

PREVISÃO DE CURTO PRAZO DA SÉRIE TEMPORAL CANADIAN LYNX VIA REDES NEURAIS ARTIFICIAIS E DECOMPOSIÇÃO WAVELET

Luiz Albino Teixeira Júnior

Instituto Latino-Americano de Tecnologia, Infraestrutura e Território
Universidade Federal da Integração Latino-Americana - UNILA
luiz.a.t.junior@gmail.com

Edgar Manuel Carreño Franco

Departamento de Engenharia Elétrica
Universidade Estadual do Oeste do Paraná - UNIOESTE
emfra.uioeste@gmail.com

Rafael Morais de Souza

Departamento de Ciências Contábeis
Universidade Federal de Minas Gerais - UFMG
rmsouza@face.ufmg.br

RESUMO

Neste artigo, é proposta uma abordagem alternativa híbrida que combina uma Rede Neural Artificial e a Decomposição *Wavelet* (DW), referida por RNA-WD, com o objetivo de produzir previsões de curto prazo para a série temporal de *Canadian Lynx*. Em termos gerais, o método RNA-DW é dividido em dois passos: (1) é realizada a decomposição de nível p da série temporal a ser prevista, gerando-se $p+1$ componentes wavelet; e (2) os sinais defasados em J instantes das $p+1$ componentes wavelet geradas em (1) são simultaneamente utilizados como entradas de uma RNA cuja saída é uma previsão em um instante de tempo t para a série temporal em questão. Os resultados obtidos mostraram que a RNA-WD obteve desempenho preditivo superior a uma RNA tradicional.

PALAVRAS-CHAVE: Séries Temporais, Decomposição *Wavelet*, Redes Neurais Artificiais, Previsões.

ABSTRACT

In this paper, it is proposed an alternative hybrid approach that combines an Artificial Neural Networks (ANN) and the Wavelet Decomposition (WD), referred as ANN-WD, with the aim to produce short-term Canadian lynx time series forecasting. In general terms, the ANN-WD method is divided into two steps: (1) the wavelet decomposition of p level of the Canadian Lynx is made; and (2) the lagged signals at J instants of the $p+1$ wavelet components generated in (1) are simultaneously used as inputs of an ANN whose outcome is a forecast at some instant t to the time series. The obtained results showed that the ANN-WD has reached forecasting performance better than the traditional ANN.

KEYWORDS: Time Series, Wavelet Decomposition, Artificial Neural Networks, Forecasts.

1. INTRODUÇÃO

Desde 1927, quando George Udny Yule propôs interpretar uma sequência de dados indexados pelo tempo (isto é, uma série temporal) como uma realização de uma coleção de variáveis aleatórias probabilisticamente distribuídas (chamada de processo estocástico) e, no caso da existência de auto-correlação e estacionariedade, projetá-la através de modelos auto-regressivos (AR's) [Hamilton, 2009], 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 [Hamilton, 2009]. 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, 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, advém importantes métodos de pré-processamento de dados que fazem o seu tratamento antes de sua efetiva modelagem. Comumente, os referidos métodos consistem em decompor [Donoho e Jonhstone, 1994] ou filtrar [Donoho *et al.*, 1995] os dados. A decomposição wavelet de nível r (que é o mais comum), em linhas gerais, consiste em expandir uma série temporal em termos de sua componente wavelet (CW) de aproximação e de suas p CWs de detalhes, resultando em $p+1$ CWs, com níveis de resolução fixos e distintos [Teixeira Jr, 2013]. A filtragem wavelet (wavelet shrinkage), por outro lado, objetiva 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 Júnior *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.

Assim, o objetivo deste artigo, é o de utilizar um método (denotado por RNA-DW) que propõe uma abordagem alternativa na combinação de uma Rede Neural Artificial (RNA) feedforward MLP com a Decomposição Wavelet (DW) para a previsão da série temporal de Canadian Lynx.

Este artigo está organizado em seis seções, além desta introdução. Nas seções 2 e 3, são introduzidos aspectos teóricos da Teoria Wavelet e das Redes Neurais Artificiais, respectivamente. O método RNA-DW é detalhado na seção 4. O Experimento Computacional, bem como os seus principais resultados, são apresentados, na seção 5. Na seção 6, têm-se as conclusões da pesquisa.

2. ANÁLISE WAVELET

2.1 PRELIMINARES

De acordo com Kubrusly (2001), um espaço de Hilbert H é um espaço produto interno completo. A coleção l^2 de todas as sequências infinitas de números complexos quadraticamente somáveis (ou seja, $l^2 := \{f: \mathbb{Z} \rightarrow \mathbb{C}: \sum_{t \in \mathbb{Z}} |f(t)|^2 < \infty\}$), munida de um produto interno $\langle ; \rangle$ (onde, $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$) - ou, simplesmente, o par ordenado $(l^2, \langle ; \rangle)$ - é um exemplo de espaços de Hilbert (KUBRUSLY, 2001). Baseado em (Kubrusly e Levan, 2002), um subconjunto $\{h_n\}_{n \in \mathbb{Z}}$ de $(l^2, \langle ; \rangle)$ é uma base ortonormal se, e somente se, satisfaz aos seguintes axiomas:

- (i) ortogonalidade: $\langle h_{n'}, h_m \rangle = 0$, sempre que $n' \neq m$, onde $n', m \in \mathbb{Z}$;
- (ii) normalização: $\|h_{n'}\| = 1$, onde $n' \in \mathbb{Z}$; e
- (iii) completamento: $\langle x, h_{n'} \rangle = 0$ se, e somente se, $x = 0$.

De acordo com o Teorema da Série de Fourier (Kubrusly, 2001), se o subconjunto $\{h_n\}_{n \in \mathbb{Z}}$ de l^2 é uma base ortonormal, em relação ao produto interno $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$, então a identidade, em (1), é uma expansão única x sobre o espaço de Hilbert $(l^2, \langle ; \rangle)$, em termos da base ortonormal $\{h_n\}_{n \in \mathbb{Z}}$, e é chamada série de Fourier.

$$x = \sum_{n \in \mathbb{Z}} \langle x, h_n \rangle h_n. \quad (1)$$

2.2 FUNÇÕES WAVELET

Tome um espaço de Hilbert $(l^2, \langle ; \rangle)$. Um elemento $\omega(.) \in l^2$ - com um produto interno $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$ - é chamado de função wavelet se, e somente se, as funções $\omega_{m,n}(.) := 2^{\frac{m}{2}} \omega(2^m(.) - n)$, onde $n, m \in \mathbb{Z}$, formam uma base ortonormal para o espaço de Hilbert $(l^2, \langle ; \rangle)$. De acordo com Levan e Kubrusly (2003), qualquer função $f(.)$ em $(l^2, \langle ; \rangle)$ admite a expansão via Série de Fourier, em termos de uma base ortonormal wavelet $\{\omega_{m,n}(.)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ de l^2 , conforme em (2):

$$f(.) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle f(.), \omega_{m,n}(.) \rangle \omega_{m,n}(.) \quad (2)$$

onde m é chamado de parâmetro de escala e n , de parâmetro de translação (OGDEN, 1997).

De acordo com Levan e Kubrusly (2003), a projeção de $f(.)$ sobre $\omega_{m,n}(.)$ pode ser interpretada como uma variação de detalhes de $f(.)$, na escala m e translação n . De acordo com Mallat (1998), o subespaço fechado (Kubrusly, 2001) $W_m(\omega) := \left(\text{span}\{\omega_{m,n}(.)\}_{n \in \mathbb{Z}} \right)^-$ de $(l^2, \langle ; \rangle)$ é chamado de subespaço de detalhes (na escala m). Por sua vez, a projeção de $f(.)$ sobre o subespaço (fechado) de detalhes $W_m(\omega)$, denotada por $f_{W_m(\omega)}(.)$, é definida pela soma parcial descrita em (3):

$$f_{W_m(\omega)}(.) := \sum_{n \in \mathbb{Z}} \langle f(.), \omega_{m,n}(.) \rangle \omega_{m,n}(.) \quad (3)$$

De acordo com Levan e Kubrusly (2003), a projeção $f_{W_m(\omega)}(.)$ pode ser interpretada como uma componente de detalhes de $f(.)$, na escala m , sobre $(W_m(\omega), \langle ; \rangle)$. Como consequência, dada a identidade, em (1), segue que $f(.)$ pode ser interpretada como a soma de todas suas componentes de detalhes $f_{W_m(\omega)}(.)$, em todas as escalas inteiras m ,

sobre o subespaço fechado $((\sum_{n \in \mathbb{Z}} W_m(\omega))^\perp, \langle ; \rangle)$ de $(l^2, \langle ; \rangle)$. Tautologicamente, tem-se que $((\sum_{n \in \mathbb{Z}} W_m(\omega))^\perp, \langle ; \rangle) = (l^2, \langle ; \rangle)$.

Por sua vez, um elemento $\phi(.) \in l^2$ - com um produto interno $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$ - é chamado de função wavelet escala (ou, simplesmente, função escala) se, e somente se, as funções $\phi_{m,n}(.):= 2^{\frac{m}{2}} \phi(2^m(.) - n)$, onde $n, m \in \mathbb{Z}$, são tais que $\langle \phi_{m',n'}(.), \phi_{j,k}(.) \rangle = 0$, sempre que $m' = j$ e $n' \neq k$, e $\langle \phi_{m',n'}(.), \phi_{j,k}(.) \rangle \neq 0$, caso contrário. De acordo com Mallat (1998), o subespaço fechado $V_m(\phi):= \left(\text{span}\{\phi_{m,n}(.)\}_{n \in \mathbb{Z}} \right)^\perp$ de $(l^2, \langle ; \rangle)$ é chamado de subespaço de aproximação (na escala m). A projeção de $f(.)$ sobre o subespaço (fechado) de aproximação $V_m(\phi)$ é definida pela soma parcial descrita em (4):

$$f_{V_m(\phi)}(.):= \sum_{n \in \mathbb{Z}} \langle f(.), \phi_{m,n}(.) \rangle \phi_{m,n}(.) \quad (4)$$

De acordo com Mallat (1998), $f_{V_m(\phi)}(.)$ pode ser interpretada como uma componente de aproximação de $f(.)$, na escala m , sobre o subespaço $(V_m(\phi), \langle ; \rangle)$ de $(l^2, \langle ; \rangle)$.

Por transformada wavelet sobre $(l^2, \langle ; \rangle)$, entende-se como sendo o produto interno $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$ entre uma função $f(.) \in l^2$ e uma função wavelet $\omega_{m,n}(.) \in W_m(\omega)$ ou uma função escala $\phi_{m,n}(.) \in V_m(\phi)$, $(m, n) \in \mathbb{Z} \times \mathbb{Z}$. De acordo com Mallat (1998), as transformadas wavelet podem ser classificadas e agrupadas em dois conjuntos disjuntos: coeficientes de detalhes, denotado por $\{d_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, e coeficientes de aproximação, denotado por $\{a_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$. Para cada par ordenado $(m, n) \in \mathbb{Z} \times \mathbb{Z}$, tem-se que as transformadas wavelet $d_{m,n}$ e $a_{m,n}$ são definidas, respectivamente, por $d_{m,n}:= \langle f(.), \omega_{m,n}(.) \rangle = \sum_{t \in \mathbb{Z}} f(t) \omega_{m,n}(t)$ e $a_{m,n}:= \langle f(.), \phi_{m,n}(.) \rangle = \sum_{t \in \mathbb{Z}} f(t) \phi_{m,n}(t)$.

2.3 DECOMPOSIÇÃO WAVELET

De acordo com Levan e Kubrusly (2003), uma cadeia de subespaços de aproximação $\{V_m(\phi)\}_{m \in \mathbb{Z}}$ em $(l^2, \langle ; \rangle)$ é uma Análise Multirresolução Wavelet (ou, simplesmente, AMR wavelet), com função escala $\phi(.) \in l^2$, se, e somente se, os seguintes axiomas (a), (b), (c), (d), e (e) são satisfeitos:

(a) $V_m(\phi) \subset V_{m+1}(\phi), \forall m \in \mathbb{Z}$;

(b) $\bigcap_{m \in \mathbb{Z}} V_m(\phi) = \{0\}$;

(c) $\left(\bigcup_{m \in \mathbb{Z}} V_m(\phi) \right)^\perp = H$;

(d) $v \in V_m(\phi) \Leftrightarrow Dv \in V_{m+1}(\phi), m \in \mathbb{Z}$; e

(e) $\left\{ 2^{\frac{m}{2}} \phi(2^m(.) - n) \right\}_{n \in \mathbb{Z}}$ é uma base ortonormal de $V_m(\phi), m \in \mathbb{Z}$.

Em (Kubrusly e Levan, 2002), mostra-se que um espaço l^2 pode ser expandido ortogonalmente tal como $l^2 = \left(\sum_{m \in \mathbb{Z}} W_m(\omega) \right)^\perp$ e, em Levan e Kubrusly (2003), mostra-se, utilizando-se os axiomas de uma AMR wavelet $\{V_m(\phi)\}_{m \in \mathbb{Z}}$, que a identidade $V_{m_0}(\phi) = \left(\sum_{-\infty}^{m_0-1} W_m(\omega) \right)^\perp$, para todo $m_0 \in \mathbb{Z}$, é verdadeira. Com base nas identidades

$l^2 = \left(\sum_{m \in \mathbb{Z}} W_m(\omega) \right)^{-}$ e $V_{m_0}(\phi) = \left(\sum_{m=m_0}^{+\infty} W_m(\omega) \right)^{-}$, bem como no Teorema da Estrutura Ortogonal (Kubrusly, 2001), mostra-se, em (Kubrusly e Levan, 2002), que o espaço l^2 pode ser ortogonalmente expandido conforme em (5):

$$l^2 = V_{m_0}(\phi) + \left(\sum_{m=m_0}^{+\infty} W_m(\omega) \right)^{-} \quad (5)$$

Como consequência, segue que $f(\cdot)$ possui decomposição ortogonal (única), sobre o espaço de Hilbert $(l^2, \langle \cdot, \cdot \rangle)$, conforme em (6):

$$f(\cdot) = f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{+\infty} f_{W_m(\omega)}(\cdot) \quad (6)$$

Dadas as definições das componentes wavelet $f_{V_{m_0}(\phi)}(\cdot)$ e $f_{W_m(\omega)}(\cdot)$ e as identidades (1) e (6), segue que a Série de Fourier da função $f(\cdot)$, sobre o espaço de Hilbert $(l^2, \langle \cdot, \cdot \rangle)$, em termos da base ortonormal wavelet $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{n,m}(\cdot)\}_{(m,n) \in \{m\}_{m_0}^{+\infty} \times \mathbb{Z}}$, é dada por:

$$f(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot) + \sum_{m=m_0}^{+\infty} \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{n,m}(\cdot). \quad (7)$$

Onde: $a_{m,n} := \sum_{t \in \mathbb{Z}} f(t) \phi_{m,n}(t)$, $d_{m,n} := \sum_{t \in \mathbb{Z}} f(t) \omega_{n,m}(t)$, onde $m_0 \leq m < +\infty$ e $m_0 \in \mathbb{Z}$.

3. REDES NEURAIS ARTIFICIAIS

De acordo com Haykin (2001), as Redes Neurais Artificiais são sistemas paralelos distribuídos e compostos por unidades de processamento simples, denominadas neurônios artificiais, e dispostas em uma ou mais camadas interligadas por um grande número de conexões (sinapses), geralmente, unidirecionais (feedforward) e com pesos sinápticos para ponderação das entradas recebidas por cada neurônio. Na Figura 1(a), tem-se uma ilustração da arquitetura mais comum de uma RNA (feedforward) multicamadas perceptron (MLP) com três camadas: camada de entrada (input), camada escondida (ou intermediária) e camada de saída (output).

A primeira camada da rede neural artificial é a camada de entrada, sendo a única exposta aos padrões de entrada. A camada de entrada transmite os valores dos padrões de entrada para os neurônios da camada intermediária para que estes extraiam padrões e transmitam os resultados para a camada de saída (última camada da RNA). A definição do número de neurônios na camada intermediária é realizada de forma empírica HAYKIN (2001).

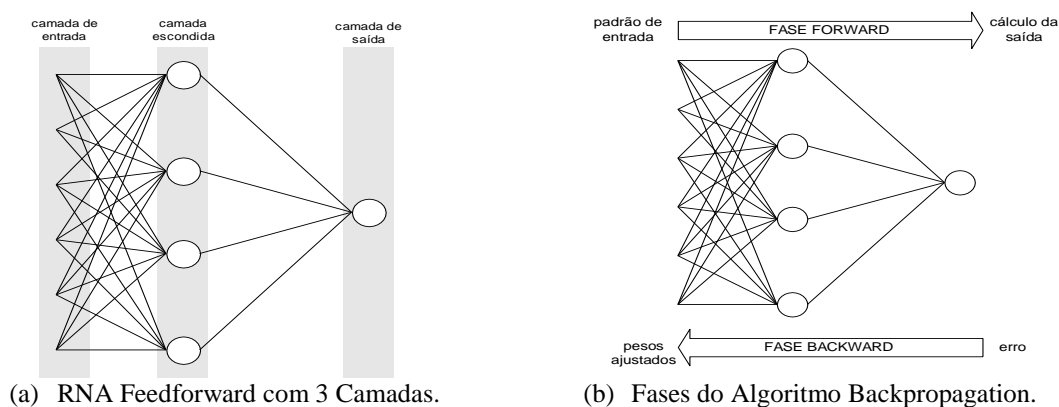


Figura 1 - Rede Neural Artificial Feedforward MLP e Algoritmo Backpropagation.

O principal algoritmo de treinamento de redes neurais artificiais é o backpropagation Haykin (2001), cujo ajuste dos pesos sinápticos ocorre por meio de um processo de otimização realizado em duas fases: forward e backward, conforme ilustrado na Figura 1(b). 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. Ao longo do treinamento da RNA, os vários padrões de entrada e as respectivas respostas desejadas são apresentados à RNA, de forma que os pesos sinápticos sejam tais que minimizem a soma dos erros quadráticos (MSE).

A previsão de valores futuros de uma série temporal, por meio de uma RNA, inicia-se com a montagem do conjunto de padrões de treinamento (pares entrada/saída), o qual depende da definição do tamanho da janela L de tempo (para os valores passados da própria série temporal que se deseja prever e para as variáveis explicativas) e do horizonte de previsão h . Em um processo autoregressivo (linear ou não linear), o padrão de entrada é formado pelos valores passados da própria série que se deseja prever.

Por sua vez, o padrão de saída desejada é o valor da observação da série temporal no horizonte de previsão. Na Figura 2, ilustra-se como geralmente é construído o conjunto de treinamento no caso da previsão basear-se nos 4 últimos valores passados. Note que a construção dos padrões de treinamento da rede consiste em mover as janelas de entrada e saída ao longo de toda série temporal, de modo que cada par de janelas (entrada/saída) funciona como um padrão de treinamento e deve ser apresentado repetidas vezes até que o algoritmo de aprendizado convirja.

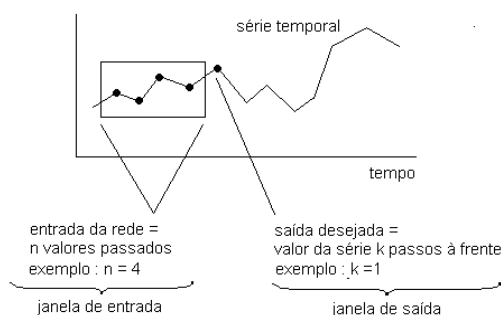


Figura 2 - Montagem do Conjunto de Treinamento.

4. MÉTODO RNA-CW

A combinação de uma Rede Neural Artificial (RNA) e uma decomposição wavelet (DW) pode ser realizada de diferentes maneiras. Por exemplo, pode-se aplicar a decomposição wavelet em uma série temporal de interesse e se utilizar alguma RNA tradicional para projetar cada componente wavelet, separadamente, e, ao final, somar suas previsões, a fim de se obter a previsão da série temporal supracitada. Outra opção consiste em usar as funções wavelet (normalizadas no intervalo $[0,1]$) como funções de ativação dos neurônios de uma RNA tradicional e utilizar os padrões de entrada decompostos através de DW.

Neste artigo, todavia, optou-se por um método de combinação de ambas (denotado por RNA-DW), no qual as componentes wavelet de uma série temporal de interesse constituem os padrões de entrada de uma RNA feedforward MLP cuja saída fornece a previsão da série temporal supracitada (conforme o esquema da Figura 3). Basicamente, A abordagem considerada pode ser dividida nas etapas (1) [descrita na seção 4.1] e (2) [descrita na seção 4.2]:

- (1) Decomposição Wavelet de nível p [7, 15, 10] de uma série temporal $f(\cdot)$; e
- (2) Apresentação das componentes wavelet de $f(\cdot)$ (oriundas da etapa 1) como entradas de uma RNA para a previsão da série temporal de interesse.

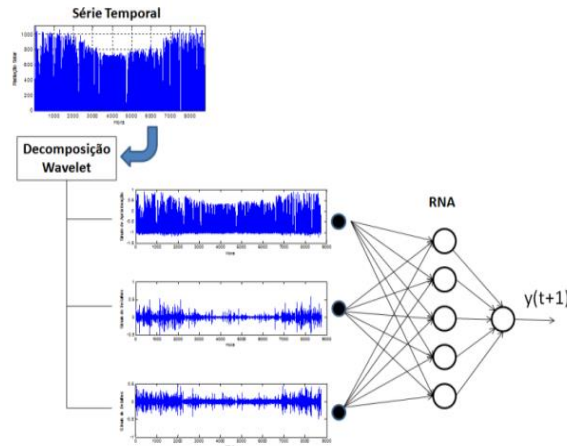


Figura 3 - Combinação da RNA com Decomposição Wavelet.

4.1 DECOMPOSIÇÃO WAVELET DE NÍVEL P

Sejam $f(\cdot)$ uma série temporal em $(l^2, \langle \cdot, \cdot \rangle)$ e $\{\phi_{m_0, n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{n, m}(\cdot)\}_{(m, n) \in \{m\}_{m_0}^{+\infty} \times \mathbb{Z}}$, uma base ortonormal wavelet do espaço de Hilbert $(l^2, \langle \cdot, \cdot \rangle)$. De acordo com a identidade em (7), a decomposição wavelet de nível p Teixeira Júnior *et al.* (2011) de $f(\cdot)$, onde p é um número natural no intervalo $1 \leq p < \infty$, é representada pela série de Fourier (aproximada) descrita em (8).

$$f(\cdot) \simeq \tilde{f}(\cdot) = \sum_{n=1}^{n_{m_0}} a_{m_0, n} \phi_{m_0, n}(\cdot) + \sum_{n=1}^{n_m} \sum_{m=m_0}^{m_0+(p-1)} d_{m, n} \omega_{m, n}(\cdot). \quad (8)$$

Os valores ótimos dos parâmetros m_0 , n_{m_0} e $\{n_m\}_{m=m_0}^{m_0+(p-1)}$ são tais que minimizam a métrica euclidiana (Kubrusly, 2001) entre a série temporal $f(\cdot)$ e sua

aproximação $\tilde{f}(\cdot)$. As componentes wavelet $f_{V_{m_0}(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot)$ e $f_{W_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot)$ são classificadas, respectivamente, como componente de aproximação (na escala m_0) e componente de detalhe (na escala m) da série temporal $f(\cdot)$ em $(l^2, \langle \cdot, \cdot \rangle)$. Dada a expansão em (8), segue que a série temporal $f(\cdot)$ pode ser expandido ortogonalmente sobre $(l^2, \langle \cdot, \cdot \rangle)$, conforme em (9).

$$f(\cdot) \simeq f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{m_0+(p-1)} f_{W_m(\omega)}(\cdot). \quad (9)$$

onde $f_{V_{m_0}(\phi)}(\cdot) = \{f_{V_{m_0}(\phi)}(t)\}_{t \in \mathbb{Z}}$, para um inteiro fixo m_0 , e $f_{W_m(\omega)}(\cdot) = \{f_{W_m(\omega)}(t)\}_{t \in \mathbb{Z}}$, onde m é um inteiro no intervalo $m_0 \leq m \leq m_0 + (p - 1)$, sendo p o nível de decomposição wavelet.

4.2 APRESENTAÇÃO DOS SINAIS WAVELET À RNA

Tome uma RNA feedforward MLP. O conjunto de sinais temporais wavelet $\left\{ \left\{ f_{V_{m_0}(\phi)}(t) \right\}_{t=1}^T \right\} \cup \left\{ \left\{ f_{W_m(\omega)}(t) \right\}_{t=1}^T \right\}_{m=m_0}^{m_0+(p-1)}$, oriundos de $p+1$ componentes wavelet de uma série temporal de interesse $\{f(t)\}_{t=1}^T$ [seção 4.1] são tais que constituem o conjunto de padrões de entrada de uma RNA feedforward MLP para realização do processo de treinamento. Considerando uma janela de tamanho igual a L valores passados, a previsão da série temporal (saída de RNA) para cada instante t' (nas amostras de treino, validação e de teste) é obtida a partir do conjunto de padrões de entrada descrito em (10).

$$\left\{ f_{V_{m_0}(\phi)}(t) \right\}_{t=t'-L}^{t'-1} \cup \left\{ \left\{ f_{W_m(\omega)}(t) \right\}_{t=t'-L}^{t'-1} \right\}_{m=m_0}^{m_0+(p-1)} \quad (10)$$

5. EXPERIMENTOS NUMÉRICOS

A série temporal Canadian Lynx (lincos canadenses), a qual é utilizada na parte de experimentos numéricos, refere-se ao registro anual do número de Lynx Canadensis (que é uma espécie de felino natural da América do Norte) capturados no distrito de Rio Mackenzie no Norte do Canadá e pode ser obtida em: <https://datamarket.com/>.

Há, na literatura, vários artigos que a utilizam para testar metodologias preditivas, como, por exemplo, Zhang (2003); Subba e Sabr (1984); Tang e Ghosal (2007); Wong e Li (2000), entre outros. No entanto, é creditado a Moran (1953) como sendo o primeiro. O gráfico da série temporal supracitada, com 114 observações, no período de 1821 a 1934, pode ser visualizado na Figura 4.

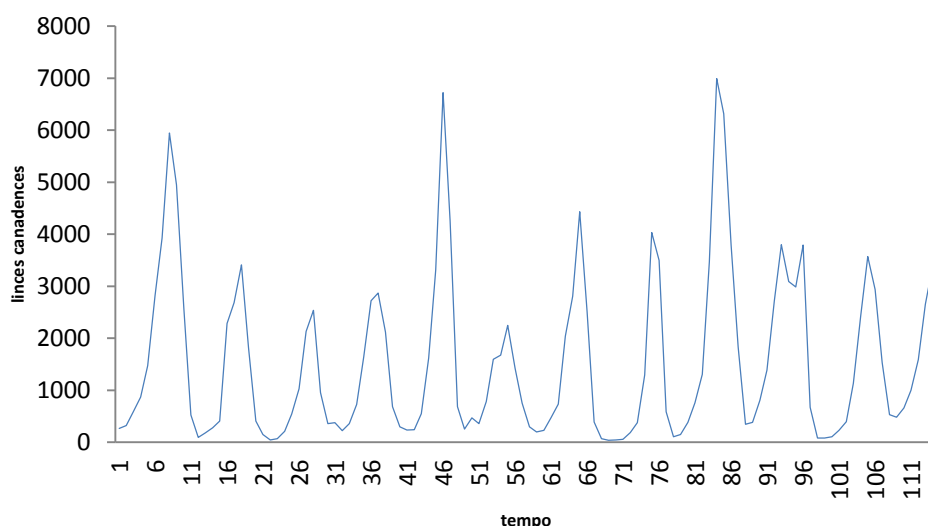


Figura 4 - Gráfico da série temporal de lincas canadenses no período 1821-1934.

Usualmente, os autores que publicam trabalhos sobre previsão da série dos lincas canadenses aplicam logaritmo de base dez ao conjunto de dados originais e, em seguida, a divide em: conjunto de treinamento (100 primeiros anos: 1821-1920); e conjunto de teste (últimos 14 anos). Além disso, fazem previsões um passo a frente para o horizonte de quatorze anos. A fim de manter o critério, o presente artigo utiliza a mesma abordagem. Para análise de desempenho, se utiliza geralmente as estatísticas MAE (mean absolute error) e MSE (mean square error), fora da amostra.

5.1 MODELAGEM

Previamente à sua modelagem, a série temporal Cadian Lynx foi ortogonalmente expandida através da decomposição wavelet de nível 2, definida em termos da base ortonormal da família Haar. Quanto à RNA (feedforward perceptron multicamada), as suas configurações foram: normalização premnmx; tamanho de janela igual a 6; uma camada escondida com 6 neurônios com função de ativação tangente hiperbólica; uma camada de saída com um neurônio com função de ativação linear; algoritmo de treinamento de Levenberg-Marquardt (trainlm).

Tabela 1 - Comparativo entre métodos de previsão, considerando o MSE e o MAE.

Método	MSE	MAE
ARIMA (Zhang, 2003)	0,0205	0,1123
RNA (Zhang, 2003)	0,0205	0,1121
Híbrido (Zhang, 2003)	0,0172	0,1040
SETAR (Kajitani <i>et al.</i> , 2005)	0,0140	-
RNA (Khashei e Bijari, 2010)	0,0136	0,0896
FFNN (Kajitani <i>et al.</i> , 2005)	0,0090	-
Híbrido (Aladag <i>et al.</i> , 2009)	0,0090	-
SVR (Ismail e Shabri, 2014)	0,0085	0,0746
RNA-DW	0,00699	0,06597

Na tabela 1, têm-se os resultados das estatísticas de aderência MSE e MAE, fora da amostra, do método RNA-DW (proposto) e os de outras abordagens na literatura. Note

que o método proposto obteve resultado superior a todos os outros. Na Figura 5, pode-se visualizar que as previsões do método RNA-DW e os valores reais, na amostra de teste, apresentam forte correlação, o que significa um nível elevado de precisão.

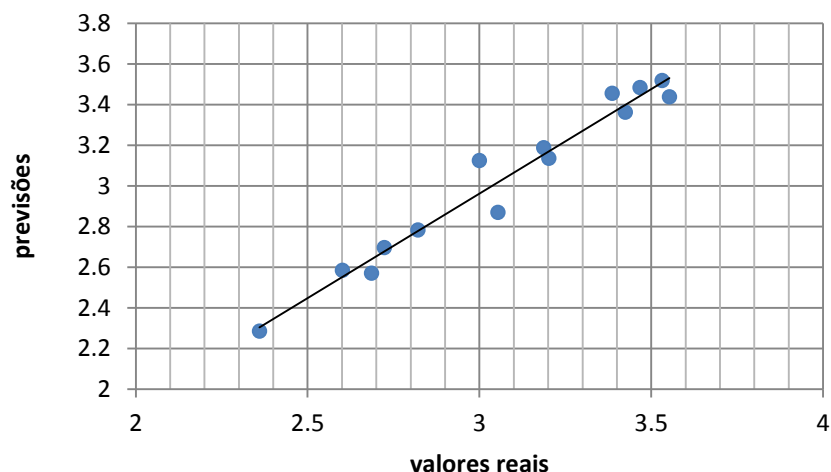


Figura 5 – Gráfico de dispersão entre as previsões da RNA-DW e as observações.

6. CONCLUSÕES

Neste artigo, foi utilizado um método (denotado por RNA-DW) que propõe uma abordagem alternativa na combinação de uma Rede Neural Artificial (RNA) feedforward MLP com a Decomposição Wavelet (DW) para a previsão da série temporal de Canadian Lynx.

Na amostra de teste, as previsões oriundas do método RNA-DW possuem, em relação a outras metodologias, maior correlação (o que significa maior acurácia) com os dados reais. Destaca-se ainda que o método RNA-DW pode ser operacionalmente implementado, de forma direta e imediata, com o auxílio do software MATLAB para gerar previsões de qualquer série temporal.

Por fim, é importante salientar que, para se realizar de uma modelagem adequada e eficiente, é importante considerar a forma como uma série temporal de interesse é apresentada ao método preditivo (como, por exemplo, uma RNA). Em outras palavras, a escolha do pré-processamento dos dados de entrada de previsor (como, por exemplo, a decomposição wavelet) é tão importante quanto a escolha deste. Nesta perspectiva, tem-se que o método RNA-DW contempla estes dois aspectos, o que o torna mais genérico e sofisticado na modelagem de qualquer série temporal.

7. REFERÊNCIAS

- [1]. 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**. Congresso Brasileiro de Automática. Campina Grande, 2012.
- [2]. ALADAG, C. H.; EGRIÖGLU, E.; KADILAR, C. Forecasting Nonlinear Time Series With a Hybrid Methodology. **Applied Mathematics Letters**, v.22, p.1467-1470, 2009.

- [3]. BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.
- [4]. DONOHO, D.L.; JONHSTONE, I.M.; KERKYACHARIAN, G.; PICARD, D. **Universal near Minimality of Wavelet Shrinkage**. Stanford University Technical Report, 1995.
- [5]. DONOHO, D. L.; JONHSTONE, I. M. Ideal Spatial Adaptation by Wavelet Shrinkage. **Biometric**, 81,425-455, 1994.
- [6]. GOOIJER, J. G.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, p.443– 473, 2006.
- [7]. ISMAIL, S.; SHABRI, A. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. **Jurnal Teknologi**, p.11-15, 2014.
- [8]. HAYKIN, S. **Redes Neurais Princípios e Aplicações**, Segunda Edição. Porto Alegre, 2001.
- [9]. KAJITANI, Y.; KEITH, W. H.; 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.
- [10]. KASHEI, M.; BIJARI, M. An Artificial Neural Network (p,d,q) Model for Times Forecasting. **Expert Systems With Applications**, v. 37, p. 479-489, 2010.
- [11]. KUBRUSLY, C. S. **Elements of Operator Theory**. Birkhäuser, Boston, 2001.
- [12]. KUBRUSLY, C. S., LEVAN, N. Dual-Shift Decomposition of Hilbert Space. **Semigroups of Operators: Theory and Application**, v.2, p. 145-157, 2002.
- [13]. LEVAN, N.; KUBRUSLY, C. S. A Wavelet “Time-Shift-Detail” Decomposition. **Mathematics and Computers in Simulation**, v.63, p. 73-78, 2003.
- [14]. MALLAT, S. **A Wavelet Tour of Signal Processing**. Academic Press. San Diego, 1998.
- [15]. MORAN, P. A. P. The Statistical Analysis of the Canadian Lynx Cycle. **Australian Journal of Zoology**, v.1, p. 163-173, 1953.
- [16]. OGDEN, R.T. **Essential Wavelet for Statistical Applications and Data Analysis**. Birkhäuser. Boston, 1997.
- [17]. SUBBA, R. T.; SABR, M. M. An Introduction to Bispectral Analysis and Bilinear Time Series Models. **Lecture Notes in Statistics**, v. 24, 1984.
- [18]. TANG, Y.; GHOSAL, S. A Consistent Nonparametric Bayesian procedure for Estimating Autoregressive Conditional Densities. **Computational Statistics & Data Analysis**, v.51, p. 4424-4437, 2007.
- [19]. TEIXEIRA JÚNIOR, L. A. **Combinação SSA-Wavelet de métodos preditivos com ajuste numérico MINIMAX, na geração de previsões e de cenários**. Rio de Janeiro, 114 p. Tese (Doutorado em Engenharia Elétrica). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.
- [20]. TEIXEIRA JÚNIOR, L. A.; MENESES, M. L.; PESSANHA, J. F. M. SOUZA, R. C.; CASSIANO, K. M. **Redes Neurais Artificiais Wavelet Causal na Previsão da Radiação Solar Direta**. Simpósio Brasileiro de Pesquisa Operacional. Rio de Janeiro, 2012.

- [21]. TEIXEIRA JÚNIOR, L.A.; PESSANHA, J.F.M.; SOUZA, R.C. **Análise Wavelet e Redes Neurais Artificiais na Previsão da Velocidade de Vento**. Simpósio Brasileiro de Pesquisa Operacional. Ubatuba, 2011.
- [22]. WANG, A.B.W., VAN GELDER, P.H.A.J.M., VRIJLING, J.K.; MA, J. Forecasting daily streamflow using hybrid ANN models. Amsterdam: **Journal of Hydrology**, v. 324, p. 383-399, 2006.
- [23]. WONG, C. S.; LI, W. K. On a Mixture Autoregressive Model. **Journal of the Royal Statistical Society Series B**, v.62, p. 91-115, 2000.
- [24]. ZHANG, G.P. Times Series Forecasting Using Hybrid ARIMA and Neural Network Model. **Neurocomputing**, v. 50, p.159-175, 2003.