GESTÃO & PRODUÇÃO

ISSN 0104-530X (Print)
ISSN 1806-9649 (Online)

ORIGINAL ARTICLE

Comparison of artificial neural networks learning methods to evaluate supply chain performance

Comparação entre métodos de aprendizagem de redes neurais artificiais aplicados à avaliação de desempenho de cadeias de suprimentos

Antonio Ricardo Lunardi[1], Francisco Rodrigues Lima Junior[2]*

[1]Universidade Tecnológica Federal do Paraná – ITFPR, Departamento Acadêmico de Eletrotécnica, Grupo Tomada de Decisão em Gestão de Operações, Curitiba, PR, Brasil. E-mail: lunardi@alunos.utfpr.edu.br

[2]Universidade Tecnológica Federal do Paraná – UTFPR, Programa de Pós-Graduação em Administração – PPGA, Grupo Tomada de Decisão em Gestão de Operações, Curitiba, PR, Brasil. E-mail: eng.franciscojunior@gmail.com

How to cite: Lunardi, A. R., Lima Junior, F. R. Comparison of artificial neural networks learning methods to evaluate supply chain performance. *Gestão & Produção*, 28(3), e5450. https://doi.org/10.1590/1806-9649-2021v28e5450

Abstract: The supply chain performance evaluation is a critical activity to continuously improve operations. Literature presents several performance evaluation systems based on multi-criteria methods and artificial intelligence. Among them, the systems based on artificial neural networks (ANN) excel due to their capacity of modeling non-linear relationships between metrics and allowing adaptations to a specific environment by means of historical performance data. These systems' accuracy depend directly on the adopted training algorithm, and no studies have been found that assess the efficiency of these algorithms when applied to supply chain performance evaluation. In this context, the present study evaluates four ANNs learning methods in order to investigate which one is the most adequate to deal with supply chain evaluation. The algorithms tested were Gradient Descendent Momentum, Levenberg-Marquardt, Quasi-Newton and Scale Conjugate Gradient. The performance metrics were extracted from SCOR®, which is a reference model used worldwide. The random sub-sampling cross-validation method was adopted to find the most adequate topological configuration for each model. A set of 80 topologies was implemented using MATLAB®. The prediction accuracy evaluation was based on the mean square error. For the four level 1 metrics considered, the Levenberg-Marquardt algorithm provided the most precise results. The results of correlation analysis and hypothesis tests reinforce the accuracy of the proposed models. Furthermore, the proposed computational models reached a prediction accuracy higher than previous approaches.

Keywords: Artificial neural networks; Supervised learning methods; Supply chain performance evaluation; SCOR® model; Multilayer perceptron.

Resumo: A avaliação de desempenho de cadeias de suprimentos é uma atividade crítica para a melhoria contínua das operações. A literatura apresenta diversos sistemas de avaliação de desempenho baseados em métodos multicritério e técnicas de inteligência artificial. Dentre esses, os sistemas baseados em redes neurais se destacam por sua capacidade de modelar relacionamentos não lineares entre as métricas e por permitirem a adaptação ao ambiente de uso por meio de dados históricos de desempenho. Embora a acurácia desses sistemas dependa diretamente do algoritmo de aprendizagem adotado, não são encontrados estudos que avaliem o

Received Apr. 24, 2019 - Accepted Jan 22, 2020
Financial support: None.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

desempenho destes algoritmos quando aplicados nesse domínio de problema. Nesse contexto, o presente estudo avalia quatro métodos de aprendizagem de redes neurais com o objetivo de investigar qual deles é mais adequado para apoiar a avaliação de cadeias de suprimentos. Foram testados os algoritmos *Gradient Descendent Momentum*, *Levenberg-Marquardt*, *Quasi-Newton* e *Scale Conjugate Gradient*. As métricas de desempenho foram extraídas do SCOR®, um modelo de referência mundialmente utilizado. O método de validação cruzada com amostragem aleatória foi adotado para encontrar a configuração topológica mais adequada para cada modelo. Um conjunto de 80 topologias foi implementado usando MATLAB. A avaliação da acurácia de predição foi baseada no erro quadrático médio. Para os quatro indicadores de nível 1 considerados, o algoritmo *Levenberg-Marquardt* forneceu resultados mais precisos. Os resultados da análise de regressão e do coeficiente de correlação ressaltam a eficácia dos modelos propostos. Ademais, os modelos computacionais propostos alcançaram acurácia superior às abordagens anteriores.

Palavras-chave: Redes neurais artificiais; Métodos de aprendizagem supervisionada; Avaliação de desempenho de cadeias de suprimentos; Modelo SCOR®; *Perceptron* multicamada.

1 Introduction

Mentzer et al. (2001) define supply chain management as “the strategic and systematic coordination of business traditional functions and tactical actions in a company and through its businesses along the chain,” aimed at enhancing the long-term performance of member companies. Supply chain management involves finance flow, services, goods, information and interorganizational relationships. Considering this, collaborative management tends to generate a sinergy condition, in which the entire supply chain becomes more efficient (Mentzer et al. 2001; Shafiee et al. 2014).

Many studies emphasize the relevance of measuring supply chain management performance as a way of planning development and managing strategies (Marchand & Raymond, 2008; Estampe et al., 2013). Supply chain performance evaluation includes many factors that work together in order to achieve certain goals. Thus, it demands the usage of intra and inter organizational processes, as well as updated, integrated, and easily accessible data for decision making. Some benefits from supply chain management are the effective monitoring of results, improvements in understanding key processes, identification of potential problems, and the perception to formulate future improvement actions. However, there are many factors that make supply chain management a difficult task. Commonly there are hindrances such as decentralized historical data, as well as the fact that many of the existing performance metrics do not have well-defined causal relationships.

The literature about supply chain management has studies that propose models for supply chain performance evaluation based on qualitative (Gunasekaran et al., 2001) and quantitative approaches (Akkawuttiwanich & Yenradee, 2018). There are also studies that present systematic reviews of the literature (Maestrini et al., 2017), analysis of metrics adopted for supply chain performance evaluation (Ahi & Searcy, 2015) and of some existing models (Estampe et al., 2013). Over the last decade researchers have developed growing interest in quantitative models of supply chain performance evaluation. Dozens of methods are being tested, including multi-criteria decision-making (MCDM), mathematical programming and artificial intelligence (AI) techniques. Despite AI techniques being an emerging tendency and less frequent in the literature, they excel by presenting new evaluation model capabilities.

Among these models two approaches based on artificial neural networks distinguish themselves from MCDM models by permitting the usage of non-linear relationships between elements of input and output. Furthermore, they are able to adapt themselves to a specific environment by using historical performance data with a supervisioned training algorithm.

Fan et al. (2013) proposed a supply chain evaluation system using a combination of Balanced Scorecard with multilayer perceptron neural networks. Lima Jr. & Carpinetti applied neural networks to predict the level 1 SCOR® metrics (Supply Chain Operation Reference). SCOR® is a reference model of supply chain management widely adopted by practitioners worldwide. Fan et al. (2013) adopted the Levenberg-Marquardt training algorithm while Lima-Junior & Carpinetti (2019) applied a backpropagation algorithm instead.

The development of tools based on artificial neural networks involves the choice of a topological configuration and an adequate training algorithm. It requires performing a series of empirical tests and may become time-consuming and costly (Tkác & Verner, 2016). The learning method directly affects the accuracy of predictions and the network training time (Mukherjee & Routroy, 2012). Thus, comparative studies among learning methods are necessary to identify the ones that show best performances for certain application types. Moreover, they can help researchers and analysts in the creation of smart solutions to support supply chain management, in order to guide the solution development process and make it more agile. However, after researching in the main data basis and analysing literature review studies (Maestrini et al., 2017; Lima-Junior & Carpinetti, 2017), comparative studies among learning methods applied on supply chain performance evaluation were not found.

Considering this context, the present study evaluates four supervised learning methods of artificial neural networks in order to find which is the most adequate to support supply chain performance evaluation. Since causal relationships are well-defined, a set of performance metrics proposed by SCOR® was adopted as input and output variables for the neural network models. It is important to note that this study continues the work of Lima Junior & Carpinetti (2019), by testing other learning methods in order to achieve better accuracy. Regarding the structure of this paper it goes as follows: section 1 is the introduction; section 2 focuses on SCOR® model; section 3 explains the work of ANNs; section 4 presents the methodological procedures; section 5 discusses the results of the computational implementation of the ANN models; section 6 shows the hypothesis tests results, and section 7 presents the conclusion and suggestions for further studies.

2 SCOR® Model

The SCOR® model was developed by the Supply Chain Council, a non-profitable organization of supply chain professionals. It is a pioneer for its inter-enterprise framework to evaluate and make improvements in supply chain management processes (SCC, 2012). The SCOR® model is subdivided into four sections: metrics, processes, practices, and people. “Metrics” introduces standard metrics to describe the processes’s performance and define strategic goals. “Processes” determines a process's structure of management and describes the relationships between these processes. “Practices” suggests management practices that result in performance levels significantly improving. “People” addresses the required abilities to execute supply chain procedures (SCC, 2012).

The management processes suggested by SCOR® are plan, source, make, deliver, return, and enable. They integrate the different tiers of a supply chain. Each process has performance metrics associated with it, which permit the ability to monitor and optimize these metrics based on a comparison between the achieved performance results and the goals defined for each metric (Akkawuttiwanich & Yenradee, 2018). The SCOR® section on performance evaluation has two categories: attribute and metrics. An attribute is a group of indicators to express a particular strategy. A metric is a standard to measure the performance of a supply chain or process. SCOR® proposes five performance attributes:

reliability, responsiveness, costs, agility, and assets. Reliability refers to the ability to execute tasks according to expectations. Responsiveness measures the speed that tasks are done. Costs assesses the operation costs from supply chain processes. Agility consists of the response ability to external stimulus and the change based on these stimulus. Asset is the ability of efficiently using assets (SCC, 2012, Dissanayake & Cross, 2018).

Figure 1 shows the suggested attributes by SCOR® as well as its level 1 and 2 respective metrics. The measures of different hierarchical levels have quantifiable cause and effect relationships, which makes it possible to predict the metric values of a superior level based on the metrics of the immediate lower level. Thereby the level 3 metrics can be used to predict the level 2 metrics, while the level 2 metrics can be applied to predict level 1 metric values. This characteristic contributes to explain why the SCOR® metrics is frequently adopted in quantitative models for supply chain performance evaluation. SCOR® does not recommend that a focus-company use all the suggested metrics but gives priority to the ones that are critical for success, based on the need to implement data collection mechanisms (SCC, 2012).

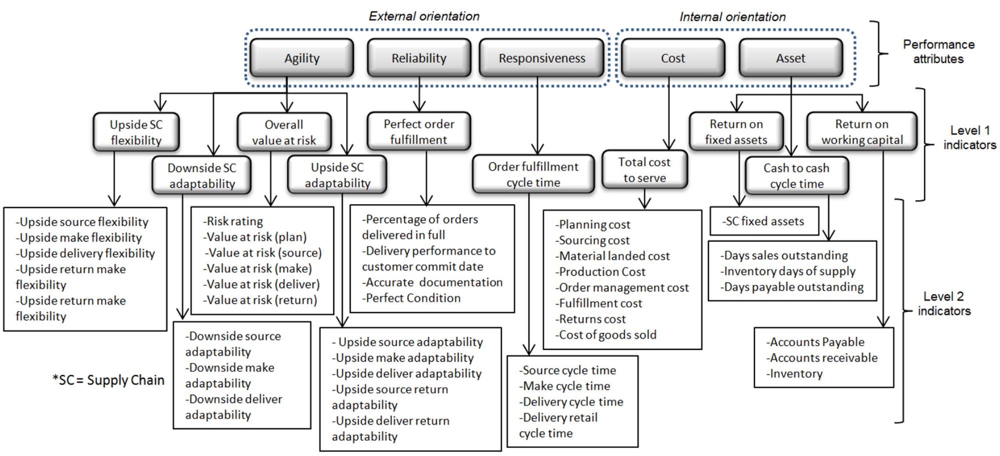

Figure 1. Attributes and metrics of performance suggested by SCOR®. Source: Adapted from Supply Chain Council (SCC, 2012) and Lima-Junior & Carpinetti (2019).

Chart 1 displays the SCOR® techniques used in studies to propose quantitative models for supply chain performance evaluation. Even though these models have made several contributions to the literature on supply chain performance evaluation, the adopted techniques have some limitations and difficulties. In the case of approaches based on pairwise comparison as proposed by Clivillé & Berrah (2012), Yang & Jiang (2012), Kocaoğlu et al. (2013), Bukhori et al. (2015), Sellitto et al. (2015) and Dissanayake & Cross (2018), the greater the metrics and supply chain considered in the evaluation, the greater the difficulty in ensuring data consistency. Another problem of the models based on multicriteria methods (Golparvar & Seifbarghy, 2009; Kocaoğlu et al., 2013; Moharamkhani et al., 2017; Akkawuttiwanich & Yenradee, 2018) is that they generate an output value based on a weighted linear combination of input values. Thus, these values are not suitable to deal with causal non-linear relationships between metrics. Only the models based on AI techniques have this capability. However, the difficulty in using models based on fuzzy inference (Ganga & Carpinetti, 2011) refers to the necessity of parameterizing and manually updating hundreds of decision rules based on specialist opinions, in order to adjust the causal relationships between metrics. Therefore, among all found models, only the ones based on ANN are capable of making automatic adjustments to the adaptive parameters using historical performance data.

Chart 1. Techniques used in quantitative models for performance evaluation based on SCOR®.

Authors	Method(s)
Golparvar & Seifbarghy (2009)	TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution)
Ganga & Carpinetti (2011)	Mamdani Inference Fuzzy System
Jalalvand et al. (2011)	DEA (Data Envelopment Analysis) and PROMETHEE II
Clivillé & Berrah (2012)	MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique)
Yang & Jiang (2012)	New Method based on Fuzzy Numbers and M(1,2,3)
Kocaoğlu et al. (2013)	AHP (Analytic Hierarchy Process) and TOPSIS
Bukhori et al. (2015)	AHP
Sellitto et al. (2015)	AHP
Moharamkhani et al. (2017)	Interval-valued Fuzzy TOPSIS
Liu & Liu (2017)	DEA
Akkawuttiwanich & Yenradee (2018)	Fuzzy QFD (Quality Function Deployment)
Dissanayake & Cross (2018)	AHP and Structural Equation Modelling
Lima-Junior & Carpinetti (2019)	Multilayer Perceptron Networks

Souce: Author.

3 Multilayer perceptron neural networks

Artificial neural networks (ANN) are intelligent systems of distributed processing that imitate neural biological systems (Kurtgoz et al., 2017). According to the review study developed by Tkác & Verner (2016), multilayer perceptron (MLP) is the most used type of ANN. MLP networks can be applied to several kinds of problems, such as function approximation, standard recognition, and prediction. As shown in Figure 2, a MLP is constituted of an input layer, one or more hidden layers and an output layer (Abdi-Khanghah et al., 2018). Each layer has processing basic units called neurons; this structure is illustrated in Figure 2. The connections between the neurons have different weights. Initial values from these parameters are given randomly, which are then modified by the network training process. Each neuron has a bias that helps to enhance the accuracy of the results (Kurtgoz et al., 2017).

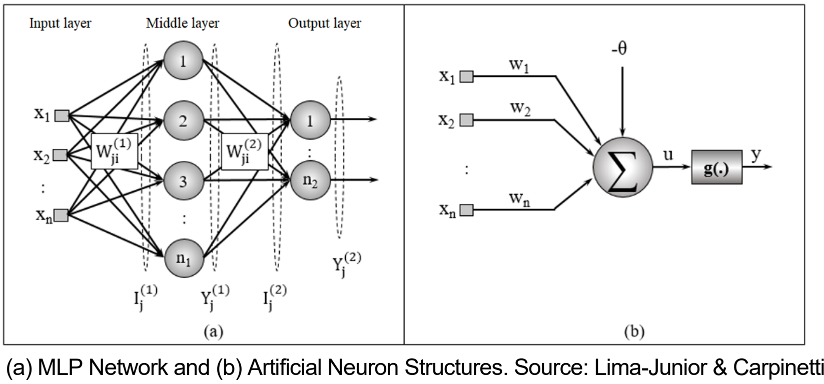

Figure 2. (a) MLP Network and (b) Artificial Neuron Structures. Source: Lima-Junior & Carpinetti (2019).

In Figure 2 all input signals ($x_1, x_2,..., x_n$) are represented as are the matrices of synaptic weights (W_{ji}^L) that link the neurons (j) of each layer (L) to their predecessor layer (i). Also the weighted inputs (I_j^L) from the neurons and the outputs produced by them (Y_j^L) are highlighted. Network training is traditionally made by using a learning algorithm called backpropagation, which is applied in two steps. This training process requires a set of samples that are subdivided in training samples and validation samples. The recommended quantity for the training is from 60% up to 90% of the samples. These samples are processed by the network in a number of times called an epoch. An epoch can be a criterion to stop the training process (Silva et al., 2010; Rezaee et al., 2018).

In the forward step of the backpropagation algorithm, the input signals (x_i) are weighted by the weights of the middle layer $W_{ji}^{(1)}$. After that, this input vector is modified as in Equation 1, by an activation function, such as the hyperbolic tangent represented in Equation 2, which generates the vector $I_j^{(1)}$ values. The procedures that are made in the posterior layers are similar. However, in these cases, the input signals from these layers refer to the outputs from the previous layers (Silva et al., 2010; Rezaee et al., 2018).

$$u = \sum_{i=1}^{n} w_i \cdot x_i - \theta \tag{1}$$

$$g(u) = \frac{1 - e^{-\beta u}}{1 + e^{-\beta u}} \tag{2}$$

In the backward step, the results generated by the network for each sample are compared to the respective output value of the training subset (expected values). The backpropagation algorithm’s main objective consists of finding the synaptic weights’ optimized values and biases to minimize the mean square error (MSE) resultant of the difference between the expected outputs and the predicted values. An adjustment of these parameters is made based on this difference in order to minimize the error. This adjustment of parameters begins with the output layer and follows to the middle layer. The process is repeated until a number of epochs are reached. At the end of the training process, the parameters are tuned determining a quantitative relationship between the output and input variables (Bilgehan, 2011).

In order to carry out the training process and select the most suitable network topology for each model, several studies apply a procedure known as cross-validation method, which consists of a set of empirical tests (Tkác & Verner, 2016; Rezaee et al., 2018). In each test, many combinations of values are tried for the network parameters in order to choose the one that results with a lesser MSE in the validation step. This procedure is also frequently applied to evaluate the accuracy of learning methods in order to select the most suitable one (Silva et al., 2010).

3.1 Training algorithms

There is a wide range of learning methods that can be applied to carry out the supervisioned training of MLP networks. In order to improve the performance of the original version of the backpropagation algorithm, new algorithms have been proposed to make the training faster and to reach higher prediction accuracies. Some of the most applied algorithms are described in this topic and were adopted in the present study: *Gradient Descendent Momentum* (GDM), *Levenberg-Marquardt* (LM), *Quasi-Newton*

(BFGS), and *Scale Conjugate Gradient* (SCG). The main difference among them is the parameter direction adjustment and the magnitude of this adjustment.

In the GDM algorithm, equation 3 is applied to tune the weights and biases, in which η is the learning rate. The value of α, named momentum coefficient, is an adjustable parameter that defines the magnitude of iterative tunings. The local gradient $\delta_j^{(L)}$ is defined for the *j*-th neuron of the output layer, as in equation 4. In the LM algorithm, the adjustment is made with the gradient calculated by equation 5. The parameter μ is the tuning rate of convergence. $J(W)$ represents a jacobian matrix (second order derivative matrix), and $J^T(W)$ is its transposed version. I is the identity matrix (Silva et al., 2010).

$$W_{ji}^{(L)}(t+1) = W_{ji}^{(L)}(t) + \alpha \cdot \left(W_{ji}^{(L)}(t) - W_{ji}^{(L)}(t-1) \right) + \eta \cdot \delta_j^{(L)} \cdot Y_i^{(L-1)} \tag{3}$$

$$\delta_j^{(L)} = \left(d_j - Y_j^{(L)} \right) . g' \left(I_j^{(L)} \right) \tag{4}$$

$$\Delta W = \left(J^T(W) . J(W) + \mu . I \right)^{-1} . J^T(W) . \left(d_j - Y_i^{(L)} \right) \tag{5}$$

In the case of the BFGS algorithm, the tuning is based on equation 6, considering $\nabla^2 J(.)$ is a hessian matrix and a_t is a scalar that defines the magnitude of tuning adjustment intensity. The algorithm SCG applies the gradient shown in equation 7, in which d_t establishes the tuning direction (Mukherjee & Routroy, 2012).

$$W(t+1) = W(t) - a_t . \left[\nabla^2 J(W(t)) \right]^{-1} . \nabla J(W(t)) \tag{6}$$

$$W(t+1) = W(t) + a_t . d_t \tag{7}$$

Literature presents comparative studies among training algorithms considering different problem domains. Tripathy & Kumar (2009) developed a comparative study aiming to find the most adequate algorithm to predict the temperature variation of ailment products in solar drying. In this study, SCG attained a better accuracy than LM and BFGS. In an application on control of grinding processes, Mukherjee & Routroy (2012) analyzed the algorithms BFGS and LM and concluded that the first converges faster and is more accurate. Maroufpoor et al. (2019) compared GDM, SCG and LM. They concluded that LM is the most suitable to deal with the modeling of uniform water distribution. All these studies prove that the performance of each training algorithm depends on its application. Thus, development of comparative studies among learning methods is needed to determine which one provides better accuracy when applied to supply chain performance evaluation.

4 Research method

The method of research adopted in this study may be classified as modeling and computational simulation, in view of the fact that it uses computational ANNs modeling that has causal relations between input and output variables (Bertrand & Fransoo, 2002). The first stage of the research was a bibliographic review about supply chain evaluation, ANN, and supervisioned training algorithms. Research papers were collected from the data basis *Web of Science, Emerald Insight, Scopus, Springer, Taylor & Francis*, and *IEEE-Xplore* using combinations of the strings "*supply chain performance evaluation,*" "*supply chain performance measurement,*" "*neural networks,*" "*learning method,*" "*training algorithm,*" and "SCOR."

The literature review allowed us to identify the research gap and support the stage of modeling and computational simulation. In this stage, the samples of training and validation were created with MS Excel. Based on the procedure proposed by Lima-Junior & Carpinetti (2019), level 2 metrics were randomly generated and posteriorly normalized in the interval [0,1]. Level 1 metrics were obtained through the expressions suggested by Supply Chain Council (SCC, 2012). The modeling, training, and validation of computational models were done with MATLAB® (nntool toolbox). Following Silva et al. (2010), the random sub-sampling cross-validation method was applied to implement and evaluate the candidate topologies and learning training.

The prediction accuracy of the models was measured through the mean square error (MSE). It was calculated in the validation step based on the difference between the estimated value and the expected value for each level 1 metric. Additionally, Pearson's correlation coefficient and linear regression tests were calculated. Lastly, hypothesis tests with paired samples were done to investigate if there were significant differences between the expected and the predicted values to each network topology chosen.

5 Results and discussion

Figure 3 shows the architecture of the proposed system for the supply chain performance evaluation, which was developed based on Supply Chain Council (SCC, 2012) and Lima-Junior & Carpinetti (2019), in order to carry out this comparative study. The system is composed of four computational models based MLP neural networks. The input variables are defined by the level 2 SCOR® metrics, while the output variables refer to the level 1 SCOR® metrics. Chart 2 describes briefly these metrics. More details about these metrics can be consulted in the SCOR® model (SCC, 2012). The architecture shown in Figure 3 was used to evaluate comparatively the accuracy of four training algorithms. Therefore, to select the most accurate topology, 20 different configurations were tested on each MLP model for a total of 80 computational models.

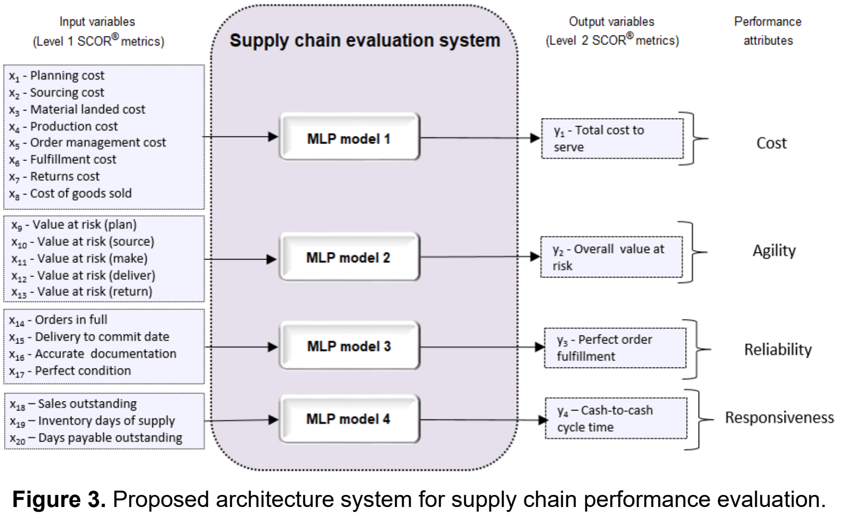

Figure 3. Proposed architecture system for supply chain performance evaluation.
Source: Proposed by authors

Chart 2. Metrics that compose the performance evaluation system description

Variables		Description
Model 1	x_1	***Sourcing cost:*** the total cost associated with managing the ordering, receiving, inspecting and warehousing of materials, products, merchandise and services
	x_2	***Planning cost***: the total costs of personnel, automation, assets and overhead associated with supply chain planning processes
	x_3	***Material landed cost***: the total cost associated with buying and making purchased materials, products or merchandise available to the location of use (location-of-use)
	x_4	***Production cost:*** the total cost associated with the production managing and performing processes
	x_5	***Order management cost***: the total cost of personnel, automation and assets responding to inquiries and quotes, order entry and maintenance, transportation scheduling, order tracking and tracing, delivery, installation, and invoicing
	x_6	***Fulfillment cost***: the total cost of personnel, automation, assets, and overhead cost associated with fulfillment orders
	x_7	***Returns cost***: the disposition cost from returned materials due to planning errors, supplier quality problems, delivery, or production
	x_8	***Cost of goods sold***: the cost of direct materials, labor force and general costs related to production or acquiring finished products
	y_1	***Total cost to serve***: the sum of the supply chain costs to deliver products and services to customers
Model 2	x_9	***Value at risk (plan):*** the sum of the monetized risks related to the process "*plan*"
	x_{10}	***Value at risk (source):*** the sum of the monetized risks related to the process "*source*"
	x_{11}	***Value at risk (make):*** the sum of the monetized risks related to the process "*make*"
	x_{12}	***Value at risk (deliver):*** the sum of the monetized risks related to the process "*deliver*"
	x_{13}	***Value at risk (return):*** the sum of the monetized risks related to the process "*return*"
	y_2	***Overall value at risk:*** the sum of the occurrences of risk probability that may affect the chain processes multiplied by the monetary impact of these occurrences.
Model 3	x_{14}	***Orders delivered in full:*** the percentage of orders where all of the items are received by customers in the quantities committed.
	x_{15}	***Delivery performance to customer commit date***: the percentage of orders that are fulfilled on the customer's originally committed date.
	x_{16}	***Documentation accuracy***: the percentage of orders with on time and accurate documentation supporting the order, including packing slips, bills of lading, invoices, etc.
	x_{17}	***Perfect condition***: the percentage of orders delivered in an undamaged state that meet specification, have the correct configuration, are faultlessly installed (as applicable) and accepted by the customer.
	y_3	***Perfect order fulfillment***: the percentage of orders meeting delivery performance with complete and accurate documentation and no delivery damage. Components include all items and quantities on-time using the customer's definition of on-time, and documentation - packing slips, bills of lading, invoices, etc.
Model 4	x_{18}	***Days sales outstanding***: the length of time from when a sale is made until cash for it is received from customers. The amount of sales outstanding is expressed in days.
	x_{19}	***Inventory days of supply***: the amount of inventory (stock) expressed in days of sale.
	x_{20}	***Days payable outstanding***: the length of time from purchasing materials, labor and/or conversation resources until cash payments must be made: expressed in days.
	y_4	***Cash-to-cash cycle time***: the time it takes for an investment to flow back into a company after it has been spent for raw materials.

Source: Based on Supply Chain Council (SCC, 2012).

In the interest of evaluating the candidate network topologies and the training algorithms, the random sub-sampling cross-validation method was applied through the following steps (Silva et al., 2010): 1) random division of the samples into subsets of training and validation; 2) definition of the candidate topology (number of neurons in the middle layer and type of activation function) parameters; 3) choose the training algorithms and parameter values; 4) execute the training processes aimed at tuning the weights and bias; 5) validate the topologies using an error measure based on the difference between the values predicted by the network and the output values of the validation subset; 6) select the candidate topology that presents the smallest error in the validation stage. If no topology accomplishes satisfactory accuracy, the procedure needs to restart and define new candidate topologies and training parameters until the desired accuracy level is reached.

5.1 Definition of topological configuration and training parameters

The candidate topologies were defined based on the variation in the number of neurons in the middle layer and the learning algorithms. For each MLP model the following algorithms were tested: GDM (*Gradient Descent Momentum*), LM (*Levenberg-Marquardt*), BFGS (*Quasi-Newton*) e SCG (*Scale Conjugate Gradient*). These algorithms were chosen based on Tkác & Verner (2016) and Mathworks (2018) who point out adequate algorithms for function approximation applications. The training parameters of the algorithms LM and GDM were chosen by performing many empirical tests. For BFGS and SCG, the suggested values of MATLAB® were used. The size of the training subset was determined by means of empirical tests. For each of the four models, 500 samples were generated with 350 applied for training and 150 for validation. Following Bilgehan (2011), the number of epochs defined was 20,000.

Chart 3 shows the topologies tested using GDM and LM algorithms. Chart 4 presents the candidate topologies using BFGS and SCG. As proposed for Patuwo et al. (1993), the number of tested neurons in the middle layers was determined according to the quantity of input variables in each MLP model. Therefore, considering n the number of input variables, the following quantities of neurons were tested in the middle layer: $n-2, n-1, n, n+1 e n+2$. Based on Lima-Junior & Carpinetti (2019), hyperbolic tangent was adopted in the middle layer and linear function in the output layer. It is important to highlight that these authors concluded that hyperbolic tangents present better results when compared to other alternative functions for evaluating performance of level 1 metrics.

Chart 3. Candidate topologies for each model (Gradient Descent Momentum and Levenberg-Marquardt).

MLP model and quantity of input variables (*n*)	Candidate topology	Number of neurons in the middle layer	Training algorithm
1 – Total cost to serve (n=8)	1	6	GDM
	2	7	GDM
	3	8	GDM
	4	9	GDM
	5	10	GDM
	6	6	LM
	7	7	LM
	8	8	LM
	9	9	LM
	10	10	LM

Chart 3. Continued…

MLP model and quantity of input variables (*n*)	Candidate topology	Number of neurons in the middle layer	Training algorithm
2 – Value at risk (n=5)	11	3	GDM
	12	4	GDM
	13	5	GDM
	14	6	GDM
	15	7	GDM
	16	3	LM
	17	4	LM
	18	5	LM
	19	6	LM
	20	7	LM
3 – Perfect order fulfillment (n=4)	21	2	GDM
	22	3	GDM
	23	4	GDM
	24	5	GDM
	25	6	GDM
	26	2	LM
	27	3	LM
	28	4	LM
	29	5	LM
	30	6	LM
4 - Cash-to-Cash Cycle Time (n=3)	31	1	GDM
	32	2	GDM
	33	3	GDM
	34	4	GDM
	35	5	GDM
	36	1	LM
	37	2	LM
	38	3	LM
	39	4	LM
	40	5	LM

Source: Proposed by authors.

Chart 4. Candidate topologies for each model (Quasi-Newton and Scale Conjugate Gradient).

MLP model and quantity of input variables (*n*)	Candidate topology	Number of neurons in the middle layer	Training algorithm
1 – Total cost to serve (n=8)	41	6	BFGS
	42	7	BFGS
	43	8	BFGS
	44	9	BFGS
	45	10	BFGS
	46	6	SCG
	47	7	SCG
	48	8	SCG
	49	9	SCG
	50	10	SCG

Chart 4. Continued…

MLP model and quantity of input variables (*n*)	Candidate topology	Number of neurons in the middle layer	Training algorithm
2 – Value at risk (n=5)	51	3	BFGS
	52	4	BFGS
	53	5	BFGS
	54	6	BFGS
	55	7	BFGS
	56	3	SCG
	57	4	SCG
	58	5	SCG
	59	6	SCG
	60	7	SCG
3 – Perfect order fulfillment (n=4)	61	2	BFGS
	62	3	BFGS
	63	4	BFGS
	64	5	BFGS
	65	6	BFGS
	66	2	SCG
	67	3	SCG
	68	4	SCG
	69	5	SCG
	70	6	BFGS
4 - Cash-to-Cash Cycle Time (n=3)	71	1	BFGS
	72	2	BFGS
	73	3	BFGS
	74	4	BFGS
	75	5	BFGS
	76	1	BFGS
	77	2	BFGS
	78	3	SCG
	79	4	SCG
	80	5	SCG

Source: Proposed by authors.

5.2 The learning process results

Tables 1 and 2 present the MSE values obtained in the validation stage, as well as the correlation coefficient R, calculated using the predicted values and the expected values for each level 1 metric. Among all implemented models, the smallest error ($2.8761 \cdot 10^{-34}$) was reached by MLP 3 using an LM algorithm with 5 neurons in the middle layer (topology 28). This result is probably due to the fact that the input variables are binary values (0 or 1), which implies a very simple output function, formed by five discrete positions. Among all selected topologies, the smallest accuracy ($7,2260 \cdot 10^{-3}$) was reached by MLP 1 using a GDM algorithm, with nine neurons in the middle layer (topology 4). It is important to notice that this model has eight input variables providing the function with more complex mapping.

Table 1. MSE and R for the evaluated topologies using GDM and LM algorithms.

Model	Topology number	MSE	R
1 – Total cost to serve	1	1.7277×10^{-4}	0.99998
	2	7.7338×10^{-4}	0.99991
	3	3.2612×10^{-4}	0.99995
	4	7.2260×10^{-3}	0.99921
	5	2.3676×10^{-4}	0.99997
	6	1.7868×10^{-14}	1
	7	4.8606×10^{-16}	1
	8	5.3279×10^{-14}	1
	9	$\mathbf{4.6739 \times 10^{-16}}$	**1**
	10	3.9927×10^{-15}	1
2 – Value at risk	11	7.6484×10^{-4}	0.99998
	12	3.2225×10^{-5}	0.99999
	13	1.4334×10^{-5}	0.99999
	14	1.4722×10^{-4}	0.99989
	15	2.9898×10^{-4}	0.99993
	16	5.2190×10^{-18}	1
	17	4.4490×10^{-17}	1
	18	1.9933×10^{-17}	1
	19	$\mathbf{4.1473 \times 10^{-18}}$	**1**
	20	8.4310×10^{-17}	1
3 – Perfect order fulfillment	21	8.1722×10^{-5}	0.99932
	22	4.2627×10^{-5}	0.99951
	23	3.0120×10^{-6}	0.99997
	24	2.1076×10^{-6}	0.99998
	25	1.2399×10^{-6}	0.99999
	26	4.3141×10^{-33}	1
	27	2.8761×10^{-34}	1
	28	$\mathbf{1.2326 \times 10^{-34}}$	**1**
	29	2.8761×10^{-34}	1
	30	2.0954×10^{-33}	1
4 - Cash-to-cash cycle time	31	6.2712×10^{-5}	0.99951
	32	1.8256×10^{-4}	0.99859
	33	7.8624×10^{-5}	0.99948
	34	3.0968×10^{-4}	0.99779
	35	6.4169×10^{-5}	0.99956
	36	4.5336×10^{-13}	1
	37	2.5943×10^{-16}	1
	38	2.6543×10^{-16}	1
	39	$\mathbf{9.3256 \times 10^{-18}}$	**1**
	40	9.8038×10^{-18}	1

Source: Proposed by authors.

Table 2. MSE and R for the evaluated topologies using BFGS and SCG algorithms.

Model	Topology number	MSE	R
1 – Total cost to serve	41	7.9847×10^{-15}	1
	42	$\mathbf{1.1622 \times 10^{-15}}$	**1**
	43	2.2807×10^{-14}	1
	44	2.5033×10^{-14}	1
	45	2.6726×10^{-14}	1
	46	2.4602×10^{-8}	1
	47	2.4359×10^{-7}	1
	48	2.3544×10^{-7}	1
	49	1.9213×10^{-6}	1
	50	1.0869×10^{-6}	1

Table 2. Continued…

Model	Topology number	MSE	R
2 – Value at risk	51	1.6209×10^{-13}	1
	52	1.9715×10^{-14}	1
	53	1.2598×10^{-14}	1
	54	$\mathbf{4.2228 \times 10^{-15}}$	**1**
	55	1.2472×10^{-14}	1
	56	8.6542×10^{-8}	1
	57	4.2247×10^{-8}	1
	58	2.1301×10^{-8}	1
	59	1.3591×10^{-8}	1
	60	1.5838×10^{-8}	1
3 – Perfect order fulfillment	61	6.7315×10^{-22}	1
	62	8.7578×10^{-23}	1
	63	2.9078×10^{-21}	1
	64	2.1911×10^{-13}	1
	65	4.6783×10^{-12}	1
	66	1.4719×10^{-25}	1
	67	1.0689×10^{-29}	1
	68	9.9189×10^{-26}	1
	69	$\mathbf{3.0556 \times 10^{-31}}$	**1**
	70	3.0918×10^{-31}	1
4 - Cash-to-cash cycle time	71	3.8212×10^{-11}	1
	72	3.8505×10^{-13}	1
	73	1.5034×10^{-13}	1
	74	$\mathbf{1.6149 \times 10^{-14}}$	**1**
	75	5.1497×10^{-14}	1
	76	4.6171×10^{-9}	1
	77	1.6531×10^{-9}	1
	78	4.4888×10^{-9}	1
	79	1.1387×10^{-9}	1
	80	7.7939×10^{-9}	1

Source: Proposed by authors.

For the MLP models 1, 2 and 4, the best accuracy was reached by the topologies 9 ($4,6739 \times 10^{-16}$), 19 ($4,1473 \times 10^{-18}$) and 39 ($9,3256 \times 10^{-18}$), using 9, 6 and 4 neurons in the middle layer, respectively. Thereby, it is concluded that Levenberg-Marquardt achieved the best accuracy for all level 1 metrics considered in this study. Hence, this algorithm is the most adequate to be applied on SCOR® based performance evaluation among those tested.

Figure 4 shows the linear regression analysis results with the correlation coefficient R for each topology. The horizontal axis represents the expected outputs (targets) and the vertical axis shows the values obtained by each topology. In all cases a perfect positive correlation rate was reached between the output values of the validation subset and the predicted MLP proposed models. Furthermore, in all equations that define two data sets' relationship, the angular coefficient is equivalent to 1, while the linear coefficient is close to zero. These results reinforce the accuracy prediction of the proposed models, as well as the adequacy of the Levenberg-Marquardt algorithm to approach supply chain evaluation based on SCOR® metrics.

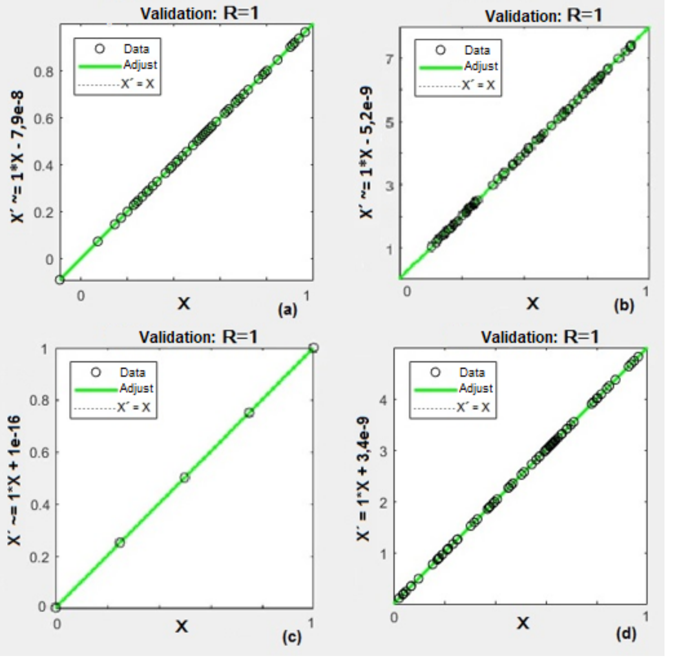

Figure 4. Regression analysis and R for MLP 1(a), 2(b), 3(c) and 4(d). Source: Proposed by authors.

6 Validation of results using the hypothesis tests

The four hypothesis tests were performed in order to verify if there is a significant difference between the expected performance values (calculated based on SCOR®) and the ones that were estimated using the LM algorithm. The tests were conducted using t-test with paired samples, which is adequate when the observations of two populations are collected in a paired way. The mean of populations 1 and 2 are respectively μ_1 and μ_2. The difference of each pair is $D_j = X_j - Y_j$, being $j = 1, 2, ..., n$. The paired t-test procedure consists of analyzing if the difference between the means (μ_D) of two populations results in a specific value Δ_0. If there is no significant difference between the two populations, so the difference of the means must be zero ($\mu_D = \Delta_0 = 0$). Therefore, as shown in Chart 5, for a significance test level α, the null hypothesis is given by H_1: $\mu_D \neq 0$. The alternative hypothesis is represented by H_1: $\mu_D \neq 0$. It is worth noting that in the statistic test T_0, the μ_D parameter is estimated by the sample mean of the differences ($\bar{D}$). For testing the rejection criterion the tabulated value $t_{\alpha/2,\ n-1}$ should be considered (Montgomery & Runger, 2009). The significance $\alpha = 0.05$ was adopted in all tests.

Chart 5. Analyzed hypothesis, test statistic and rejection criterion of null hypothesis.

Null hypothesis $H_0 : \mu_D = \Delta_0$
Alternative hypothesis $H_1 : \mu_D \neq \Delta_0$
Region of rejection: $t_0 > t_{\alpha/2,n-1}$ ou $t_0 < t_{\alpha/2,n-1}$
Test statistic: $T_0 = \frac{D - \Delta_0}{S_D / \sqrt{n}}$, being $S_D = \sqrt{\frac{\Sigma_{i=1}^{n}\left(D_i - D\right)^2}{n-1}}$

Source: Montgomery & Runger (2009).

Table 3 shows the expected values of the 30 samples and the predicted values by each model. Due to the space limitations of this paper, the presented values on this table were limited to five decimal places. However, for the calculations, all decimal places of the predicted values were considered (17 places).

Table 3. Sample values used in hypothesis tests.

MLP model 1		MLP model 2		MLP model 3		MLP model 4	
Expected value	**Predicted value**	**Expected value**	**Predicted value**	**Expected value**	**Predicted value**	**Expected value**	**Predicted value**
0.00000	0.00000	0.00000	0.00000	0.50000	0.50000	1.00000	1.00000
1.00000	1.00000	1.00000	1.00000	0.25000	0.25000	-0.09091	-0.09091
0.48500	0.48500	0.59800	0.59800	0.25000	0.25000	0.74545	0.74545
0.80750	0.80750	0.65000	0.65000	0.25000	0.25000	0.76364	0.76364
0.58000	0.58000	0.87000	0.87000	0.50000	0.50000	0.23636	0.23636
0.76130	0.76125	0.79600	0.79600	1.00000	1.00000	0.62727	0.62727
0.64630	0.64625	0.32000	0.32000	0.25000	0.25000	0.42727	0.42727
0.18000	0.18000	0.30600	0.30600	0.50000	0.50000	0.71818	0.71818
0.22500	0.22500	0.40800	0.40800	0.25000	0.25000	0.92727	0.92727
0.29630	0.29625	0.16200	0.16200	0.50000	0.50000	0.42727	0.42727
0.10130	0.10125	0.12400	0.12400	0.75000	0.75000	0.14545	0.14545
0.95750	0.95750	0.89600	0.89600	0.50000	0.50000	0.46364	0.46364
0.59750	0.59750	0.38000	0.38000	0.50000	0.50000	0.06364	0.06364
0.70880	0.70875	0.61000	0.61000	0.50000	0.50000	0.48182	0.48182
0.37250	0.37250	0.41600	0.41600	0.25000	0.25000	0.63636	0.63636
0.71380	0.71375	0.76600	0.76600	0.50000	0.50000	0.87273	0.87273
0.59750	0.59750	0.80400	0.80400	0.50000	0.50000	0.93636	0.93636
0.30880	0.30875	0.39200	0.39200	0.25000	0.25000	0.67273	0.67273
0.28880	0.28875	0.00000	0.00000	0.25000	0.25000	0.79091	0.79091
0.28000	0.28000	1.00000	1.00000	0.75000	0.75000	0.41818	0.41818
0.12630	0.12625	0.04600	0.04600	0.75000	0.75000	0.50000	0.50000
0.82500	0.82500	0.94800	0.94800	0.00000	0.00000	0.23636	0.23636
0.62130	0.62125	0.53800	0.53800	0.50000	0.50000	0.22727	0.22727
0.68630	0.68625	0.64200	0.64200	0.75000	0.75000	0.28182	0.28182
0.36000	0.36000	0.60600	0.60600	0.50000	0.50000	0.39091	0.39091
0.79500	0.79500	0.79000	0.79000	0.75000	0.75000	0.56364	0.56364
0.53880	0.53875	0.56800	0.56800	0.50000	0.50000	0.93636	0.93636
0.32630	0.32625	0.37800	0.37800	0.25000	0.25000	0.43636	0.43636
0.46750	0.46750	0.20600	0.20600	0.75000	0.75000	0.70000	0.70000
0.19000	0.19000	0.21000	0.21000	0.25000	0.25000	0.10909	0.10909

Source: Proposed by authors.

Table 4 displays the results of the hypothesis tests for the four MLP models. In this table, $\bar{D}$ is the distribution mean of the differences and S_D is the standard deviation. In all cases, the p-value is bigger than the significance (α) adopted for the test. Moreover, all the values of T_0 are outside the region of rejection of the null hypothesis. These results demonstrate that the null hypothesis cannot be rejected, which indicated that there is no significant difference between the expected values and the predicted values for each level 1 metric. Thus, it confirms that the LM algorithm is suitable to deal with supply chain performance evaluation based on level 1 SCOR® metrics.

Table 4. Results of hypothesis tests for the MLP models.

Model	$\bar{D}$	S_D	T_0	p-value	$t_{\alpha/2,n\text{-}1}$
1	$-8.03914.10^{-11}$	$1.31516.10^{-14}$	$-3.30264.10^{6}$	0.94311	2.04523
2	$8.52000.10^{-9}$	$2.09590.10^{-15}$	$7.98267.10^{8}$	0.66530	2.04523
3	$1.85000.10^{-18}$	$2.87832.10^{-33}$	$9.86887.10^{16}$	0.32558	2.04523
4	$2.93667.10^{-10}$	$8.89652.10^{-18}$	$1.80799.10^{8}$	0.71776	2.04523

Source: Proposed by authors.

7 Conclusion

This study compared four artificial neural networks learning methods when applied on supply chain performance evaluation based on SCOR® metrics. The cross-validation method was used to evaluate the candidate topologies and choose the most appropriate number of neurons for each model. The LM algorithm obtained greater prediction accuracy in the four level 1 metrics. Results suggest that LM and SCG algorithms present best performance in the models where the number of neurons in the middle layer is one unity bigger than the number of input variables. There was no similar behavior for BFGS and GDM algorithms. It is important to highlight that the GDM algorithm has the lowest accuracy among those evaluated, but did generate more precise results than the original backpropagation algorithm used by Lima-Junior & Carpinetti (2019). The regression analysis and correlation coefficient results reinforce the suitability of the LM algorithm to support the supply chain performance evaluation based on level 1 SCOR® metrics.

The results of this study are useful to aid researchers in the creation of new performance evaluation models based on ANN, especially in respect to the definition of topological parameters, learning methods and the accuracy level that can be reached for each level 1 metric. It can also be useful to guide developers of Machine Learning tools that aim to create new solutions for decision-making, which is an imminent demand in the industry 4.0 era.

A limitation of this study is related to the use of simulated data, since there was no possibility to collect real data due to the required amount (500 samples for each metric). It is important to highlight that the difficulty in collecting data to evaluate supply chain performance is mentioned in various studies (Didehkhani, et al., 2009; Brandenburg et al., 2014; Dias & Ierapetritou, 2017; Lima-Junior & Carpinetti, 2017). However, factors such as a greater integration of processes and performance measurement systems across supply chain tiers, as well as popularization of data management technologies such as Big Data and Data Warehouse, may contribute to increasing data availability and facilitate the implementation of ANN models in the next years.

Future studies can compare the performance of training algorithms that were not tested yet in supply chain performance evaluation. Another suggestion is to compare the performance of other learning methods and consider the level 1 and level 3 metrics that were not tested in this study.

References

Abdi-Khanghah, M., Bemani, A., Naserzadeh, Z. & Zhang, Z. (2018). Prediction of solubility of N-alkanes in supercritical CO2 using RBF-ANN and MLP-ANN. Journal of CO2 Utilization, 25, 108-119. https://doi.org/10.1016/j.jcou.2018.03.008.

Ahi, P., & Searcy, C. (2015). Assessing sustainability in the supply chain: A triple bottom line approach. *Applied Mathematical Modelling*, 39(10-11), 10-11. http://dx.doi.org/10.1016/j.apm.2014.10.055.

Akkawuttiwanich, P., & Yenradee, P. (2018). Fuzzy QFD approach for managing SCOR performance indicators. *Computers & Industrial Engineering*, 122, 189-201. http://dx.doi.org/10.1016/j.cie.2018.05.044.

Bertrand, J. W. M., & Fransoo, J. (2002). Operations management research methodologies using quantitative modeling. *International Journal of Operations & Production Management*, 22(2), 241-264. http://dx.doi.org/10.1108/01443570210414338.

Bilgehan, M. (2011). Comparison of ANFIS and NN models – With a study in critical buckling load estimation. *Applied Soft Computing*, 11(4), 3779-3791. http://dx.doi.org/10.1016/j.asoc.2011.02.011.

Brandenburg, M., Govindan, K., Sarkis, J., & Seuring, S. (2014). Quantitative models for managing supply chain risks. *European Journal of Operational Research*, 233, 299-312. http://dx.doi.org/10.1016/j.ejor.2013.09.032.

Bukhori, I. B., Widodo, K. H., & Ismoyowati, D. (2015). Evaluation of Poultry Supply Chain Performance in XYZ Slaughtering House Yogyakarta using SCOR and AHP Method. *Agriculture and Agricultural Science Procedia*, 3, 221-225. http://dx.doi.org/10.1016/j.aaspro.2015.01.043.

Clivillé, V., & Berrah, L. (2012). Overall performance measurement in a supply chain: towards a supplier-prime manufacturer based model. *Journal of Intelligent Manufacturing*, 23(6), 2459-2469. http://dx.doi.org/10.1007/s10845-011-0512-x.

Dias, L. S., & Ierapetritou, M. G. (2017). From process control to supply chain management: an overview of integrated decision making strategies. *Computers & Chemical Engineering*, 106, 826-835. http://dx.doi.org/10.1016/j.compchemeng.2017.02.006.

Didehkhani, H., Jassbi, J., & Pilevari, N. (2009). Assessing flexibility in supply chain using adaptive neuro fuzzy inference system. In *IEEE International Conference On Industrial Engineering And Engineering Management (IEEM 2009)*. Hong-Kong: IEEE. http://dx.doi.org/10.1109/IEEM.2009.5373292

Dissanayake, C. K., & Cross, J. A. (2018). Systematic mechanism for identifying the relative impact of supply chain performance areas on the overall supply chain performance using SCOR Model and SEM. *International Journal of Production Economics*, 201, 102-115. http://dx.doi.org/10.1016/j.ijpe.2018.04.027.

Estampe, D., Lamouri, S., Paris, J., & Brahim-Djelloul, S. (2013). A framework for analysing supply chain performance evaluation models. *International Journal of Production Economics*, 142(2), 247-258. http://dx.doi.org/10.1016/j.ijpe.2010.11.024.

Fan, X., Zhang, S., Wang, L., Yang, Y., & Hapeshi, K. (2013). An evaluation model of supply chain performances using 5DBSC and LMBP Neural Network Algorithm. *Journal of Bionics Engineering*, 10(3), 383-395. http://dx.doi.org/10.1016/S1672-6529(13)60234-6.

Ganga, G. M. D., & Carpinetti, L. C. R. (2011). A fuzzy logic approach to supply chain performance management. *International Journal of Production Economics*, 134(1), 177-187. http://dx.doi.org/10.1016/j.ijpe.2011.06.011.

Golparvar, M., & Seifbarghy, M. (2009). Application of SCOR Model in an Oil- producing Company. *Journal of Industrial Engineering*, 4, 59-69.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management*, 21(1-2), 71-87. http://dx.doi.org/10.1108/01443570110358468.

Jalalvand, F., Teimoury, E., Makui, A., Aryanezhad, M. B., & Jolai, F. (2011). A method to compare supply chains of an industry. *Supply Chain Management*, 16(2), 82-97. http://dx.doi.org/10.1108/13598541111115347.

Kocaoğlu, B., Gülsün, B., & Tanyaş, M. (2013). A SCOR based approach for measuring a benchmarkable supply chain performance. *Journal of Intelligent Manufacturing*, 24(1), 113-132. http://dx.doi.org/10.1007/s10845-011-0547-z.

Kurtgoz, Y., Karagoz, M., & Deniz, E. (2017). Biogas engine performance estimation using ANN. *Engineering Science and Technology*, 20(6), 1563-1570. http://dx.doi.org/10.1016/j.jestch.2017.12.010.

Lima-Junior, F. R., & Carpinetti, L. C. R (2019). Predicting supply chain performance based on SCOR® metrics and multilayer perceptron neural networks. *International Journal of Production Economics*, 212, 19-38. http://dx.doi.org/10.1016/j.ijpe.2019.02.001.

Lima-Junior, F. R., & Carpinetti, L. C. R. (2017). Quantitative models for supply chain performance evaluation: A literature review. *Computers & Industrial Engineering*, 113, 333-346. http://dx.doi.org/10.1016/j.cie.2017.09.022.

Liu, F. F., & Liu, Y. C. (2017). A methodology to assess the supply chain performance based on virtual-gap measures. *Computers & Industrial Engineering*, 110, 550-559. http://dx.doi.org/10.1016/j.cie.2017.06.010.

Maestrini, V., Luzzini, D., Maccarrone, P., & Caniato, F. (2017). Supply chain performance measurement systems: A systematic review and research agenda. *International Journal of Production Economics*, 183, 299-315. http://dx.doi.org/10.1016/j.ijpe.2016.11.005.

Marchand, D., & Raymond, L. (2008). Researching performance measurement systems – An information system perspective. *International Journal of Operations & Production Management*, 28(7), 663-686. http://dx.doi.org/10.1108/01443570810881802.

Maroufpoor, S., Shiri, J., & Maroufpoor, E. (2019). Modeling the sprinkler water distribution uniformity by data-driven methods based on effective variables. *Agricultural Water Management*, 215(20), 63-73. http://dx.doi.org/10.1016/j.agwat.2019.01.008.

Mathworks. (2018). *Site de suporte do software MATLAB*. Retrieved in 20 de agosto de 2018, from http://www.mathworks.com/help/nnet/ref/traingd.html

Mentzer, J. T., DeWitt, W., Keebler, J. S., Min, S., Nix, N. W., Smith, C. D., & Zacharia, Z. G. (2001). Defining supply chain management. *Journal of Business Logistics*, 22(2), 1-25. http://dx.doi.org/10.1002/j.2158-1592.2001.tb00001.x.

Moharamkhani, A., Amiri, A. B., & Mina, H (2017). Supply chain performance measurement using SCOR model based on interval-valued fuzzy TOPSIS. *International Journal of Logistics Systems and Management*, 27(1), 115. http://dx.doi.org/10.1504/IJLSM.2017.083225.

Montgomery, D. C., & Runger, G. C. (2009). *Estatística aplicada e probabilidade para engenheiros* (4a ed.). Rio de Janeiro: LTC.

Mukherjee, I., & Routroy, S. (2012). Comparing the performance of neural networks developed by using Levenberg–Marquardt and Quasi-Newton with the gradient descent algorithm for modelling a multiple response grinding process. *Expert Systems with Applications*, 39(3), 2397-2407. http://dx.doi.org/10.1016/j.eswa.2011.08.087.

Patuwo, E., Hu, M. Y., & Hung, M. S. (1993). Two-group classification using neural networks. *Decision Sciences*, 26, 749-779. http://dx.doi.org/10.1111/j.1540-5915.1993.tb00491.x.

Rezaee, M., Jozmaleki, M., & Valipour, M. (2018). Integrating dynamic fuzzy C-means, data envelopment analysis and artificial neural network to online prediction performance of companies in stock exchange. *Physica*, 489, 78-93. http://dx.doi.org/10.1016/j.physa.2017.07.017.

Sellitto, M. A., Pereira, G. M., Borchardt, M., Silva, R., & Viegas, C. V. (2015). A SCOR-based model for supply chain performance measurement: application in the footwear industry. *International Journal of Production Research*, 53(16), 4917-4926. http://dx.doi.org/10.1080/00207543.2015.1005251.

Shafiee, M., Lotfi, F. H., & Saleh, H. (2014). Supply chain performance evaluation with data envelopment analysis and balanced scorecard approach. *Applied Mathematical Modelling*, 38(21-22), 5092-5112. http://dx.doi.org/10.1016/j.apm.2014.03.023.

Silva, I. N., Spati, D. H., & Flauzino, R. A. (2010). *Redes Neurais Artificiais: para engenharia e ciências aplicadas*. São Paulo: Artliber.

Supply Chain Council – SCC. (2012). *Supply Chain Operations Reference Model, version 11.0*. Supply Chain Council.

Tkác, M., & Verner, R. (2016). Artificial neural networks in business: two decades of research. *Applied Soft Computing*, 38, 788-804. http://dx.doi.org/10.1016/j.asoc.2015.09.040.

Tripathy, P. P., & Kumar, S. (2009). Neural network approach for food temperature prediction during solar drying. *International Journal of Thermal Sciences*, 48(7), 1452-1459. http://dx.doi.org/10.1016/j.ijthermalsci.2008.11.014.

Yang, J., & Jiang, H. (2012). Fuzzy evaluation on supply chains' overall performance based on AHM and M(1,2,3). *Journal of Software*, 12(7), 2779-2786. http://dx.doi.org/10.4304/jsw.7.12.2779-2786.