Article, Published Version

Stylianoudaki, Christina; Trichakis, Ioannis; Karatzas, George P.

Artificial neural networks for the prediction of groundwater nitrate contamination

Hydrolink

ARTIFICIAL NEURAL NETWORKS FOR THE PREDICTION OF GROUNDWATER NITRATE CONTAMINATION

BY CHRISTINA STYLIANOUDAKI, IOANNIS TRICHAKIS & GEORGE P. KARATZAS

An artificial neural network (ANN) model is proposed for the determination of groundwater nitrate contamination, based on an approach of easily measurable and cost-effective water quality parameters (pH, electrical conductivity, HCO_3^-, Cl^-, Ca^{2+}, Mg^{2+}, Na^+, K^+, SO_4^{2-}). The data used, derived from the chemical analyses of groundwater samples, from wells located in the Kopaidian Plain, Greece. The results of the model described in this article indicate that ANNs could be used as an alternative method for the estimation of groundwater contamination problems.

The rapid increase in population, as well as industrialization and intensification of agricultural activities, have led to significant quantitative and qualitative degradation of groundwater resources worldwide [1]. The situation is expected to be burdened by climate change, which will cause changes in rainfall patterns and an increase in average surface temperature, especially in drought-prone areas [2]. Surface and groundwater contamination due to the presence of nitrates (NO_3^-) is considered as one of the most common environmental problems [3], [4]. The major anthropogenic source of nitrogen in the environment is the application of nitrogen fertilizer [5]. Other anthropogenic sources include industrial wastes, deforestation (leading to conversion to agricultural land) domestic wastewater and septic tanks [6]. According to Greek and EU legislation, nitrate concentration shall not exceed 50 mg /l for nitrates (NO_3^-), or 11 mg /l for nitrate-nitrogen (NO_3^-N) [7]. On a global scale, concentrations of nitrate in groundwater exceed the limits that have been set, and it is estimated that in the last three decades nitrate pollution has increased by 36%. In the eastern Mediterranean and Africa, the situation is even more worrying, as nitrate levels seem to have more than doubled [6].

Nitrate ions have a toxic effect with proven effects on human health and have been statistically associated with various forms of cancer [8], [9], [10]. In addition, increased indices of thyroid diseases have been recorded in areas with high nitrate levels in water supplies [5]. In order to protect public health, sustainable management of groundwater resources is required. However, techniques for detecting and measuring nitrate concentrations in water can be characterized by high cost and high time demand [11], while the portable devices used for this purpose are not of sufficient accuracy. Furthermore, in the various methods used for chemical analysis of water, the detection of nitrates is affected by the presence of other ions, especially Cl^- [12]. Physics-based models for the analysis of groundwater contamination problems have developed significantly in recent years. However, these models require extensive data that are often not available. In many applications, there is a need to develop surrogate, easy to use models that can rapidly analyze groundwater contamination, without the limitations of more complex models. The scope of this study is to describe a model for the easy estimation of nitrate groundwater contamination based on easily measurable and cost effective input parameters.

Artificial neural networks are data driven models that treat the system being studied as a 'black box' [13]. ANNs have the ability to correlate variables whose relationship is not known or is very complex [14], [15], without the use of physical data, such as porosity or hydraulic conductivity [16].

> **"It is estimated that in the last three decades nitrate pollution has increased by 36%"**

ANNs have widely found applications in hydrology and have been successfully used in groundwater quality modeling [16],[17]. Several studies have presented ANNs for the estimation of the water level by using water budget variables as input parameters [15], [18], [19], [20], [21]. Regarding nitrate contamination, ANN models using water quality parameters or/and water budget variables as input parameters, have been proposed [22], [23], [24]. Comprehensive reviews over the applications of ANNs in hydrology can be found in [17], [25] and [26].

An ANN consists of a number of fully connected processors called neurons, which accept, analyze, and exchange information over a network of weighted connections [26]. The feed-forward neural network was the first type of ANN, where information moves in a forward direction. The information is processed at different layers, divided in three categories: input, output and hidden layers. Each input x_i presented in a neuron, is weighted by a synaptic weight w_i and the results are summed. The sum is introduced in an activation function; in the case it exceeds a certain threshold value,. The basic function of an ANN is the training process, which is performed by a learning rule that modifies the weights of the connections in order to minimize the difference between the calculated output of the network and the desired output (real value) [25]. The efficiency of the model is evaluated by its generalization ability, i.e. its ability to give the correct output even for examples not included in the training set. More information regarding ANNs and their operation is presented by [16] and [27].

In this study, a feed-forward neural network consisting of three layers was developed, using MATLAB R2010 software. A Levenberg-Marquardt regularization algorithm was

Table 1: Maximum, minimum and mean values of the input and output parameters used in the model

	NO_3^- (mg/l)	pH	COND	Ca^{2+} (mg/l)	Mg^2 (mg/l)	Na^+ (mg/l)	K^+ (mg/l)	HCO_3^- (mg/l)	Cl^- (mg/l)	SO_4^{2-} (mg/l)
Min	5	6.4	234	2.4	4.4	2.3	0.4	49	5.3	10
Max	167	9.1	2750	236	121.1	303.5	13	585	560.2	148.9
Mean	27.31	7.66	791.03	66.66	38.40	36.62	1.94	324.10	58.01	38.04

Figure 1. Model's results - R coefficient

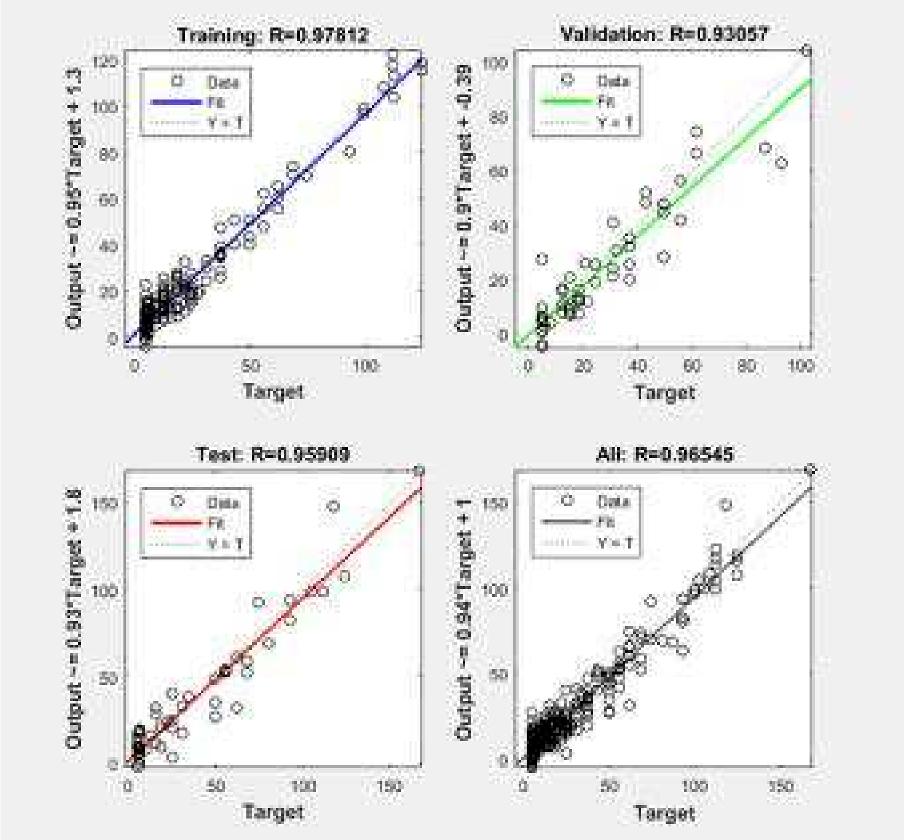


Figure 2. Observed versus simulated values

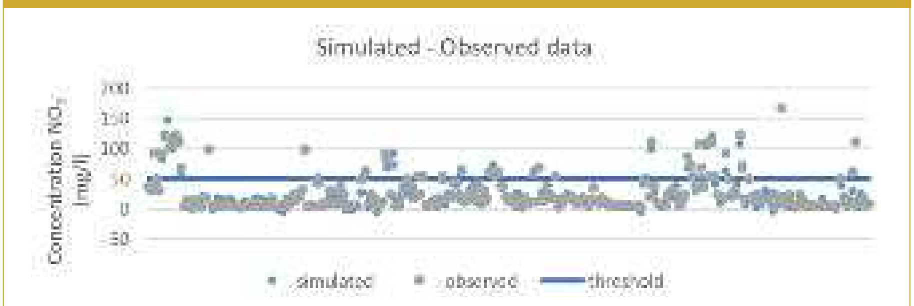


Table 3. Calculated statistical indicators

Index	All	Test	Validation
RMSE (mg/l)	7.75	10.94	9.14
MAE (mg/l)	5.70	8.07	6.90
Bias (mg/l)	-0.65	-0.77	-3.07
NSE	0.9878	0.9193	0.9969
St. Deviation	29.70	38.90	23.64

employed for the training procedure. The ANN's architecture was determined through a trial and error procedure, based on the correlation coefficient (R) between the real data and the simulated values by the model. The best architecture came out to be that of one hidden layer with 10 nodes, a sigmoid function in the first layer and a linear in the output layer as activation functions. Of the available data, 60% was used in the training process, 20% in the testing process and the remaining 20% was used for the evaluation of the model's performance (generalization ability).
For the analysis of the results, four error indicators were calculated: the Root Mean Square Error RMSE), the Mean Absolute Error (MAE), the bias (mean error) and the Nash-Sutcliffe Model Efficiency (NSME).

The data used for the ANN's training and validation, were derived from a set of 263 measurements of typical water quality parameters, obtained from sampling in the Kopaidian Plain. The area has been designated as a vulnerable zone with respect to nitrogen pollution from agricultural runoff, according to the requirements of the European Union Directive 91/676/EEC [28], due to the intensive agricultural, livestock and industrial activities that take place in it. The input parameters to the model were pH, electrical conductivity, bicarbonate (HCO_3^-) and Cl^-, Ca^{2+}, Mg^{2+}, Na^+, K^+, SO_4^{2-}.
Table 1 presents the maximum, minimum and the mean value of the NO_3^- concentrations and of all the parameters used as inputs to the model.

The simulation results are presented in Figure 1. The Pearson coefficient (R index) is shown on top of each chart, respectively, for the training data in the top left part of the figure, for the validation set in the top right, for the testing data set in the bottom left and for the full data set in the bottom right. In the plot, the simulated values (output – vertical axis) are plotted against the observed values (target – horizontal axis), and their best-fit equation, which describes the solid line in the graph, is shown on the vertical axis title. As shown in the charts of Figure 1 the ANN has delivered very good results.

A high correlation is observed, between the simulated and actual values, for every data set, with a correlation index for the full data set of 0.96545. In the test set, R is equal to 0.95909 and in the validation set, R= 0.93057. Considering that these data have not been used

in the training process, R values signify a very good generalization ability of the model.
In Figure 2, the simulated values by the model, together with the real data are presented, along with the threshold value of 50 mg/l for nitrate concentration. As already expected from the R value, Figure 2 confirms that the simulated values are very close to their observed counterparts.

The calculated indicators for an additional evaluation of the model's performance are shown in Table 3.

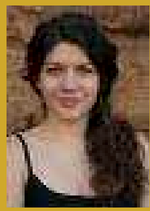
For the full data set, the NSE is equal to 0.9878, for the test set, $NSE_{test} = 0.9193$ and for the validation set, $NSE_{vald}=0.9969$. As shown by the indicators, the model has produced remarkably satisfactory results. NSE values in the range ($0.75 < NSE < 1$), indicate very good performance of the model being assessed [29]. Therefore, taking into account the NSE index, the simulation can be characterized successful. Moreover, according to [30], RMSE and MAE values less than half of the standard deviation of the observed data are considered low. In addition, the small difference between RMSE and MAE (7.754874 mg/l - 5.702638 mg/l) indicates the absence of extreme errors. Lastly, it is worth pointing out that according to the Bias index, the model tends to underestimate the observed values but not by much. The calculated indices suggest that the model is highly accurate.

Conclusions

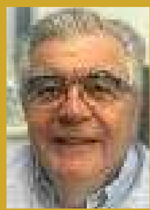
In hydrological applications, there is a need to develop simple models that can capture the main relationships between parameters without the need to develop complex physics-based models that are difficult to solve. ANNs have the essential advantage that they can track the hidden relationship between variables – without the need to assume linearity- and so, available data that are not usually used in conventional techniques can be exploited. The results of the study described in this article demonstrate that ANNs are a potentially powerful modelling method, more economical and less time consuming. ■

References

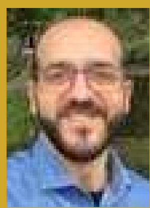
[1] Kazakis, Nerantzis & Voudouris, K. (2015). Groundwater vulnerability and pollution risk assessment of porous aquifers to nitrate: Modifying the DRASTIC method using quantitative parameters. Journal of Hydrology. 525. 13-25.
[2] IPCC, 2014: Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, 151 pp.
[3] Spalding R.F., Exner M.E. (1993) Occurrence of nitrate in groundwater—a review. Journal of Environmental Quality 22. 392–402.
[4] Rivett, Michael & R Buss, Stephen & Morgan, Philip & Smith, Jonathan & D Bemment, Chrystina. (2008). Nitrate Attenuation in Groundwater: A Review of Biogeochemical Controlling Processes. Water research. 42. 4215-32.
[5] Ward, Mary & Jones, Rena & Brender, Jean & de Kok, Theo & J Weyer, Peter & T Nolan, Bernard & M Villanueva, Cristina & Breda, Simone. (2018). Drinking Water Nitrate and Human Health: An Updated Review. International Journal of Environmental Research and Public Health. 15. 1557.
[6] Shukla, Saurabh & Saxena, Abhishek. (2018). Global Status of Nitrate Contamination in Groundwater: Its Occurrence, Health Impacts, and Mitigation Measures. *Handbook of Environmental Materials Management*. Springer, Cham.
[7] WHO (World Health Organization) (2011) Guidelines for drinking-water quality, 4th edn. World Health Organization, Geneva. Available: Accessed 10 Apr 201.
[8] Hunder.H. Comly. (1945). Cyanosis in infants caused by nitrates in well water. JAMA. 129. 112-116
[9] Morales-Suárez-Varela, Maria & Llopis-Gonzalez, Augustin & L. Tejerizo-Perez, María. (1995). Impact of nitrates in drinking water on cancer mortality in Valencia, Spain. European Journal of Epidemiology. 11. 15-21.
[10] Palli, Domenico & Saieva, Calogero & Coppi, Claudio & Del Giudice, Giuseppe & Magagnotti, Cinzia & Nesi, Gabriella & Orsi, Federica & Airoldi, Luisa. (2001). O 6 -Alkylguanines, Dietary N-Nitroso Compounds, and Their Precursors in Gastric Cancer. Nutrition and cancer. 39. 42-9.
[11] Azmi, Aizat & Amsyar Azman, Ahmad & Ibrahim, Sallehuddin & Md Yunus, Mohd Amri. (2017). Techniques in advancing the capabilities of various nitrate detection methods: A review. International Journal on Smart Sensing and Intelligent Systems. 10. 223-261.
[12] Alahi, Md & Mukhopadhyay, S.C. (2018). Detection methods of Nitrate in water: A Review. Sensors and Actuators A: Physical. 280. 210-221. https://doi.org/10.1016/j.sna.2018.07.026
[13] Tapoglou, Evdokia & Karatzas, George & Trichakis, Ioannis & Varouchakis, Emmanouil. (2014b). A spatio-temporal hybrid neural network-Kriging model for groundwater level simulation. Journal of Hydrology. 519. 3193-3203.
[14] Almahallawi, Khamis & Jacky, Mania & HANI, Azzedine & Shahrour, Isam. (2011). Using of neural networks for the prediction of nitrate groundwater contamination in rural and agricultural areas. Environmental Earth Sciences. 65. 917-928.
[15] Ghose, Dillip & Das, Umesh & Roy, Parthajit. (2018). Modeling response of runoff and evapotranspiration for predicting water table depth in arid region using dynamic recurrent neural network. Groundwater for Sustainable Development. 6. 263-269, https://doi.org/10.1016/j.gsd.2018.01.007
[16] ASCE, 2000a. Artificial neural networks in hydrology. I: Preliminary concepts. Journal of Hydrologic Engineering. 5. 115-123.
[17] ASCE, 2000b. Artificial Neural Networks in hydrology. II: Hydrologic applications. Journal of Hydrologic Engineering. 5. 124-137.
[18] Nayak, P., Rao, Y. & Sudheer, K. (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resources. Management. 20. 77–90.
[19] Trichakis, I., Nikolos, I., & Karatzas, G. (2009). Optimal selection of artificial neural network parameters for the prediction of a karstic aquifer's response. Hydrological Processes. 23. 2956-2969. https://doi.org/10.1002/hyp.7410
[20] Trichakis, I.C., Nikolos, I.K. & Karatzas, G.P. (2011). Artificial Neural Network (ANN) Based Modeling for Karstic Groundwater Level Simulation. Water resources management. 25. 1143-1152. https://doi.org/10.1007/s11269-010-9628-6
[21] Tapoglou, Evdokia & Trichakis, Ioannis & Dokou, Zoi & Nikolos, Ioannis & Karatzas, George. (2014a). Groundwater-level forecasting under climate change scenarios using an artificial neural network trained with particle swarm optimization. Hydrological Sciences Journal. 59. 1225-1239. https://doi.org/10.1080/02626667.2013.838005
[22] Yesilnacar M.I., Sahinkaya E., Naz M. and. Ozkaya B. (2008). Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. Environmental Geology. 56. 19–25. DOI: 10.1007/s00254-007-1136-5
[23] Ehteshami, M., Farahani, N.D. & Tavassoli, S. (2016). Simulation of nitrate contamination in groundwater using artificial neural networks. Modeling Earth Systems and Environment. 2: 28. https://doi.org/10.1007/s40808-016-0080-3
[24] Wagh, Vasant & Panaskar, Dipak & Muley, Aniket & Mukate, Shrikant & Gaikwad, Satyajit. (2018). Neural Network Modelling for Nitrate Concentration in groundwater of Kadava River basin, Nashik, Maharashtra, India. Groundwater for Sustainable Development. DOI: 10.1016/j.gsd.2017.12.012
[25] Tanty, Rakesh & S. Desmukh, Tanweer. (2015). Application of Artificial Neural Network in Hydrology- A Review. International Journal of Engineering Research. 4. DOI: 10.17577/IJERTV4IS060247
[26] Oyebode, O, Stretch, D. (2019). Neural network modeling of hydrological systems: A review of implementation techniques. *Natural Resource Modeling*. 32:e12189. https://doi.org/10.1111/nrm.12189
[27] Haykin, S.S., (1999). Neural networks: a comprehensive foundation. Prentice Hall
[28] Ministry of Environment and Energy, Nitrate Pollution, http://www.ypeka.gr/Default.aspx?tabid=250&locale=en-US&language=el-GR



Christina Stylianoudaki is an Environmental Engineer, MSc and Ph.D. student in the School of Environmental Engineering at the Technical University of Crete, in the field of water resources management and climate change. Currently she works in the Geoenvironmental Engineering Laboratory, as a researcher on the development of a flood risk management system. Her research interests include hydrological processes modelling, environmental systems optimization, simulation of flood events and assessment of climate change's impact on water resources.



Professor **George P. Karatzas** received his Ph.D. in 1992 from the Department of Civil Engineering of Rutgers University, USA in a combined Ph.D. with Princeton University, USA. He is a Professor in the School of Environmental Engineering at the Technical University of Crete, Greece since 2006. His area of research includes numerical simulation in porous media, groundwater management, subsurface remediation techniques, global optimization techniques, surface water modeling and floods. He has been principle investigator and co-investigator in several national and international research projects related to groundwater flow and contaminant transport, groundwater management, surface hydrology and floods.



Ioannis Trichakis studied Environmental Engineering at the Technical University of Crete (1999-2004) and did postgraduate studies at the National Technical University of Athens (2004-06). He received his PhD from the University of Crete (2006-2011) where he worked on research projects on water resources and flood management. He has worked as a researcher at the European Commission's Joint Research Center on water resources management programs in European and developing countries and on simulating groundwater and surface water systems. He has extensive experience in technical support for research programs, writing progress reports and coordinating a large number of partners from different countries.

[29] Moriasi, Daniel & Gitau, Margaret & Pai, Naresh & Daggupati, Prasad. (2015). Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. Transactions of the ASABE (American Society of Agricultural and Biological Engineers). 58. 1763-1785. DOI: 10.13031/trans.58.10715
[30] Singh, J., H. V. Knapp, and M. Demissie. 2004. Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004-08. Champaign, Ill.: Illinois State Water Survey. DOI:10.1111/j.1752-1688.2005.tb03740.x