

Estatística Clássica e Inteligência Computacional na Previsão de Séries Temporais

Juliana C. Carvalho de Araújo

Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro Brasil
juliana@ele.puc-rio.br

Reinaldo Castro Souza

Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro Brasil
reinaldo@ele.puc-rio.br

Resumo — O objetivo da análise de séries temporais é identificar padrões não aleatórios na série temporal de uma variável de interesse, e a observação de como esse comportamento passado pode permitir a realização de previsões sobre o futuro, tanto orientando a tomada de decisões no curto, quanto no médio e longo prazo. Neste trabalho, são utilizadas técnicas computacionais e de estatística clássica e suas particularidades no tratamento de séries temporais. Desse modo, foram selecionadas, 11 séries temporais com diferentes padrões de comportamento de sazonalidade, de tendência e de periodicidade que compõem o conjunto reduzido da Competição M3. Os resultados demonstram que diferentes métodos podem ser muitas vezes usados conjuntamente ou alternativamente para análises de diversas séries temporais e apontam para a necessidade de aprimoramento dos métodos de previsão a fim de fornecer resultados mais precisos.

Palavras-Chave— *previsão, séries temporais, inteligência computacional.*

I. INTRODUÇÃO

Séries temporais é uma sequência escalar de dados numéricos, y_0, y_1, y_2, \dots observados em um intervalo de tempo equidistante e que apresentam dependência serial entre si. Devidamente arranjado, um conjunto de dados de séries temporais escalares, $y_{t-(n-1)\tau}, \dots, y_{t-2\tau}, y_{t-\tau}, y_t, y_w$, constitui um vetor. O subscrito t representa o tempo atual, τ para tempo de atraso (também chamado lag temporal), $w = t + T$ para tempo de previsão, T para o tempo futuro (também chamado horizonte de previsão), e n para a incorporação da dimensão euclidiana do espaço de estado onde cada vetor é um ponto. Em uma definição mais rigorosa define-se séries temporais como uma realização qualquer de um processo estocástico.

A previsão de séries temporais está presente em diversas áreas, como o setor elétrico, financeiro, econômico e industrial. O objetivo da análise de séries temporais é identificar padrões não aleatórios na série temporal de uma variável de interesse, e a observação deste comportamento passado pode permitir a realização de previsões sobre o futuro, orientando a tomada de decisões no curto, médio e longo prazo.

Os métodos convencionais de previsão de séries temporais, em sua grande maioria exigem a estacionariedade da série, o que não ocorre com as séries financeiras e econômicas, por

exemplo, reduzindo assim a acurácia dos resultados. Nesse sentido, técnicas de inteligência computacional têm sido cada vez mais aplicadas em problemas de previsão de séries temporais, com destaque para as Redes Neurais Artificiais (RNA) [1][2] e a Programação Genética (PG) [3] como uma alternativa aos métodos tradicionais por apresentar resultados promissores [4][5][6].

Como em geral as técnicas da Inteligência Artificial apresentam erros de previsão inferiores aos obtidos pelos métodos tradicionais. O objetivo deste trabalho é o aprimoramento da qualidade da previsão de séries temporais, utilizando programação multigênica e posterior comparação com os métodos de estatística clássica (amortecimento exponencial e Box-Jenkins) e inteligência computacional (redes neurais e programação genética).

O presente trabalho está dividido em cinco seções: a seção 2 apresenta uma breve introdução sobre as técnicas de previsão utilizadas. A seção 3 contém informações sobre a base utilizada. Na seção 4, são apresentados os experimentos realizados e os resultados obtidos. Na seção 5, são apresentadas as conclusões e trabalhos futuros.

II. TÉCNICAS DE PREVISÃO DE SÉRIES TEMPORAIS

Para a implementação dos procedimentos estatísticos, foram utilizados neste estudo os seguintes softwares computacionais: Projeto R [7] e Matlab [8].

Uma breve descrição dos métodos estatísticos e de inteligência computacional trabalhados será apresentada a seguir.

A. Amortecimento Exponencial.

Há inúmeras técnicas estatísticas para previsão de séries temporais. A maioria se baseia na decomposição estrutural de um ou mais componentes da série temporal [9]: nível, tendência, sazonalidade, ciclo e ruído.

Modelos de suavização exponencial ou amortecimento exponencial (AES), são métodos de decomposição baseados na ideia de que observações passadas contêm informações sobre o padrão da série temporal. O propósito dos métodos é distinguir um padrão de comportamento de qualquer outro ruído que possa estar contido nas observações da série e então usar esse padrão para prever valores futuros da série.

Nível, tendência e sazonalidade podem ser combinados através de operadores de adição ou multiplicação. Assim 24 variações da suavização exponencial podem ser identificados [10]. Entretanto há três variações básicas de métodos de suavização exponencial que são mais comumente usados: suavização exponencial simples [11]; suavização exponencial [12] com correção de tendência; e método de Holt- Winters [13].

Para este estudo específico utilizamos o modelo de Holt para as séries anuais e de Holt Winters para as séries trimestrais e mensais.

B. Modelos de Box-Jenkins.

Box-Jenkins (B&J) considera um conjunto de processos estocásticos denominados de ARIMA (*Autoregressive Integrated Moving Average*), resultantes da combinação de componentes auto-regressivo (AR), integração (I) e médias móveis (MA). É representado por ARIMA (p, d, q), onde p é o número de defasagens da série, d é a ordem de integração para tornar a série estacionária e q o número de defasagens dos erros aleatórios [14].

Esses modelos são usados para analisar as séries temporais não estacionárias, transformando-as em estacionárias, com média e variância constante ao longo do tempo, através da ordem de integração [14]. A necessidade da estacionariedade advém do fato de que qualquer modelo inferido a partir desses dados possa ser considerado estável, fornecendo uma base válida para previsão. Os dados que compõem a série devem possuir distribuição normal e variância constante.

A metodologia desenvolve modelos seguindo um ciclo iterativo, considerando os próprios dados [14]. O ciclo iterativo é composto de:

- Identificação - Analisam-se a FAC e FACP, e tenta-se identificar o modelo.
- Estimação - Os modelos ajustados são comparados utilizando alguns critérios, como o da parcimônia. Uma das formas de melhorar o grau de ajuste desse modelo aos dados da série temporal se dá com a inclusão de defasagens adicionais nos processos AR(p), MA(q), ARMA (p,q), ARIMA(p,d,q) e SARIMA (p,d,q)*(P,D,Q).
- Verificação - Se os resíduos são autocorrelacionados, então, a dinâmica da série não é completamente explicada pelos coeficientes do modelo ajustado.
- Previsão - As previsões podem ser *ex-ante*, feitas para calcular valores futuros, de curto prazo, da variável em estudo. Ou, *ex-post* realizada para gerar valores dentro do período amostral. Quanto melhor forem essas últimas, mais eficiente será o modelo estimado.

Se o modelo atender a todos os pressupostos das etapas do ciclo iterativo, o mesmo é adequado para realizar previsões.

C. Programação Genética.

A programação genética (PG) [15] é uma subárea da computação evolutiva que consiste em evoluir programas

computacionais com o objetivo de encontrar uma função ou um programa que realize o melhor ajuste entre um conjunto de dados de entrada e de saída.

Os indivíduos que serão evoluídos, são programas representados por estruturas de árvores sintáticas que possuem funções e terminais, que determinam suas características e definem seu comportamento no ambiente em que foi desenvolvido. Cada função é um ramo da árvore, e cada terminal é uma folha do indivíduo. As funções podem ser, por exemplo, operações lógicas ou aritméticas, ou funções iterativas, e os terminais podem ser variáveis, constantes, ou funções que não recebem argumentos. Além do conjunto de dados, um conjunto de funções ou comandos válidos deve ser fornecido. Este último conjunto deve ter uma importante propriedade conhecida como *closure* [15], que pode ser basicamente definida como a consistência entre o cruzamento aleatório de subárvores, garantido que uma subárvore pode ser usada como argumento de qualquer função.

Um exemplo de um indivíduo pode ser observado na figura 1.

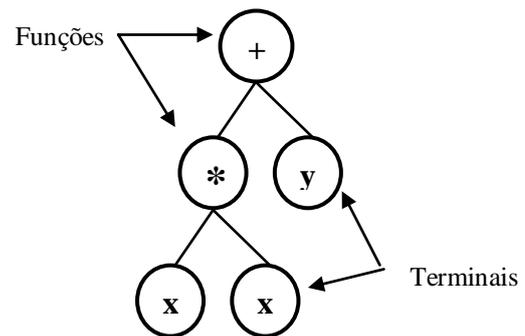


Fig. 1: Exemplo de indivíduo.

Operadores genéticos, como cruzamento, mutação e reprodução, atuam nos cromossomos, que neste caso, são programas representados, usualmente, por estrutura em árvores. Esses cromossomos devem ser avaliados a partir da sua execução para todo o conjunto de dados de entrada. O resultado desta execução é comparado com os dados de saída, segundo alguma métrica de acurácia que mensure a proximidade entre o valor encontrado e o valor real [16].

Tipicamente, a PG usa uma população de estruturas em árvores. Para cada uma é codificada uma equação matemática que prediz e descreve um vetor de saídas y_i , usando uma matriz (X_{ij}) de variáveis explicativas. A PG Multigênica [17] perfaz uma combinação linear (regressão simbólica) de uma quantidade de n saídas quaisquer de programas, onde cada programa pode ser considerado um "gene" em um indivíduo.

Neste trabalho, em especial, utilizaremos uma regressão simbólica que aproxime uma função que melhor se ajuste a uma séries temporal e posteriormente tente prever seus valores futuros.

A formulação do problema considerou o uso do mesmo conjunto de variáveis da competição M3 utilizado na

modelagem estatística e seus valores defasados. A função de avaliação dos indivíduos pode ser formulada em termos do erro quadrático médio (EQM):

$$EQM = \sum_{t=1}^N (y_t - \hat{y}_t)^2,$$

onde y_t = valor atual, \hat{y}_t = valor previsto e N = número de observações.

D. Rede Neural Não-Linear Autorregressiva

As redes neurais artificiais (RNA) são modelos paramétricos não-lineares bio-inspirados nas conexões neurológicas do cérebro humano. As RNAs caracterizam-se por um processador maciçamente e paralelamente distribuído de unidades de processamento simples, capaz de armazenar conhecimento experimental e torná-lo disponível para o uso [18].

Uma das principais aplicações das redes neurais artificiais (RN) esta relacionada com o reconhecimento de padrões temporais e previsão, sendo para isso utilizada uma arquitetura de rede denominada como rede neural recorrente. A característica da RN recorrente é possuir um ou mais laços de realimentação

O modelo de rede neural não-linear autorregressiva (NAR) e dado pela seguinte expressão:

$$y(t) = f(y(t-1), y(t-2), \dots, y(t-p)) + \epsilon(t)$$

Onde f é uma função suave desconhecida, $y(t)$ é a saída, t é o vetor temporal, e p é o número de termos autorregressivos, e $\epsilon(t)$ é o termo de error. Assume-se que o termo de error é independente e identicamente distribuído (i.i.d.) com média condicional zero e variância finita e constante [19][20].

A grande vantagem de se usar RN está na sua capacidade de lidar com perturbações temporais não lineares [20]. Redes neurais apresentam características de aproximadores universais, ou seja, uma rede neural pode detectar implicitamente qualquer relação não-linear entre a variável saída e as variáveis de entrada. Adicionalmente, a não necessidade de independência e normalidade das variáveis em estudo, bem como a sua grande capacidade de aprendizado a partir do ambiente, a aplicação de redes neurais artificiais na análise de problemas complexos de reconhecimento de padrões temporais. Por outro lado, uma potencial desvantagem das redes neurais, está associada ao fato dos parâmetros (pesos sinápticos) não terem uma interpretação imediata, exigindo análise adicional para se compreender a forma com que a informação é extraída.

III. COMPETIÇÃO M3 E DADOS UTILIZADOS

Foram selecionadas, 11 séries temporais com diferentes padrões de comportamento de sazonalidade, de tendência e de periodicidade que compõem o conjunto reduzido do *Forecasting Competition M3* organizado pelo *International Institute of Forecasters* e publicado no *International Journal of Forecasting*.

A. Competição M3.

A terceira competição, chamada M3-Competition, destinava-se a replicar e estender as funcionalidades do M-Competition e M2-Competition, através da inclusão de mais métodos e pesquisadores (particularmente pesquisadores na área de redes neurais) e mais séries temporais. Foi usado um total de 3.003 séries temporais [21].

Na tabela 1 abaixo, encontra-se o número de série temporal com base no intervalo de tempo e no domínio.

TABELA 1: DESCRIÇÃO DOS DADOS DA COMPETIÇÃO M3.

Intervalo	Domínio da Série Temporal						Total
	Micro	Indústria	Macro	Finanças	Demografia	Outros	
Anual	146	102	83	58	245	11	645
Trimestral	204	83	336	76	57	-	756
Mensal	474	334	312	145	111	52	1428
Outro	4	-	-	29	-	141	174
Total	828	519	731	308	413	204	3003

As cinco medidas utilizadas para avaliar a precisão das previsões dos diferentes métodos na competição M3 foram: SMAPE (erro percentual absoluto médio simétrico), ranking médio, melhor porcentagem, MdRAE (erro absoluto mediano relativo ao naïve sazonal) e MdSAPE (erro percentual absoluto mediano simétrico) [21].

B. Base de Dados.

Das 11 séries selecionadas, três apresentam 47 pontos de dados, quatro apresentam 72 pontos de dados e quatro 144 pontos de dados. Assim como na competição, foram calculadas previsões para 6, 8 e 18 passos à frente.

Na tabela 2 abaixo, encontra-se o número de série temporal com base no número de observações, no horizonte de previsão, no intervalo de tempo e na categoria.

TABELA 2: DESCRIÇÃO DA BASE DE DADOS UTILIZADA.

Série	Tipo de Dados de Série Temporal			
	Observações	Horizonte	Intervalo	Categoria
N 156	47	6	Anual	Indústria
N 336	47	6	Anual	Finanças
N 407	47	6	Anual	Demografia
N 903	72	8	Trimestral	Indústria
N 911	72	8	Trimestral	Indústria
N 1316	72	8	Trimestral	Finanças
N 1353	72	8	Trimestral	Demografia
N 1956	144	18	Mensal	Indústria
N 2270	144	18	Mensal	Macroeconomia
N 2493	144	18	Mensal	Macroeconomia
N 2602	144	18	Mensal	Finanças

C. Métricas de Comparação de Modelos.

De forma a validar a qualidade de ajustamento dos resultados obtidos foi calculado o SMAPE (Symmetric mean absolute percentage error) ou erro percentual absoluto médio simétrico definido como:

$$SMAPE = \frac{2}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{\hat{y}_t + y_t} \cdot 100$$

Onde representa \hat{y}_t o conjunto de valores previstos e y_t os valores reais.

IV. RESULTADOS

Efetuada as previsões com a aplicação dos 4 (quatro) métodos, os resultados obtidos foram comparados entre si com relação ao erro percentual absoluto simétrico.

A tabela 3 expõe o SMAPE de cada método para cada uma das 11 séries selecionadas.

TABELA 3: COMPARAÇÃO DO ERRO PERCENTUAL ABSOLUTO MÉDIO SIMÉTRICO (SMAPE) DOS MÉTODOS TESTADOS.

Série	Método			
	AES	B&J	NAR	PG
N 156	15,76%	7,93%	8,70%	7,28%
N 336	12,88%	13,44%	31,10%	27,48%
N 407	1,65%	1,46%	2,66%	1,46%
N 903	28,08%	39,20%	43,36%	42,19%
N 911	11,74%	13,85%	10,00%	9,00%
N1316	26,13%	31,38%	44,47%	30,38%
N1353	3,23%	4,03%	7,21%	5,28%
N1956	28,42%	23,57%	20,96%	10,07%
N2270	23,60%	20,90%	19,90%	20,62%
N2493	23,04%	26,18%	31,28%	25,70%
N2602	96,90%	92,88%	100,59%	99,90%

De acordo com os resultados expostos na tabela acima, o método de amortecimento exponencial (AES) foi o que melhor previu as séries N336, N903, N1316, N1353 e N2493. O método de programação genética apresentou menor SMAPE para as séries N407, N156, N911 e N1956. As redes NAR e o apresentaram melhor desempenho para as série N2270 e Box&Jenkins, para as séries N407 e N2602.

Todavia ao analisarmos os erros de previsão um passo à frente para cada uma das séries separadamente (figura 2), o método de alisamento exponencial obteve melhor desempenho nas séries N903, N2493 e N2602. O método de Box-Jenkins, para as séries N336, N407 e N1353.

Com relação aos métodos de inteligência computacional, a programação genética apresentou-se mais competitiva nas séries N156, N911, N1316 e N1956 e as redes neurais autorregressivas na série N2270.

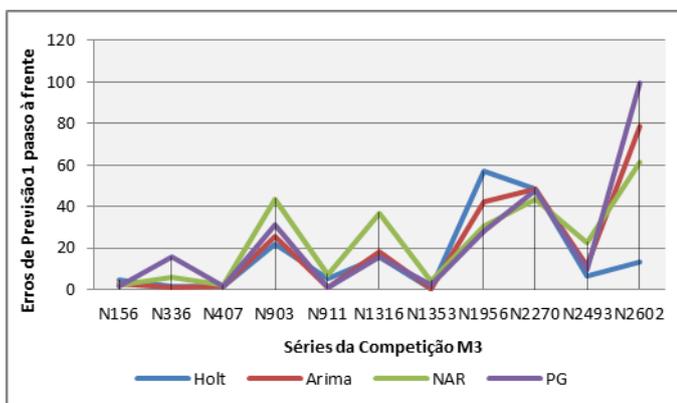


Fig. 2. Gráfico de comparação dos modelos para o horizonte de previsão 1 passo à frente.

Entretanto tão importante quanto mostrar a acurácia dos modelos com relação a métrica adotada, é demonstrar a qualidade desses métodos em relação a um modelo escolhido como ponto de referência. Para este trabalho, adotou-se um modelo ingenuo (equivalente a um SARIMA(0,0,0)*(0,1,0)) que apenas captura a sazonalidade nos dados.

Na Figura 3, abaixo encontra-se a comparação do vários métodos de previsão adotados por categoria (indústria, finanças, demografia e macroeconomia) em relação ao modelo de referência.

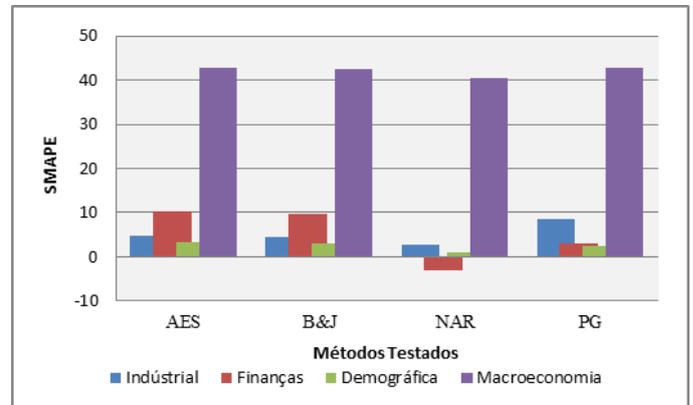


Fig. 3. Gráfico de comparação dos modelos por categoria.

Pela métrica de comparação adotada, para as séries industriais e macroeconômicas, o melhor modelo foi a programação genética com diferenças de 8,49% e 42,79% para o modelo de referência, respectivamente. Para as séries financeiras e demográficas, destacou-se, como o melhor método, aquele baseado em decomposição clássica, que se constitui em uma mescla de modelo analítico e método de previsão. O amortecimento exponencial apresentou diferenças significativas de 10,33% e 3,36% para o ingênuo.

A Figura 4, apresenta o SMAPE médio para o horizonte de previsão de 1 passo à frente e para o intervalo de 1 a 4, 1 a 6, 1 a 8, 1 a 12, 1 a 15 e 1 a 18 passos à frente. Os resultados demonstram que para o horizonte de previsão e todos os intervalos adotados, o método ingênuo obteve o pior desempenho comparado aos demais, exceto para o intervalo de 1 a 4 para o NAR. O sinal negativo implica que esse método obteve previsões inferiores ao modelo ingênuo adotado como referência.

As redes neurais do tipo não-linear autorregressivas, são as mais simples do método de redes neurais. Por analisar apenas o comportamento da estrutura autorregressiva, a rede pode não ter capturado todas as perturbações inerentes a algumas séries, como as financeiras, por exemplo.

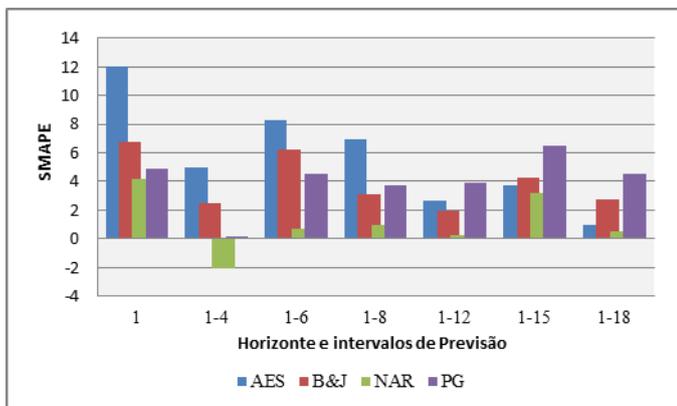


Fig. 5. Gráfico de comparação dos modelos no horizonte de 1 passo à frente e por intervalos de previsão.

O modelo de amortecimento exponencial mostrou-se em média competitivo na previsão das séries selecionadas no horizonte de 1 passo à frente e nos intervalos de médio e curto prazo (1 a 4, 1 a 6 e 1 a 8). No entanto para previsões de longo prazo, o método de programação genética apresentou maior diferença em relação ao método ingênuo.

Todos os métodos apresentam vantagens e desvantagens., estando a sua utilização condicionada aos tipos de dados históricos obtidos. Ao analisar as previsões por periodicidade (Figura 4), o modelo mais competitivo para as séries anuais foi o Box&Jenkins, com diferença de 11,73% para o modelo ingênuo. Para as séries trimestrais e mensais, destacam-se os métodos de amortecimento exponencial (10,92%) e de programação genética (23,12%).

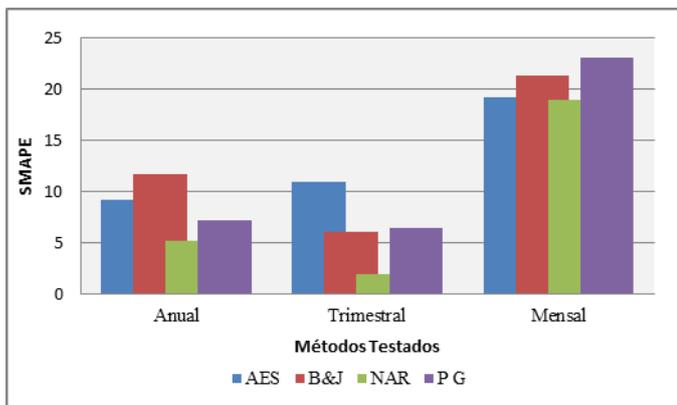


Fig. 5. Gráfico de comparação dos modelos por periodicidade.

A escolha do melhor método passa por analisar as medidas de erro de cada um. Assim é possível chegar a um modelo que reduza ao máximo a componente residual sem prejuízo da respectiva aleatoriedade.

V. CONCLUSÃO

A previsão é um valioso instrumento na tomada de decisões. Neste trabalho, são abordadas técnicas computacionais e de estatística clássica e suas particularidades no tratamento de séries temporais.

Os resultados apontam para a necessidade de aprimoramento dos métodos preditivos a fim de fornecer

resultados mais precisos. A exceção do método NAR para séries financeiras, todos os métodos apresentam diferenças significativas em comparação ao modelo ingênuo sazonal, principalmente em consideração as séries macroeconômicas.

A capacidade computacional, atualmente, permite modelar, de forma rápida, vários métodos. A partir das saídas desses modelos o decisor poderá optar por aquele que conduza a menores erros de previsão quando tiver de tomar a sua decisão.

REFERÊNCIAS

- [1] Thiesing, F.; Vornberger, O. "Sales Forecasting using neural networks". IEEE International Conference on Neural Networks, 1997, pp. 2125-2128.
- [2] Zhang, G.; Kline, D. "Quarterly time-series forecasting with neural networks". IEEE Transactions on Neural Networks, 2007, pp. 1800-1814.
- [3] Koza, John R. "Genetic Programming: On the Programming of Computers by Means of Natural Selection". Cambridge, MA: The MIT Press, 1992.
- [4] Kaboudan, M. "A Measure of Time Series Predictability Using Genetic Programming Applied to Stock Returns". Journal of Forecasting, vol. 18, 1999, pp. 345 - 357.
- [5] Ambera, K.P.; Aslamb, M.W.; Hussainc, S.K.. "Electricity consumption forecasting models for administration buildings of the UK higher education". Energy Build, vol. 90, 2015, p.p. 127-136.
- [6] Cortez, P. A. R. "Modelos Inspirados na Natureza para a Previsão de Séries Temporais". Tese de doutorado. Universidade do Minho. Portugal, 2002.
- [7] R Development Core Team. Writing R Extensions. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>, 2011.
- [8] Searson, D. P.. "Genetic Programming & Symbolic Regression for MATLAB". <http://gptips.sourceforge.net>, 2009.
- [9] Makridakis, S.; Wheelwright, S.C.; McGee, V.E.. "Forecasting: Methods and Applications", 2nd Ed. New York: John Wiley and Sons, 1983.
- [10] Hyndman, R. J., Koehler, A. B., Snzder, R. D., & Grose, S. (2002). "A state space framework for automatic forecasting using exponential smoothing methods". International Journal of Forecasting, vol. 18, 2002, pp. 439 - 454.
- [11] Brown, R. G. (1959). "Statistical forecasting for inventor control". New York: McGraw-Hill, 1959.
- [12] Holt, C. C. (1957). "Forecasting Trends and Seasonal by Exponential Weighted Averages". ONR Memorandum No. 52, Carnegie Institute of Technologz, Pittsburgh, USA (published in International Journal of Forecasting, vol. 20, 2004, pp. 5-13).
- [13] Winters, P. R. "Forecasting sales by exponential weighted moving averages. Management Science, vol. 6, 1960, pp.324 - 342.
- [14] Box, G.E.P.; Jenkins, G.M., "Time Series Analysis: Forecasting and Control", Revised Edition, San Francisco: Holden-Daz, 1976.
- [15] Koza, John R. "Genetic Programming: On the Programming of Computers by Means of Natural Selection". Cambridge, MA: The MIT Press, 1992.
- [16] Poli, R., Langdon, W. B., and McPhee, N. F. "A field guide to genetic programming". <http://lulu.com>, 2008.
- [17] Searson, D. P., Willis, M. J., and Montague, G. A. (2007). "Co-evolution of non-linear pls model components". Journal of Chemometrics, 2007, pp. 592-603.
- [18] Hazkin, S. "Neural networks: a comprehensive foundation". Prentice Hall, 1999.
- [19] FRANCES, P. H. "Time Series models for business and economic forecasting.". Cambridge, 1998.

- [20] J.T. Connor; R.D. Martin; L.E. Atlas. "Recurrent neural networks and robust time series prediction". IEEE Transactions on Neural Networks, 5 (2), 1994, pp. 240–254.
- [21] Makridakis, Spzros; Hibon, Michele. "The M-3 Competition: results, conclusions, and implications". International Journal of Forecasting, vol. 16, 2000, pp. 451–476.