

An Evolutionary Approach to Design Dilation-Erosion Perceptrons for Stock Market Indices Forecasting

Ricardo de A. Araújo, Adriano L. I. Oliveira, Sergio Soares, Silvio Meira
Informatics Center, Federal University of Pernambuco, Recife, PE, Brazil.
{raa, alio, scbs, srlm}@cin.ufpe.br

ABSTRACT

In this work we present an evolutionary learning process using the covariance matrix adaptation evolutionary strategy (CMAES) to design the dilation-erosion perceptron (DEP) for stock market indices forecasting. Also, we have included an automatic phase fix procedure (APFP) into proposed learning process to eliminate time phase distortions observed in some forecasting problems. The main advantage of the DEP model designed by our learning process, apart from its higher forecasting performance, is do not request any methodology to overcome the nondifferentiability of morphological operators needed into classical gradient-based learning process of the DEP model. Besides, we present an experimental analysis using two stock market indices, where five well-known performance metrics and an evaluation function are used to assess forecasting performance.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods, and Search; G.3 [Probability and Statistics]: Time series analysis; I.6.5 [Simulation and Modeling]: Model Development

General Terms

Algorithms; Experimentation; Verification

Keywords

Evolutionary Learning Process, Covariance Matrix Adaptation Evolution Strategy, Dilation-Erosion Perceptrons, Stock Market Indices Forecasting, Random Walk Dilemma. Time Phase Adjustment.

1. INTRODUCTION

Morphological artificial neurons, referred to as morphological perceptrons (MPs) [1], represent a particular class of neural processing units based on the framework of mathematical morphology (MM) [2, 3] with algebraic foundations in the complete lattice theory (CLT) [4–6]. Several MPs have been proposed in the literature, including morphological perceptrons [1], morphological

perceptrons with dendrites [7], morphological-rank-linear perceptrons [8], constructive morphological perceptrons [9], increasing hybrid morphological-linear perceptron [10, 11], dilation-erosion perceptron [12], amongst other.

Such morphological approach for neurocomputing was successfully applied to a large number of problems [9]. Recently, the forecasting problem have been solved with the use of some morphological approaches [12–17]. However, a dilemma arises from all these models when we work with financial problems, known as random walk dilemma [18], where the forecasts have one step delay regarding the real data, that is, a time phase distortion in the reconstruction of phase space of financial phenomena [12–17].

According to Araújo [12], the dilation-erosion perceptron (DEP), a hybrid perceptron which uses elementary operators from MM [2, 3] on CLT [4–6], is currently considered the more accurate morphological model presented in the literature of financial forecasting. The classical DEP learning process, called DEP(GB), employs a gradient steepest descent method using the back propagation (BP) algorithm [12]. However, we can see that the main drawback of DEP(GB) is the need of a systematic methodology to overcome the problem of nondifferentiability of morphological operators in training equations [12].

In this sense, this work presents an alternative learning process, called DEP(CMAES), using the covariance matrix adaptation evolutionary strategy (CMAES) [19] to design the dilation-erosion perceptron (DEP) [12] for stock market indices forecasting. Also, we have included an automatic phase fix procedure (APFP) [12–17, 20] into proposed learning process in the attempt to adjust time phase distortions that occur in some forecasting problems. Furthermore, an experimental analysis is presented using two complex nonlinear problems of the stock market indices forecasting: Dow Jones Industrial Average Index and Standard & Poor 500 Index. Also, five well-known performance metrics and an evaluation function are used to assess forecasting performance. At the end, the obtained results by the DEP model using the proposed learning process are compared with those generated by DEP model using the classical gradient-based learning process, where we can see an improved forecasting performance of the proposed DEP(CMAES).

2. FUNDAMENTALS

In this section the fundamentals and theoretical concepts necessary for this work will be presented.

2.1 The Time Series Forecasting

A time series is a sequence of observations about a given phenomenon observed in a discrete or continuous space. In this work all time series will be considered time discrete and equidistant, and

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'11, July 12–16, 2011, Dublin, Ireland.

Copyright 2011 ACM 978-1-4503-0557-0/11/07 ...\$10.00.

formally defined by

$$\mathbf{x} = \{x_t \in \mathbb{R} \mid t = 1, 2, \dots, N\}, \quad (1)$$

where t is the temporal index, which is called time and defines the granularity of observations of a given phenomenon, and N is the number of observations.

The aim of forecasting techniques applied to a given time series is to provide a mechanism that allows, with certain accuracy, the forecasting of the future values of \mathbf{x} , given by x_{t+h} , $h = 1, 2, \dots, H$, where h represents the forecasting horizon of H steps ahead. These techniques try to identify certain regular patterns present in the data set, creating a model capable of generating the next temporal patterns, where, in this context, a most relevant factor for an accurate forecasting performance is the correct choice of the past window, or the time lags, considered for the representation of a given time series. In mathematical sense, the relationship which involves time series historical data defines a d -dimensional phase space, where d is the minimum dimension capable of representing such relationship. Therefore, a d -dimensional phase space can be built so that it is possible to unfold its corresponding time series. Takens [21] proved that if d is sufficiently large, such phase space is homeomorphic to the phase space that generates the series. The Takens' Theorem [21] is the theoretical justification that it is possible to rebuild a phase space using the correct time lags.

2.2 The Random Walk Dilemma

A naive forecasting strategy is to define the last observation of a time series as the best forecasting of its next future value ($x_{t+1} = x_t$). This kind of model is known as the random walk (RW) model [18], which is defined by

$$x_t = x_{t-1} + z_t, \quad (2)$$

where x_t is the current observation, x_{t-1} is the immediate observation before x_t , and z_t is a noise term with a Gaussian distribution of zero mean and standard deviation $sd(z_t \approx N(0, sd))$. The model above clearly implies that, as the information set consists of past time series data, the future data is unpredictable. Therefore, on average, the value x_{t-1} is indeed the best forecasting of value x_t , and proof of this statement is given in Araújo [12].

It is possible to verify that the use of an arbitrary model to make forecasts have an intrinsic limitation, since the generated forecasts have a characteristic one step ahead delay regarding the original time series values, in which this behavior is common in the finance and economics and is called random walk dilemma or random walk hypothesis [18]. Therefore, in these conditions, to escape of the random walk dilemma is a hard task [12].

2.3 Morphological Perceptrons Background

Morphological perceptrons (MPs) [1] are defined as a particular kind of artificial neurons which perform morphological operators on complete lattices (algebraic framework for mathematical morphology [4–6]). Next, we present some background information in lattice theory [22], minimax algebra [23], and mathematical morphology on complete lattices [4–6].

A *partially ordered set* is defined as a set that is a reflexive, anti-symmetric and transitive, where a binary relation " \leq " is defined. For simplicity purpose, it is common to assume a partially ordered set as a non-empty set [9]. Note that if $a \leq b$ or $b \leq a$ in a partially ordered set \mathbb{L} then \mathbb{L} is defined as *totally ordered* and known as *chain*. A partially ordered set \mathbb{L} is defined as *lattice* if every finite and non-empty subset of \mathbb{L} has an infimum and a supremum in \mathbb{L} [22]. Note that every chain is a lattice [9].

For any $L \subseteq \mathbb{L}$, the supremum and infimum of L , given respectively by $\bigvee L$ and $\bigwedge L$, are formally defined by [9]

$$\bigvee L = \bigvee_{j \in J} l^j \quad \text{and} \quad \bigwedge L = \bigwedge_{j \in J} l^j, \quad (3)$$

where $L = \{l^j, j \in J\}$ for a index set J .

An *increasing mapping* $\Psi : \mathbb{L} \rightarrow \mathbb{M}$, in which \mathbb{L} and \mathbb{M} are lattices, is defined for all $a, b \in \mathbb{L}$ and given by [9]

$$a \leq b \Rightarrow \Psi(a) \leq \Psi(b). \quad (4)$$

A *bounded lattice* \mathbb{L} is defined when the lattice have at least one element denoted by $0_{\mathbb{L}}$ and one greatest element denoted by $1_{\mathbb{L}}$. A *complete lattice* \mathbb{L} is defined when every non-empty subset has an infimum and a supremum in \mathbb{L} [22]. Note that every complete lattice is *bounded complete*. Particular examples of complete lattices are the extended real numbers $\mathbb{R}_{\pm\infty}$ and the unit interval $[0, 1]$.

If δ and ε are operators from a complete lattice \mathbb{L} to a complete lattice \mathbb{M} , and $L \subseteq \mathbb{L}$, then δ and ε can be algebraic *dilation* and *erosion* operators, respectively, if and only if [9]

$$\delta(\bigvee L) = \bigvee_{l \in L} \delta(l) \quad \text{and} \quad \varepsilon(\bigwedge L) = \bigwedge_{l \in L} \varepsilon(l). \quad (5)$$

However, elementary morphological operators require an additional algebraic complete lattice structure, which uses additional operations from infimum and supremum lattice operations [9]. In this way, it is possible to define two kinds of matrix products, the *max-product* ($C = A \vee B$) and the *min-product* ($D = A \wedge B$), in terms of matrices $A \in \mathbb{R}_{\pm\infty}^{m \times p}$ and $B \in \mathbb{R}_{\pm\infty}^{p \times n}$, which are respectively given by

$$c_{ij} = \bigvee_{k=1}^p (a_{ik} + b_{kj}) \quad \text{and} \quad d_{ij} = \bigwedge_{k=1}^p (a_{ik} + ' b_{kj}), \quad (6)$$

where the main differences between " $+'$ " and " $+$ " are given by the following rules:

$$(-\infty) + (+\infty) = (+\infty) + (-\infty) = -\infty, \quad (7)$$

and

$$(-\infty) + ' (+\infty) = (+\infty) + ' (-\infty) = +\infty. \quad (8)$$

Therefore, we can define the elementary morphological operators in terms of matrix products. Let $\delta_A, \varepsilon_A : \mathbb{R}_{\pm\infty}^n \rightarrow \mathbb{R}_{\pm\infty}^m$ for $A \in \mathbb{R}_{\pm\infty}^{n \times m}$. Then ε_A and δ_A can be respectively defined by

$$\delta_A(B) = A^T \vee B \quad \text{and} \quad \varepsilon_A(B) = A^T \wedge B, \quad (9)$$

where $B \in \mathbb{R}_{\pm\infty}^{n \times m}$, and term T denotes transposition.

As a particular case of the theorem proved in [9], it is possible to conclude that the operators ε_A and δ_A represent respectively an algebraic dilation and an algebraic erosion from a complete lattice $\mathbb{R}_{\pm\infty}^n$ to a complete lattice $\mathbb{R}_{\pm\infty}^m$. Therefore, every dilation $\delta : \mathbb{R}_{\pm\infty}^n \rightarrow \mathbb{R}_{\pm\infty}^m$ is of the form δ_A and every erosion $\varepsilon : \mathbb{R}_{\pm\infty}^n \rightarrow \mathbb{R}_{\pm\infty}^m$ is of the form ε_A .

This work have the main focus on $\mathbb{R}_{\pm\infty}$, because according to Araújo [12], financial forecasting problems can be modeled in terms of increasing functions $\Psi : \mathbb{R}_{\pm\infty}^d \rightarrow \mathbb{R}_{\pm\infty}$ (d represents the minimum necessary dimension to determining the characteristic phase space that generates the time series phenomenon, or, the time lags dimensionality), which can be approximated in terms of vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$, and given by

$$\Psi \simeq \delta_{\mathbf{a}} \quad \text{and} \quad \Psi \simeq \varepsilon_{\mathbf{b}}. \quad (10)$$

Therefore, this argumentation provides the basis for stock market indices forecasting using morphological-based perceptrons. Further details can be found in [12].

3. ANALYZED TIME SERIES

In this section we present the investigated stock market indices, where their graphics and lagplots are analyzed. The lagplot [24] consists of dispersion graph constructions relating the different time lags of the time series $(x_t \times x_{t-1}, x_t \times x_{t-2}, x_t \times x_{t-3}, \dots)$, and allows observations of possible relative strong relationships between any pair of time lags (when a structured appearance is shown in the graph). Although such technique is very limited since it depends on human interpretation of the graphs. However, its simplicity is a strong argument for its utilization [12].

3.1 Dow Jones Industrial Average Index (DJI Series)

The Dow Jones Industrial Average Index is the main important international stock market index, which shows how thirty large, publicly-owned companies based in the United States have traded during a standard trading session in the stock market. The DJI series corresponds to daily records of Dow Jones Industrial Average Index from 1998/01/01 to 2003/08/26. Figure 1 shows the DJI series graphic.

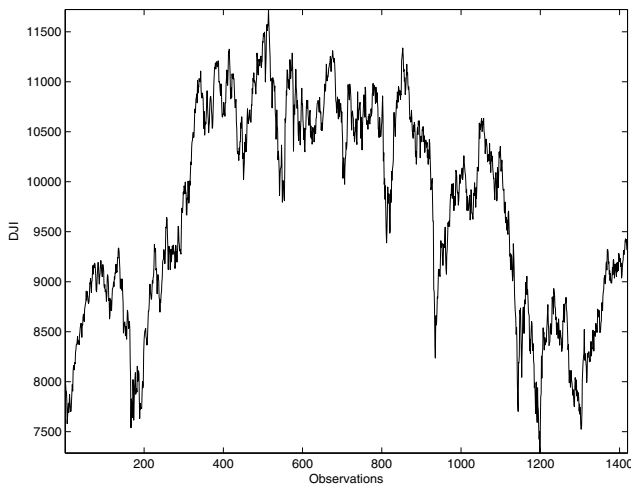


Figure 1: DJI series graphics.

In the Figure 2, it is presented the DJI series lagplot.

According to Figure 2, it is possible to see that for the first time lags of DJI series there is a clear dominant linear relationship among the lags. However, with the increase in the time lag degree, the appearance of a complex structure indicates a subdominant nonlinear relationship among the lags.

3.2 Standard & Poor 500 Stock Index (GSPC Series)

The Standard & Poor 500 Stock Index represents a free-float capitalization-weighted index of the prices of five hundred large-cap common stocks actively traded in the United States stock market. The GSPC series corresponds to monthly records of Standard & Poor 500 Stock Index from 1970/01 to 2003/08. Figure 3 illustrates the GSPC series graphic.

In the Figure 4, it is presented the GSPC series lagplot.

It is possible to see that, according to the Figure 4, for the first time lags of GSPC series there is a clear dominant linear relationship among the lags. However, once again, with the increase in the time lag degree, the appearance of a complex structure indicates a subdominant nonlinear relationship among the lags.

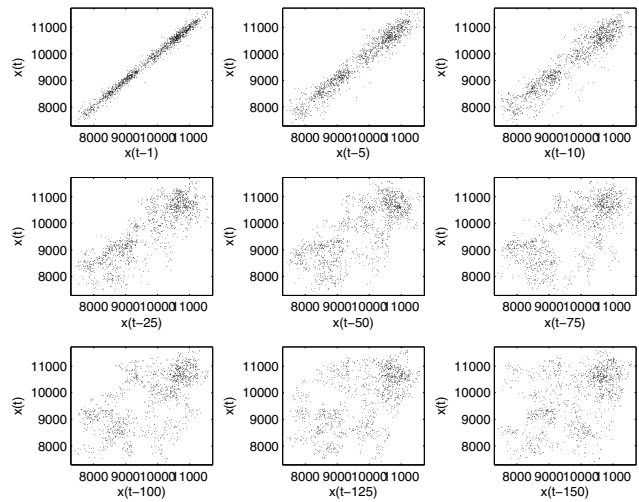


Figure 2: DJI series lagplot.

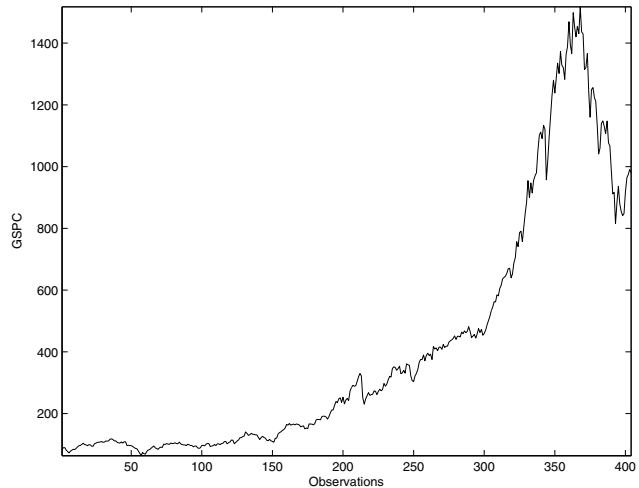


Figure 3: GSPC series graphics.

4. THE DILATION-EROSION PERCEPTOR (DEP)

Let $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^d$ a real-valued input signal inside an d -point moving window of the time series and let y the output of the DEP. Then, the DEP is defined by a translation invariant morphological system with local signal transformation rule $\mathbf{x} \rightarrow y$, given by

$$y = \lambda\alpha + (1 - \lambda)\beta, \quad \lambda \in [0, 1], \quad (11)$$

with

$$\alpha = \delta_{\mathbf{a}}(\mathbf{x}) = \bigvee_{i=1}^d (x_i + a_i), \quad (12)$$

and

$$\beta = \varepsilon_{\mathbf{b}}(\mathbf{x}) = \bigwedge_{i=1}^d (x_i + b_i), \quad (13)$$

where $\lambda \in \mathbb{R}$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$. Terms $\mathbf{a} = (a_1, a_2, \dots, a_d)$ and $\mathbf{b} = (b_1, b_2, \dots, b_d)$ represent the structuring elements of morphological operators of dilation and erosion, respectively, in which

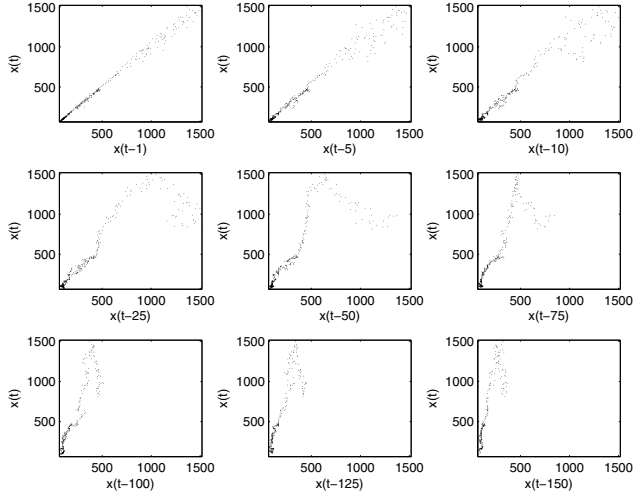


Figure 4: GSPC series lagplot.

d is the time lags dimensionality. In this way, it is worth mentioning that the DEP have a convex combination of its components, where when it increases the contribution of one component, the other one tends to decrease. The structure of DEP is illustrated in Figure 5.

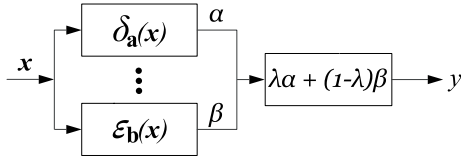


Figure 5: The proposed DEP structure.

5. THE PROPOSED EVOLUTIONARY LEARNING PROCESS

According to the DEP definition, we can see that the main objective of its design is to determine a set of parameters defined by $\mathbf{a} \in \mathbb{R}^d$, $\mathbf{b} \in \mathbb{R}^d$ and λ . Therefore, the weight vector ($\mathbf{w} \in \mathbb{R}^n$ with $n = 2d + 1$) to be used in the learning process is given by

$$\mathbf{w} = (\mathbf{a}, \mathbf{b}, \lambda). \quad (14)$$

During the proposed evolutionary learning process, the weights of the DEP are adjusted according to an error criterion until convergence or until the end of evolutionary process generations. Each i -th individual from population represents a candidate weight vector (denoted by \mathbf{w}_i) for the DEP model. The scheme to adjust the weight vector is initially to define a fitness function $f(\mathbf{w}_i)$ (which must reflect the solution quality achieved by the parameters configuration of the system), given by

$$f(\mathbf{w}_i) = \frac{1}{M} \sum_{j=1}^M e^2(j), \quad (15)$$

where M is the number of input patterns and $e(j)$ is the instantaneous error, given by

$$e(j) = t(j) - y(j), \quad (16)$$

where $t(k)$ and $y(j)$ are the target output and the actual model output for the j -th training pattern, respectively.

The evolutionary learning process proposed in this work, called DEP(CMAES), employ a covariance matrix adaptation evolutionary strategy (CMAES) based on the work of Hansen [19], which uses the idea of covariance matrix adaptation from a multivariate normal distribution into the evolutionary process, to train the DEP model. According to previous experiments with some CMAES versions, we decided to use the strategy $(\mu_{\mathbf{R}}, \gamma)$ -CMAES. A further discussion about this choice is beyond the scope of the paper, and we refer the reader to an upcoming journal paper.

The CMAES procedure starts with its parameters definition. We can note that there are eight initial parameters to be defined in the $(\mu_{\mathbf{R}}, \gamma)$ -CMAES [19]: i) γ is the population size, ii) μ is the parent individuals amount, that is, the search space points amount chosen into the population, iii) $\mathbf{r} = (r_1, r_2, \dots, r_\mu)$ is the recombination weight vector, iv) c_σ is a learning rate for the cumulation of step size control, v) e_σ is a damping parameter for step size update, vi) c_c is a learning rate for the cumulation of rank-one update of covariance matrix, vii) μ_{cov} is a weighting parameter between rank-one and rank- μ update, viii) c_{cov} represents the learning rate for covariance matrix update. The standard values of these parameters presented by Hansen [19] were used in this paper.

Next, the evolution paths $\mathbf{p}_c^{(0)} = \mathbf{0}$ and $\mathbf{p}_\sigma^{(0)} = \mathbf{0}$, the covariance matrix $\mathbf{C}^{(0)} = \mathbf{I}$, the evolution path step size $\sigma^{(0)} \in \mathbb{R}_+$ and the mean of search distribution $\mathbf{m}^{(0)} \in \mathbb{R}^n$ are initiated. Note that the search space is initially restricted within a cube $(\mathbf{m}^{(0)} \pm 2\sigma^{(0)}\mathbf{1})$ with $\mathbf{1} = (1, \dots, 1)^T$. Then, the CMAES starts a loop containing several steps to minimize the fitness function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, which is defined by Equation 15. Recall that term n represents the dimensionality of the DEP model weight vector, which is given by $2d + 1$.

Into this loop, the first step is to build a set of candidate individuals, or search space points, denoted by $\mathbf{w}_i^{(g)} \in \mathbb{R}^n$ with $i = 1, 2, \dots, \gamma$ and $g = 1, 2, \dots$, and sampled through a multivariate normal distribution. Then, the i -th population individual of the g -th CMAES generation is given by

$$\mathbf{w}_i^{(g)} \sim \mathcal{N}(\mathbf{m}^{(g)}, (\sigma^{(g)})^2 \mathbf{C}^{(g)}), \quad (17)$$

where \sim represents the equality among distributions. The term $\mathcal{N}(\mathbf{m}^{(g)}, (\sigma^{(g)})^2 \mathbf{C}^{(g)}) \sim \mathbf{m}^{(g)} + \sigma^{(g)} \mathcal{N}(\mathbf{0}, \mathbf{C}^{(g)}) \sim \mathbf{m}^{(g)} + \sigma^{(g)} \mathbf{B}^{(g)} \mathbf{D}^{(g)} \mathcal{N}(\mathbf{0}, \mathbf{I})$ represents a multi-variate normal distribution with mean $\mathbf{m}^{(g)}$ and covariance matrix $(\sigma^{(g)})^2 \mathbf{C}^{(g)}$, in which $\mathbf{B}^{(g)} \in \mathbb{R}^n$ is an orthogonal matrix of the eigenvectors from $\mathbf{C}^{(g)}$, $\mathbf{D}^{(g)} \in \mathbb{R}^n$ is a diagonal matrix of the squared roots of the eigenvalues from $\mathbf{C}^{(g)}$, and $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is a multi-variate normal distribution with zero mean and unitary covariance matrix.

The next steps definition is to update the parameters $\mathbf{m}^{(g)}$, $\mathbf{p}_\sigma^{(g)}$, $\sigma^{(g)}$, $\mathbf{p}_c^{(g)}$ and $\mathbf{C}^{(g)}$.

The mean of search distribution $\mathbf{m}^{(g)}$ can be updated by [19]

$$\mathbf{m}^{(g)} = \sum_{i=1}^{\mu} r_i \mathbf{w}_{i:\gamma}^{(g)}, \quad (18)$$

with

$$\sum_{i=1}^{\mu} r_i = 1, \quad r_i > 0 \text{ for } i = 1, \dots, \mu, \quad (19)$$

where $\mu \leq \gamma$, $r_{i=1, \dots, \mu} \in \mathbb{R}_+$ with $r_1 \geq r_2 \geq \dots \geq r_\mu$, and $\mathbf{w}_{i:\gamma}^{(g)}$ the i -th best individual. Note that the index $1 : \gamma$ denotes the index of i -th individual and $f(\mathbf{w}_{1:\gamma}^{(g)}) \leq f(\mathbf{w}_{2:\gamma}^{(g)}) \leq \dots \leq f(\mathbf{w}_{\gamma:\gamma}^{(g)})$.

The Equation 18 implements the recombination process using a weighted sum of μ individuals, as well as implements the selection

operator choosing $\mu < \gamma$ and setting distinct values of the recombination weights vector \mathbf{r} [19].

The conjugate evolution path $\mathbf{p}_\sigma^{(g)} \in \mathbb{R}^n$ can be updated by [19]

$$\mathbf{p}_\sigma^{(g)} = (1 - c_\sigma)\mathbf{p}_\sigma^{(g-1)} + \sqrt{c_\sigma(2 - c_\sigma)\mu_{eff}} \mathbf{C}_\sigma^{(g-1)^{-\frac{1}{2}}} \frac{\mathbf{m}^{(g)} - \mathbf{m}^{(g-1)}}{\sigma^{(g-1)}}, \quad (20)$$

where $(1 - c_\sigma)$ is an adjustment factor, $\sqrt{c_\sigma(2 - c_\sigma)\mu_{eff}}$ is a normalization constant, $\mathbf{C}_\sigma^{(g-1)^{-\frac{1}{2}}} \stackrel{def}{=} \mathbf{B}^{(g-1)} \mathbf{D}^{(g-1)^{-1}} \mathbf{B}^{(g-1)T}$ with $\mathbf{C}_\sigma^{(g-1)} = \mathbf{B}^{(g-1)} (\mathbf{D}^{(g-1)})^2 \mathbf{B}^{(g-1)T}$ is an eigendecomposition from $\mathbf{C}_\sigma^{(g-1)}$, and $\mathbf{B}^{(g-1)}$ is an orthonormal basis of eigenvectors. The elements of diagonal matrix $\mathbf{D}^{(g-1)}$ are the squared roots of its correspondent positive eigenvalues.

The evolution path $\sigma^{(g)}$ can be updated by [19]

$$\sigma^{(g)} = \sigma^{(g-1)} \exp \left[\frac{c_\sigma}{e_\sigma} \left(\frac{\|\mathbf{p}_\sigma^{(g)}\|}{E\|\mathcal{N}(\mathbf{0}, \mathbf{I})\|} - 1 \right) \right], \quad (21)$$

in which

$$E\|\mathcal{N}(\mathbf{0}, \mathbf{I})\| = \frac{\sqrt{2}\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \approx \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right). \quad (22)$$

Equations 20 and 21 implement the evolution step control [19].

The evolution path $\mathbf{p}_c^{(g)}$ can be updated by [19]

$$\mathbf{p}_c^{(g)} = (1 - c_c)\mathbf{p}_c^{(g-1)} + \vartheta_\sigma^{(g)} \sqrt{c_c(2 - c_c)\mu_{eff}} \frac{\mathbf{m}^{(g)} - \mathbf{m}^{(g-1)}}{\sigma^{(g-1)}}, \quad (23)$$

in which

$$\vartheta_\sigma^{(g)} = \begin{cases} 1, & \text{if } \frac{\|\mathbf{p}_\sigma^{(g)}\|}{\sqrt{1 - (1 - c_\sigma)}} < \left(1.5 + \frac{1}{n - 0.5} \right) E\|\mathcal{N}(\mathbf{0}, \mathbf{I})\| \\ 0, & \text{otherwise} \end{cases} \quad (24)$$

At the end, the covariance matrix $\mathbf{C}^{(g)}$ can be updated by [19]

$$\mathbf{C}^{(g)} = (1 - c_{cov})\mathbf{C}^{(g-1)} + \frac{c_{cov}}{\mu_{cov}} \left[\mathbf{p}_c^{(g)} \mathbf{p}_c^{(g)T} + \zeta(\vartheta_\sigma^{(g)})\mathbf{C}^{(g-1)} \right] + c_{cov} \left(1 - \frac{1}{\mu_{cov}} \right) r\mu u, \quad (25)$$

in which

$$\zeta(\vartheta_\sigma^{(g)}) = (1 - \vartheta_\sigma^{(g)})c_c(2 - c_c). \quad (26)$$

and

$$r\mu u = \sum_{i=1}^{\mu} \left[r_i \left(\frac{\mathbf{w}_{i:\gamma}^{(g)} - \mathbf{m}^{(g-1)}}{\sigma^{(g-1)}} \right) \left(\frac{\mathbf{w}_{i:\gamma}^{(g)} - \mathbf{m}^{(g-1)}}{\sigma^{(g-1)}} \right)^T \right]. \quad (27)$$

Further details of CMAES can be found in [19].

Besides, in order to automatically adjust time phase distortions in some time series representation, we have included an automatic phase fix procedure (APFP) [12] in the proposed learning process of the DEP model. Figure 6 presents the APFP.

According to Figure 6, in the first step an input pattern \mathbf{x} is presented to DEP generating the output y_1 . The first output y_1 is used to rebuild the input pattern in the second step. This reconstructed pattern is presented to the same DEP generating the second output y_2 , which is the phase fixed forecasting.

Figure 7 presents the steps of our proposed evolutionary learning process including the APFP.

It is worth mentioning that three stop conditions are used in the proposed learning process:

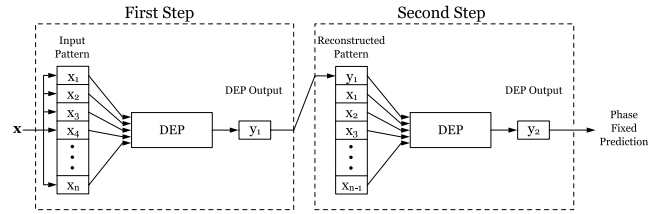


Figure 6: Automatic phase fix procedure.

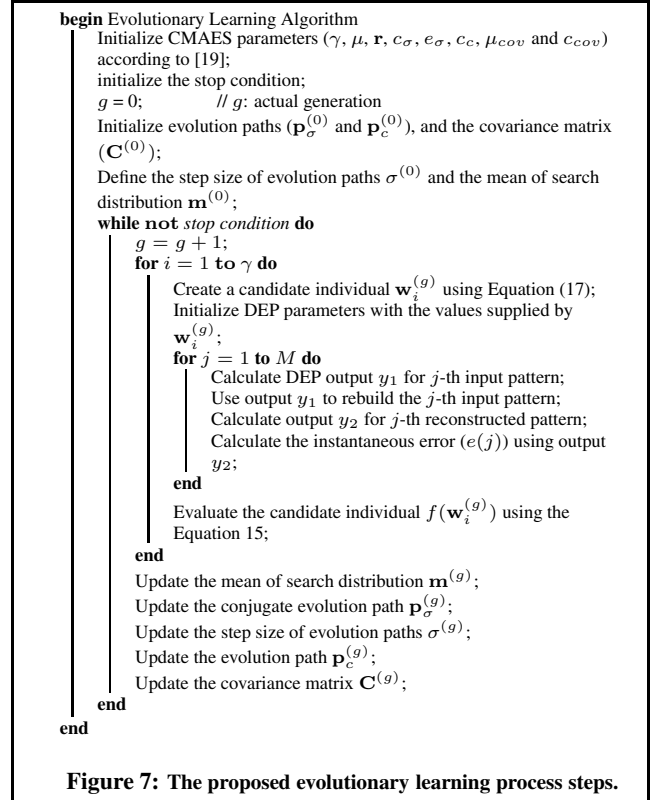


Figure 7: The proposed evolutionary learning process steps.

1. The maximum generation number: $g = 10^4$;
2. The decrease in the training error process training (Pt) [25] of the cost function: $Pt \leq 10^{-6}$.
3. The increase in the validation error or generalization loss (Gl) [25] of the cost function: $Gl > 5\%$;

6. EXPERIMENTAL RESULTS

Two real world time series (DJI and GSPC) were used as a test bed for evaluation of the proposed model. Both time series were normalized to lie within the range $[0, 1]$ and divided in three sets according to Prechelt [25]: training set (50% of the data points), validation set (25% of the data points) and test set (25% of the data points).

In order to establish a performance study, results previously published in the literature with the DEP model, using the classical gradient-based learning process (DEP(GB)), are employed in our comparative analysis, where we investigate the same time series under the same conditions. The procedure used to generate the DEP(GB) is the same presented in the work of Araújo [12]. These

procedures were not formally described because all of them are well-defined and explained in the paper previously mentioned.

Additionally, we use five well-known evaluation metrics formally defined in [12] to assess the forecasting performance: mean square error (MSE), mean absolute percentage error (MAPE), u of their statistic (UTS), prediction of change in direction (POCID) and average relative variance (ARV). Also, we use an evaluation function (EF) defined in [12] to serve as a global performance indicator for arbitrary forecasting models.

For each time series, five experiments were performed, where we calculate the mean (MEAN) and the standard deviation (STD) in the attempt to obtain an average forecasting performance of the proposed DEP(CMAES) model. Also, we calculate all confidence intervals (CI) with the assumption of normal distribution with 99% of certainty degree. In addition, we include in our analysis an additional measure referred to as the percentage gain (PG), which measures, in percentage terms, how much better is the DEP(CMAES) regarding DEP(GB). The PG is formally defined by

$$PG = 100 - 100 \frac{pmm}{imm}, \quad (28)$$

or

$$PG = 100 \frac{pmm}{imm} - 100, \quad (29)$$

in which pmm and imm represent the evaluation metric value found by DEP(CMAES) and by DEP(GB), respectively. Note that the Equation 28 must be used to measure the obtained gains for MSE, MAPE, UTS and ARV metrics, while the Equation 29 must be used to measure the obtained gains for POCID and EF metrics.

6.1 DJI Series

For the DJI series forecasting (with one step ahead of forecasting horizon $-H = 1$), we use the same time lags presented in [12] to create the input patterns (lags 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 – note that here $d = 10$). The Table 1 shows the experiments performed with the DEP(CMAES) model, where we calculate all evaluation metrics, as well as their MEAN, STD and CI.

According to the Table 1, we can see that in all experiments the POCID metric is greater than 50%, indicating that the proposed model had much better performance than a “coin-tossing” experiment, and the obtained UTS measure ($\approx 1.0e-003$) indicated that the DEP(CMAES) model overcame the random walk dilemma. We can also verify that the DEP(CMAES) obtained small STD values, demonstrating the stability of the CMAES to train the DEP model.

Table 2 presents a performance study with the best results obtained by the proposed DEP(CMAES) with those previously presented in the literature with the DEP(GB) for DJI series.

Table 2: Best results (test set) for DJI series with DEP(BP) and DEP(CMAES) models.

| Evaluation Metrics | DEP(GB) | DEP(CMAES) |
|--------------------|-------------|-------------|
| MSE | 3.5086e-007 | 1.4031e-008 |
| MAPE | 2.3699e-003 | 4.7393e-004 |
| UTS | 1.6324e-002 | 1.6718e-005 |
| ARV | 1.4364e-005 | 5.7440e-007 |
| POCID | 99.14 | 99.43 |
| EF | 97.3193 | 99.3846 |

Analyzing the Table 2, we can verify that the proposed model overcame the DEP(GB) in this work. However, to take more precise indications of the best performance of the proposed model, we present in Table 3 the obtained PG of the DJI series.

Table 3: Percentage gain (test set) for DJI series of the DEP(CMAES) regarding the DEP(GB).

| Evaluation Metrics | PG (%) DEP(CMAES) / DEP(GB) |
|--------------------|-----------------------------|
| MSE | 96.00 |
| MAPE | 80.00 |
| UTS | 99.90 |
| ARV | 96.00 |
| POCID | 0.29 |
| EF | 2.12 |

According to Table 3 we can verify a better forecasting performance of the DEP(CMAES) regarding the DEP(GB) (having a PG greater than 80% for all metrics, except for POCID metric, having a PG greater than 0.2%). In addition, assessing the DEP(CMAES) in terms of overall forecasting performance (using EF metric), we have a PG around 2.0% regarding DEP(GB) model.

Finally, we present in Figure 8 a comparative graphic between real (solid line) and predicted (dashed line) values generated by DEP(CMAES) for the last ten points of the DJI series test set. We can see that the predicted values are superimposed to the real values of the DJI series, where the one step delay regarding the forecasting values did not occur, that is, the time phase distortion that causes the random walk dilemma was successfully adjusted.

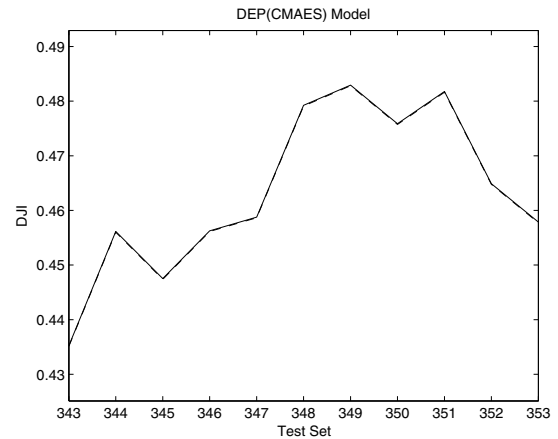


Figure 8: Forecasting results of DJI series (last ten points of the test set): actual values (solid line) and predicted values (dashed line).

6.2 GSPC Series

For the GSPC series forecasting (with one step ahead of forecasting horizon $-H = 1$), we use the same time lags presented in [12] to create the input patterns (lags 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 – once again $d = 10$). In Table 4 we present the experiments performed with the DEP(CMAES) model, including all evaluation metrics, as well as their MEAN, STD and CI.

According to the Table 4, we can note that in all experiments the POCID measure showed that the DEP(CMAES) model had much better performance than a “coin-tossing” experiment (POCID = 98.99%), and the obtained UTS measure indicated that the proposed model overcame the random walk dilemma ($\approx 3.0e-003$). Once again, analyzing the STD values obtained by the proposed model, we can verify the high stability of the CMAES to train the DEP model.

Table 1: Results for all experiments with the proposed DEP(CMAES) model for DJI series (test set).

| Statistics | Evaluation Metrics | | | | | |
|------------|--------------------|--------------|--------------|--------------|-------|---------|
| | MSE | MAPE | UTS | ARV | POCID | EF |
| | 1.4031e-008 | 4.7393e-004 | 1.6718e-005 | 5.7440e-007 | 99.43 | 99.3846 |
| | 2.7583e-006 | 6.6451e-003 | 3.2867e-003 | 1.1292e-004 | 99.43 | 98.4443 |
| | 4.4454e-007 | 2.6677e-003 | 5.2970e-004 | 1.8199e-005 | 99.43 | 99.1147 |
| | 6.0964e-007 | 3.1240e-003 | 7.2643e-004 | 2.4958e-005 | 99.43 | 99.0495 |
| | 1.3153e-006 | 4.5887e-003 | 1.5673e-003 | 5.3847e-005 | 99.43 | 98.8196 |
| MEAN | 1.0284e-006 | 3.4999e-003 | 1.2254e-003 | 4.2101e-005 | 99.43 | 98.9625 |
| STD | 4.5680e-006 | 2.2950e-003 | 1.2806e-003 | 4.3999e-005 | 0.00 | 0.3528 |
| CI | ±8.9102e-006 | ±2.6480e-003 | ±1.4776e-003 | ±5.0766e-005 | ±0.00 | ±0.4070 |

Table 4: Results for all experiments with the proposed DEP(CMAES) model for GSPC series (test set).

| Statistics | Evaluation Metrics | | | | | |
|------------|--------------------|--------------|--------------|--------------|-------|---------|
| | MSE | MAPE | UTS | ARV | POCID | EF |
| | 2.6348e-006 | 2.6858e-003 | 1.9710e-003 | 7.5439e-005 | 98.99 | 98.5234 |
| | 2.3350e-006 | 2.5284e-003 | 1.7468e-003 | 6.6855e-005 | 98.99 | 98.5617 |
| | 2.1143e-005 | 7.6083e-003 | 1.5817e-002 | 6.0537e-004 | 98.99 | 96.6649 |
| | 3.3089e-007 | 9.5181e-004 | 2.4754e-004 | 9.4741e-006 | 98.99 | 98.8704 |
| | 1.2974e-007 | 5.9599e-004 | 9.7056e-005 | 3.7147e-006 | 98.99 | 98.9210 |
| MEAN | 5.3147e-006 | 2.8741e-003 | 3.9758e-003 | 1.5217e-004 | 98.99 | 98.3083 |
| STD | 8.9207e-006 | 2.8041e-003 | 6.6735e-003 | 2.5542e-004 | 0.00 | 0.9357 |
| CI | ±1.0293e-005 | ±3.2354e-003 | ±7.7000e-003 | ±2.9471e-004 | ±0.00 | ±1.0797 |

Table 5 presents a performance study regarding the best results obtained by the proposed DEP(CMAES) with those previously presented in the literature with the DEP(GB) for GSPC series.

Table 5: Best results (test set) for GSPC series with DEP(BP) and DEP(CMAES) models.

| Evaluation Metrics | DEP(GB) | DEP(CMAES) |
|--------------------|-------------|-------------|
| MSE | 5.6819e-006 | 1.2974e-007 |
| MAPE | 2.9699e-003 | 5.9599e-004 |
| UTS | 0.2531e-000 | 9.7056e-005 |
| ARV | 4.3707e-004 | 3.7147e-006 |
| POCID | 94.44 | 98.99 |
| EF | 75.1604 | 98.9210 |

According to Table 5 we can see that the proposed DEP(CMAES) model had best forecasting performance than the DEP(GB). However, in order to have more precise indication of the best performance of the proposed model, we present in Table 6 the obtained PG of the GSPC series.

Table 6: Percentage gain (test set) for GSPC series of the DEP(CMAES) regarding the DEP(GB).

| Metrics | PG (%) DEP(CMAES) / DEP(GB) |
|---------|-----------------------------|
| MSE | 97.72 |
| MAPE | 79.93 |
| UTS | 99.96 |
| ARV | 99.15 |
| POCID | 4.82 |
| EF | 31.61 |

When we analyze the Table 6, we can verify a better forecasting performance of the DEP(CMAES) regarding the DEP(GB) (having a PG greater than 79% for all metrics, except for POCID metric, having a PG greater than 4.0%. In addition, assessing the

DEP(CMAES) in terms of overall predictive performance (using EF metric), we have a PG around 31.0% regarding DEP(GB) model.

In the Figure 9, we present a comparative graphic between real (solid line) and predicted (dashed line) values generated by the proposed model for the last ten points of the GSPC series test set. Once again, we can not that the predicted values are very close to the real values of the GSPC series, where the one step delay regarding the prediction values did not occur, that it, the time phase distortion that causes the random walk dilemma was successfully adjusted.

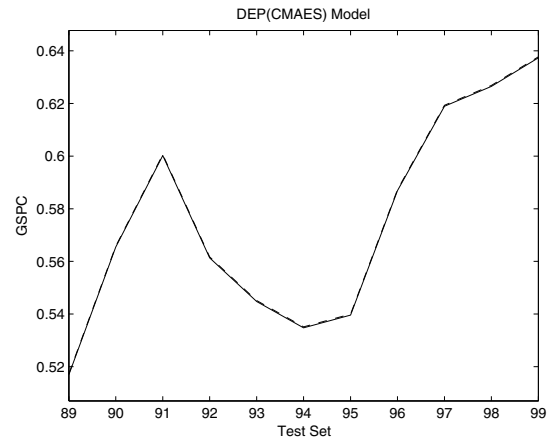


Figure 9: Prediction results of GSPC series (last ten points of the test set): actual values (solid line) and predicted values (dashed line).

7. CONCLUSION

In this work we presented an evolutionary learning process, called DEP(CMAES), using the covariance matrix adaptation evolutionary strategy (CMAES) to design the dilation-erosion perceptron

(DEP), in the attempt to solve the stock market indices forecasting problem. The main motivation of the proposed learning process arisen from the drawback of classical gradient-based learning of the DEP model, called DEP(GB), which requires a systematic methodology to overcome the problem of nondifferentiability of morphological operators in training equations. Also, due to time phase distortions that occur in the reconstruction of phase space of some temporal phenomena, we have included an automatic phase fix procedure (APFP) into proposed learning process.

The evaluation performance of the proposed model regarding to DEP(GB) model was assessed in terms of five well-known performance measures and using two stock market indices time series (with all their dependencies on exogenous and uncontrollable variables): DJI and GSPC series. In addition, an evaluation function served as a global indicator for the quality of solutions achieved by the investigated models.

The experimental results demonstrated a consistently better performance, of the proposed learning process, for training the DEP model in comparison to the classical gradient-based learning process. With the inclusion of the APFP into the proposed learning process of the DEP model, we succeeded in automatically correcting the time phase distortions that typically occur in some forecasting problems, where our forecasts have not any one step delay regarding real time series values. A feasible explanation for such behavior is that the APFP depend on the information complexity contained in the time series data and the ability to accurately define the best forecasting model parameters to estimate the real time series values, in other words, the success of the APFP is strongly dependent on an accurate adjustment of the forecasting model parameters. Therefore, we can verify that the CMAES was more efficient and stable, regarding the gradient-based learning, to train the DEP model, adjusting more precisely the time phase distortions that occur in the analyzed time series in this work.

Future works will consider further studies, in terms of risk and financial return, to determine the additional economical benefits, for an investor, with the use of the DEP(CMAES) model in stock market applications, as well as a particular study about the computing complexity and CPU time of the proposed learning process to establish a complete cost-performance evaluation.

8. REFERENCES

- [1] G. X. Ritter and P. Sussner. Morphological perceptrons. In *ISAS'97, Intelligent Systems and Semiotics*, Gaithersburg, Maryland, 1997.
- [2] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, London, 1982.
- [3] P. Maragos. A representation theory for morphological image and signal processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:586–599, 1989.
- [4] J. Serra. *Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances*. Academic Press, New York, 1988.
- [5] C. Ronse. Why mathematical morphology needs complete lattices. *Signal Processing*, 21(2):129–154, October 1990.
- [6] H. J. A. M. Heijmans. *Morphological Image Operators*. Academic Press, New York, NY, 1994.
- [7] G. X. Ritter and G. Urcid. Lattice algebra approach to single-neuron computation. *IEEE Transactions on Neural Network*, 14(2):282–295, March 2003.
- [8] L. F. C. Pessoa and P. Maragos. Neural networks with hybrid morphological rank linear nodes: a unifying framework with applications to handwritten character recognition. *Pattern Recognition*, 33:945–960, 2000.
- [9] P. Sussner and E. L. Esmi. Morphological perceptrons with competitive learning: Lattice-theoretical framework and constructive learning algorithm. *Information Sciences*, 2009.
- [10] R. de A. Araújo and P. Sussner. 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*. IEEE, 2010.
- [11] R. de A. Araújo and P. Sussner. An increasing hybrid morphological-linear perceptron with evolutionary learning and phase correction for financial time series forecasting. In *International Conference on Hybrid Artificial Intelligence Systems*, 2010.
- [12] R. de A. Araújo. A class of hybrid morphological perceptrons with application in time series forecasting. *Knowledge-Based Systems*, 2011. . Accepted for Publication.
- [13] R. de A. Araújo and T. A. E. Ferreira. An intelligent hybrid morphological-rank-linear method for financial time series prediction. *Neurocomputing*, 72(10-12):2507–2524, 2009.
- [14] R. de A. Araújo. A hybrid intelligent morphological approach for stock market forecasting. *Neural Processing Letters*, 31(3):195–217, 2010.
- [15] R. de A. Araújo. Hybrid intelligent methodology to design translation invariant morphological operators for brazilian stock market prediction. *Neural Networks*, 23(10):1238–1251, 2010.
- [16] R. de A. Araújo. Swarm-based translation-invariant morphological method for financial time series forecasting. *Information Sciences*, 180(24):4784–4805, 2010.
- [17] R. de A. Araújo. Translation invariant morphological time-lag added evolutionary forecasting method for stock market prediction. *Expert Systems with Applications*, 38(3):2835–2848, 2011.
- [18] R. Sitte and J. Sitte. Neural networks approach to the random walk dilemma of financial time series. *Applied Intelligence*, 16(3):163–171, May 2002.
- [19] N. Hansen. The CMA evolution strategy: a comparing review. In J.A. Lozano, P. Larranaga, I. Inza, and E. Bengoetxea, editors, *Towards a new evolutionary computation. Advances on estimation of distribution algorithms*, pages 75–102. Springer, 2006.
- [20] R. de A. Araújo and T. A. E. Ferreira. A morphological-rank-linear evolutionary method for stock market prediction. *Information Sciences*, 2010. . In Press.
- [21] F. Takens. Detecting strange attractor in turbulence. In A. Dold and B. Eckmann, editors, *Dynamical Systems and Turbulence*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381, New York, 1980. Springer-Verlag.
- [22] G. Birkhoff. *Lattice Theory*. American Mathematical Society, Providence, 3rd edition, 1993.
- [23] R. Cuninghame-Green. *Minimax Algebra: Lecture Notes in Economics and Mathematical Systems 166*. Springer-Verlag, New York, 1979.
- [24] H. Kantz and T. Schreiber. *Nonlinear Time Series analysis*. Cambridge University Press, New York, NY, USA, second edition, 2003.
- [25] L. Prechelt. Proben1: A set of neural network benchmark problems and benchmarking rules. Technical Report 21/94, 1994.