

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Encountered Problems of Time Series with Neural Networks: Models and Architectures

Paola Andrea Sánchez-Sánchez, José Rafael García-González and Leidy Haidy Perez Coronell

Abstract

The growing interest in the development of forecasting applications with neural networks is denoted by the publication of more than 10,000 research articles present in the literature. However, the high number of factors included in the configuration of the network, the training process, validation and forecasting, and the sample of data, which must be determined in order to achieve an adequate network model for forecasting, converts neural networks in an unstable technique, given that any change in training or in some parameter produces great changes in the prediction. In this chapter, an analysis of the problematic around the factors that affect the construction of the neural network models is made and that often present inconsistent results, and the fields that require additional research are highlighted.

Keywords: time series, prediction of neural networks, learning algorithms

1. Introduction

The time series forecasting has received a lot of attention in recent decades, due to the growing need to have effective tools that facilitate decision making and overcome the theoretical, conceptual, and practical limitations presented by traditional approaches. The classification of forecasting methods from a statistical point of view, in general, has two aspects, one oriented to causal methods, such as regression and intervention models, and the other to time series, where mobile averages, exponential smoothing, ARIMA models, and neural networks are included. Under this current, the forecast is oriented only to the task of predicting the behavior, prioritizing forward vision and thus obviating many important steps in the model construction process; while the modeling is oriented to find the global structure, model and formulas, which explain the behavior of the data generating process and can be used to predict trends of future behavior (long term), as well as to understand the past. This last vision allows the construction of solid models in its foundation and under which the forecast is seen as an additional step.

The representation of time series with dynamics of nonlinear behavior has acquired great weight in the last decades, because many authors agree in affirming that the real world series present nonlinear behaviors, and the approximation that can be done with linear models, it is inadequate [1–3]. Although approximations have been made with statistical models (an extensive compilation of these is

presented by [4–6]), its representation is difficult to restrict its use to a functional form a priori, for which neural networks have proven to be a valuable tool since they allow to extract the unknown nonlinear dynamics present between the explanatory variables and the series, without the need to perform any assumptions.

The growing interest in the development of forecasting applications with neural networks is denoted by the publication of more than 10,000 research articles in the literature [7]. However, as stated by Zhang et al. [8], inconsistent results about the performance of neural networks in the prediction of time series are often reported in the literature. Many conclusions are obtained from empirical studies, thus presenting limited results that often cannot be extended to general applications and that are not replicable. Cases where the neural network presents a worse performance than linear statistical models or other models may be due to the fact that the series studied do not present high volatilities, that the neural network used to compare was not adequately trained, that the criterion of selection of the best model is not comparable, or that the configuration used is not adequate to the characteristics of the data. Whereas, many of the publications that indicate superior performance of neural networks are related to novel paradigms or extensions of existing methods, architectures, and training algorithms, but lack a reliable and valid evaluation of the empirical evidence of their performance. The high number of factors included in the configuration of the network, the training process, validation and forecast, and the sample of data, which is required to determine to achieve a suitable network model for the forecast, makes neural networks a technique unstable, given that any change in training or in some parameter produces large changes in the prediction [9]. In this chapter, an analysis of the problematic environment is made to the factors that affect the construction of neural network models and that often present inconsistent results.

Empirical studies that allow the prediction of time series with particular characteristics such as seasonal patterns, trends, and dynamic behavior have been reported in the literature [10–12]; however, few contributions have been made in the development of systematic methodologies that allow representing time series with neural networks on specific conditions, limiting the modeling process to ad-hoc techniques, instead of scientific approaches that follow a methodology and process of replicable modeling.

In the last decade, there has been a considerable number of isolated contributions focused on specific aspects, for which a unified vision has not been presented; Zhang et al. [8] made a deep revision until 1996. This chapter is an effort to evaluate the works proposed in the literature and clarify their contributions and limitations in the task of forecasting with neural networks, highlighting the fields that require additional research.

Although some efforts aimed at the formalization of time series forecasting models with neural networks have been carried out, at a theoretical level, there are few advances obtained [13], which evidences a need to have systematic research about modeling and forecasting of time series with neural networks.

The objective of this chapter is to delve into the problem of forecasting time series with neural networks, through an analysis of the contributions present in the literature and an identification of the difficulties underlying the task of forecasting, thus highlighting the open field research.

2. Motivation of the study

The time series forecasting is considered a generic problem to many disciplines, which has been approached with different models [14]. Formally, the objective of

the time series forecasting is to find a flexible mathematical functional form that approximates with sufficient precision the data generating process, in such a way that it appropriately represents the different regular and irregular patterns that the series may present, allowing the constructed representation to extrapolate future behavior [15]. However, the choice of the appropriate model for each series depends on the characteristics of the time series, and its usefulness is associated with the degree of similarity between the dynamics of the series generating process and the mathematical formulation that is made of it under the premise that the data dictate the tool to be used [16].

As pointed out by Granger and Terasvirta [2], the construction of a model that relates a variable to its own history and/or to the history of other explanatory variables of its behavior can be carried out through a variety of alternatives. These depend both on the functional form by which the relationship is approximated and on the relationship between these variables. Although, each modeler is autonomous in the choice of the modeling tool, in cases where there are relations of a non-linear order, there are limitations in the use of certain types of tools, moreover, this same reason leads to the absence of a method that is the best for all cases. The question that arises is then, how to properly specify the functional form in the presence of non-linear relationships between the time series and the explanatory variables of its behavior.

The representation of time series with dynamics of nonlinear behavior has acquired great weight in the last decades, because many authors agree in affirming that the real world series present nonlinear behaviors, and the approximation that can be done with linear models, it is inadequate [1–3], among others. The approach of series with the stated characteristics has been made, among others, from statistical models, combined or hybrid models and neural networks. The complexity in the representation of non-linear relationships lies in the fact that in most cases, there are not enough physical or economic laws that allow us to specify a suitable functional form for their representation.

The literature has proposed a wide range of statistical models for the representation of series with nonlinear behavior such as bilinear models autoregressive thresholds—TAR, autoregressive soft transition—STAR [17, 18], autoregressive conditional heteroscedasticity—ARCH [19], and its generalized form—GARCH [20]; a comprehensive compilation of these is presented by [4–6]. Although the stated models have proved useful in particular problems, they are not universally applicable, since they limit the form of non-linearity present in the data to empirical specifications of the characteristics of the series based on the available information [2]; its success in practical cases depends on the degree to which the model used manages to represent the characteristics of the series studied. However, the formulation of each family of these models requires the specification of an appropriate type of non-linearity, which is a difficult task compared to the construction of linear models, since there are many possibilities (wide variety of possible non-linear functions), more parameters to be calculated, and more errors can be made [21, 22].

Likewise, in the prediction of time series, it is universally accepted that a simple method is not the best in all situations [23–25]. This is because real-world problems are often complex in nature and a model of this kind may not be adequate to capture different patterns. Empirical studies suggest that by combining different models, the accuracy of the representation may be better than for the individual case [26–28]. Therefore, the union of models with different characteristics increases the possibility of capturing different patterns in the data and provides a more appropriate representation of the time series. The hybrid modeling then arises, naturally as the union of similar or different techniques with complementary characteristics.

In the forecast literature, several combinations of methods have been proposed. However, many of them use similar methods, and this is how different studies about hybrid linear modeling techniques are found in the traditional literature. Although this type of combinations has demonstrated its ability to improve the accuracy of the representations made, it is considered that a more effective route could be based on models with different characteristics. Both theoretical and empirical evidence suggest that the combination of dissimilar models, or those that strongly disagree with others, leads to a decrease in model errors [29, 30] and allows, in addition, to reduce the uncertainty of this one [31]. The hybrid model is thus, more robust to estimate the possible changes in the structure of the data.

Numerous applications have been proposed in the literature based on combinations of linear models with computational intelligence [32–39]. However, the main criticisms of these works is that they do not contemplate the need to integrate subjective information into models, which, like traditional statistical models, require a preprocessing of the series, which is aimed at eliminating the visible components of this one and that require the determination of a large number of parameters, which are not economically explainable.

Neural networks seen as a non-parametric non-linear regression technique have emerged as attractive alternatives to the problem posed, since they allow extracting the unknown nonlinear dynamics present between the explanatory variables and the series, without the need to make any kind of assumptions. From this family of techniques, multi-layer perceptron networks—MLP, understood as a non-linear statistical regression model, have received great attention among researchers from the computational intelligence and statistics community.

The attractiveness of neural networks in the prediction of time series is their ability to identify hidden dependencies based on a finite sample, especially of a non-linear order, which gives them the recognition of universal approximation of functions [3, 40–42]. Perhaps the main advantage of this approach over other models is that they do not start from a priori assumptions about the functional relationship of the series and its explanatory variables, a highly desirable characteristic in cases where the mechanism generating the data is unknown and unstable [43], in addition to its high generalization capacity allows to learn behaviors and extrapolate them, which leads to better forecasts [5].

For artificial intelligence, as well as for operation research, the time series forecasting with neural networks is seen as a problem of error minimization, which consists of adjusting the parameters of the neural network in order to minimize the error between the real value and the output obtained. Although, this criterion allows obtaining models whose output is increasingly closer to the desired one, it is to the detriment of the parsimony of the model, since it leads to more complex representations (a large number of parameters). From the statistical point of view, a criterion based solely on the reduction of the error is not the most optimal, it is necessary a development oriented to the formalization of the model, which requires the fulfillment of certain properties that are not always taken into account, such as the stability of the calculated parameters, the coherence between the series and the model, the consistency with the previous knowledge and the predictive capacity of the model.

The evident interest in the use of neural networks in the prediction of time series has led to the emergence of an enormous research activity in the field. Crone and Kourentzes [7] reveal more than 5000 publications in prediction of time series with neural networks (see also publications [39, 44, 45]), and journals in fields with econometrics, statistics, engineering, and artificial intelligence, even being the central topic in special editions, such as the case of *Neurocomputing* with “*Special issue on evolving solution with neural networks*” published in October 2003 [46] and

the *International Journal of Forecasting* with “Special issue on forecasting with artificial neural networks and computational intelligence” published in 2011.

In order to establish the relevance of the prediction of time series with neural networks, a search was made through *Science Direct* of the Journals that publish articles related to the topic. **Table 1** and **Figure 1** present a compilation of the 10 Journals with more publications and also relate the number of articles published in the years 2015–2019, 2010–2014, 2005–2009, 2000–2004 and 1999 and earlier, which is identified using keywords: (*Forecasting o Prediction, Neural Networks, and Time Series*).

An analysis of **Table 1** and **Figure 1** shows the following facts:

- The number of publications reported on the subject is increasing, being representative the drastic growth reported in the last 5 years (2015–2019), which is evident in all the magazines listed.
- There is a greater participation in journals pertaining to or related to the fields of engineering and artificial intelligence.
- Journals with high number of published articles, *Neurocomputing*, *Applied Soft Computing*, *Procedia Computer Science*, and *Expert Systems with Applications*, are closely related to the topic, both from contributions in the field of neural networks, and time series forecasting.

Many comparisons have been made between neural networks and statistical models in order to measure the prediction performance of both approaches. As stated by Zhang et al. [8]:

“There are many inconsistent reports in the literature on the performance of ANNs for forecasting tasks. The main reason is that a large number of factors including network structure, training method, and sample data may affect the forecasting ability of the networks.”

| Journal | Articles identified using keywords (forecasting or prediction, neural networks, and time series) | | | | | Total |
|---|--|-----------|-----------|-----------|----------------|-------|
| | 2015–2019 | 2010–2014 | 2005–2009 | 2000–2004 | 1999 and antes | |
| <i>Energy</i> | 308 | 83 | 17 | 3 | 2 | 413 |
| <i>Applied energy</i> | 297 | 90 | 10 | 4 | — | 401 |
| <i>Neurocomputing</i> | 254 | 148 | 88 | 46 | 37 | 573 |
| <i>Renewable and sustainable energy reviews</i> | 241 | 74 | 14 | — | 1 | 330 |
| <i>Applied soft computing</i> | 238 | 132 | 41 | 5 | — | 416 |
| <i>Journal of hydrology</i> | 233 | 166 | 102 | 34 | 5 | 540 |
| <i>Expert systems with applications</i> | 226 | 364 | 188 | 20 | 11 | 809 |
| <i>Procedia computer science</i> | 212 | 77 | — | — | — | 289 |
| <i>Renewable energy</i> | 191 | 49 | 22 | 5 | 3 | 270 |
| <i>Energy procedia</i> | 155 | 41 | — | — | — | 196 |
| | 2355 | 1224 | 482 | 117 | 59 | 4237 |

Table 1.
 Journals that publish time series forecast articles with neural networks.

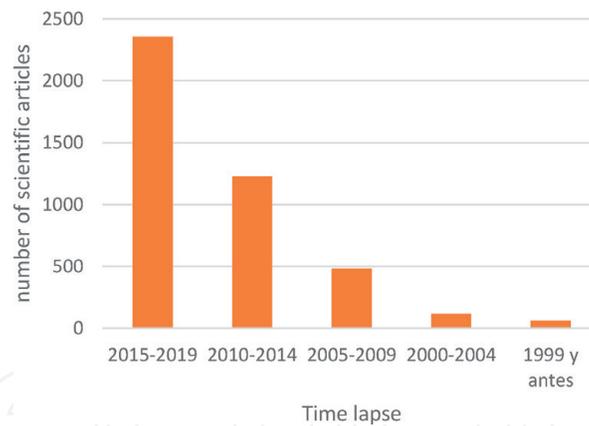


Figure 1.
Published articles for forecasting time series with neural networks.

Such inconsistencies make neural networks an unstable method, given that any change in training or in some parameter produces large changes in prediction [9]. Some key factors where mixed results are presented are:

- Need for data preprocessing (scaling, transformation, simple and seasonal differentiation, etc.) [10–12, 47, 48].
- Criteria for the selection of input variables [15, 22].
- Criteria for the selection of the network configuration. Complexity vs. Parsimony (number of internal layers [40–42], neurons in each layer [22]).
- Estimation of the parameters (learning algorithms, stop criteria, etc.).
- Criteria for selecting the best model [43].
- Diagnostic tests and acceptance.
- Tests on the residuals. Consistency of linear tests.
- Properties of the model: stability of the parameters, mean and variance series versus model.
- Predictive capacity of the model.
- Presence of regular patterns such as: trends, seasonal, and cyclical patterns [10–12].
- Presence of irregular patterns such as: structural changes, atypical data, effect of calendar days, etc. [3, 49, 50].

Cases where the neural network presents a worse performance than linear statistical models or other models, may be due to the fact that the series studied do not present a great disturbance, that the neural network used to compare was not adequately trained, that the criterion of selection of the best model is not comparable, or that the configuration used is not adequate to the characteristics of the data. Many conclusions about the performance of neural networks are obtained from empirical studies, thus presenting limited results that often cannot be extended to general applications.

However, there are few systematic researches about the modeling and prediction of time series with neural networks and the theoretical advances obtained [13], and this is perhaps the primary cause of the inconsistencies reported in the literature.

Many of the optimistic publications that indicate superior performance of neural networks are related to novel paradigms or extensions of existing methods, architectures and training algorithms, but lack a reliable and valid evaluation of the empirical evidence of their performance. Few contributions have been made in the systematic development of methodologies that allow representing time series with neural networks on specific conditions, limiting the modeling process to ad-hoc techniques, instead of scientific approaches that follow a methodology and replicable modeling process. A consequence of this is that, despite the empirical findings, neural network models are not fully accepted in many forecast areas. The previous discussion leads us to think that, although progress has been made in the field, there are still topics open to investigate. The question of whether, because, and on what conditions the models of neural networks are better is still valid.

3. Difficulties in the prediction of time series with neural networks

The design of an artificial neural network is intended to ensure that for certain network inputs, it is capable of generating a desired output. For this, in addition to a suitable network topology (architecture), a learning or training process is required, which allows modifying the weights of the neurons until finding a configuration according to the relationship measured by some criterion and thus estimating the parameters of the network, a process that is considered critical in the field of neural networks [8, 43]. Model selection is not a trivial task in forecasting linear models and is particularly difficult in non-linear models, such as neural networks. Because the set of parameters to be estimated is typically large, neural networks often suffer from over-training problems. That is, they fit the training data very well but produce poor results in the forecast.

To mitigate the effect of over-training, the available data set is often divided into three parts: training, validation, and testing or prediction. The training and validation sets are used to build the neural network model and then be evaluated with the test set. The training set is used to estimate the parameters of an alternative number of neural network specifications (networks with different numbers of inputs and hidden neurons). The generalization capacity of the network is evaluated with the validation set. The network model that performs best in the validation set is selected as the final model. The validity and utility of the model is then tested using the test set. Often this last set is used for forecasting purposes, and the network's generalization capacity for unknown data is evaluated.

The criterion of selecting the model based on the best performance of the validation set, however, does not guarantee that the model has a good fit in the forecast set, and the selection of the appropriate amount of data in each set can also affect performance. This is how a large training set can lead to over-training. Granger [21] suggests that at least 20% of the data be used as a test set; however, there is no general guide on how to partition the set of observations, so that optimal results are guaranteed.

Zhang et al. [22] states that the size of the training set has limited effects on the performance of the network, where, for the sizes investigated by the authors, there is no significant difference in the performance of the forecast. These results are perhaps due to the forecasting method used, with little difference for prediction one step ahead, and marked for multi-step forecast, in which case large differences in the results are expected in the case of different sizes of the training, validation, and test sets.

Although, as a criterion for the selection of the best model, the minimization of some error function is often used, such as mean square error (MSE), absolute average deviation (MAD), cost functions [51], or even expert knowledge [52], because the performance of each measure is not the same, since they can favor or penalize certain characteristics in the data, and that, in the case of expert knowledge is not always easy to acquire; approaches based on the use of machine learning [53, 54] and meta-learning [55–59] have been reported in the literature, which show advantages by allowing an automatic process of model selection based on the parallel evaluation of multiple network architectures, but they are limited to the execution of certain architectures and their implementation is complex. Other studies related to the topic include Qi and Zhang [43] who investigate the well-known criteria of AIC [60], BIC [61], square root of the mean square error (RMSE), absolute average percentage deviation (MAPE), and direction of occurrence (DA). The amplified panorama of the techniques for selecting the best model reflects that, despite the effort made, there is not a strong criterion for adequate selection.

Another widespread criticism that is often made to neural networks is the high number of parameters that must be experimentally selected to generate the desired output, such as: the selection of input variables to the neural network from a usually large set of possible entries; the selection of the internal architecture of the network; and the estimation of the values associated with the weights of the connections. For each of the problems mentioned, different approaches to its solution have been proposed in the literature.

The selection of the input variables depends to a large extent on the knowledge that the modeler possesses about the time series, and it is the task of the latter to choose according to some previously fixed criterion the need of each variable within the model. Although there is no systematic way to determine the set of inputs accepted by the research community, recent studies have suggested the use of rational procedures, based on the use of decisional analysis, or traditional statistical methods, such as autocorrelation functions [62]; however, the use of the latter is disregarded since the functions are based on linear approaches and not neural networks do not express by themselves the components of moving averages (MA) of the model. Mixed results about the benefits of including many or few input variables are also reported in the literature. Tang et al. [63] report the benefits of using a large set of input variables, while Lachtermacher and Fuller [15] report the same results for multistep forecasting, but opposed in forecasting a step forward. Zhang et al. [22] said that the number of input variables in a neural network model for prediction is much more important than the number of hidden neurons. Other techniques based on heuristic analysis of the importance of each lag, statistical tests of non-linear dependence Lagrange multipliers, [64, 65]; radio of likelihood, [66]; Biespectro, [67], criteria for identifying the model, such as AIC [5] or evolutionary algorithms [68, 69] they have also been proposals.

The selection of the internal configuration of the neural network (number of hidden layers and neurons in each layer) is perhaps the most difficult process in the construction of the model where more different approaches have been proposed in the literature, demonstrating in this way the interest of the scientific community to solve this problem.

Regarding the number of hidden layers, theoretically a neural network with a hidden capacity and a sufficient number of neurons can approximate the accuracy of a continuous function in a compact domain. However, in practice, some authors say that the use of a hidden layer when the time series is continuous, and twice if there is some type of discontinuity [41, 42]. However, other research has shown that a network with two hidden layers can result in a more compact architecture and with a high efficiency than networks with a single hidden layer [70–72]. Increasing the number of hidden layers only increases computational time and the danger of overtraining.

With respect to the number of hidden neurons, a small number means that the network cannot adequately learn the relationships in the data, while a large number causes the network to memorize the data with a poor generalization and little utility for prediction. Some authors propose that the number of hidden neurons should be based on the number of input variables; however, this criterion is in turn related to the extension of the time series and the sets of training, validation, and prediction. Given that the value of the weights in each neuron depends on the degree of error between the desired value and that predicted by the network, the selection of the optimal number of hidden neurons is directly associated with the training process used.

The training of a neural network is a problem of non-restricted non-linear minimization in which the weights of the network are iteratively modified in order to minimize the error between the desired output and the obtained one. Several methods have been proposed in the literature for the training of the neural network, going through the classical gradient descendant techniques [73], which have convergence problems and are robust, adaptive dynamic optimization [74, 75], Quickprop [76], Levenberg–Marquardt [77], Cuasi-Newton, BFGS, GRG2 [78], among others. However, the joint selection of hidden neurons and the training process has led to the development of fixed, constructive, and destructive methods, where those based on constructive algorithms have certain advantages over others, since they allow evaluating the convenience of adding or not adding a new one. Neuron to the network, during training, according to it decreases the term of the error, which makes them more efficient methods, although with high computational cost [79]. Other developments such as pruning algorithms (*pruning algorithm*) [77, 80–82], Bayesian algorithms, based on Genetic algorithms as the GANN, neural networks with rugged assemblies, assembled learning [83–86], and meta-learning [9, 87] have also shown good results in the task of finding the optimal architecture of the network; however, these methods are usually more complex and difficult to implement. Furthermore, none of them can guarantee to find the optimal global solution and they are not universally applicable for all real forecasting problems, thus designing a proper neural network.

The efficiency of the prediction with neural networks has been evidenced through the applications published in the literature; however, the power of the prediction produced is limited to the degree of stability of the time series and can fail when it presents complex dynamic behaviors. This is how representations that use dynamic character models, such as neural networks with recurrence Elman, Jordan, etc., emerge as an alternative solution [88–91], which due to the possibility of accumulating dynamic behaviors are able to allow more adequate forecasts. The recurrence feature allows forward and backward connections (recurrent or feedback), forming cycles within the network architecture, which uses previous states as a basis for the current state, and allowing to preserve an internal memory of the behavior of the data, which facilitates the learning of dynamic relationships. However, their main criticism lies in the need they impose an efficient training algorithm that allows them to capture the dynamics of the series, its use being computationally complex. Potentially useful models to address the problem of series with dynamic behavior arise from the combination of different architectures in the input of the multilayer perceptron.

The problem that arises goes beyond the simple estimation of each model in light of the characteristics of each series. Although it is recognized that there is much experience gained in multilayer perceptron neural networks, there are still many theoretical, methodological, and empirical problems open about the use of such models. These general problems are related to the aspects listed below, and for which many of the recommendations given in the literature are contradictory [92–98]:

- There is no systematic way accepted in the literature to determine the appropriate set of inputs to the neural network.
- There is no general guide to partition the set of observations in training, validation and forecast, in such a way that optimal results are guaranteed.
- The effects that factors such as partition in training sets, validation and forecast, preprocessing, transfer function, etc., in different forecasting methods are unknown or unclear.
- There are no clear indications that allow to express a priori which transfer function should be used in the neural network model according to the characteristics of the time series.
- There is no clarity about procedures oriented to the selection of neurons in the hidden layer that in turn allow to minimize the training time of the network.
- There are no empirical, methodological or theoretical reasons to prefer a specific model among several alternatives.
- There is no agreement on how to select the final model when several alternatives are considered.
- It is not clear when and how to transform the data before performing the modeling.
- There is no clarity about the necessity of eliminating or not eliminating trend and seasonal components in neural network models.
- It is difficult to incorporate qualitative, subjective, and contextual information in the forecasts.
- There is little understanding of the statistical properties of different neural network architectures.
- There is no clarity about which are the most adequate procedures for the estimation, validation, and testing of different neural network architectures.
- There is no clarity on how to combine forecasts from several alternative models, and if there are gains derived from this practice.
- There are no or no clarity in the criteria for evaluating the performance of different neural network architectures.
- There is no clarity about whether and under what criteria, different architectures of neural networks allow the handling of dynamic behaviors in the data.

4. Conclusions

In this chapter, the need to have adequate models of neural networks for the prediction of time series has been identified, and this task has been exposed as a difficult, relevant, and timely problem.

A critical step in the forecast process is the selection of the set of input variables. At this point the decision of which lags of the series to include is fundamental for the result and depends on the available information and knowledge. Obviating the need for prior knowledge about the series, the choice of candidate lags to be included in the model should be based on a heuristic analysis of the importance of each lag, a statistical test of non-linear dependence, criteria for identifying the model or evolutionary algorithms, however, before such options the mixed results reported in the literature show that there is no consensus about what is the appropriate procedure for this purpose.

As previously emphasized, in the literature there are no clear indications about the best practices for choosing the size of the training, test, and prediction sets. Often the size is a predefined parameter in the construction of the neural network model or it is chosen randomly; however, there is no study that demonstrates the effect that this decision entails, moreover, this may be related to the forecasting method used.

Likewise, there is a close relationship between the selection of the internal configuration, especially the hidden neurons, and the training process of the neural network. The consensus about the use of a hidden layer when the data of the time series are continuous and two when there is discontinuity, and of the advantages of the functions sigmoidia and hyperbolic tangent in the transfer of the hidden layer, reflects a deep investigation of such topics.

It is often used as a criterion for the selection of the best model based on the error of prediction, expert knowledge or criteria of information; however, the limitations that they manifest and the mixed results reported in their use, in addition to the limited results reported with other techniques, which do not allow conclusive conclusions about their use.

The consideration of characteristic factors of the time series that can affect the evolution of the neural network model such as the length of the time series, the frequency of the observations, the presence of regular and irregular patterns, and the scale of the data, must be included in the process of building the neural network model. The discussion of whether a preprocessing oriented to the stabilization of the series is necessary in non-linear models, and even more, in neural networks, is a topic that is still valid, and depends to a large extent on the type of data that is modeled. The abilities exhibited by neural networks allow, in the first instance, to avoid pre-processing via data transformation. However, it is not yet clear whether, under a correct network construction and training procedure, a prior process of elimination of seasonal trends and patterns is necessary. Scaling is always preferable given its advantages of reducing training patterns and leading to more accurate results.

Likewise, the benefits that different neural network architectures have in relation to nonlinear relationships in the data have been discussed. Neural network models, by themselves, facilitate the representation of non-linear characteristics, without the need for a priori knowledge about such relationships, and such consideration is always desirable in models for real time series; however, it is not. In addition, their performance in the face of dynamic behavior in the data, the exposed architectures have been developed as an extension of neural network models and not explicitly as time series models, so there is no theoretical foundation for the construction of these, nor rigorous studies that allow to assess their performance in time series with the stated characteristics.

Conflict of interest

The authors declare no conflict of interest.

IntechOpen

IntechOpen

Author details

Paola Andrea Sánchez-Sánchez*, José Rafael García-González
and Leidy Haidy Perez Coronell
Universidad Simón Bolívar, Barranquilla, Colombia

*Address all correspondence to: psanchez9@unisimonbolivar.edu.co

IntechOpen

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Zhang P. An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*. 2001;**28**(12):1183-1202
- [2] Granger C, Terasvirta T. *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press; 1993
- [3] Franses P, Van Dijk D. *Non-Linear Time Series Models in Empirical Finance*. UK: Cambridge University Press; 2000
- [4] Tong H. *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford Statistical Science Series; 1990
- [5] De Gooijer I, Kumar K. Some recent developments in non-linear modelling, testing, and forecasting. *International Journal of Forecasting*. 1992;**8**:135-156
- [6] Peña D. Second-generation time-series models: A comment on 'Some advances in non-linear and adaptive modelling in time-series analysis' by Tiao and Tsay. *Journal of Forecasting*. 1994;**13**:133-140
- [7] Crone S, and Kourentzes N. Input-variable Specification for Neural Networks - An Analysis of Forecasting low and high Time Series Frequency. *Proceedings of the International Joint Conference on Neural Networks, (IJCNN'09)*. in press. 2009
- [8] Zhang P, Patuwo B, Hu M. Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*. 1998;**14**(1):35-62
- [9] Yu L, Wang S, Lai K. A neural-network-based nonlinear metamodeling approach to financial time series forecasting. *Applied Soft Computing*. 2009;**9**:563-574
- [10] Franses P, Draisma G. Recognizing changing seasonal patterns using artificial neural networks. *Journal of Econometrics*. 1997;**81**(1):273-280
- [11] Qi M, Zhang P. Trend time-series modeling and forecasting with neural networks. *IEEE Transactions on Neural Networks*. 2008;**19**(5):808-816
- [12] Zhang P, Qi M. Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*. 2005;**160**:501-514
- [13] Trapletti A. *On Neural Networks as Time Series Models*. Universidad Técnica de Wien; 2000
- [14] Kasabov N. *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*. 2nd ed. Massachusetts: The MIT Press Cambridge; 1998
- [15] Lachtermacher G, Fuller J. Backpropagation in time-series forecasting. *Journal of Forecasting*. 1995;**14**:381-393
- [16] Meade N. Evidence for selection of forecasting methods. *Journal of Forecasting*. 2000;**19**:515-535
- [17] Granger C, Anderson A. *An Introduction to Bilinear Time Series Models*. Gottingen: Vandenhoeck and Ruprecht; 1978
- [18] Tong H, Lim K. Threshold autoregressive, limit cycles and cyclical data. *Journal of the Royal Statistical Society, Series B*. 1980;**42**(3):245-292
- [19] Engle R. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica*. 1982;**50**:987-1008
- [20] Bollerslev T. Generalised autoregressive conditional heteroscedasticity. *Journal of Econometrics*. 1986;**31**:307-327

- [21] Granger C. Strategies for modelling nonlinear time-series relationships. *Economic Record*. 1993;**69**(206):233-238
- [22] Zhang P, Patuwo E, Hu M. A simulation study of artificial neural networks for nonlinear time-series forecasting. *Computers and Operations Research*. 2001;**28**(4):381-396
- [23] Chatfield C. What is the “best” method of forecasting? *Journal of Applied Statistics*. 1988;**15**:19-39
- [24] Jenkins G. Some practical aspects of forecasting in organisations. *Journal of Forecasting*. 1982;**1**:3-21
- [25] Makridakis S, Anderson A, Carbone R, Fildes R, Hibon M, Lewandowski R, et al. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*. 1982;**1**:111-153
- [26] Clemen R. Combining forecasts: A review and annotated bibliography with discussion. *International Journal of Forecasting*. 1989;**5**:559-608
- [27] Makridakis S, Chatfield C, Hibon M, Lawrence M, Mills T, Ord K, et al. The M2 competition: A real-time judgmentally based forecasting competition. *Journal of Forecasting*. 1993;**9**:5-22
- [28] Newbold P, Granger C. Experience with forecasting univariate time series and the combination of forecasts (with discussion). *Journal of the Royal Statistical Society*. 1974;**137**:131-164
- [29] Granger C. Combining forecasts- twenty years later. *Journal of Forecasting*. 1989;**8**:167-173
- [30] Krogh A, Vedelsby J. Neural network ensembles, cross validation, and active learning. *Advances in Neural Information Processing Systems*. 1995;**7**:231-238
- [31] Chatfield C. Model uncertainty and forecast accuracy. *Journal of Forecasting*. 1996;**15**:495-508
- [32] Bates J, Granger C. The combination of forecasts. *Operational Research Quarterly*. 1969;**20**:451-468
- [33] Davison M, Anderson C, Anderson K. Development of a hybrid model for electrical power spot prices. *IEEE Transactions on Power Systems*. 2002;**2**:17
- [34] Luxhoj J, Riis J, Stensballe B. A hybrid econometric-neural network modeling approach for sales forecasting. *International Journal of Production Economics*. 1996;**43**:175-192
- [35] Makridakis S. Why combining works? *International Journal of Forecasting*. 1989;**5**:601-603
- [36] Palm F, Zellner A. To combine or not to combine? issues of combining forecasts. *Journal of Forecasting*. 1992;**11**:687-701
- [37] Reid D. Combining three estimates of gross domestic product. *Economica*. 1968;**35**:431-444
- [38] Winkler R. Combining forecasts: A philosophical basis and some current issues. *International Journal of Forecasting*. 1989;**5**:605-609
- [39] Zhang P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*. 2003;**50**:159-175
- [40] Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*. 1989;**2**:303-314
- [41] Hornik K. Approximation capability of multilayer feedforward networks. *Neural Networks*. 1991;**4**:251-257
- [42] Hornik K, Stinchcombe M, White H. Multilayer feedforward

- networks are universal approximators. *Neural Networks*. 1989;2(5):359-366
- [43] Qi M, Zhang P. An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research*. 2001;132:666-680
- [44] Adya M, Collopy F. How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*. 1998;17:481-495
- [45] Hill T, O'Connor M, Remus W. Neural network models for time series forecasts. *Management Science*. 1996;42:1082-1092
- [46] Fanni A, Uncini A. Special issue on evolving solution with neural networks. *Neurocomputing*. 2003;55(3-4):417-419
- [47] Faraway J, Chatfield C. Time series forecasting with neural networks: a comparative study using the airline data. *Applied Statistics*. 1998;47:231-250
- [48] Nelson M, Hill T, Remus T, O'Connor M. Time series forecasting using NNs: Should the data be deseasonalized first? *Journal of Forecasting*. 1999;18:359-367
- [49] Hill T, Marquez L, O'Connor M, Remus W. Artificial neural networks for forecasting and decision making. *International Journal of Forecasting*. 1994;10:5-15
- [50] Tkacz G, Hu S. Forecasting GDP Growth Using Artificial Neural Networks. Bank of Canada; 1999
- [51] Tashman L. Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting*. 2000;16:437-450
- [52] Adya M, Collopy F, Armstrong J, Kennedy M. Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting*. 2001;17(2):143-157
- [53] Arinze B. Selecting appropriate forecasting models using rule induction. *Omega-International Journal of Management Science*. 1994;22(6):647-658
- [54] Venkatachalan A, Sohl J. An intelligent model selection and forecasting system. *Journal of Forecasting*. 1999;18:167-180
- [55] Giraud-Carrier R, Brazdil P. Introduction to the special issue on meta-learning. *Machine Learning*. 2004;54(3):187-193
- [56] Santos P, Ludermir T, Prudencio R. Selection of time series forecasting models based on performance information. In: *Proceedings of the 4th International Conference on Hybrid Intelligent Systems*. 2004. pp. 366-371
- [57] Santos P, Ludermir T, Prudencio R. Selecting neural network forecasting models using the zoomed-ranking approach. In: *Proceedings of the 10th Brazilian Symposium on Neural Networks SBRN '08*. 2008. pp. 165-170
- [58] Soares C, Brazdil P. Zoomed Ranking – Selection of classification algorithms based on relevant performance information. *Lecture Notes in Computer Science*. 1910;2000:126-135
- [59] Vilalta R, Drissi Y. A perspective view and survey of meta-learning. *Journal of Artificial Intelligence Review*. 2002;18(2):77-95
- [60] Akaike H. A new look at statistical model identification. *IEEE Transactions on Automatic Control*. 1974;9:716-723
- [61] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978;6:461-464

- [62] Tang Z, Fishwick P. Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing*. 1993;5(4):374-385
- [63] Tang Z, Almeida C, Fishwick P. Time series forecasting using neural networks vs Box-Jenkins methodology. *Simulation*. 1991;57(5):303-310
- [64] Luukkonen R, Saikkonen P, Terasvirta T. Testing linearity in univariate time series models. *Scandinavian Journal of Statistics*. 1988;15:161-175
- [65] Saikkonen P, Luukkonen R. Lagrange multiplier tests for testing non-linearities in time series models. *Scandinavian Journal of Statistics*. 1988;15:55-68
- [66] Chan W, Tong H. On tests for non-linearity in time series analysis. *Journal of Forecasting*. 1986;5:217-228
- [67] Hinich M. Testing for Gaussianity and linearity of a stationary time series. *Journal of Time Series Analysis*. 1982;3:169-176
- [68] Happel B, Murre J. The design and evolution of modular neural network architectures. *Neural Networks*. 1994;7:985-1004
- [69] Schiffmann W, Joost M, Werner R. Application of genetic algorithms to the construction of topologies for multilayer perceptron. In: *Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms*. 1993. pp. 675-682
- [70] Srinivasan D, Liew A, Chang C. A neural network short-term load forecaster. *Electric Power Systems Research*. 1994;28:227-234
- [71] Zhang X. Time series analysis and prediction by neural networks. *Optimization Methods and Software*. 1994;4:151-170
- [72] Chester D. Why two hidden layers are better than one. In: *Proceedings of the International Joint Conference on Neural Networks*. 1990. pp. 1265-1268
- [73] Bishop C. *Neural Networks for Pattern Recognition*. Oxford University Press; 1995
- [74] Pack D, El-Sharkawi M, Marks R, Atlas L. Electric load forecasting using an artificial neural network. *IEEE Transactions on Power Systems*. 1991;6(2):442-449
- [75] Yu X, Chen G, Cheng S. Dynamic learning rate optimization of the backpropagation algorithm. *IEEE Transactions on Neural Networks*. 1995;6(3):669-677
- [76] Falhman S. Faster-learning variations of back-propagation: An empirical study. In: *Proceedings of the 1988 Connectionist Models Summer School*. 1989. pp. 38-51
- [77] Cottrell M, Girard B, Girard Y, Mangeas M, Muller C. Neural modeling for time series: a statistical stepwise method for weight elimination. *IEEE Transactions on Neural Networks*. 1995;6(6):1355-1364
- [78] Lasdon L, Waren A. *GRG2 User's Guide*. Austin: School of Business Administration, University of Texas; 1986
- [79] Weigend A, Rumelhart D, Huberman B. Generalization by weight-elimination with application to forecasting. *Advances in Neural Information Processing Systems*. 1991;3:875-882
- [80] Karnin E. A simple procedure for pruning back-propagation trained neural networks. *IEEE Transactions on Neural Networks*. 1990;1(2):239-245

- [81] Reed R. Pruning algorithms a survey. *IEEE Transactions on Neural Networks*. 1993;**4**:740-747
- [82] Siestema J, Dow R. Neural net pruning – why and how. In: *Proceedings of the IEEE International Conference on Neural Networks*. Vol. 1. 1998. pp. 325-333
- [83] Breiman L. Combining predictors de Combining Artificial Neural Nets—Ensemble and Modular Multi-Net Systems. Berlin: Springer; 1999. pp. 31-50
- [84] Carney J, Cunningham P. Tuning diversity in bagged ensembles. *International Journal of Neural Systems*. 2000;**10**:267-280
- [85] Hansen L, Salamon P. Neural network ensembles. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1990;**12**:993-1001
- [86] Naftaly U, Intrator N, Horn D. Optimal ensemble averaging of neural networks. *Network: Computation in Neural Systems*. 1997;**8**:283-296
- [87] Chan P, Stolfo S. Metalearning for multistrategy and parallel learning. In: *Proceedings of the Second International Workshop on Multistrategy Learning*. 1993. pp. 150-165
- [88] Connor J, Atlas L, Martin D. Recurrent Networks and NARMA Modeling de Advances in Neural Information Processing Systems. Morgan Kaufmann Publishers, Inc. 1991;**119**:301-308
- [89] Kuan C, Liu T. Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics*. 1995;**10**:347-364
- [90] Najand M, Bond C. Structural models of exchange rate determination. *Journal of Multinational Financial Management*. 2000;**10**:15-27
- [91] Tenti P. Forecasting foreign exchange rates using recurrent neural networks. *Applied Artificial Intelligence*. 1996;**10**:567-581
- [92] Caire P, Hatabian G, Muller C. Progress in forecasting by neural networks. In: *Proceedings of the International Joint Conference on Neural Networks*. Vol. 2. 1992. pp. 540-545
- [93] Ong P, Zainuddin Z. Optimizing wavelet neural networks using modified cuckoo search for multi-step ahead chaotic time series prediction. *Applied Soft Computing*. 2019;**80**:374-386
- [94] Zhanga Y, Wanga X, Tang H. An improved Elman neural network with piecewise weighted gradient for time series prediction. *Neurocomputing*. 2019;**359**:199-208
- [95] Wang L, Wang Z, Qu H, Liu S. Optimal forecast combination based on neural networks for time series forecasting. *Applied Soft Computing*. 2018;**66**:1-17
- [96] Lopez-Martin M, Carro B, Sanchez-Esguevillas A. Neural network architecture based on gradient boosting for IoT traffic prediction. *Future Generation Computer Systems*. 2019;**100**:656-673
- [97] Zurbarán M, Sanmartin P. Efectos de la Comunicación en una Red Ad-Hoc. *Investigación e Innovación en Ingenierías*. 2016;**4**(1):26-31
- [98] Tealab A. Time series forecasting using artificial neural networks methodologies: A systematic review. *Future Computing and Informatics Journal*. 2018;**3**(2):334-340