

Previsões para a série de tempo *Canadian Lynx* usando método que combina redes neurais artificiais, encolhimento e decomposição *wavelet*

Forecasts for the Canadian Lynx time series using method that combine neural networks, wavelet shrinkage and decomposition

Levi Lopes Teixeira Correio¹ - Universidade Tecnológica Federal do Paraná, Departamento de Matemática e Estatística
Paulo Henrique Siqueira² - Universidade Federal do Paraná, Departamento de Expressões Gráficas
Luiz Albino Teixeira Jr³ - Universidade Federal da Integração Latino Americana, Departamento de Estatística
Samuel Bellido Rodrigues⁴ - Universidade Tecnológica Federal do Paraná, Departamento de Matemática e Estatística
Arinei Carlos Lindbeck da Silva⁵ - Universidade Federal do Paraná, Departamento de Matemática

RESUMO A previsão de séries temporais é largamente utilizada nas diversas áreas do conhecimento humano, principalmente no planejamento e direcionamento estratégico das empresas. O sucesso desta tarefa depende das técnicas de previsões aplicadas. Neste artigo, é proposta uma metodologia híbrida para se projetar séries temporais. Para a validação da metodologia foi escolhida uma série de tempo já modelada por outros autores, possibilitando a comparação dos resultados. A metodologia proposta integra as seguintes técnicas: encolhimento wavelet, decomposição wavelet de nível r e redes neurais artificiais (RNAs). Primeiramente, uma série temporal a ser prevista é submetida ao método de filtragem wavelet proposto, o qual a decompõe em componentes de tendência e de resíduo linear. Em seguida, ambas são decompostas via decomposição de nível r , gerando, para cada uma, $r+1$ componentes wavelet (CWs); e, em seguida, cada CW é individualmente modelada por uma RNA. Finalmente, as previsões para todas as CWs são linearmente combinadas, produzindo as previsões para a série temporal supracitada. Para avaliá-lo, a série temporal de *Canadian Lynx* foi usada e todos os resultados alcançados pelo método proposto foram melhores do que outros existentes na literatura.

Palavras-chave Encolhimento Wavelet. Decomposição Wavelet. Redes neurais artificiais.

ABSTRACT *Time series forecasting is widely used in various areas of human knowledge, especially in the planning and strategic direction of companies. The success of this task depends on the forecasting techniques applied. In this paper, a hybrid approach to project time series is suggested. To validate the methodology, a time series already modeled by other authors was chosen, allowing the comparison of results. The proposed methodology includes the following techniques: wavelet shrinkage, wavelet decomposition at level r , and artificial neural networks (ANN). Firstly, a time series to be forecasted is submitted to the proposed wavelet filtering method, which decomposes it to components of trend and linear residue. Then, both are decomposed via level r wavelet decomposition, generating $r + 1$ Wavelet Components (WCs) for each one; and then each WC is individually modeled by an ANN. Finally, the predictions for all WCs are linearly combined, producing forecasts to the underlying time series. For evaluating purposes, the time series of *Canadian Lynx* has been used, and all results achieved by the proposed method were better than others in existing literature.*

Keywords *Wavelet Shrinkage. Wavelet Decomposition. Artificial neural networks.*

1. Av. Brasil, 4232, 85884-000, Medianeira - PR, prof.levilopes@gmail.com
2. paulohs@ufpr.br
3. luiz.a.t.junior@gmail.com
4. prof.samuel@gmail.com
5. arineids@gmail.com

CORREIO, L. L. T.; SIQUEIRA, P. H.; TEIXEIRA JR, L. A.; RODRIGUES, S. B.; SILVA, A. C. L. Previsões para a série de tempo *Canadian Lynx* usando método que combina redes neurais artificiais, encolhimento e decomposição wavelet. **GEPROS. Gestão da Produção, Operações e Sistemas**, Bauru, Ano 10, nº 4, out-dez/2015, p. 157-172.

DOI: 10.15675/gepros.v10i4.1249

1. INTRODUÇÃO

Desde 1927, quando George Udny Yule propôs interpretar uma sequência de dados indexado pelo tempo (isto é, uma série temporal) como uma realização de uma coleção de variáveis aleatórias probabilisticamente distribuídas (chamada de processo estocástico) e, em casos de existência de auto-correlação e estacionariedade, projetá-la através de modelos auto-regressivos (ARs), muitos foram os avanços para se alcançar maior acurácia nas previsões pontuais e intervalares.

Box e Jenkins (1970) apresentaram uma classe de modelos estatísticos clássicos que consiste, basicamente, em se ajustar um modelo auto-regressivo integrado de médias móveis (ARIMA) a uma série temporal que apresente auto-correlação e estacionariedade (de segunda ordem), a partir de três estágios: (1) identificação das ordens dos modelos iniciais plausíveis via análise de perfil das funções de auto-correlação parcial e simples; (2) estimação dos parâmetros; e (3) verificação das estatísticas de diagnósticos. Além desta metodologia, a utilização dos modelos de Redes Neurais Artificiais (RNAs) é vastamente difundida. Wang *et al.* (2006) e Abreu *et al.* (2012) são apenas alguns, de diversos outros exemplos, existentes na literatura. Revisando 25 anos (1985-2005) de pesquisa sobre a aplicabilidade das RNA na predição de séries temporais, De Gooijer e Hyndman (2006) concluíram que importantes progressos foram alcançados e salientam que, na maioria dos estudos, as RNAs obtiveram maior acurácia preditiva. Teixeira Jr (2013), destaca que as RNAs possuem uma vantagem sobre os modelos ARIMA, que é a de serem aplicáveis às séries temporais que exibem estruturas de auto-dependência linear e não-linear, bem como àquelas que não apresentam estacionariedade.

Por sua vez, da Teoria *Wavelet*, advêm importantes métodos de pré-processamento de dados que fazem o seu tratamento antes de sua efetiva modelagem. Comumente, os referidos consistem em decompô-los (DONOHO; JOHNSTONE, 1994), ou filtrá-los (DONOHO *et al.*, 1995). A decomposição *wavelet* de nível r (que é a mais comum), em linhas gerais, consiste em expandir uma série temporal em termos de sua componente *wavelet* (CW) de aproximação e de suas r CWs de detalhes, resultando $r+1$ CWs, com níveis de resolução fixos e distintos, (TEIXEIRA JR, 2013). A filtragem *wavelet* (*wavelet shrinkage*), por outro lado, objetiva a extrair perturbações (ou ruídos, ou inovações) - os quais não possuem estrutura de auto-dependência -, que podem afetar, de forma negativa, o processo de geração de previsões. Baseado em Lei e Ran (2008) e Teixeira Jr *et al.* (2012), bem como em diversas outras referências, verifica-se que ganhos preditivos substanciais podem ser alcançados, quando o tratamento *wavelet* é implementado adequadamente.

No âmbito empresarial, planejamentos estratégicos carecem de projeções de variáveis significativas nas tomadas de decisões. A previsão de séries de tempo pode auxiliar no planejamento energético, análises de índices da bolsa de valores, taxas de câmbios, entre outras áreas. Fazer a previsão de demandas, por exemplo, é de grande importância para um eficiente planejamento operacional e estratégico de organizações de diversos setores, reduzindo riscos inerentes ao processo decisório das atividades empresariais.

Diante disso, propõem-se uma metodologia híbrida que se apresenta como opção de ferramenta auxiliadora dos tomadores de decisão. A metodologia integra a filtragem *wavelet*, a decomposição *wavelet* de nível r e nos modelos de RNAs para projetar séries temporais. Em linhas gerais, pode-se dividi-la, nos seguintes passos: (1) tomar uma série temporal a ser prevista e filtrá-la via

abordagem *wavelet*; (2) fazer a decomposição *wavelet* de nível r da série de tempo filtrada em (1), gerando-se suas $r+1$ CWs; (3) modelar individualmente cada CW, do passo 2, via RNAs; e (4) a previsão da série temporal considerada em (1) é obtida por meio da soma das previsões de cada CW obtidas no passo 3. A fim de avaliar a qualidade de suas previsões, foi utilizada a série temporal *Canadian Lynx* e os resultados auferidos comparados com os dos seguintes autores: Aladag *et al.* (2009), Adhikari; Agrawal (2013), Ismail; Shabri (2014), Kajitani *et al.* (2005), Karnaboopathy; Venkatesan (2012), Karnaboopathy; Venkatesan (2012), Khashei; Bijari (2011), Khashei; Bijari (2012), Zhang (2003) e Zheng; Zhong (2011).

2. ANÁLISE WAVELET

A análise *wavelet* propicia a decomposição de uma série temporal em componentes, de forma a facilitar o estudo da mesma. Embora a análise de Fourier seja uma das formas mais tradicionais para o tratamento de sinais, as transformadas *wavelets* são mais completas, fornecendo uma melhor descrição do sinal, (KARTHIKEYAN; NAGESH KUMAR, 2013). Em Tiwari e Chatterjee (2010), Kisi (2010), Kisi e Cimen (2011) e Seo *et al.* (2015), utilizaram decomposição *wavelet* em seus trabalhos de previsão de séries temporais e concluíram o bom desempenho desta técnica.

2.1. Decomposição *Wavelet*

Na Análise de Fourier, uma série temporal (unidimensional) de cardinalidade igual a T , denotada por $f(\cdot) := (f(1), f(2), \dots, f(T))$, é decomposta ortogonalmente, em termos de senos e cossenos definidos em diversas bandas de frequências (MORETTIN; TOLOI, 2006); enquanto que, na Análise *Wavelet*, $f(\cdot)$ é, de forma ortogonal, decomposta, simultaneamente, a partir de informações de escala e de translação capturadas a partir de um conjunto $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z}^2}$ de funções *wavelet*, onde $\omega_{m,n}(\cdot) := 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$ é gerado por meio da dilatação (associada ao parâmetro m) e translação (associada ao parâmetro n) de uma função *wavelet* $\omega(\cdot)$ original (LEVAN; KUBRUSLY, 2003 e TEIXEIRA JR *et al.* 2013). Segundo Mallat (2008), o parâmetro m é chamado de parâmetro de escala diádica e n , de parâmetro de translação unitária.

Em Teixeira Jr *et al.* (2013), pode-se verificar, empiricamente, que as funções *wavelet* $\omega_{m,n}(\cdot)$ estão associadas às componente *wavelet* de detalhe (isto é, as de alta frequência) de $f(\cdot)$; e, teórico e empiricamente, que a componente *wavelet* de aproximação (isto é, a de baixa frequência) de $f(\cdot)$ é gerada de uma combinação linear de funções escala em $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}}$, onde: cada mapa $\phi_{m_0,n}(\cdot) := 2^{\frac{m_0}{2}} \phi(2^{m_0}(\cdot) - n)$, sendo que o parâmetro m_0 toma um valor inteiro fixo, é gerado por meio da translação unitária n de uma função escala original $\phi(\cdot)$, (MALLAT, 2008). De acordo com Daubechies (1988), por definição, as funções *wavelet* e as funções escala consistem em ondas de curta duração, geralmente assimétricas, que permitem a sua dilatação diádica m (positiva ou negativa) e translação unitária n (positiva ou negativa) de seu domínio. Não obstante, definições mais formais e gerais, do ponto de vista matemático, podem ser verificadas em Teixeira Jr *et al.* (2013). As famílias de bases ortonormais *wavelet*, em geral, são geradas por convoluções de $\omega(\cdot)$ e $\phi(\cdot)$. Em particular, as bases de Haar, Daubechies, Coiflet e Symelet são exemplos de famílias que, frequentemente, são utilizadas na modelagem de séries temporais, para fins de geração de previsões.

Baseado em Kubrusly e Levan (2006) e Mallat (2008), tem-se que a decomposição *wavelet* de uma série temporal $f(\cdot)$ é dada, genericamente, em (1).

Equação 1 – Decomposição *wavelet* de uma série temporal.

$$f(\cdot) = f_{A_{m_0}}(\cdot) + \sum_{m=m_0}^{\infty} f_{D_m}(\cdot) \quad (1)$$

Onde: $f_{A_{m_0}}(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot)$ é a componente de aproximação de nível m_0 de $f(\cdot)$; $f_{D_m}(\cdot) = \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot)$ é a componente de detalhe de nível m de $f(\cdot)$; $d_{m,n}$ e $a_{m_0,n}$ são, respectivamente, os coeficientes de detalhe e de aproximação e são calculados, ainda nesta ordem, a partir do produto interno usual entre $f(\cdot)$ e $\omega_{m,n}(\cdot)$ (isto é, $d_{m,n} = \sum_{t=1}^T f(t) \omega_{m,n}(t)$) e entre $f(\cdot)$ e $\phi_{m_0,n}(\cdot)$ (isto é, $a_{m_0,n} = \sum_{t=1}^T f(t) \phi_{m_0,n}(t)$), para todo $m_0, m, n \in \mathbb{Z}$.

2.2. Encolhimento *Wavelet*

Se $f(\cdot) = (f(1), f(2), \dots, f(T))$ for fortemente ruidosa, é recomendada a sua filtragem estatística via algoritmos de encolhimento *wavelet*. Note que $f(\cdot)$ é tal que, para cada instante t , pode ser teoricamente decomposta como: $f(t) = f(t) + e_t$, onde: (i) e_t é um estado (ou realização) de uma variável aleatória (independente e identicamente distribuída (iid), (LIU, 2006), com média zero e variância constante, denotada por $e_t \sim D(0, \sigma^2)$, onde D é uma distribuição arbitrária; e (ii) $f(t)$ é a componente determinística (que, de fato, é passível de ser prevista pontualmente) de $f(t)$. Devido a ausência de estrutura de auto-dependência na sequência (e_1, e_2, \dots, e_T) de perturbações (ou inovações) da série temporal $f(\cdot)$, não se pode modelá-la. Baseado nisso, o objetivo dos algoritmos de encolhimento (estatístico) *wavelet* é o que transformar (isto é, filtrar) $f(\cdot)$ ao passo de deixá-la, o mais próximo possível, da sequência teórica $\tilde{f}(\cdot) = (\tilde{f}(1), \tilde{f}(2), \dots, \tilde{f}(T))$ de componentes determinísticas, (MORETTIN, 1999; DONOHO; JOHNSTONE, 1994 e DONOHO *et al.*, 1995).

Basicamente, as três etapas do encolhimento *wavelet* são: (1) decomposição *wavelet* de nível r de $f(\cdot)$; (2) selecionar limiares (*threshold*) que reduzem os ruídos em $f(\cdot)$; (3) aplicar a transformada *wavelet* inversa, definida em (1), para se obter uma série temporal menos ruidosa (o mais próximo possível de $\tilde{f}(\cdot)$). Das várias regras de limiarização encontradas na literatura, talvez as mais tradicionais são: o limiar duro (*hard threshold*) e limiar suave (*soft threshold*), representadas em (2) e (3), respectivamente.

Equação 2 – Limiar duro.

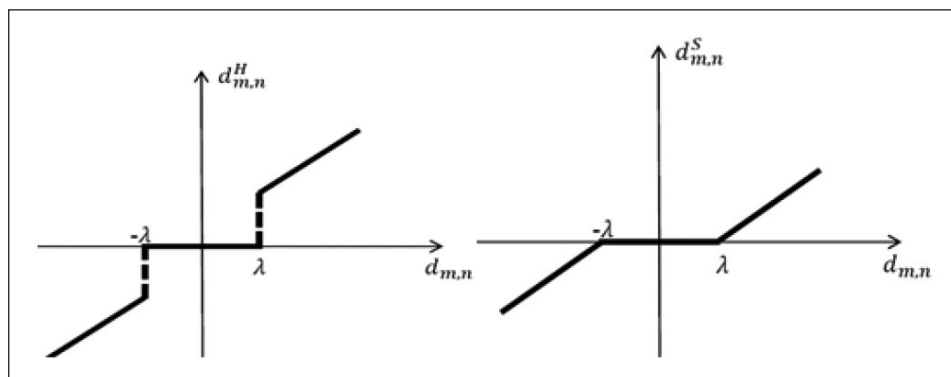
$$d_{m,n}^H = \begin{cases} d_{m,n}, & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases} \quad (2)$$

Equação 3 – Limiar suave.

$$d_{m,n}^s = \begin{cases} \text{sign}(d_{m,n})(|d_{m,n}| - \lambda), & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases} \quad (3)$$

Onde: $d_{m,n}$ é o coeficiente de detalhe dotado de ruído; $d_{m,n}^H$ e $d_{m,n}^S$ são coeficientes de detalhes filtrados; $\text{sign}(d_{m,n})$ denota o sinal transformado de $d_{m,n}$; e λ o limiar. O *hard threshold* gera uma descontinuidade, ou seja, um “salto” em λ , eliminando ou mantendo o coeficiente de detalhe. Enquanto o limiar suave atenua por $(|d_{m,n}| - \lambda)$ os valores de $d_{m,n}$, se $|d_{m,n}| > \lambda$; e, caso contrário, os anula. Nos gráficos da Figura 1 podem ser observadas tais características.

Figura 1 – Gráficos das funções limiar duro e suave.



Fonte: Elaborado pelos autores (2015).

Neste artigo, porém, o cálculo do limiar λ foi realizado a partir de duas abordagens: o universal e o SURE (*Stein's Unbiased Risk Estimate*).

De acordo com Donoho e Johnstone (1998), o cálculo do limiar universal é dado por $\lambda = \sigma \sqrt{2 \log(T)}$, onde: σ o desvio-padrão da série de perturbações; e T a cardinalidade de $f(\cdot)$. Note que o limiar universal independe do parâmetro de escala m , assumindo, portanto, um valor global. Por sua vez Donoho e Johnstone (1995), propuseram selecionar um limiar, pela minimização do estimador não-viesado do risco Stein (SURE), para cada escala m . Sendo N_m o número de coeficientes para cada valor de escala m , segue que o limiar de SURE é definido conforme em (4).

Equação 4 – Limiar de SURE.

$$\lambda_m = \operatorname{argmin}_{0 \leq \lambda \leq \sqrt{2 \log(Nm)}} \operatorname{SURE}(y_m, \lambda) \quad (4)$$

Onde: λ_m é o limiar associado ao nível m ; argmin significa “argumento que minimiza”; e $\operatorname{SURE}(y_m, \lambda)$ é uma função de dois argumentos cuja definição pode ser visualizada em Morettin (1999).

3. REDES NEURAS ARTIFICIAIS

As Redes Neurais Artificiais (RNAs) foram idealizadas com base nos processos de aprendizagem do cérebro humano, de forma que a ser capaz de simular as conexões sinápticas. De acordo com Haykin (2001), uma RNA consiste em um sistema de processamento paralelo formado por unidades capazes de armazenar conhecimento e disponibilizá-lo para o futuro, sendo formada basicamente por neurônios artificiais distribuídos em camadas. Um dos tipos mais usados é a *multilayer feedforward* que é composta por várias camadas e os sinais são propagados apenas da entrada para a saída, nunca ao contrário.

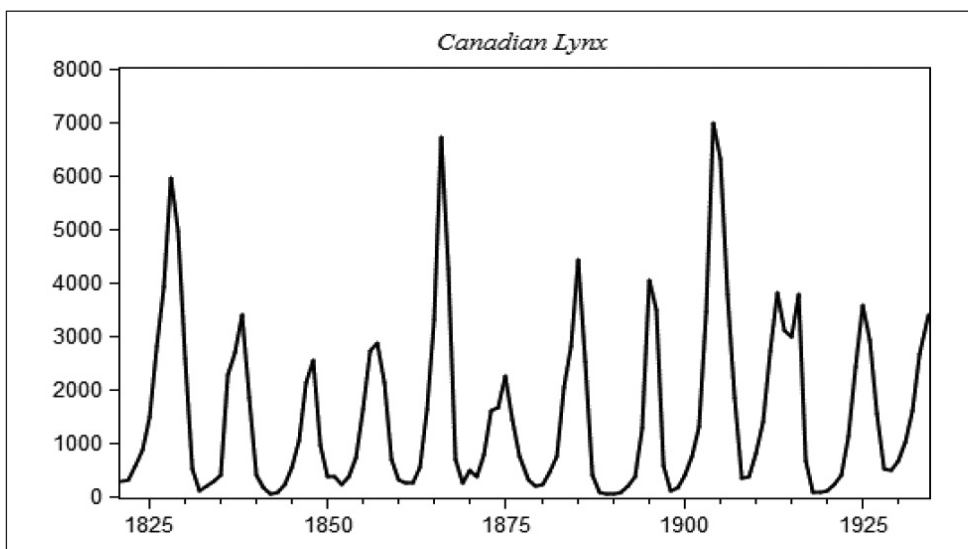
O principal algoritmo de treinamento de redes neurais artificiais é o *backpropagation*, cujo ajuste dos pesos sinápticos ocorre por meio de um processo de otimização realizado em duas fases: *forward* e *backward*. Na fase *forward*, é calculada a resposta fornecida pela RNA para determinado padrão de entrada. Na fase *backward*, o desvio (ou erro) entre a resposta da RNA e a resposta desejada é utilizado no processo de ajuste dos pesos sinápticos, (HAYKIN, 2001). A previsão de uma série temporal $f(\cdot)$, com T observações, através de RNA consiste na utilização de dados observados para se prever o estado $f(t + k)$, onde: $t + k$ é o instante de tempo futuro. Uma vez definido o tamanho da janela (n) e o valor de k , para se prever $f(t + k)$, é inserida, na RNA, a sequência $\{f(t + k - n), \dots, f(t + k - 2), f(t + k - 1)\}$ de sinais temporais resultando na previsão $\hat{f}(t + k)$, (TEIXEIRA JR *et al.*, 2012).

4. MATERIAIS E MÉTODOS

4.1. Materiais

A série de dados *Canadian Lynx* (lincas canadenses) refere-se ao registro anual do número de lincas canadenses (que é uma espécie de felino natural da América do Norte) capturados no distrito de Rio Mackenzie no Norte do Canadá, a qual pode ser obtida no seguinte endereço eletrônico: <https://datamarket.com/>. Existem na literatura vários trabalhos que fazem análise e previsão da série de tempo *Canadian Lynx*, credita-se a Moran (1953), como sendo o primeiro trabalho de análise desta. Pode-se encontrar metodologias que a modelam nas seguintes referências: Subba e Sabr (1984), Tang e Ghosal (2007), Zhang (2003), Wong e Li (2000), entre outros. O gráfico da série *Canadian Lynx* (com 114 observações no período: 1821-1934) é graficamente representado na Figura 2, onde se pode observar uma periodicidade de aproximadamente 10 anos, (STONE; HE 2007).

Figura 2 – Gráfico da série temporal de linces canadenses no período 1821-1934.



Fonte: Elaborado pelos autores (2015).

É comum entre os autores que publicaram trabalhos de previsão da série dos linces canadenses, Aladag *et al.* (2009), Ismail; Shabri (2014), Zhang (2003) entre outros, aplicarem logaritmo de base dez ao conjunto de dados originais, dividindo o conjunto obtido em amostras de treinamento (100 primeiros anos: 1821-1920) e teste (últimos 14 anos), com previsões um passo à frente para o horizonte de quatorze anos. Por conta disso, o presente artigo utiliza a mesma abordagem.

4.2. Método proposto

Tome série temporal, $f(\cdot) = (f(1), f(2), \dots, f(T))$ de cardinalidade T a ser prevista h passos à frente e assuma que $f(t) = \hat{f}(t) + e_t$, onde $e(t)$ é uma realização de uma variável aleatória linearmente independente (isto é temporalmente descorrelatada) e identicamente distribuída $\varepsilon_t \sim D(0, \sigma^2)$, onde D é uma distribuição arbitrária – conforme a Seção 2.2. Assim, inicialmente, realiza-se a filtragem de $f(\cdot)$ por meio de um algoritmo de encolhimento *wavelet*, obtendo-se a sua versão estatisticamente filtrada, denotada por $\tilde{f}(\cdot)$ e chamada de componente de tendência, e uma sequência de ruídos lineares (e_1, e_2, \dots, e_T) – ou seja, possuem média zero, variância constante e não são dotados de estruturas de auto-dependência linear –, onde: $e_t = f(t) - \tilde{f}(t)$, para todo t . Para garantir tais condições, análises nos intervalos de confiança das FAC e FACPs até os lags 36 (que equivale a três anos), de estacionariedade Dickey-Fuller e o teste t foram aplicados sobre (e_1, e_2, \dots, e_T) . Note que, a rejeição da hipótese de existência de auto-dependência linear nos resíduos, a um nível α de significância, não garante inexistência de estruturas de auto-dependência não-linear. Desse modo, a previsão $\hat{f}(t)$ para o estado $f(t)$ é definida por: $f(t) = \hat{f}(t) + \hat{e}_t$, para todo t , onde: $\hat{f}(t)$ é a previsão do estado $f(t)$; e \hat{e}_t , do estado e_t . Denotando, genericamente, $\tilde{y}_{t,s}$ para denotar $f(t)$ ou e_t , dependendo do contexto, segue que, baseado em Teixeira Jr *et al.* (2012), e na Seção 2.1, tem-se que a decomposição *wavelet* de nível r da série temporal $\tilde{y}_{t,s}$ é definida em (5).

Equação 5 – Decomposição *wavelet* de nível r da série temporal $\tilde{y}_{t,S}$.

$$\tilde{y}_{t,S} \cong \tilde{\tilde{y}}_{t,S} = \sum_{n \in \Gamma_{V_{m_0}}} a_{m_0,n} \phi_{m_0,n}(t) + \sum_{m=m_0}^{m_0+(r-1)} \left(\sum_{n \in \Gamma_{W_m}} d_{m,n} \omega_{m,n}(t) \right) \quad (5)$$

Onde: $\Gamma_{V_{m_0}}$ e Γ_{W_m} são subconjuntos de índices em Z . A determinação de tais conjuntos é feita através da minimização do erro quadrático médio (MSE) entre $\tilde{y}_{t,S}$ e $\tilde{\tilde{y}}_{t,S}$. A filtragem *wavelet* e o cálculo das componentes *wavelet* de aproximação $\sum_{n \in \Gamma_{V_{m_0}}} a_{m_0,n} \phi_{m_0,n}(t)$ e de detalhe $\sum_{n \in \Gamma_{W_m}} d_{m,n} \omega_{m,n}(t)$, provenientes da decomposição das séries temporais $f(\cdot)$ e de resíduo linear (e_1, e_2, \dots, e_T) , foram feitos no software Matlab 8.0 – o qual dispõe de famílias de bases ortonormais *wavelet*, como, por exemplo: daubechies (db), coiflets (coif) e symlets (sym), (DAUBECHIES, 1992). Uma vez obtidas as séries temporais de tendência $f(\cdot)$ e de resíduos (e_1, e_2, \dots, e_T) , ambas foram individualmente decompostas via *wavelet* de nível 2, resultando, para cada uma, em duas componentes *wavelet* de detalhe, genericamente denotadas por $D_{1,S}$ e $D_{2,S}$ e uma componente *wavelet* de aproximação, genericamente denotada por $A_{2,S}$.

Em seguida, suas componentes *wavelet* foram modeladas separadamente via redes neurais artificiais *feedforward multilayer perceptron* (RNA-MLP), produzindo, para cada t , previsões para as componentes *wavelet* de detalhe, denotadas por $\hat{D}_{1,S}$ e $\hat{D}_{2,S}$ e para a de aproximação, denotada por $\hat{A}_{2,S}$. Assim, a construção da previsão final para a série temporal original $f(\cdot)$, é dada pela seguinte combinação de previsões.

Equação 6 – Previsão final para a série temporal original $f(\cdot)$.

$$f(t) = \hat{A}_{2,EW}(t) + \hat{D}_{1,EW}(t) + \hat{D}_{2,EW}(t) + \hat{A}_{2,R}(t) + \hat{D}_{1,R}(t) + \hat{D}_{2,R}(t), t = 1, \dots, T + h \quad (6)$$

Onde, em t : $\hat{A}_{2,EW}(t)$ e $\hat{A}_{2,R}(t)$, $\hat{D}_{2,EW}(t)$ e $\hat{D}_{2,R}(t)$, e $\hat{D}_{1,EW}(t)$ e $\hat{D}_{1,R}(t)$ consistem, respectivamente, nas previsões das CWs de aproximação e de detalhes de nível 2 e das de nível 1, respectivamente, das componentes de nível $\hat{f}(t)$ e de ruído linear (e_1, e_2, \dots, e_T) .

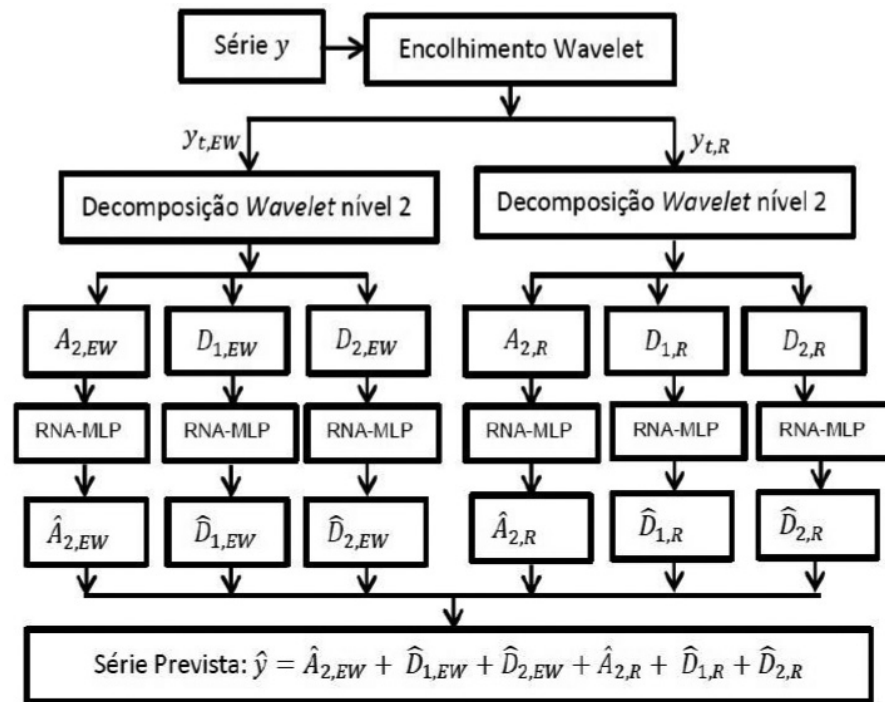
Em (6), tem-se que – dado $f(t) = y_t = \tilde{f}(t) + e_t$ e que $\hat{f}(t)$ e \hat{e}_t denotam as previsões das componentes de tendência $\tilde{f}(t)$ de ruído linear e_t –, segue que: $\hat{f}(t) = \hat{A}_{2,EW}(t) + \hat{D}_{2,EW}(t) + \hat{D}_{1,EW}(t)$ e $\hat{e}_t = \hat{A}_{2,R}(t) + \hat{D}_{2,R}(t) + \hat{D}_{1,R}(t)$, para cada instante t . A Figura 3 apresenta, de forma resumida, um fluxograma com as todas as etapas da metodologia proposta.

Como estatísticas de desempenho, foram utilizadas as estatísticas de aderência MSE (*mean square error*) e MAE (*mean absolute error*), definidos em (7). Note que $y_t = f(t)$, para fins de notação.

Equação 7 – Estatísticas de aderência MSE e MAE.

$$MSE = \frac{1}{T} \sum_{t=t_0}^T (y_t - \hat{y}_t)^2 \text{ e } MAE = \frac{1}{T} \sum_{t=t_0}^T |y_t - \hat{y}_t| \quad (7)$$

Figura 3 – Fluxograma com as etapas do método proposto no experimento computacional.



Fonte: Elaborado pelos autores (2015).

5. RESULTADOS NUMÉRICOS

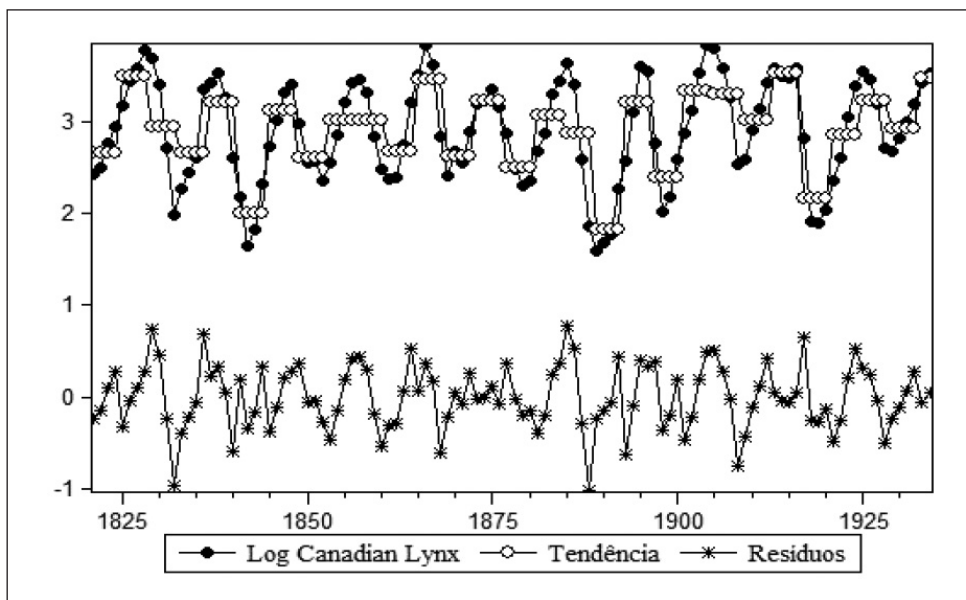
Para o encolhimento *wavelet* foram usadas as bases de Daubechies (db), Symlet (sym) e Coiflet (coif), com regra de limiarização suave e limiar universal e SURE. As séries temporais de tendência e de resíduos lineares oriundas do algoritmo *wavelet* de filtragem foram decompostas via expansão *wavelet*, com base Daubechies de momento nulo igual a 10 (db 10), com nível 2. Uma vez obtidas as componentes *wavelet*, estas foram modeladas separadamente, via rede neural artificial *feedforward multilayer perceptron* (RNA-MLP). Os testes exaustivos realizados indicaram que, de maneira geral, apresentariam resultados satisfatórios a RNA-MLP com a seguinte estrutura topológica: tamanho de janela de entrada igual a 14; 1 camada escondida, com quatorze neurônios e função de ativação tangente hiperbólica; e 1 camada de saída, com um neurônio e de ativação linear (ZHANG, 2003). Os quatro melhores resultados para a amostra de teste estão expostos na Tabela 1, onde se encontram as bases ortonormais *wavelet*, os níveis de decomposição utilizados no encolhimento e os valores de MSE e MAE obtidos na previsão da amostra de teste.

Tabela 1 – Configurações no encolhimento da série temporal *Canadian Lynx* e erros de previsões da amostra de teste.

Modelagem	Base	Nível p	Amostra de Teste	
			MSE	MAE
1	db 1	2	0,00016	0,0091
2	db 2	3	0,00027	0,0140
3	coif 1	2	0,00026	0,0124
4	sym 2	2	0,00023	0,0123

Fonte: Elaborado pelos autores (2015).

Para efeitos de comparação foi adotada a modelagem 1, com MSE e MAE iguais a 0,00016 e 0,0091, respectivamente. Na Figura 4 estão representadas as séries de tempo, onde, em t , tem-se: logaritmo de *Canadian Lynx*, denotada por y_t ou $f(t)$; a tendência, denotada por $y_{t,EW}$ ou $\hat{f}(t)$; e resíduo linear, representado por $y_{t,R}$ ou e_t . O formato de onda apresentada pela série tendência é consequência da base de Haar (db 1), usada no encolhimento, cuja o gráfico da função *wavelet* é uma onda quadrada.

Figura 4 – Gráficos das séries log *Canadian Lynx*, tendência e resíduos.

Fonte: Elaborado pelos autores (2015).

Na Tabela 2, têm-se os dez últimos pontos na amostra de teste do logaritmo, no base 10, da série temporal de *Canadian Lynx*, suas componentes de tendência e de resíduos, como também as suas respectivas previsões. Assim o sendo, no instante t , tem-se que: y_t denota o log da observação de *Canadian Lynx*; $y_{t,EW}$ e $y_{t,R}$ representam, respectivamente, as componentes de tendência e de resíduos lineares obtidos via procedimento de encolhimento *wavelet*; $\hat{y}_{t,EW}$ e $\hat{y}_{t,R}$ são, respectivamente, as previsões pontuais dos estados $\hat{y}_{t,EW}$ e $\hat{y}_{t,R}$, onde ambas foram obtidas por meio de duas RNAs distintas; e \hat{y}_t , a projeção pontual do logaritmo, na base 10, da série temporal subjacente.

Tabela 2 – Previsões para as 10 últimas observações das amostras de teste.

Ano	y_t	$y_{t,EW}$	$y_{t,R}$	$\hat{y}_{t,EW}$	$\hat{y}_{t,R}$	\hat{y}_t
1925	3,552155	3,232723	0,320431	3,222249	0,337599	3,559849
1926	3,467608	3,232723	0,234885	3,242404	0,214561	3,456966
1927	3,186674	3,232723	-0,046050	3,248391	-0,065520	3,182871
1928	2,723456	3,232723	-0,509270	3,244703	-0,521390	2,723317
1929	2,685742	2,926999	-0,241260	2,917521	-0,248040	2,669486
1930	2,820858	2,926999	-0,106140	2,916774	-0,096990	2,819781
1931	3,000000	2,926999	0,073000	2,890266	0,093990	2,984254
1932	3,201397	2,926999	0,274390	2,898945	0,305235	3,204181
1933	3,424392	3,477679	-0,053290	3,470920	-0,048300	3,422621
1934	3,530968	3,477679	0,053290	3,475408	0,065862	3,541271

Fonte: Elaborado pelos autores (2015).

Como mencionado anteriormente, os resultados obtidos pelo método proposto foram comparados, na Tabela 3, com as seguintes abordagens: (I) Zhang (2003), que comparou resultados obtidos com os modelos ARIMA e RNA com o híbrido ARIMA-RNA; (II) Kajitani *et al.* (2005), que utilizaram na modelagem da *Canadian Lynx* via uma RNA *Feed-Forward* (FFNN – *Feed-Forward Neuron Networks*) e modelo SETAR (*Self-Exciting Threshold Auto-Regression*); (III) Aladag *et al.* (2009), que realizaram a combinação de RNAs recorrentes Elman e um modelo ARIMA; (IV) Zheng; Zhong (2011), que apresentaram uma metodologia que combina a Função de Base Radial (RBF – *Radial Basis Function*) de uma rede neural artificial baseado em alisamento binomial (BS-RBF) e uma abordagem que combina a função de base radial (RBF) de uma rede neural com modelo auto-regressivo (AR), utilizando a técnica alisamento binomial (BS), denominado BS-RBFAR; (V) Khashei e Bijari (2011), os quais propuseram um modelo híbrido RNA/ARIMA e o compararam com a modelagem ANN (p,d,q), (VI) Khashei e Bijari (2012), que sugeriram metodologias híbridas com base em um classificador (redes neurais probabilísticas (PNNS)), gerando os previsores ARIMA/PNN e ANN/PNN, (VII) Karnaboopathy e Venkatesan (2012), que utilizaram em

sua modelagem o modelo *Full Range Auto-regressive* (FRAR); (VIII) Adhikari e Agrawal (2013), os quais propuseram uma metodologia combinando vários algoritmos de treinamento denominado (*Ensemble of Multiple ANN Training Methods*) e a comparou com as modelagens ARIMA e Máquinas de Vetores Suporte (SVM) para Regressão; e (IX) Ismail; Shabri (2014), que usaram Máquinas de Vetores Suporte (SVM) para Regressão e SVM com mínimos quadrados (LSSMV).

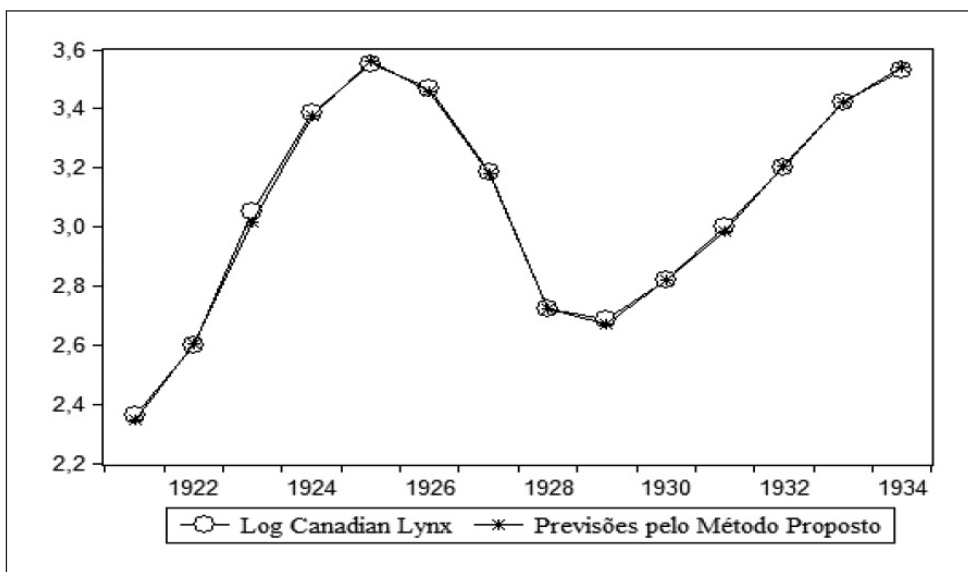
Tabela 3 – Comparativo entre métodos de previsão aplicados à série temporal *Canadian Lynx*, utilizando as estatísticas de aderência MSE e MAE.

Autores	Métodos	h=14	
		MSE	MAE
Zhang (2003)	ARIMA	0,020486	0,112255
	ANN	0,020466	0,112109
	HYBRID	0,017233	0,103972
Kajitani <i>et al.</i> (2005)	FFNN	0,0090	-
	SETAR	0,0140	-
Aladag <i>et al.</i> (2009)	HYBRID	0,00900	-
Zheng; Zhong (2011)	BS-RBF	0,002809	-
	BS-RBFAR	0,002199	-
Khashei; Bijari (2011)	ANN(p,d,q)	0,013609	0,089625
	ANNs/ARIMA	0,00999	0,085055
Khashei; Bijari (2012)	ARIMA/PNN	0,011461	0,084381
	ANN/PNN	0,014872	0,079628
Karnaboopathy; Venkatesan (2012)	FRAR	0,00455	-
Adhikari; Agrawal (2013)	ARIMA	0,01285	-
	SVR	0,05267	-
	ENSAMBLE	0,00715	-
Ismail; Shabri (2014)	SVR	0,0085	0,07460
	LSSVM	0,00300	0,04180
Teixeira <i>et al.</i> (2015)	Método Proposto	0,00016	0,0091

Fonte: Elaborado pelos autores (2015).

Na Figura 5, pode-se visualizar que as previsões do método proposto, na amostra de teste, apresentam forte correlação com as observações (o que indica alto nível de acurácia preditiva), o que ratifica os bons resultados de MSE e MAE, na Tabela 3.

Figura 5 – Gráficos da série temporal logaritmo *Canadian Lynx* e de suas respectivas previsões produzidas pelo método proposto, na amostra de teste.



Fonte: Elaborado pelos autores (2015).

6. CONSIDERAÇÕES FINAIS

No presente trabalho foram feitas previsões para a série de tempo log *Canadian Lynx*, através de uma metodologia híbrida que integra os métodos de filtragem (ou encolhimento) *wavelet*, decomposição *wavelet* de nível p e modelagem RNA-MLP.

Baseado na Tabela 3, bem como nas outras análises, conclui-se a integração de métodos distintos incorreu na redução relevante nas estatísticas de aderência MSE e MAE, em relação a todas as outras abordagens consideradas. Em particular, confrontando-o com o segundo melhor resultado (a saber, BS-RBFAR), a redução na estatística MSE foi de 92,72%, aproximadamente. Em relação à MAE do LSSVM, a redução foi cerca de 78,22%. Os resultados auferidos com a série temporal log de *Canadian Lynx* são motivadores da aplicação do método proposto a séries de tempo de outras áreas.

Os ganhos preditivos obtidos pela metodologia proposta frente a diversos métodos individuais e híbridos permitem inferir que a adoção da metodologia por empresas implicará em planejamentos mais precisos, desde que os gestores se atinem para a qualidade dos dados fornecidos aos modelos matemáticos.

Em última análise, salienta-se que, embora o embasamento teórico do método proposto seja matematicamente complexo, a utilização do software Matlab viabiliza a sua utilização, ao passo de poder ser operacionalizado de forma relativamente simples.

REFERÊNCIAS

- ABREU, T.; ARAÚJO, K. A. M.; LOPES, M. L. M.; LOTUFO, A. D. P. Metodologia híbrida utilizando os modelos arima e redes neurais artificiais para previsão de cargas elétricas. *In: XIX CBA - Congresso Brasileiro de Automática*, Campina Grande, Paraíba, 2012. **Anais... XIX CBA**. SBA: Campinas, São Paulo, p. 2229–2236, 2012.
- ADHIKARI, R.; AGRAWAL, R. K. A Homogeneous Ensemble of Artificial Neural Networks for Time Series Forecasting. **International Journal of Computer Applications**, v. 32, n. 7, p. 8, 2013.
- ALADAG, C. H.; EGRIOGLU, E.; KADILAR, C. Forecasting nonlinear time series with a hybrid methodology. **Applied Mathematics Letters**, v. 22, n. 9, p. 1467–1470, 2009.
- BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.
- AUBECHIES, I. Orthonormal bases of compactly supported wavelets. **Communications on Pure and Applied Mathematics**, v. 41, n. 7, p. 909–996, 1988.
- DAUBECHIES, I. Ten lectures on wavelets. **Philadelphia – Society for industrial and applied mathematics**, v. 61, p. 198–202, 1992.
- DONOHO, D.; JOHNSTONE, I. Adapting to Unknown Smoothness via Wavelet Shrinkage. **Journal Of The American Statistical Association**, v. 90, n. 432, p. 1200–1224, 1995.
- DONOHO, D. L.; JOHNSTONE, I. M. Minimax estimation via wavelet shrinkage. **The Annals of Statistics**, v. 26, n. 3, p. 879–921, 1998.
- DONOHO, D. L.; JOHNSTONE, I. M.; KERKYACHARIAN, G.; PICARD, D. Wavelet shrinkage: asymptopia. **Journal Royal Statistical Society**, v. 57, p. 301–369, 1995.
- DONOHO, D. L.; JOHNSTONE, J. M. Ideal spatial adaptation by wavelet shrinkage. **Biometrika**, v. 81, n. 3, p. 425–455, 1994.
- DE GOOIJER, J. G.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, v. 22, n. 3, p. 443–473, 2006.
- HAYKIN, S. S. **Redes Neurais**. 2. ed. Porto Alegre: Bookman, 2001.
- ISMAIL, S.; SHABRI, A. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. **Jurnal Teknologi**, v. 1, p. 11–15, 2014.
- KAJITANI, Y.; HIPEL, K. W.; MCLEOD, A. I. Forecasting nonlinear time series with feed-forward neural networks: A case study of Canadian lynx data. **Journal of Forecasting**, v. 24, p. 105–117, 2005.

- KARNABOOPATHY, R.; VENKATESAN, D. Data mining in canadian lynx time series. **Journal of Reliability and Statistical Studies**, v. 5, n. 1, p. 1–06, 2012.
- KARTHIKEYAN, L.; NAGESH KUMAR, D. Predictability of nonstationary time series using wavelet and EMD based ARMA models. **Journal of Hydrology**, v. 502, p. 103–119, 2013.
- KHASHEI, M.; BIJARI, M. Which Methodology is Better for Combining Linear and Nonlinear Models for. **Time Series Forecasting**, v. 4, n. 4, p. 265–285, 2011.
- KHASHEI, M.; BIJARI, M. A new class of hybrid models for time series forecasting. **Expert Systems with Applications**, v. 39, n. 4, p. 4344–4357, 2012.
- KISI, O. Wavelet regression model for short-term streamflow forecasting. **Journal of Hydrology**, v. 389, n. 3-4, p. 344–353, 2010.
- KISI, O.; CIMEN, M. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. **Journal of Hydrology**, v. 399, n. 1-2, p. 132–140, 2011.
- KUBRUSLY, C. S.; LEVAN, H. Abstract wavelets generated by Hilbert space shift operators. **Advances in mathematical and applications**, v. 16, p. 643–660, 2006.
- LEI, C.; RAN, L. Short-term wind speed forecasting model for wind farm based on wavelet decomposition. *In: DRPT - Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*. Nanjing, China, 2008. **Anais... III DRPT**. IEEE: New York, NY, p.2525–2529, 2008.
- LEVAN, N.; KUBRUSLY, C. S. A wavelet “time-shift-detail” decomposition. **Mathematics and Computers in Simulation**, v. 63, n. 2, p. 73–78, 2003.
- LIU, L.-M. **Time Series Analysis and Forecasting**. 2 ed. Chicago, IL: Scientific Computing Associates Corporation, 2006.
- MALLAT, S. **A Wavelet Tour of Signal Processing: The Sparse Way**. 3 ed. 2008.
- MORAN, P. The statistical analysis of the Canadian Lynx cycle. **Australian Journal of Zoology**, v. 1, n. 3, p. 291, 1953.
- MORETTIN, P. A. **Ondas e Ondaletas: Da análise de Fourier à Análise**. São Paulo: EDUSP, 1999.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de séries temporais**. Edgard Blucher, 2006.
- SEO, Y.; KIM, S.; KISI, O.; SINGH, V. P. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. **Journal of Hydrology**, v. 520, p. 224–243, 2015.
- STONE, L.; HE, D. Chaotic oscillations and cycles in multi-trophic ecological systems. **Journal of theoretical biology**, v. 248, n. 2, p. 382–90, 2007.

SUBBA, R. T.; SABR, M. M. An Introduction to Bispectral Analysis and Bilinear Time Series Models. **Lecture Notes in Statistics**, v. 24, p. 31-42, 1984.

TANG, Y.; GHOSAL, S. A consistent nonparametric Bayesian procedure for estimating autoregressive conditional densities. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4424-4437, 2007.

TEIXEIRA JR, L. A. **Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e de Cenários**. PUC-Rio. 2013.

TEIXEIRA JR, L. A.; DE MENEZES, M. L.; CASSIANO, K. M.; PESSANHA, J. F. M.; SOUZA, R. C. Residential Electricity Consumption Forecasting Using a Geometric Combination Approach. **International Journal of Energy and Statistics**, v. 1, n. 2, p. 113-125, 2013.

TEIXEIRA JR, L. A.; PESSANHA, J. F. M.; MENEZES, M. L. DE; CASSIANO, K. M.; SOUZA, R. C. Redes neurais artificiais e decomposição wavelet na previsão da radiação solar direta. *In: XVI CLAIO - Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional*. **Anais... XVI CLAIO**. SOBRAPO: São Paulo, SP, p.1401-1412, 2012.

TIWARI, M. K.; CHATTERJEE, C. Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. **Journal of Hydrology**, v. 394, n. 3-4, p. 458-470, 2010.

WANG, W.; GELDER, P. H. A. J. M. VAN; VRIJLING, J. K.; MA, J. Forecasting daily streamflow using hybrid ANN models. **Journal of Hydrology**, v. 324, n. 1-4, p. 383-399, 2006.

WONG, C. S.; LI, W. K. On a mixture autoregressive model. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 62, n. 1, p. 91-115, 2000.

ZHANG, G. P. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.

ZHENG, F.; ZHONG, S. Time series forecasting using a hybrid RBF neural network and AR model based on binomial smoothing. **World Academy of Science Engineering and Technology**, p. 1125-1129, 2011.