

PREVISOR SVR-LSSVR WAVELET NA PROJEÇÃO DE SÉRIES TEMPORAIS

Samuel Bellido Rodrigues, UTFPR, samuelbellido@hotmail.com
Arinei Carlos Lindbeck da Silva, UFPR, arineicls@gmail.com
Luiz Albino Teixeira Júnior, UNILA, luiz.a.t.junior@gmail.com
Edgar Manuel Carreño Franco, UNIOESTE, emfra.uioeste@gmail.com
Rafael Morais de Souza, UFMG, rafael.morais.souza@gmail.com

Resumo: Na literatura, é bem conhecido que os métodos preditivos (ou previsores) *support vector regression* (SVR) e *least square support vector regression* (LSSVR) consistem em alternativas eficientes na projeção de séries temporais (estocásticas) e que a decomposição wavelet oferece vantagens atrativas no processo preditivo. Assim o sendo, utilizando programação não linear e combinação linear de previsões, propõe-se neste artigo um previsor híbrido que integra as seguintes abordagens: SVR, LSSVR e decomposição wavelet. A fim de ilustrá-lo, é utilizada a série temporal Canadian lynx. Os resultados alcançados pelo previsor híbrido proposto alcançou maior nível de acurácia que dezoito métodos competitivos.

Palavras-chave: SVR, LSSVR, decomposição wavelet, programação não linear, combinação de previsões.

FORECAST THROUGH THE COMBINATION WAVELET SVR LSSVR

Abstract: In literature, it is well known that support vector regression (SVR) and least square support vector regression (LSSVR) predictive methods (or forecasters) consist of efficient alternatives in the projection of (stochastic) time series, and the wavelet decomposition offers attractive advantages in predictive process. So being, using non-linear programming and linear combination of forecasts, this paper proposes a hybrid forecaster that integrates the following approaches: SVR, LSSVR and wavelet decomposition. In order to illustrate it, the Canadian lynx time series is used. The results achieved by the proposed hybrid forecaster have achieved higher level of accuracy than eighteen competitive methods.

Keywords: SVR, LSSVR, wavelet decomposition, non-linear programing, combination of forecasts.

1. INTRODUÇÃO

Um dos objetivos no estudo de séries temporais (estocásticas) é modelá-las com a finalidade de se produzir previsões, com o maior nível de acurácia possível. Não obstante, para prevê-las, é necessária a utilização de observações passadas que são analisadas com o intuito de se desenvolver um modelo matemático que as descreva, por meio da relação de autodependência existente entre elas. A acurácia nas projeções temporais é de suma importância em muitos processos de tomada de decisão, o que motiva a realização de pesquisas, cada vez maiores, neste campo do conhecimento.

Nesta perspectiva, o método preditivo individual *support vector regression* (SVR), cuja versão inicial foi proposta por Vapnik (1995), consiste em uma metodologia preditiva baseada na teoria da aprendizagem estatística, a qual adere ao princípio da minimização do risco estrutural. Tal abordagem que vem sendo frequentemente utilizada, nos últimos anos, na projeção de séries temporais, principalmente aquelas exibem não linearidade em sua estrutura de autodependência, com extraordinário sucesso, como mostram, dentre outros, Vapnik et al. (1996), Hong (2011), Anandhi e Chezian (2013). Derivado do SVR, o método *least squares support vector regression* (LSSVR), proposto por Suykens e Vandewalle (1999), possui as mesmas capacidades que o SVR, no tocante a mapear estruturas de autodependência não lineares; porém, um de seus diferenciais reside no fato de que utiliza restrições de igualdade,

ao invés de restrições de desigualdade. Além disso, adota o sistema de mínimos quadrados lineares como a função objetivo (a ser otimizada), o que o torna, do ponto de vista computacional e operacional, atraente.

Da Análise Wavelet, (Mallat, (2009)), obtém-se uma decomposição ortogonal de uma série temporal em termos de *componentes wavelet* (CWs). As CWs consistem em subséries temporais com frequências espectrais diferentes e constantes, que tendem a exibir melhores padrões de comportamento que a série temporal original. Diversos artigos propõem a decomposição wavelet como um método auxiliar à modelagem de sinais temporais e que apresentam ganhos preditivos relevantes como mostrado, por exemplo, em Teixeira Jr et al. (2012), Sang (2013), Teixeira Jr et al. (2015) e Teixeira et al. (2015). Em Teixeira Jr et al. (2012), é apresentada uma metodologia em que as CWs de uma série temporal constituem os padrões de entrada de uma *rede neural artificial multilayer perceptron* (RNA-MLP), cuja camada de saída neural produz as suas previsões, dentro e fora da amostra. Os autores mostram que os resultados preditivos da RNA-MLP integrada com a decomposição wavelet alcançaram resultados preditivos muito superiores aos de uma RNA-MLP convencional (que não utiliza o pré-processamento dos padrões de entrada via a decomposição wavelet).

Por sua vez, é bem conhecido na literatura de séries temporais que a combinação linear de previsões pode obter ganhos de acurácia relevantes, no processo preditivo. Em Granger (1989), por exemplo, pode ser verificado, na modelagem de diversas séries temporais, oriundas de diferentes áreas, que a combinação linear de previsões incorreu em ganhos de acurácia preditiva. Faria e Mubwandarikwa (2008), por outro lado, salientam que a previsão combinada tende a ser mais informativa, uma vez que pode ser visualizada como uma agregadora de informações oriundas de diferentes fontes (a saber, métodos preditivos individuais, a saber). Em Wallis (2011) e em Mancuso e Werner (2013), em que são fornecidas revisões históricas sobre métodos preditivos combinados, são referidos e comentados muitos dos principais artigos que tratam de combinações lineares de previsões. Recentemente, várias outras contribuições atestam e incentivam à utilização de previsões linearmente combinadas, como, por exemplo em: Babu e Reddy (2014); Kosanan e Kantanantha (2014); Rodrigues et al. (2014); Teixeira et al. (2014); Rodrigues et al. (2015) e Teixeira et al. (2015).

Dado que a combinação linear de previsões e a decomposição wavelet podem contribuir substancialmente na obtenção de ganhos no processo de geração de previsões, este artigo propõe uma metodologia preditiva (aplicável à modelagem de qualquer série temporal que exiba estrutura de autodependência não linear) que integra as seguintes abordagens: decomposição wavelet, SVR, LSSVR e combinação linear de previsões. A escolha do SVR e LSSVR ocorreu por serem aplicáveis à modelagem de quaisquer séries temporais e pela sua comprovada eficiência ao fazê-lo. Informações distintas são capturadas por cada abordagem, de forma que as previsões combinadas são agregadoras de informações.

Desse modo, este artigo é dividido, como segue: na Seção 1, é feita uma contextualização; na Seção 2, realiza-se uma sucinta revisão da literatura; na Seção 3, a metodologia proposta é descrita; na Seção 4, são apresentados os principais resultados numéricos, com aplicações à série temporal Canadian Lynx, seguindo a abordagem proposta por Zhang (2003); e, finalmente, na Seção 5, têm-se as principais conclusões.

2. REVISÃO DA LITERATURA

2.1 Decomposição Wavelet

De acordo com Morettin e Tolo (2006), na Análise de Fourier Canônica, uma série temporal com T observações, denotada por $y(\cdot) := (y(1), y(2), \dots, y(T))$, pode ser ortogonalmente decomposta em termos de senos e cossenos definidos em diversas bandas de frequências; enquanto que, de acordo com Teixeira Jr et al. (2015), na Análise Wavelet, $y(\cdot)$ pode, também de forma ortogonal, ser decomposta, simultaneamente, a partir de informações de frequência e de tempo, ambas capturadas a partir de um conjunto de *funções wavelet* $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, onde \mathbb{Z} é o conjunto dos números inteiros, sendo que cada mapa $\omega_{m,n}(\cdot)$ é tal que $\omega_{m,n}(\cdot) = 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$, o qual é gerado por meio de convoluções de dilatação (associada ao parâmetro m) e de translação (ao parâmetro n) de uma função wavelet $\omega(\cdot)$ original (Mallat, (2009)). Ainda de acordo com Mallat (2009), o parâmetro m é chamado de parâmetro de escala diádica e n , de parâmetro de translação unitária.

Em Teixeira Jr et al. (2012), verifica-se, empiricamente, que cada função wavelet $\omega_{m,n}(\cdot)$ está associada ao mapeamento das componentes wavelet (CWs) de detalhe (que são as CW de alta frequência) da série temporal $y(\cdot)$; e, teórica e empiricamente, que a CW de aproximação (que é a CW de baixa frequência) da série temporal $y(\cdot)$ está associada ao conjunto de *funções escala* $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}}$ (Teixeira Jr et al., (2015)). Cada mapa $\phi_{m_0,n}(\cdot) = 2^{\frac{m_0}{2}} \phi(2^{m_0}(\cdot) - n)$, onde m_0 é um número inteiro fixo, é gerado por meio da convolução de translação unitária n de uma função escala original $\phi(\cdot)$.

De acordo com Mallat (2009), as famílias wavelet, em geral, são apresentadas em pares constituídos de uma função wavelet $\omega(\cdot)$ e uma função escala $\phi(\cdot)$ (originais). As famílias de Haar, de Daubechies, de Coifelet e de Symelet são exemplos que, comumente, são utilizadas na decomposição de sinais temporais, antes de sua efetiva modelagem.

Baseado em Kubrusly e Levan (2006) e Teixeira Jr et al. (2015), tem-se que a *decomposição wavelet* de uma série temporal $y(\cdot)$ é dada, genericamente, em (1).

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

Onde: (i) $y_{A_{m_0}}(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot)$ é a CW de aproximação de nível m_0 do estado $y(\cdot)$; $y_{D_m}(\cdot) = \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot)$ é a CW de detalhe de nível m do estado $y(\cdot)$; (ii) $d_{m,n}$ 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 $y(\cdot)$ e $\omega_{m,n}(\cdot)$ (isto é, $d_{m,n} = \sum_{t=1}^T y(t) \omega_{m,n}(t)$) e entre $y(\cdot)$ e $\phi_{m_0,n}(\cdot)$ (isto é, $a_{m_0,n} = \sum_{t=1}^T y(t) \phi_{m_0,n}(t)$), para todo $m_0, m, n \in \mathbb{Z}$.

2.2 Support vector regression

O método preditivo individual *support vector regression* (SVR) é fundamentado nas *support vector machines* (SVMs), propostas por Vapnik (1995). O modelo produzido pelo SVR depende apenas da própria série temporal e de um objetivo chamado de *função custo* (o qual é utilizado para a construção do SVR).

Assim, assumamos que $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$ consiste em uma amostra de dados temporais de treinamento, onde X indica o espaço dos padrões de entrada. Aqui, o objetivo é

encontrar uma função $f(x)$ que apresente, no máximo, um desvio ε dos alvos y_i ($i=1,\dots,l$) de treinamento. A função linear de aproximação $f(x)$ é definida por $f(x)=\langle w, x \rangle + b$, com $w \in X, b \in \mathbb{R}$, onde $\langle \cdot, \cdot \rangle$ denota o produto interno usual sobre o espaço X . Não obstante, nem sempre é possível garantir a viabilidade do problema, uma vez que existem pontos que violam as restrições. Por conta disso, a formulação em (2), proposta por Vapnik (1995), utiliza variáveis de folga ξ_i e ξ_i^* .

$$\begin{aligned} \text{Minimize} \quad & \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^l \xi_i + \xi_i^* \right) \\ \text{sujeito a:} \quad & \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, l \end{cases} \end{aligned}$$

(2)

Para a construção da formulação dual do problema descrito em (2), são introduzidos alguns multiplicadores de Lagrange (α_i, α_i^*) não negativos, conforme destaca Vapnik (1995), de modo que o mesmo é convertido em um problema de maximização, na sua forma dual, conforme é descrito em (3).

$$\begin{aligned} \text{Maximizar} \quad & \left(-\frac{1}{2} \right) \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ \text{Sujeito a} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \tag{3}$$

Assim, a função linear $f(x)$ de aproximação reescrita como segue:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b$$

A formulação dual do problema do SVR fornece como alternativa fazer modelagens em um espaço de alta dimensionalidade. Adicionalmente, pode-se realizar um mapeamento não linear de dados temporais de entrada para espaços de dimensão maior, nos quais a regressão linear torna-se possível, desde que sejam utilizadas abordagens do SVR baseadas em *funções Kernel* (não lineares), as quais são denotadas, genericamente, por $K(x, x') := \langle \phi(x), \phi'(x') \rangle$, (Smola e Schölkopf (1998)). Consequentemente, o problema descrito em (3) pode é reeditado conforme a formulação em (4).

$$\begin{aligned} \text{Maximizar} \quad & -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ \text{Sujeito a} \quad & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \tag{4}$$

Com efeito, a função de aproximação da SVR consiste em um modelo não-linear e é reeditada tal como segue: $f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b$.

Dentre as funções Kernel mais utilizadas na modelagem SVR, destacam-se: o linear, o polinomial, o sigmoidal e o gaussiano, cujas definições podem ser encontradas em Smola e Schölkopf (1998).

2.3 Least squares support vector regression

O método *least squares support vector regression* (LSSVR) consiste em uma variante do SVR e foi proposto por Suykens e Vandewalle (1999). De acordo com Suykens et al. (2002), a principal modificação do LSSVR, em relação ao SVR de Vapnik (1995), ocorre em dois pontos: (1º) em vez de restrições de desigualdade, são consideradas restrições de igualdade, ou seja, no LSSVR, encontra-se a solução através da resolução de um conjunto de equações lineares, ao invés de um problema quadrático (conforme equação 2); e (2º) a função de perda quadrática é tomada para a variável e_i (erro de aproximação) que desempenham um papel semelhante ao das variáveis de folga ξ_i e ξ_i^* , na formulação do SVR.

Desse modo, seja $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$ uma amostra de dados temporais de treinamento, onde X representa o espaço de padrões de entrada (por exemplo, $X = \mathbb{R}^d$). Aqui, o objetivo é ajustar numericamente um modelo $f(x) = \langle w, x \rangle + b$, com $w \in X$, $b \in \mathbb{R}$, onde $\langle \cdot, \cdot \rangle$ denota o produto interno usual em X , considerando o problema em (5).

$$\begin{aligned} \text{Minimizar} \quad & \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 \\ \text{s.a} \quad & (y_i - (\langle w, x_i \rangle + b)) = e_i, \quad i = 1, 2, \dots, l \end{aligned} \quad (5)$$

Onde: w e b são os vetores de pesos; γ é o parâmetro de regularização; e_i é o i -ésimo erro de aproximação. Para resolver o problema (5), considera-se a função de Lagrange, definida em (6).

$$L(w, b, e, \alpha) = \frac{1}{2} \|w\|^2 + \frac{\gamma}{2} \sum_{i=1}^l e_i^2 - \sum_{i=1}^l \alpha_i (\langle w, x_i \rangle + b + e_i - y_i) \quad (6)$$

Onde α_i ($i=1, \dots, l$) são os multiplicadores de Lagrange, que podem ser positivos ou negativos devido às restrições de igualdade como decorre das condições de Kuhn-Tucker, Fletcher (1987).

No tocante às condições de otimalidade da equação (6), as mesmas são dadas em (7).

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^l \alpha_i x_i \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i, \quad i = 1, 2, \dots, l \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow w^T x_i + b + e_i - y_i = 0 \quad i = 1, 2, \dots, l \end{cases} \quad (7)$$

Onde, eliminando-se os vetores w e e , obtém-se um sistema de equações lineares, descrito em (8), ao invés de um problema de programação quadrática (descrito em (2)).

$$\begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 & I_v^T \\ I_v & K + \gamma^{-1}I \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$$

(8)

Onde: $y := [y_1, y_2, \dots, y_l]^T$, $I_v := [1, 1, \dots, 1]^T$, $\alpha := [\alpha_1, \alpha_2, \dots, \alpha_l]^T$ e $K(x, x') := \langle \phi(x), \phi'(x) \rangle$, que é a função Kernel (que satisfaz às condições de Mercer, (Smola e Schölkopf (2004))), que realiza um mapeamento não linear dos dados de entrada para um espaço de dimensão maior, onde a regressão linear torna-se possível.

Com efeito, o modelo LSSVR é definido em termos dos coeficientes α , b e da função Kernel (K). Genericamente, tem-se que $f(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b$. Para mais detalhes, veja ainda Suykens et al. (2002) e Guo; Bai (2009).

2.4 Combinação linear de previsões

A combinação linear de previsões, de acordo com Clemen (1989), consiste em uma metodologia atraente na obtenção de previsões mais acuradas que as oriundas de previsores individuais. Assim o sendo, ao invés de se escolher o melhor método preditivo individual, baseado em algum critério de seleção, o problema agora passa a ser definir quais os previsores individuais a serem utilizados na obtenção das previsões combinadas

De acordo com Teixeira Jr (2013), as previsões linearmente combinadas necessitam de algum mecanismo ponderador tal que vise à minimização da variância dos resíduos combinados ou outra função residual. Ainda de acordo com o autor, os pesos adaptativos lineares (responsáveis por ponderar as previsões geradas por previsores individuais) podem ser fixos ou variáveis e, além disso, podem não necessariamente ser positivos ou somar uma unidade. Para mais detalhes, veja Faria e Mubwandarikwa (2008) e Teixeira et al. (2015).

3. METODOLOGIA PROPOSTA

Seja $(y(t))_{t=1}^T$ uma série temporal de cardinalidade T a ser prevista. O método híbrido proposto é executado de acordo com os quatro passos seguintes.

Passo 1 - Decomposição wavelet da série original: uma decomposição wavelet de nível r , (Teixeira Júnior et al., 2015) da série temporal $(y(t))_{t=1}^T$ é realizada, gerando-se $r+1$ CWs - isto é, uma CW de aproximação de nível m_0 , denotada por $(y_{A_{m_0}}(t))_{t=1}^T$, e r CWs de detalhe de níveis $m_0, m_0+1, \dots, m_0+(r-1)$, denotadas por $(y_{D_{m_0}}(t))_{t=1}^T, (y_{D_{m_0+1}}(t))_{t=1}^T, \dots, (y_{D_{m_0+(r-1)}}(t))_{t=1}^T$ respectivamente.

Na prática, usualmente se adota para o parâmetro de nível m_0 valor igual ao nível de decomposição r . As CWs de aproximação e de detalhe foram calculadas no software *Matlab 8.0*, o qual oferece várias opções de bases wavelet, tais como: as famílias de Haar, Daubechies, coiflets e symlets.

Passo 2 - Modelagem das componentes wavelet via SVR e LSSVR: as CW obtidas no passo 1 são modeladas através de um SVR e de um LSSVR.

Para a obtenção dos modelos individuais SVR, foi utilizada uma biblioteca do software *Matlab 8.0*, denominada LIBSVM (versão 3.2), a qual foi desenvolvida por Chang e Lin (2015) e se encontra disponível em <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

Para modelagem via LSSVR, foi utilizada uma biblioteca para *Matlab 8.0* denominada LS-SVMlab, (versão 1.8), a qual se encontra disponível no link <http://www.esat.kuleuven.be/sista/lssvmlab/>. Esta foi desenvolvida por Brabanter et al. (2011), baseado nos trabalhos de Suykens et al. (2002) e de Suykens e Vandewalle (1999).

Passo 3 - Combinação linear das previsões dos métodos preditivos SVR e LSSVR: Uma vez obtidas as previsões da componente wavelet de aproximação e as r previsões das componentes de detalhe, na etapa anterior, através dos métodos SVR e LSSVR, faz-se a combinação linear das previsões, com o objetivo de se obter a previsão da componente de aproximação, e de cada uma das r componentes de detalhe.

Assim, para cada uma das CW, a referida combinação é dada conforme a Equação (9).

$$(CL_{CW_{SVR-LSSVR}})_{t=T'}^T = \rho_1 \times (y_{CW,SVR_{CW}}(t))_{t=T'}^T + \rho_2 \times (y_{CW,LSSVR_{CW}}(t))_{t=T'}^T \quad (9)$$

Onde: $y_{CW,SVR_{CW}}$ são as componentes wavelet previstas por SVR, $y_{CW,LSSVR_{CW}}$ são as componentes wavelet LSSVR, ρ_1 e ρ_2 denotam os pesos adaptativos cujos valores ótimos são determinados através da solução do problema de programação não linear, em 10.

$$\begin{aligned} \text{Min } MSE &= \frac{1}{T'} \sum_{t=T'}^T (y_{CW} - CL_{CW_{SVR-LSSVR}})^2 \\ \text{s.a. } &\rho_1 \text{ e } \rho_2 \text{ são irrestritas} \end{aligned} \quad (10)$$

Passo 4 – Soma das combinações lineares: somam-se, para cada instante t , as previsões linearmente combinadas obtidas no passo anterior, conforme a Equação (11).

$$(\hat{y}_H(t))_{t=T''}^{T+h} = CL_{A_{m_0}, SVR-LSSVR} + CL_{D_{m_0}, SVR-LSSVR} + CL_{D_{m_0+1}, SVR-LSSVR} + \dots + CL_{D_{m_0+(r-1)}, SVR-LSSVR} \quad (11)$$

Nesta etapa as previsões linearmente combinadas são somadas para cada instante t , de forma a gerar as previsões combinadas, dentro e fora da amostra, que será denotado por $(\hat{y}_H(t))_{t=T^n}^{T+h}$.

Os passos do método híbrido proposto aqui são esquematizados no fluxograma da Figura 1.

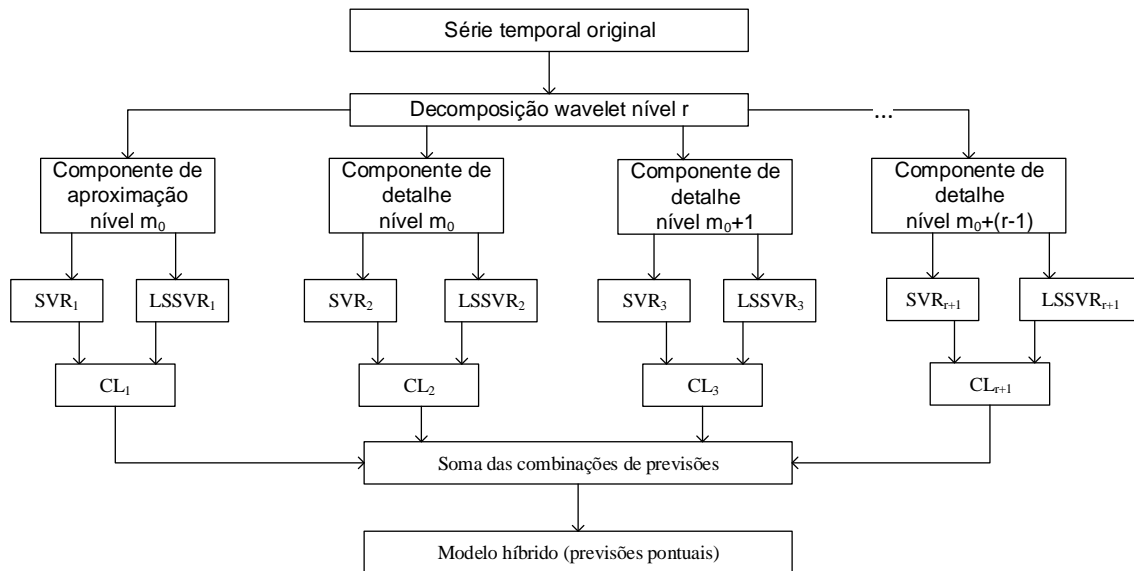


Figura 1 - Fluxograma da metodologia híbrida proposta.

4. RESULTADOS NUMÉRICOS

Nesta seção, a série temporal anual do Canadian Lynx (lince canadense) foi usada para mostrar a eficácia e o poder preditivo do método híbrido proposto no processo de geração de previsões. A série temporal supracitada se refere ao número de lincos preso por ano no distrito de Rio Mackenzie do Norte do Canadá para o período 1821-1934 e pode ser obtida em <https://datamarket.com/>. Existem na literatura vários trabalhos que fazem análise e previsão da série de tempo Canadian Lynx, credita-se, porém, a Moran (1953), como sendo o primeiro trabalho de análise desta. O gráfico da série Canadian Lynx (com 114 observações, relativas ao período 1821-1934) é exibida graficamente, na Figura 2. Pode-se observar uma periodicidade de aproximadamente 10 anos, conforme destaca Zhang (2003).

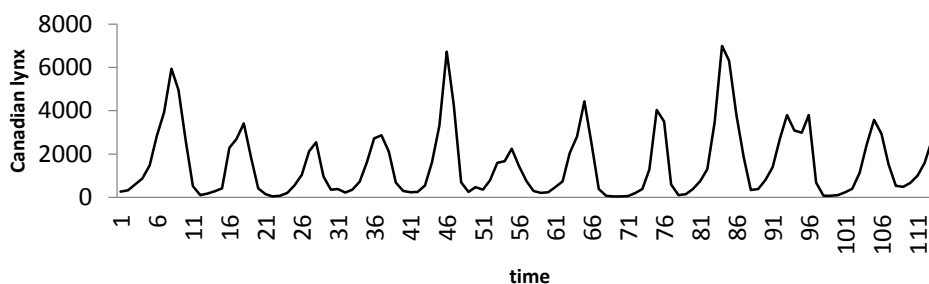


Figura 2 - Série temporal Canadian Lynx (1821-1934).

Neste experimento, foram produzidas previsões um passo à frente, com um horizonte de previsão de 14 passos à frente (ou seja, $h = 14$), já que o presente método é comparado aos resultados de outros autores, que se referem exclusivamente a este tipo de projeção. Além disso, é considerada a versão logarítmica de base 10 da série temporal do Canadian Lynx,

para fins de modelagem. Este conjunto de dados possui cardinalidade 114 observações, sendo que as primeiras 100 observações foram utilizadas para treinamento, enquanto que as 14 remanescentes, para teste. Para tanto, no primeiro passo, foi utilizada uma decomposição wavelet de nível 2, gerando uma CW de aproximação de nível 2 (cuja notação é A2) e duas CWs de detalhe de níveis 2 e 3 (cujas notações são respectivamente D2 e D3). A base wavelet escolhida para a referida decomposição foi a da família de Daubechies com momento nulo igual a 8 (cuja notação é dada por “db 8”). As 3 componentes wavelet geradas podem ser visualizadas na Figura 3.

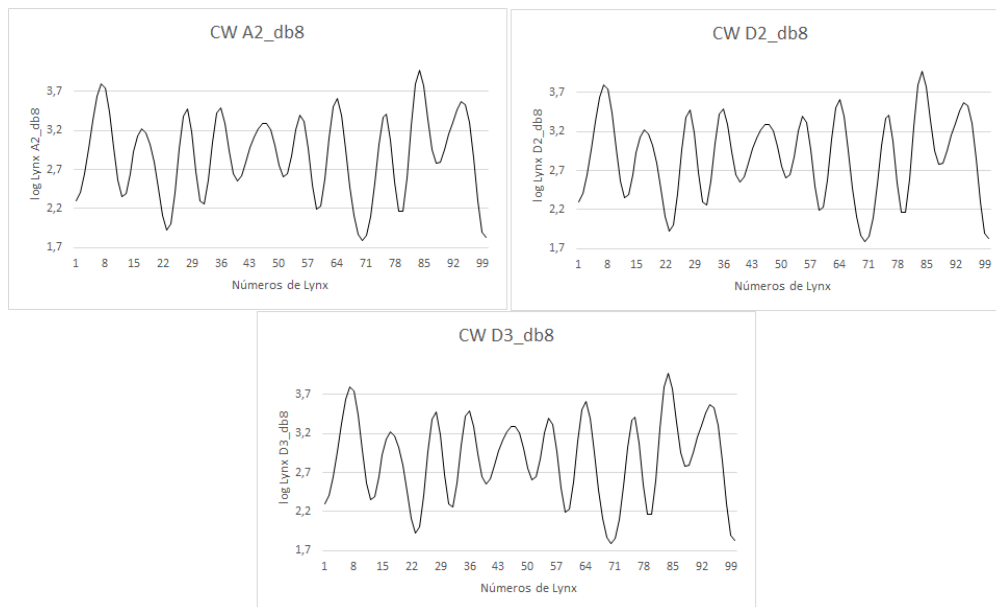


Figura 3: CWs da decomposição wavelet de nível com base db8.

Quanto à modelagem SVR, foi utilizado o Kernel polinomial, definido por $(\gamma * \langle u, v \rangle + \text{coef } 0)^{\text{degree}}$ e o Kernel RBF, dado por $\exp(-\gamma * \|u - v\|^2)$. É conhecido que o desempenho de generalização de um modelo SVR depende de um bom ajuste do parâmetro de regularização (C) e do parâmetro ε da função perda, bem como dos parâmetros relacionados ao Kernel, escolhido conforme Smola e Schölkopf (1998).

A biblioteca utilizada aqui permite que os parâmetros sejam definidos pelo usuário. A fim de melhor avaliar o desempenho da abordagem proposta, para cada CW a ser modelada, os parâmetros C e ε , e os parâmetros do Kernel foram configurados em uma busca por grade (grid search) conforme Hsu et al. (2010), e selecionado aquele que apresentou um melhor desempenho de validação cruzada 10 vezes 10-fold no conjunto de treino, (Viana et al., (2009)).

Para o LSSVR, é necessário o ajuste de dois parâmetros, o de regularização (γ) e o do Kernel (σ^2), sendo que nesse caso somente o Kernel gaussiano foi utilizado, o qual é dado por $\exp\left(-\frac{\|u - v\|^2}{2\sigma^2}\right)$. A fim de se determinar a melhor configuração, foi programado um algoritmo de busca por grade e selecionado aquele que apresentou um melhor desempenho de validação cruzada 10 vezes 10-fold no conjunto de treinamento.

Os resultados selecionados, após o treino baseados no menor Cross Validation Mean squared error (CVMSE) (Ibrahim e Wibowo, (2014)) relativos à modelagem SVR podem ser visualizados na Tabela 1; e os relativos à modelagem LSSVR, na Tabela 2

Tabela 1- Resultados dos parâmetros selecionados SVR.

Série	Kernel	Gamma	Degree	coef0	Cost	epsilon	janela
A2	Polinomial	0.36	3	3	121	0.001	10
D2	Polinomial	0.01	3	4	1094	0.01	10
D3	Polinomial	0.21	3	4	73	0.01	10

Tabela 2- Resultados dos parâmetros selecionados na modelagem LSSVR.

Série	Kernel	γ	σ^2	janela
A2	Rbf	47345	40.76	4
D1	Rbf	5157098	21742.47	10
D2	Rbf	82012812838944	1865591127	10

Os resultados estatísticos obtidos pelo método proposto são exibidos, na Tabela 3, e comparados 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 e 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) Karnabopathy 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 e Shabri (2014), que usaram Máquinas de Vetores Suporte (SVM) para Regressão e SVM com mínimos quadrados (LSSMV).

Quando comparado com os resultados de outros métodos nas mesmas condições, a metodologia híbrida proposta alcançou melhores resultados que outros 18 métodos preditivos competidores, em relação às medidas de acurácia MSE (mean square error) e MAE (mean absolute error), que são definidas em Hamilton (1994). Note que, em relação ao segundo melhor método competidor (a saber, LSSVM), o ganho relativo na estatística MSE foi de 59%; enquanto que foi de 37% na métrica MAE.

Por sua vez, a Figura 4 mostra os valores de previsão na amostra de teste, com as últimas 14 observações. A linha contínua representa os dados reais da série subjacente, enquanto que a linha pontilhada, as previsões produzidas pelo método proposto. É possível

verificar que as previsões e as observações exibem alta correlação, evidenciado pela proximidade de seus pontos no plano.

Tabela 3 - Comparativo entre métodos de previsão aplicados à série temporal log 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 et al. (2005)	FFNN	0,0090	-
	SETAR	0,0140	-
Aladag et al. (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	-
	ARIMA	0,01285	-
Adhikari; Agrawal (2013)	SVR	0,05267	-
	ENSAMBLE	0,00715	-
Ismail; Shabri (2014)	SVR	0,00850	0,07460
	LSSVM	0,00300	0,04180
Os autores	Método Proposto	0,00123	0,02626

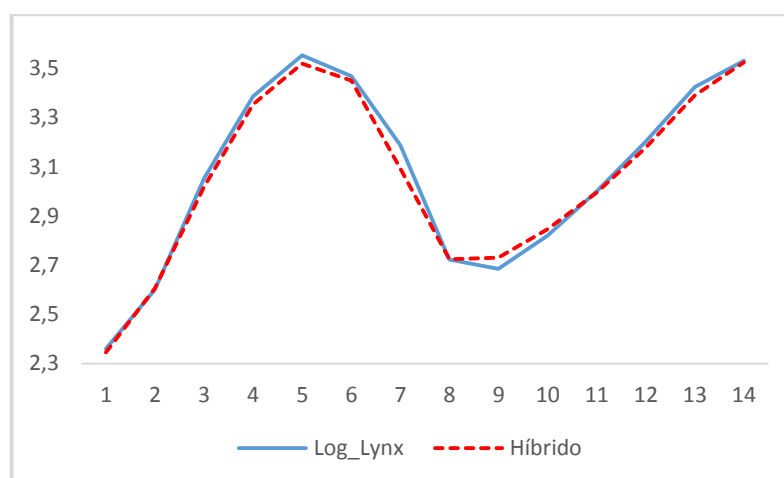


Figura 4: Observações e previsões, na amostra de teste

5. CONCLUSÃO

Nesse trabalho, foi proposta um método híbrido para previsão de séries temporais usando a decomposição Wavelet, a modelagem via Support Vector Regression (SVR) e Least Square Support Vector Regression (LSSVR), programação não linear e a combinação de previsões.

A eficácia do método híbrido proposto foi verificada por meio da aplicação na modelagem da série temporal Canadian Lynx, e comparada com diversas outras modelagens,

nas mesmas condições, a qual mostrou-se superior quando comparadas as medidas de acurácia MSE e MAE. É importante salientar que, embora os fundamentos teóricos associados à metodologia proposta sejam complexo, a sua implementação é relativamente simples, quanto utilizados os programas e pacotes mencionados no texto.

AGRADECIMENTOS

Às instituições: Itaipu Binacional, PPGMNE (Programa de Pós Graduação em Métodos Numéricos em Engenharia), CEASB (Centro de Estudos Avançados em Segurança de Barragens), FPTI (Fundação Parque Tecnológico Itaipu).

REFÊRENCIAS

- 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. Elsevier Ltd.
- ANANDHI, V.; CHEZIAN, R. M. Support Vector Regression to Forecast the Demand and Supply of Pulpwood. **International Journal of Future Computer and Communication**, v. 2, n. 3, p. 1–4, 2013.
- BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. **Applied Soft Computing**, v. 23, p. 27–38, 2014. Elsevier B.V.
- BRABANTER, K. DE; KARSMARKERS, P.; OJEDA, F.; et al. **LS-SVMlab Toolbox User 's Guide**. 2011.
- CHANG, C.-C.; LIN, C.-J. LIBSVM -- A Library for Support Vector Machines. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>>. Acesso em: 2/1/2015.
- CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, 1989.
- FARIA, A. E.; MUBWANDARIKWA, E. The geometric combination of bayesian forecasting models. **Journal of Forecasting**, v. 27, n. 1, p. 519–535, 2008.
- FLETCHER, R. Practical methods of optimization; (2nd ed.). , 1987. Wiley-Interscience.
- GRANGER, C. W. J. Combining forecasts – twenty years later. **Journal of Forecasting**, v. 8, p. 167–73, 1989.
- GUO, Z.; BAI, G. Application of Least Squares Support Vector Machine for Regression to Reliability Analysis. **Chinese Journal of Aeronautics**, v. 22, n. 2, p. 160–166, 2009. Chinese Journal of Aeronautics.
- HAMILTON, J. D. **Time Series Analysis**. Princeton University Press, 1994.
- HONG, W.-C. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. **Energy**, v. 36, n. 9, p. 5568–5578, 2011.
- HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. **A Practical Guide to Support Vector Classification**. 2010.
- IBRAHIM, N.; WIBOWO, A. Support vector regression with missing data treatment based variables selection for water level prediction of Galas River in Kelantan Malaysia. **WSEAS Transactions on Mathematics**, v. 13, p. 69–78, 2014.
- 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.
- 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. Elsevier Ltd.

- KOSANAN, O.; KANTANANTHA, N. Thailand's Para Rubber Production Forecasting Comparison. Proceedings of the International MultiConference of Engineers and Computer Scientists 2014. **Anais...** v. II, 2014. Hong Kong.
- KUBRUSLY, C. S.; LEVAN, N. Abstract wavelets generated by Hilbert space shift operators. **Advances in mathematical and applications**, v. 16, p. 643–660, 2006.
- MALLAT, S. **A Wavelet Tour of Signal Processing: The Sparse Way**. Third Edit ed.2009.
- MANCUSO, A. C. B.; WERNER, L. Review of Combining Forecasts Approaches. **Independent Journal of Management & Production**, v. 4, n. June, p. 248–277, 2013.
- MORAN, P. The statistical analysis of the Canadian Lynx cycle. **Australian Journal of Zoology**, v. 1, n. 3, p. 291, 1953.
- MORETTIN, P. A.; TOLOI, C. M. DE C. **Análise de séries temporais**. Edgard Blucher, 2006.
- RODRIGUES, S. B.; TEIXEIRA, L. L.; CARLOS, A.; SIQUEIRA, P. H. Combinação Linear De Redes Neurais Artificiais E Máquinas De Vetores De Suporte Para Regressão Nas Previsões De Vazões Mensais No Posto 266-Itaipu. XVII Encontro de Modelagem Computacional V Encontro de Ciência e Tecnologia de Materiais Universidade Católica de Petrópolis (UCP). **Anais...**, 2014.
- RODRIGUES, S. B.; TEIXEIRA, L. L.; TEIXEIRA JÚNIOR, L. A.; et al. Forecasting of the relative displacements in block I11 of the Itaipu hydroelectric plant dam through a ARIMA-SVR hybrid approach. **Espaço Energia**, v. 22, n. 22, p. 28–37, 2015.
- SANG, Y.-F. A review on the applications of wavelet transform in hydrology time series analysis. **Atmospheric Research**, v. 122, p. 8–15, 2013.
- SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, 1998.
- SMOLA, A.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, p. 199–222, 2004.
- SUYKENS, J. A. K.; GESTEL, T. VAN; BRABANTER, J. DE. **Least Squares Support Vector Machines**. 2002.
- SUYKENS, J. A. K.; VANDEWALLE, J. Least Squares Support Vector Machine Classifiers. **Neural Processing Letters**, v. 9, n. 3, p. 293–300, 1999. Kluwer Academic Publishers.
- TEIXEIRA JR, L. A. **Combinação SSA-Wavelet de Métodos Preditivos com Ajuste Numérico MINIMAX, na Geração de Previsões e de Cenários**, 2013. Puc-Rio.
- 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. Congresso Latino-Iberoamericano de Investigación Operativa e Simpósio Brasileiro de Pesquisa Operacional. **Anais...** p.1401–1412, 2012.
- TEIXEIRA JR, L. A.; SOUZA, R. M.; MENEZES, L. M.; et al. Artificial Neural Network and Wavelet decomposition in the Forecast of Global Horizontal Solar Radiation. **Sobrapo**, v. 35, n. 1, p. 1–16, 2015.
- TEIXEIRA, L. L.; RODRIGUES, S. B.; SIQUEIRA, P. H.; SILVA, A. C. L. DA; TEIXEIRA JR, L. A. Comparativo Entre Arima, Redes Neurais Artificiais, Máquinas De Vetores Suporte Para Regressão E Combinação Linear De Previsões De Vazões Mensais No Posto 266- Itaipu. Simpósio de Engenharia de Produção. **Anais...** p.1–15, 2014. XXI SIMPEP.
- TEIXEIRA, L. L.; TEIXEIRA JUNIOR, L. A.; SIQUEIRA, P. H. Forecasts of monthly flows in the post-266 Itaipu using neural networks with wavelet shrinkage. **Espaço Energia**, v. 22, n. 22, p. 8–17, 2015.
- VAPNIK, V.; GOLOWICH, S. E.; SMOLA, A. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. **Advances in Neural Information Processing Systems**, v. 9, p. 281–287, 1996.
- VAPNIK, V. N. **The Nature of Statistical Learning Theory**. 1995.
- VIANA, F. A. C.; HAFTKA, R. T.; STEFFEN, V. Multiple surrogates: How cross-validation errors can help us to obtain the best predictor. **Structural and Multidisciplinary Optimization**, v. 39, n. 4, p. 439–457, 2009.
- WALLIS, K. F. Combining forecasts : forty years later. **Applied Financial Economics**, p. 33–41, 2011.
- 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.