Computers & Industrial Engineering 139 (2020) 106191

Contents lists available at ScienceDirect

Computers & Industrial Engineering

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

An adaptive network-based fuzzy inference system to supply chain performance evaluation based on SCOR® metrics

Francisco Rodrigues Lima-Junior[a], Luiz Cesar Ribeiro Carpinetti[b,*]

[a] Department of Management and Economics, University: Federal Technological University of Paraná, Av. Sete de Setembro, 3165, Rebouças, 80230-901 Curitiba, Paraná, Brazil

[b] Production Engineering Department, University: São Carlos School of Engineering – University of São Paulo, Av. Trabalhador São-Carlense, 400, Parque Arnold Schimidt, 13566-590 São Carlos, São Paulo, Brazil

ARTICLE INFO

Keywords:
Supply chain performance evaluation
SCOR® model
ANFIS
Neuro-fuzzy systems
Supply chain management

ABSTRACT

Evaluating the performance of supply chains (SC) is a critical activity to enhance the outcomes of operations along the SC tiers. In order to support this evaluation process, several studies have proposed the application of artificial intelligence techniques combined with the performance metrics suggested by the SCOR® model (Supply Chain Operations Reference). However these propositions present some limitations. While the systems based on Mamdani fuzzy inference do not allow adaptation to the environment of use based on historical performance data, the systems based on artificial neural networks are not adequate to deal with imprecise data and qualitative metrics. In order to overcome these limitations, this paper presents a new approach to support SC performance evaluation based on the combination between the SCOR® metrics with an adaptive network-based fuzzy inference systems (ANFIS). In total, 56 candidate topologies were implemented and assessed using MATLAB. The random subsampling cross-validation method was applied to select the most appropriate topological parameters for each ANFIS model. The mean square error between the target values and the values estimated by each ANFIS model demonstrate its greater accuracy of prediction. In addition, results of the hypothesis tests based on paired samples indicate that the proposed ANFIS models are adequate to support SC performance evaluation. The proposed system can help managers to develop improvement actions plans based on the outcomes of the evaluation process. When compared to previous approaches, it presents advantages such as greater accuracy of prediction, learning ability based on historical data, suitability to support decision making under uncertainty, better interpretability of results, among others.

1. Introduction

Supply chain management (SCM) has become a critical issue for any company seeking global growth and profitability (Lakri, Dallery, & Jemai, 2015). A supply chain can be seen as a set of integrated business processes that encompass all activities related to the flow and transformation of goods, from the raw material stage to the delivery of the final product to the end customer (Handfield & Nichols, 1999). SCM involves establishing strategic partnership with suppliers and customers, long-term relationships, sharing information, and work together to improve products and processes (Qi, Huo, Wang, & Yeung, 2017). As a result, it can bring advantages such as reduction of inventory, improvement on the use of resources, greater customer satisfaction, among others (Ko, Tiwari, & Mehnen, 2010; Chorfi, Benabbou, & Berrado, 2018). Therefore, measuring supply chain performance in order to assess effectiveness of SCM strategies and practices has become even more important (Chorfi et al., 2018; Lakri et al., 2015).

Evaluating supply chain performance requires the adoption of a set of financial and non-financial metrics related to processes that cross member companies (SCC, 2012). Over the last two decades, a large variety of studies on supply chain performance evaluation have been proposed, including case studies (Sundarakani, Razzak, & Manikandan, 2018), surveys about the most frequently applied metrics, literature reviews (Najmi, Gholamian, & Makui, 2013; Lima-Junior & Carpinetti, 2017; Guersola, Lima, & Steiner, 2018), tools to select performance metrics (Osiro, Lima-Junior, & Carpinetti, 2018), conceptual frameworks (Lakri et al., 2015) and quantitative models for supply chain performance measurement (Akkawuttiwanich & Yenradee, 2018). Faced with the need to make rational and more automated decisions, quantitative models have received special attention by researchers and practitioners in recent years. In general, these models apply multi-criteria decision making methods, statistical or artificial intelligence

* Corresponding author.
E-mail address: carpinet@sc.usp.br (L.C.R. Carpinetti).

https://doi.org/10.1016/j.cie.2019.106191
Received 30 May 2019; Received in revised form 10 October 2019; Accepted 17 November 2019
Available online 19 November 2019

(AI) techniques to estimate the supply chain performance based on multiple measures (Lima-Junior & Carpinetti, 2017). Desirable features of these models comprise: adoption of leading and lagging metrics related to different aspects of SCM (Elgazzar, Tipi, & Jones, 2019); ability to deal with qualitative metrics and decision making processes under uncertainty (Najmi et al., 2013); learning capacity to adapt to the environment of use (Lima-Junior & Carpinetti, 2017); and compatibility of metrics to support benchmarking between global supply chains (Elgazzar et al., 2019).

Akyuz and Erkan (2010) reviewed 24 studies on supply chain performance measurement published on relevant journals. In addition to suggesting some requirements for performance metrics, they have highlighted the significance of the SCOR® (Supply Chain Operations Reference) model to support the development of systems for supply chain performance evaluation. Najmi et al. (2013) reviewed 42 papers published from 1998 to 2010 on supply chain performance measurement; they pointed out that AHP (Analytic Hierarchy Process) and DEA (Data Envelopment Analysis) are the most applied decision making methods, while metrics related to cost, customer, and internal process are the most popular. In another literature review study, Guersola et al. (2018) analysed about 100 papers published between 1999 and 2014, including theoretical studies and models for supply chain performance measurement. These authors also concluded that the most used techniques are AHP and DEA, followed by techniques based on fuzzy logic. However, despite the large number of proposed models, these authors highlight that few of them have been validated using appropriated approaches. Lima-Junior and Carpinetti (2017) identified 84 supply chain performance evaluation models published in journals from 2003 to 2015. They concluded that more than half of the models are based on multicriteria decision making techniques, while only 11 studies adopted AI techniques in order to add new intelligent abilities into performance measurement systems. Similarly to the results presented by Guersola et al. (2018), Lima-Junior and Carpinetti (2017) pointed out that only 33.3% of the studies have applied a validation approach such as statistical tests or sensitivity analysis to evaluate the effectiveness of the obtained results. Therefore, although these supply chain performance evaluation models have brought several significant contributions to SCM, most of them suffer from some limitations such as:

(1) They do not support comparative assessments of performance results against other global supply chains, because they make use of a specific set of metrics generally chosen by experts or selected from previous studies (Didehkhani, Jassbi, & Pilevari, 2009; Jassbi, Seyedhosseini, & Pilevari, 2010; Wang, 2013);
(2) Since some performance metrics are calculated using nonlinear expressions, it is necessary to adopt techniques able to quantitatively model nonlinear causal relationships between input and output metrics. However, most of the quantitative models for supply chain performance evaluation are not suitable for this purpose, since many of them are based on multicriteria decision making techniques that make a weighted linear combination between the input data to produce the output values (Akkawuttiwanich & Yenradee, 2018; Bukhori, Widodo, & Ismoyowati, 2015; Golparvar & Seifbarghy, 2009; Jalalvand, Teimoury, Makui, Aryanezhad, & Jolai, 2011; Kocaoglu, Gülsün, & Tanya, 2013; Moharamkhani, Bozorgi-Amiri, & Mina, 2017; Sellitto, Pereira, Borchardt, Silva, & Viegas, 2015; Theeranuphattana & Tang, 2008; Yang & Jiang, 2012). This approach is often appropriate for problems involving selection, ordering, and categorization of alternatives, but is not suitable for performance prediction problems, since the output variables represent specific performance measures that are derived from specific equations;
(3) Another desirable feature of a supply chain performance evaluation system is the ability to incorporate performance metrics already used by the companies, holding the cause and effect relationships between them, which avoids that measures need to be modified due to the deployment of this new system (Melnyk, Stewart, & Swink, 2004). Thus, it is convenient to adopt artificial intelligence techniques that have the ability to learn the cause and effect relationships from historical performance data, which becomes even more necessary given the trend of increasing data availability due to the use of data management technologies such as big data (Büyüközkan & Göçer, 2018). It is worth to note that supply chain performance evaluation models based on multicriteria methods and mathematical programming methods do not present learning ability (Theeranuphattana & Tang, 2008; Golparvar & Seifbarghy, 2009; Jalalvand et al., 2011; Bai, Sarkis, Wei, & Koh, 2012; Clivillé & Berrah, 2012; Yang & Jiang, 2012; Kocaoglu et al., 2013; Zhang & Reimann, 2014; Bukhori et al., 2015; Sellitto et al., 2015; Moharamkhani et al., 2017), which makes them unsuitable to adapt themselves to a particular environment of use.

Among the studies found in the literature reviewed, only the supply chain performance evaluation systems proposed by Ganga and Carpinetti (2011), and Lima-Junior and Carpinetti (2019) do not present these drawbacks. These three studies combine the performance metrics proposed by the SCOR® model with AI techniques in order to make predictive evaluation systems. In the model developed by Ganga and Carpinetti (2011), a set of fuzzy inference rules are used to represent the cause-and-effect relationships implicit among the SCOR® metrics, which permit to predict performance figures of lagging metrics. Another advantage of using the SCOR® metrics is to compatibility of metrics to carry out benchmarking among global supply chains by means of the SCORmark database (SCC, 2012). However, as the fuzzy inference systems proposed by Ganga and Carpinetti (2011) require collecting the opinion of experts to tune hundreds of inference rules, they are not able to adjust the relationships among the metrics using historical performance data. Another drawback is that choosing and updating the system parameters based on the experts' opinion is very time-consuming. Regarding the multilayer perceptron neural network-based models proposed by Lima-Junior and Carpinetti (2019), although it enables adjusting the system to the environment of use by means of historical performance data, it does not seem appropriate to support decision making processes under uncertainty. Since these models represent and process information from input and output variables using crisp values, they do not have any procedure to deal with the subjectivity of decision makers, qualitative metrics, or imprecise data.

On the other hand, neuro-fuzzy systems bring together supervised learning and decisions under uncertainty, advantages of neural and fuzzy systems respectively. Therefore, this paper proposes a new supply chain performance evaluation system based on neuro-fuzzy systems of the ANFIS type to predict the performance figures of SCOR® level-1 metrics based on the values of level-2 metrics. In total, 56 computational models were implemented and assessed using the random subsampling cross-validation method so as to choose the topologies that present higher accuracy of prediction. Statistical experiments using hypothesis tests were performed to check if there are significant statistical differences between the data predicted by the ANFIS models and the expected values for each metric. Following this introduction, Section 2 reviews the literature about supply chain management, supply chain performance evaluation, and the SCOR® model. Section 3 details the fundamentals on modelling, training, and validation of ANFIS. Section 4 presents the proposal system, while Section 5 details an illustrative application. Section 6 focuses on the validation of the results based on hypothesis tests. Lastly, Section 7 presents the conclusions and recommendations for further works.

2. Supply chain management

The literature presents several definitions regarding the concept of supply chain management (SCM). In a seminal review study about this issue, Mentzer et al. (2001) define supply chain management as "the systemic, strategic coordination of the traditional business functions and the tactics across these business functions within a particular company and across businesses within the supply chain, for the purposes of improving the long-term performance of the individual companies and the supply chain as a whole". Büyüközkan and Göçer (2018) point out that traditional supply chain has a lack of certain attributes that are needed in today's and tomorrow's business requirements. Thus, faced with the need to consider new business requirements, the concepts of supply chain and supply chain management have evolved over the last decades.

In the last decades, the concepts and practices of sustainable supply chain management have gained the attention of academics and practitioners (Dubey et al., 2017). Seuring and Müller (2008) conceptualizes sustainable supply chain management as "the management of material, information and capital flows as well as cooperation among companies along the supply chain while taking goals from all three dimensions of sustainable development, i.e. economic, environmental and social, into account which are derived from customer and stakeholder requirements". While there is no single definition of sustainable supply chain management, existing definitions promote environmental, economic, and social sustainability along the supply chain members aiming for long-term growth (Dubey et al., 2017).

More recently, in the face of changes related to industry 4.0 such as digitalization, virtualization, interoperability, and autonomous decision-making, the concepts of digital supply chain and supply chain 4.0 emerged (Makris, Hansen, & Khan, 2019). According to Büyüközkan and Göçer (2018), digital supply chain "is an intelligent best-fit technological system that is based on the capability of massive data disposal and excellent cooperation and communication for digital hardware, software, and networks to support and synchronize interaction between organizations by making services more valuable, accessible and affordable with consistent, agile and effective outcomes". Makris et al. (2019) defines Supply Chain 4.0 as "a supply chain which involves close collaboration of different stakeholders (e.g. suppliers and customers) and is built on digital technology, including but not limited to, webenabled technology, cloud computing and the Internet of Things". Although there are different terminologies in use in the literature, Makris et al. (2019) highlight that the terms digital supply chain and supply chain 4.0 are often employed interchangeably.

Managing supply chain requires decision making that influence the outcomes in the operational, tactical and strategic levels (Gunasekaran, Patel, & Mcgaughey, 2004; Melnyk et al., 2004). Such management often involves professionals of purchasing and supply management, quality managers, logistic managers, engineers, among others, who act as decision makers (Bals, Schulze, Kelly, & Stek, 2019). Thus, evaluating the performance achieved as a consequence of these decisions is an essential activity to support the supply chain management (Gunasekaran, Patel, & Tirtiroglu, 2001).

2.1. Supply chain performance evaluation

Beamon (1999) conceptualizes supply chain performance as "the evaluation of supply chain management, including resources, output, and flexibility factors". For Chopra and Meindl (2013), supply chain performance is "the result of how supply chain is managed and how well the logistical drivers (facilities, inventory, transportation) and cross functional drivers (information, sourcing and pricing) interact together to determine the level of performance in terms of supply chain's responsiveness and efficiency". Based on SCC (2012), supply chain performance is defined in this study as the performance outcomes of processes of supply chain member companies regarding measures related to reliability, cost, responsiveness, agility, and asset management.

According to Melnyk, Narasimhan, and Decampos (2014), managing supply chain performance involves evaluation of the differences among actual and desired outcomes to identify and flag those performance gaps that are critical. It is also necessary understanding the root causes as well as implementing and monitoring improvement action plans. The adoption of performance evaluation systems is frequently recommended to facilitate implementation of strategies to improve supply chain performance (Melnyk et al., 2014; Qi et al., 2017). Elgazzar et al. (2019) claim that choosing adequate performance metrics may drive managers allocate resources to the most relevant improvement actions. These authors also highlight that it is essential to adopt metrics that allow evaluation of supply chain performance from different perspectives in order to afford a balanced assessment as well as to set performance targets to reflect a company's strategy and objectives. However, most firms failed in developing a system providing a clear portrait of supply chain performance and supporting elaboration of action plans (Lakri et al., 2015).

Metrics adopted for supply chain performance evaluation influence decision making at strategic, tactical, and operational levels (Gunasekaran et al., 2001; Melnyk et al., 2004). The strategic level measures are related to the top level management decisions, very often reflecting "investigation of broad based policies, corporate financial plans, competitiveness and level of adherence to organizational goals" (Gunasekaran et al., 2004). Strategic level measures encompass metrics regarding quality level, order lead time, total cash flow time, cost saving initiatives and lead time against industry standards. The tactical level involves decisions related to resource allocation and performance evaluation against targets in order to attain outcomes specified at the strategic level. Tactical level metrics comprise order cycle time, cash flow and capacity flexibility. In the operational level, performance evaluation requires precise data to measure the outcomes of decisions of low level managers. Performance evaluation of operational level includes measures regarding human resource productivity, percentage of defects, quality of delivered goods, on time delivery of goods, delivery reliability performance. Managers and workers must set operational objectives, which if achieved will carry out to accomplishment of tactical goals (Gunasekaran et al., 2001, 2004).

Elgazzar et al. (2019) classified several studies that propose supply chain performance measurement systems. Based on the analysis of their main scope, tools, measures and practical implications provided, they subdivided the studies into four categories: process-focused systems, prioritization frameworks, causal systems, and integration frameworks. Process-focused systems can be used to identify processes that need improvement and then link corresponding measures to goals. Prioritisation frameworks are designed to identify measures that are below targets and attention. Causality systems may be adopted to assess the impact of enablers on the supply chain performance as well predict performance based on the quantitative relationships among the input and output metrics. Additionally, an integration framework can be applied to embed new functions in the supply chain performance measurement systems, such as connecting metrics to strategy (Elgazzar et al., 2019).

Few studies in the literature propose quantitative forward-looking (predictive) performance evaluation systems (Unahabhokha, Platts, & Tan, 2007). Melnyk et al. (2004) point out that the use of systems based predictive metrics is suitable when the objective is in preventing the occurrence of problems, rather than correcting them. Among the 84 quantitative models for supply chain performance evaluation analyzed by Lima-Junior and Carpinetti (2017), only 4 support the prediction of performance (Didehkhani et al., 2009; Fan, Zhang, Wang, Yang, & Hapeshi, 2013; Ganga & Carpinetti, 2011; Jassbi et al., 2010). In this type of system, the outcome of some measures (lagging metrics) are the consequence of some others (leading metrics) (Unahabhokha et al., 2007; Lima-Junior & Carpinetti, 2019). According to Unahabhokha et al. (2007), a leading metric "is an indicator that measures the drivers of future performance", whereas a lagging metric "is an indicator that measures the output of success of past activity". In general, supply chain performance prediction systems are based on artificial

Table 1
Performance attributes suggested by the SCOR® model.

Attribute	Description
Reliability	"The ability to perform tasks as expected. Reliability focuses on the predictability of the outcome of a process."
Responsiveness	"The speed at which tasks are performed. The speed at which a supply chain provides products to the customer."
Agility	"The ability to respond to external influences, the ability to respond to marketplace changes to gain or maintain competitive advantage."
Assets	"The ability to efficiently utilize assets. Asset management strategies in a supply chain include inventory reduction and in-sourcing vs. outsourcing."
Costs	"The costs of operating the supply chain processes. This includes labor costs, material costs, management and transportation costs."

Source: SCC (2012).

intelligence methods that maps the mathematical expressions aiming to quantify the causal relationships between the input and output metrics (Lima-Junior & Carpinetti, 2017). It is made by using supervised learning algorithms (Didehkhani et al., 2009; Fan et al., 2013; Jassbi et al., 2010) or fuzzy inference rules tuned by specialists' judgments (Ganga & Carpinetti, 2011). Thus, the use of supply chain performance prediction systems allows managers to estimate the values of lagging metrics and to compare them against performance target or standards in order to detect performance gaps and then develop reactive action plans (Unahabhokha et al., 2007; Ganga & Carpinetti, 2011; Lima-Junior & Carpinetti, 2019).

In our review, we identified just two prediction systems based on neuro-fuzzy systems. According to Ko et al. (2010), benefits of employing artificial intelligence techniques such as neuro-fuzzy systems include the "capability to deal with imprecision, uncertainty, and partial truth to achieve tractability and robustness on simulating human decision making behavior with low cost". Kar, Das, and Ghosh (2014) highlight the ability to explore interpretable decision rules. Another advantage is the self-adaptive ability to learn from experience as well as to generalize results from sample data (Ko et al., 2010). Didehkhani et al. (2009) proposed a neuro-fuzzy model to predict supply chain flexibility, which was applied in an automotive enterprise. Jassbi et al. (2010) developed a neuro-fuzzy system to predict supply chain agility and applied them in a leading car manufacturing in Iran. In both studies, some experiments using hypothesis tests were carried out aiming to check if there was a significant statistical difference between the values predicted and the performance figures for each company. Despite the contributions of these studies to the literature on SCM, these models estimate performance figures of just one output metric. In addition, since they use a set of measures limited to the evaluation of a single performance dimension, these models are not adequate to support a broad and balanced evaluation of supply chain performance.

Fan et al. (2013) developed a supply chain performance prediction system based on the combination of 5 Dimensional Balanced Scorecard with Levenberg–Marquardt backpropagation neural network. Wang (2013) proposes a predictive system based on a new neural network algorithm to evaluate supply chain performance of fresh agricultural products. Similarly to Didehkhani et al. (2009) and Jassbi et al. (2010), Fan et al. (2013) and Wang (2013) makes use of a specific set of measures defined by experts or chosen from previous studies. Thus, the performance metrics adopted by these studies are not compatible to benchmark the performance values obtained by a focus company against of other companies within a global supply chains. An approach to overcome these limitations is by using the metrics suggested by the SCOR® model combined with neuro-fuzzy systems. However, in the reviewed literature, there are no predictive performance evaluation systems based on such an approach.

2.2. SCOR® model

The SCOR® model was proposed by the Supply Chain Council in order to link business processes, best practices, performance metrics, people, and technology into a unified structure (SCC, 2012; Chorfi et al., 2018). It defines a practice as a unique way of configuration of a process or activity and categorizes the practices as emerging, best practices, standard practices and declining practices (SCC, 2012). The SCOR® model allows describing the way the processes interact along a supply chain and how they are set from a supplier's supplier to a customer's customer, which enables practitioners to identify the features that contribute to client satisfaction (Ntabe, LeBel, Munson, & Santa-Eulalia, 2015). It has been universally recognized and widely applied in industry, since a common reference model allows companies to communicate using a common terminology that is understandable within and across organizations (Akkawuttiwanich & Yenradee, 2018). SCOR® is well suited to support tactical and operational management for the implementation of strategic planning decision-making (Estampe, Lamouri, Paris, & Brahim-Djelloul, 2013).

As shown in Table 1, the SCOR® model suggests five attributes to manage the supply chain performance, named reliability, responsiveness, agility, assets, and costs (SCC, 2012). The first three attributes are customer-focused, dealing with the effectiveness of the supply chain processes, whereas the others are internally-focused attributes, dealing with efficiency (Chorfi et al., 2018). These attributes should be adopted to establish a strategic direction (SCC, 2012). In order to measure the success of the implementation of these strategies, the SCOR® model proposes the use of some performance metrics associated to each attribute, which are arranged in three hierarchical levels. Fig. 1 illustrates the performance metrics of levels 1 and 2. Level-2 metrics serve as diagnostics to identify the root cause of performance gaps of level-1 metrics (SCC, 2012). Similarly, level-3 performance metrics may be used to diagnose level-2 metrics. The SCOR® model reinforces that this performance diagnosis is an important activity that helps in identifying the processes that need further investigation and improvement action.

In addition to the diagnostic of performance based on causal relationships, the adoption of the SCOR® metrics enables a focus company to compare their performance results against of other global supply chains. A tool called SCORmark, which contains performance figures of over 1,000 companies and 2,000 supply chains, is able to support this benchmarking process. This tool allows three types of analysis, including monitoring and comparison of performance measures on a historical basis, external benchmarking, and comparative analysis between firms of the same group. To carry out benchmarking using SCORmark, managers should track the following steps (APICS, 2018): identify supply chain(s); determine strategic objectives; select metrics; source/collect data; submit data; data validation; and receive final report. Application of the SCORmark brings benefits that comprise identification of performance gaps; setting of performance goals; and helping in the formulation company-specific roadmaps aiming to enhance performance outcomes (SCC, 2012; Lima-Junior & Carpinetti, 2019).

Several applications based on the SCOR® model have been made to deal with problems related to supply chain management. Ntabe et al. (2015) review 45 applications of the SCOR® model published between 2000 and 2012. They concluded that this model is not only appropriated to supply chain financial performance assessment, but it is also a useful decision support tool for environmental performance evaluation along the chain supply. Delipinar and Kocaoglu (2016) identified 27 studies published from 2006 to 2014 that present applications of the SCOR® model, including case studies and mathematical models. These authors reinforce its appropriateness to deal with process modelling

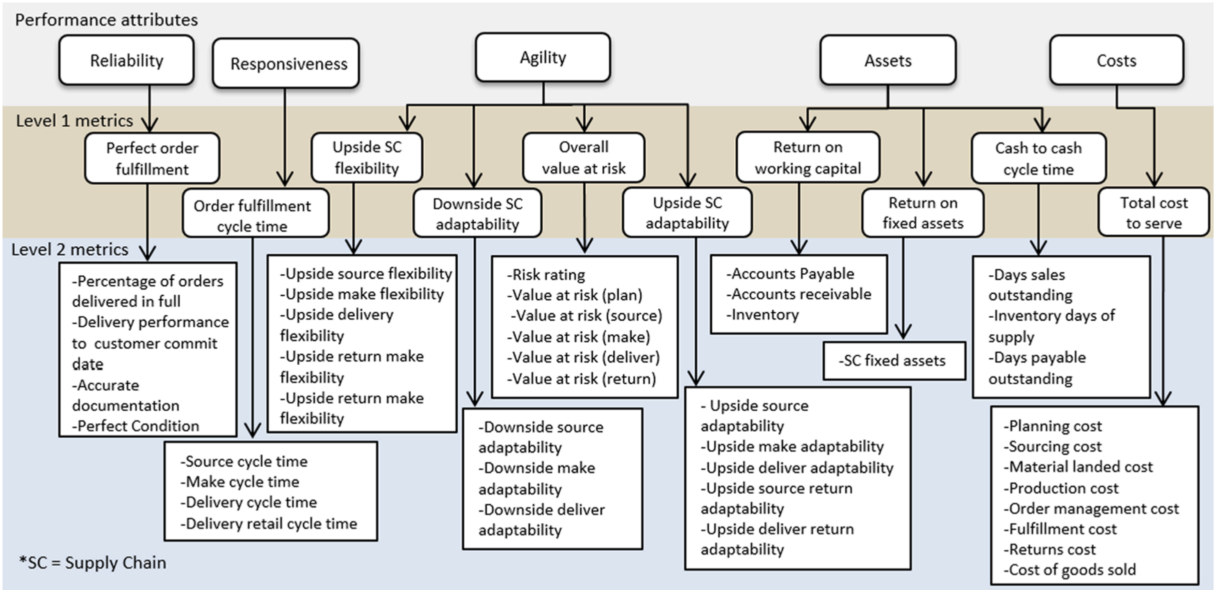

Fig. 1. Level 1 and 2 metrics proposed by the SCOR® model.

and performance measurement. Based on the results of a literature review on supply chain performance evaluation, Guersola et al. (2018) have also recommend using the SCOR® model to support development of new measurement systems.

Table 2 briefly describes the studies found in the literature that propose quantitative models for supply chain performance evaluation based on the SCOR® metrics. There are few models that use mathematical programming methods, such as DEA (Jalalvand et al., 2011; Liu & Liu, 2017), integer programming and multi-objective optimization (Zhang & Reimann, 2014). Artificial intelligence techniques as fuzzy inference systems (Ganga & Carpinetti, 2011) and multilayer perceptron neural networks (Lima-Junior & Carpinetti, 2019) have also been proposed. There is a predominance traditional multicriteria decision making methods such as AHP (Bukhori et al., 2015; Dissanayake & Cross, 2018; Kocaoglu et al., 2013; Sellitto et al., 2015), MACBETH (Clivillé & Berrah, 2012), PROMETHEE (Jalalvand et al., 2011), and TOPSIS (Golparvar & Seifbarghy, 2009; Kocaoglu et al., 2013). However, multicriteria models based on pairwise comparisons between fuzzy numbers (Theeranuphattana & Tang, 2008; Yang and Jiang, 2012) and hybrid decision methods such as interval-valued Fuzzy TOPSIS (Moharamkhani et al., 2017) and Fuzzy QFD (Akkawuttiwanich & Yenradee, 2018) have also been reported.

Although the models showed in Table 1 have brought various contributions to the advancement of the literature and practice on performance evaluation of supply chains, they present some limitations related to the techniques adopted. In the case of the approaches based on pairwise judgments from experts proposed by Theeranuphattana and Tang (2008), Clivillé and Berrah (2012), Yang and Jiang (2012), Kocaoglu et al. (2013), Bukhori et al. (2015), Sellitto et al. (2015), and Dissanayake and Cross (2018), lack of consistency of judgments is a problem. Moreover, the use of these approaches can limit the number of metrics and supply chains considered in the evaluation process. Another problem is that, as the models based on multicriteria methods (Golparvar & Seifbarghy, 2009; Kocaoglu et al., 2013; Moharamkhani et al., 2017; Akkawuttiwanich & Yenradee, 2018) produce an output value based on a weighted linear combination of inputs, they are not adequate to deal with non-linear causal relationships between the SCOR® level-1 and level-2 metrics. It is worth to note that only models based on artificial intelligence techniques (Ganga & Carpinetti, 2011; Lima-Junior & Carpinetti, 2019) present this ability. However, the models based on Mamdani fuzzy inference (Ganga & Carpinetti, 2011) need manual tuning and updating of hundreds of decisions rules based on specialists opinion. Even though neural network-based models (Lima-Junior & Carpinetti, 2019) enables adjusting of topological parameters based on historical performance data, they are not able to deal with qualitative variables and imprecise data. Thus, adoption of ANFIS neuro-fuzzy systems in combination with SCOR® metrics permits to bring together the advantages from neural networks and fuzzy inference systems in order to overcome these limitations.

3. Adaptive network-based fuzzy inference system

There are different computational methods that combine abilities of fuzzy systems with artificial neural networks, which are generically called neuro-fuzzy systems (Kar et al., 2014; Rajab & Sharma, 2018). Jang (1993) proposed the most popular type of neuro-fuzzy system, named Adaptive Network-based Fuzzy Inference System (ANFIS). In order to approximate the human reasoning way, ANFIS combines the architecture of Takagi-Sugeno fuzzy inference systems with the supervised learning ability from radial basis function neural network. The use of fuzzy sets and decision rules enables neuro-fuzzy systems to deal with "impreciseness of input data and domain knowledge and also allows quick and often sufficiently good approximations of desired solutions" (Ko et al., 2010). Such characteristics contribute to the high performance of this type of intelligent system in problems involving function approximation, pattern classification, real-time applications, among others (Kar et al., 2014; Rajab & Sharma, 2018). In a literature review study, Kar et al. (2014) identified dozens of applications of neuro-fuzzy systems in economic systems, manufacturing, electrical and electronic systems, traffic control, image processing and medical systems. Rajab and Sharma (2018) presented a review study focused on neuro-fuzzy applications in business. These authors have identified several ANFIS-based models to support decision making problems in areas such as finance, business planning, marketing, production and operations, human resource management and information systems.

As shown in Fig. 2, the architecture of an ANFIS is composed of five layers. The nodes within the same layer perform similar tasks. Nodes represented by circles are called fixed nodes while those represented by squares are denoted adaptive nodes. The outputs of adaptive nodes depend on modifiable parameters within each node, which are tuned using a set of sample values in the training stage of the model (Jang, 1993; Özkan & Inal, 2014). Unlike artificial neural network models, connections between the nodes of different layers do not present synaptic weights. As in traditional fuzzy inference systems, a type of

Table 2
Brief description of the performance evaluation models based on SCOR® metrics.

Authors	Method(s)	Proposal	Metrics
Theeranuphattana and Tang (2008)	Pairwise comparisons between fuzzy numbers	A conceptual model of performance measurement for supply chains.	Adopts some SCOR® level-2 metrics as subcriteria and the performance attributes proposed by SCOR® as criteria. It proposes the use of an overall supply chain performance index as output variable.
Golparvar and Seifbarghy (2009)	TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution)	Application of SCOR® model to analyse the supply chain of oil-producing company.	Uses the performance attributes proposed by SCOR® as metrics to select projects in order to improve the performance supply chain.
Ganga and Carpinetti (2011)	Mamdani fuzzy inference	An approach based on fuzzy rules decision to predict supply chain performance.	Utilizes some SCOR® level-2 metrics to predict the performance figures of the level-1 metrics.
Jalalvand et al. (2011)	DEA (Data envelopment analysis) and PROMETHEE II	A method to compare supply chains of an industry.	Adopts some level-2 metrics to estimate the values of the level-1 metrics related to the process Plan, Source, Make, and Deliver.
Clivillé and Berrah (2012)	MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique)	A model to quantify overall performance of a supply chain.	Proposes the use of some particular metrics not extracted from SCOR®, which have been linked to the processes Plan, Source, Make, Deliver, and Return. It also proposes the use of an overall supply chain performance measure.
Yang and Jiang (2012)	New method based on fuzzy numbers and M(1,2,3)	Fuzzy evaluation on supply chains' overall performance based on analytic hierarchical model	Aggregate the values of some of the level-1 SCOR® metrics to estimate a measure of overall supply chain performance proposed by the authors.
Kocaoglu et al. (2013)	AHP (Analytic Hierarchy Process) and TOPSIS	A SCOR® based approach for evaluating supply chain performance	Applies some of the metrics of levels 1 and 2 to estimate the "supply chain performance efficiency".
Zhang and Reimann (2014)	Integer programming and multi-objective optimization	Multi-objective performance assessment and optimization of a multi-period two-echelon supply chain.	Utilizes the performance attributes Costs, Agility, Assets, and Reliability as variables of a mathematical programming model.
Bukhori et al. (2015)	AHP	Evaluation of poultry supply chain performance in XYZ Slaughtering House Yogyakarta.	Adopts some of the level-1 metrics related to the attributes Reliability, Responsiveness, Flexibility, and Costs.
Sellitto et al. (2015)	AHP	Performance assessment model in the footwear industry.	Applies a set particular of metrics not extracted from SCOR®, which have been linked to the processes and attributes proposed by SCOR®.
Moharamkhani et al. (2017)	Interval-valued Fuzzy TOPSIS	Supply chain performance evaluation based on the combination between SCOR® metrics and interval-valued fuzzy TOPSIS.	Adopts the performance attributes proposed by SCOR® as metrics to rank supply chains based on their overall performance.
Liu and Liu (2017)	DEA	A methodology to assess the supply chain performance based on gap-based measures.	Utilizes a set particular of metrics not extracted from SCOR®, which have been linked to the processes Plan, Source, and Make.
Akkawuttiwanich and Yenradee (2018)	Fuzzy QFD (Quality Function Deployment)	An approach to manage key performance indicators of the SCOR® model for performance improvements.	Adopts some level 1 and 2 metrics related to the attributes Reliability, Responsiveness, Agility, Costs, and Assets.
Dissanayake and Cross (2018)	AHP and structural equations modelling	Systematic mechanism for identifying the relative impact of supply chain performance areas on the overall supply chain performance.	Applies a set particular of metrics not extracted from SCOR®, which have been linked to the attributes Reliability, Responsiveness, Agility, and Asset Management.
Lima-Junior and Carpinetti (2019)	Multilayer perceptron neural networks	An intelligent system for performance prediction of SCOR® level-1 metrics based on historical performance data.	Applies some level-2 metrics to predict the performance figures of the level-1 metrics.

Source: Proposed by Authors.

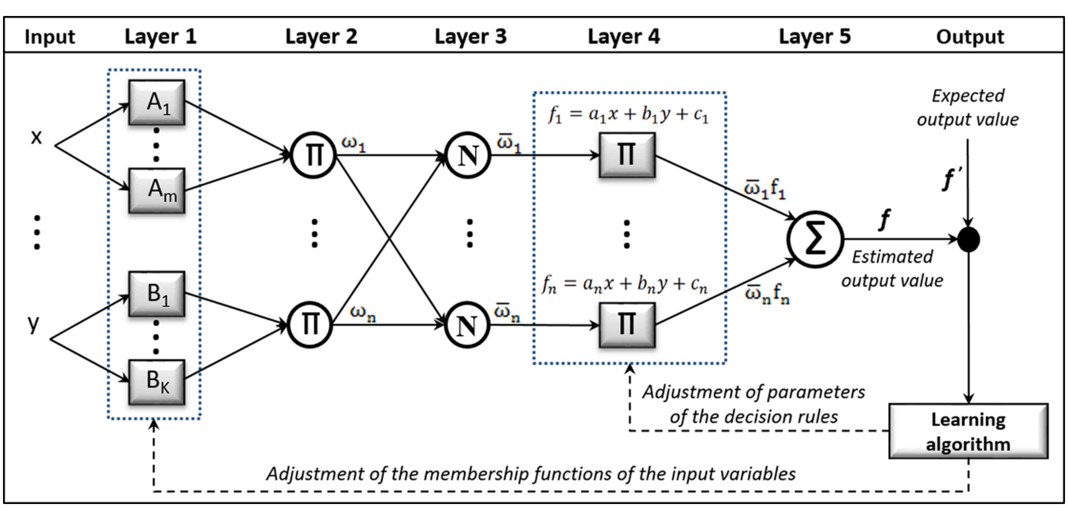

Fig. 2. Architecture of an adaptive network-based fuzzy inference system. Source: Based on Jang (1993).

$$\overbrace{\text{IF } x = A_1 \text{ AND } y = B_1}^{\text{Antecedent}} \text{ THEN } \overbrace{f_1 = a_1x + b_1y + c_1}^{\text{Consequent}}$$
(a)

$$\text{IF } x = A_m \text{ AND } y = B_k \text{ THEN } f_n = a_nx + b_ny + c_n$$
(b)

Fig. 3. The first (a) and last (b) inference rule of an ANFIS. Source: Based on Güneri et al. (2011).

membership functions such as triangular, trapezoidal, or gaussian should be chosen for each input variable. In several applications involving decision making, the membership functions are associated to linguistic terms defined by specialists (Aengchuan & Phruksaphanrat, 2018; Güneri, Ertay, & Yücel, 2011; Jassbi et al., 2010).

In layer 1, the objective is to fuzzify the input values (x and y) of the model, that is, to convert a set of crisp numerical values into one or more equivalent fuzzy sets. The output of this layer consists of a set of membership values corresponding to the level of activation of each one of the membership functions of the input variables: $\{\mu_{A_1}(x), \cdots, \mu_{A_m}(x)\}$ and $\{\mu_{B_1}(y), \cdots, \mu_{B_k}(y)\}$. In layer 2, each node corresponds to the antecedent part of a inference rule, which represents possible combinations between the membership functions of the first layer. The objective of layer 2 is to compose the logical relationships between the activated membership functions in order to determine the weight of each rule (ω_i). As shown in Fig. 3, inference rules of ANFIS are similar to ones of a first-order Takagi-Sugeno fuzzy inference system. Fig. 3(a) illustrates the first rule of an ANFIS, while Fig. 3(b) shows the last one. A t-norm operator such as minimum or algebraic product is applied to calculate the activation degree of each inference rule, as in Eqs. (1) and (2), respectively. Layer 3 nodes are dedicated exclusively to the normalization of weights of the activated rules by means of Eq. (3) (Jang, 1993; Güneri et al., 2011).

$$\min\{\mu_A(x), \mu_B(y)\} \tag{1}$$

$$\mu_A(x)\cdot\mu_B(y) \tag{2}$$

$$\bar{\omega}_i = \frac{\omega_i}{\omega_1 + \cdots + \omega_i + \cdots + \omega_n} \tag{3}$$

Layer 4 is composed of a set of adaptive nodes that represent the consequents of the inference rules and provide the outputs of each rule. Each consequent is represented by a linear function or a constant value. In the first case, as indicated in Fig. 2, the crisp values of the input variables (x and y) serve as parameters of the function. The output of each node of the layer 4 is computed by multiplication between the consequent and the weight of the activated rule. Finally, using the Eq. (4), layer 5 performs the aggregation of the outputs of each layer 4 node by means of a weighted sum, which provides the final output of the system (Jang, 1993; Aengchuan & Phruksaphanrat, 2018).

$$f = \sum_{i=1}^{n} \bar{\omega}_i \cdot f_i \tag{4}$$

Fig. 4 illustrates the inference process of an ANFIS. In this example, two fuzzy input variables are partitioned by triangular membership functions. In the fuzzification step, the crisp value x activates the fuzzy sets A_1 and A_2, whereas y activates B_1 and B_2, with different membership degrees. The combination of the activated fuzzy sets triggers two inference rules, which weights (ω_1 and ω_2) are calculated by the minimum or algebraic product operators based on the values of $\mu_{A_1}(x)$, $\mu_{B_1}(x)$, $\mu_{A_2}(x)$ and $\mu_{B_2}(x)$. In the last step, following Eqs. (3) and (4), the output values of the activated rules (f_1 and f_2) are weighted using ω_1 and ω_2 in order to produce the final output f (Jang, 1993; Aengchuan & Phruksaphanrat, 2018).

Modelling of ANFIS is an empirical process that requires several computational tests aiming to choose an appropriate network topology for each model. A set of topological parameters that directly affects the prediction accuracy of an ANFIS, such as the number of membership functions of the input variables, the type of membership function, the type of consequent, and the logical operator of the inference rules should be evaluated in each test (Aengchuan & Phruksaphanrat, 2018; Özkan & Inal, 2014). Several studies have adopted methods of cross-validation to support the modelling of the ANFIS process in order to develop topologies that attain superior prediction accuracy (Dewan, Huggett, Liao, Wahab, & Okeil, 2016; Erdas, Andac, Gurkan-Alp, Alpaslan, & Buyukbingol, 2013; Kolus, Imbeau, Dubé, & Dubeau, 2015). Such methods evaluate distinct combinations of values of parameters with the purpose of selecting those that achieve the minor error in the validation stage of the model. Examples of cross-validation methods comprise leave-one-out cross-validation (Dewan et al., 2016), k-fold cross-validation (Kolus et al., 2015) and random subsampling cross-validation (Erdas et al., 2013), which have been applied in this study.

Application of the random subsampling cross-validation method involves several learning processes that require a set of sample values of the input and output variables. Each learning process is composed of two stages, named training and validation. The total set of samples should be subdivided in two subsets. While the first subset is applied to tune the adaptive parameters during the training stage, the second is employed to evaluate the prediction accuracy of each candidate topology. The percentage of samples recommended is 60 to 90% for the training subset and 10 to 40% for validation subset (Özkan & Inal, 2014; Lima-Junior & Carpinetti, 2019). One of the learning methods most frequently applied to ANFIS tuning is a hybrid algorithm proposed by Jang (1993). In the first step of this algorithm, called forward, while the initial values of the membership functions of the input variables are kept fixed, the Least Mean Squares (LMS) method is applied to tune the adaptive parameters of the layer 4 (represented in Figs. 3 and 4 by a_i, b_i and c_i). In the backward step, the algorithm backpropagation adjusts the membership functions of the input variables based on the minimization of the error between the output value yield by the ANFIS model and the output value of each training sample. The number of times the training subset is presented to the model, called number of epochs, serves as a stopping criterion for the training process. After completing training, in the validation stage, accuracy of the model is assessed using measures such as the Mean Square Error (MSE) or Mean Absolute Error (MAE). If none of the tested candidate topologies reaches the required accuracy level, new topologies should be created in order to test new values for the parameters of the model.

4. A system for supply chain performance evaluation

The system proposed by this study to aid supply chain performance evaluation is based on Jang (1993), Ganga and Carpinetti (2011), SCC (2012), and Lima-Junior and Carpinetti (2019). As depicted in Fig. 5, it utilizes a set of seven neuro-fuzzy systems to model the causal relationships defined by SCOR®, making it possible to estimate the values of level-1 metrics based on level-2 metrics. The main objective of the model is to support a predictive diagnosis that allows identifying which level-1 metric(s) underperform so as to develop action plans to fulfil the identified performance gaps. The inputs of the neuro-fuzzy models consist of level-2 metrics, whereas the output variables represent level-1 metrics. In the case of the ANFIS model 5, it uses the level-2 metrics "inventory ", "accounts receivable", and "accounts payable" to predict the performance values of variable y_5. This variable is applied to calculate the performance figures of level-1 metric "return on capital working capital", which is estimated by model 3 also considering "total cost to serves" and "supply chain revenue". It is worth to note that since the set of measures adopted in this study is entirely based on the SCOR® metrics, the proposed system does not calculate an overall performance score.

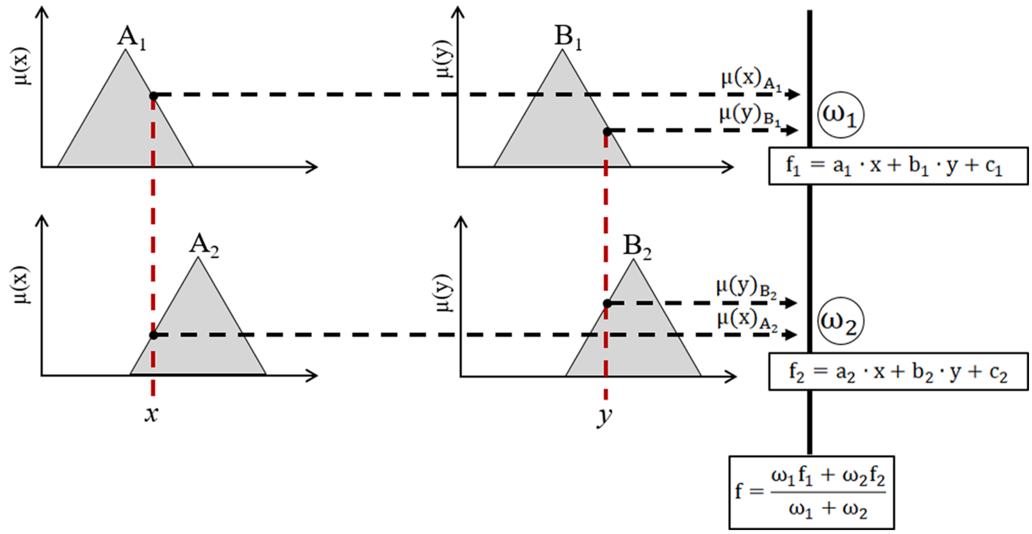

Fig. 4. Inference process of an ANFIS. Source: Jang (1993).

To apply the proposed system to the performance evaluation of a supply chain, the following steps should be taken:

i. Define set of SCOR® metrics: choose a set of SCOR® level-1 metrics aligned with the competitive strategies and objectives defined to manage the supply chain. The SCOR® model (SCC, 2012) suggests choosing at least one metric related to each performance attribute in order to promote a balanced assessment and a decision making process based on different perspectives. Realistic performance targets should also be established for each metric;

ii. Train ANFIS: gather or estimate a set of performance figures from the chosen level-2 metrics. The values of level-1 metrics should also be collected or estimated by decision makers according to the values defined for level-2 metrics. The SCOR® model (SCC, 2012) highlights that these data should be obtained based on estimation or historical data extracted from traditional organizational systems. It is essential to ensure that at least one sample containing the upper and lower limit values of all input and output variables is included in the training subset. These samples should be applied to carry out the supervised learning processes of the ANFIS models using the random subsampling cross-validation method. After completing the learning processes, when the cause and effect relationships between input and output metrics is defined by the ANFIS models, the system is able to predict the values of the SCOR® level-1 metrics;

iii. Predict SC performance: to predict performance related to level-

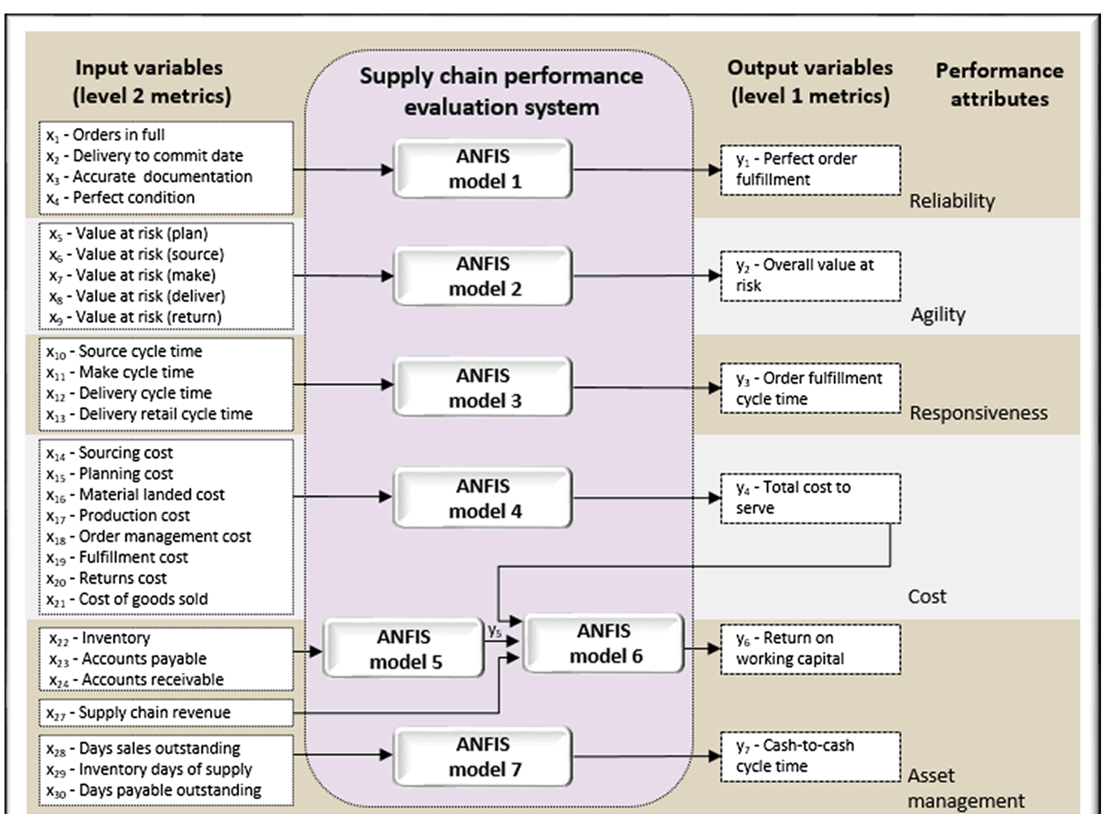

Fig. 5. Proposed system for supply chain performance evaluation.

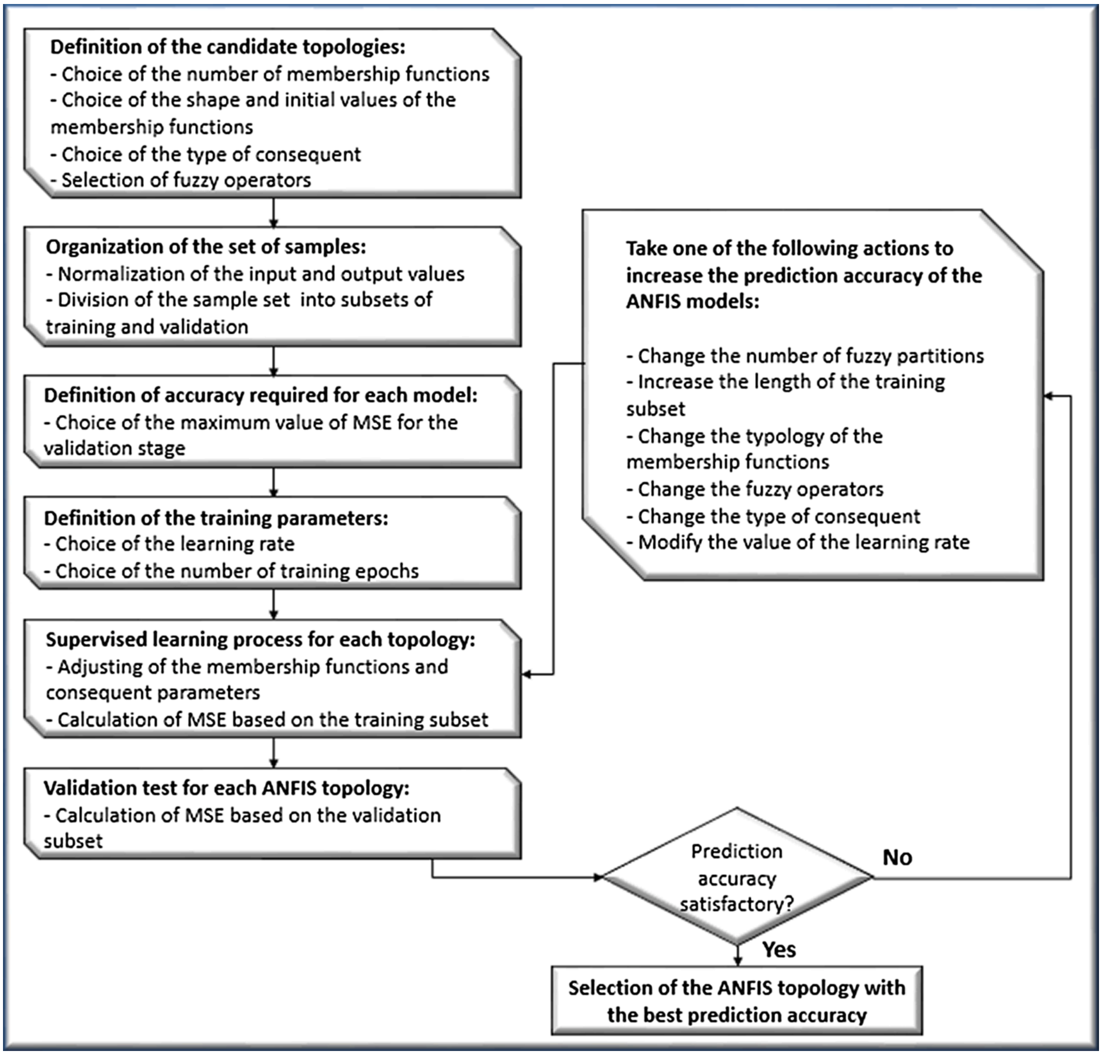

Fig. 6. Sequence of steps for application of the random subsampling cross-validation method.

1 metrics, at least one sample of the current performance in each of the level-2 metrics needs to be presented to the ANFIS models. These samples should be normalized using Eq. (5).

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{5}$$

iv. Plan improvement actions: based on the comparison between the predicted performance values and the targets defined for each level-1 metric, managers can identify the gaps of performance so as to elaborate and implement action plans aiming to improve outcomes of the supply chain operations. These action plans should be focused on the enhancement of the supply chain processes related to the level 1 and 2 metrics whose performance is below of the expected level. Monitoring of the effectiveness of these plans can also be done using the proposed system.

Regarding the learning process mentioned in step 2, Fig. 6 suggests a sequence of steps to apply the random subsampling cross-validation method in order to select the candidate topologies that present the superior accuracy. Initially, samples values of SCOR® level 1 and 2 metrics should be normalized according to Eq. (5). It is also divided into two subsets of samples, with about 70% of the samples being applied in the training stage and the remainder in the validation stage. After organizing the samples, a set of candidate topologies should be created by varying the settings of the same parameters that directly affect the accuracy of the ANFIS models, including the number of fuzzy partitions, the type of membership functions, the type of consequent and fuzzy operator used in the layer 2. Values for training parameters, such as learning rate and number of epochs, should also be defined. The selection of the values of these parameters should be made based on the literature and empirical computational tests. With respect to the initial values of parameters of the membership functions, it can be defined based on knowledge from specialists or by symmetrically dividing the universe of discourse of the input variables by the number of linguistic terms adopted. The type of consequent and fuzzy operators to be tested in the inference rule base should also be indicated as well as the values for the prediction performance accuracy desired for each ANFIS model.

Training the ANFIS models should be carried out using the hybrid algorithm proposed by Jang (1993), in which the least squares method is used to adjust the parameters of the consequent part of the inference rules, whereas the backpropagation algorithm is applied to adjust the membership functions of the input variables. At each training epoch, the values of the input variables of the training subset are computed by the ANFIS models in order to estimate an output value for each sample. Then, the estimated values are compared with the output values of the training subset so as to calculate the values of MSE that are used to tuning the adaptive parameters of the neuro-fuzzy models. When the training stage is finished, the values of MSE are computed based on the validation subset. If at least one candidate topology reaches a satisfactory prediction accuracy, the one that produces the lowest MSE in the validation stage should be chosen. When none of the topologies

Table 3
Description of the metrics of the proposed performance evaluation system.

ANFIS model	Brief description	Universe of discourse	Measurement unit
1	**x_1 - orders delivered in full:** "evaluates if all items on the order line are received by customer in the specified quantities."	[0; 1] (SCC, 2012)	Dimensionless
	x_2 - delivery performance to customer commit date: "an order is considered delivered to the customer in the commitment date if: the order is received on time as defined by the customer; and delivery is made to the correct location and customer entity."	[0; 1] (SCC, 2012)	Dimensionless
	x_3 - documentation accuracy: "an order is considered to have accurate documentation when the following are accepted by the customer: shipping documentation, payment documentation, compliance documentation, and other required documentations."	[0; 1] (SCC, 2012)	Dimensionless
	x_4 - perfect condition: "orders delivered in an undamaged state that meet specification, have the correct configuration, are faultlessly installed (as applicable) and accepted by the customer."	[0; 1] (SCC, 2012)	Dimensionless
	y_1 - perfect order fulfilment: "quantifies the orders meeting delivery performance with complete and accurate documentation and no delivery damage. For an order line to be considered as perfect, all of the individual components must be perfect. Each component receives a score of 1 if it is judged to be perfect."	[0; 4] (SCC, 2012)	Dimensionless
2	**x_5 - value at risk (plan):** "is the sum of the monetized risks related to the plan process."	[10,000; 100,000] (Lima-Junior & Carpinetti, 2019)	$
	x_6 - value at risk (source): "is the sum of the monetized risks related to the source process."	[50,000; 200,000] (Lima-Junior & Carpinetti, 2019)	$
	x_7 - value at risk (make): "is the sum of the monetized risks associated to the make process."	[50,000; 300,000] (Lima-Junior & Carpinetti, 2019)	$
	x_8 - value at risk (deliver): "is the sum of the monetized risks related to the deliver process."	[20,000; 200,000] (Lima-Junior & Carpinetti, 2019)	$
	x_9 - value at risk (return): "is the sum of the monetized risks associated to the return process."	[20,000; 200,000] (Lima-Junior & Carpinetti, 2019)	$
	y_2 - overall value at risk: "is the sum of the probability of risk events times the monetary impact of the events which can impact any core supply chain functions (e.g. Plan, Source, Make, Deliver or Return) or key dependences. The use of VaR allows organizations to look at all potential supply chain risks through one metric and helps prioritize mitigation efforts."	[150,000; 1,000,000] (SCC, 2012)	$
3	**x_{10} - source cycle time:** "the sum of the time spent to identify sources of supply, select the supplier (s), negotiate cycle time, schedule product delivery, receive, verify and transfer product, and authorize supplier payment."	[1; 6] (Ganga & Carpinetti, 2011)	Days
	x_{11} - make cycle time: "the sum of time spent to finalize production engineering, schedule production activities, material unloading, produce, test, package and release product."	[1; 7] (Ganga & Carpinetti, 2011)	Days
	x_{12} - delivery cycle time: "the average time associated with deliver processes."	[1; 7] (Ganga & Carpinetti, 2011)	Days
	x_{13} - delivery retail cycle time: "the average processes cycle time to acquire, merchandise, and sell finished goods at a retail store."	[7; 20] (Lima-Junior & Carpinetti, 2019)	Days
	y_3 - **order fulfilment cycle time**: "the average actual cycle time achieved to fulfil customer orders. For each individual order, this cycle starts from the receipt of the order and ends with the acceptance of the order by the customer."	[10; 40] (SCC, 2012)	Days
4	**x_{14} - sourcing cost:** "is the total cost associated with managing and ordering, receiving, inspection and warehousing of material, products, merchandise and services."	[140,000; 300,000] (Ganga & Carpinetti, 2011; Jalalvand et al., 2011)	$
	x_{15} - planning cost: "is the total cost of personnel, automation, assets and overhead associated with the supply chain planning processes."	[25,000; 50,000] (Ganga & Carpinetti, 2011; Jalalvand et al., 2011)	$
	x_{16} - material landed cost: "the total cost associated with buying materials, products or merchandize and making them available to the location of use."	[70,000; 150,000] (Ganga & Carpinetti, 2011; Jalalvand et al., 2011)	$
	x_{17} - production cost: "the total cost associated with managing and performing production processes."	[150,000; 380,000] (Jalalvand et al., 2011)	$
	x_{18} - order management cost: "the total personnel, automation, assets and overhead cost associated with the fulfilment of orders."	[220,000; 480,000] (Ganga & Carpinetti, 2011)	$
	x_{19} - fulfilment cost: "the total cost of personnel, automation and assets associated with the responding to inquiries and quotes, order entry and maintenance, scheduling transportation, order tracking and tracing, delivery, installation and invoicing."	[45,000; 70,000] (Ganga & Carpinetti, 2011; Jalalvand et al., 2011)	$
	x_{20} - returns cost: "the total cost of disposition of materials returned due to planning errors, supplier quality, product, order management and delivery errors."	[50,000; 200,000] (Lima-Junior & Carpinetti, 2019)	$
	x_{21}- cost of goods sold: "the cost of direct materials, direct labour and overhead associated with the acquisition or production of finished goods."	[1,300,000; 1,900,000] (Jalalvand et al., 2011; Ganga & Carpinetti, 2011)	$
	y_4 - total cost to serve: "the sum of the direct and indirect costs to deliver products and services to customers. Total cost to serve is the sum of planning cost, sourcing cost, material landed cost, production cost, order management cost, fulfilment cost, and returns cost."	[2,000,000; 3,530,000] (SCC, 2012)	$
5	**x_{22} - inventory**: "the amount of inventory (stock) expressed in dollars."	[100,000; 2,000,000] (Lima-Junior & Carpinetti, 2019)	$
	x_{23} - accounts receivable: "the amount of accounts receivable outstanding expressed in dollars."	[500,000; 2,000,000] (Lima-Junior & Carpinetti, 2019)	$
	x_{24} - accounts payable: "the amount of purchased materials, labour and/or conversion resources that are to be paid."	[500,000; 2,000,000] (Lima-Junior & Carpinetti, 2019)	$
	y_5 – "it quantifies the denominator used in the calculations of "return on working capital."	[−1,400,000; 3,500,000] (SCC, 2012)	$

(*continued on next page*)

Table 3 (*continued*)

ANFIS model	Brief description	Universe of discourse	Measurement unit
6	x_{25} - **total cost to serve:** "it is the output variable (y) of the ANFIS model 4."	[2,000,000; 3,530,000] (SCC, 2012)	$
	x_{26} - "it is the output variable (y) of the ANFIS model 5."	[−1,400,000; 3,500,000] (SCC, 2012)	$
	x_{27} - **supply chain revenue:** "refers to the operating revenue generated from a supply chain. This does not include non-operating revenue, such investments and sale of office buildings."	[3,500,000; 10,000,000] (Lima-Junior & Carpinetti, 2019)	$
	y_6 - **return on working capital**: "refers to the magnitude of investment relative to a company's working capital position versus the revenue generated from a supply chain."	[−15; 100] (SCC, 2012)	%
7	x_{28} - **days sales outstanding**: "the length of time from when a sale is made until cash for it is received from customers."	[25; 70] (Ganga & Carpinetti, 2011)	Days
	x_{29} - **inventory days of supply**: "the amount of inventory (stock) expressed in days of sales."	[27; 80] (Ganga & Carpinetti, 2011)	Days
	x_{30} - **days payable outstanding**: "the length of time from purchasing materials, labour and/or conversion resources until cash payments must be made."	[30; 72] (Ganga & Carpinetti, 2011)	Days
	y_7 - **cash-to-cash cycle time**: "the time it takes for an investment made to flow back into a company after it has been spent for raw materials."	[22; 120] (SCC, 2012)	Days

Source: SCC (2012).

accomplishes the desired accuracy level, one of the actions listed in Fig. 6 should be taken following the sequence in which they are listed. In this way, the learning process of all the candidate topologies should be restarted.

5. Illustrative application case

An illustrative application case was developed with the objective of: select the best topologies for each one of the seven ANFIS models that compose the proposed performance evaluation system; assess the suitability of these models to estimate the performance figures of the SCOR® level-1 metrics. All of the SCOR® metrics shown in Fig. 5 have been considered in this application case. Table 3 details these metrics and their measurement units according to SCC (2012) and Lima-Junior and Carpinetti (2019). It also indicates the universe of discourse defined for each metric based on some studies from literature. Table 4 details the parameters to be set during the implementation of the proposed model, as well as the values tested in this study using the random subsampling cross-validation method. In order to minimize the amount of inference rules of each ANFIS model, the number of membership functions of the input variables was initially defined as two based on Özkan and Inal (2014). Following Akkoç (2012) and Aengchuan and Phruksaphanrat (2018), triangular and Gaussian membership functions were evaluated since they are more sensitive to variations in the values of the performance metrics. A symmetric partitioning of the input variables was made in order to define the initial values of the membership functions. Based on Jang (1993) and Güneri et al. (2011), the minimum and algebraic product operators were tested to perform the logical operations "AND" in the inference rules. Regarding the types of consequent, linear functions and constant crisp values was evaluated. As proposed by Jang (1993), the weighted average operator was adopted to aggregate the results provide by the rules in all of the ANFIS models implemented. Table 5 presents the set of candidate topologies tested for each ANFIS model. As can be seen, eight candidate topologies were implemented and assessed for each model.

The data used in the training and validation stages of the models was generated by means of the following steps (Lima-Junior & Carpinetti, 2019): (i) Using MS Excel, the performance values of the level-2 metrics were randomly generated considering the intervals of the universe of discourse shown in Table 3; (ii) Based on the expressions suggested by the SCOR® (SCC, 2012), the values of level-1 metrics were obtained using the generated values in the previous step; (iii) The total set of data was subdivided, with 70% of the samples applied in the training stage and 30% used in the validation of the models. In each training subset, some samples of the lower limit values of the universe of discourse of each input variable were included. Samples of the upper limit values of these variables were also inserted; (4) All the sample values were normalized using Eq. (5). The length of the subsets of training and validation for each model was defined based on empirical tests and studies from literature such as Fan et al. (2013) and Lima-Junior and Carpinetti (2019). Table 6 present the amount of samples generated to train and validate each of the ANFIS models, as well as

Table 4
Summary of the topological and training parameters of the ANFIS models.

Parameters	Relevance	Tested values
Number of fuzzy membership functions of the input variables	Defines the partition granularity of the fuzzy variables.	2 (Özkan & Inal, 2014)
Type and initial values of the membership functions	Determine the behavior of input variables. The membership functions are used in the steps of fuzzification and inference.	Gaussian (Akkoç, 2012) Triangular (Aengchuan & Phruksaphanrat, 2018)
Type of consequent of the inference rules	Produces an output value for each activated inference rule.	Linear function (Jang, 1993) Constant value (Jang, 1993)
Fuzzy operator used in layer 2	Aggregate the degrees of membership functions activated in the fuzzification step.	Minimum (Jang, 1993) Algebraic product (Güneri et al., 2011)
Aggregation operator used in layer 5	Aggregate the individual contributions from the activated inference rules into a single crisp value.	Weighted average (Jang, 1993)
Desired accuracy of prediction for each model	Refers to the maximum MSE value tolerable by managers in the stage of use of the system. It also serves as acceptance criterion of candidate topologies in the validation stage.	1×10^{-3} (Fan et al., 2013; Lima-Junior & Carpinetti, 2019)
Learning rate	Determines the magnitude of the gradient used to adjust the membership functions. It also influences how fast the ANFIS model will reach the point of minimization of MSE in the training stage.	0.01 (MathWorks, 2019)
Number of epochs	Defines how many times the training subset should be presented to the ANFIS model.	200 (MathWorks, 2019)

Table 5
Candidate topologies assessed for each of the ANFIS models.

ANFIS model	Candidate topology	Type of membership function	T-norm operator	Type of consequent
1	Topology 1	Triangular	Minimum	Linear function
	Topology 2	Triangular	Minimum	Crisp value
	Topology 3	Triangular	Product	Linear function
	Topology 4	Triangular	Product	Crisp value
	Topology 5	Gaussian	Minimum	Linear function
	Topology 6	Gaussian	Minimum	Crisp value
	Topology 7	Gaussian	Product	Linear function
	Topology 8	Gaussian	Product	Crisp value
2	Topology 9	Triangular	Minimum	Linear function
	Topology 10	Triangular	Minimum	Crisp value
	Topology 11	Triangular	Product	Linear function
	Topology 12	Triangular	Product	Crisp value
	Topology 13	Gaussian	Minimum	Linear function
	Topology 14	Gaussian	Minimum	Crisp value
	Topology 15	Gaussian	Product	Linear function
	Topology 16	Gaussian	Product	Crisp value
3	Topology 17	Triangular	Minimum	Linear function
	Topology 18	Triangular	Minimum	Crisp value
	Topology 19	Triangular	Product	Linear function
	Topology 20	Triangular	Product	Crisp value
	Topology 21	Gaussian	Minimum	Linear function
	Topology 22	Gaussian	Minimum	Crisp value
	Topology 23	Gaussian	Product	Linear function
	Topology 24	Gaussian	Product	Crisp value
4	Topology 25	Triangular	Minimum	Linear function
	Topology 26	Triangular	Minimum	Crisp value
	Topology 27	Triangular	Product	Linear function
	Topology 28	Triangular	Product	Crisp value
	Topology 29	Gaussian	Minimum	Linear function
	Topology 30	Gaussian	Minimum	Crisp value
	Topology 31	Gaussian	Product	Linear function
	Topology 32	Gaussian	Product	Crisp value
5	Topology 33	Triangular	Minimum	Linear function
	Topology 34	Triangular	Minimum	Crisp value
	Topology 35	Triangular	Product	Linear function
	Topology 36	Triangular	Product	Crisp value
	Topology 37	Gaussian	Minimum	Linear function
	Topology 38	Gaussian	Minimum	Crisp value
	Topology 39	Gaussian	Product	Linear function
	Topology 40	Gaussian	Product	Crisp value
6	Topology 41	Triangular	Minimum	Linear function
	Topology 42	Triangular	Minimum	Crisp value
	Topology 43	Triangular	Product	Linear function
	Topology 44	Triangular	Product	Crisp value
	Topology 45	Gaussian	Minimum	Linear function
	Topology 46	Gaussian	Minimum	Crisp value
	Topology 47	Gaussian	Product	Linear function
	Topology 48	Gaussian	Product	Crisp value
7	Topology 49	Triangular	Minimum	Linear function
	Topology 50	Triangular	Minimum	Crisp value
	Topology 51	Triangular	Product	Linear function
	Topology 52	Triangular	Product	Crisp value
	Topology 53	Gaussian	Minimum	Linear function
	Topology 54	Gaussian	Minimum	Crisp value
	Topology 55	Gaussian	Product	Linear function
	Topology 56	Gaussian	Product	Crisp value

instances of samples and their normalized values. All candidate topologies were trained using the learning method originally proposed by Jang (1993). As indicated in Table 4, the value adopted for the learning rate was 0.01 in all the performed trainings, and the maximum number of training epochs was set to 200. The results of the learning process of the candidate topologies are detailed next.

5.1. Results of training and validation of the ANFIS models

The learning processes of the evaluated topologies were carried out using MATLAB© software on a desktop computer with Intel Quad Core i5 processor and 6 Gigabytes of RAM. Table 7 presents the MSE values obtained in the training and validation stages for each of the candidate topologies. For models 1, 2, 3, and 4, these values were obtained using only two membership functions. For models 5, 6, and 7, following the actions recommended in Fig. 6, the number of membership functions used in each input variable was gradually increased in order to achieve a satisfactory level of accuracy. Thus, the results shown in Table 7 for these models were achieved using 5, 4, and 3 membership functions, respectively. As can be seen in Table 7, several of the tested topologies have achieved a value of MSE lower that the maximum value defined based on the literature. The topologies highlighted in bold were chosen since they have accomplished the best prediction accuracy. Among the selected topologies, those that estimate the figures of the performance metric perfect order fulfilment (ANFIS model 1) have reached the best prediction accuracy, with MSE in validation stage equals to 1.9203×10^{-16}. On the other hand, the topology that predicts the performance metric "return on working capital" (ANFIS model 6) has obtained the largest MSE value, equivalent to 3.6508×10^{-7}. The topologies with the second best level of accuracy for each ANFIS model are also highlighted in bold in Table 7. They may be adopted alternatively as they have reached an accuracy level close to the best topologies.

Some linear regression tests were executed in order to analyse the relationship between the target values and the performance figures predicted by the selected topologies. The coefficient of determination (R^2) was computed in order to estimate the proportion of the total variability of the dependent variable *y* (predicted performance figures) that can be explained by the independent variable *x* (output values of the validation subset). The value of R^2 correspond to the square of the correlation coefficient. It is given as a percentage varying from 0 to 1 (Montgomery & Runger, 2011). Fig. 7 present the values for R^2 and the outcomes of regression tests using MS Excel. The target values for each sample are represented in the horizontal axis, whereas the performance figures predicted are in the vertical axis. The expression that indicates the relationship pattern between these data for each ANFIS model is also shown in Fig. 7. In these equations, the linear coefficient indicates the point at which the line crosses the vertical axis, while the angular coefficient represent how much the average of the estimated values by the ANFIS models changes by an increase of one unit of the target values. The results presented in Fig. 7 point out that the performance figures predicted by the ANFIS models are very near to the target

Table 6
Amount of samples generated to train and validate the proposed ANFIS models.

ANFIS model	Training subset	Validation subset	Instance of sample values	Normalized data
1	210	90	$x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 0, y_1 = 3.$	$x_1 = 1, x_2 = 1, x_3 = 1, x_4 = 0, y_1 = 0.75.$
2	350	150	$x_5 = 21500, x_6 = 83749, x_7 = 68365, x_8 = 41208, x_9 = 20825, y_2 = 235647.$	$x_5 = 0.1278, x_6 = 0.2250, x_7 = 0.0735, x_8 = 0.1178, x_9 = 0.0046, y_2 = 0.1008.$
3	350	150	$x_{10} = 3, x_{11} = 7, x_{12} = 2, x_{13} = 12, y_3 = 24.$	$x_{10} = 0.4, x_{11} = 1, x_{12} = 0.1667, x_{13} = 0.3846, y_3 = 0.4667.$
4	700	300	$x_{14} = 163952, x_{15} = 33586, x_{16} = 95529, x_{17} = 180538, x_{18} = 305101, x_{19} = 69307, x_{20} = 130847, x_{21} = 1413240, y_4 = 2392100.$	$x_{14} = 0.1497, x_{15} = 0.3434, x_{16} = 0.3191, x_{17} = 0.1328, x_{18} = 0.3273, x_{19} = 0.9722, x_{20} = 0.5390, x_{21} = 0.1887, y_4 = 0.2563.$
5	350	150	$x_{22} = 309284, x_{23} = 1157403, x_{24} = 869462, y_5 = 21343.$	$x_{22} = 0.1101, x_{23} = 0.4383, x_{24} = 0.2463, y_5 = 0.2901.$
6	350	150	$x_{25} = 2392100, x_{26} = 21343, x_{27} = 3716903, y_6 = 62.072$	$x_{25} = 0.2562, x_{26} = 0.2901, x_{27} = 0.0334, y_6 = 0.6702.$
7	350	150	$x_{28} = 30, x_{29} = 41, x_{30} = 38, y_7 = 33.$	$x_{28} = 0.1111, x_{29} = 0.2642, x_{30} = 0.1905, y_7 = 0.1122.$

Table 7
MSE values obtained for each candidate topology in the stages of training and validation.

ANFIS model	Candidate topology	Training MSE	Validation MSE
1	Topology 1	1.6475×10^{-16}	2.0165×10^{-16}
	Topology 2	1.5647×10^{-15}	1.8201×10^{-15}
	Topology 3	1.6176×10^{-16}	$\mathbf{1.9404 \times 10^{-16**}}$
	Topology 4	1.3906×10^{-15}	1.5924×10^{-15}
	Topology 5	2.5653×10^{-16}	3.4220×10^{-16}
	Topology 6	3.0418×10^{-15}	4.0348×10^{-15}
	Topology 7	1.6622×10^{-16}	$\mathbf{1.9203 \times 10^{-16*}}$
	Topology 8	1.4229×10^{-15}	1.6182×10^{-15}
2	Topology 9	2.7756×10^{-10}	4.1946×10^{-9}
	Topology 10	1.7065×10^{-3}	2.4517×10^{-3}
	Topology 11	4.6719×10^{-10}	$\mathbf{1.5515 \times 10^{-9**}}$
	Topology 12	4.8820×10^{-10}	$\mathbf{9.4665 \times 10^{-10*}}$
	Topology 13	2.5697×10^{-10}	5.2399×10^{-9}
	Topology 14	8.3463×10^{-4}	1.1446×10^{-3}
	Topology 15	2.8656×10^{-10}	1.2049×10^{-8}
	Topology 16	1.4544×10^{-6}	1.9502×10^{-6}
3	Topology 17	1.2229×10^{-10}	2.0424×10^{-10}
	Topology 18	1.6617×10^{-3}	2.3317×10^{-3}
	Topology 19	1.3498×10^{-10}	$\mathbf{1.9474 \times 10^{-10**}}$
	Topology 20	1.5501×10^{-10}	$\mathbf{1.6800 \times 10^{-10*}}$
	Topology 21	1.2099×10^{-10}	2.0461×10^{-10}
	Topology 22	4.3418×10^{-4}	4.8223×10^{-4}
	Topology 23	1.1525×10^{-10}	2.0589×10^{-10}
	Topology 24	2.0359×10^{-6}	2.6265×10^{-6}
4	Topology 25	2.6105×10^{-10}	$\mathbf{1.0611 \times 10^{-6**}}$
	Topology 26	1.3923×10^{-1}	4.5675×10^{-1}
	Topology 27	3.3481×10^{-11}	2.2704×10^{-5}
	Topology 28	6.9080×10^{-12}	$\mathbf{4.4936 \times 10^{-11*}}$
	Topology 29	1.9717×10^{-9}	2.6455×10^{-3}
	Topology 30	6.3605×10^{-2}	8.2100×10^{-1}
	Topology 31	4.7528×10^{-10}	1.1400×10^{-3}
	Topology 32	1.9573×10^{-5}	4.5882×10^{-4}
5	Topology 33	4.0340×10^{-11}	4.0340×10^{-11}
	Topology 34	5.6632×10^{-3}	5.6632×10^{-3}
	Topology 35	2.0062×10^{-12}	$\mathbf{2.0062 \times 10^{-12**}}$
	Topology 36	5.8447×10^{-15}	$\mathbf{5.8447 \times 10^{-15*}}$
	Topology 37	3.5283×10^{-10}	3.5283×10^{-10}
	Topology 38	2.4738×10^{-3}	2.4738×10^{-3}
	Topology 39	2.3991×10^{-10}	2.3991×10^{-10}
	Topology 40	4.7138×10^{-6}	4.7138×10^{-6}
6	Topology 41	1.1141×10^{-4}	1.2540×10^{-4}
	Topology 42	1.1302×10^{-3}	1.5504×10^{-3}
	Topology 43	9.0421×10^{-5}	$\mathbf{9.4790 \times 10^{-5**}}$
	Topology 44	9.1039×10^{-4}	1.2121×10^{-3}
	Topology 45	1.5568×10^{-4}	1.8536×10^{-4}
	Topology 46	3.4380×10^{-3}	3.9489×10^{-3}
	Topology 47	1.4888×10^{-7}	$\mathbf{3.6508 \times 10^{-7*}}$
	Topology 48	1.4342×10^{-4}	1.2485×10^{-4}
7	Topology 49	6.8779×10^{-12}	2.1364×10^{-11}
	Topology 50	5.0631×10^{-4}	7.2481×10^{-4}
	Topology 51	2.7664×10^{-13}	$\mathbf{4.8616 \times 10^{-13**}}$
	Topology 52	1.0227×10^{-13}	$\mathbf{1.1935 \times 10^{-13*}}$
	Topology 53	1.5607×10^{-11}	4.4410×10^{-11}
	Topology 54	2.2740×10^{-4}	2.5449×10^{-4}
	Topology 55	1.1842×10^{-11}	3.5011×10^{-11}
	Topology 56	9.7333×10^{-7}	1.5973×10^{-5}

*Topologies with the best prediction accuracy.
**Topologies with the second best prediction accuracy.

values, since the values of the linear coefficient are close to zero and most of the angular coefficients are equals to 1. Additionally, in the majority of the tests, the value for R^2 is equal to 1. The only exception is the ANFIS model 6, whose value of R^2 correspond to 0.9954.

With the purpose of illustrating the inference process using the ANFIS models, Fig. 8 presents the decision rule base of model 2 after the learning process. This model is used to estimate the performance figures of the overall value at risk (y_2) based on the risks of planning (x_5), sourcing (x_6), make (x_7), delivery (x_8), and return (x_9). The antecedent part of the decision rules are in the first up to the fifth columns, which are composed by the input variables with their respective triangular membership functions. The consequent part is depicted in the sixth column. In this example, the values $x_5 = 1$, $x_6 = 0.5$, $x_7 = 0.5$, $x_8 = 0.5$, and $x_9 = 1$ are fuzzified so as to activate the rules 18, 20, 22, 24, 26, 28, 30, and 32. The aggregation of the values yielded by these rules using the weighted average operator resulted in the value 0.7 for the metric y_2. The surface plots presented in Fig. 9 illustrate how the variation of some input variables affect the output values of the ANFIS models 2 and 5 (indicated as (a) and (b) in Fig. 9). The plot shown in Fig. 9(a) indicates a linear relationship between the two input variables and the output metric overall value at risk, whereas the surface response represented in Figure (b) depicts a nonlinear relationship between the variable x_{25} and the output metric return on working capital.

Table 8 summarizes the results of the computational implementation and details some of the characteristics of the selected topologies. The values obtained for MSE in the validation stage for all the ANFIS models are much lower than the limit value defined based on the literature (1.0×10^{-3}), which evidences that the proposed models present a greater accuracy of prediction. Although 200 training epochs have been performed throughout the training of each of the candidate topologies, the process of minimizing the MSE was finished between 3 and 20 epochs in all the cases. The computational time consumed in the training stage of each model was low. The training of model 4 was the most time consuming, which can be explained by the fact the this model comprises a greater amount of input variables and decision rules.

The use of triangular membership functions in the input variables and constant values in the consequents yielded the best results in the most of the implemented models. In the case of the topologies chosen for the ANFIS models 1 and 6, the use of Gaussian membership functions and linear consequent seems to be more suitable. Regarding the t-norm operator applied to modelling the logical connectives of the inference rules, the algebraic product operator provided better results than the minimum operator in all the chosen topologies. Among the actions suggested in Fig. 6 to increase the accuracy of ANFIS models, it was only needed an increase in the number of membership functions of the input variables of models 5, 6, and 7 in order to attain a satisfactory accuracy. Thus, for most of the models, the use of only two membership functions produced more accurate results.

When compared with the studies that applied artificial intelligence techniques to evaluate suppliers or supply chains based on a particular set of metrics, the MSE values reached by the chosen topologies are lower than the ones achieved by Didehkhani et al. (2009), Jassbi et al. (2010), Güneri et al. (2011), Efendigil and Önüt (2012), Fan et al. (2013), Özkan and Inal (2014), and Tavana, Fallahpour, Caprio, and Santos-Arteaga (2016). The MSE values obtained by these topologies have been also smaller than the ones accomplished by Lima-Junior and Carpinetti (2019) when using multilayer perceptron neural networks to predict the performance figures of the SCOR® level-1 metrics. Thus, the proposed ANFIS models reached a level of accuracy better than the previous similar studies found in the literature.

6. Validation of the results using paired t-test

In order to check whether there is a significant difference between the performance values estimated by the ANFIS models (Y_i') and the target values for each SCOR® level-1 metric (Y_i), seven hypotheses tests were executed using paired t test. According Montgomery and Runger (2011), this type of test is suitable when the data of the populations are gathered in pairs. Let (Y_1, Y_1'), (Y_2, Y_2'), …, (Y_n, Y_n') be a set of n paired samples, where μ_1 represents the mean of the population represented by Y and μ_2 symbolize the mean of the population Y'. The differences among each pair of observations are represented by $D_i = Y_i - Y_i'$ ($i = 1, 2, \ldots, n$). Assuming that the values of D_i are normally distributed with average $\mu_D = \mu_1 - \mu_2$ and standard deviation S_D, the objective of the paired t-test is to verify whether μ_D equals to a given value Δ_0. Thus, if there is no significant difference between the two sets of samples, the values of μ_D must be zero ($\mu_D = \Delta_0 = 0$). Table 9 describe the test as

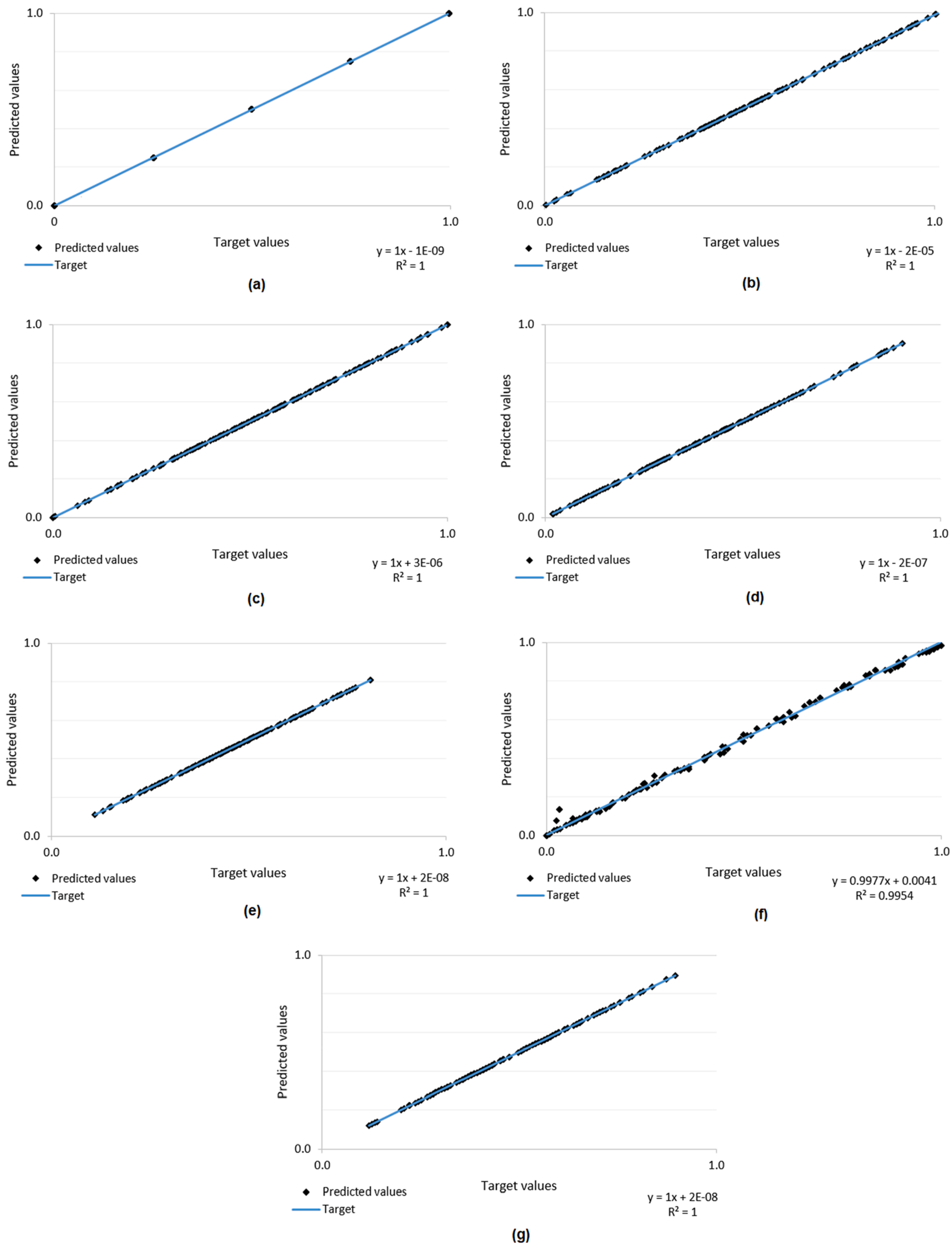

Fig. 7. Results of the linear regression tests and values of R^2 calculated for the ANFIS models 1(a), 2(b), 3(c), 4(d), 5(e), 6(f), and 7(g).

well the rejection criterion of the null hypothesis for a test with significance level α.

In order to carry out the hypothesis tests, a set of 30 samples of the target values (Y) and the performance figures estimated using the chosen topologies (Y') was adopted. The target values was obtained following the procedure described in section 5 for data generation, while the estimated values were calculated in the validation stage of the learning process of the ANFIS models. Table 10 presents the target

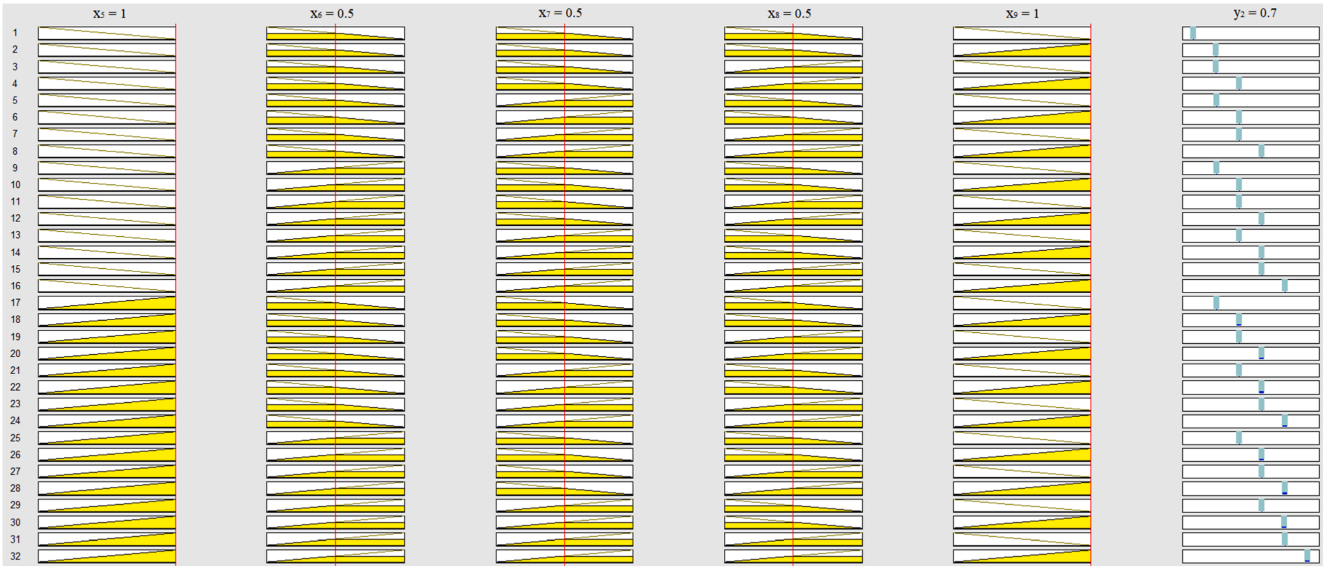

Fig. 8. Inference rules of the ANFIS model 2 after the learning process.

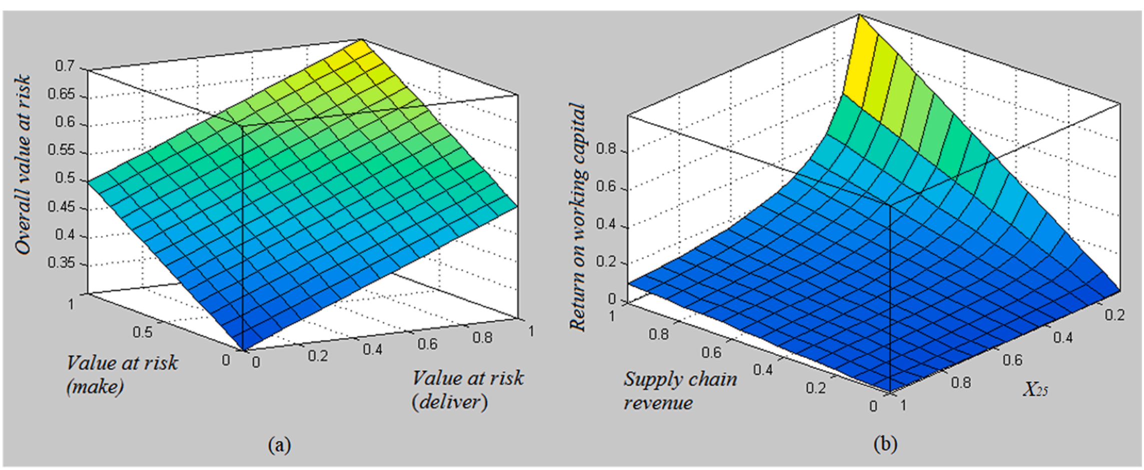

Fig. 9. Surface plots of some variables of the ANFIS models 2 (a) and 5 (b).

values and the estimated values for the metrics “perfect order fulfilment”, “overall value at risk”, and “order fulfilment cycle time”. The samples of “total cost to serve”, “return on working capital”, and “cash-to-cash cycle time” are presented in Table 11, which also shows the output values of the ANFIS model 5 that are applied to estimate the performance figures of “return on work capital”.

Table 12 presents the results of the hypothesis tests considering a level of significance $\alpha = 0.05$. It shows the values for the mean and the standard deviation of the distribution of the differences between the pairs of observations showed in Tables 10 and 11, as well as $\mathbf{t}_{\alpha/2,\mathbf{n}-1}$, T_0 and *p-value*. In all the tests, the calculated value for *p-value* is greater than the level of significance adopted. In addition, all values obtained for T_0 are outside the rejection region of the null hypothesis. Therefore, since the null hypothesis cannot be rejected, it is concluded that there is no significant difference between the target values and the estimated performance figures, which reinforce the accuracy of prediction of the proposed ANFIS models by this study.

Table 8
Results of the implementation of the ANFIS models using the random subsampling cross-validation method.

ANFIS model	Chosen topology	Computational time of training process	Type of membership function	Type of consequent	T-norm operator	Number of partitions of the input variables	Number of inference rules	Validation MSE
1	Topology 7	13 sec	Gaussian	Linear function	Product	2	16	1.9203×10^{-16}
2	Topology 12	14 sec	Triangular	Crisp value	Product	2	32	9.4665×10^{-10}
3	Topology 20	7 sec	Triangular	Crisp value	Product	2	16	1.6800×10^{-10}
4	Topology 28	549 sec	Triangular	Crisp value	Product	2	256	4.4936×10^{-11}
5	Topology 36	84 sec	Triangular	Crisp value	Product	5	125	5.8447×10^{-15}
6	Topology 47	9 sec	Gaussian	Linear function	Product	4	64	3.6508×10^{-7}
7	Topology 52	9 sec	Triangular	Crisp value	Product	3	27	1.1935×10^{-13}

Table 9
Statistical test and rejection criterion of the null hypothesis.

Null hypothesis: H_0: $\mu_D = 0$
Test statistic: $T_0 = \frac{\bar{D}}{S_D/\sqrt{n}}$
Alternative hypothesis: H_1: $\mu_D \neq 0$
Rejection region (for two-tailed test): $t_0 > t_{\alpha/2,n-1}$ or $t_0 < -t_{\alpha/2,n-1}$
Reject H_0 if the *p-value* is $< \alpha$

Source: Montgomery and Runger (2011).

Table 10
Values used in the hypothesis tests (ANFIS models 1–4).

Perfect order fulfilment (ANFIS model 1)		Overall value at risk (ANFIS model 2)		Order fulfilment cycle time (ANFIS model 3)	
Y_1	Y_1'	Y_2	Y_2'	Y_3	Y_3'
0.75000	0.75000	0.02910	0.02913	0.35975	0.35975
0.75000	0.75000	0.16070	0.16068	0.66578	0.66580
0.75000	0.75000	0.37320	0.37313	0.37256	0.37258
0.75000	0.75000	0.55200	0.55201	0.34703	0.34703
0.50000	0.50000	0.68430	0.68433	0.23563	0.23565
0.50000	0.50000	0.47450	0.47450	0.50199	0.50195
0.50000	0.50000	0.48640	0.48632	0.49515	0.49515
0.25000	0.25000	0.40940	0.40942	0.57437	0.57435
0.25000	0.25000	0.47240	0.47242	0.60917	0.60915
0.00000	0.00000	0.40580	0.40573	0.47743	0.47743
1.00000	1.00000	0.60280	0.60274	0.68433	0.68433
0.75000	0.75000	0.59090	0.59083	0.44109	0.44108
0.75000	0.75000	0.37090	0.37088	0.52120	0.52120
0.75000	0.75000	0.47980	0.47982	0.56321	0.56320
0.75000	0.75000	0.52750	0.52755	0.57010	0.57010
0.50000	0.50000	0.42460	0.42455	0.66933	0.66933
0.50000	0.50000	0.63780	0.63784	0.46043	0.46045
0.50000	0.50000	0.56670	0.56672	0.64970	0.64970
0.25000	0.25000	0.34980	0.34978	0.70267	0.70268
0.25000	0.25000	0.57030	0.57032	0.56194	0.56195
0.00000	0.00000	0.50880	0.50879	0.71196	0.71195
1.00000	1.00000	0.45580	0.45579	0.67680	0.67678
0.75000	0.75000	0.53660	0.53650	0.35284	0.35288
0.75000	0.75000	0.59260	0.59261	0.71811	0.71812
0.75000	0.75000	0.47040	0.47041	0.33342	0.33340
0.75000	0.75000	0.56740	0.56745	0.35337	0.35338
0.50000	0.50000	0.77130	0.77132	0.49136	0.49138
0.50000	0.50000	0.70790	0.70788	0.14617	0.14617
0.50000	0.50000	0.50230	0.50238	0.58883	0.58880
0.25000	0.25000	0.40400	0.40405	0.60729	0.60728

7. Conclusion

This study proposed a new intelligent system for supply chain performance evaluation based on the combination between the SCOR® level 1 and 2 metrics with ANFIS neuro-fuzzy models. In total, 56 topologies were assessed using the random subsampling cross-validation method in order to choose the most appropriate one for each ANFIS model. For all the studied topologies, the algebraic product operator generated better outcomes than the minimum operator. In most of the cases, the utilization of triangular membership functions for the input variables and constant values in the consequent part of the inference rules resulted in a greater accuracy of prediction. In addition, in four of the seven proposed ANFIS models, it was more adequate to adopt fewer membership functions. It contributed to minimize the number of inference rules, decrease the amount of adaptive parameters and consequently made the learning process faster.

Several tests were carried out to assess the accuracy of the proposed models. The MSE values obtained by each topology throughout the learning process (Tables 7 and 8) indicate that the ANFIS model 1 reached the higher accuracy of prediction (1.9203×10^{-16}), while the model 6 achieved the lowest one (3.6508×10^{-7}). The lowest accuracy of model 6 may be due to the fact that it models a nonlinear relationship between the input and output metrics, which is more complicated than the linear relationships quantified by the other models. The results of the hypothesis tests using paired samples indicate that there is no significant difference between the target values and the performance values estimated by each chosen topology, which strengthens the suitability of usage of ANFIS neuro-fuzzy systems for the evaluation of supply chain performance.

The proposed system enables operations managers to quantify the supply chain performance under different perspectives, providing information that allows analyzing the gap between the expected and achieved performance level in each SCOR® level-1 metric. In this way, managers can elaborate action plans aiming at enhancing the outcomes of the metrics that present underperformance. Use of the proposed system allows managers to evaluate the effectiveness of their strategies, thus contributing to the focus company being more proactive in the search for better performance results. Over time, managers can switch the metrics used in the proposed evaluation system. However, as with other intelligent systems based on fuzzy inference and artificial neural networks, changes of metrics require that the ANFIS models be reconstructed and trained using samples of new variables. In comparison to the supply chain performance evaluation models found in the literature, the system proposed by this study has the following advantages:

- Differently of the models proposed by Didehkhani et al. (2009), Jassbi et al. (2010), and Fan et al. (2013), the proposed system uses a broad set of metrics related to different performance attributes, including reliability, agility, responsiveness, cost, and asset management. The use of the metrics suggested by Supply Chain Council (SSC, 2012) contributes to a better alignment, standardization, and integration of performance measures across different tiers of a supply chain. Another advantage is the compatibility of the SCOR® metrics with the SCORmark database, which enables benchmarking based on the comparison between the estimated performance figures and performance database from several supply chains around the world;
- In contrast with the supply chain performance evaluation models that combine SCOR® metrics with mathematical programming approaches (Liu & Liu, 2017; Zhang & Reimann, 2014) or multicriteria decision methods (Akkawuttiwanich & Yenradee, 2018; Bukhori et al., 2015; Clivillé & Berrah, 2012; Dissanayake & Cross, 2018; Golparvar & Seifbarghy, 2009; Jalalvand et al., 2011; Kocaoglu et al., 2013; Moharamkhani et al., 2017; Sellitto et al., 2015; Theeranuphattana & Tang, 2008; Yang & Jiang, 2012), the application of ANFIS models allows modelling non-linear causal relationships between input and output metrics, as well as enables adaptation to environment of use based on historical data;
- Unlike the model based on Mamdani fuzzy inference developed by Ganga and Carpinetti (2011), with manual adjustment of the parameters based on the judgments of specialists, the proposed system is able to automatically tune the fuzzy input variables and the inference rules based on a supervised learning procedure. When compared to models based on artificial neural networks (Fan et al., 2013; Wang, 2013; Lima-Junior & Carpinetti, 2019), ANFIS models require fewer training times to adjust their adaptive parameters. In addition, it requires a smaller number of training because it does not contain random parameters such as neural networks. Therefore, in addition to better prediction accuracy, the use of ANFIS makes the learning process faster and contributes to a greater agility of the modelling process;
- One benefit over the models based on artificial neural networks (Fan et al., 2013; Wang, 2013; Lima-Junior & Carpinetti, 2019) and other techniques based on crisp numbers (Golparvar & Seifbarghy, 2009; Liu & Liu, 2017; Zhang & Reimann, 2014) refers to the ability to modelling imprecise values and uncertain variables by means of the fuzzy representation of the input variables. The ANFIS models seeks

Table 11
Values used in the hypothesis tests (ANFIS models 4–7).

Total cost to serve (ANFIS model 4)		Input used to predict return on work capital (ANFIS model 5)		Return on work capital (ANFIS model 6)		Cash-to-cash cycle time (ANFIS model 7)	
Y_4	Y_4'	Y_5	Y_5'	Y_6	Y_6'	Y_7	Y_7'
0.31500	0.31500	0.27267	0.27267	0.24000	0.23967	0.53333	0.53333
0.30625	0.30625	0.43200	0.43200	0.58330	0.60263	0.28889	0.28889
0.45375	0.45375	0.39800	0.39800	0.00000	0.00071	0.31852	0.31852
0.88125	0.88125	0.28000	0.28000	0.60000	0.59000	0.58519	0.58519
0.46000	0.46000	0.57233	0.57233	0.40000	0.40771	0.51852	0.51852
0.10250	0.10250	0.43100	0.43100	0.35000	0.34646	0.89630	0.89630
0.34750	0.34750	0.54533	0.54533	0.28000	0.27592	0.29630	0.29630
0.56250	0.56250	0.47733	0.47733	0.22860	0.23714	0.42963	0.42963
0.54625	0.54625	0.46933	0.46933	0.76920	0.77037	0.22222	0.22222
0.59250	0.59250	0.43300	0.43300	0.85710	0.85538	0.22222	0.22222
0.39000	0.39000	0.14800	0.14800	0.68180	0.69397	0.13333	0.13333
0.51250	0.51250	0.25400	0.25400	0.73330	0.75158	0.55556	0.55556
0.12125	0.12125	0.66267	0.66267	0.46000	0.45064	0.60000	0.60000
0.11000	0.11000	0.80767	0.80767	0.53330	0.55339	0.40741	0.40741
0.30000	0.30000	0.61933	0.61933	0.13500	0.12164	0.51852	0.51852
0.77500	0.77500	0.11000	0.11000	0.36000	0.34320	0.42963	0.42963
0.35125	0.35125	0.28667	0.28667	0.24570	0.26560	0.37778	0.37778
0.53875	0.53875	0.22467	0.22467	0.10000	0.10736	0.70370	0.70370
0.18500	0.18500	0.53533	0.53533	0.20000	0.19383	0.42222	0.42222
0.53625	0.53625	0.42400	0.42400	0.03330	0.13525	0.22222	0.22222
0.41375	0.41375	0.46033	0.46033	0.81820	0.82549	0.53333	0.53333
0.30875	0.30875	0.69700	0.69700	0.08330	0.08666	0.45185	0.45185
0.62625	0.62625	0.51800	0.51800	0.90000	0.88589	0.45185	0.45185
0.49125	0.49125	0.19433	0.19433	0.60000	0.61313	0.57037	0.57037
0.50500	0.50500	0.20400	0.20400	0.50000	0.48720	0.57778	0.57778
0.24625	0.24625	0.47633	0.47633	0.40000	0.38908	0.25185	0.25185
0.26000	0.26000	0.37367	0.37367	0.00000	−0.00021	0.41482	0.41481
0.41000	0.41000	0.15167	0.15167	1.00000	0.98221	0.40741	0.40741
0.09625	0.09625	0.35933	0.35933	0.27270	0.27526	0.59259	0.59259
0.17125	0.17125	0.72333	0.72333	0.19170	0.19395	0.37037	0.37037

Table 12
Results of the hypothesis tests using paired *t*-test.

ANFIS model	$\bar{D}$	S_D	$t_{\alpha/2,n-1}$	T_0	p-value
1	1.371×10^{-8}	2.027×10^{-17}	2.04523	−1.41930	0.16647
2	3.507×10^{-5}	6.769×10^{-10}	2.04523	0.51918	0.60758
3	1.173×10^{-5}	1.046×10^{-10}	2.04523	0.52546	0.60326
4	6.730×10^{-7}	3.816×10^{-13}	2.04523	−0.17958	0.85870
5	6.712×10^{-9}	2.871×10^{-17}	2.04523	−0.43393	0.66755
6	1.220×10^{-2}	3.280×10^{-4}	2.04523	−1.05507	0.30010
7	3.020×10^{-7}	2.740×10^{-14}	2.04523	0.19818	0.84429

to imitate human reasoning by establishing cause and effect relationships between variables as well as considering uncertain scenarios. In this way, the output value of the system is calculated considering the combination of all the scenarios represented by the triggered rules with their respective degrees of activation;

- Another benefit is that the procedure for calculating the performance values of the output variables using ANFIS models is more transparent and easy to understand than those used in the artificial neural networks, since it allows to identify which decision rules justify the yielded results. Therefore, the greater interpretability of information provided by the decision rules can bring more confidence to operations managers when dealing with decision making processes aiming at improving the supply chain performance.

Another contribution of this study is to suggest the most appropriate topological parameters when using the SCOR® level 1 and 2 metrics, as well as to provide some guidelines for the creation of ANFIS models and for conducting the learning process using the random subsampling cross-validation method. Thus, the steps suggested in Fig. 6 can be adopted by other studies that aim to develop new intelligent decision support systems based on ANFIS models, which is even more relevant in face of the trend towards intelligent systems in the context of industry 4.0.

Regarding limitations of the proposed performance evaluation system, the main one is the difficulty of collecting enough data to carry out the learning process of the ANFIS models. Due to this difficulty, it has not yet been possible to apply the proposed system in a real case. The difficulty of collecting data to assess the supply chain performance has been pointed out in several studies (Didehkhani et al., 2009; Brandenburg, Govindan, Sarkis, & Seuring, 2014; Dias & Ierapetritou, 2017; Lima-Junior & Carpinetti, 2017). Dias and Ierapetritou (2017) points out that currently most companies use a variety of different IT tools in order to manage their supply chains. However, such tools are often not integrated with each other. Thus, data from the different levels of decision making are generally collected and kept in different functional areas within the company. As a result, the different decision makers involved in the supply chain management do not have access to all the data they require to make optimal decisions (Dias & Ierapetritou, 2017). However, factors such as greater integration of processes and performance measurement systems across the supply chain tiers, as well as the popularization of data management technologies such as Big Data and Data Warehouse, can contribute to increase data availability and to facilitate implementation of the proposed system in the coming years (Büyüközkan & Göçer, 2018; Lima-Junior & Carpinetti, 2019).

Another limitation of the proposed system is related to the amount of input variables and fuzzy partitions that can be adopted for each ANFIS model. If many partitions are adopted, there will be a large number of possible combinations between them, and consequently the number of inference rules will also be high. In this case, a larger amount of training samples to tune the topological parameters may be needed. In addition, it may affect the accuracy of the results provided by the system.

Further studies can implement the proposed system in real cases aiming to evaluate its usability by users who are not specialists in neuro-fuzzy systems, as well assess its applicability to benchmarking using SCORmark. In these applications, different learning algorithms can be tested. Other suggestions for further studies are: the development of ANFIS models to estimate the values of the SCOR® level-2 metrics based on the performance figures of the SCOR® level-3 metrics; the application of other computational intelligence techniques to deal with supply chain performance evaluation, including the recent methods based on the fuzzy hesitant sets theory; and the comparative evaluation of artificial intelligence techniques aiming to identify advantages and disadvantages of use when applied to assess the performance of supply chains.

Acknowledgement

To FAPESP (2016/14618-4) and CAPES (Phd scholarship – code 001) for funding this study.

References

Akkoç, S. (2012). An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: The case of Turkish credit card data. *European Journal of Operational Research, 222*, 168–178.

APICS, 2018. American Production and Inventory Control Society. Benchmark your Supply Chain: Close Performance Gaps. http://www.apics.org/sites/apics-supply-chain-council/products-and-services/benchmarking.

Aengchuan, P., & Phruksaphanrat, B. (2018). Comparison of fuzzy inference system (FIS), FIS with artificial neural networks (FIS + ANN) and FIS with adaptive neuro-fuzzy inference system (FIS + ANFIS) for inventory control. *Journal of Intelligent Manufacturing, 29*, 905–923.

Akkawuttiwanich, P., & Yenradee, P. (2018). Fuzzy QFD approach for managing SCOR performance indicators. *Computers & Industrial Engineering, 122*, 189–201.

Akyuz, G. A., & Erkan, T. E. (2010). Supply chain performance measurement: A literature review. *International Journal of Production Economics, 48*, 5137–5155.

Bai, C., Sarkis, J., Wei, X., & Koh, L. (2012). Evaluating ecological sustainable performance measures for supply chain management. *Supply Chain Management: An International Journal, 17*, 78–92.

Bals, L., Schulze, H., Kelly, S., & Stek, K. (2019). Purchasing and supply management (PSM) competencies: Current and future requirements. *Journal of Purchasing and Supply Management, 25* 100572.

Beamon, B. M. (1999). Measuring supply chain performance. *International Journal of Operations & Production Management, 19*, 275–292.

Brandenburg, M., Govindan, K., Sarkis, J., & Seuring, S. (2014). Quantitative models for managing supply chain risks. *Europen Journal of Operational Research, 233*, 299–312.

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.

Büyüközkan, G., & Göçer, F. (2018). Digital supply chain: Literature review and a proposed framework for future research. *Computers and Industrial Engineering, 97*, 157–177.

Chopra, S., & Meindl, P. (2013). *Supply chain management: Strategy, planning and operation* (5th ed.). Pearson, Upper Saddle River, N.J: Prentice Hall.

Chorfi, Z., Benabbou, L., & Berrado, A. (2018). An integrated performance measurement framework for enhancing public health care supply chains. *Supply Chain Forum: An International Journal, 19*, 191–203.

Clivillé, V., & Berrah, L. (2012). Overall performance measurement in a supply chain: Towards a supplier-prime manufacturer based model. *Journal of Intelligent Manufacturing, 23*, 2459–2469.

Delipinar, G. E., & Kocaoglu, B. (2016). Using SCOR model to gain competitive advantage: A Literature review. *Procedia - Social and Behavioral Sciences, 229*, 398–406.

Dewan, M. W., Huggett, D. J., Liao, T. W., Wahab, M. A., & Okeil, A. M. (2016). Prediction of tensile strength of friction stir weld joints with adaptive neuro-fuzzy inference system (ANFIS) and neural network. *Materials and Design, 92*, 288–299.

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.

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

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.

Dubey, R., Gunasekaran, A., Papadopoulos, T., Childe, S. J., Shibin, K. T., & Wamba, S. F. (2017). Sustainable supply chain management: Framework and further research directions. *Journal of Cleaner Production, 142*, 1119–1130.

Efendigil, T., & Önüt, S. (2012). An integration methodology based on fuzzy inference systems and neural approaches for multi-stage supply-chains. *Computers & Industrial Engineering, 62*, 554–569.

Elgazzar, S., Tipi, N., & Jones, G. (2019). Key characteristics for designing a supply chain performance measurement system. *International Journal of Productivity and Performance Management, 68*, 296–318.

Erdas, O., Andac, C. A., Gurkan-Alp, A. S., Alpaslan, F., & Buyukbingol, E. (2013). Compressed images for affinity prediction (CIFAP): A study on predicting binding affinities for checkpoint kinase 1 protein inhibitors. *Chemometrics, 27*, 155–164.

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*, 247–258.

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 Bionic Engineering, 10*, 383–395.

Ganga, G. M. D., & Carpinetti, L. C. R. (2011). A fuzzy logic approach to supply chain performance management. *International Journal of Production Economics, 134*, 177–187.

Golparvar, M., & Seifbarghy, M. (2009). Application of SCOR model in an oil- producing company. *Journal of Industrial Engineering, 4*, 59–69.

Guersola, M., Lima, E. P., & Steiner, M. T. A. (2018). Supply chain performance measurement: A systematic literature review. *Int. J. Logistics Systems and Management, 31*, 109–137.

Gunasekaran, A., Patel, C., & Mcgaughey, R. E. (2004). A framework for supply chain performance measurement. *International Journal of Production Economics, 87*, 333–347.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations & Production Management, 21*, 71–87.

Güneri, 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.

Handfield, R. B., & Nichols, E. L. (1999). *Introduction to supply chain management.* Prentice-Hall.

Jalalvand, F., Teimoury, E., Makui, A., Aryanezhad, M. B., & Jolai, F. (2011). A method to compare supply chains of an industry. *Supply Chain Management: An International Journal, 16*, 82–97.

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

Jassbi, J., Seyedhosseini, S. M., & Pilevari, N. (2010). An adaptive neuro fuzzy inference system for supply chain agility evaluation. *International Journal Of Industrial Engineering & Production Research, 20*, 187–196.

Kar, S., Das, S., & Ghosh, P. K. (2014). Applications of neuro fuzzy systems: A brief review and future outline. *Applied Soft Computing, 15*, 243–259.

Ko, M., Tiwari, A., & Mehnen, J. (2010). A review of soft computing applications in supply chain management. *Applied Soft Computing, 10*, 661–674.

Kocaoglu, B., Gülsün, B., & Tanya, M. (2013). A SCOR based approach for measuring a benchmarkable supply chain performance. *Journal of Intelligent Manufacturing, 24*, 113–132.

Kolus, A., Imbeau, D., Dubé, P., & Dubeau, D. (2015). Adaptive neuro-fuzzy inference systems with k-fold cross-validation for energy expenditure predictions based on heart rate. *Applied Ergonomics, 50*, 68–78.

Lakri, S., Dallery, Y., & Jemai, Z. (2015). Measurement and management of supply chain performance: Practices in today's large companies. *Supply Chain Forum: An International Journal, 16*, 16–30.

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.

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.

Liu, F. F., & Liu, Y. (2017). A methodology to assess the supply chain performance based on gap-based measures. *Computers & Industrial Engineering, 110*, 550–559.

Makris, D., Hansen, Z. N. L., & Khan, O. (2019). Adapting to supply chain 4.0: An explorative study of multinational companies. *Supply Chain Forum: An International Journal, 20*, 116–131.

Mathworks. MATLAB® Fuzzy Logic Tool™ - Users'Guide. The MathWorks, Inc., USA, March 2019.

Melnyk, S. A., Narasimhan, R., & Decampos, H. A. (2014). Supply chain design: Issues, challenges, frameworks and solutions. *International Journal of Production Research, 52*, 1887–1896.

Melnyk, S. A., Stewart, D. M., & Swink, M. S. A. (2004). Metrics and performance measurement in operations management: Dealing with the metrics maze. *Journal of Operations Management, 22*, 209–217.

Mentzer, J. T., Witt, W., Keebler, J. S., Min, S., Nix, N., & Smith, C. D. (2001). Defining supply chain management. *Journal of Business Logistics, 22*, 1–25.

Moharamkhani, A., Bozorgi-Amiri, A., & Mina, H. (2017). Supply chain performance measurement using SCOR model based on interval-valued fuzzy TOPSIS. *International Journal of Logistics Systems and Management*, 27.

Montgomery, D. C., & Runger, G. C. (2011). *Applied statistics and probability for engineers* (5th ed.). John Wiley & Sons.

Najmi, A., Gholamian, M. R., & Makui, A. (2013). Supply chain performance models: A literature review on approaches, techniques, and criteria. *Journal of Operations and Supply Chain Management, 6*, 94–113.

Ntabe, E. N., LeBel, L., Munson, A. D., & Santa-Eulalia, L. A. (2015). A systematic literature review of the supply chain operations reference (SCOR) model application with special attention to environmental issues. *International Journal of Production Economics, 109*, 310–332.

Osiro, L., Lima-Junior, F. R., & Carpinetti, L. C. R. (2018). A group decision model based on quality function deployment and hesitant fuzzy for selecting supply chain sustainability metrics. *Journal of Cleaner Production, 183*, 964–978.

Özkan, G., & Inal, M. (2014). Comparison of neural network application for fuzzy and ANFIS approaches for multi-criteria decision making problems. *Applied Soft Computing, 24*, 232–238.

Qi, Y., Huo, B., Wang, Z., & Yeung, H. Y. J. (2017). The impact of operations and supply chain strategies on integration and performance. *International Journal of Production Economics, 185*, 162–174.

Rajab, S., & Sharma, V. (2018). A review on the applications of neuro-fuzzy systems in business. *Artificial Intelligence Review, 49*(4).

SCC - Supply chain council. (2012). Supply chain operations reference model, version 11. 0. Supply Chain Council, 2012.

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*, 4917–4926.

Seuring, S., & Müller, M. (2008). From a literature review to a conceptual framework for sustainable supply chain management. *Journal of Cleaner Production, 16*, 1699–1710.

Sundarakani, B., Razzak, H. A., & Manikandan, S. (2018). Creating a competitive advantage in the global flight catering supply chain: A case study using SCOR model. *International Journal of Logistics Research and Applications, 21*, 481–501. https://doi.org/10.1080/13675567.2018.1448767.

Tavana, M., Fallahpour, A., Caprio, D., & Santos-Arteaga, F. (2016). A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. *Expert Systems with Applications, 61*, 129–144.

Theeranuphattana, A., & Tang, J. C. S. (2008). A conceptual model of performance measurement for supply chains: Alternative considerations. *Journal of Manufacturing Technology Management, 19*, 125–148.

Unahabhokha, C., Platts, K., & Tan, H. T. (2007). Predictive performance measurement system: A fuzzy expert system approach. *Benchmarking: An International Journal, 14*, 77–91.

Wang, H. (2013). Research on supply chain performance evaluation of fresh agricultural products. *INMATEH - Agricultural Engineering, 40*, 36–42.

Yang, J., & Jiang, H. (2012). Fuzzy evaluation on supply chains' overall performance based on AHM and M(1,2,3). *Journal of software, 12*, 2779–2786.

Zhang, W., & Reimann, M. (2014). Towards a multi-objective performance assessment and optimization model of a two-echelon supply chain using SCOR metrics. *Central European Journal of Operations Research, 22*, 591–622.