Expert Systems with Applications 42 (2015) 4081–4096

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A hybrid model for high-frequency stock market forecasting

Ricardo de A. Araújo [a,b,*], Adriano L.I. Oliveira [b], Silvio Meira [b]

[a] Informatics Department, Federal Institute of Sertão Pernambucano, Ouricuri, PE, Brazil
[b] Informatics Center, Federal University of Pernambuco, Recife, PE, Brazil

ARTICLE INFO

Article history:
Available online 17 January 2015

Keywords:
Artificial neuron
Descending gradient-based learning
Forecasting
High-frequency stock market

ABSTRACT

Several models have been presented to solve the financial time series forecasting problem. However, even with sophisticated techniques, a dilemma arises from all these models, called the random walk dilemma (RWD). In this context, the concept of time phase adjustment was proposed to overcome such dilemma for daily frequency financial time series. However, the evolution of trading platforms had increased the frequency for performing operations on the stock market to fractions of seconds, which makes needed the analysis of high-frequency financial time series. In this way, this work presents a model, called the increasing decreasing linear neuron (IDLN), for high-frequency stock market forecasting. Also, a descending gradient-based method with automatic time phase adjustment is presented for the proposed model design. Besides, an experimental analysis is conduced using a set of high-frequency financial time series from the Brazilian stock market, and the achieved results overcame those obtained by establishing forecasting models in the literature.

1. Introduction

Information is an invaluable asset for modern society. The identification of needed information to build a predictor is a key element for its success. In this context, there are several challenges to the development of efficient solutions for the financial time series forecasting problem, since it can be found complex behaviors within this kind of time series, such as irregularities, volatility, trends and noise (de A. Araújo, 2012b).

Several approaches have been investigated for the development of forecasting models to estimate the future of temporal phenomena. Linear and nonlinear statistical predictors have been proposed to deal with a wide kind of time series. A particular model, called the autoregressive integrated moving average (ARIMA) (Box, Jenkins, & Reinsel, 1994), has been commonly employed to predict financial time series. However, the ARIMA is a linear model, and this represents a restriction of its practical use, since nonlinear relationships can be found within financial time series (Clements, Franses, & Swanson, 2004). In this context, a wide number of nonlinear statistical predictors have been presented to overcome this limitation (Clements et al., 2004). However, Clements et al. (2004) argue that nonlinear statistical predictors have similar performance than linear statistical predictors, having high mathematical complexity and the need of a human specialist to build the model, and this fact restricts the development of automatic predictors.

Artificial neural networks (ANNs) can be viewed as an alternative promising approach, regarding statistical models, for time series forecasting (Aznarte, Alcala-Fdez, Arauzo-Azofra, & Benítez, 2012; de A. Araújo, 2007; de Oliveira, Zarate, & Nobre, 2011; Cheng & Wei, 2014; Feng & Chou, 2011; Khashei & Bijari, 2014; Kourentzes, Barrow, & Crone, 2014; Kristjanpoller, Fadic, & Minutolo, 2014; Ma, Dai, & Liu, 2015; Menezes & Barreto, 2013; Nassirtoussi, Aghabozorgi, Wah, & Ngo, 2014; Park, Kim, & Lee, 2014). However, it has a main limitation, which is the definition of the parameters set (topology, architecture, amount of hidden units within hidden layer and training algorithm are some of these parameters) (de A. Araújo, 2007). Furthermore, the particular case of the time series forecasting problem, ANNs have the need of time lags definition to represent temporal phenomena (de A. Araújo, 2007). A particular class of ANNs presented in the literature to solve several kinds of problems are the morphological neural networks (MNNs) (Ritter & Sussner, 1996), which are based on fundamental concepts of mathematical morphology (MM) (Maragos, 1989; Serra, 1982) and lattice theory (Banon & Barrera, 1993; Heijmans, 1994; Ronse, 1990; Serra, 1988), and employ elementary morphological operators in their processing units. In the literature, it is possible to find some promising MNNs to deal with forecasting problems (de A. Araújo, 2012b; Madeiro, Sousa, Ferreira, & Pessoa, 2006, 2007, 2011; Sussner & Valle, 2007; Sussner & Esmi, 2011a, 2011b).

* Corresponding author at: Informatics Department, Federal Institute of Sertão Pernambucano, Ouricuri, PE, Brazil, and Informatics Center, Federal University of Pernambuco, Recife, PE, Brazil.
E-mail addresses: ricardo.araujo@ifsertao-pe.edu.br, raa@cin.ufpe.br (R.d.A. Araújo), alio@cin.ufpe.br (A.L.I. Oliveira), srlm@cin.ufpe.br (S. Meira).

http://dx.doi.org/10.1016/j.eswa.2015.01.004
0957-4174/

In this context, hybrid forecasting models, composed of evolutionary algorithms (EAs) (Hansen, 2006; Leung, Lam, Ling, & Tam, 2003; Vandenbergh & Engelbrecht, 2004) and ANNs or MNNs have been proposed in the literature (Asadi, Hadavandi, Mehmanpazir, & Nakhostin, 2012; de A. Araújo, 2007, 2011b, 2012a, 2013; de A. Araújo & Ferreira, 2013; Du, Leung, & Kwong, 2014; Huang, Wang, Fang, Liu, & Dou, 2013; Khashei & Bijari, 2014; Shahrabi, Hadavandi, & Asadi, 2013; Wong & Versace, 2012). The hybrid approach employs EAs to choose the time lags, parameters, structure and architecture of forecasting models, together with a descending gradient-based method to train them. However, there are some restrictions regarding the practical use of such approach, since it uses a wide amount of computing resources to achieve accurate forecasting results (de A. Araújo, 2011a, 2012b). Also, de A. Araújo (2011a, 2012b) argues that hybrid forecasting models not always outperform traditional forecasting models, since the increase in costs does not imply in the increase of forecasting performance.

However, even with sophisticated forecasting models presented in the literature, a limitation arises in the particular case of financial time series forecasting problem, called the random walk dilemma (RWD). In this dilemma, it is possible to verify that the predictions generated have a delay of one step ahead regarding real values of the time series (Malkiel, 2003; Sitte & Sitte, 2002). Due to this behavior, Sitte and Sitte (2002) argue that financial time series are generated by random processes and cannot be predicted. However, Sitte and Sitte (2002) do not develop any conclusive mathematical proofs to support such a claim (this is currently supported by the fact that there is not any model able to predict financial time series without the delay of one step ahead regarding real values of the series) (de A. Araújo, 2011a, 2012b).

In this context, a research initially conducted by Ferreira, Vasconcelos, and Adeodato (2008) raised the assumption of a subdominant nonlinear component embedded within a dominant linear component in financial time series. Based on such assumption, Ferreira et al. (2008) presented a procedure, called the phase adjustment procedure (PAP), for ANN models (multilayer perceptron like) built using a hybrid learning process composed by GA (Leung et al., 2003) and gradient-based techniques (Battiti, 1992; Hagan & Menhaj, 1994; Moller, 1993; Riedmiller & Braun, 1993), which solves of a practical way the RWD. It is worth mentioning that the PAP is conditioned to the application of a behavioral test after the learning process of the ANN, from which a manual verification is needed to determine the use the PFP (Ferreira et al., 2008). From the introduction of the concept of phase adjustment in financial time series, it is possible to find some studies investigating its use in ANN models (de A. Araújo, 2007; Salgado, Lima, Ferreira, & Cavalcanti, 2010).

In an attempt to formalize the assumption presented by Ferreira et al. (2008), Araújo (de A. Araújo, 2010a, 2010b, 2010c, 2010d, 2011a, 2011b, 2012b; de A. Araújo, Oliveira, Soares, & Meira, 2012; de A. Araújo & Ferreira, 2013) conducted several detailed analyzes of the lagplot graph of financial time series, where nonlinear structures in high-order time lags could be identified, and then confirming the hypothesis of nonrandomness of this particular kind of time series. Besides, de A. Araújo (2010d) and de A. Araújo and Susner (2010) have demonstrated that the nonlinear component in the generator phenomenon financial time series can be approximated in terms of a nonlinear mapping with increasing behavior.

In this way, several works have been proposed in the literature investigating the use of increasing models to solve the financial time series forecasting problem, among which stand out: (i) A quantum-evolutionary approach to design translation invariant increasing morphological operators (de A. Araújo, 2010b), (ii) A hybrid intelligent method to design translation invariant increasing morphological operators (de A. Araújo, 2010c), (iii) A hybrid approach to design increasing modular-morphological operators (de A. Araújo, 2010d), (iv) An evolutionary approach to design morphological-rank-linear operators (de A. Araújo, 2010a; de A. Araújo & Ferreira, 2013), (v) An increasing morphological-linear model based on Matheron decomposition (de A. Araújo & Susner, 2010), (vi) A hybrid method to design increasing morphological operators (de A. Araújo, 2011b), (vii) An increasing morphological perceptron based on fundamental morphological operators (de A. Araújo, 2011a), (viii) A quantum-evolutionary approach to design increasing morphological perceptrons (de A. Araújo et al., 2012), and (ix) An increasing morphological perceptron based on Matheron decomposition (de A. Araújo, 2012b). It is worth mentioning that the demonstration presented by de A. Araújo (2010d) and de A. Araújo and Susner (2010) has a limitation, since it only considers financial time series with daily frequency. This justifies why increasing models have better performance to estimate the generator phenomena of this kind of time series, since a long-term analysis of stock market sampled in daily frequency revels a natural increasing behavior (de A. Araújo, 2012b).

However, with the evolution of trading platforms, the frequency for performing operations on the stock market had increased to fractions of seconds (Bagheri, Peyhani, & Akbari, 2014; Son, Noh, & Lee, 2012). Since year of 2009 the BM&F Bovespa (the Brazilian stock exchange) has worked in high-frequency, and according to the last annual report of the securities and exchange commission (*comissão de valores mobiliários*, CVM), the number of high-frequency operations has grown from 2.5% in 2009 to 36.5% in 2013. In the American stock market this percentage is around 80%. These percentages suggest that the high-frequency stock market (HFSM) is a global trend, which makes needed the analysis of this kind of time series because as a typical financial time series it is expected to occur the RWD. Besides, it is also expected that high-frequency generator phenomena have distinct behavior with respect to daily frequency, which emphasizes the limitation of the work developed by de A. Araújo (2010d) and de A. Araújo and Susner (2010), since the demonstration of increasing behavior present in daily frequency financial time series cannot be generalized for any kind of financial time series. In this way, efforts must be made to the development of forecasting models able to overcome the RWD for high-frequency financial time series.

In this way, this work presents a hybrid forecasting model to overcome the random walk dilemma for high-frequency financial time series forecasting. Also, it is presented a descending gradient-based method with automatic time phase adjustment to design the proposed model using ideas from the back propagation (BP) algorithm. An experimental analysis is conduced using relevant high-frequency financial time series from the Brazilian stock market, where five relevant metrics and an evaluation function are used to assess the forecasting performance of the proposed model in high-frequency stock market.

This paper is organized as follows. Section 2 presents the fundamentals of the time series forecasting problem and the random walk dilemma. Section 3 describes the high-frequency financial time series investigated in this work. Section 4 presents the proposed model and its learning process. In Section 5 it is presented the experimental analysis with the proposed model, as well as a comparison between the obtained results with those given by classical prediction models. The final remarks of this work are presented in Section 6.

2. The time series forecasting problem

According to Box et al. (1994), a time series is a sequence of observations, sampled in a discrete or continuous space, about a

phenomenon that evolves with time. Therefore, it can be formally defined by Box et al. (1994)

$$\mathbf{x} = \{x_t \in \mathbb{R} |\ t = 1, 2, 3 \ldots N\}, \quad (1)$$

where t is the temporal index and N is the number of observations. The term $\mathbf{x}$ is the set of all temporal observations equally spaced and temporally ordered by a chronological index t, which is called time and defines the granularity of observations of a temporal phenomenon.

The aim to predict a time series $\mathbf{x}$ is to build a mapping able to estimate the future behavior of a temporal phenomenon, that is, to build a mechanism that allows, with certain accuracy, to estimate the future values of the time series, given by x_{t+h}, $h = 1, 2, \ldots, H$, where h represents the prediction horizon of H steps ahead. Takens (1980) presented an alternative definition of the time series prediction problem: Let M a phase space, let $\varphi(x)$ a trajectory defined by temporal phenomenon over the phase space M, and let $f(\varphi(x))$ an observation of each point regarding the temporal evolution of this phenomenon, that is, the time series. In this way, it is possible to verify that the time series prediction problem is related to the correct definition of a mapping $\Phi_{(\varphi,f)} : M \rightarrow \mathbb{R}^n$ among observations of the temporal phenomenon and the phase space M, formally given by

$$\Phi_{(\varphi,f)} = [f(x), f(\varphi(x)), \ldots, f(\varphi^n(x))], \quad (2)$$

in which n represents the dimension of phase space M.

According to Takens (1980), the relationship among observations of a temporal phenomenon constitutes a n-dimensional phase space, where n is the minimum dimension able to represent such relationship. Therefore, a n-dimensional phase space can be rebuilt to represent the needed information to characterize a time series. Takens (1980) proved that if n is sufficiently large, the rebuilt phase space is topologically equivalent to real phase space that generates the time series, that is, the phase space dynamics is topologically identical to phase space dynamics that generates the time series.

The identification of relevant dependencies among observations (time lags) can be performed using the autocorrelation function (ACF) (Box et al., 1994) and the partial autocorrelation function (PACF) (Box et al., 1994) only when there are linear relationships among temporal phenomenon observations. However, the lagplot analysis of financial time series reveals nonlinear relationships, which tends to make the definition of time lags a complex process (de A. Araújo, 2012b). Therefore, the main problem to rebuild the phase space is the correct choice of the variable n, or more specifically, the correct choice of the relevant time lags needed for the system dynamics characterization, where several techniques have been presented in the literature to determine the relevant time lags (Pi & Peterson, 1994; Savit & Green, 1991; Tanaka, Okamoto, & Naito, 2001).

2.1. The random walk dilemma

Sitte and Sitte (2002) and Malkiel (2003) shown that a future observation (x_t) of a financial time series can be defined by

$$x_t = x_{t-1} + r_t, \quad (3)$$

where term x_t is the current observation, the term x_{t-1} is the observation before x_t, and the term r_t is a noise with Gaussian distribution of zero mean and standard deviation σ ($r_t \approx N(0, \sigma)$), that is, a white noise.

According to Eq. (3), as future values of a financial time series are given by past values (and term r_t is a white noise), the future is unpredictable, since x_{t-1} is indeed to be the prediction of x_t. This behavior occurs in financial time series problems and is known as the random walk dilemma (RWD) or the random walk hypothesis (RWH) (Sitte & Sitte, 2002; Malkiel, 2003). In this dilemma, the prediction of any model for financial time series would has a one step ahead delay regarding the time series values. This affirmation is according to results reported in the literature (Sitte & Sitte, 2002; Malkiel, 2003) and its proof is developed as follows.

Assuming that an accurate forecasting model is used to build an estimated value of x_t, denoted by $\widehat{x_t}$. Then the expected value ($E[\cdot]$) of the difference between $\widehat{x_t}$ and x_t must tend to zero, as follows

$$E[\widehat{x_t} - x_t] \rightarrow 0. \quad (4)$$

According to de A. Araújo (2011a, 2012b), the generator phenomenon of financial time series is composed by a balanced combination between a linear dominant component and a nonlinear subdominant component (denoted by $g(t)$), and $E[r_t] = 0$ and $E[r_i r_j] = 0$ ($\forall\ i \neq j$). Then, the expected value of the difference between $\widehat{x_t}$ and x_t (assuming that $x_t = x_{t-1} + g(t) + r_t$) will be

$$E[\widehat{x_t}] \rightarrow E[x_{t-1}] + E[g(t)]. \quad (5)$$

As the nonlinear component has small magnitude regarding to linear component (de A. Araújo, 2011a, 2012b), it can be assumed that

$$E[g(t)] \rightarrow 0 \quad (6)$$

and hence

$$E[x_{t-1}] + E[g(t)] \rightarrow E[x_{t-1}]. \quad (7)$$

Therefore, Eq. (5) can be rewritten by

$$E[\widehat{x_t}] \rightarrow E[x_{t-1}]. \quad (8)$$

In this way, it can be verified that overcoming the random walk dilemma is a hard task, since generator phenomenon of financial time series is very similar to random walk generator phenomenon, and any prediction model tends to by guided by this behavior (Malkiel, 2003; Sitte & Sitte, 2002). However, according to de A. Araújo (2011a, 2012b), an alternative procedure can be employed within training algorithm of prediction models to overcome the random walk dilemma in the particular case of financial time series with daily frequency, called the phase fix procedure (PFP).

According to the Fig. 1 the PFP has two steps: *(i)* input pattern $\{x_1, x_2, \ldots, x_n\}$ (time lags) is presented to the prediction model to create an output (y_1), and *(ii)* the input pattern is re-arranged including the output generated in the step *(i)*, given by $\{y_1, x_1, x_2, \ldots, x_{n-1}\}$, and is presented to the same prediction model, where another output (y_2) is created. It is worth mentioning that y_2 represents the phase fixed prediction, that is, a prediction without the delay of one step ahead regarding time series values. The concept of phase adjustment assumes that the financial time series are generated from a balanced combination of a linear dominant component and a nonlinear subdominant component (de A. Araújo, 2011a, 2012b). Note that if the time series is generated by a

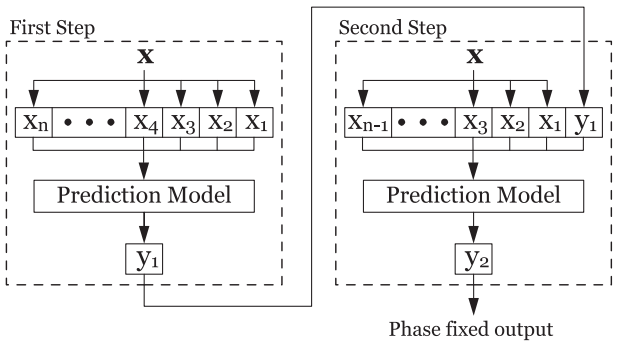

Fig. 1. Phase fix procedure.

random walk (Eq. (3)), the phase adjustment should not work (de A. Araújo, 2011a, 2012b).

2.2. Considerations

Recent researches (de A. Araújo, 2011a, 2012b) which employ the concept of phase adjustment within prediction models have considered only financial time series with daily frequency, and a long-term analysis of stock market sampled in daily frequency revels a natural increasing behavior. However, there is no reason to generalize the increasing behavior for financial time series from the high-frequency stock market, since it is possible to find both increasing and decreasing behavior in this kind of time series. In this way, this work assumes that high-frequency financial time series are generated by $x_t = x_{t-1} + g(t) + r_t$, composed by linear dominant component and nonlinear subdominant component with increasing and decreasing behavior. Therefore, this claim supports the fact that high-frequency financial time series are not generated by a random walk and it can be efficiently predicted using a forecasting model that uses the concept of phase adjustment and is able to represent a mapping composed by a balanced combination of linear and nonlinear (with increasing and decreasing behavior) operators.

3. High-frequency financial time series

This section describes three high-frequency financial time series investigated in this work: Banco do Brasil SA (BBAS3), Brasil Foods SA (BRFS3) and BR Malls Participacoes SA (BRML3). These time series represent the open value of the stock price, with frequency of one second, from the most important Brazilian stock market index, called *Bolsa de Valores de São Paulo* (BOVESPA). All time series investigated in this work is composed by high-frequency observations of the stock price at 2013/02/01, and they are illustrated in the Fig. 2. Some statistics of these time series are presented in the Table 1.

Based on Takens (1980) it is possible to conclude that the main problem in the reconstruction of the phase space of a temporal phenomenon is the choice of the relevant time lags (dimensionality n) to correctly represent the time series (system dynamics). The lagplot graphic is then used to determine and to analyze the relationships among time lags to define the dimensionality d. The Fig. 3 presents the lagplot graphic of all time series.

Through the analysis of Fig. 3, it can be verified the existence of a linear dominant relationship among time lags 1 and 25. However, in high order time lags (from dimensionality 50 to 1000) it is possible to identify a nonlinear subdominant relationship embedded within linear relationship present in time lags with low order. Such fact confirms the hypotheses of high-frequency financial time series are not generated by random processes, but by a balanced combination of linear dominant relationship and a nonlinear subdominant relationship. Furthermore, through the analysis of Fig. 2, it can be identified that this kind of time series has increasing and decreasing behavior, that is, the nonlinear subdominant component found in high-frequency financial time series can be estimated using increasing and decreasing operators, since a small, mid and long-term analysis of high-frequency stock market reveal a temporal phenomenon with increasing and decreasing components.

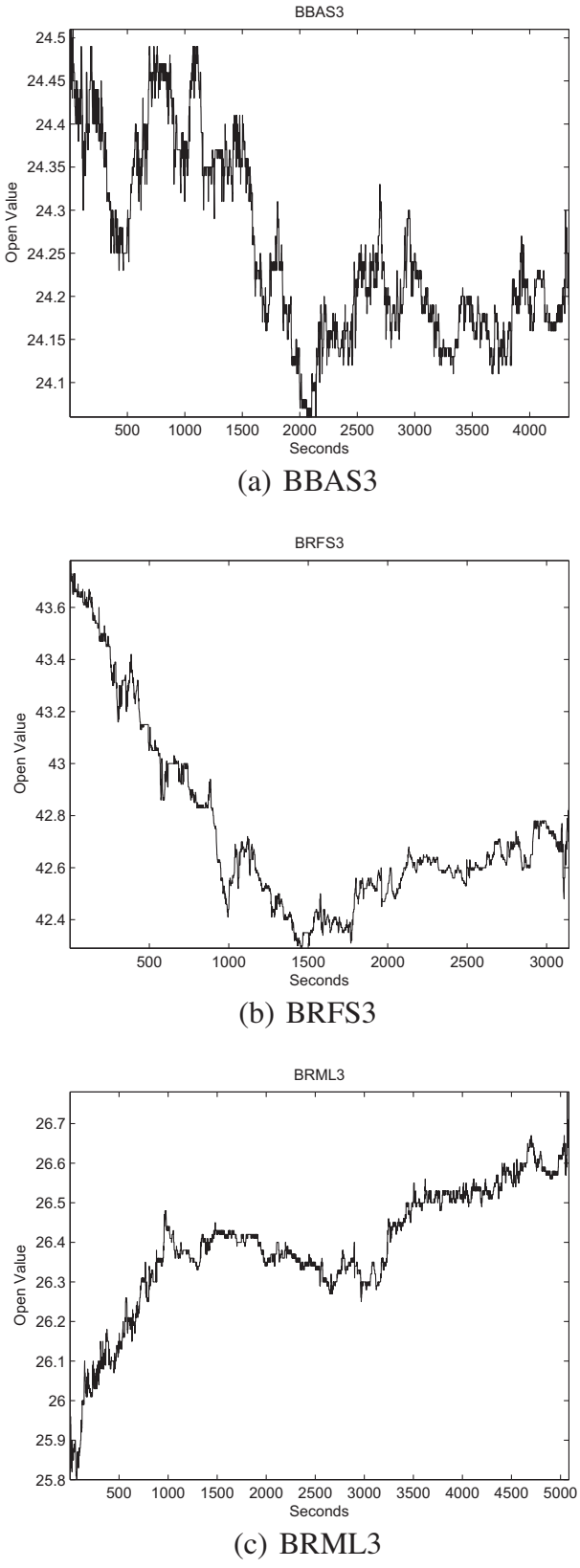

Fig. 2. Graphic of the analyzed time series.

4. The proposed model

According to lagplot analysis presented in the Section 3, it is possible to identify a linear dominant relationship within low order time lags and a nonlinear subdominant relationship (with increasing and decreasing behavior) within high order time lags, which leads to consider that high-frequency time series can be described by $x_t = x_{t-1} + g(t) + r_t$, where x_{t-1} represents the linear component, $g(t)$ represents the nonlinear component with

Table 1
Statistics of the analyzed time series.

Statistic	BBAS3	BRFS3	BRML3
Amount of observations	4341.00	3140.00	5085.00
Maximum value	24.51	43.78	26.78
Minimum value	24.06	42.29	25.80
Mean	24.2536	42.7463	26.3862
Variance	0.0119	0.1186	0.0260
Standard deviation	0.1089	0.3443	0.1613

increasing and decreasing behavior, and r_t represents the noise term generated by Gaussian distribution with zero mean and standard deviation σ $(\mathcal{N}(0,\sigma))$. Besides, it can be verified that the nonlinear component has lower magnitude than linear component magnitude, since it is dominant regarding nonlinear component. In this way, it is expected that a prediction model for this kind of time series has the ability to estimate, of a balanced way, both linear and nonlinear components, since the use of linear component is greater than the use of nonlinear component.

Therefore, this section describes a hybrid prediction model with all these characteristics for high-frequency financial time series. The proposed model, called the increasing decreasing linear neuron (IDLN), consists of a combination among linear operators (finite impulse response like), nonlinear increasing operators (dilation and erosion morphological operators) and nonlinear decreasing operators (anti-dilation and anti-erosion morphological operators), where their background and relevant theories are presented in the Appendix A.

4.1. Fundamentals

From the analysis presented in the Section 3, it can be assumed that the subdominant nonlinear component $g(t)$ present in high-frequency financial time series can be estimated in terms of a balanced combination of increasing and decreasing operators. In this context, Eqs. (A.10) and (A.11) proposed by Banon and Barrera (1993) (Appendix A) for increasing operators reveal that there are A^i and B^j for an index set I and J so that

$$\Psi = \bigwedge_{i\in I} \delta_{A^i} \tag{9}$$

and

$$\Psi = \bigvee_{j\in J} \varepsilon_{B^j}. \tag{10}$$

Note that both Eqs. (9) and (10) suggest that an increasing operator $\Psi : \mathbb{R}^n \to \mathbb{R}$, can be estimated in terms of vectors $\mathbf{a}^i, \mathbf{b}^j \in \mathbb{R}^n$ for a finite index set $\bar{I}$ and $\bar{J}$:

$$\Psi \simeq \bigwedge_{i\in I} \delta_{\mathbf{a}^i} \tag{11}$$

and

$$\Psi \simeq \bigvee_{j\in J} \varepsilon_{\mathbf{b}^j}. \tag{12}$$

When the index set $\bar{I} = 1$ and $\bar{J} = 1$, an increasing operator $\Psi : \mathbb{R}^n \to \mathbb{R}$ can be estimated in terms of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$:

$$\Psi \simeq \delta_{\mathbf{a}} \tag{13}$$

and

$$\Psi \simeq \varepsilon_{\mathbf{b}}. \tag{14}$$

In the same way, Eqs. (A.12) and (A.13) proposed by Banon and Barrera (1993) (Appendix A) for decreasing operators reveal that there are A^i and B^j for an index set I and J so that

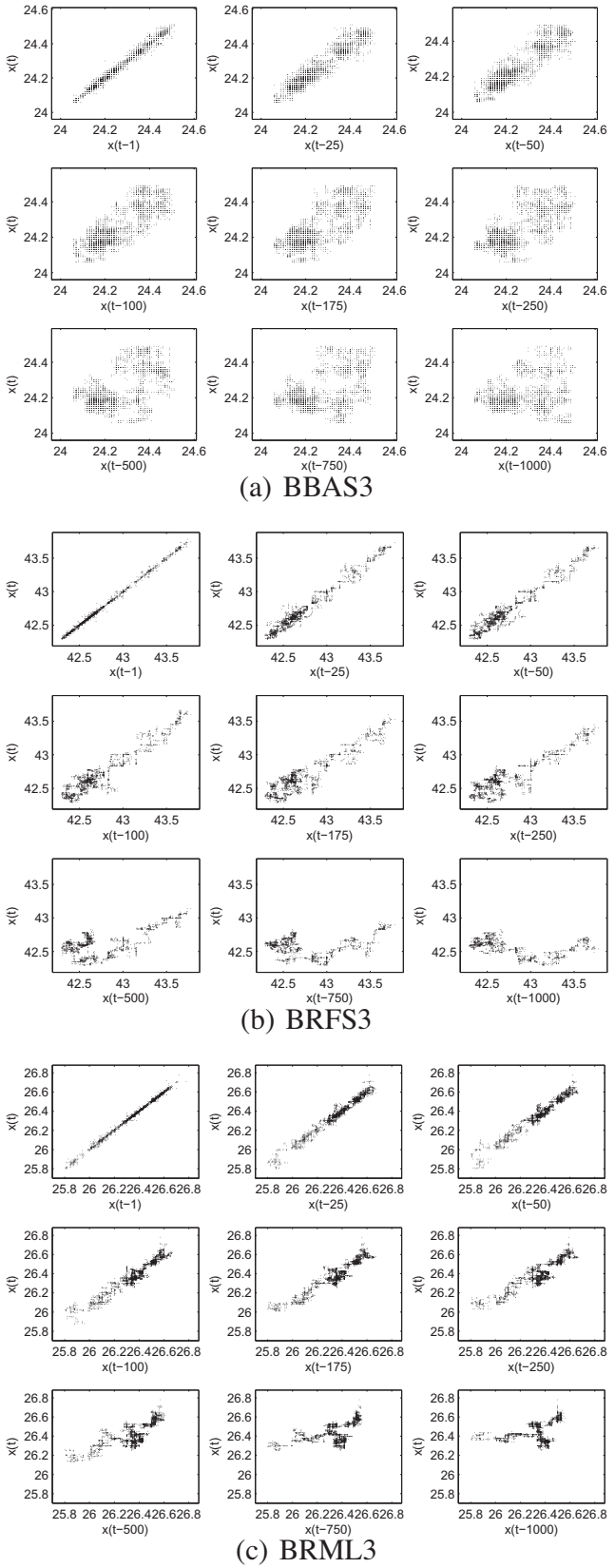

Fig. 3. Lagplot graphic of the analyzed time series.

$$\Psi = \bigvee_{i\in I} \overline{\delta}_{A^i}, \tag{15}$$

and

$$\Psi = \bigwedge_{j\in J} \overline{\varepsilon}_{B^j}. \tag{16}$$

Note that both Eqs. (15) and (16) suggest that a decreasing operator $\Psi : \mathbb{R}^n \to \mathbb{R}$, can be estimated in terms of vectors $\mathbf{a}^i, \mathbf{b}^j \in \mathbb{R}^n$ for a finite index set I and J:

$$\Psi \simeq \bigvee_{i\in I} \overline{\delta}_{\mathbf{a}^i} \tag{17}$$

and

$$\Psi \simeq \bigwedge_{j\in J} \overline{\varepsilon}_{\mathbf{b}^j}. \tag{18}$$

When the index set $I = 1$ and $J = 1$, a decreasing operator $\Psi : \mathbb{R}^n \to \mathbb{R}$ can be estimated in terms of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$:

$$\Psi \simeq \overline{\delta}_{\mathbf{a}} \tag{19}$$

and

$$\Psi \simeq \overline{\varepsilon}_{\mathbf{b}}. \tag{20}$$

Therefore, Eqs. (13), (14), (19) and (20) provides the basis to estimate the nonlinear subdominant component using increasing (dilation and erosion) and decreasing (anti-dilation and anti-erosion) morphological operators.

4.2. Definition

The proposed model, called the increasing decreasing linear neuron (IDLN), consists of a combination among increasing (dilation and erosion) and decreasing (anti-dilation and anti-erosion) morphological nonlinear operators and a linear operator (finite impulse response like). Next it is presented the mathematical definition of the proposed model.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ be an input pattern, represented by the time lags, inside an ith moving window and let y be the output, represented by the forecasting, of the IDLN, which is defined by a hybrid morphological-linear operator whose local signal transformation rule $\mathbf{x} \to y$ is given by

$$y = \lambda\alpha + (1-\lambda)\beta, \quad \lambda \in [0,1], \tag{21}$$

where

$$\beta = \sum_{i=1}^{n} x_i p_i \tag{22}$$

and

$$\alpha = \theta\tau + (1-\theta)\kappa, \quad \theta \in [0,1], \tag{23}$$

in which

$$\tau = \varphi\delta + (1-\varphi)\varepsilon, \quad \varphi \in [0,1] \tag{24}$$

and

$$\kappa = \omega\overline{\delta} + (1-\omega)\overline{\varepsilon}, \quad \omega \in [0,1], \tag{25}$$

with

$$\delta = \delta_{\mathbf{a}}(\mathbf{x}) = \bigvee_{i=1}^{n} (x_i + a_i), \tag{26}$$

$$\varepsilon = \varepsilon_{\mathbf{b}}(\mathbf{x}) = \bigwedge_{i=1}^{n} (x_i +' b_i), \tag{27}$$

$$\overline{\delta} = \overline{\delta}_{\mathbf{c}}(\mathbf{x}) = \bigwedge_{i=1}^{n} (x_i^* +' c_i), \tag{28}$$

$$\overline{\varepsilon} = \overline{\varepsilon}_{\mathbf{d}}(\mathbf{x}) = \bigvee_{i=1}^{n} (x_i^* + d_i), \tag{29}$$

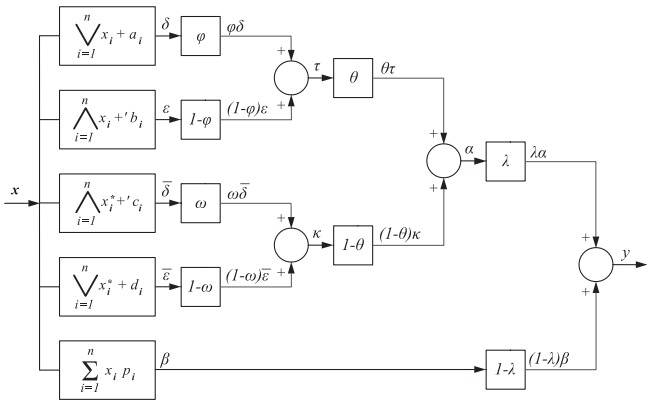

Fig. 4. The architecture of the proposed model.

where term n represents the input pattern dimensionality, terms $\lambda, \theta, \varphi, \omega \in \mathbb{R}$, and terms $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p} \in \mathbb{R}^n$. The term $\mathbf{p}$ represents the coefficients (weights) of the linear operator and term β represents the output of linear module. The term α represents the nonlinear module, given by a linear combination (the mixture term is given by θ) between increasing nonlinear module (defined by τ) and decreasing nonlinear module (defined by κ). Note that term τ represents a linear combination (the mixture term is given by φ) between dilation (δ) and erosion (ε) morphological operators. In the same way, the term κ represents a linear combination (the mixture term is given by ω) between anti-dilation ($\overline{\delta}$) and anti-erosion ($\overline{\varepsilon}$) morphological operators. The vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$ and $\mathbf{d}$ represent the structuring elements (weights) of dilation ($\delta_{\mathbf{a}}(\mathbf{x})$) and erosion ($\varepsilon_{\mathbf{b}}(\mathbf{x})$) (employed within increasing nonlinear module) and anti-dilation ($\overline{\delta}_{\mathbf{c}}(\mathbf{x})$) and anti-erosion ($\overline{\varepsilon}_{\mathbf{d}}(\mathbf{x})$) (employed within decreasing nonlinear module). At the end, the output y of the IDLN is given by a linear combination between linear and nonlinear modules (the mixture term is given by λ). Fig. 4 describes the IDLN architecture.

4.3. Fundamentals of the learning process

The rth rank function of a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ is given by the rth component of the vector $\mathbf{x}$ ordered by decreasing way ($x_{(1)} \geqslant x_{(2)} \geqslant \cdots \geqslant x_{(n)}$), defined by Pessoa and Maragos (2000)

$$\mathscr{R}_r(\mathbf{x}) = x_{(r)}, \quad r = 1, 2, \ldots, n. \tag{30}$$

The unit impulse function ($q(x)$) is defined by Pessoa and Maragos (2000)

$$q(x) = \begin{cases} 1, & \text{if } x = 0, \\ 0, & \text{otherwise}, \end{cases} \tag{31}$$

in which $x \in \mathbb{R}$.

The unit impulse function of a vector ($Q(\mathbf{x})$) is given by Pessoa and Maragos (2000)

$$Q(\mathbf{x}) = [q(x_1), q(x_2), \ldots, q(x_n)]. \tag{32}$$

The rth rank indicator vector ($\mathbf{c}$) of a vector $\mathbf{x}$ is defined by Pessoa and Maragos (2000)

$$\mathbf{c}(\mathbf{x}, r) = \frac{Q(z \cdot \mathbf{1} - \mathbf{x})}{Q(z \cdot \mathbf{1} - \mathbf{x}) \cdot \mathbf{1}^T}, \tag{33}$$

where $z = \mathscr{R}_r(\mathbf{x})$ and $\mathbf{1} = (1, 1, \ldots, 1)$. It is worth mentioning that the rank indicator vector marks the locations of the vector $\mathbf{x}$ where the value of z occurs.

Let $\mathbf{x} \in \mathbb{R}^{\mathbb{n}}, \mathbb{r} \in \{\mathbb{1}, \mathbb{2}, \ldots, \mathbb{n}\}$ and $\mathbf{c} = \mathbf{c}(\mathbf{x}, r)$. Then, it has that (Pessoa & Maragos, 2000):

(1) $\mathbf{c} \cdot \mathbf{1}^T = 1$ (unit area)
(2) $\mathbf{c} \cdot \mathbf{x}^T = \mathscr{R}_r(\mathbf{x})$ (representation via inner product)
(3) $\mathbf{c} \cdot (z\mathbf{1} - \mathbf{x})^T = 0$ where $z = \mathscr{R}_r(\mathbf{x})$
(4) If r is fixed then $\mathbf{c}$ is a piecewise constant function of the vector $\mathbf{x}$ with $2^n - 1$ possible values. Furthermore, for all points $\mathbf{x}_0 \in \mathbb{R}^n$ with unequal components, $x_{0,i} \neq x_{0,j} \forall i \neq j$, it is possible to verify a neighborhood around them, so that

$$\|\mathbf{x} - \mathbf{x}_0\|_\infty < \frac{1}{2}\min_{i \neq j}|x_{0,i} - x_{0,j}|, \tag{34}$$

where term $\mathbf{c}$ is constant, so that $\mathbf{c}(\mathbf{x}, r) = \mathbf{c}(\mathbf{x}_0, r)$.

Let $\mathbf{c} = \mathbf{c}(\mathbf{x}, r)$. For a given fixed value of r, if $\mathbf{c}$ is constant in a neighborhood of some $\mathbf{x}_0$, then the $r$the rank function $\mathscr{R}_r(\mathbf{x})$ is differentiable in $\mathbf{x}_0$ (Pessoa & Maragos, 2000):

$$\left.\frac{\partial \mathscr{R}_r(\mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0} = \left.\frac{\partial}{\partial \mathbf{x}}(\mathbf{c}(\mathbf{x}_0, r) \cdot \mathbf{x}^T)\right|_{\mathbf{x}=\mathbf{x}_0} = \mathbf{c}(\mathbf{x}_0, r). \tag{35}$$

Otherwise, the rank function $\mathscr{R}_r(\mathbf{x})$ is not differentiable (Pessoa & Maragos, 2000). In this case, it is possible to make it differentiable assigning to the vector $\mathbf{c}$ an unilateral value of the discontinuity $\frac{\partial z}{\partial \mathbf{x}}$. Therefore, the rank indicator vector can be used to estimate the gradients based on the discontinuity of the unit impulse function. The proof of this affirmation can be found in Pessoa and Maragos (2000).

As the proposed model in this work uses morphological operators of dilation, erosion, anti-dilation and anti-erosion, which are not differentiable of an usual way, the main issue is how to estimate their gradients at the points of non-differentiability. An extension of the rank indicator vector can be used to overcome the problem for dilation, erosion, anti-dilation and anti-erosion operators. It can be verified that these operators can be seen as particular cases of the rank function ($\mathscr{R}_r(\mathbf{x} + \mathbf{a}) = \bigvee \mathbf{x} + \mathbf{a}$ when $r = 1$ (dilation), $\mathscr{R}_r(\mathbf{x} +' \mathbf{b}) = \bigwedge \mathbf{x} +' \mathbf{b}$ when $r = n$ (erosion), $\mathscr{R}_r(\mathbf{x}^* +' \mathbf{c}) = \bigwedge \mathbf{x}^* +' \mathbf{c}$ when $r = n$ (anti-dilation), and $\mathscr{R}_r(\mathbf{x}^* + \mathbf{d}) = \bigvee \mathbf{x}^* + \mathbf{d}$ when $r = 1$ (anti-erosion) with $\mathbf{x}, \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^n$). However, the extension of rank indicator vector to estimate the gradients of dilation, erosion, anti-dilation and anti-erosion operators can lead to abrupt changes and compromising the numerical robustness of the gradient estimation (Pessoa & Maragos, 2000).

In this way, the unit impulse function ($q(x)$) can be replaced by the smoothed impulse function ($q_\sigma(x)$), which depends of a scale factor $\sigma \geqslant 0$, to overcome this problem. This approach was originally proposed by Pessoa and Maragos (2000) for rank operations and extended by Sousa, Carvalho, Assis, and Pessoa (2000) for maximum and minimum operations. Note that smoothed impulses boils down to approximating the rank function (and consequently the morphological operators used within the proposed model) in terms of differentiable functions to compute their gradients. According to Pessoa and Maragos (2000), the function $exp[\frac{1}{2}(\frac{x}{\sigma})^2]$ is a good choice for $q_\sigma(x)$. The smoothed impulse must has at least the following properties (Pessoa & Maragos, 2000):

$$\begin{aligned} &q_\sigma(x) = q_\sigma(-x)(\text{symmetry}),\\ &q_\sigma(x) \to q(x)\ \forall x \text{ with } \sigma \to 0,\\ &q_\sigma(x) \to 1\ \forall x \text{ with } \sigma \to \infty. \end{aligned} \tag{36}$$

In the same way of unit impulse function of a vector $Q(\mathbf{x})$, the smoothed unit impulse function of a vector is given by Pessoa and Maragos (2000)

$$Q_\sigma(\mathbf{x}) = [q_\sigma(x_1), q_\sigma(x_2), \ldots, q_\sigma(x_n)]. \tag{37}$$

The rth smoothed rank function is defined by Pessoa and Maragos (2000)

$$\mathscr{R}_{r,\sigma}(\mathbf{x}) = \mathbf{c}_\sigma \cdot \mathbf{x}^T, \tag{38}$$

where

$$\mathbf{c}_\sigma(\mathbf{x}, r) = \frac{Q_\sigma(z \cdot \mathbf{1} - \mathbf{x})}{Q_\sigma(z \cdot \mathbf{1} - \mathbf{x}) \cdot \mathbf{1}^T}, \quad z = \mathscr{R}_r(\mathbf{x}), \tag{39}$$

in which $\mathbf{c}_\sigma$ represents the smoothed rank indicator vector.

Note that the smoothed rank function $\mathbf{c}_\sigma$ represents an approximation of rank function $\mathbf{c}$. Let $\mathbf{x} \in \mathbb{R}^n, r \in \{1, 2, \ldots, n\}, \mathbf{c}_\sigma = \mathbf{c}_\sigma(\mathbf{x}, r), z = \mathscr{R}_r(\mathbf{x})$ and $z_\sigma = \mathscr{R}_{r,\sigma}(\mathbf{x})$. Then, it has that

(1) $\mathbf{c}_\sigma \cdot \mathbf{1}^T = 1$ (unit area);
(2) $\lim_{\sigma \to 0} \mathbf{c}_\sigma = \mathbf{c}$;
(3) $\lim_{\sigma \to 0} z_\sigma = z$;
(4) $\lim_{\sigma \to \infty} \mathbf{c}_\sigma = \frac{1}{n}$;
(5) $\lim_{\sigma \to \infty} z_\sigma = \frac{1}{n}\sum_{j=1}^n t_j$.

4.4. Learning process with automatic time phase adjustment

After the definition of IDLN model, it can be verified that it requires the adjustment of the following parameters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$, $\mathbf{p} \in \mathbb{R}^n$ and $\lambda, \theta, \varphi, \omega \in \mathbb{R}$. Therefore, the weight vector $\mathbf{w}$ (note that $\mathbf{w} \in \mathbb{R}^{5n+4}$) of the IDLN model is defined by

$$\mathbf{w} = (\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p}, \lambda, \theta, \varphi, \omega). \tag{40}$$

At the learning process, the IDLN weights are iteratively adjusted according to an error criterium until achieve convergence. In this way, it is necessary to define a cost function in terms of the weights $J(\mathbf{w})$:

$$J(\mathbf{w}) = \sum_{m=1}^{M} e^2(m), \tag{41}$$

where M represents the amount of training patterns, and $e(m)$ represents the instantaneous error for the mth training pattern, and defined by

$$e(m) = t(m) - y(m), \tag{42}$$

in which $t(m)$ and $y(m)$ are the desired output and the actual output, respectively.

It is worth mentioning that cost function generates an error surface which lies in $\mathbb{R}^{5n+4}$. The main problem in minimizing the cost function $J(\mathbf{w})$ is to find an optimal point in this space that minimizes the error between the model output and the desired output, that is, determining $\mathbf{w}$ when $\arg min[J(\mathbf{w})]$. In this work it is proposed a supervised learning approach, employing a descending gradient-based method with automatic time phase adjustment to design the proposed model using ideas from the back propagation (BP) algorithm (Haykin, 1998, 2007). Therefore, the weight vector $\mathbf{w}$ for the mth training pattern is updated according to the iterative formula:

$$\mathbf{w}(i+1) = \mathbf{w}(i) - \mu \nabla J(\mathbf{w}), \tag{43}$$

in which $i \in \{1, 2, \ldots\}$ and term μ represents the step size or learning rate, which is responsible for regulating the tradeoff between stability and speed of convergence of the iterative procedure. The term $\nabla J(\mathbf{w})$ is given by the gradient of J with respect to $\mathbf{w}$ at the points where the gradient exists. In this case it has that

$$\nabla J(\mathbf{w}) = \frac{\partial J}{\partial \mathbf{w}} = -2e(m)\frac{\partial y}{\partial \mathbf{w}}, \tag{44}$$

in which

$$\frac{\partial y}{\partial \mathbf{w}} = \left(\frac{\partial y}{\partial \mathbf{a}}, \frac{\partial y}{\partial \mathbf{b}}, \frac{\partial y}{\partial \mathbf{c}}, \frac{\partial y}{\partial \mathbf{d}}, \frac{\partial y}{\partial \mathbf{p}}, \frac{\partial y}{\partial \lambda}, \frac{\partial y}{\partial \theta}, \frac{\partial y}{\partial \varphi}, \frac{\partial y}{\partial \omega}\right)^T. \tag{45}$$

The existence of the gradient of J with respect to $\mathbf{w}$ only hinges on the existence of the gradients $\frac{\partial y}{\partial \mathbf{a}}, \frac{\partial y}{\partial \mathbf{b}}, \frac{\partial y}{\partial \mathbf{c}}, \frac{\partial y}{\partial \mathbf{d}}, \frac{\partial y}{\partial \mathbf{p}}, \frac{\partial y}{\partial \lambda}, \frac{\partial y}{\partial \theta}, \frac{\partial y}{\partial \varphi}$ e $\frac{\partial y}{\partial \omega}$. The following equations present the formulas to calculate the gradients.

The term $\frac{\partial y}{\partial \lambda}$ is given by

$$\frac{\partial y}{\partial \lambda} = \alpha - \beta. \tag{46}$$

The term $\frac{\partial y}{\partial \theta}$ is given by

$$\frac{\partial y}{\partial \theta} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \theta}, \tag{47}$$

in which

$$\frac{\partial y}{\partial \alpha} = \lambda \tag{48}$$

and

$$\frac{\partial \alpha}{\partial \theta} = \tau - \kappa. \tag{49}$$

The term $\frac{\partial y}{\partial \varphi}$ is given by

$$\frac{\partial y}{\partial \varphi} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \tau}\frac{\partial \tau}{\partial \varphi} = \lambda\frac{\partial \alpha}{\partial \tau}\frac{\partial \tau}{\partial \varphi}, \tag{50}$$

where

$$\frac{\partial \alpha}{\partial \tau} = \theta \tag{51}$$

and

$$\frac{\partial \tau}{\partial \varphi} = \delta - \varepsilon. \tag{52}$$

The term $\frac{\partial y}{\partial \omega}$ is given by

$$\frac{\partial y}{\partial \omega} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \kappa}\frac{\partial \kappa}{\partial \omega} = \lambda\frac{\partial \alpha}{\partial \kappa}\frac{\partial \kappa}{\partial \omega}, \tag{53}$$

in which

$$\frac{\partial \alpha}{\partial \kappa} = 1 - \theta \tag{54}$$

and

$$\frac{\partial \tau}{\partial \varphi} = \overline{\delta} - \overline{\varepsilon}. \tag{55}$$

The term $\frac{\partial y}{\partial \mathbf{p}}$ is given by

$$\frac{\partial y}{\partial \mathbf{p}} = \frac{\partial y}{\partial \beta}\frac{\partial \beta}{\partial \mathbf{p}}, \tag{56}$$

where

$$\frac{\partial y}{\partial \beta} = 1 - \lambda \tag{57}$$

and

$$\frac{\partial \beta}{\partial \mathbf{p}} = \mathbf{x}, \tag{58}$$

in which $\mathbf{x}$ represents the mth training pattern.

It is worth mentioning that terms $\frac{\partial y}{\partial \mathbf{a}}, \frac{\partial y}{\partial \mathbf{b}}, \frac{\partial y}{\partial \mathbf{c}}$ e $\frac{\partial y}{\partial \mathbf{d}}$ are estimated using the concept of smoothed rank indicator vector defined in the Section 4.3, due to the nondifferentiability of dilation, erosion, anti-dilation and anti-erosion morphological operators employed within IDLN, where it is employed the smoothed impulse function $Q_\sigma(\mathbf{x}) = [q_\sigma(x_1), q_\sigma(x_2), \ldots, q_\sigma(x_d)]$, given by

$$q_\sigma(x_i) = exp\left[\frac{1}{2}\left(\frac{x_i}{\sigma}\right)^2\right], \quad \forall i = 1, \ldots, d. \tag{59}$$

Also, it can be verified that the choice of scale factor σ affects the estimation and interpolation of $\frac{\partial y}{\partial \mathbf{a}}, \frac{\partial y}{\partial \mathbf{b}}, \frac{\partial y}{\partial \mathbf{c}}$ e $\frac{\partial y}{\partial \mathbf{d}}$. However, the learning process of the IDLN even works with $\sigma \to 0$, since in this particular case the gradient will be estimated in terms of the usual rank indicator vector. The following equations present the formulas to calculate the gradients.

The term $\frac{\partial y}{\partial \mathbf{a}}$ is given by

$$\frac{\partial y}{\partial \mathbf{a}} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \tau}\frac{\partial \tau}{\partial \delta}\frac{\partial \delta}{\partial \mathbf{a}} = \lambda\theta\frac{\partial \tau}{\partial \delta}\frac{\partial \delta}{\partial \mathbf{a}}, \tag{60}$$

in which

$$\frac{\partial \tau}{\partial \delta} = \varphi \tag{61}$$

and

$$\frac{\partial \delta}{\partial \mathbf{a}} = \frac{Q_\sigma(\delta \cdot \mathbf{1} - (\mathbf{x} + \mathbf{a}))}{Q_\sigma(\delta \cdot \mathbf{1} - (\mathbf{x} + \mathbf{a})) \cdot \mathbf{1}^T}. \tag{62}$$

The term $\frac{\partial y}{\partial \mathbf{b}}$ is given by

$$\frac{\partial y}{\partial \mathbf{b}} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \tau}\frac{\partial \tau}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \mathbf{b}} = \lambda\theta\frac{\partial \tau}{\partial \varepsilon}\frac{\partial \varepsilon}{\partial \mathbf{b}}, \tag{63}$$

where

$$\frac{\partial \tau}{\partial \varepsilon} = 1 - \varphi \tag{64}$$

and

$$\frac{\partial \varepsilon}{\partial \mathbf{b}} = \frac{Q_\sigma(\varepsilon \cdot \mathbf{1} - (\mathbf{x} +' \mathbf{b}))}{Q_\sigma(\varepsilon \cdot \mathbf{1} - (\mathbf{x} +' \mathbf{b})) \cdot \mathbf{1}^T}. \tag{65}$$

The term $\frac{\partial y}{\partial \mathbf{c}}$ is given by

$$\frac{\partial y}{\partial \mathbf{c}} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \kappa}\frac{\partial \kappa}{\partial \overline{\delta}}\frac{\partial \overline{\delta}}{\partial \mathbf{c}} = \lambda(1-\theta)\frac{\partial \kappa}{\partial \overline{\delta}}\frac{\partial \overline{\delta}}{\partial \mathbf{c}}, \tag{66}$$

in which

$$\frac{\partial \kappa}{\partial \overline{\delta}} = \omega \tag{67}$$

and

$$\frac{\partial \overline{\delta}}{\partial \mathbf{c}} = \frac{Q_\sigma(\overline{\delta} \cdot \mathbf{1} - (\mathbf{x}^* +' \mathbf{c}))}{Q_\sigma(\overline{\delta} \cdot \mathbf{1} - (\mathbf{x}^* +' \mathbf{c})) \cdot \mathbf{1}^T}. \tag{68}$$

The term $\frac{\partial y}{\partial \mathbf{d}}$ is given by

$$\frac{\partial y}{\partial \mathbf{d}} = \frac{\partial y}{\partial \alpha}\frac{\partial \alpha}{\partial \kappa}\frac{\partial \kappa}{\partial \overline{\varepsilon}}\frac{\partial \overline{\varepsilon}}{\partial \mathbf{d}} = \lambda(1-\theta)\frac{\partial \kappa}{\partial \overline{\varepsilon}}\frac{\partial \overline{\varepsilon}}{\partial \mathbf{d}}, \tag{69}$$

where

$$\frac{\partial \kappa}{\partial \overline{\varepsilon}} = 1 - \omega \tag{70}$$

and

$$\frac{\partial \overline{\varepsilon}}{\partial \mathbf{d}} = \frac{Q_\sigma(\overline{\varepsilon} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{d}))}{Q_\sigma(\overline{\varepsilon} \cdot \mathbf{1} - (\mathbf{x}^* + \mathbf{d})) \cdot \mathbf{1}^T}. \tag{71}$$

In order to automatically adjust time phase distortions that occur in the problem of high-frequency financial time series, it is included the phase fix procedure, defined in the Section 2.1, in the proposed learning process of the IDLN. The steps of the proposed learning process is illustrated in the Fig. 5.

According to the Fig. 5 it is necessary to define the stop condition. For that, it is used three stop conditions defined in Prechelt (1994): (i) Maximum training epochs number, (ii) Process training (PT), and (iii) Generalization loss (Gl).

5. Experimental results

A set of three high-frequency financial time series (BBAS3, BRFS3 and BRML3) was used as a test bed for evaluation of the proposed model. All these times series were normalized to lie within

```
begin Learning Process
    initialize parameters (μ, σ, w and the stop condition);
    epoch = 0;
    while not stop condition do
        epoch = epoch + 1;
        for i = 1 to M do
            calculate y1 using the i-th training pattern;
            calculate the instantaneous error (e(i));
            calculate the gradient (∇J(w));
            use y1 to rebuild the i-th training pattern;
            calculate y2 using the i-th reconstructed training pattern;
            calculate the cost function (J(w)) using y2;
            update weight vector w;
        end
    end
end
```

Fig. 5. The proposed learning process.

the range $[0,1]$ and divided in three sets according to Prechelt (1994): training set (50% of the data points), validation set (25% of the data points) and test set (25% of the data points). The most accurate prediction models, according to Clements et al. (2004) and de A. Araújo (2007, 2012b), from four approaches presented in the literature were elected to establish a performance study: (i) ARIMA (Box et al., 1994) (statistical model), (ii) MLP (Haykin, 1998, 2007) (artificial neural network model), (iii) IMP (de A. Araújo, 2012b) (morphological neural network model), and (iv) SHIF (de A. Araújo, 2007) (hybrid model). It is worth mentioning that the phase fix procedure was applied to all forecasting models investigated in this work in order to ensure a fair comparison among them. The following metrics are used to evaluate prediction performance (de A. Araújo, 2012b): mean squared error (MSE), mean absolute percentage error (MAPE), theil statistics (THEIL), prediction of change in direction (POCID), average relative variance (ARV) and evaluation function (EF).

For the experiments with the ARIMA $(p;q;d)$ model, it is used the ARMASA toolbox to determine both AR and MA coefficients. Note that the term $d=1$ as suggested by Box et al. (1994). For the experiments with the MLP $(I;H;O)$ model, it is used the Neural Network toolbox to determine the MLP parameters. The term I is given by the amount of time lags using the empirical methodology described in de A. Araújo (2007), where values $1,3,5$ and 10 are investigated. The term H is given by the amount of hidden units, where values $1,5,10,25$ e 50 are investigated. Note that the term $O=1$ since this work have focus on predictions with unitary horizon. It is used sigmoidal units for all MLP layers, since this configuration achieves improved prediction performance (de A. Araújo, 2007), as well as the Levenberg Marquardt algorithm (Hagan & Menhaj, 1994) for the MLP learning process with the following stop criteria (Prechelt, 1994): (i) $epoch=10,000$, (ii) $Pt\leqslant 10^{-6}$, and (iii) $Gl\leqslant 5\%$. For the experiments with the IMP $(lags;k)$ model, it is used the time lags 2 to 151 $(lags=2-151)$ as suggested by de A. Araújo (2012b). The term k is given by amount of morphological decompositions, where values $1,5,10,25$ and 50 are investigated. It is used the decent gradient-based method (de A. Araújo, 2012b) for the IMP learning process with the following stop criteria (Prechelt, 1994): (i) $epoch=10,000$, (ii) $Pt\leqslant 10^{-6}$, and (iii) $Gl\leqslant 5\%$. For the experiments with SHIF ($lags$; MLP $(I;H;O)$; $trainalg$) model, it is fixed the following parameters (de A. Araújo, 2007): minimum acceptable fitness equals to 40, method's iterations equals to 10, maximum amount of time lags equals to 10, maximum amount of hidden units equals to 20. Also, it is fixed the following parameters of the particle swarm optimizer (PSO), which is used within SHIF model (de A. Araújo, 2007): PSO iterations equals to 10,000, acceleration coefficients equals to 2.05, inertia weight equals to 0.9, maximum velocity equals to 2.048 and swarm size equals to 10. Note that in the SHIF model each PSO particle is represented by a MLP model with maximum architecture defined by $I=10,H=20$ and $O=1$. Each particle is additionally trained by the algorithms resilient backpropagation (Riedmiller & Braun, 1993), Levenberg Marquardt (Hagan & Menhaj, 1994), scaled conjugate gradient (Moller, 1993) and one step secant conjugate gradient (Battiti, 1992) with the following stop criteria (Prechelt, 1994): (i) $epoch=10,000$, (ii) $Pt\leqslant 10^{-6}$, and (iii) $Gl\leqslant 5\%$.

For the experiments with the proposed model, it is necessary to define a basic architecture, defined as IDLN $(lags;\mu;\sigma)$, in which

Table 2
Prediction performance for BBAS3 time series (test set).

Model	Statistic	Metric					
		ARV	MAPE	MSE	POCID	THEIL	EF
ARIMA	MED	0.2714	0.0567	0.0002	82.64	2.9895	19.1391
	DP	0.0000	0.0000	0.0000	0.00	0.0000	0.0000
	MIN	0.2714	0.0567	0.0002	82.64	2.9895	19.1391
	MAX	0.2714	0.0567	0.0002	82.64	2.9895	19.1391
	IC	±0.0000	±0.0000	±0.0000	±0.00	±0.0000	±0.0000
MLP	MED	0.3999	0.0552	0.0004	85.04	4.4108	25.9414
	DP	0.55738	0.0399	0.0005	5.06	6.1475	12.5406
	MIN	0.0985	0.0267	0.0001	82.64	1.0860	4.2699
	MAX	1.7473	0.1436	0.0017	94.64	1.9273	37.3696
	IC	±0.4547	±0.0325	±0.0004	±4.13	±5.0155	±10.2315
IMP	MED	0.0950	0.0271	0.00009	82.64	1.0474	38.0888
	DP	0.00006	0.00003	0.00000006	0.00	0.0007	0.0148
	MIN	0.0949	0.0271	0.00009	82.64	1.0466	38.0628
	MAX	0.0951	0.0272	0.00009	82.64	1.0487	38.1055
	IC	±0.00005	±0.00003	±0.00000005	±0.00	±0.0006	±0.0121
SHIF	MED	0.0907	0.0243	0.00009	82.64	1.0003	39.0643
	DP	0.0001	0.0006	0.0000001	0.00	0.0021	0.0541
	MIN	0.0903	0.0226	0.00009	82.64	0.9957	39.0359
	MAX	0.0909	0.0245	0.00009	82.64	1.0015	39.1889
	IC	±0.0001	±0.0005	±0.0000001	±0.00	±0.0017	±0.0441
IDLN	MED	0.0000006	0.00007	0.0000000006	100.00	0.000007	99.9921
	DP	0.0000007	0.00004	0.0000000007	0.00	0.000008	0.0058
	MIN	0.000000002	0.000004	0.000000000002	100.00	0.00000002	99.9803
	MAX	0.000002	0.0001	0.000000002	100.00	0.00002	99.9995
	IC	±0.0000006	±0.00004	±0.0000000006	±0.00	±0.000006	±0.0047

lags represents the time lags used to represent the temporal phenomenon, μ represents the learning rate, and σ represents the scale factor of the smoothed impulse function. The choice of the time lags was based on lagplot analysis presented in the Section 3, where time lags 2 to 1001 ($lags = 2 - 1001$) are fixed for all experiments. Note that the first time lag is not used because a random walk structure is necessary to use the phase fix procedure, since the key step of this procedure is the two step prediction to adjust time phase distortions. It is investigated the values $0.001, 0.01$ and 0.1 for the term μ and the values $0.005, 0.05$ and 0.5 for the term σ. The initial values of the IDLN weight vector are $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}, \mathbf{p} \in [-1, 1]$ and $\lambda, \theta, \varphi, \omega \in [0, 1]$. Three stop conditions are used in the IDLN learning process: (i) $epoch = 10,000$, (ii) $Pt \leqslant 10^{-6}$, and (iii) $Gl \leqslant 5\%$. For each model investigated in this work, ten experiments were performed, where the mean (MED), the standard deviation (DP), minimum (MIN) and maximum (MAX) statistics were calculated, as well as all confidence intervals (IC) with the assumption of normal distribution with 99% of certainty degree.

5.1. Banco do Brasil SA (BBAS3)

The best configuration of the investigated models for the BBAS3 time series are: ARIMA (6;5;1), MLP (5;10;1), IMP (2–151;25), SHIF (1, 3, 4, 5, 6, 9; MLP(6;9;1); Levenberg Marquardt) and IDLN (2–1001;0.1;0.005). The Table 2 presents the prediction

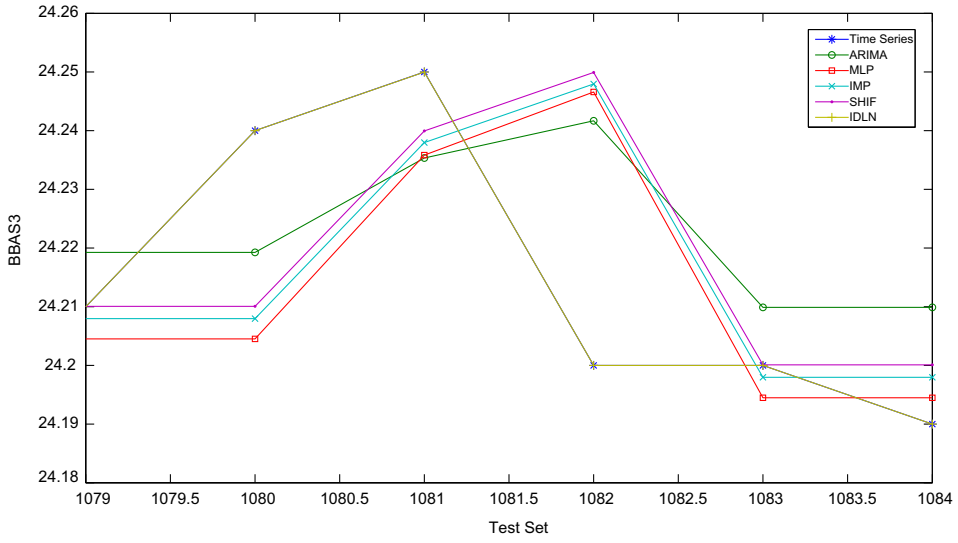

Fig. 6. Prediction results of BBAS3 time series (last five points of the test set).

Table 3
Prediction performance for BRFS3 time series (test set).

Model	Statistic	Metric					
		ARV	MAPE	MSE	POCID	THEIL	EF
ARIMA	MED	0.1689	0.0564	0.0007	87.74	3.8474	17.2936
	DP	0.0000	0.0000	0.0000	0.00	0.0000	0.0000
	MIN	0.1689	0.0564	0.0007	87.74	3.8474	17.2936
	MAX	0.1689	0.0564	0.0007	87.74	3.8474	17.2936
	IC	±0.0000	±0.0000	±0.0000	±0.00	±0.0000	±0.0000
MLP	MED	0.1312	0.0450	0.0006	87.74	2.9896	24.6767
	DP	0.0727	0.0183	0.0003	0.00	1.6557	10.0871
	MIN	0.0509	0.0222	0.0002	87.74	1.1624	11.7426
	MAX	0.2688	0.0774	0.0012	87.74	6.1243	39.2409
	IC	±0.0593	±0.0149	±0.0002	±0.00	±1.3508	±8.2297
IMP	MED	0.0438	0.0160	0.0002	87.74	1.0010	42.5686
	DP	0.000009	0.00006	0.00000004	0.00	0.0002	0.0059
	MIN	0.0438	0.0159	0.0002	87.74	1.0007	42.5605
	MAX	0.0439	0.0161	0.0002	87.74	1.0012	42.5765
	IC	±0.000007	±0.00005	±0.00000003	±0.00	±0.0001	±0.0048
SHIF	MED	0.0438	0.0161	0.0002	87.74	0.9997	42.5932
	DP	0.000008	0.00006	0.00000003	0.00	0.0001	0.0054
	MIN	0.0438	0.0160	0.0002	87.74	0.9993	42.5901
	MAX	0.0438	0.0161	0.0002	87.74	0.9998	42.6045
	IC	±0.000006	±0.00005	±0.00000003	±0.00	±0.0001	±0.0044
IDLN	MED	0.00001	0.0005	0.00000006	100.00	0.0003	99.9144
	DP	0.00001	0.0002	0.00000005	0.00	0.0002	0.0526
	MIN	0.0000002	0.00007	0.000000001	100.00	0.000005	99.8157
	MAX	0.00003	0.0009	0.0000001	100.00	0.0008	99.9918
	IC	±0.000009	±0.0002	±0.00000004	±0.00	±0.0002	±0.0429

performance of investigated models for the BBAS3 time series (test set). The Fig. 6 presents a comparative graphic between real and predicted values generated by investigated models in this work for the BBAS3 time series (the last five points of the test set).

5.2. Brasil Foods SA (BRFS3)

The best configuration of the investigated models for the BRFS3 time series are: ARIMA (3;2;1), MLP (5;10;1), IMP (2–151;10), SHIF (1, 7; MLP(2;5;1); Levenberg Marquardt) and IDLN (2–1001;0.1;0.005). The Table 3 presents the prediction performance of investigated models for the BRFS3 time series (test set). The Fig. 7 presents a comparative graphic between real and predicted values generated by investigated models in this work for the BRFS3 time series (the last five points of the test set).

5.3. BR Malls Participacoes SA (BRML3)

The best configuration of the investigated models for the BRML3 time series are: ARIMA (3;2;1), MLP (10;10;1), IMP (2–151;50), SHIF (1, 3, 4, 6, 8, 10; MLP(6;10;1); scaled conjugate gradient) and IDLN (2–1001;0.1;0.005). The Table 4 presents the prediction performance of investigated models for the BRML3 time series (test set). The Fig. 8 presents a comparative graphic between real and predicted values generated by investigated models in this work for the BRML3 time series (the last five points of the test set).

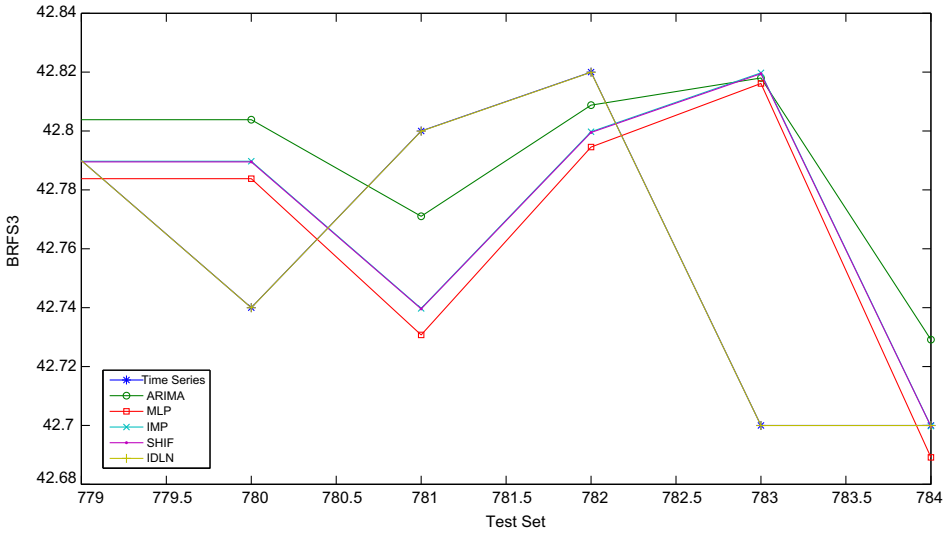

Fig. 7. Prediction results of BRFS3 time series (last five points of the test set).

Table 4
Prediction performance for BRML3 time series (test set).

Model	Statistic	Metric					
		ARV	MAPE	MSE	POCID	THEIL	EF
ARIMA	MED	0.0744	0.0343	0.0001	82.74	1.5121	31.5685
	DP	0.0000	0.0000	0.0000	0.00	0.0000	0.0000
	MIN	0.0744	0.0343	0.0001	82.74	1.5121	31.5685
	MAX	0.0744	0.0343	0.0001	82.74	1.5121	31.5685
	IC	±0.0000	±0.0000	±0.0000	±0.00	±0.0000	±0.0000
MLP	MED	0.0690	0.0325	0.0001	82.74	1.4025	33.0512
	DP	0.0022	0.0007	0.000004	0.00	0.0451	0.6224
	MIN	0.0653	0.0311	0.0001	82.74	1.3279	31.5329
	MAX	0.0745	0.0343	0.0001	82.74	1.5149	34.1268
	IC	±0.0018	±0.0006	±0.000003	±0.00	±0.0368	±0.5078
IMP	MED	0.0491	0.0201	0.00009	82.74	0.9998	39.9853
	DP	0.000003	0.00006	0.000000006	0.00	0.00006	0.0028
	MIN	0.0491	0.0201	0.00009	82.74	0.9998	39.9822
	MAX	0.0491	0.0202	0.00009	82.74	0.9999	39.9890
	IC	±0.000002	±0.00005	±0.000000005	±0.00	±0.00005	±0.0023
SHIF	MED	0.0494	0.0214	0.00009	82.74	1.0070	39.8191
	DP	0.0001	0.0005	0.0000002	0.00	0.0028	0.0678
	MIN	0.0490	0.0197	0.00009	82.74	0.9990	39.7957
	MAX	0.0495	0.0215	0.00009	82.74	1.0080	40.0117
	IC	±0.0001	±0.0004	±0.0000002	±0.00	±0.0022	±0.0553
IDLN	MED	0.0009	0.0047	0.000001	100.00	0.0183	97.6682
	DP	0.0004	0.0013	0.0000008	0.00	0.0084	0.9771
	MIN	0.0001	0.0020	0.0000003	100.00	0.0033	96.6666
	MAX	0.0013	0.0059	0.000002	100.00	0.0271	99.4397
	IC	±0.0003	±0.0010	±0.0000006	±0.00	±0.0068	±0.7972

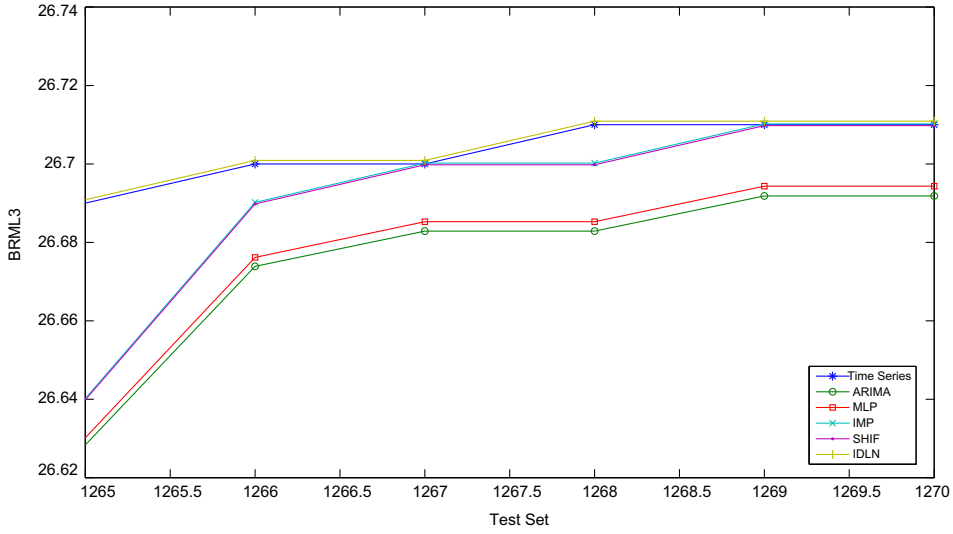

Fig. 8. Prediction results of BRML3 time series (last five points of the test set).

Table 5
Statistics of the THEIL metric for all investigated models (test set).

Metric	Statistic	Model				
		ARIMA	MLP	IMP	SHIF	IDLN
THEIL	MED	2.7460	2.9100	1.0280	1.0018	0.0039
	DP	0.8584	1.0841	0.0408	0.0046	0.0080
	MIN	1.5100	1.4000	1.0000	0.9990	0.000007
	MAX	3.8500	4.4100	1.0900	1.0100	0.0183
	IC	±0.9904	±1.2509	±0.0471	±0.0053	±0.0092

Table 6
Statistics of the parameters λ and θ used within IDLN model.

Statistic	Parameter	
	λ	θ
MED	0.0306	0.4966
DP	0.0271	0.0356
MIN	0.0092	0.4352
MAX	0.0764	0.5273
IC	±0.0313	±0.0411

5.4. Considerations

It can be verified that, for all high-frequency financial time series investigated in this work, the proposed model obtained better predictive performance (with respect to all considered metrics) when compared to the best prediction models reported in the literature (under the same experimental condition). Note that the proposed model overcame, of practical way, the random walk dilemma for high-frequency financial time series investigated in this work, since according to Table 5, the averaged value for THEIL metric is too closer to 0. In this way, the proposed model can be considered an accurate prediction model with high prediction performance. Besides, it is not possible to identify the presence of a one step ahead delay regarding the real time series values, that is, the time phase distortion that generates the random walk dilemma is properly adjusted. It is possible to verify that ARIMA, MLP, IMP and SHIF models are not able to overcome the random walk dilemma, since THEIL $\geqslant 1$.

Also, through the analysis of the parameters of the proposed model, presented in the Table 6, it is possible to confirm the hypothesis that high-frequency financial time series are generated by a dominant linear component and a subdominant nonlinear component (note that the averaged value of the term λ is 0.0306, which means that the IDLN used 3.06% of nonlinear component and 96.94% of linear component). Also, it is possible to confirm the hypothesis that the nonlinear component has increasing and decreasing behavior (note that the averaged value of term θ is 0.4966, which means that the IDLN used 49.66% of increasing module and 50.34% of decreasing module).

Therefore, there are three main evidences that justify the fact that the proposed model has found far superior prediction performance as the solution of the random walk dilemma for high-frequency financial time series: (i) the hypothesis, confirmed by lagplot analysis and by IDLN parameters λ and θ, that temporal phenomenon of this kind of time series is given by a combination of a dominant linear component and a subdominant nonlinear component with increasing and decreasing behavior, (ii) the ability to determine, through a balanced estimation, the percentage of use of linear and nonlinear components, as well as the percentage of use of increasing and decreasing modules, and (iii) the development of a gradient-based learning process with time phase adjustment.

6. Conclusion

This paper presented a hybrid forecasting model, called the increasing–decreasing-linear neuron (IDLN), for high-frequency financial time series forecasting. A gradient-based learning process with automatic time phase adjustment was presented to design the proposed model using ideas from the back propagation algorithm and using a systematic approach to overcome the problem of non-differentiability of morphological operators.

The results were collected with three high-frequency financial time series from the Brazilian stock market (BBAS3, BRFS3 and BRML3). The experimental results demonstrated a consistent much better performance of the proposed model when compared to the most common statistical, neural and hybrid prediction models. The reason for that is due to linear component and nonlinear component with increasing and decreasing behavior, as well as by the inclusion of the time phase adjustment within the learning process of the proposed model. In other words, the success of the proposed model was strongly dependent on an accurate adjustment of the predictive model parameters and on the model itself used for

forecasting. Therefore, the it can be considered a viable model to predict high-frequency financial time series.

Also, it is worth mentioning that all prediction models investigated in this work were not able to overcome the random walk dilemma for high-frequency financial time series (THEIL $\geqslant 1$). However, according to the obtained results with the proposed model, it can be verified that the proposed model was able to overcome the random walk dilemma in high-frequency financial time series (THEIL ~ 0). A more detailed explanation of such behavior is that the proposed model was built with similar characteristics of high-frequency financial temporal phenomena, which implies in the minimization of the search space for the model design. Furthermore, the inclusion of the phase fix procedure within learning process of the proposed model leads to a more precise time phase adjustment, since the relationship of high-frequency financial time series was directly mapped in the proposed model equations.

Note that the main advantages of the proposed model (apart from its superior predictive performance), are: (i) it has both linear and nonlinear components (a distinct percentage of the linear and nonlinear components can be used according to the characteristics of the problem), (ii) it has both increasing and decreasing nonlinear components (it means that the model can use a distinct percentage of the increasing and decreasing operators), and (iii) it is quite attractive due to their simpler computational complexity when compared to other models investigated in this work.

Besides, according to the analysis presented in the Section 3, it was possible to verify that the temporal phenomenon of this kind of time series can be represented by a balanced combination between a dominant linear component and a subdominant nonlinear component. Note that the lagplot analysis revealed a complex nonlinear subdominant relationship embedded in linear dominant relationship: (i) there is a linear relationship within low order time lags, and (ii) with the increase in the time lags degree, there is a complex nonlinear subdominant relationship. Also, considering the relationships of high-frequency financial generator phenomena, it could be supposed, through time series graphic, of the existence of increasing and decreasing behavior within nonlinear component in this kind of time series. Note that both hypotheses could be confirmed analyzing the mixture terms (λ and θ) of the proposed model, which have used around 97% of the linear component and around 3% of nonlinear component, as well as used around 50% of both increasing and decreasing nonlinear modules.

As future works, an additional study must be developed to formalize and to explain the properties of the proposed model, as well as to determine its limitations with other kind time series. Also, further studies, in terms of risk and financial return, must be developed in order to determine the additional economical benefits, for an investor, with the use of the proposed model. Besides, a particular study about the computing complexity and CPU time of the proposed model must be done in order to establish a complete cost performance evaluation of the proposed model. According to this investigation, it will be possible to relate, in terms of cost, the necessary time to generate an optimal forecasting model.

Appendix A. Fundamentals concepts and theories

Let A be a nonempty set, and let $\leqslant$ be a binary relation in A. Therefore, $\leqslant$ is a partial order if and only if the following properties are valid ($\forall\ a, b, c \in A$): (i) $a \leqslant a$ (reflexivity), (ii) $a \leqslant b$ e $b \leqslant a \Rightarrow a = b$ (anti-symmetry), and (iii) $a \leqslant b$ e $b \leqslant c \Rightarrow a = c$ (transitivity). Note that $(A, \leqslant)$ is denoted by a partially ordered set. If $a \leqslant b$ or $b \leqslant a\ \forall\ a, b \in A$, then A is denoted by a totally ordered set known as chain.

Let $(A, \leqslant)$ be a partially ordered set, and let B be a subset of A ($B \subseteq A$), then the element $a \in A$ is called upper bound of B if $b \leqslant a, \forall\ b \in B$. The lowest upper bound of B, when it exists, is called the supremum of B and denoted by $\bigvee B$. In other words, $a \in A$ is the supremum of B if a is an upper bound of B and let c be another upper bound of B, then $a \leqslant c$. In this way, the supremum of B is given by $\bigvee_{i\in I} b^i$ instead of $\bigvee B$ when $B = \{b^i, i \in I\}$ for an index set I. In the same way, the element $a \in A$ is called lower bound of B if $a \leqslant b, \forall\ b \in B$. The greatest lower bound of B, when it exists, is called the infimum of B and denoted by $\bigwedge B$. In other words, $a \in A$ is the infimum of B if a is an lower bound of B and let c be another lower bound of B, then $c \leqslant a$. In this way, the infimum of B is given by $\bigwedge_{i\in I} b^i$ instead of $\bigwedge B$ when $B = \{b^i, i \in I\}$ for an index set I.

A partially ordered set A is a lattice if and only if every finite subset of A have a supremum and an infimum in A, that is, $\forall B \subseteq A$ it has $\bigvee B \in A$ and $\bigwedge B \in A$. According to Birkhoff (1993), every chain is a lattice. Note that if $A_1, \ldots, A_n$ are lattices, then a partial order over $A^n = A_1 \times \cdots \times A_n$ can be defined by (Birkhoff, 1993)

$$(a_1, \ldots, a_n) \leqslant (b_1, \ldots, b_n) \Longleftrightarrow a_i \leqslant b_i, \quad i = 1, \ldots, n, \tag{A.1}$$

in which $a_i, b_i \in A_i$. According to Sussner and Esmi (2011a, 2011b), a partially ordered set A^n is a lattice, and denoted by product of lattices with components $A_1, \ldots, A_n$, which correspond to the product of n copies of A.

A lattice A is bounded if and only if it has a smallest element denoted by 0_A and a greatest element denoted by 1_A (Sussner & Esmi, 2011a, 2011b). Note that the product of lattices $A^n = A_1 \times \cdots \times A_n$ is bounded if and only if their lattices components $A_i, i = 1, \ldots, n$, are bounded. A lattice A is conditionally complete if and only if each nonempty bounded subset has a supremum and an infimum in A (Sussner & Esmi, 2011a, 2011b). The set of real numbers $\mathbb{R}$ is an example of a conditionally complete lattice (Sussner & Esmi, 2011a, 2011b). A lattice A is complete if and only if all finite or infinite subset has a supremum and an infimum in A (Birkhoff, 1993), that is, $\forall\ B \subseteq A$ finite of infinite it has that $\bigvee B \in A$ and $\bigwedge B \in A$. According to Birkhoff (1993), all complete lattice is bounded. Sussner and Esmi (2011a, 2011b) shown that if A is a complete lattice then the product of complete lattices is also complete, since for each set of indexes I and $a^i \in A$ it has that

$$\bigwedge_{i\in I} a^i = \bigwedge_{i\in I} (a_1^i, \ldots a_n^i) = \left(\bigwedge_{i\in I} a_1^i, \ldots, \bigwedge_{i\in I} a_n^i \right) \tag{A.2}$$

and

$$\bigvee_{i\in I} a^i = \bigvee_{i\in I} (a_1^i, \ldots a_n^i) = \left(\bigvee_{i\in I} a_1^i, \ldots, \bigvee_{i\in I} a_n^i \right). \tag{A.3}$$

According to Sussner and Esmi (2011a, 2011b), the set of extended real numbers $\mathbb{R}_{\pm\infty}$ and the unit interval $[0, 1]$ are examples of bounded complete lattices, which are chains.

Let A and B lattices. An operator $\Psi : A \to B$ is increasing if and only if the following condition is valid ($\forall a, b \in A$) (de A. Araújo, 2010d):

$$a \leqslant b \Rightarrow \Psi(a) \leqslant \Psi(b). \tag{A.4}$$

Let A and B lattices. An operator $\Psi : A \to B$ is decreasing if and only if the following condition is valid ($\forall a, b \in A$) (de A. Araújo, 2010d):

$$a \leqslant b \Rightarrow \Psi(b) \leqslant \Psi(a). \tag{A.5}$$

Let A and B complete lattices, let $\delta, \varepsilon, \overline{\delta}$ and $\overline{\varepsilon}$ operators from A to B, and let $X \subseteq A$. Then, the operators $\delta, \varepsilon, \overline{\delta}$ and $\overline{\varepsilon}$ are algebraic dilation, erosion, anti-dilation and anti-erosion, respectively, if and only if

$$\delta\left(\bigvee X\right) = \bigvee_{x\in X} \delta(x), \tag{A.6}$$

$$\varepsilon\left(\bigwedge X\right) = \bigwedge_{x\in X}\varepsilon(x), \tag{A.7}$$

$$\overline{\delta}\left(\bigvee X\right) = \bigwedge_{x\in X}\overline{\delta}(x), \tag{A.8}$$

$$\overline{\varepsilon}\left(\bigwedge X\right) = \bigvee_{x\in X}\overline{\varepsilon}(x). \tag{A.9}$$

It can be verified that δ (dilation) and ε (erosion) are increasing operators and $\overline{\delta}$ (anti-dilation) and $\overline{\varepsilon}$ (anti-erosion) are decreasing operators (Banon & Barrera, 1993).

According to Sussner and Esmi (2011a, 2011b) composition among elementary morphological operators also represent elementary operators. Let $\delta, \varepsilon, \overline{\delta}$ and $\overline{\varepsilon}$ elementary morphological operators of dilation, erosion, anti-dilation and anti-erosion, respectively, from a complete lattice A to a complete lattice B, e let an operator $f : A \to C$, in which C is a complete lattice. Then: (i) if f is a dilation, then $f \circ \delta$ is a dilation, (ii) if f is an erosion, then $f \circ \varepsilon$ is an erosion, (iii) if f is an erosion, then $f \circ \overline{\delta}$ is an anti-dilation, and (iv) if f is a dilation, then $f \circ \overline{\varepsilon}$ is an anti-erosion.

Banon and Barrera (1993) demonstrated that the main issue in this context is the decomposition of mappings among complete lattices in term of elementary morphological operators, where several theorems were developed. Among these theorems, this work addresses the decomposition of increasing and decreasing operators, since the analysis of high-frequency time series presented in Section 3 revels increasing and decreasing behavior of this kind of time series.

Let $\Psi : A \to B$ be an increasing operator among complete lattices A and B. Then, there are dilations δ^i and erosions ε^i for an index set I so that

$$\Psi = \bigwedge_{i\in I}\delta^i, \tag{A.10}$$

$$\Psi = \bigvee_{i\in I}\varepsilon^i. \tag{A.11}$$

Let $\Psi : A \to B$ be a decreasing operator among complete lattices A and B. Then, there are anti-dilations $\overline{\delta}^i$ and anti-erosions $\overline{\varepsilon}^i$ for an index set I so that

$$\Psi = \bigvee_{i\in I}\overline{\delta}^i. \tag{A.12}$$

$$\Psi = \bigwedge_{i\in I}\overline{\varepsilon}^i. \tag{A.13}$$

However, elementary morphological operators of dilation, erosion, anti-dilation and anti-erosion require an additional algebraic structure, called the minimax algebra (Cuninghame-Green, 1995). Therefore, the morphological operators must be redefined using additional operators of infimum and supremum (Sussner & Esmi, 2011a, 2011b). This work focused on an extension of such approach for matrixes and vectors proposed by Cuninghame-Green (1995). Note that, as nonlinear relationship of high-frequency time series can be modeled in terms of increasing and decreasing operators of kind $\mathbb{R}^n_{\pm\infty} \to \mathbb{R}_{\pm\infty}$ (n represents the time lags dimensionality), the main focus of this work is given to the chain $\mathbb{R}_{\pm\infty}$.

Let A be a nonempty set, and let $\leqslant$ partial order relation. Then A is a group if and only if A is a lattice and $\forall a, b, x, y \in A$, it has that $a \leqslant b \Rightarrow xay \leqslant xby$. Note that the chain $\mathbb{R}_{\pm\infty}$ is a group, and $\mathbb{R}$, the set of finite elements of $\mathbb{R}_{\pm\infty}$, is a group with conventional addition operation (Sussner & Esmi, 2011a, 2011b). According to Cuninghame-Green (1979, 1995), the minimax algebra considers a bounded lattice ordered group (BLOG), which is given by a bounded lattice A whose set of finite elements builds a group defined by $A \setminus \{+\infty, -\infty\}$, where $+\infty$ and $-\infty$ denotes $\bigvee A$ and $\bigwedge A$, respectively. Note that when A is a complete lattice whose elements form a group, then A is defined by a complete lattice ordered group (Sussner & Esmi, 2011a, 2011b). Also, when a conditionally complete lattice can form a group, it can be defined by a conditionally complete lattice ordered group, where it is necessary to define the operation "+" over $(A \times A) \setminus (B \times B)$, given by Sussner and Esmi (2011a, 2011b):

$$b + (+\infty) = +\infty + b = +\infty, \forall b \in B \cup +\infty \tag{A.14}$$

and

$$b + (-\infty) = -\infty + b = -\infty, \forall b \in B \cup -\infty. \tag{A.15}$$

However, the isotonicity of translations group using $+\infty$ cannot be maintained in A, even we choose to define $(-\infty) + (+\infty) = (+\infty) + (-\infty)$. There are two similar alternatives (Sussner & Esmi, 2011a, 2011b): (i) $(-\infty) + (+\infty) = (+\infty) + (-\infty) = +\infty$, or (ii) $(-\infty) + (+\infty) = (+\infty) + (-\infty) = -\infty$. In this way, Sussner and Esmi (2011a, 2011b) have shown that an additional group operation "$+'$" have to be defined over $(A \times A) \setminus \{(+\infty, -\infty), (-\infty, +\infty)\}$ and given by

$$(-\infty) + (+\infty) = (+\infty) + (-\infty) = -\infty \tag{A.16}$$

and

$$(-\infty) +'(+\infty) = (+\infty) +'(-\infty) = +\infty. \tag{A.17}$$

According to Sussner and Esmi (2011a, 2011b), a bounded lattice ordered group is defined by $(A, \vee, \wedge, +, +')$, as well as the complete lattice ordered group by $(\mathbb{R}_{\pm\infty}, \vee, \wedge, +, +')$ and $(\mathbb{Z}_{\pm\infty}, \vee, \wedge, +, +')$, which are chains, are called complete totally ordered group. Birkhoff (1993) shown that operations $\vee, \wedge, +$ e $+'$ defined over $\mathbb{R}$ can be extended to $\mathbb{R}^n$ with $n > 1$, including the greatest and lowest element in $\mathbb{R}^n$, given by "$+\infty$" and "$-\infty$", respectively, and defining "$+'$" and "+" over $(\mathbb{R}^n \times \mathbb{R}^n)$, given by $(\mathbb{R}^n \cup \{-\infty, +\infty\}, \vee, \wedge, +, +')$. A bounded lattice ordered group can be built using an algebraic structure called division belts, based on the concept of conjugation (Sussner & Esmi, 2011a, 2011b).

Let A and B be complete lattices, then an operator v is a negation in A if it reverts the partial order in A. Let operator $\Phi : A \to B$ and let v_A and v_B the negation in A and B, respectively, then the operator $\Phi^v : A \to B$ is the negation of Φ regarding v_A and v_B, and given by Sussner and Esmi (2011a, 2011b)

$$\Phi^v = v_B \circ \Phi \circ v_A. \tag{A.18}$$

Let A be a complete lattice ordered group, then the conjugate of an element $a \in A$, given by a^*, is defined by Sussner and Esmi (2011a, 2011b)

$$a^* = \begin{cases} -a, & \text{if} \quad a \in A \setminus \{-\infty, +\infty\}, \\ +\infty, & \text{if} \quad a = -\infty, \\ -\infty, & \text{if} \quad a = +\infty, \end{cases}, \tag{A.19}$$

in which $-a$ represents the inverse of a. According to Sussner and Esmi (2011a, 2011b), the conjugation operation leads to a negation v_* of A^n which maps the ith component x_i of $X \in A^n$ to its conjugate:

$$(v_*(X))_i = (x_i)^*, \quad \forall i = 1, 2, \ldots, n. \tag{A.20}$$

According to Cuninghame-Green (1979, 1995), fundamental morphological operators can be defined in terms of matrix products from minimax algebra. Let A be a complete lattice ordered group, and let the matrixes $X \in A^{m\times p}$ and $Y \in A^{p\times n}$, then the maximum product of X and Y is given by

$$B = X \vee Y. \tag{A.21}$$

An element of the matrix B can be defined by

$$b_{ij} = \bigvee_{k=1}^{p}(x_{ik} + y_{kj}). \tag{A.22}$$

In the same way, the minimum product of X and Y is given by

$$C = X \wedge Y. \tag{A.23}$$

An element of the matrix C can be defined by

$$c_{ij} = \bigwedge_{k=1}^{p} (x_{ik} +' y_{kj}). \tag{A.24}$$

Therefore, let A be a complete lattice ordered group, and let operators $\delta_X, \varepsilon_X, \overline{\delta}_X, \overline{\varepsilon}_X : A^n \to A^m$ for $X, Y \in A^{n\times m}$, then it has

$$\delta_X(Y) = Y^T \vee X, \tag{A.25}$$

$$\varepsilon_X(Y) = Y^T \wedge X, \tag{A.26}$$

$$\overline{\delta}_X(Y) = Y^{*^T} \wedge X, \tag{A.27}$$

$$\overline{\varepsilon}_X(Y) = Y^{*^T} \vee X, \tag{A.28}$$

where the term "^{T}" represents the transposition operation, and the term "*" represents the conjugation operation. Note that operators $\delta_X, \varepsilon_X, \overline{\delta}_X$ and $\overline{\varepsilon}_X$ represent dilation, erosion, anti-dilation and anti-erosion operators, respectively, from A^n to A^m.

References

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

Aznarte, J. L., Alcala-Fdez, J., Arauzo-Azofra, A., & Benítez, J. M. (2012). Financial time series forecasting with a bio-inspired fuzzy model. *Expert Systems with Applications, 39*(16), 12302–12309.

Bagheri, A., Peyhani, H. M., & Akbari, M. (2014). Financial forecasting using {ANFIS} networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications, 41*(14), 6235–6250.

Banon, G. J. F., & Barrera, J. (1993). Decomposition of mappings between complete lattices by mathematical morphology, part 1. general lattices. *Signal Processing, 30*(3), 299–327.

Battiti, R. (1992). One step secant conjugate gradient. *Neural Computation, 4*, 141–166.

Birkhoff, G. (1993). *Lattice theory*. Providence: American Mathematical Society.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: Forecasting and control* (3rd ed.). New Jersey: Prentice Hall.

Cheng, C., & Wei, L. (2014). A novel time-series model based on empirical mode decomposition for forecasting {TAIEX}. *Economic Modelling, 36*, 136–141.

Clements, M. P., Franses, P. H., & Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting, 20*, 169–183.

Cuninghame-Green, R. (1979). *Minimax algebra: Lecture notes in economics and mathematical systems* (Vol. 166). New York: Springer-Verlag.

Cuninghame-Green, R. (1995). Minimax algebra and applications. In P. Hawkes (Ed.). *Advances in imaging and electron physics* (Vol. 90, pp. 1–121). New York, NY: Academic Press.

de A. Araújo, R. (2007). Swarm-based hybrid intelligent forecasting method for financial time series prediction. *Learning and Nonlinear Models, 5*, 137–154.

de A. Araújo, R. (2010a). Adjusting time phase distortions in financial forecasting via morphological-rank-linear evolutionary approach. *Learning and Nonlinear Models, 8*, 157–162.

de A. Araújo, R. (2010b). Hybrid intelligent methodology to design translation invariant morphological operators for brazilian stock market prediction. *Neural Networks, 23*, 1238–1251.

de A. Araújo, R. (2010c). A hybrid intelligent morphological approach for stock market forecasting. *Neural Processing Letters, 31*, 195–217.

de A. Araújo, R. (2010d). Swarm-based translation-invariant morphological method for financial time series forecasting. *Information Sciences, 180*, 4784–4805.

de A. Araújo, R. (2011a). A class of hybrid morphological perceptrons with application in time series forecasting. *Knowledge-Based Systems, 24*, 513–529.

de A. Araújo, R. (2011b). Translation invariant morphological time-lag added evolutionary forecasting method for stock market prediction. *Expert Systems with Applications, 38*, 2835–2848.

de A. Araújo, R. (2012a). Hybrid morphological methodology for software development cost estimation. *Expert Systems with Applications, 39*, 6129–6139.

de A. Araújo, R. (2012b). A morphological perceptron with gradient-based learning for brazilian stock market forecasting. *Neural Networks, 28*, 61–81.

de A. Araújo, R. (2013). Evolutionary learning processes to design the dilation-erosion perceptron for weather forecasting. *Neural Processing Letters, 37*, 303–333.

de A. Araújo, & R., Susner, P. (2010). An increasing hybrid morphological-linear perceptron with pseudo-gradient-based learning and phase adjustment for financial time series prediction. In *IEEE international joint conference on neural networks*.

de A. Araújo, R., Madeiro, F., Sousa, R., Ferreira, T., & Pessoa, L. (2006). An evolutionary morphological approach for financial time series forecasting. In *IEEE congress on evolutionary computation*.

de A. Araújo, R., Sousa, R., & Ferreira, T. (2007). An intelligent hybrid approach for designing increasing translation invariant morphological operators for time series forecasting. In *International symposium on neural networks*.

de A. Araújo, R., & Ferreira, T. (2013). A morphological-rank-linear evolutionary method for stock market prediction. *Information Sciences, 237*, 3–17.

de A. Araújo, R., Oliveira, A., & Meira, S. (2011). A shift-invariant morphological system for software development cost estimation. *Expert Systems with Applications, 38*, 4162–4168.

de A. Araújo, R., Oliveira, A., Soares, S., & Meira, S. (2012). A quantum-inspired evolutionary learning process to design dilation-erosion perceptrons for financial forecasting. *Learning and Nonlinear Models, 10*, 192–201.

de Oliveira, F. A., Zarate, L. E., de A. Reis, M., Nobre, C. N. (2011). The use of artificial neural networks in the analysis and prediction of stock prices. In *IEEE international conference on systems, man, and cybernetics (SMC)*.

Du, W., Leung, S. Y. S., & Kwong, C. K. (2014). Time series forecasting by neural networks: A knee point-based multiobjective evolutionary algorithm approach. *Expert Systems with Applications, 41*(18), 8049–8061.

Feng, H. M., & Chou, H. C. (2011). Evolutional {RBFNs} prediction systems generation in the applications of financial time series data. *Expert Systems with Applications, 38*(7), 8285–8292.

Ferreira, T. A. E., Vasconcelos, G. C., Adeodato, & P. J. L. (2008). A new intelligent system methodology for time series forecasting with artificial neural networks. In *Neural processing letters* (Vol. 28, pp. 113–129).

Hagan, M., & Menhaj, M. (1994). Training feedforward networks with the marquardt algorithm. *IEEE Transactions on Neural Networks, 5*(6), 989–993.

Hansen, N. (2006). The CMA evolution strategy: A comparing review. In J. Lozano, P. Larranaga, I. Inza, & E. Bengoetxea (Eds.), *Towards a new evolutionary computation. Advances on estimation of distribution algorithms* (pp. 75–102). Springer.

Haykin, S. (1998). *Neural networks: A comprehensive foundation*. New Jersey: Prentice Hall.

Haykin, S. (2007). *Neural networks and learning machines*. Canada: McMaster University.

Heijmans, H. J. A. M. (1994). *Morphological image operators*. New York, NY: Academic Press.

Huang, D., Wang, X., Fang, J., Liu, S., & Dou, R. (2013). A hybrid model based on neural networks for financial time series. In *Mexican international conference on artificial intelligence*.

Khashei, M., & Bijari, M. (2014). Fuzzy artificial neural network (p, d, q) model for incomplete financial time series forecasting. *Journal of Intelligent and Fuzzy Systems, 26*(2), 831–845.

Kourentzes, N., Barrow, D. K., & Crone, S. F. (2014). Neural network ensemble operators for time series forecasting. *Expert Systems with Applications, 41*(9), 4235–4244.

Kristjanpoller, W., Fadic, A., & Minutolo, M. C. (2014). Volatility forecast using hybrid neural network models. *Expert Systems with Applications, 41*(5), 2437–2442.

Leung, F. H. F., Lam, H. K., Ling, S. H., & Tam, P. K. S. (2003). Tuning of the structure and parameters of the neural network using an improved genetic algorithm. *IEEE Transactions on Neural Networks, 14*(1), 79–88.

Ma, Z., Dai, Q., & Liu, N. (2015). Several novel evaluation measures for rank-based ensemble pruning with applications to time series prediction. *Expert Systems with Applications, 42*(1), 280–292.

Malkiel, B. G. (2003). In *A random walk down wall street. Completely revised and updated edition*. W.W. Norton & Company.

Maragos, P. (1989). A representation theory for morphological image and signal processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11*, 586–599.

Menezes, J. M. P., & Barreto, G. A. (2013). Redes baseadas em projeções aleatórias para predição não-linear de séries temporais caóticas. *Simpósio Brasileiro de Automação Inteligente*.

Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural Networks, 6*, 525–533.

Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications, 41*(16), 7653–7670.

Park, H., Kim, N., & Lee, J. (2014). Parametric models and non-parametric machine learning models for predicting option prices: Empirical comparison study over {KOSPI} 200 index options. *Expert Systems with Applications, 41*(11), 5227–5237.

Pessoa, L. F. C., & Maragos, P. (2000). Neural networks with hybrid morphological rank linear nodes: A unifying framework with applications to handwritten character recognition. *Pattern Recognition, 33*, 945–960.

Pi, H., & Peterson, C. (1994). Finding the embedding dimension and variable dependences in time series. *Neural Computation, 6*, 509–520.

Prechelt, L. (1994). Proben1: A set of neural network benchmark problems and benchmarking rules. Tech. Rep. 21/94.

Riedmiller, M., & Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The rprop algorithm. In *Proceedings of the IEEE international conference on neural networks (ICNN), San Francisco* (pp. 586–591).

Ritter, G. X., & Sussner, P. (1996). An introduction to morphological neural networks. In *Proceedings of the 13th international conference on pattern recognition, Vienna, Austria* (pp. 709–717).

Ronse, C. (1990). Why mathematical morphology needs complete lattices. *Signal Processing, 21*(2), 129–154.

Salgado, P., Lima, A., Ferreira, T., & Cavalcanti, G. (2010). An intelligent perturbative approach for the time series forecasting problem. In *International joint conference on neural networks*.

Savit, R., & Green, M. (1991). Time series and dependent variables. *Physica D, 50*, 95–116.

Serra, J. (1982). *Image analysis and mathematical morphology*. London: Academic Press.

Serra, J. (1988). Image analysis and mathematical morphology. *Theoretical advances* (Vol. 2). New York: Academic Press.

Shahrabi, J., Hadavandi, E., & Asadi, S. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. *Knowledge-Based Systems, 43*, 112–122.

Sitte, R., & Sitte, J. (2002). Neural networks approach to the random walk dilemma of financial time series. *Applied Intelligence, 16*(3), 163–171.

Son, Y., Noh, D., & Lee, J. (2012). Forecasting trends of high-frequency {KOSPI200} index data using learning classifiers. *Expert Systems with Applications, 39*(14), 11607–11615.

Sousa, R. P., Carvalho, J. M., Assis, F. M., & Pessoa, L. F. C. (2000). Designing translation invariant operations via neural network training. In *Proceedings of the IEEE international conference on image processing, Vancouver, Canada*.

Sussner, P., & Esmi, E. L. (2011a). Morphological perceptrons with competitive learning: Lattice-theoretical framework and constructive learning algorithm. *Information Sciences, 181*(10), 1929–1950.

Sussner, P., & Esmi, E. L. (2011b). Morphological perceptrons with competitive learning: Lattice-theoretical framework and constructive learning algorithm. *Information Sciences, 181*(10), 1929–1950.

Sussner, P., & Valle, M. E. (2007). Morphological and certain fuzzy morphological associative memories for classification and prediction. In V. G. Kaburlassos & G. X. Ritter (Eds.). *Computational intelligence based on lattice theory* (Vol. 67, pp. 149–173). Heidelberg, Germany: Springer Verlag.

Takens, F. (1980). Detecting strange attractor in turbulence. In *Dynamical systems and turbulence. Lecture notes in mathematics* (Vol. 898, pp. 366–381). New York: Springer-Verlag.

Tanaka, N., Okamoto, H., & Naito, M. (2001). Estimating the active dimension of the dynamics in a time series based on a information criterion. *Physica D, 158*, 19–31.

Vandenbergh, F., & Engelbrecht, A. P. (2004). A cooperative approach to particle swarm optimization. *IEEE Trans. Evolutionary Computation, 8*(3), 225–239.

Wong, C., & Versace, M. (2012). Cartmap: A neural network method for automated feature selection in financial time series forecasting. *Neural Computing and Applications, 21*(5), 969–977.