DISPERSION COEFFICIENT ASSESSMENT BY MEANS OF DIFFERENT NEURAL NETWORKS

Adam PIOTROWSKI and Jarosław J. NAPIÓRKOWSKI

Institute of Geophysics, Polish Academy of Sciences
ul. Księcia Janusza 64, 02-452 Warszwa, Poland
e-mail: adampp@igf.edu.pl , jnn@igf.edu.pl

Abstract

Accurate application of the longitudinal dispersion model requires that specially designed experimental studies are performed in the river reach under consideration. Such studies are usually very expensive, so in order to quantify the longitudinal dispersion coefficient, numerous empirical formulae, based on hydraulic and morphometric characteristics, have been proposed as an alternative approach. The paper presents the application of three artificial neural networks as a parameter estimation technique. The networks were trained for a special arrangement of input nodes, namely: channel depth, channel width, cross-sectionally averaged water velocity, shear velocity and sinuosity index.

Keywords
Artificial neural networks, longitudinal dispersion, pollutant transport, rivers

1   Introduction

This paper discusses modeling of the transport of pollution in rivers by means of a longitudinal dispersion model. For one-dimensional case, after a substance has become fully mixed across the depth and width of a river, transport of the passive solutes in the main stream may be described by advection-dispersion equation (Fickian model):

$$\frac{\partial C}{\partial t} + u\frac{\partial C}{\partial x} = \frac{1}{A}\frac{\partial}{\partial x}\left(DA\frac{\partial C}{\partial x}\right) \tag{1}$$

where: x – longitudinal direction, t-time, C – cross-sectional averaged solute concentration, u-cross-sectional averaged velocity of water, D – longitudinal dispersion coefficient, A – cross-sectional area of the channel.

The solution domain of eq.(1) is the plane Oxt constrained by inequalities $0 \le x \le L$ and $t \ge 0$, where L is the length of the modelled channel reach. The model equations are complemented by the following:

- initial conditions:

$$C(x, t = 0) = C_p(x) \tag{2}$$

- and boundary conditions:

$$C(x = 0, t) = C_0(t),... \quad \left.\frac{\partial C}{\partial x}\right|_{x=L} = 0 \tag{3}$$

where C_p is the initial distributions of solute concentration along the channel reach in the main stream and C_0 describes the inflow of admixture at the initial cross-section.

To treat the Fickian model as a predictive tool, one needs to know the way to relate the usually unknown longitudinal dispersion coefficient to basic characteristics of the natural stream under consideration. Empirical studies are usually very expensive, so in order to quantify longitudinal dispersion coefficient, various researchers proposed numerous empirical

formulae based on hydraulic and morphometric characteristics, as an alternative approach. A discussion is going on which expression is most useful. There are numerous publications dealing with different forms and methods of the evaluation of dispersion coefficients dependent on various hydraulic conditions (e.g. [4], [8], [14], [6], [2], [3]). Since all these formulae have simple forms which have been designed to fit the data obtained from specific type of rivers and do not take into account such an important characteristic like bed sinuosity, they fail to assess so called longitudinal dispersion coefficient in many actual cases.

The paper presents successful application of the so-called "intelligent data analysis", namely three types of Artificial Neural Networks: Multi Layer Perceptron Neural Network (MLP) (the most popular one in hydrological sciences), Fuzzy Neural Network (FNN) and Radial Basis Function Neural Network (RBF).

Following a well established practice in the application of ANNs to work on combinations of the channel measurements, all networks were trained for specific arrangements of: mean channel depth H, mean channel width B, cross-sectional averaged water velocity U(m/s) and shear velocity U^* (m/s), namely for nodes described as $I_1 = 3UB$, $I_2 = U/U^*$, $I_3 = B/H$. Shear velocity U^* is expressed as

$$U^* = \sqrt{gSH} \qquad\qquad (4)$$

where $S(-)$ is a channel slope, and g is acceleration of gravity. Moreover, the ratio of the length of the main riverbed to the length of the valley, i.e. sinuosity index S was considered as an additional input node $I_4 = S$. Please note, that $3UW$ is assumed to be a very rough assessment of dispersion coefficient for straight rivers [3]. Constant '3' above was not omitted as it does not cause any impact on neural network performance.

Database is collected from 81 experiments performed in USA [3] and Moldova [14]. This data set includes a wide variety of cases, from one of the world biggest rivers (Mississippi in Louisiana) to small creeks (with flow as low as 0.04m^3/s) and inland channels (for example Chicago Ship Canal). Dispersion coefficients from these 81 experiments range from 1486 to 0.2m^2/s. Because of data scarcity, the observations are randomly divided into two samples: training (50 experiments) and validation (31) one.

2 Neural Networks

Artificial Neural Networks have been developed by looking for the analogies to the behavior and functioning of brain and nervous system of living organisms. The most important feature imitating brain is the ability to learn from experience and to utilize the gained knowledge to solve new problems. It is a kind of ability to generalize which, we hope, could help identify the dispersion coefficient in water quality models.

2.1 Multi-layer Perceprtrons

The easiest way of transforming the input vector is to introduce one or more layers of artificial neurons in the perceptron architecture. Networks with more than one layer of artificial neurons, where only forward connection from the input towards the output are allowed, are called Multi-layer Perceptron (MLP). The topology of the neural network used in our study is presented in Fig. 1.

The Rowinski et al. [11] showed recently, based on wide range of data, that MLP is able to evaluate longitudinal dispersion coefficient better than linear techniques and that implementing sinuosity index as input variable plays crucial role in improving the model performance.

Multi Layer Perceptron (MLP)

w_{ij}, v_j - parameters

hidden nodes

$$H_j = = \left[1 + \exp\left(-\sum_i I_i \cdot w_{ij}\right)\right]^{-1}$$

$I_1 = 3UB$ I_1

$I_2 = U/U^*$ I_2

w_{ij}

$I_3 = B/H$ I_3

$I_4 = sin$ I_4

H_1

H_2
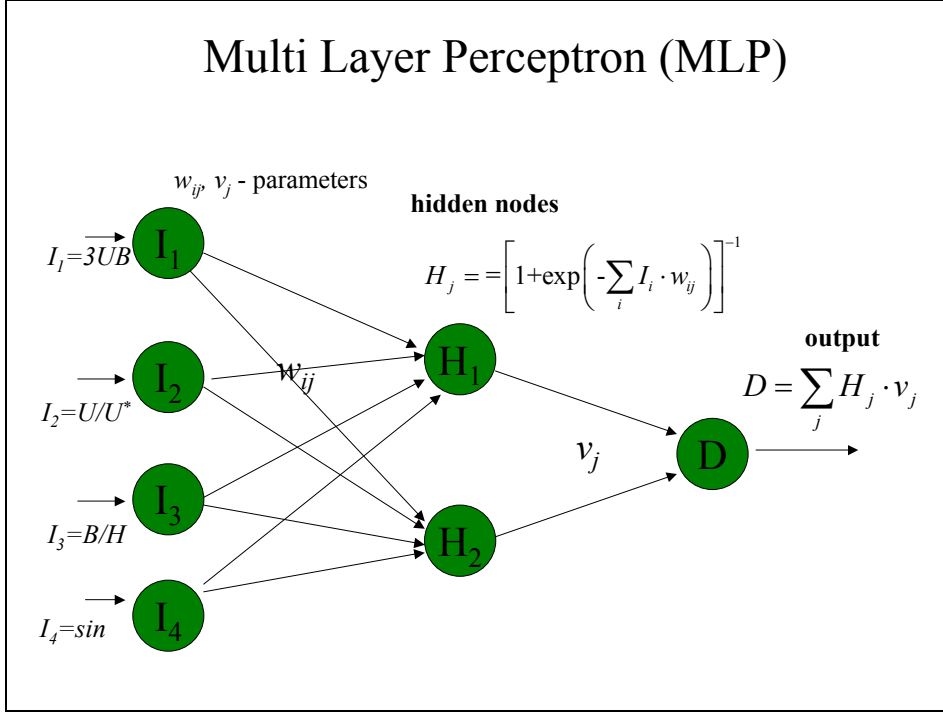
v_j

output

$$D = \sum_j H_j \cdot v_j$$

D

Figure1. A two-layer feedforward neural network

The number of input nodes is the same as the number of input variables (four in this study), number of hidden neurons should be found "as optimal" for the solution of a given problem, the number of output nodes is equal to the number of output variables (one output variable, namely longitudinal dispersion coefficient is considered in this study). The applied network consists of only one hidden layer due to the general observation that neural networks with only one hidden layer and finite number of nodes are able to approximate every continuous, bounded, everywhere differentiable function (e.g. [7]). Simple computational elements (the nodes) are linked via weighted connections. The values of those connections are adaptively modified during the process of training the network. Each node processes a weighted sum of its inputs and filters it through a given, so-called activation function. Following a number of other authors a sigmoidal function was used for this purpose.

Dozens of networks with different structures and different initial parameters value have been trained. To optimise these randomly selected parameters Levenberg-Marquardt algorithm was applied. Having in mind small size of data set available to the authors, the best network is formed by 3 hidden nodes, i.e. 19 parameters to be optimised. Results obtained from this network are presented in the final section.

2.2 Radial Basis Function Network

Radial Basis Function architecture (e.g. [10], [13]) includes one hidden layer of special units, that pre-process the input and feed a single-layer perceptron (Fig.2). Each unit k in the hidden layer contains the prototype c_k of the given region of the input space. The corresponding nonlinear activation function $\phi()$ expresses, by means of distance measure, the similarity between any input I and the prototype c_k. The most commonly adopted Basic Function is Gaussian:

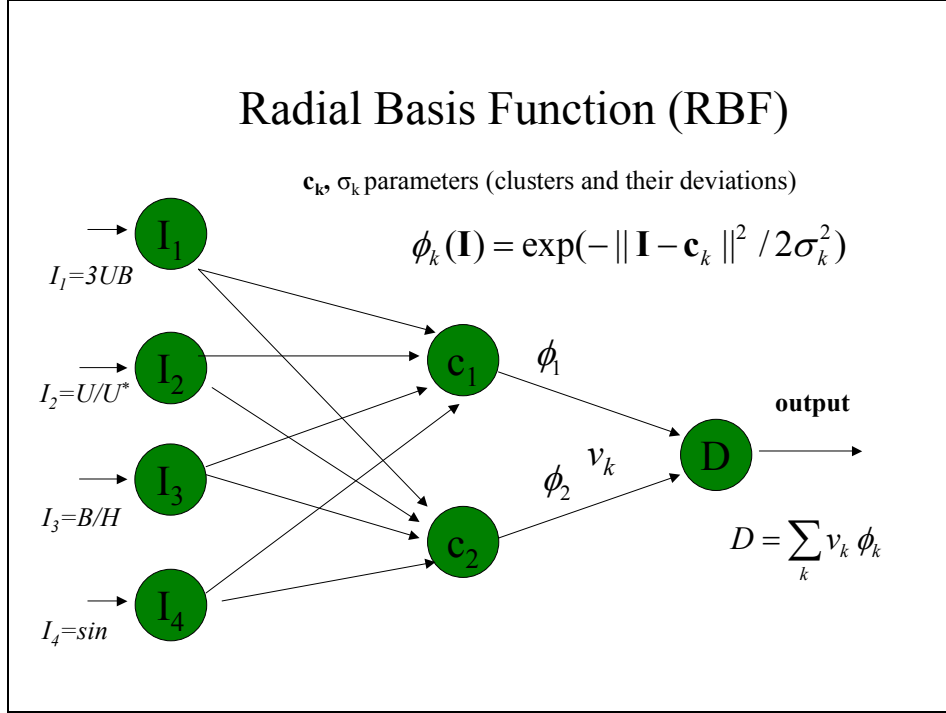$$\phi_k(\mathbf{I}) = \exp\left(-\|\mathbf{I} - \mathbf{c}_k\|^2 / 2\sigma_k^2\right) \tag{5}$$

Figure 2. The architecture of a Radial Basis Function network

The RBF networks with different number of centers have been tested. For the dataset described in the introduction, the best RBF network consists of 8 centres. Hence 48 parameters, namely 8 pairs (σ_k, c_k) (where vector c_k is composed of 4 elements), 8 weighting coefficients v_k, and 1 parameter representing the threshold value were optimised. The degree of fit of the longitudinal dispersion calculated by means of RBF network to measured value is shown in the final section.

2.3 Fuzzy Neural Network

In this section fuzzy rules [1] are used to characterize imprecise dependencies between combinations of the channel measurements and longitudinal dispersion. Using linguistic variables, the knowledge is described as fuzzy rules:

$$R^k : IF(I_1 \, is \, A_{1j}^k(j=1 \quad OR \quad j=2) \, AND...I_{4j} \, is \, A_{4j}^k(j=1 \quad OR \quad j=2)) \quad THEN \, (D^k \, is \, B^k) \, (6)$$

where k denotes the rule number; number of all possible rules is $2^4=16$. The structure of Fuzzy Neural Network applied for dispersion coefficient evaluation, shown in Fig.3, was adapted from [12]. For any input variables I_i ($i = 1,...,4$), there are two fuzzy sets A_{ij} ($j = 1,2$), defined in our case by Gausian membership function

$$\mu_{A_{ij}} = \exp\left[-\left(\tfrac{I_i - \overline{I_{ij}}}{\sigma_{ij}}\right)^2\right] \tag{7}$$

that represent two fuzzy values, namely "big" (GF) and "small" (Gf). Hence the number of that pre-processed elements in the first layer (L1), where a degree of membership of input data to appropriate fuzzy sets are evaluated, is equal 8. The second layer (L2) represents the inference unit. Note that k-th node in the second layer representing k-th rule is connected to all nodes in the first layer with corresponding fuzzy sets in the antecedent of this rule. The last two layers (L3 and L4) form defuzzification unit. In the third layer (L3) is 16 weights, so one has to find 32 unknown parameters.
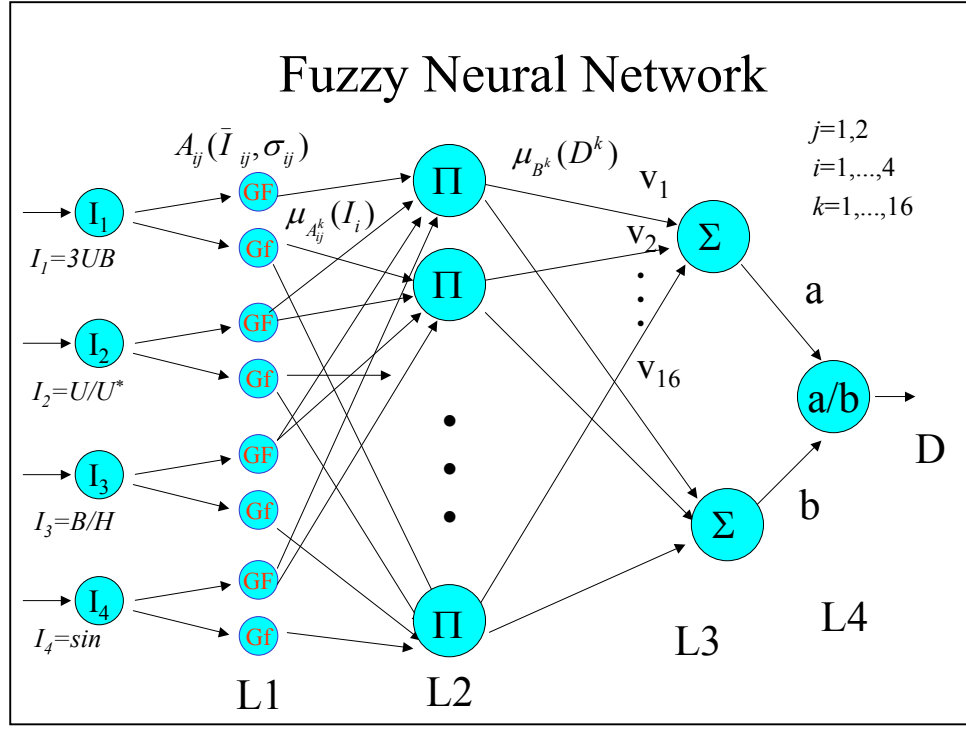
4

Figure 3. The architecture of a Fuzzy Neural network

All network parameters were optimized using gradient method with multiple random initialization of its initial values.

3   RESULTS AND DISCUSSION

To compare results obtained from different networks three measures are proposed to quantify the error. In the following equations D_m is measured and D_f is a calculated dispersion coefficient.

1) Percentage of the mean error (PME) is defined [9]:

$$PME = \frac{\frac{1}{N}\sum_{i=1}^{N}\left|\log\frac{D_{f_i i}}{D_{m_i}}\right|}{\frac{1}{N}\sum_{i=1}^{N}\log D_{m_i}}\cdot 100\% \tag{8}$$

The lower the PME index is, the better model performance.

2) Mean relative error sensitive to outliers proposed by the first author

$$MRE_O = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|D_{m_i}-D_{f_i}\right|}{min(D_{f_i},D_{m_i})} \tag{9}$$

3) Number of D_f cases (NC index) belonging to the interval $D_m/s<D_f<D_m\cdot s$ (for assumed $s>1$) is a very simple statistic. It allows to compare results given by different models, especially if a few intervals with different s values are considered. The more cases match the interval, the better the performance of considered model is.

All mentioned measures for MLP, RBF, FNN and linear regression (REG) are presented in table 1.

Table 1: Errors of dispersion coefficient assessment obtained from different neural networks

Network type	Data set	PME (%)	MRE_O	NC (%) $(D_m/s<D_f<D_m \cdot s)$	
				s=1.5	s=3
MLP		*7.12*	*0.36*	*86*	96
RBF	Training	9.60	0.49	68	*98*
FNN		10.03	0.57	68	96
REG		12.05	0.65	58	96
MLP		*10.27*	*0.46*	*77*	94
RBF	Verification	12.13	0.51	58	*97*
FNN		13.43	0.57	68	*97*
REG		17.34	0.83	55	94

Note, that all neural networks are better than simple linear regression and in the case of verification data set the difference between neural networks and linear regression is even more significant. In both cases, training and verification, the best results were obtained by means of MLP network.

NC index in case of MLP network is much higher for lower s=1.5 than for other models, so dispersion coefficients are evaluated with smaller errors. However, in the verification set two outliers occurred (only one in case of RBF and FNN), which reflected in the increase of NC index.

Based on the data set available to the authors ([3], [11], [14]) MLP network should be considered as the most reliable, RBF network better than FNN and much better than linear regression.

Although the results obtained with the use of artificial neural networks are not fully satisfying, they are far less costly than physically-based models allowing for the prediction of longitudinal dispersion coefficient and, consequently, the pattern of pollution spread in rivers. The neural networks may be very useful in situations where the data cannot be easily provided. The performance of neural networks methodology was very much improved when the river sinuosity index was added to the input data. Then the results turned out to be better than those based on any other method.

References

[1] M. Berthold: *Fuzzy Logic*, in M. Berthold and D.J. Hand (eds.) Intelligent Data Analysis, pp.321-350, 2003.
[2] Z.Q. Deng, V.P. Singh, L. Bengtsson: *Longitudinal dispersion coefficient in straight rivers*. J. Hydraul. Engng., ASCE 127 (11), pp. 919-927, 2001.
[3] Z.Q. Deng, L. Bengtsson, V.P. Singh: *Longitudinal dispersion coefficient in single-channel streams*. J. Hydraul. Engng., ASCE 128 (10), pp. 901-916, 2002.
[4] B.H. Fischer: *The mechanics of dispersion in natural streams*. J. Hydraul. Div. ASCE 93 (6), pp. 187-216, 1967.

[5] E. Gaume, R. Gosset: *Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology?*, Hydrology and Earth Sciences 7 (5), 693-706, 2003.

[6] I. Guymer: *Longitudinal Dispersion in a Sinuous Channel with Changes in Shape*, J. Hydraul. Engng. ASCE 124 (1), pp. 33-40, 1998.

[7] S. Haykin: *Neural networks, a comprehensive foundation*. Macmillan College Publishing Comp., New York, 1994.

[8] E.R. Holley, G.H. Jirka: *Mixing in rivers*. Technical Report E-86-11, Environmental and water quality operational studies. US Army Corps of Engineers, Washington D.C., 1986.

[9] S.M. Kashefipour, R.A. Falconer, B. Lin: *Modeling longitudinal dispersion in natural channel flows using ANNs*, in: *River Flow 2002*, Bousmar and Zech, Swets and Zeitlinger (eds.), Lisse, A.A. Balkema Publishers, pp. 111-116, 2002.

[10] S. Osowski: *Sieci Neuronowe do przetwarzania informacji*. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa, 2000.

[11] P.M. Rowiński, A. Piotrowski, J.J. Napiórkowski: *Are artificial neural networks techniques relevant for the estimates of longitudinal dispersion coefficient in rivers?* Hydrological Sciences Journal 50 (1), pp.175-187, 2005.

[12] D. Rutkowska, M. Piliński, L. Rutkowski: *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*. Wyd. nauk. PWN, Warszawa, 1999.

[13] R. Silipo, "Neural Networks", in M. Berthold and D.J. Hand (eds.) Intelligent Data Analysis, 2003, pp.270-320.

[14] A.N. Sukhodolov, V.I. Nikora, P.M. Rowiński, W. Czernuszenko: *A case study of longitudinal dispersion in small lowland rivers,* Water Env. Res. 69 (7), pp. 1246-1253, 1997.