.
13. N. Mohsenilar, M. Mohsenilar, K. Mohsenilar, Using Artificial Neural Network (ANN) for estimating rainfall relationship with river pollution, Adv. Environ. Biol. 5 (2011) 1202–1206.
14. S. Nemati, L. Naghipour, M. H. F. Fard, Artificial neural network modeling of total dissolved solid in the Simineh River, Iran, J. Civ. Eng. Urban. 4 (1) (2014) 8–13.
15. A. H. Zare, V. M. Bayat, A. P. Daneshkare, Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models, Int. Agrophys. 25 (2011) 187–192.
16. F. A. Pour, K. Mohsenilar, E. Pazira, Affect of drought on pollution of Lenj station of Zayandehrood River by Artificial Neural Network (ANN), Adv. Environ. Biol. 5 (7) (2011)
1461–1464.

17. A. Abdallaoui, H. El Badaoui, Prédiction des teneurs en métaux lourds des sédiments à partir de leurs caractéristiques physico-chimiques, J. Phys. Chem. News 58 (2011) 90–97.

18. L. Cavas, Z. Karabay, H. Alyuruk, H. Doğan, G. K. Demir, Thomas and artificial neural network models for the fixed-bed adsorption of methylene blue by a beach waste Posidonia oceanica (L.) dead leaves, Chem. Eng. J. 171 (2) (2011) 557–562, doi: https://doi.org/10.1016/j.cej.2011.04.030.

19. R. M. Aghav, S. Kumar, S. N. Mukherjee, Artificial neural network modeling in competitive adsorption of phenol and resorcinol from water environment using some carbonaceous adsorbents, J. Hazard. Mater. 188 (1–3) (2011) 67–77, doi: https://doi.org/10.1016/j.jhazmat.2011.01.067.

20. H. Ousmana, A. E. Hmaid, M. Berrada, B. Dammati, I. Etabaai, A. Essahlaoui, Development of a Neural Network Approach for Predicting nitrate and sulfate concentration in three lakes: Ifrah, Iffer and Afourgargh, Middle Atlas Morocco, Mor. J. Chem. 6 (2) (2018) 245–255, url: https://revues.imist.ma/index.php/journal?mo=jourchem&page=article&op=view&path%5B%5D=5939.

21. M. Sediri, S. Hanini, H. Cherif, M. Laidi, S. A. Turki, Dynamic Adsorption Modelling of P-Nitrophenol in Aqueous Solution Using Artificial Neural Network, J. Mater. Environ. Sci. 8 (7) (2017) 2282–2287.

22. J. Rodier, B. Legube, N. Merlet, R. Brunet, L’analyse de l’eau: Eaux naturelles, eaux résiduaires, eau de mer, Dunod, Paris, 2009, 1526 p.

23. R. Yacef, A. Mellit, S. Belaid, Z. Şen, New combined models for estimating daily global solar radiation from measured air temperature in semi-arid climates: application in Ghardaïa, Algeria, Energ. Convers. Manage. 79 (2014) 606–615, doi: https://doi.org/10.1016/j.enconman.2013.12.057.

24. H. El Badaoui, A. Abdallaoui, I. Manssouri, L. Lanculet, Elaboration of models mathématiques stochastiques pour la prédicttion des teneurs en métaux lourds des eaux superficielles en utilisant les réseaux de neurones artificiels et la régression lineaire multiple, J. Hydrocarb. Mine. Environ. Res. 3 (2012) 31–36.

25. I. Manssouri, M. Manssouri, B. El Kihel, Fault detection by K-NN algorithm and MLP neural networks in a distillation column: Comparative study, J. Inf. Intellig. Knowl. 3 (3) (2011) 201.

26. I. Manssouri, A. El Hmaid, T. E. Manssouri, B. El Mouumni, Prediction levels of heavy metals (Zn, Cu and Mn) in current Holocene deposits of the eastern part of the Mediterranean Moroccan margin (Alboran Sea), IOSR J. Comp. Eng. 16 (2014) 117–123, doi: https://doi.org/10.9790/0661-1618117123.

27. O. R. Dolling, E. A. Varas, Artificial neural networks for streamflow prediction, J. Hydraul. Res. 40 (5) (2002) 547–554, doi: https://doi.org/10.1080/00221680209499899.

28. A. D. Belsley, E. Kuh, R. E. Welsch, Regression diagnostics: identifying influential data and sources of collinearity, Wiley, 1980, doi: https://doi.org/10.1002/0471725153.

29. S. H. Hong, M. W. Lee, D. S. Lee, J. M. Park, Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks, Biochem. Eng. J. 35 (3) (2007) 365–370, doi: https://doi.org/10.1016/j bj.2007.01.033.

30. C. Voyant, M. Muselli, C. Paoli, M.-L. Nivet, Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation, Energy 36 (1) (2011) 348–359, doi: https://doi.org/10.1016/j. energy.2010.10.032.

SAŽETAK

Predviđanje količine bikarbonata u pitkoj vodi regije Médéa modeliranjem umjetnom neuronskom mrežom

Hichem Tahraoui, Abd-Elmonime Belhadj i Adhya-Édoume Hamitouche

Regija Médéa (Alžir) smještena na poljoprivrednom zemljištu zahtijeva veliku količinu pitke vode te je stoga analiza vode od iznimne važnosti. Da bi se ispitao razvoj kvalitete pitke vode u toj regiji, najprije je napravljen eksperimentalni protokol za dobivanje skupa podataka uzimajući u obzir nekoliko fizikalno-kemijskih parametara. Zatim je dobiven skup podataka podijeljen na dva dijela: jedan za testiranje, a drugi za vali-daci modela. Dobiveni model neuronske mreže dao daleko bolje rezultate, jer je točniji a njegova relativna pogreška je mala s koeficijentom korelacije blizu 1. Konačno, zaključeno je da taj model može učinkovito predvidjeti brzinu topljivosti bikarbonata u vodi za piće u regiji Médéa.

Ključne riječi: Pitka voda, fizikalno-kemijski parametri, bikarbonat, modeliranje, umjetne neuronske mreže

Laboratory of Biomaterials and Transport Phenomenon (LBMP), Faculty of Technology, University Yahia Fares of Médéa, 26 000, Alžir

Center for Scientific and Technical Research in Physico-chemical Analysis CRAPC, BP 384, Bou-Ismail, RP 42 004, Tipaza, Alžir

Izvorni znanstveni rad Prispjelo 1. siječnja 2020. Prihvaćeno 19. veljače 2020.