XVIII International Conference on Water Resources
CMWR 2010
J. Carrera (Ed)

SENSITIVITY ANALYSIS ON AN ADAPTATIVE NEURO FUZZY INFERENCE SYSTEM (ANFIS) FOR HYDRAULIC HEAD INTERPOLATION: ORGEVAL EXPERIMENTAL SITE/FRANCE

B. Kurtulus[*,†], N. Flipo[†] and P. Goblet[†]

[†]Geoscience Department, MINES ParisTech
35, rue Saint-Honoré, 77305 Fontainebleau, France
e-mail: Nicolas.Flipo@mines-paristech.fr, web page: http://www.geosciences.ensmp.fr

[*]Mugla University Geological Engineering Department
48000 Kotekli Mugla
e-mail: bkurtulus@mu.edu.tr, web page: http://geoe.mu.edu.tr

Key words: Sensitivity, ANFIS, Kriging, Hydraulic Head, Orgeval

1 INTRODUCTION

Hydraulic head mapping is useful in many environmental applications. For instance, estimates of the hydraulic head distribution are frequently used to determine the capture zone of pumping wells. Another use of the hydraulic head distribution is also to initialize distributed models, which are nowadays critical tools for managing water resources at the basin scale [5, 4, 11].

Understanding the temporal and spatial variations of the depth to groundwater is a prerequisite to achieve sustainable water use in basin. Point measurements of watertable levels are available, but what is needed are groundwater surfaces based on these measurements. A technique often used in earth sciences and especially in hydrogeology is kriging with a digital elevation model (DEM) of the ground surface as an external drift which seems to be a satisfactory hhydraulic head interpolation method [3, 10, 16].

Over recent years hydrologists have started, to use soft computing and especially Adaptive neuro-fuzzy inference system (ANFIS). ANFIS has been successfully used to interpolate hydraulic head distribution in aquifers [8, 9, 10, 12] albiet really few studies focus on ANFIS uncertainty analysis. The few sensitivity analysis concerns data selection in termes of input variables or of dataset splitting amoung training, validation and test subsets [13, 17]. Only one study analysed the ANFIS structure itself and pointed out the crucial step of membership functions shaping [20].

As understanding uncertainties due to methods is of primary importance [2, 14], this paper aims at better characterizing ANFIS uncertainties in hydrogeology. [10] have show the ability of ANFIS to map hydraulic head distribution at the Avenelles watershed scale (an experimental sub bassin of the Seine Basin, France) for high water regime using three inputs (cartesian coordinates and elevation of the soil). In this paper different ANFIS models with different membership functions (gaussian, bell shaped and triangular) are tested. After selecting the best model, the interpolation robustness of ANFIS on a 50*50m grid is assessed by slightly perturbing either the coordinates of the center of the interpolation grid and the elevation of the soil.

2 EXPERIMENTAL SITE AND DATA

With an area of 104 km^2, the Orgeval experimental basin (Fig. 1) is located 70 km east from Paris [1, 4].The Orgeval basin is influenced by the aquifer system, which is composed of two main geological formations: the Oligocene (see Rupelian limestone, Fig. 1) and the Eocene (from Priabonian to Ypresian claystones, Fig. 1). These two aquifer units are separated by a clayey aquitard. Most of the basin is covered with table-land loess about 2-5m in thickness. These unconsolidated deposits are essentially composed of sand and loam lenses of low permeability but they seem to be more or less connected to the Rupelian limestone. The basin is relatively flat with slopes increasing near the small valley at the river mouth (80 % of the territory spans between 130 and 170 m above mean sea level). In this work we will focus on hydraulic head distribution in the eastern part of the basin covering the Avenelles watershed (Fig. 1).

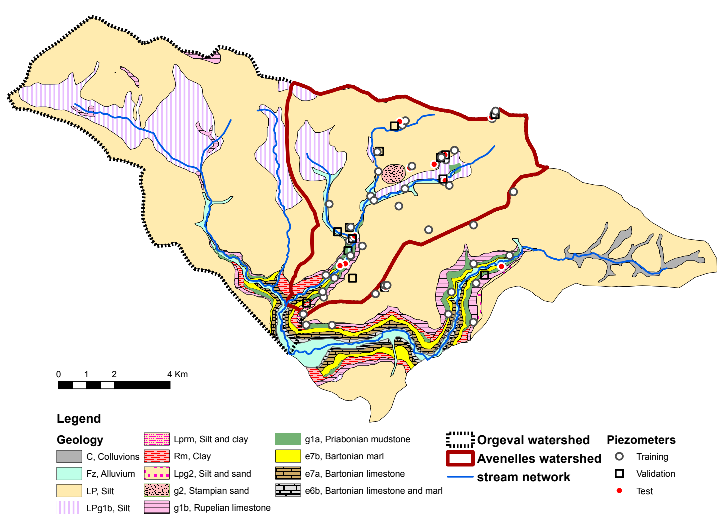

Figure 1: Geological Map of the Orgeval watershed, location of wells and springs divided into training, validation and testing sets

The dataset is composed of three different types of data (Fig. 1). The first one consists of three piezometers, the second one of water levels measured in wells. The 41 wells were sampled on april 16, 2009 during a snapshot campaign. Our goal was to determine the hydraulic head distribution of the subsurface aquifer unit - silt connected to the rupelian limestone. Due to the complex geometry of the aquifer system at the outlet of the Avenelles basin and in the south-eastern part of the area of interest (Fig. 1), we needed to complete the piezometers and wells dataset in this part of the domain of interest. To do so we used a DEM (100 $\times$ 100 m) of the top of the Priabonian mudstone. The elevation of the limit between Priabonian mudstone and rupelian limestone was then implemented inside the datasets as springs. Finally the overall dataset is composed of 70 hydraulic heads measurements.

3 ANFIS

3.1 Theoretical reminders

Adaptive neuro-fuzzy inference system (ANFIS) [6, 7, 15, 18] is a modelling technique based on fuzzy sets [19], which assumes that input and output data are ill-defined with uncertainty that can not be exactly assess with probability theory based on a two-valued logic. A fuzzy set is a set of elements with an imprecise (vague) boundary [15]. A fuzzy set does not have a crisp boundary. That is, the transition from "belonging to the set" to "not belonging to the set" is gradual and is characterized by membership functions (MF). A fuzzy set $A(x)$ is then represented by a pair of two things - the first one is the constituent elements x and their associated membership values $\mu_A(x)$ (that is their degree of belongingness):

$$A(x) = \{(x, \mu_A(x)), x \in X\} \tag{1}$$

Where X is the Universal set consisting of all possible elements. The membership function μ_A ranges between 0 and 1. If the value of the membership function is restricted to either 0 and 1, the fuzzy set is then reduced to classical crisp set with a known boundary. The fuzziness does not come from the randomness of the constituent members of the sets, but from the uncertain and imprecise nature of the abstract thoughts and concepts [7].

In ANFIS the relationship between input and output are expressed in the form of If-Then rules. ANFIS used for the present work is based on Sugeno fuzzy model [18] which formalizes a systematic approach to generating fuzzy rules from an input-output dataset. A typical fuzzy rule in a Sugeno fuzzy model has the format: If $x \in A$ and $y \in B$ then $z = f(x, y)$, where A and B are fuzzy sets in the antecedent and $f(x, y)$ is a crisp function in the consequent. Usually f is a polynomial function.

ANFIS uses a hybrid learning algorithm that combines the back-propagation gradient descent and least squares methods to create a fuzzy inference system whose membership functions are iteratively adjusted according to a given set of input and output data [6]. For each iteration, the back propagation method involves minimization of an objective function using the steepest gradient descent approach in which the network weights and biases are adjusted by moving a small step in the direction of negative gradient. The

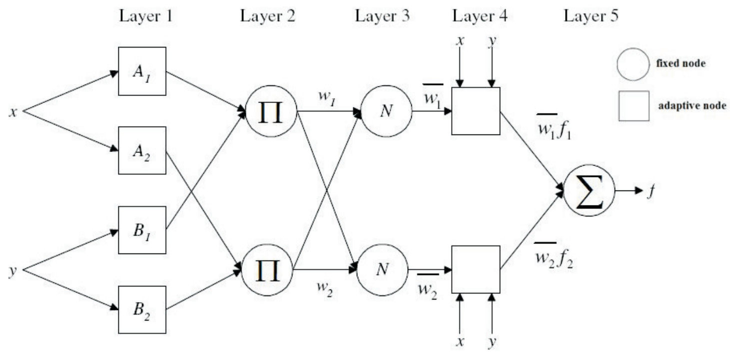

Figure 2: ANFIS architecture for two inputs x, y. Layer 1: generates membership grades. Layer 2: Fuzzy rules. Layer 3: calculates weights of rules named firing strengths. Layer 4: product of the normalized firing strengths. Layer 5: fuzzy results transformed into a traditional output by summation

iterations are repeated untill a convergence criteria or a specified number of iterations is achieved. It has the advantage of allowing the extraction of fuzzy rules from numerical data and adaptively constructs a rule base. The architecture of the ANFIS systems is composed of five layers (Fig. 2). Each layer consists in different nodes described by node function. The output signal from nodes of a layer is the input signal of the next layer.

3.2 Implementation of anfis

Input data are X,Y coordinates of piezometers plus the elevation of the ground. Hydraulic head is the ANFIS output. Input data are pre-processed to obtain input vectors for which coordinates are in the same range of variations. ANFIS is run with variables that are reduced and centered [9, 10]. The selection of appropriate input parameters is done iteratively with a square process splitting [9, 10] to obtain the most similar training, validation and test sets in terms of high and low values as well as concerning the statistical distribution. Finally 45 points are assigned to the training subset (64 %), 14 points to the validation subset (20 %), and the remaining 11 points to the test subset (16 %) (see Fig. 1).

Before using the model to interpolate unknown outputs (hydraulics head), its actual predictive performance must be tested by comparing outputs estimated by the calibrated models with known outputs. At each phase (training, validation and test), the ANFIS performance is measured by the determination of the coefficient of goodness-of-fit R^2, and the root mean square error (RMSE).

4 ANFIS SENSITIVITY TO MF

To built an ANFIS model, the first step is to select the type and the number of membership functions (MF) for the first layer of the model (Fig. 2). In this work we have test the sensitivity of the ANFIS model to gaussian (Gaussmf), bell shaped (Gbellmf) and triangular (Trimf) MF (eq. 2, 3, 4).

$$gaussmf : f(x,\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}} \tag{2}$$

$$gbellmf : f(x,a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{3}$$

$$trimf : f(x,a,b,c) = max\left(min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right), with\ a < b < c \tag{4}$$

Once the MF selected, each ANFIS model is run with 2, 3, 4 and 5 MF for each node of input data. Table 1 summarizes each ANFIS model performance. It indicates that ANFIS is very sensitive to the type of MF and the number. Increasing the number of MF per nodes does not increase model performance and usually leads to the contrary (Table 1). The test performance is used to select the five best ANFIS models, which are : Trimf with 2 and 3 MF, Gbellmf with 2 and 2 MF and Gaussmf with 2 MF.

		RMSE			R^2		
MF		Trimf	Gbellmf	Gaussmf	Trimf	Gbellmf	Gaussmf
222	it	95	3000	200	95	3000	200
	Tr	0,65	0,25	0,39	1,00	1,00	1,00
	Val	1,64	3,37	3,13	0,99	0,96	0,97
	Test	2,77	2,02	2,38	0,98	0,99	0,99
333	it	120	750	40	120	750	40
	Tr	0,01	0,00	0,01	1,00	1,00	1,00
	Val	2,15	2,50	2,45	0,98	0,98	0,98
	Test	2,49	2,19	3,08	0,99	0,99	0,98
444	it	500	200	1000	500	200	1000
	Tr	0,00	0,00	0,00	1,00	1,00	1,00
	Val	2,41	4,07	2,06	0,98	0,95	0,99
	Test	3,90	3,70	5,22	0,96	0,97	0,94
555	it	250	20	23	250	20	23
	Tr	0,00	0,00	0,00	1,00	1,00	1,00
	Val	2,04	2,45	2,15	0,99	0,98	0,98
	Test	3,79	4,50	3,90	0,97	0,95	0,96

Table 1: RMSE ans R^2 for different number of MF. it: number of iteration, Tr: Training, Val: Validation, T: Test data sets

The five best models have good test performances: RMSE between 2 and 3 m, and R^2 higher than 0.98 (Tab. 1). In order to select the best model, each one was used to interpolate the hydraulic head distribution on a 50m×50m grid. For each cell the hydraulic head is compared to the elevation of the soil. Table 2 summarizes the percentage of cells where the hydraulic head is above the soil elevation, which is a wrong estimate for the Avenelles basin where no artesian springs are observed. Finally even if ANFIS model with two triangular MF does not perform best in terms of statistical criteria, it interpolates the best hydraulic head distribution (Fig. 3).

	gbellmf2	gbellmf3	gaussmf2	Trimf2	Trimf3
0 m	2,69	14,38	0,90	0,00	2,99
2 m	1,76	10,01	0,72	0,00	1,03
5 m	1,02	6,09	0,55	0,00	0,08

Table 2: Percentage of cells where hydraulic head is N meters above the elevation of the soil

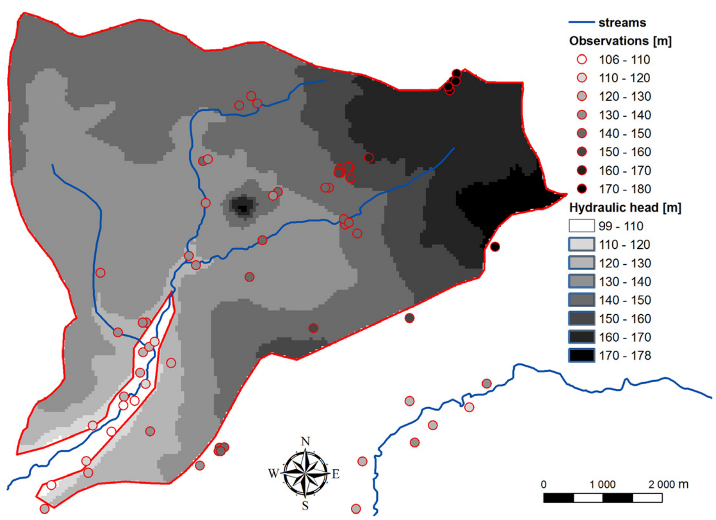

Figure 3: Interpolated hydraulic head distribution with triangular MF. Dots are observations.

5 SENSITIVITY ANALYSIS OF ANFIS TO INPUT DATA

5.1 Sensitivity to soil elevation

Once the best ANFIS model is selected a sensitivity analysis was made on input data. First we assess ANFIS sensitivity to soil elevation. To do so, an error of -2m, -1m, 1m and 2m was added to the soil elevation of each cell of the 50m×50m grid. The ANFIS model with two triangular MF per node was then used to assess the resulting hydraulic head distribution. Finally the resulting hydraulic head was compared to the reference one (Fig.

3) by calculating at each cell center the difference between the two estimates. Table 3 indicates the mean of the difference and the standard deviation of each simulation.

Systematic error	+ 2m	+ 1m	- 1m	- 2m
Min	0,694	0,347	-1,134	-2,258
Max	2,584	1,132	-0,347	-0,694
Mean	1,714	0,853	-0,848	-1,691
Std	0,173	0,087	0,092	0,189

Table 3: Sensistivity Statistics to soil elevation

ANFIS is very sensitive to soil elevation. First the sign of the error is propagated by ANFIS. Indeed an uderestimated soil elevation leads to an estimation of the hydraulic head distribution lower than the reference one. Contrarly an overestimated soil elevation leads to a higher estimate of the hydraulic head distribution. It is interesting to notice that the mean error on hydraulic head is in the same order of magnitude than the error on soil elevation (Tab 3).

5.2 Sensitivity to XY coordinates

The same method was used to assess ANFIS sensitivity to XY coordinates by propagating an error of -10m, -5m, -1m, +1m, +5m, +10m. The analysis reveals that ANFIS model is not sensitive at to errors of this range on XY coordinates. Indeed the absolute maximum difference is 10 and 20 cm for an error on XY coordinates of 5m and 10m, respectively.

6 CONCLUSION

This work aimed at analysing ANFIS method to interpolate hydraulic head distribution using XY coordinates and soil elevation as input data. ANFIS is very sensitive to the type and number of Membership functions. When using this interpolator it is thus of primary importance to first identify a satisfactory type of MF and then to use the right number of those per input nodes. The sensitivity analysis revealed that ANFIS model is mostly sensitive to soil elevation. XY coordinates are forcing data at larger scale than the soil elevation.

7 ACKNOWLEDGMENT

This work was funded by the ANR Carnot MINES “Neuro N'Eaudyssée”, the PIREN Seine research program, and the FIRE FR3020 (Federation Ile de France de Recherche en Environnement). It is also a contribution to the GIS Oracle that maintains the experimental basin of the Orgeval. We kindly thank the BRGM for providing the DEM of the top of the aquifer system.

References

[1] F. Anctil, M. Filion, and J. Tournebize. A neural network experiment on the simulation of daily nitrate-nitrogen and suspended sediment fluxes from a small agricultural catchment. *Ecol. Model.*, 220:879–887, 2009.

[2] K. Beven. Changing ideas in hydrology. The case of physically-based model. *J. of Hydrology*, 105:157–172, 1989.

[3] A. J. Desbarats, C. E. Logan, M. J. Hinton, and D. R. Sharpe. On the kriging of water table elevations using collateral information from a digital elevation model. *Journal of Hydrology*, 255(1-4):25–38, 2002.

[4] N. Flipo, S. Even, M. Poulin, S. Théry, and E. Ledoux. Modelling nitrate fluxes at the catchment scale using the integrated tool CaWaQS. *Sci Total Environ*, 375:69–79, 2007.

[5] N. Flipo, N. Jeannée, M. Poulin, S. Even, and E. Ledoux. Assessment of nitrate pollution in the Grand Morin aquifers (France): combined use of geostatistics and physically-based modeling. *Environ Pollut*, 146(1):241–256, 2007.

[6] J. Jang. ANFIS adaptive-network-based fuzzy inference systems. *IEEE Trans. Systems, Man Cybern*, 23(3):665–685, 1993.

[7] J. Jang. Neuro-fuzzy modeling and control. *Proceedings of the IEEE 83*, 3:378–406, 1995.

[8] M. Kholghi and S. M. Hosseini. Comparison of groundwater level estimation using neuro-fuzzy and ordinary kriging. *Environ Model Assess*, 2008.

[9] B. Kurtulus, N. Flipo, G. Vilain, J. Tournebize, G. Tallec, and P. Goblet. Comparison of ANFIS and ordinary kriging to assess hydraulic head distribution: the Orgeval case study. In *Proceedings of ICNC, Madeira, Oct 5-7, 2009*, 2009.

[10] B. Kurtulus, N. Flipo, G. Vilain, J. Tournebize, G. Tallec, P. Goblet, and A. Johannet. Assessing the spatial hydraulic head distribution in surface aquifer using different interpolation methods. *J. of Hydrology*, submitted, 2009.

[11] E. Ledoux, E. Gomez, J. Monget, C. Viavattene, P. Viennot, A. Ducharne, M. Benoit, C. Mignolet, C. Schott, and B. Mary. Agriculture and groundwater nitrate contamination in the Seine basin. The STICS-MODCOU modelling chain. *Sci Total Environ*, 375:33–47, 2007.

[12] G.-F. Lin and L.-H. Chen. A spatial interpolation method based on radial basis function networks incorporating a semivariogram model. *Journal of Hydrology*, 288(3-4):288–298, 2004.

[13] R. Marcé, M. Comerma, J. García, and J. Armengo. A neuro-fuzzy modeling tool to estimate fluvial nutrient loads in watersheds under time-varying human impact. *Limnol. Oceanogr.: Methods*, 2:342–355, 2004.

[14] E. Polus, N. Flipo, C. de Fouquet, and M. Poulin. Geostatistics for assessing the efficiency of distributed physically-based water quality model. application to nitrates in the seine river. *Hydrological Processes*, submitted, 2009.

[15] D. Pratihar. *Soft Computing.* Alpha Science International Ltd, 2008.

[16] M. Rivest, D. Marcotte, and P. Pasquier. Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information. *Journal of Hydrology*, 361(3-4):349–361, 2008.

[17] G. Sahoo, C. Ray, and H. Wade. Pesticide prediction in ground water in north carolina domestic wells using artificial neural networks. *Ecol. Model.*, 183:29–46, 2005.

[18] T. Takagi and M. M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Trans. Systems Man and Cybernetics*, 15(1):116–132, 1985.

[19] L. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.

[20] C. Zanchettin, F. Minku, and T. Ludernir. Design of experiments in neuro-fuzzy systems. In IEEE, editor, *Proceedings of the Fifth International Conference on Hybrid Intelligent Systems (HIS05)*, 2005.