Expert Systems With Applications 61 (2016) 129–144

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection

Madjid Tavana[a,b,*], Alireza Fallahpour[c], Debora Di Caprio[d,e], Francisco J. Santos-Arteaga[f,g]

[a] Business Systems and Analytics Department, Distinguished Chair of Business Systems and Analytics, La Salle University, Philadelphia, PA 19141, USA
[b] Business Information Systems Department, Faculty of Business Administration and Economics, University of Paderborn, D-33098 Paderborn, Germany
[c] Department of Mechanical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia
[d] Department of Mathematics and Statistics, York University, Toronto, M3J 1P3, Canada
[e] Polo Tecnologico IISS G. Galilei, Via Cadorna 14, 39100, Bolzano, Italy
[f] School of Economics and Management, Free University of Bolzano, 39100 Bolzano, Italy
[g] Instituto Complutense de Estudios Internacionales, Universidad Complutense de Madrid, Campus de Somosaguas, 28223, Pozuelo, Spain

ARTICLE INFO

Article history:
Received 11 January 2016
Revised 26 April 2016
Accepted 17 May 2016
Available online 18 May 2016

Keywords:
Supplier selection
Artificial neural network
Adaptive neuro fuzzy inference system
Criteria selection
Prediction

ABSTRACT

Supplier evaluation and selection constitutes a central issue in supply chain management (SCM). However, the data on which to base the corresponding choices in real life problems are often imprecise or vague, which has led to the introduction of fuzzy approaches. Predictive intelligent-based techniques, such as Artificial Neural Network (ANN) and Adaptive Neuro Fuzzy Inference System (ANFIS), have been recently applied in different research fields to model fuzzy multi-criteria decision processes where the understanding and learning of the relationships between the input and output data are the key to select suitable solutions. In this paper, a hybrid ANFIS-ANN model is proposed to assist managers in their supplier evaluation process. After aggregating the data set through the Analytical Hierarchy Process (AHP), the most influential criteria on the suppliers' performance are determined by ANFIS. Then, Multi-Layer Perceptron (MLP) is used to predict and rank the suppliers' performance based on the most effective criteria. A case study is presented to illustrate the main steps of the model and show its accuracy in prediction. A battery of parametric tests and sensitivity analyses has been implemented to evaluate the overall performance of several models based on different effective criteria combinations.

1. Introduction

In today's competitive world, decision making represents a very complicated process at basically all levels. It is undoubtedly hard for decision makers to select the best alternative based on several attributes/criteria/factors, particularly when the information on which to rely is incomplete or imprecise. Multi-Criteria Decision Making (MCDM) is the main research field dealing with the complexity of both evaluation and selection problems and their possible solution methods.

Several MCDM techniques have been presented to support decision makers through their decision making processes, including the Analytical Hierarchy Process (AHP) and Analytical Network Process (ANP), Data Envelopment Analysis (DEA), Mathematical Programming (MP) and hybrid models combing the previous and other ranking techniques. Each one of these methods has its advantages and limitations. For instance, AHP and ANP are easy to use, but rely heavily on human judgment, especially in determining the weights of criteria (Vahdani, Iranmanesh, Mousavi, & Abdollahzade, 2012). DEA is sensitive to outliers and statistical noise (Azadeh, Saberi, & Anvari, 2011) while MP models are very precise but do not consider qualitative attributes (Golmohammadi, 2011). We will often refer to Golmohammadi (2011) through the current paper. Thus, we will denote this reference by GHD henceforth.

In the last decades, Artificial Intelligence (AI) has been receiving more attention in the decision making literature. The AI approach is usually twofold: prediction, which constitutes the focus of this paper, and optimization. Prediction is carried out by implementing different techniques such as Artificial Neural Network (ANN), Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Machine (SVM), which enable decision makers to take the best possible decision (output) based on past and present information and future predictions (input) (GHD).

Decisions concerning the evaluation and selection of the right suppliers play a fundamental role in all manufacturing activities.

* Corresponding author. Fax: +1 267 295 2854; Web: http://tavana.us/.
E-mail addresses: tavana@lasalle.edu (M. Tavana), fallahpour.a@gmail.com (A. Fallahpour), dicaper@mathstat.yorku.ca, debora.dicaprio@istruzione.it (D. Di Caprio), fsantosarteaga@unibz.it, fransant@ucm.es (F.J. Santos-Arteaga).

http://dx.doi.org/10.1016/j.eswa.2016.05.027
0957-4174/

Two main problems arise when selecting the best possible alternative: determining the proper criteria and assessing the performance of the alternatives on the basis of the selected criteria. Theoretically, there might be many factors affecting the performance of the available alternatives but, in practice, only a few criteria determine the evaluation process. An unnecessary large number of inputs not only weakens the clarity of the underlying model, but also increases its computational complexity (Jang, 1996). At the same time, the existence of conflicting criteria (i.e. cost and quality, responsiveness and flexibility) complicates the identification of the main ones determining the performance of the alternatives. Therefore, a powerful enough method is needed to find the most important criteria (inputs) and use them to identify the best performing alternative (output).

The main objective of the current paper is to illustrate how two ANN approaches such as ANFIS and the ANN architecture known as Multi-Layer Perceptron (MLP) can complement each other in the design of a supplier selection framework with incomplete information. In this regard, the strategic integration of ANFIS within the initial stages of an MLP-based supplier evaluation model constitutes the main contribution of this paper. The capacity of ANN to replicate the behavior of the alternatives becomes particularly important when data on a given alternative are missing. At the same time, ANFIS allows us to identify the most influential combinations of characteristics that can be used to replicate the performance of the potential suppliers and evaluate their behavior. That is, consider a situation where the decision maker has limited information processing capacity or faces a very large set of supplier characteristics. In this case, the decision maker can use ANFIS to select a subset of characteristics that allow him to classify the alternatives via MLP with the same confidence as if he was able to use the entire set of data.

In order to show its applicability to decision making situations, the proposed model has been implemented in a case study involving an automotive industry whose alternatives are the suppliers of the company. Moreover, in order to test its effectiveness, the results derived from our ANFIS-ANN framework have been compared with those obtained using only the ANN model. This comparison shows that the proposed model is more accurate. As a result, it can be applied to a large class of MCDM problems and implemented by managers to rank alternatives without having to continuously update their judgments (data set) or when lacking a subset of observations from a given alternative.

The paper proceeds as follows. Section 2 provides a short review of the related literature on decision making. Section 3 presents the methodology, while Section 4 illustrates the case study discussing the data collection and the results derived from the model. In Section 5, the performance of the model is evaluated using different statistical tests and compared with the results provided by other techniques. Section 6 discusses the applicability of our hybrid model to real-life supplier evaluation problems. Finally, Section 7 concludes.

2. Literature review

2.1. Supply chains and their evolution

The academic research on supply chains has evolved through time while emphasizing the following characteristics of the chain:

- Initially, it mainly focused on the pecuniary incentives motivating the vertical integration and quasi-integration of firms (Blois, 1972; Houssiaux, 1957). In particular, researchers acknowledged the fact that the costs of integrating operations within the chain could be smaller than those of contracting due to the existence of appropriable quasi-rents and the possibility of default by other firms (Aoki, 1988; Klein, Crawford, & Alchian, 1978).
- The literature moved towards emphasizing cooperation in the organization of vertical markets when accounting for the circulation of information and technology across the chain (Colombo & Mariotti, 1998; Esposito & Passaro, 2009). The emergence of trust between firms (Day, Fawcett, Fawcett, & Magnan, 2013; Sako, 1992), and the existence of different knowledge and technology capacities within the chain followed as main research topics (Cannavacciuolo, Iandoli, Ponsiglione, & Zollo, 2015; Esposito & Raffa, 1994).
- The literature has lately emphasized the skill specificity acquired by suppliers through learning and technological investments (Asanuma, 1989; Day, Lichtenstein, & Samouel, 2015) together with the resilience of the chain to disruptions caused by unforeseen events (Heckmann, Comes, & Nickel, 2015; Hohenstein et al., 2015).

Recently, criticisms have been raised regarding the standard approaches focusing on data transmission to measure the interactions across supply chain partners. It has been argued that increasing the cooperation and integration of the different elements composing the chain (Stevens, 1989) requires a complex framework that goes beyond data and information transmission and considers the exchange of opinions, expertise and knowledge (Bessant, Kaplinsky, & Lamming, 2003; Frohlich & Westbrook, 2001; Gressgård & Hansen, 2015; Opengart, 2015). As a result, fuzzy methods have gained incremental attention when applied to manage the sharing of subjective opinions and expertise reports provided by different agents along the chain (Bruno, Esposito, Genovese, & Simpson, 2016; Labib, 2011).

2.2. MCDM and AI methods for supplier selection

MCDM methods have been successfully implemented in different research areas such as the textile industry, the automotive industry, civil engineering, banking and supply chains (Guneri, Ertay, & YüCel, 2011; Guneri, Yucel, & Ayyildiz, 2009; Opricovic & Tzeng, 2004; Wu, 2009; Wu, Yang, & Liang, 2006; GHD; Büyüközkan & Çifçi, 2012a, 2012b).

When considering supplier selection, the literature initially emphasized costs as the main selection criterion but the complexity of the supplier evaluation problem has led to the implementation of MCDM techniques (Bhutta, 2003), whose application has evolved significantly in the latter years (Parthiban, Zubar, & Katakar, 2013). In this regard, AHP has become a standard technique used to determine the relative importance of the selection criteria (Bruno, Esposito, Genovese, & Passaro, 2012). Moreover, AHP can be easily combined with other techniques such as DEA (Weber, Current, & Benton, 1991) and ANN (Ha & Krishnan, 2008) to obtain a final ranking of the alternatives. The use of AI-based methods remains generally limited in relation to other techniques (de Boer, Labro, & Morlacchi, 2001), a tendency that pervades nowadays (Chai, Liu, & Ngai, 2013). However, due to their capacity to replicate the behavior of suppliers, these latter methods become particularly useful when firms face incomplete information environments.

Several papers implementing AI-based methods such as ANN or ANFIS within different decision making environments are recalled below.

Choy, Lee, and Lo (2003) proposed a combined ANN-based model to choose and benchmark potential partners of Honeywell Consumer Products Limited in Hong Kong. Lee and Ou-Yang (2009) introduced an ANN-based predictive decision model for assessing vendors' performance. Guneri et al. (2011) proposed a predictive ANFIS-based model for both criteria selection and perfor-

mance evaluation. They implemented the model in a textile industry study case. GHD proposed an ANN-based model for evaluating the performances of different alternatives. In his research, AHP pair-wise comparisons were used to rank the alternatives based on their performance. Then, the relationships between the selected criteria and the performance of the alternatives were determined by ANN.

Vahdani et al. (2012) presented a linear neuro-fuzzy model for supplier assessment in a cosmetic industry. First, they selected the appropriate criteria for assessing the vendors. Then, they gathered the data set for the criteria and performance evaluation using a numerical scale. After collecting historical data on attributes and performances, they divided the resulting data set into two parts for implementing the proposed model and testing its predictive ability. In order to illustrate the estimation capacity of the model, the results were compared with the results obtained by a Radial Basis Function (RBF) neural network, a Multi-Layer Perceptron (MLP) neural network and a Least Square-Support Vector Machine (LS-SVM).

The current paper considers a fuzzy decision environment to which incorporates an information selection mechanism designed to replicate the performance of suppliers using AI methods. Our model aims at enhancing the capacity of firms to take decisions when part of the knowledge and information required is missing, i.e. in incomplete information settings. However, AI methods can be complex to understand and implement by firm managers and decision makers. Therefore, we will assume that if a company decides to implement an ANN architecture such as MLP, it should be able to include other AI methods such as ANFIS in its evaluation and decision process.

3. Methodology

This section starts with a brief overview of ANFIS. Then, the MATLAB command that implements ANFIS, which will be used to determine the most important criteria (inputs) affecting the performance of the alternatives (output), is provided and the ANN architecture known as Multi-Layer Perceptron (MLP) reviewed. Finally, the integrated ANFIS-ANN model is explained in detail.

3.1. Adaptive neuro fuzzy inference system (ANFIS)

Jang (1993) hybridized a Fuzzy Inference System (FIS) with a neural network to present ANFIS. The structure of ANFIS consists of if-then rules and input-output data processing where the learning algorithm of a neural network is used for training. ANFIS is a methodology employed to simulate complex nonlinear mappings using neural network learning and fuzzy inference methodologies (Bektas Ekici & Aksoy, 2011). It adjusts membership functions and the related parameters towards the target data sets (Wu, Hsu, & Chen, 2009).

Conventionally, an ANFIS structure includes five layers (Admuthe & Apte, 2010): a fuzzified layer, product layer, normalized layer, defuzzified layer, and a total output layer. A simple ANFIS structure is the one associated with the Sugeno Fuzzy model. This model is known for allowing to generate fuzzy rules from an input-output data set, a typical fuzzy rule being "*if x is A and y is B then* $z = f(x, y)$", where A and B are fuzzy labels and f is a crisp function (see Jang, 1996). The corresponding ANFIS structure includes two inputs, x and y, and one output and it is represented in Fig. 1. The nodes within the same layer of this network perform functions of the same type. The layers are described more in detail below.

Layer 1. Every node in this layer is an adaptive node endowed with a node function. More precisely, this layer consists of four nodes. The first two nodes are represented by fuzzy linguistic labels, A_1 and A_2, that apply to the input x. The other two nodes are fuzzy linguistic labels, B_1 and B_2, associated with the input y. Each node produces an output by means of the membership function associated with the label characterizing the node.

Thus, there are two kind of outputs in Layer 1: the output $O^x_{1,i}$ of the node i (where $i = 1, 2$) with antecedent x:

$$O^x_{1,i} = \mu_{A_i}(x), \quad i = 1, 2 \tag{1}$$

and the output $O^y_{1,i}$ of the node i (where $i = 1, 2$) with antecedent y:

$$O^y_{1,i} = \mu_{B_i}(y), \quad i = 1, 2 \tag{2}$$

where μ_{A_i} and μ_{B_i} are the membership functions of A_i and B_i (regarded as fuzzy sets), respectively. In ANFIS, the generalized Bell function is usually used as a membership function:

$$\mu_\Lambda(\lambda) = \frac{1}{1 + \left|\frac{\lambda - c_\Lambda}{a_\Lambda}\right|^{2b_\Lambda}} \tag{3}$$

where $\Lambda \in \{A_1, A_2, B_1, B_2\}$ and a_Λ, b_Λ, c_Λ are real values determining the shape of the membership function of the fuzzy set Λ. These values are known as the "premise parameters".

Layer 2. This layer consists of two nodes. The output of node i (where $i = 1, 2$) is the firing strength w_i of the i-th rule and it is obtained as the product of the outputs of Layer 1:

$$O_{2,i} = w_i = \mu_{A_i}(x)\ \mu_{B_i}(y), \quad i = 1, 2 \tag{4}$$

Layer 3. (The normalized layer). This layer consists of two nodes. Each node i (where $i = 1, 2$) in this layer computes the ratio of the i-th rule's firing strength to the sum of all rules' firing strengths.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \tag{5}$$

Layer 4. In this layer every node i (where $i = 1, 2$) is adaptive and endowed with a node function f_i. The output of node i is given by:

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i), \quad i = 1, 2 \tag{6}$$

where $\bar{w}_i$ is the output of Layer 3 and p_i, q_i and r_i are referred to as the linear consequent parameters.

Layer 5. In this layer the overall output of ANFIS is computed as the sum of all incoming signals from Layer 4.

$$O_{5,i} = \sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{\sum_{i=1}^{2} w_i} \tag{7}$$

ANFIS applies a hybrid learning rule algorithm which combines the back propagation algorithm with the least squared method: the former is used for the parameters in Layer 1 while the latter is employed for training the parameters (Ho, Lee, Chen, & Ho, 2002).

3.2. The command of ANFIS in MATLAB software for criteria selection

In order to identify the most influential criteria (or factors) affecting the output, i.e., the performance of the different alternatives, using ANFIS, the "exhsrch" MATLAB command is implemented. This command is executed as follows:

exhsrch(1, trn_data, test_data, input_name)

where "trn_data" and "test_data" correspond to training data and testing data, respectively, while "input_name" stands for the list of all inputs. To evaluate a higher number of input combinations, it suffices to replace "1″ with any other number in the "exhsrch" command argument. As will be done when implementing the ANN section of the model below, we use half of the data available for training while the other half is used for testing purposes. For further information regarding how the "exhsrch" command works, the reader may refer to Jang (1996).

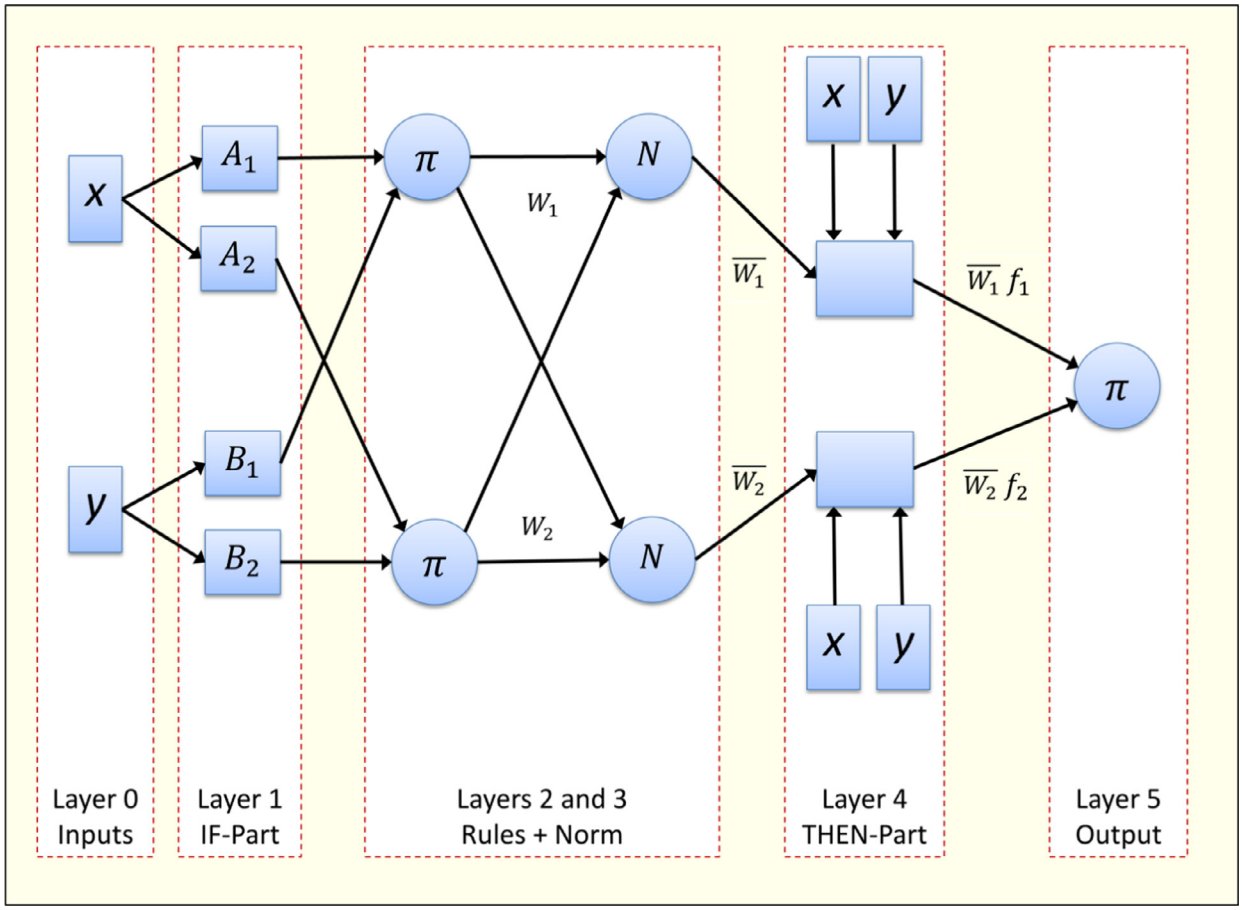

Fig. 1. ANFIS structure.

3.3. Artificial neural network (ANN)

Artificial neural network (ANN) is a widely used non-linear functional approximator (Asadi, Hadavandi, Mehmanpazir, & Nakhostin, 2012). This subsection provides a brief review of one of the main variants of ANN known as Multi-Layer Perceptron (MLP).

MLP is a feed-forward-based architecture of ANN (Cakir & Yilmaz, 2014) which is usually trained using the Back Propagation (BP) learning algorithm. There are at least three layers in a MLP network including an input layer, at least one hidden layer of neurons (or processing units), and an output layer. Each one of these layers has several processing units and each unit is fully interconnected through weighted connections to units in the subsequent layer (Gandomi & Alavi, 2011). These units are the nodes of the layer. To calculate the output of the j-th neuron in a hidden layer, all the inputs received from the neurons in the previous layer are multiplied by the corresponding connection weight to this j-th node (Mirzahosseini, Aghaeifar, Alavi, Gandomi, & Seyednour, 2011). More precisely, the output h_j of the j-th neuron in a fixed hidden layer is defined as follows:

$$h_j = f\left(\sum_{i=1}^{n} x_i w_{ij} + b\right) \tag{8}$$

where f is the activation function (i.e. hyperbolic tangent sigmoid or log-sigmoid), n is the number of neurons in the previous layer, x_i (with $i = 1, 2, \ldots, n$) is the activation input received from the i-th node in the previous layer, w_{ij} is the weight of the connection joining the j-th neuron in the fixed layer with the i-th neuron in the previous layer, and b is the bias for the neuron.

3.4. The proposed ANFIS-ANN model

As stated above, this study proposes a hybrid AI-based model that combines ANFIS with ANN to improve the criteria selection process and, hence, facilitate the corresponding decision making process for managers. The model is shown in Fig. 2.

Except for the third step (ANFIS-step), the other steps of the model are similar to those implemented in other intelligent-based models (Alavi & Gandomi, 2011; Alavi, Aminian, Gandomi, & Esmaeili, 2011; GHD; Vahdani et al., 2012). The steps of the current model are defined as follows:

- Choosing the reference criteria to analyze the performance of the alternatives.
- Gathering the complete data set (inputs and outputs): historical and recent input data are collected and the output (performance) data obtained using AHP.
- Using ANFIS to identify the most influential criteria accounting for the maximum collective effects on the performance of the alternatives.
- Implementing the optimized ANN model using training and testing data in order to replicate the ranking results obtained through AHP.

Following GHD's approach, the data collected are separated randomly and 50% of the data set is used for training. The remaining 50% is used for testing and assessing the accuracy of the mathematical model in predicting the performance of the alternatives.

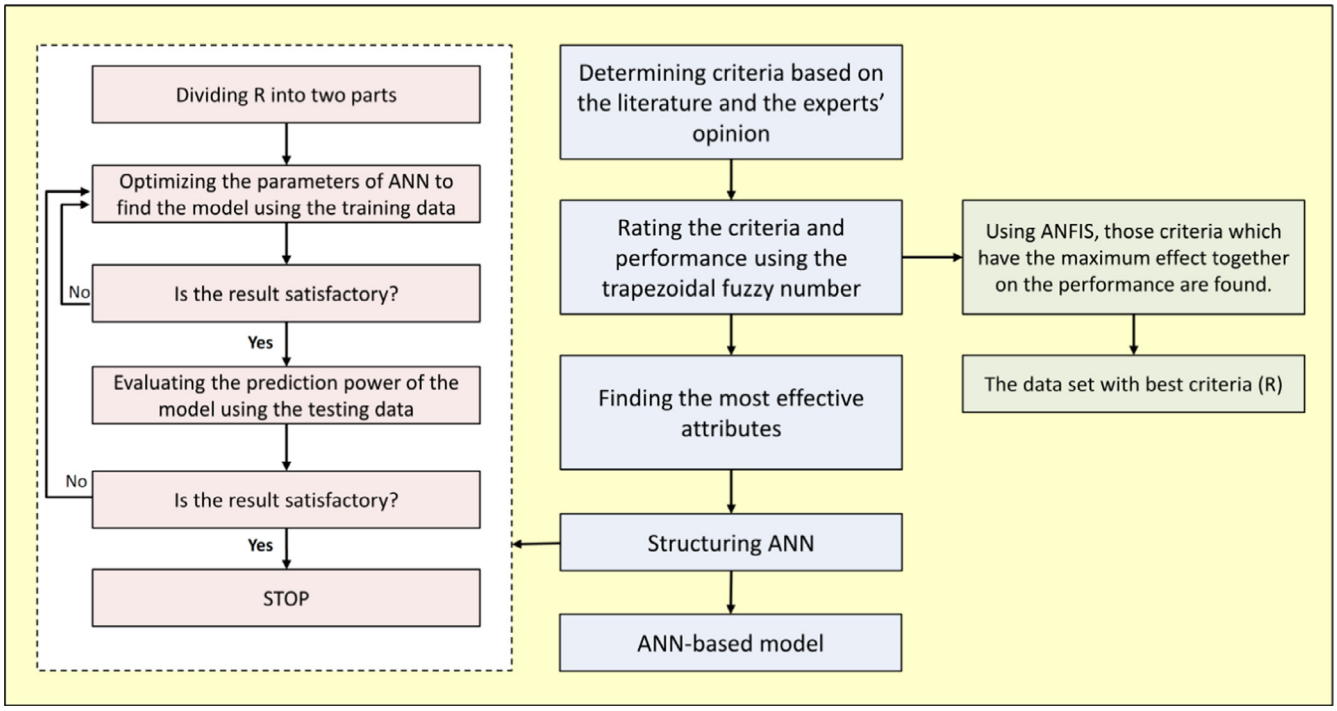

Fig. 2. Proposed ANFIS-ANN model.

3.5. Statistic indicators

The correlation coefficient (R) and the mean squared error (MSE) are used to measure the accuracy of the proposed ANFIS-ANN model in estimating the performance of the alternatives:

$$R = \frac{\sum_{\alpha=1}^{\Delta}(r_\alpha - \bar{r})(t_\alpha - \bar{t})}{\sqrt{\sum_{\alpha=1}^{\Delta}(r_\alpha - \bar{r})^2 \sum_{\alpha=1}^{\Delta}(t_\alpha - \bar{t})^2}} \tag{9}$$

$$MSE = \frac{\sum_{\alpha=1}^{\Delta}(r_\alpha - t_\alpha)^2}{\Delta} \tag{10}$$

where Δ is the number of alternatives, r_α and t_α ($\alpha = 1, 2, \ldots, \Delta$) are the actual and predicted values of the α-th output (i.e. the performance of the α-th alternative), and $\bar{r}$ and $\bar{t}$ are the averages of the actual and predicted performances, respectively. It should be noted that the Root MSE (RMSE) is also generally employed when implementing ANFIS.

Overall, the following more general features of the proposed ANFIS-ANN model can be identified:

a. It shows how combining ANFIS and ANN constitutes a significant contribution to the area of decision making.
b. ANFIS handles the complex task of determining the most important criteria/factors in a decision making process without imposing any functional form on the ANN.
c. ANN allows to predict the behavior of decision making units, enabling managers to monitor and evaluate their corresponding performances even when data on a subset of criteria from an alternative are missing.

4. Case study

4.1. Data set and evaluation criteria

This paper uses one of the data sets (i.e. the "improved" data set) collected by Golmohammadi. GHD's case study analyzes a company in the automotive industry that must choose among 31 suppliers for 8 products. The products are denoted by A, B, C, D, E, F, G, H and each supplier may deliver more than one product. The first column of Table 1 shows the list of suppliers with the corresponding contract types analyzed in GHD. After identifying the type of contract that a supplier can sign with the letter of the corresponding product, 33 supplier-contract pairs are obtained and can be used as alternatives for the implementation of the ANFIS-ANN model proposed in the current study. These alternatives are presented in the last column of Table 1.

Based on the experts' opinion, Quality (Q), Delivery (D), Technology (T), Price (P) and Location (L) were determined as the appropriate criteria to analyze the suppliers' performance and, hence, rank the suppliers. To measure the suppliers' performance on each of these criteria, historical and recent suppliers' data were collected and several functions defined to convert the raw data into input data to implement the model. The input functions applied by GHD for each criterion are described below.

- *Quality.* This criterion accounts for the quality history of the suppliers. It is defined as the ratio of the defective parts to the total number of parts supplied. That is:

$$Q = \frac{\sum_{\gamma=1}^{\Gamma} d_\gamma}{\rho} \tag{11}$$

where Q is the input function for quality, Γ is the total number of deliveries per contract, γ ($\gamma = 1, 2, \ldots, \Gamma$) is one of the contract deliveries, d_γ is the number of defective parts in the γ-th delivery, and ρ is the total number of products delivered.

- *Delivery*. This criterion includes quantity and punctuality in delivering. These two important factors are used to measure suppliers' performance as follows:

$$D_q = \sum_{\gamma=1}^{\Gamma} \frac{\frac{|q_{\gamma p} - q_{\gamma a}|}{q_{\gamma p}}}{\Gamma} \tag{12}$$

$$D_t = \sum_{\gamma=1}^{\Gamma} \frac{\frac{|t_{\gamma p} - t_{\gamma a}|}{t_{\gamma p}}}{\Gamma} \tag{13}$$

Table 1
Data set and AHP-based evaluation of suppliers' performance.

(Supplier number, Contract type) as in GHD	Inputs: GHD's data set					Output: AHP suppliers' evaluation	Suppliers' renumbering in the proposed model
	Q	D	T	P	L		
(18, E)	0.078	0.280	60	0.23	0.25	0.03	S_1
(10, A)	0.031	0.140	50	0.29	0.25	0.04	S_2
(31, H)	0.038	0.130	55	0.33	0.11	0.04	S_3
(22, H)	0.039	0.130	55	0.37	0.16	0.06	S_4
(8, B)	0.012	0.186	60	0.27	0.18	0.08	S_5
(4, C)	0.039	0.210	60	0.23	0.19	0.08	S_6
(8, A)	0.009	0.120	70	0.25	0.11	0.10	S_7
(17, E)	0.021	0.190	65	0.24	0.15	0.10	S_8
(1, C)	0.021	0.190	60	0.27	0.11	0.11	S_9
(9, D)	0.011	0.010	55	0.34	0.09	0.12	S_{10}
(26, F)	0.023	0.060	60	0.22	0.15	0.13	S_{11}
(16, C)	0.019	0.220	55	0.19	0.15	0.14	S_{12}
(23, G)	0.009	0.120	80	0.28	0.08	0.14	S_{13}
(7, E)	0.014	0.150	90	0.31	0.16	0.18	S_{14}
(5, A)	0.014	0.070	65	0.35	0.13	0.19	S_{15}
(13, B)	0.056	0.110	70	0.22	0.17	0.19	S_{16}
(15, D)	0.024	0.130	70	0.22	0.13	0.21	S_{17}
(28, G)	0.018	0.190	70	0.20	0.17	0.21	S_{18}
(12, C)	0.014	0.170	85	0.29	0.16	0.23	S_{19}
(6, B)	0.012	0.030	75	0.31	0.11	0.26	S_{20}
(20, F)	0.013	0.110	90	0.26	0.13	0.26	S_{21}
(30, H)	0.009	0.050	80	0.26	0.18	0.32	S_{22}
(5, D)	0.012	0.080	60	0.25	0.18	0.34	S_{23}
(24, G)	0.009	0.080	70	0.23	0.09	0.34	S_{24}
(27, H)	0.013	0.080	70	0.31	0.13	0.40	S_{25}
(5, B)	0.005	0.050	80	0.44	0.13	0.41	S_{26}
(11, C)	0.010	0.090	75	0.25	0.09	0.42	S_{27}
(21, F)	0.010	0.090	80	0.27	0.09	0.46	S_{28}
(3, A)	0.007	0.060	90	0.40	0.22	0.52	S_{29}
(14, E)	0.003	0.040	90	0.29	0.11	0.53	S_{30}
(25, G)	0.004	0.010	85	0.35	0.16	0.54	S_{31}
(19, H)	0.003	0.010	85	0.29	0.17	0.56	S_{32}
(29, F)	0.005	0.030	90	0.24	0.17	0.63	S_{33}

$$D = D_q + D_t \tag{14}$$

where D is the input function for delivery, Γ is the total number of deliveries per contract, γ ($\gamma = 1, 2, \ldots, \Gamma$) is one of the contract deliveries, D_q is the quantity evaluation function, D_t is the "on time" evaluation function, $q_{\gamma p}$ ($\gamma = 1, 2, \ldots, \Gamma$) is the quantity planned based on the contract, $q_{\gamma a}$ ($\gamma = 1, 2, \ldots, \Gamma$) is the actual quantity delivered, $t_{\gamma p}$ ($\gamma = 1, 2, \ldots, \Gamma$) is the delivery time planned based on the contract, and $t_{\gamma a}$ ($\gamma = 1, 2, \ldots, \Gamma$) is the actual delivery time.

- *Price*. This criterion includes discount issues and types of payment. Since the final offer price from supplier to buyer is a tangible comparison criterion, the prices offered by the suppliers are considered as the input function values.
- *Location*. Transportation cost (TC) is also a tangible criterion for comparison. Thus, the TCs of the suppliers are considered as the input function values.
- *Technology*. A scale from 10 (low) to 100 (high) is used to rate the suppliers according to this criterion. The values assigned to the suppliers are considered as the input function values.

After collecting the data set corresponding to these criteria, GHD used Saaty's (1980) well-known AHP pairwise comparisons to evaluate the suppliers' performance. Table 1 presents the data set and the AHP-based evaluations in terms of performance for all the suppliers. For further information on how the data set was gathered and the pairwise comparison matrices defined, the reader may refer to GHD.

4.2. Implementation of the proposed model

4.2.1. Finding the most influential criteria combination on the performance

After collecting the data set and ranking the suppliers' performances using AHP, ANFIS was implemented to determine the most influential criteria combinations on the performances of the suppliers. To do this, the MATLAB command, "exhsrch", was executed. As already explained in Section 3, the argument of this command is "(S, trn_data, test_data, input_name)" and can be used to analyze and compare all the criteria combinations of a fixed size S with respect to the effect they have on the suppliers' performance.

We started by identifying the most influential single criterion. The command "exhsrch" was executed with $S = 1$. The ANFIS model was run five times. The results obtained show that quality (Q) is the most influential criterion on the suppliers' performance considering both training (RMSE 0.67) and testing (checking) (RMSE 0.24) (see Fig. 3a).

To find the most influential combination of two criteria, the command "exhsrch" was executed with $S = 2$. In this part, the ANFIS model was run 10 times. In each run, two inputs were selected and combined as criteria. The results obtained show that delivery (D) and technology (T) are the two most influential criteria considering both training (RMSE 0.044) and testing (RMSE 0.018) (see Fig. 3b).

The same procedure, but with $S = 3$, was implemented to find the most influential combination of three criteria. In this case, ANFIS was run 10 times and the results (Fig. 3c) show that the combination consisting of D, T and P is the one affecting the most the suppliers' performance. The model training and testing returned RMSE = 0.005 and RMSE = 0.199, respectively, for this combination.

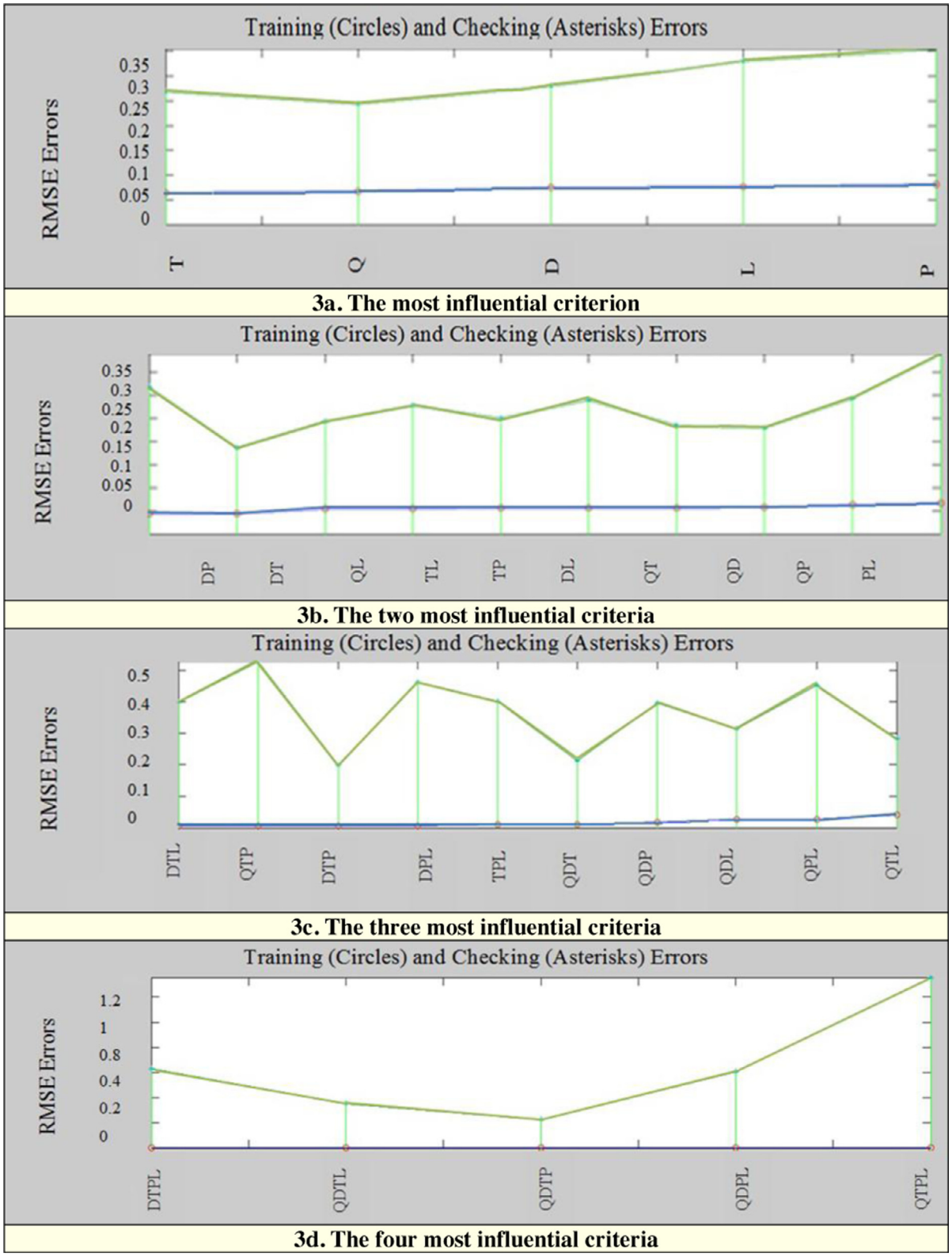

Fig. 3. Most influential criteria combinations using ANFIS.

Finally, Fig. 3d illustrates how after running ANFIS 5 times Q, D, T and P is the combination of four criteria affecting the most the suppliers' performance. The rate of RMSE in training equals 0, while in testing it is 0.263.

Given the above numerical results, it is clear that the combination of four criteria exhibits the best performance when training is considered. However, the very same combination performs worse than any of the previous ones when it comes to testing, with D, T, P and D, T arising also as potential candidates when both training and testing are considered. This is particularly the case for the three criteria combination, which exhibits a very small training error and a lower testing error than the Q, D, T, P combina-

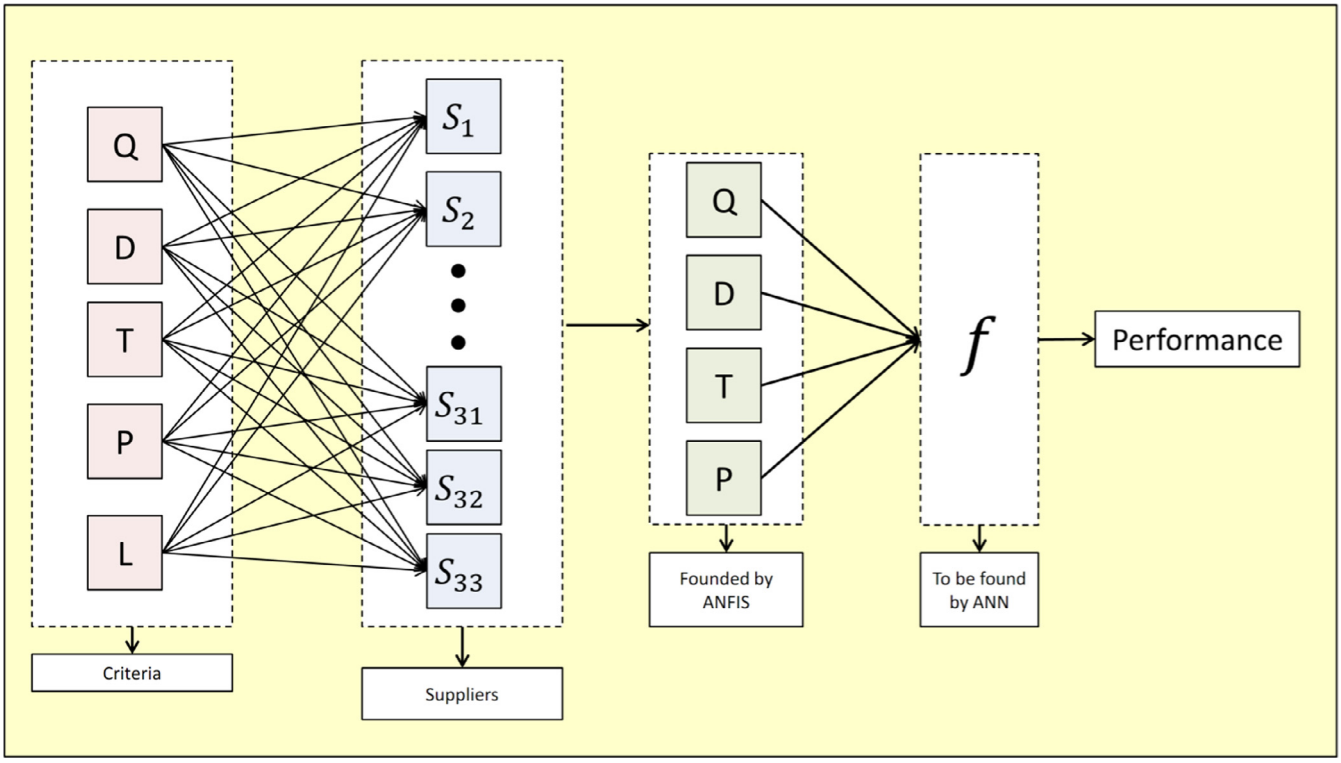

Fig. 4. Hierarchical model for suppliers' performance assessment.

tion. Therefore, we will evaluate and compare the performance of the two, three and four criteria combinations throughout the next stages of our hybrid supplier evaluation model.

4.2.2. ANN-based model for suppliers' performance

As next step of the proposed process, a neural-based model was defined to compute and evaluate the suppliers' performance. This part of the model aims at providing managers with a reliable supplier evaluation and selection method. Fig. 4 represents the performance evaluation model based on the selected combination of four criteria.

It should be noted that after finding the most influential combination of criteria using ANFIS, the data set is divided into two parts for training and testing. In order to compare the results of the proposed model with those obtained by GHD, 50% of the data was used for training and 50% for testing (the same division was implemented by GHD).

In this paper, Neurosolution 5 has been used to run the ANN model. It must be highlighted that there is no exact general rule allowing to determine the best structure (in terms of number of hidden layers, number of nodes and motivation function) for a network model. Network design is a trial and error process and may factors affect the precision of the resulting trained network model.

Therefore, in order to find the optimal structure for training the ANN model, the model was run using several different structures. After finding the best structure, the testing data were used to validate the accuracy of the ANN model in predicting the performance of suppliers. Tables 2–4 show the different architectures used to run the MLP-ANN model when the most influential combinations of two, three and four criteria were considered, respectively. The rows in bold characters indicate the best structure among the MLP models that were run in each corresponding setting.

Consider now the implementation of the fourth MLP model architecture when defining the different MLP-ANN models. Denote by ANN4 the artificial neural network designed using the Q, D, T and P criteria. Similarly, we will denote by ANN3 the artificial neural network model based on the D, T and P criteria, while ANN2 will be used to represent the neural network implemented using D and T. As can be observed in Tables 2–4, the performance of the ANN2 and ANN3 models in terms of the correlation coefficient is inferior to that of ANN4. However, and we will build some of our main results on this property, they both generate lower MSE testing values. Thus, the predicted values obtained using these models exhibit a less correlated performance but provide a higher degree of precision. This latter property will be particularly important for the ANN3 model, which exhibits a higher correlation value than the ANN2 one.

The capacity of ANN3 to compete in precision with ANN4 when considering the testing section of our model will be emphasized throughout the rest of the paper. The ANFIS results illustrated how model training was clearly dominated by the combination of four criteria, which exhibited a RMSE equal to zero. However, the four criteria combination suffered a considerable decrease in precision relative to the other configurations (including the use of only the Q criterion) in the testing section of the ANFIS model. This finding has led us to consider the alternative combinations of criteria and emphasize the importance of implementing ANFIS as a criteria selection mechanism in the early stages of the evaluation process.

4.2.3. Comparing the ANN results obtained based on different criteria combinations

Table 5 describes the predicted values obtained when implementing the different MLP-ANN models analyzed in the paper within the training and testing sections of the supplier evaluation process.

These data will be used in Figs. 5 and 6 to illustrate the performance of the ANN3 and ANN2 models relative to the real (reference) values obtained using AHP and to ANN4. Fig. 5 concentrates on the training data while Fig. 6 describes the testing results.

Table 2
MLP model architectures with two criteria and their statistical factors.

Output Performance (D and T criteria)								
MLP model	Learning method	Learning method rate	Number of hidden layers	Number of nodes (first layer/second layer)	Motivation function (first layer)	Motivation function (second layer)	R in testing	MSE in testing
1	Momentum	0.3	1	3	Sigmoid	–	0.0021	0.027
2	CG	–	2	3/4	Tan	Sigmoid	0.67	0.009
3	LM	–	1	4	Sigmoid	–	0.23	0.045
4	**Momentum**	**0.7**	**2**	**4/4**	**Tan**	**Tan**	**0.67**	**0.008**
5	Momentum	0.4	2	3/2	Sigmoid	Sigmoid	0.38	0.07
6	DBD	–	2	4	Tan	Tan	0.0013	0.091
7	Quick Prop	–	2	4/5	Linear Sigmoid	Sigmoid	0.29	0.04
8	CG	–	1	5	Linear Tan	–	0.63	0.010
9	LM	–	2	4/3	Tan	Linear	0.66	0.087
10	Momentum	0.5	2	3/3	Linear Sigmoid	Tan	0.57	0.48

Table 3
MLP model architectures with three criteria and their statistical factors.

Output Performance (D,T and P criteria)								
MLP model	Learning method	Learning method rate	Number of hidden layers	Number of nodes (first layer/second layer)	Motivation function (first layer)	Motivation function (second layer)	R in testing	MSE in testing
1	Momentum	0.3	1	3	Sigmoid	–	0.045	0.11
2	CG	–	2	3/4	Tan	Sigmoid	0.53	0.019
3	LM	–	1	4	Sigmoid	–	0.20	0.013
4	**Momentum**	**0.7**	**2**	**4/4**	**Tan**	**Tan**	**0.73**	**0.006**
5	Momentum	0.4	2	3/2	Sigmoid	Sigmoid	0.70	0.051
6	DBD	–	2	4	Tan	Tan	0.087	0.019
7	Quick Prop	–	2	4/5	Linear Sigmoid	Sigmoid	0.068	0.402
8	CG	–	1	5	Linear Tan	–	0.55	0.210
9	LM	–	2	4/3	Tan	Linear	0.60	0.079
10	Momentum	0.5	2	3/3	Linear Sigmoid	Tan	0.70	0.0084

Table 4
MLP model architectures with four criteria and their statistical factors.

Output Performance (Q, D,T and P criteria)								
MLP model	Learning method	Learning method rate	Number of hidden layers	Number of nodes (first layer/second layer)	Motivation function (first layer)	Motivation function (second layer)	R in testing	MSE in testing
1	Momentum	0.3	1	3	Sigmoid	–	0.83	0.014
2	CG	–	2	3/4	Tan	Sigmoid	0.21	0.276
3	LM	–	1	4	Sigmoid	–	0.80	0.013
4	**Momentum**	**0.7**	**2**	**4/4**	**Tan**	**Tan**	**0.87**	**0.0102**
5	Momentum	0.4	2	3/2	Sigmoid	Sigmoid	0.86	0.017
6	DBD	–	2	4	Tan	Tan	0.86	0.011
7	Quick Prop	–	2	4/5	Linear Sigmoid	Sigmoid	0.34	0.026
8	CG	–	1	5	Linear Tan	–	0.86	0.011
9	LM	–	2	4/3	Tan	Linear	0.87	0.057
10	Momentum	0.5	2	3/3	Linear Sigmoid	Tan	0.82	0.181

In both figures, the X and Y axes illustrate the absolute value differences between the reference AHP values and the values predicted by the different neural network models. The Z axis describes the relative performance of the ANN3 and ANN2 models with respect to ANN4. More precisely, we have implemented the following steps in order to define the variables represented in both figures.

- First, we have computed the absolute value differences between the training and testing values derived from the neural network models and those obtained using AHP. As stated above, the set of values used in the calculations is presented in Table 5. We have denoted by |AHP-ANN3| the absolute value difference between the training and testing values obtained using AHP and those derived from the neural network model based on the D, T and P criteria, while |AHP-ANN2| represents the corresponding values derived from the artificial neural network when it is based on D and T. The same intuition applies to the |AHP-ANN4| expression.
- Then, we have sorted the difference values obtained and used the resulting pairs to describe the dispersion generated by these methods among themselves and relative to the AHP reference values.

The intuition defining these figures can be divided in two differentiated effects. The first one implies that the closer the observations are to the origin of the X and Y axes, the lower is the distance between the AHP values and the predicted ones. Note how the training data obtained using ANN4 and presented in the Y axis of Fig. 5 has a lower dispersion relative to AHP than the training values generated by ANN2 and ANN3. Note also how the dispersion with respect to the real values is much lower in Fig. 6, which represents a more concentrated set of data obtained when testing the respective models.

At the same time, the Z axis provides an illustrative measure of how much better is the approximation of ANN4 to AHP relative to those of ANN3 (defined using red circles) and ANN2 (rep-

Table 5
AHP and ANN-predicted values for each supplier based on the different criteria combinations considered.

Supplier number and contract type	AHP score	ANN4 score (Q,D,T and P criteria)	ANN3 score (D,T and P criteria)	ANN2 score (D and T criteria)
		Training		
(18, E)	0.078	0.110201632	0.060032044	0.131344112
(10, A)	0.04	0.070163326	0.095053536	0.104467986
(31, H)	0.08	0.119650905	0.089044796	0.131344112
(22, H)	0.1	0.198004871	0.210383768	0.230134395
(8, B)	0.11	0.116789131	0.087207354	0.131344112
(4, C)	0.13	0.173508092	0.177862568	0.131344112
(8, A)	0.14	0.197004987	0.314723736	0.358860964
(17, E)	0.19	0.214585868	0.234674848	0.172692291
(1, C)	0.21	0.185256176	0.193793564	0.230134395
(9, D)	0.23	0.136492209	0.303178144	0.407525693
(26, F)	0.26	0.288405989	0.401452751	0.441827305
(16, C)	0.34	0.270021582	0.162070019	0.131344112
(23, G)	0.4	0.332805238	0.269899359	0.230134395
(7, E)	0.42	0.364518997	0.296641105	0.296678323
(5, A)	0.52	0.545533686	0.447274148	0.441827305
(13, B)	0.5	0.491720276	0.458457011	0.407525693
(15, D)	0.63	0.565913509	0.462344544	0.441827305
		Testing		
(28, G)	0.04	0.05	0.06986641	0.08751475
(12, C)	0.06	0.058	0.096362186	0.104467986
(6, B)	0.08	0.084	0.07959085	0.131344112
(20, F)	0.1	0.119896	0.110780141	0.172692291
(30, H)	0.12	0.1995478	0.200439271	0.104467986
(5, D)	0.14	0.1245786	0.064167705	0.104467986
(24, G)	0.18	0.1500598	0.369674212	0.441827305
(27, H)	0.19	0.09023659	0.215684466	0.230134395
(5, B)	0.21	0.180096	0.137798197	0.230134395
(11, C)	0.26	0.510235	0.37769023	0.296678323
(21, F)	0.32	0.55563	0.389758622	0.358860964
(3, A)	0.34	0.36014256	0.253737712	0.230134395
(14, E)	0.41	0.3785476	0.396583038	0.358860964
(25, G)	0.46	0.364518997	0.348093569	0.358860964
(19, H)	0.53	0.545533686	0.459920236	0.441827305
(29, F)	0.56	0.557462	0.457219459	0.407525693

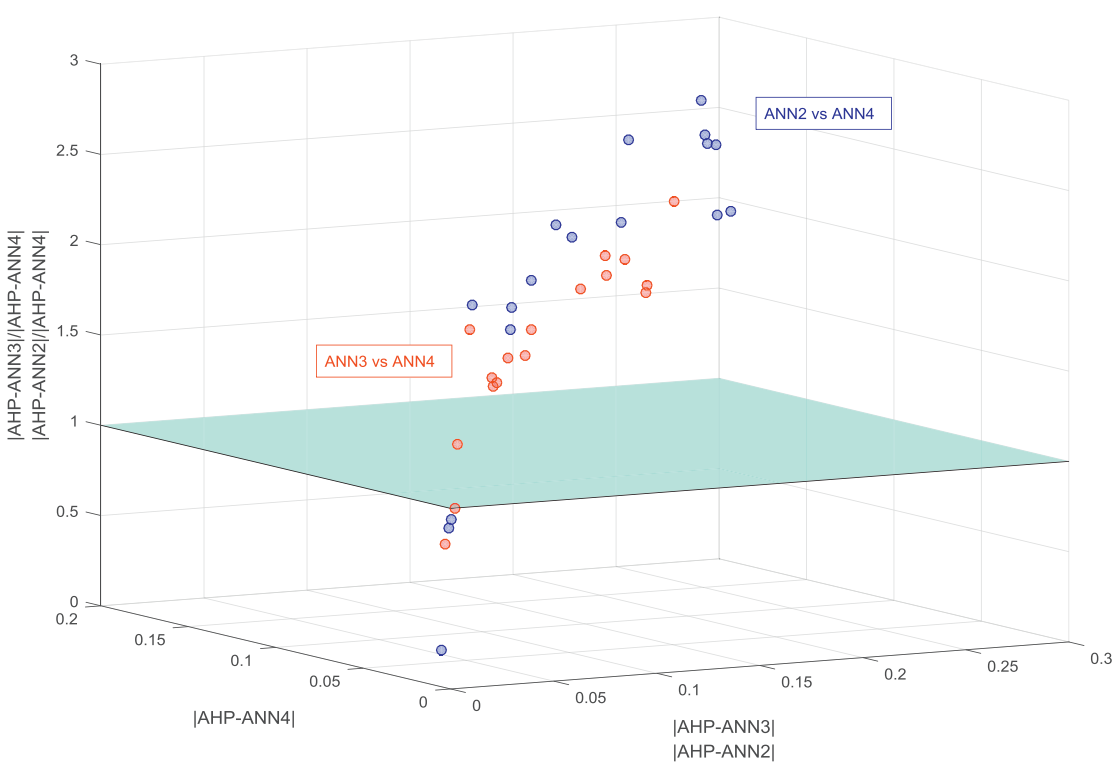

Fig. 5. Training scenario: relative dispersion of the different ANN models with respect to AHP and among themselves.

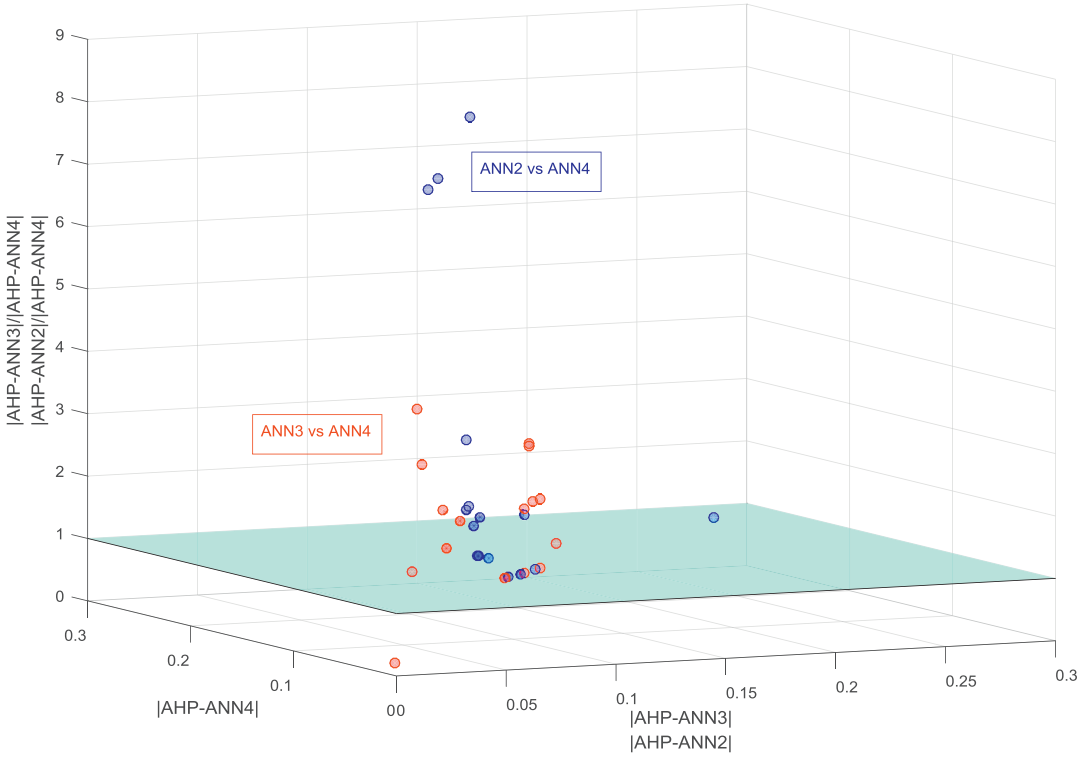

Fig. 6. Testing scenario: relative dispersion of the different ANN models with respect to AHP and among themselves.

resented by blue circles). That is, the Z axis represents the values of $\frac{|AHP-ANN3|}{|AHP-ANN4|}$ and $\frac{|AHP-ANN2|}{|AHP-ANN4|}$. These equations describe the relative dispersion of ANN3 and ANN2 with respect to ANN4 when comparing their respective accuracies. Therefore, whenever the figures illustrate a value located above one, represented by the plane introduced in both figures, the ANN4 model provides a closer approximation to the AHP values than either ANN2 or ANN3. Note also that the ANN3 model performs better than ANN2 in both the training and testing cases, with the blue observations located consistently above the red ones.

It should be highlighted that the computation of the correlation and MSE values that will be performed in the next section to validate statistically the accuracy of the different models is not fully related to the results represented in these figures. However, we have introduced these figures to provide an intuitive description of the relative dispersion generated by the different ANN models when training and testing. The former setting is clearly dominated by ANN4 while the latter presents a less disperse framework, with ANN2 and ANN3 performing better than in the training scenario and ANN3 exhibiting a lower dispersion than ANN2.

5. Performance of the model

This section describes the different statistical tests performed on the results obtained from the proposed ANFIS-ANN model. The model has been evaluated from three different viewpoints. First, a sensitivity analysis was performed to verify the relative importance of the different criteria combinations selected by ANFIS on the performance of the alternatives composing the decision setting. Second, Golbraikh and Tropsha (2002) and Smith (1986) tests were run to evaluate the prediction capacity of the model. Third, the results derived from the proposed model were compared with those obtained using GHD's ANN model. Both results have been compared to illustrate the relative increase in efficiency that follows from implementing ANFIS in the initial stages of the model.

5.1. Sensitivity analysis

A sensitivity analysis has been performed to show that the different combinations of criteria selected by ANFIS have an actual (predictive) influence on the results obtained by ANN-modeling. In such analysis, the input relative to one criterion is excluded from the ANN4 model while the other inputs remain unchanged and their values given by those of the corresponding collected data sets. In this regard, the sensitivity analysis performed by Rezaei and Ortt (2013) aims at identifying the relative importance of the criteria by computing their individual contribution to the performance of the suppliers. The process of analysis is composed of three steps, which adapted to the current setting are defined as follows:

1. Compute the performance of each alternative considering all four criteria, P, Q, D and T. We will denote this value by Y_j, with the subindex $j = 1, \ldots, 33$ accounting for the number of alternatives.
2. Remove the information of the $k-th$ criterion and compute the performance of each alternative based on the three remaining criteria. Denote this value by $Y(k)_j$, with $k = P, Q, D, T$ referring to the criterion removed in the computation. The resulting value will be used as a reference against which to subtract the performance initially obtained with all four criteria.
3. Calculate the average of the absolute value difference between the performances obtained when criterion k is removed and those computed in the first step

$$\bar{\Delta}_k = \frac{\sum_k |Y(k)_j - Y_j|}{33}, \text{ with } k = P, Q, D, T \tag{15}$$

The variable $\bar{\Delta}_k$ describes the contribution of the $k-th$ criterion to the performance of the alternatives and can be interpreted as the relative importance of the corresponding criterion (Rezaei & Ortt, 2013). For example, in order to calculate the relative importance of prices (P), we have removed the price variable from the ANN4 model and run the MLP based only on Q, D and T for

Table 6
Contribution of the P, Q, D and T criteria to the performance of the suppliers.

$\bar{\Delta}_P$	$\bar{\Delta}_Q$	$\bar{\Delta}_D$	$\bar{\Delta}_T$
0.067361038	0.063239446	0.058297375	0.048076314

Table 7
Statistic indicators from the regressions on the results provided by each ANN model.

	Training			Testing		
	ANN4	ANN3	ANN2	ANN4	ANN3	ANN2
R	0.959	0.815	0.680	0.865	0.861	0.818
MSE	0.002	0.006	0.010	0.010	0.006	0.006

all alternatives. Then, we have computed the average of the absolute value differences between the new performance value of each alternative and the ones obtained using the MLP based on P, Q, D and T. The resulting values of $\bar{\Delta}_k$, which increase in the importance of the criterion, are presented in Table 6.

Note that the difference between the contribution of the most important criterion, P, and the less important one, T, is relatively small. Thus, all the criteria can be assumed to have a similar effect on the performance of the suppliers. That is, as explained by Rezaei and Ortt (2013), when the differences in contribution across inputs (criteria) are not substantial, then all of them can be assumed to have a similar importance on the performance of the alternatives. For instance, the testing performance of the model is not significantly modified when shifting from ANN4 to ANN3, despite the fact that the second most influential criterion (Q) is removed in the process. Indeed, as illustrated in Table 6, the difference in contribution between the second (Q) and the third (D) criterion is not significant.

However, the elimination of the next criterion (P) has a stronger negative effect on the testing performance of the model. That is, the ANN3 model retains the explanatory capacity of ANN4 but ANN2 exhibits a weaker performance. This is not a surprising outcome, since ANFIS had already provided some intuition in this direction. In this regard, the introduction of ANFIS in the early stages of the model allows us to select the different combinations of criteria with higher explanatory capacity. This property is particularly relevant since the chosen combinations of criteria determine the ability of the MLP to replicate the performance of the alternatives. Thus, as shown in the current paper, ANFIS can be used to complement and improve the performance of MLP.

The results obtained imply that ANFIS can be considered a reliable meta-heuristic for input selection, that is, it shortens the time of modeling, decreases the complexity of the modeling process and identifies the most important criteria. Moreover, as we will see in the next subsection, the performance of ANN3 is equivalent to that of ANN4 in the testing section of the model. Therefore, ANN3 can be used as a valid testing approximation to ANN4 in cases where a limited amount of information is available on a particular supplier.

Finally, we will show in the following subsection that even though the performance of ANN2 is inferior to that of ANN3 in the testing section of the model, its correlation coefficient and MSE value suffice to consider ANN2 as a valid testing approximation to ANN4 when a substantial subset of data is unavailable on a particular supplier.

5.2. Smith and Golbraikh & Tropsha tests for evaluating the performance of the model in prediction

We describe now the different statistical tests run to evaluate the performance of the ANN models and validate the proposed ANFIS-ANN model. These tests have been implemented using SPSS software.

Smith (1986) recommended the following (R-value-based) statistical criterion for assessing the performance of a predictive model:

a. If a model delivers $|R| > 0.8$, a strong correlation exists between the predicted and real (reference) values.
b. If a model delivers $0.2 < |R| < 0.8$, a correlation exists between the predicted and real values.
c. If a model delivers $|R| < 0.2$, a weak correlation exists between the predicted and real values.

In all cases, the error values (i.e. MSE) should be minimal for the model to have predictive ability (Mostafavi, Mostafavi, Jaafari, & Hosseinpour, 2013). Table 7 presents the numerical values obtained for the statistic indicators used to evaluate the significance of the results provided by the different ANN models. In particular, Table 7 illustrates that the training section of the evaluation process is clearly dominated by ANN4, which exhibits a higher correlation coefficient and a lower MSE than both ANN3 and ANN2.

However, the testing section presents quite different results, with ANN3 providing almost the same correlation level as ANN4 but a lower MSE. That is, the slightly better performance of ANN4 is compensated by the higher precision of ANN3. This result could be intuitively observed when ANFIS was implemented in the initial stages of our evaluation model, since the criteria composing both the ANN3 and ANN2 models performed considerably better than the four criteria of ANN4 in the testing stage. The combination of criteria defining ANN4 exhibited a superior training performance, but its testing performance was considerably inferior to those of all the other criteria combinations analyzed.

At the same time, even though ANN2 performs slightly worse than ANN3 in the testing section, it also provides a lower MSE than ANN4. Figs. 7 and 8 illustrate the dispersion of the values predicted

Table 8
Statistical factors of the external validation decision model*.

Item	Formula	Condition	Testing		
			ANN4	ANN3	ANN2
1	R	$R > 0.8$	0.865	0.861	0.818
2	$k = \frac{\sum_{\alpha=1}^{\Delta}(r_\alpha \times t_\alpha)}{\sum_{\alpha=1}^{\Delta} r_\alpha^2}$	$0.85 < k < 1.15$	0.870228	0.992544	1.003037
3	$k' = \frac{\sum_{\alpha=1}^{\Delta}(r_\alpha \times t_\alpha)}{\sum_{\alpha=1}^{\Delta} t_\alpha^2}$	$0.85 < k' < 1.15$	1.053682	0.929435	0.897128
4	$m = \frac{R^2 - R_o^2}{R^2}$	$m < 0.1$	−0.227501	−0.244946	−0.343064
5	$n = \frac{R^2 - R_{o'}^2}{R^2}$	$n < 0.1$	−0.227501	−0.244946	−0.343064
Where	$R_o^2 = 1 - \frac{\sum_{\alpha=1}^{\Delta}(t_\alpha - r_\alpha^o)^2}{\sum_{\alpha=1}^{\Delta}(t_\alpha - \bar{t})^2}$,	$r_\alpha^o = k \times t_\alpha$	0.916943	0.922505	0.899853
	$R_{o'}^2 = 1 - \frac{\sum_{\alpha=1}^{\Delta}(r_\alpha - t_\alpha^o)^2}{\sum_{\alpha=1}^{\Delta}(r_\alpha - \bar{r})^2}$,	$t_\alpha^o = k' \times r_\alpha$	0.916943	0.922505	0.899853

* r_α is the actual output of supplier α and t_α is the predicted output of supplier α.

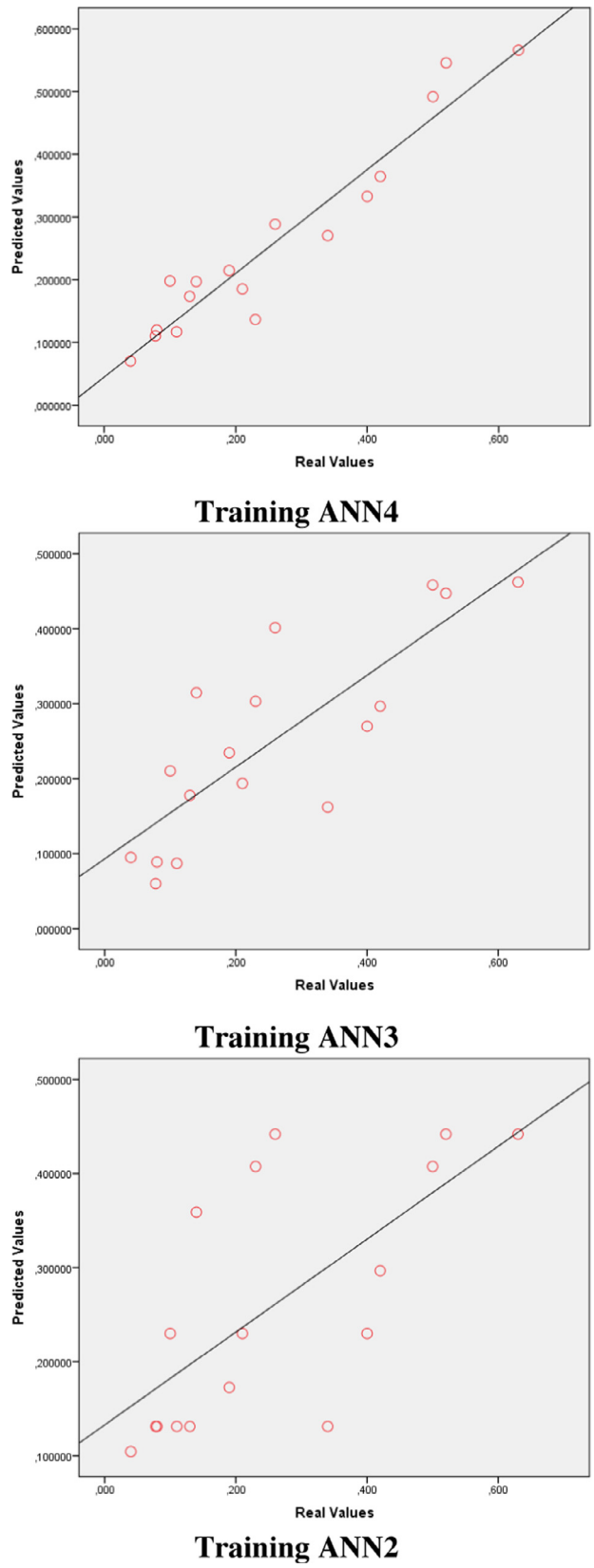

Fig. 7. Regression lines for each one of the ANN models analyzed: training section.

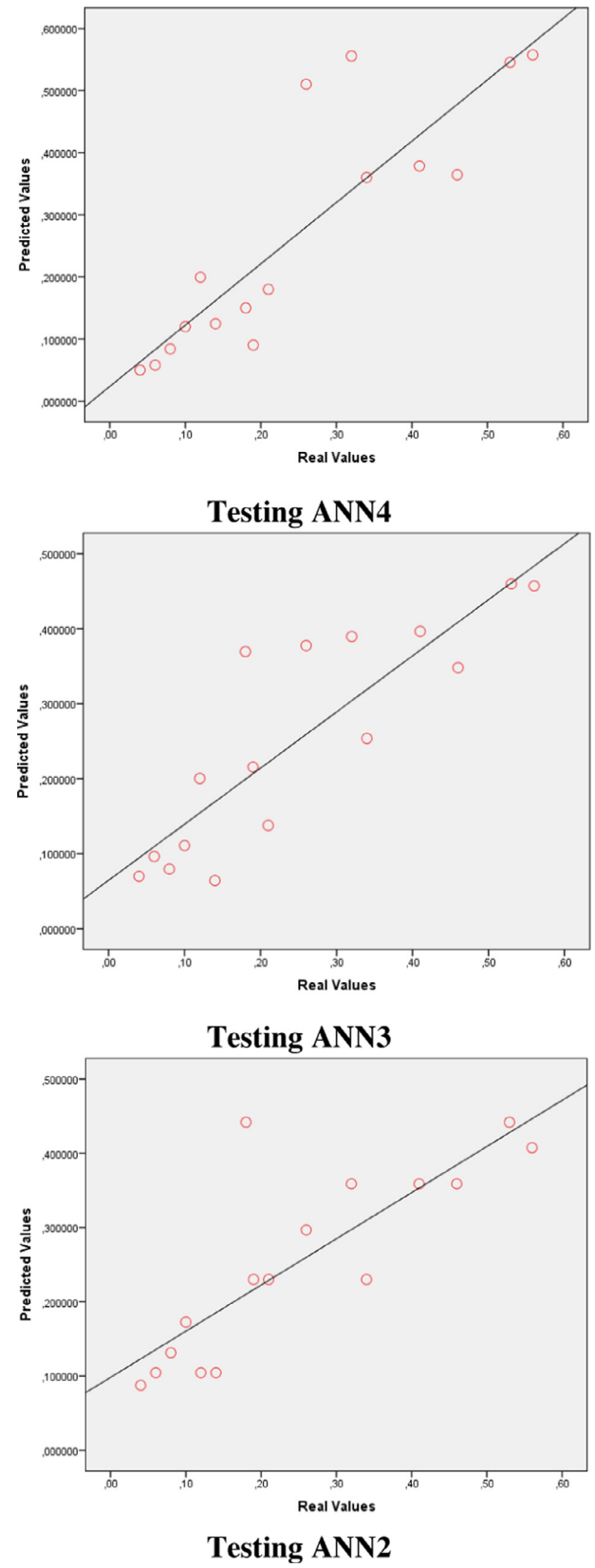

Fig. 8. Regression lines for each one of the ANN models analyzed: testing section.

relative to the regression line for each one of the different ANN models in the training and testing sections, respectively. The wider dispersion of the values predicted by the ANN3 and ANN2 models relative to the more accurate predictions of ANN4 in the training section can be directly observed in Fig. 7. However, as can be observed in Fig. 8, the superiority of the ANN4 model vanishes in the testing section, where both ANN3 and ANN2 exhibit lower MSEs and the former almost the same correlation value.

Thus, ANN3 and ANN2 could be considered as viable alternatives to ANN4 *when having a limited amount of information or data on a particular supplier*. Moreover, the testing results obtained from all the ANN models illustrate the importance of ANFIS in selecting a smaller set of data that improves upon larger amounts of criteria being used to decide among a set of alternatives.

The results derived from the remaining statistical tests performed on the ANN models are presented in Table 8. This table shows the statistic indicators employed as external validation criteria and the corresponding values obtained for the proposed models. In particular, we have used the statistic indicators recommended by Golbraikh and Tropsha (2002) to perform an external validation of the proposed models on the testing data sets. These authors recommend that at least one slope of the regression lines through the origin (k or k') is close to 1 (Mollahasani, Alavi, & Gandomi, 2011). Here, k and k' denote the slopes of the regression lines through the origin when plotting the regression of actual output (r_α) against predicted output (t_α) and that of predicted output (t_α) against actual output (r_α), i.e. $r_\alpha = kt_\alpha$ and $t_\alpha = k'r_\alpha$, respectively.

Moreover, either the squared correlation coefficient between the actual and predicted values (R_o^2), or the one between the predicted and actual values ($R_{o'}^2$) should be close to R^2 and to 1 (Alavi et al., 2011; Mostafavi et al., 2013; Mostafavi, Mousavi, & Hosseinpour, 2014). In addition, the value of the performance indexes m and n defined in Table 8 should be less than 0.1. Note how the k and k' criteria of Golbraikh and Tropsha as well as the constraints imposed by the m and n performance indexes are satisfied by all the ANN models. Thus, the validation phase ensures that the proposed ANFIS-ANN model is strongly suitable and accurate.

5.3. Comparing results with those obtained by Golmohammadi's ANN model

GHD collected a complete data set (inputs and outputs) using AHP pairwise comparisons and proposed a MLP-ANN model to predict the suppliers' performance. In order to improve the model, he defined an input function for each criterion (refer to Section 7 in GHD). As stated in Subsection 4.1, we have used the improved data set generated by Golmohammadi to implement the proposed ANFIS-ANN model. This allows for a direct comparison between our hybrid model and GHD's ANN-based one, illustrating the increase in accuracy achieved by the former.

That is, the R and MSE values obtained by GHD are equal to 0.773 and 0.193, respectively. As the R and MSE values described in Table 7 illustrate, all the ANFIS-ANN models perform better than that of GHD in the testing section. The higher R values and the lower values obtained for the MSE prove that the results delivered by the ANFIS-ANN model are in general more accurate than those offered by GHD's ANN-based model.

Note that, in this case study, eliminating up to three out of five criteria can lead to an improvement in the predictive capacity of the model with respect to the MLP considered by GHD. That is, adding variables to the simulation does not necessarily improve the performance of the MLP. More importantly, the performance and explanatory capacity of an ANN model can be guaranteed and even improved when dealing with missing data.

As described in Subsection 5.1, the four evaluation criteria contribute similarly to the performance of the alternatives. Thus, the selection of potential combinations of criteria that can improve upon the MLP of GHD while reducing its computational complexity of the model becomes a trial-and-error exercise. In this regard, ANFIS has been introduced at the beginning of our evaluation process to complement and improve the performance of the MLP model.

6. Applicability of the hybrid ANFIS-MLP model to supplier selection problems

Both methods, MLP and ANFIS, can be directly implemented using add-ins for standard software packages such as Excel and MATLAB. Thus, if a company decides to implement an ANN architecture such as MLP to help with its supplier selection problem, it should be able to include an AI-based model such as ANFIS within the corresponding decision process. This last remark introduces the discussion regarding the complexity of formal models and their applicability by companies, which has been around since Hall (1990).

The literature has illustrated how the improvement in the user-friendliness of formal models has increased their applicability to industrial decision problems in the latter years (Bowen & Hinchey, 2012; de Boer & Van der Wegen, 2003; Jeffery, Staples, Andronick, Klein, & Murray, 2015). However, companies may be reticent to implement formal AI models that help them replicating ranking results which can be directly obtained via AHP. This is particularly the case if the performance of the corresponding AI model is not completely accurate.

At the same time, companies may have to deal with a subset of missing observations from several potential suppliers. ANN-based models can be used both to replicate the performance of the suppliers being considered and to simulate the behavior of those with missing observations while constrained by the incomplete information available. Formal models such as ANFIS provide an alternative to intuition and trial-and-error when selecting the main combinations of criteria required to simulate the behavior of suppliers accurately.

Finally, as emphasized by Bruno et al. (2016), the supplier selection process of companies operating in low-complexity sectors relies on few criteria such as price, delivery time, and quality. Consequently, the resulting ranking is generally based on the experience and intuition of the corresponding decision maker. On the other hand, as the complexity of the product increases, the process of supplier selection becomes a multi-criteria problem that must be formalized and adapted to the specific context of analysis. This implies that different products require different criteria and weights in order to perform an adequate evaluation of the suppliers. As a result, companies operating in an industry such as the automotive one will tend to have different ranking lists. Our model can be applied in such a context so as to allow companies to complete their evaluations when information is missing on a subset of criteria from a particular supplier within one (or more) of these lists.

7. Conclusion

This study has proposed a new hybrid fuzzy multi-criteria decision making model combining the ANFIS and ANN approaches to solve criteria selection and alternatives' ranking problems. The proposed model deals with fuzzy data, which are common in real life situations, and consists of four main phases:

a. determining the criteria used to analyze the performance of the alternatives;
b. gathering a complete (input and output) data set through AHP;
c. identifying the most influential criteria on the performance of the alternatives using ANFIS;

d. ranking the alternatives based on their ANN performance.

A case study has been presented to illustrate the main steps of the model and to verify its accuracy in prediction. The alternatives of the case study are the suppliers of a company in the automotive industry. The data set used to implement the model is the same one used by GHD to run his ANN-based model. Both a sensitivity analysis and several statistical tests have been performed to validate the model and compare it to the ANN-based model proposed by GHD.

The ANFIS-ANN model proposed in this paper has been applied to the supplier evaluation and selection problem but it can be used to assist managers and, more in general, decision makers, in any evaluation process with fuzzy data where understanding and learning about the relationships between inputs and outputs are the key to an accurate solution. In other words, the proposed model takes advantage of the feed forward quality of both ANFIS and ANN to provide a useful predictive framework for assessing alternatives and facilitating the decision making process in numerous real life situations, such as those characterized by incomplete information on a subset of criteria from a given alternative.

References

Admuthe, L. S., & Apte, S. (2010). Adaptive neuro-fuzzy inference system with subtractive clustering: A model to predict fibre and yarn relationship. *Textile Research Journal, 80*, 841–846.

Alavi, A. H., Aminian, P., Gandomi, A. H., & Esmaeili, M. A. (2011). Genetic-based modeling of uplift capacity of suction caissons. *Expert Systems with Application, 38*, 12608–12618.

Alavi, A. H., & Gandomi, A. H. (2011). A robust data mining approach for formulation of geotechnical engineering systems. *Engineering Computations, 28*, 242–274.

Aoki, M. (1988). *Information, incentives, and bargaining in the Japanese economy*. New York: Cambridge University Press.

Asadi, S., Hadavandi, E., Mehmanpazir, F., & Nakhostin, M. M. (2012). Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowledge-Based Systems, 35*, 245–258.

Asanuma, B. (1989). Manufacturer-supply relationships in Japan and the concept of relation-specific skill. *Journal of the Japanese and International Economies, 3*, 1–30.

Azadeh, A., Saberi, M., & Anvari, M. (2011). An integrated artificial neural network fuzzy C-means-normalization algorithm for performance assessment of decision-making units: The cases of auto industry and power plant. *Computers & Industrial Engineering, 60*, 328–340.

Bektas Ekici, B., & Aksoy, U. T. (2011). Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications, 38*, 5352–5358.

Bessant, J., Kaplinsky, R., & Lamming, R. (2003). Putting supply chain learning into practice. *International Journal of Operations and Production Management, 23*, 167–184.

Bhutta, M. K. S. (2003). Supplier selection problem: Methodology literature review. *Journal of International Technology and Information Management, 12*, 53–72.

Blois, K. (1972). Vertical quasi-integration. *Journal of Industrial Economics, 20*, 37–51.

Bowen, J. P., & Hinchey, M. (2012). Ten commandments of formal methods … Ten years on. In M. Hinchey, & L. Coyle (Eds.), *Conquering complexity* (pp. 237–251). London: Springer.

Bruno, G., Esposito, E., Genovese, A., & Passaro, R. (2012). AHP-based approaches for supplier evaluation: Problems and perspectives. *Journal of Purchasing and Supply Management, 18*, 159–172.

Bruno, G., Esposito, E., Genovese, A., & Simpson, M. (2016). Applying supplier selection methodologies in a multi-stakeholder environment: A case study and a critical assessment. *Expert Systems with Applications, 43*, 271–285.

Büyüközkan, G., & Çifçi, G. (2012a). A novel hybrid MCDM approach based on fuzzy DEMATEL, fuzzy ANP and fuzzy TOPSIS to evaluate green suppliers. *Expert Systems with Applications, 39*, 3000–3011.

Büyüközkan, G., & Çifçi, G. (2012b). Evaluation of the green supply chain management practices: A fuzzy ANP approach. *Production Planning & Control, 23*, 405–418.

Cakir, L., & Yilmaz, N. (2014). Polynomials, radial basis functions and multilayer perceptron neural network methods in local geoid determination with GPS/levelling. *Measurement, 57*, 148–153.

Cannavacciuolo, L., Iandoli, L., Ponsiglione, C., & Zollo, G. (2015). Knowledge elicitation and mapping in the design of a decision support system for the evaluation of suppliers' competencies. *VINE, 45*, 530–550.

Chai, J., Liu, J. N. K., & Ngai, E. W. T. (2013). Application of decision-making techniques in supplier selection: A systematic review of literature. *Expert Systems with Applications, 40*, 3872–3885.

Choy, K. L., Lee, W., & Lo, V. (2003). Design of an intelligent supplier relationship management system: A hybrid case based neural network approach. *Expert Systems with Applications, 24*, 225–237.

Colombo, M. G., & Mariotti, S. (1998). Organizing vertical markets: The Italtel case. *European Journal of Purchasing and Supply Management, 4*, 7–19.

Day, M., Fawcett, S. E., Fawcett, A. M., & Magnan, G. M. (2013). Trust and relational embeddedness: Exploring a paradox of trust pattern development in key supplier relationships. *Industrial Marketing Management, 42*, 152–165.

Day, M., Lichtenstein, S., & Samouel, P. (2015). Supply management capabilities, routine bundles and their impact on firm performance. *International Journal of Production Economics, 164*, 1–13.

de Boer, L., Labro, E., & Morlacchi, P. (2001). A review of methods supporting supplier selection. *European Journal of Purchasing & Supply Management, 7*, 75–89.

de Boer, L., & Van der Wegen, L. L. M. (2003). Practice and promise of formal supplier selection: A study of four empirical cases. *Journal of Purchasing and Supply Management, 9*, 109–118.

Esposito, E., & Passaro, R. (2009). The evolution of supply chain relationships: An interpretative framework based on the Italian inter-industry experience. *Journal of Purchasing and Supply Management, 15*, 114–126.

Esposito, E., & Raffa, M. (1994). The evolution of Italian subcontracting firms: Empirical evidence. *European Journal of Purchasing and Supply Management, 1*, 67–76.

Frohlich, M. T., & Westbrook, R. (2001). Arcs of integration: An international study of supply chain strategies. *Journal of Operations Management, 19*, 185–200.

Gandomi, A. H., & Alavi, A. H. (2011). Applications of computational intelligence in behavior simulation of concrete materials. In X.-S. Yang, & S. Koziel (Eds.), *Computational optimization and applications in engineering and industry* (pp. 221–243). Springer.

Golbraikh, A., & Tropsha, A. (2002). Beware of q2!. *Journal of Molecular Graphics and Modelling, 20*, 269–276.

Golmohammadi, D. (2011). Neural network application for fuzzy multi-criteria decision making problems. *International Journal of Production Economics, 131*, 490–504.

Gressgård, L. J., & Hansen, K. (2015). Knowledge exchange and learning from failures in distributed environments: The role of contractor relationship management and work characteristics. *Reliability Engineering & System Safety, 133*, 167–175.

Guneri, A. F., Ertay, T., & YüCel, A. (2011). An approach based on ANFIS input selection and modeling for supplier selection problem. *Expert Systems with Applications, 38*, 14907–14917.

Guneri, A. F., Yucel, A., & Ayyildiz, G. (2009). An integrated fuzzy-lp approach for a supplier selection problem in supply chain management. *Expert Systems with Applications, 36*, 9223–9228.

Ha, H. S., & Krishnan, R. (2008). A hybrid approach to supplier selection for the maintenance of a competitive supply chain. *Expert Systems with Applications, 34*, 1303–1311.

Hall, A. (1990). Seven myths of formal methods. *Software, IEEE, 7*, 11–19.

Heckmann, I., Comes, T., & Nickel, S. (2015). A critical review on supply chain risk – Definition, measure and modeling. *Omega, 52*, 119–132.

Ho, S.-Y., Lee, K.-C., Chen, S.-S., & Ho, S.-J. (2002). Accurate modeling and prediction of surface roughness by computer vision in turning operations using an adaptive neuro-fuzzy inference system. *International Journal of Machine Tools and Manufacture, 42*, 1441–1446.

Hohenstein, N.-O., Feisel, E., Hartmann, E., & Giunipero, L. (2015). Research on the phenomenon of supply chain resilience: A systematic review and paths for further investigation. *International Journal of Physical Distribution & Logistics Management, 45*, 90–117.

Houssiaux, J. (1957). Le Concept de "Quasi-intégration" et le Role des Sous-traitants dans l'industrie. *Revue Economique*, 221–247.

Jang, J.-S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man and Cybernetics, 23*, 665–685.

Jang, J.-S. R. (1996). Input selection for ANFIS learning. In *Proceedings of the fifth IEEE international conference on fuzzy systems* (pp. 1493–1499).

Jeffery, R., Staples, M., Andronick, J., Klein, G., & Murray, T. (2015). An empirical research agenda for understanding formal methods productivity. *Information and Software Technology, 60*, 102–112.

Klein, B., Crawford, R., & Alchian, A. (1978). Vertical integration, appropriable rents and the competitive contracting process. *The Journal of Law and Economics, 21*, 297–326.

Labib, A. W. (2011). A supplier selection model: A comparison of fuzzy logic and the analytic hierarchy process. *International Journal of Production Research, 49*, 6287–6299.

Lee, C. C., & Ou-Yang, C. (2009). A neural networks approach for forecasting the supplier's bid prices in supplier selection negotiation process. *Expert Systems with Applications, 36*, 2961–2970.

Mirzahosseini, M. R., Aghaeifar, A., Alavi, A. H., Gandomi, A. H., & Seyednour, R. (2011). Permanent deformation analysis of asphalt mixtures using soft computing techniques. *Expert Systems with Applications, 38*, 6081–6100.

Mollahasani, A., Alavi, A. H., & Gandomi, A. H. (2011). Empirical modeling of plate load test moduli of soil via gene expression programming. *Computers and Geotechnics, 38*, 281–286.

Mostafavi, E. S., Mostafavi, S. I., Jaafari, A., & Hosseinpour, F. (2013). A novel machine learning approach for estimation of electricity demand: An empirical evidence from Thailand. *Energy Conversion and Management, 74*, 548–555.

Mostafavi, E. S., Mousavi, S. M., & Hosseinpour, F. (2014). Gene expression programming as a basis for new generation of electricity demand prediction models. *Computers & Industrial Engineering, 74*, 120–128.

Opengart, R. (2015). Supply chain management and learning organization: A merging of literatures. *International Journal of Commerce and Management, 25*, 183–195.

Opricovic, S., & Tzeng, G.-H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research, 156*, 445–455.

Parthiban, P., Zubar, H. A., & Katakar, P. (2013). Vendor selection problem: A multi--criteria approach based on strategic decisions. *International Journal of Production Research, 51*, 1535–1548.

Rezaei, J., & Ortt, R. (2013). Supplier segmentation using fuzzy logic. *Industrial Marketing Management, 42*, 507–517.

Saaty, T. L. (1980). *The analytic hierarchy process*. New York: McGraw-Hill.

Sako, M. (1992). *Price, quality and trust inter-firm relations in Britain & Japan*. Cambridge: Cambridge University Press.

Smith, G. N. (1986). *Probability and statistics in civil engineering: An introduction*. London: Collins.

Stevens, J. (1989). Integrating the supply chain. *International Journal of Physical Distribution & Materials Management, 19*, 3–8.

Vahdani, B., Iranmanesh, S., Mousavi, S. M., & Abdollahzade, M. (2012). A locally linear neuro-fuzzy model for supplier selection in cosmetics industry. *Applied Mathematical Modelling, 36*, 4714–4727.

Weber, C. A., Current, J. R., & Benton, W. C. (1991). Vendor selection criteria and methods. *European Journal of European Research, 50*, 2–18.

Wu, D. (2009). Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications, 36*, 9105–9112.

Wu, J.-D., Hsu, C.-C., & Chen, H.-C. (2009). An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference. *Expert Systems with Applications, 36*, 7809–7817.

Wu, D. D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert Systems with applications, 31*, 108–115.