A hybrid model for modelling the salinity of the Tafna River in Algeria

Khemissi HOUARI1) ABCDEF, Tarik HARTANI2) DEF, Boualem REMINI3) ABCDEF, Abdelouhab LEFKIR4) EF, Leila ABDA5) AB, Salim HEDDAM6) C

1) University M’hamed Bougara, Faculty of Sciences, Department of Agronomy, 35000 Boumerdes, Algeria; e-mail: houarikhemissi@gmail.com
2) orcid.org/0000-0002-2373-9516; Tipaza University Center, Agricultural Water Management Laboratory, El Harrach, Algiers, Algeria; e-mail: hartani.tarik@gmail.com
3) orcid.org/0000-0002-8735-7250; Saad Dahlab University, Department of Rural Engineering, Blida, Algeria; e-mail: reminib@yahoo.fr
4) orcid.org/0000-0002-8002-0129; National School of Public Works, LTPiTE Laboratory, Kouba, Algiers, Algeria; e-mail: a.lefkir@enstp.edut.dz
5) National Water Resources Agency, Algiers, Algeria; e-mail: abdaleila16@gmail.com
6) orcid.org/0000-0002-8055-8463; University 20 August 1955, Faculty of Science, Agronomy Department, Hydraulics Division, Skikda, Algeria; e-mail: heddamsalim@yahoo.fr

For citation: Houari K., Hartani T., Remini B., Lefkir A., Abda L., Heddam S. 2019. A hybrid model for modelling the salinity of the Tafna River in Algeria. Journal of Water and Land Development. No. 40 (I–III) p. 127–135. DOI: 10.2478/jwld-2019-0014.

Abstract
In this paper, the capacity of an Adaptive-Network-Based Fuzzy Inference System (ANFIS) for predicting salinity of the Tafna River is investigated.

Time series data of daily liquid flow and saline concentrations from the gauging station of Pierre du Chat (160801) were used for training, validation and testing the hybrid model. Different methods were used to test the accuracy of our results, i.e. coefficient of determination ($R^2$), Nash–Sutcliffe efficiency coefficient ($E$), root of the mean squared error (RSR) and graphic techniques. The model produced satisfactory results and showed a very good agreement between the predicted and observed data, with $R^2$ equal (88% for training, 78.01% validation and 80.00% for testing), $E$ equal (85.84% for training, 82.51% validation and 78.17% for testing), and RSR equal (2% for training, 10% validation and 49% for testing).

Key words: Adaptive-Network-Based Fuzzy Inference System (ANFIS), hybrid model, neuro-fuzzy, salinity, salt flow, Tafna River

INTRODUCTION

The salinity of streams and rivers is a complex hydrological and environmental phenomenon due to the large number of obscure parameters. The processes involved in salinization so complex that it is difficult to establish a general empirical analytical model to provide accurate prediction. Artificial intelligence is a viable and fully justified alternative for modelling phenomena with non-linear behaviour [BOUZERIA ET AL. 2017; CHEN ET AL. 2008; ME-
(neuro-fuzzy) “capture” the behaviour of non-linear systems quickly and accurately, even more so than other methods. It is one of the most widely used models for predicting time series and thus provides a good basis for comparison [KOUJOK et al. 2008; MARASHI et al. 2018; KESKIN et al. 2010; KISI 2006] and MOHANDES et al. [2011] confirm the performance of hybrid model prediction compared to neural networks and fuzzy systems.

Neuro-fuzzy systems are currently one of the “flavours of the month” in the neural network and fuzzy logic communities. They attempt to combine the structural and learning abilities of a neural network with the linguistic initialisation and validation aspects of a fuzzy system [BROWN 1996]. Neural networks are interesting tools for recognizing patterns; they do not explain how they reach their decisions. Likewise, for systems of fuzzy logic which can run with inaccurate information, there are interesting ways to explain their decisions but it does not automatically acquire the rules they use to make these decisions. These limitations have been a reason behind the creation of intelligent hybrid systems where two or more techniques are combined in a way to overcome the limitations of a single technique [CHANG, CHANG 2006; FULLER 2000].

The hybrid fuzzy neural networks have a tremendous potential to solve engineering problems. It is improper to expect that if the individual technologies are good then hybridization of technologies should turn out to be even better. Hybridization is performed for the purpose of investing better methods of problem solving [CHENNakesava 2008]. Using a system created on the basis of a hybrid approach, existing methods and algorithms can be improved and more effective approaches can be developed to solve existing problems [SKVORTSOVA et al. 2017].

The aim of this study is the developed a neuro-fuzzy model (ANFIS) capable of predict and estimate the salinity of the Tafha River at every moment. Three observation times over 13 years and more than, 420 pairs of fluid flow and saline concentrations were used, to evaluate the performance of this model. Once the model is validated, watershed managers can integrate it into a decision support system to control the chemical quality.

**HYBRID MODEL (ANFIS)**

Adaptive neuro-fuzzy inference system (ANFIS), first introduced by JANG [1993]. ANFIS is capable of approximating any real continuous function on a compact set-of parameters to any degree of accuracy [JANG et al. 1997]. ANFIS identifies a set of parameters through a hybrid learning rule combining back propagation gradient descent error gradient and a least squared error method. There are two main approaches to fuzzy inference systems, namely the Mamdani approaches [MAMDANI, ASSILIAN 1975] and the TAKAGI, SUGENO [1985] approaches. The artificial neural networks ANN provides connectionist structures and learning abilities to the fuzzy systems whereas the fuzzy systems offer ANN a structured framework with high level IF-THEN rule thinking and reasoning [AVCI 2008; MINH et al. 2017]. The main type of association between neural networks and fuzzy systems is when a fuzzy inference system is put in the form of a multilayered network [BUCKLEY, HAYASHI 1994], in which the weights correspond to the parameters of the system. Network architecture depending on the type of rules and methods of inference, aggregation and defuzzification chosen.

**STRUCTURE OF HYBRID MODEL**

The fuzzy inference system under consideration has two inputs \((x, y)\) and one output \((f)\). The rule base contains the fuzzy if-then rules of Takagi and Sugeno’s type [SUGENO, KANG 1988; TAKAGI, SUGENO 1983] as follows:

\[
\text{If } x \text{ is } A \text{ and } y \text{ is } B \text{ then } z = f(x, y)
\]

Where: \(A\) and \(B\) are the fuzzy sets in the antecedents and \(z = f(x, y)\) is a crisp function in the consequent. Usually \(f(x, y)\) is a polynomial for the input variables \(x\) and \(y\). But it can also be any other function that can approximately describe the output of the system within the fuzzy region as specified by the antecedent. When \(f(x, y)\) is a constant, a zero order Sugeno fuzzy model is formed which may be considered to be a special case of Mamdani fuzzy inference system [MAMDANI, ASSILIAN 1975] where each rule consequent is specified by a fuzzy singleton. If \(f(x, y)\) is taken to be a first order polynomial a first order Sugeno. For a first order, two rule Sugeno fuzzy inference system, the two rules may be stated [SENGUR 2008; ÜBEYLI 2008; YING, PAN 2008] as:

\[
\begin{align*}
\text{Rule 1: If } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \text{ then } f_1 &= p_1 x + q_1 y + r_1 \\
\text{Rule 2: If } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \text{ then } f_2 &= p_2 x + q_2 y + r_2
\end{align*}
\]

ANFIS architecture is an adaptive network that uses supervised learning on learning algorithm, which has a function similar to the model of Takagi–Sugeno fuzzy inference system. Figure 1 shows the scheme fuzzy reasoning mechanism for Takagi–Sugeno model and ANFIS architecture [SUPARTA, ALHASA 2013].

**First layer**

Every node \(i\) in this layer is adaptive with a node function

\[
Q_i^1 = \mu_{A_i}(x)
\]

Where: \(x\) is the input to node \(i\), \(A_i\) is the linguistic variable associated with this node function and \(\mu_{A_i}\) is the membership function of \(A_i\). Usually \(\mu_{A_i}(x)\) is chosen as

\[
\mu_{A_i}(x) = \frac{1}{1 + \left| \frac{x - c_i}{w_i} \right|^2}
\]

Second layer

Each node in this layer is a fixed node, which calculates the firing strength ‘\(w_i\)’ of a rule. The output of each node is the product of all the incoming signals to it and is given by

\[
Q_i^2 = w_i = \mu_{A_i}(x) \cdot \mu_{B_i}(y), \quad i = 1, 2
\]

© PAN in Warsaw, 2019; © ITP in Falenty, 2019; Journal of Water and Land Development. No. 40 (I–III)
A hybrid model for modelling the salinity of the Tafna River in Algeria

Third layer
Every node in this layer is a fixed node. Each $i^{th}$ node calculates the ratio of the $i^{th}$ rule’s firing strength to the sum of firing strengths of all the rules. The output from the $i^{th}$ node is the normalized firing strength given by:

$$Q_i^3 = \tilde{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \quad (8)$$

Fourth layer
Every node in this layer is an adaptive node with a node function given by:

$$Q_i^4 = \tilde{w}_i f_i = \tilde{w}(p_i x + q_i y + r_i), \quad i = 1, 2 \quad (9)$$

Fifth layer
The output is obtained by summing all the outputs of layer 4:

$$Q_i^5 = \Sigma_i \tilde{w}_i f_i = \frac{\Sigma_i w_i f_i}{\Sigma_i w_i} \quad (10)$$

LEARNING OF HYBRID MODEL

The adjustment of the ANFIS parameters is carried out during the learning phase. The first layer and the fourth layer contain the parameters that can be modified over time. In the first layer, it contains a nonlinear of the premises parameter while the fourth layer contains linear consequent parameters. A hybrid algorithm proposed by JANG [1993] will be used in this study to train these parameters. When these lasts are obtained, input data are passed back to the adaptive network input, and the output generated are compared with the actual output.

While backward path is run, the consequent parameters must be in a steady state. The error occurred during the comparison between the output generated with the actual output is propagated back to the first layer.

Next, after the consequent parameters are obtained, input data is passed back to the adaptive network input, and the output generated will be compared with the actual output.

While backward path is run, the consequent parameters must be in a steady state. The error occurred during the comparison between the output generated with the actual output is propagated back to the first layer. At the same time, parameter premises in the first layer are updated using learning methods of gradient descent or back propagation. With the use of hybrid learning algorithm that combines with the recursive least square estimator and the gradient descent methods, it can ensure the convergence rate is faster because it can reduce the dimensional search space in the original method of back propagation. One level of hybrid learning is called epochs [NAYAK et al. 2004; SUPARTA, ALHASA 2016].
VALIDATION OF THE MODEL

The coefficient of determination \( \left( R^2 \right) \), the root of the mean squared error \( \text{RSR} \) (ratio of the root mean square error to the standard deviation of measured) and the Nash–Sutcliffe efficiency coefficient \( (E) \) were used [NASH, SUTCLIFFE 1970]. These parameters are given by the following relationships respectively:

\[
R^2 = \left( \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \bar{Q})^2} \right)
\]

\[
\text{RSR} = \sqrt{\frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \bar{Q})^2}}\]

\[
E = 1 - \frac{\sum_{i=1}^{n} (Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n} (Q_i - \bar{Q})^2}\]

Where: \( Q_i \) is the measured value of the flow; \( \hat{Q}_i \) is the flow calculated by the model; \( \bar{Q} \) is the measured average rate and \( n \) is the number of data.

Data in Table 1 show the values of the \( \text{RSR} \), \( R^2 \) and \( E \) criteria corresponding to different degree of performance of the model at the daily scale.

| Rating            | \( E \) (%)       | \( \text{RSR} \) (%) | \( R^2 \) (%) |
|-------------------|-------------------|-----------------------|---------------|
| Very good         | \( 75 < E \leq 100 \) | \( 0 \leq \text{RSR} \leq 50 \) | \( 75 < R^2 \leq 100 \) |
| Good              | \( 65 < E \leq 7 \) | \( 50 < \text{RSR} \leq 60 \) | \( 65 < R^2 \leq 100 \) |
| Satisfactory      | \( 50 < E \leq 6 \) | \( 60 < \text{RSR} \leq 70 \) | \( 50 < R^2 \leq 100 \) |
| Unsatisfactory    | \( E \leq 50 \)   | \( \text{RSR} > 70 \) | \( R^2 \leq 50 \) |

Explanations: \( E \) = Nash–Sutcliffe efficiency coefficient; \( \text{RSR} \) = root of the mean squared error; \( R^2 \) = coefficient of determination.

STUDY AREA

The present work carried out on the Tafna River (Fig. 2). The Tafna River is 170 km long. It is located in the northwestern region of Algeria, draining the Tafna watershed with an area of 7,245 km². It is ranging from the Algerian-Moroccan border in the West to about 30 km East of Tlemcen. This river reaches the Mediterranean Sea near Beni-Saf city. It is located between 34°11' N and 35°19' N and 0°50' longitude and 2°20' W. The climate is Mediterranean with two main seasons: a long, dry, hot summer–autumn and a winter–spring with abrupt and frequent heavy rainfall. During the summer, most of the streams, especially in their downstream parts, become mostly dry between June and October [ZETTAM et al. 2017].

DATA AND METHODS

For simplicity, we assume the fuzzy inference system under consideration has two inputs daily liquid flow \( Q_l \) (m³·s⁻¹) and daily salt concentration \( C_s \) (g·dm⁻³). In this approach, 420 samples of liquid flow (m³·s⁻¹) and mean daily saline concentrations (g·dm⁻³) collected over three observation periods for 13 years (from 1977 to 1979 and from 1982 to 1991) were used (Fig. 3).
The flow is measured using a gauging station and saline concentrations are determined in the laboratory. The choice of these two parameters is based on the fact that the liquid flow is the causal factor and the saline concentrations is consequence factor. So that, we choose ‘Pierre du Chat’ hydrometric station which is based on their site, it is the outlet of the river and controls 80% of the Tafna watershed.

The dataset (420 samples) has been divided into three subsets: training, validation and testing.

- Training phase: the training data sample is a set of data that is independent follows the same population; 252 samples (60% of the population) are used for this phase from 26.09.1977 to 18.05.1979 and from 18.05.1979 to 26.01.1982.
- Test phase: the test stage, carried out with another sample of pairs (Cs and Qs) randomly drawn from the same population but for another observation period, from 18.05.1988 to 01.05.1990 with 84 samples. That is to say 20% of the population.
- Validation phase: the validation step, performed on a third sample (i.e. 20% of the population) with the same Cs and Qs, it allows to check the stability of the predictive ability from 02.05.1990 to 13.03.1991.

RESULTS AND DISCUSSION

The summary of statistical characteristics of the training, validation and test for observed and predicted saline flows are presented in Table 2. As gradually as we obtained data, the values of the saline flow of the samples ranged from 0.023 to 80.6 kg s⁻¹. This explains the influence of the runoff on the chemical quantity of the Tafna valley waters.

After the calibration and validation of ANFIS, their performance was evaluated using graphical techniques and three statistical criteria: determination coefficient ($R^2$), Nash–Sutcliffe efficiency coefficient ($E$), and the root of mean square error (RSR). Table 3 shows the best performance of the proposed hybrid model, with determination coefficient ($R^2$ = 88% for training, 78.01% for validation and 80.00% for tests), efficiency Nash–Sutcliffe coefficient ($E$ = 85.84% for training, 82.51% for validation) and root of mean squared error ($RSR$ = 2% for training, 10% for validation and 49% for the tests).

The scatter plots of observed values versus simulated saline flow rates analysed here are presented in Figure 4 respectively for the training, validation and tests phases. These statistics clearly show that the hybrid model is close to the adjustment line.

Figure 5 shows the time series of observed and expected salt fluxes during training, validation and testing. The model took precisely the non-linear pattern of salt flow during training and produced a good generalization during testing and validation.

During the study period, the flow values as well as the salt concentration of the hybrid model were used to predict and fill gaps in the salt concentration series. The hybrid model was used to estimate the amount of salinity (Fig. 6).

![Image](55x795 to 540x832)

Fig. 3. Observed data of liquid flow ($Q_l$) and salt concentration ($Cs$) from 1977 to 1979 and from 1982 to 1991; source: Agence Nationale des Ressources Hydrauliques (ANRH)

![Image](55x795 to 540x832)

Table 2. Comparison of observed and predicted daily salt flow (kg s⁻¹)

| Phase    | $Q_l$ | Average | SD   | CV   | Min  | Max  |
|----------|-------|---------|------|------|------|------|
| Training | observed | 5.11    | 8.87 | 1.74 | 0.052| 80.6 |
|          | predicted | 4.36    | 9.47 | 2.05 | 0.023| 80.6 |
| Validation | observed | 4.39    | 9.47 | 2.05 | 0.023| 80.6 |
|          | predicted | 3.43    | 8.05 | 2.05 | 0.043| 40.3 |
| Testing  | observed | 5.92    | 13.31| 2.24 | 0.038| 66.0 |
|          | predicted | 5.24    | 10.94| 2.09 | 0.024| 50.0 |

Explanations: $Q_l$ = salt flow, SD = standard deviation, CV = coefficient of variation.

Source: own study.

Table 3. Statistical parameters of ANFIS in the training, validation and testing

| Phase    | $RSR$ (%) | $E$ (%) | $R^2$ (%) |
|----------|-----------|---------|-----------|
| Training | 2         | 85.84   | 88.00     |
| Validation | 10      | 82.51   | 78.01     |
| Testing  | 49        | 78.17   | 80.00     |

Explanations: RSR = root of mean square error, $E$ = Nash–Sutcliffe efficiency coefficient, $R^2$ = determination coefficient.

Source: own study.
Fig. 4. Correlation between observed and predicted values of saline flow rates: a) training, b) validation, c) testing; source: own study

Fig. 5. Comparison of the predicted and observed values: a) training, b) validation, c) testing; source: own study
A hybrid model for modelling the salinity of the Tafna River in Algeria

CONCLUSIONS

The key objective of this study was to evaluate the accuracy of hybrid model (ANFIS) in prediction salinity of the Tafna River. Data collected during 13 years by National Water Resources Agency (ANRH) were used to establish a salinity-rating curve and also acted as inputs to the hybrid model. The ANFIS model showed a better performance results in step of daily time, with coefficient of determination ($R^2 = 0.88\%$ for training, $78.01\%$ validation and $80.00\%$ for testing), Nash–Sutcliffe efficiency coefficient ($E = 85.84\%$ for training, $82.51\%$ validation and $78.17\%$ for testing), and the root of the mean squared error ($\text{RSR} = 0.02\%$ for training, $0.10\%$ validation and $0.49\%$ for testing). These encouraging results open up a number of perspectives, where it would be interesting to try to apply hybrid models (neuro-fuzzy) on a larger scale in hydrology and the environment. This approach should be tested on a large scale for the whole country and, if successful, integrated into early warning systems in the event of Water Supply and Sanitation (WSS) water quality degradation.

REFERENCES

ALVISI S., MASCELLANI G., FRANCHINI M., BARDOSY A. 2006. Water level forecasting through fuzzy logic and artificial neural network approaches. Hydrology and Earth System Sciences. Vol. 10. Iss. 1 p. 1–17.

AVCI E. 2008. Comparison of wavelet families for texture classification by using wavelet packet entropy adaptive network based fuzzy inference system. Applied Soft Computing. Vol. 8. Iss. 1 p. 225–231. DOI 10.1016/j.asoc.2007.01.003.

BOUZERIA H., GHENIM A.N., KHANCHOUL K. 2017. Using artificial neural network (ANN) for prediction of sediment loads, application to the Mellah catchment, northeast Algeria. Journal of Water and Land Development. No. 33 p. 47–55. DOI 10.1515/jwld-2017-0018.

BUCKLEY J.J., HAYASHI Y. 1994. Fuzzy neural networks. Fuzzy sets and Systems. Vol. 66 p. 1–13. DOI 10.1016/0165-0114(94)00976-7.

BURAGOHAIN M., MAHANTA C. 2008. A novel approach for ANFIS modeling based on full factorial design. Applied Soft Computing. Vol. 8. Iss. 1 p. 609–625. DOI 10.1016/j.asoc.2007.03.010.

CHANG F.J., CHANG Y.T. 2006. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Advances in Water Resources. Vol. 29 p. 1–10. DOI 10.1016/adwatre.2005.04.015.

CHEN S.H., JAKEMAN A.J., NORTON J.P. 2008. Artificial intelligence techniques: An introduction to their use for modelling environmental systems. Mathematics and Computers in Simulation. Vol. 78. Iss. 2–3 p. 379–400.

CHENNAKESAVA R.A. 2008. Fuzzy logic and neural networks: basic concepts and applications. New Delhi. New Age Publishers. ISBN 8122421822 pp. 257.

FASHI F.H. 2016. Evaluation of adaptive neural-based fuzzy inference system approach for estimating saturated soil water content. Modeling Earth Systems and Environment. Vol. 2. Iss. 4 p. 1–6. DOI 10.1007/s40808-016-0255-y.

FULLER R. 2000. Introduction neuro-fuzzy systems. Ser. Advances in Intelligent and Soft Computing. Vol. 2. Heidelberg. Physica Verl. Springer. Springer Verl. Comp. pp. 289.

JANG J.R.S. 1993. ANFIS. Adaptive-network-based fuzzy inference systems. IEEE Transaction on Systems, Man, and Cybernetics. Vol. 23. No. 1 p. 665–685. DOI 10.1109/8.256541.

JANG J.S.R., SUNI C.T., MIZUTANI E. 1997. Neuro-fuzzy and soft computing: A computational approach to learning and machine intelligence. IEEE Transactions on Automatic Control. Vol. 42. Iss. 10 p. 1482–1484.
Khemissi HOUARI, Tarik HARTANI, Boualem REMINI, Abdelouhab LEFKIR, Leila ABDA, Salim HEDDAM

Hybrydowy model służący modelowaniu zasolenia rzeki Tafna w Algierii

STRESZCZENIE

W pracy badano zdolność systemu wnioskowania rozmytego opartego na adaptacyjnej sieci (ANFIS) do przewidywania zasolenia rzeki Tafna. Do trenowania, oceny i testowania modelu hybrydowego wykorzystano serie pomiarów dobowych przepływów płynu i stężeń soli ze stacji pomiarowej w Pierre du Chat (160801). Dokładność wyników testowano za pomocą: współczynnika determinacji ($R^2$), współczynnika wydajności Nasha–Sutcliffe’a ($E$), pierwiastka średniego błędu kwadratowego ($RSR$) i technik graficznych. Model dał zadowalające wyniki i wykazywał dobrą zgodność między danymi obserwowanymi a przewidywanymi: $R^2$ (88% w przypadku uczenia sieci, 78.01% walidacji i 80.00% testowania), $E$ (85.84% w przypadku uczenia sieci, 82.51% walidacji i 78.17% testowania) i $RSR$ (2% w przypadku uczenia sieci, 10% walidacji i 49% testowania).

Słowa kluczowe: model hybrydowy, przepływ soli, rzeka Tafna, system neuronowo-rozmyty, system wnioskowania rozmytego (ANFIS), zasolenie