

Non-Stationary Demand Forecasting Based on Empirical Mode Decomposition and Support Vector Machines

I. D. da Silva, M. C. Moura, I. D. Lins, E. López Droguett and E. Braga

Abstract— A company performance strongly depends on its ability of delivering the right quantity of a given product at the time customers require. Even though some demand forecasting techniques have been developed, they have commonly used simplifying assumptions that limit their use like assuming that the relation between the inputs and the output is linear, for example. Therefore, machine-learning techniques, such as Support Vector Machines (SVM), arise as a promising alternative for accomplishing demand forecasting. SVM has the advantage of performing well in cases where the relationship between input and output data is unknown, and thus has brought good results when applied in different contexts. However, SVM presents some limitations in predicting non-stationary series. In this context, a method called Empirical Mode Decomposition (EMD) has been adopted for decomposing non-stationary and nonlinear time series into a group of Intrinsic Mode Functions (IMFs). Moreover, SVM performance strongly depends on the values of real-valued parameters, which need to be tuned to enhance the predictive ability of the model. This situation gives rise to the model selection problem, which may be solved by heuristics such as Particle Swarm Optimization (PSO). Therefore, this work proposes a non-stationary demand forecasting methodology based on EMD-PSO-SVM. An example in the context of the food industry is presented and we compare the results obtained by the proposed methodology against the ones returned from a plain PSO-SVM. The results show that the proposed EMD-PSO-SVM presented superior performance.

Keywords— Demand Forecasting, Non-Stationary Time Series, Empirical Mode Decomposition, Support Vector Machines.

I. INTRODUÇÃO

A REALIZAÇÃO de previsões de demanda é essencial nas organizações. As decisões sobre quanto de cada produto deve ser produzido, em que momento essa produção deve ocorrer, para onde esses produtos devem ser enviados, são direcionadas pelo conhecimento do comportamento da demanda no tempo. Sendo assim, é de grande importância aplicar técnicas e ferramentas que permitam às empresas anteverem-se à demanda e serem capazes de oferecer o que é requerido com alto nível de serviço. Algumas técnicas voltadas à realização dessas previsões são apontadas por [1].

Em diferentes contextos, a utilização de métodos de previsão de séries temporais com boa precisão já é bastante comum, como os modelos: AR (*Autoregressive*), ARIMA (*Autoregressive Integrated Moving Average*), MA (*Moving-*

Average), NN (*Neural Network*). Com exceção do método NN, os demais citados consideram que a relação entre as entradas e a saída é linear, o que, em geral, não corresponde à realidade ([2], [3]). A técnica *Support Vector Machines* (SVM) tem sido utilizada para realizar previsões do comportamento de variáveis nos contextos da engenharia, de finanças, da hidrologia, da oceano-meteorologia, dentre outros ([4], [5], [6]). De acordo com [7], SVM apresenta uma série de vantagens se comparado a outras técnicas de aprendizado (por exemplo, NN), como: maior capacidade de generalização; estrutura matemática capaz de fornecer uma solução única e global; maior robustez e treinamento feito rapidamente favorecendo a construção de melhores modelos de estimação.

SVM baseia-se na Teoria de Aprendizado Estatístico [8] e é caracterizado como um método de aprendizado supervisionado, pois é desenvolvido a partir de um conjunto de dados de entrada e de saída já conhecido. Essa técnica resolve um problema de classificação, caso as saídas representem categorias, ou de regressão, caso as saídas assumam valores contínuos. Esse método é muito útil quando se tem um conjunto de dados a respeito de determinado fenômeno, porém não se conhece como as entradas são mapeadas em saídas. Sua ideia principal é a realização de um treinamento, a partir dos dados de entrada e de saída conhecidos, permitindo o aprendizado sobre o comportamento desses dados. O treinamento consiste na resolução de um problema de programação matemática convexa e quadrática e o modelo resultante deve ser capaz de prever resultados futuros a partir de novas observações.

A eficiência do SVM depende de quais valores dos parâmetros aparecem no problema de treinamento. Sendo assim, um problema a ser resolvido é a definição dos parâmetros que são melhores e mais adequados para o modelo. Para solucionar esse problema, há diversas técnicas que podem ser usadas: *grid search* [9], busca baseada em derivadas [10], meta-heurísticas probabilísticas como algoritmos genéticos e *Particle Swarm Optimization* (PSO) ([11], [12]). Nesse trabalho, será utilizado o método PSO por não requerer conhecimento de derivadas, ser adequado para problemas com variáveis de decisão contínuas (no caso, os parâmetros do SVM) e por sua simplicidade de implementação [13].

Além disso, estudos recentes têm mostrado que a aplicação da ferramenta SVM associada à técnica EMD ([14], [15], [16]) (*Empirical Mode Decomposition*) melhora a capacidade do SVM em realizar previsões. O EMD [17] tem se mostrado bastante eficiente quando se trabalha com séries não

I. D. da Silva, Universidade Federal de Pernambuco, italla@hotmail.com
M. C. Moura, Universidade Federal de Pernambuco, marcio@ceerma.org
I. D. Lins, Universidade Federal de Pernambuco, isis.lins@ceerma.org
E. L. Droguett, Universidad de Chile, elopezdroguett@ing.uchile.cl

E. Braga, Universidade Federal de Pernambuco, braga.eas@gmail.com
(Corresponding author: I. D. Lins)

estacionárias e não lineares, pois se trata de um método adaptativo, ou seja, que se adapta às características locais da série sob análise. EMD caracteriza-se, ainda, por trabalhar no próprio espaço de tempo e por ser uma técnica direta e intuitiva. A principal característica desse método é a capacidade de identificar os modos de oscilação intrínseca pela escala temporal própria dos dados empíricos e, de acordo com isso, realizar a decomposição dos dados.

O objetivo deste trabalho é, portanto, explorar a utilização conjunta das técnicas EMD-PSO-SVM como ferramenta para a realização da previsão de demanda de produtos, bem como comparar seu desempenho em relação ao método PSO-SVM apenas. Para isso, um exemplo no contexto de uma empresa do ramo alimentício no Nordeste do Brasil será utilizado para avaliar o desempenho da metodologia proposta. Um histórico da demanda agregada dos dois produtos considerados mais relevantes para a empresa, representada pelo volume de vendas, é usado como entrada dos métodos.

O restante do trabalho está organizado da seguinte forma. A Seção II apresenta a fundamentação teórica sobre EMD, PSO e SVM. Na Seção III é descrita a metodologia proposta. Na Seção IV, a aplicação a um caso real é exposta. A Seção V conclui o trabalho.

II. FUNDAMENTAÇÃO TEÓRICA

A. Empirical Mode Decomposition

O método EMD decompõe uma série de tempo em um conjunto de funções ortogonais e em um resíduo monotonicamente crescente ou decrescente ou caracterizado por apresentar apenas um extremo. A decomposição baseia-se na ideia de que, para qualquer instante de tempo, a série é composta por vários modos de oscilação simples de frequências diferentes, que estão sobrepostas umas às outras [18]. Essas funções ortogonais, conhecidas como funções de modo intrínseco (*Intrinsic Mode Functions* - IMF), correspondem a uma descrição física do tempo-frequência-energia da série analisada, onde um IMF é caracterizado por dois aspectos: (1) em todo o conjunto de dados, o número de extremos (máximos e mínimos) deve ser igual ao número de cruzamentos no zero, ou devem diferir em apenas uma unidade; (2) para qualquer dado, a média do envelope definido a partir dos valores máximo e mínimo deve ser igual a zero.

O processo que dá origem aos IMFs é conhecido pelo termo em inglês *sifting* e ocorre conforme as etapas a seguir:

1. Para uma dada série de tempo $x(t)$, identificar todos os extremos (máximos e mínimos) locais;
2. Conectar todos os extremos (máximos e mínimos) locais identificados através de uma *cubic spline*, gerando os envelopes inferior e superior (e_{min} , e_{max} , respectivamente);
3. Calcular a média dos envelopes superior e inferior, $m(t) = (e_{min} + e_{max})/2$;
4. Subtrair $m(t)$ da série original e usar o resultado, $h(t) = x(t) - m(t)$ como um novo sinal, conhecido como proto-IMF. Se o novo sinal não atender às exigências de um IMF, volta-se ao passo 1, utilizando a série $h(t)$ como sendo a série analisada, ou seja, $x(t) = h(t)$.

5. No entanto, se $h(t)$ atender às exigências de um IMF, este passará a ser um IMF, caracterizado como $c_i(t) = h(t)$, referente ao i -ésimo IMF. Assim, a série original é subtraída desse IMF, formando a nova série $r(t) = x(t) - c_i(t)$;
6. Retornar ao passo 1, utilizando a nova série $r(t)$ e guardar o $c_i(t)$ como um IMF.

Portanto, o resultado do processo de decomposição é:

$$x(t) = \sum_{j=1}^z c_j + r \quad (1)$$

onde z é o número de IMFs.

O *sifting* é realizado com o propósito de tornar a onda mais simétrica no que se refere ao zero. Desse modo, o *sifting* deve ser repetido quantas vezes forem necessárias para atender a essas condições. No entanto, ao mesmo tempo em que a repetição do *sifting* faz com que o proto-IMF se aproxime ainda mais da definição do IMF, essa repetição produz uma suavidade na amplitude da onda, produzindo IMFs de amplitude constante e, muitas vezes, sem significado físico. Sendo assim, esse processo não pode ser realizado indefinidamente, apontando para a necessidade de que um critério de parada seja adotado e, portanto, a repetição do *sifting* seja limitada a um número máximo de vezes.

Um dos critérios adotados é o de convergência de teste de Cauchy. Segundo esse critério, o valor absoluto da diferença entre duas operações de *sifting* consecutivas elevado ao quadrado, conforme indica a Equação 2, deve ser menor que um valor pré-definido.

$$SD_k = \sum_{t=0}^T \left[\frac{|(h_{1(k-1)}(t) - h_{1k}(t))|^2}{h_{1(k-1)}^2(t)} \right]. \quad (2)$$

Se SD_k estiver entre 0,2 e 0,3, é possível encontrar o primeiro IMF [12]. Além disso, mostrou-se que, caso o SD_k esteja nesse intervalo, os IMFs encontrados terão significado físico. Contudo, ao se usar o critério de Cauchy bem como suas variantes [18], não se tem garantia de que os IMFs encontrados estão realmente em conformidade com a definição do IMF, já que essa abordagem não guarda nenhuma relação direta com os dois aspectos que caracterizam um IMF ([18], [19]).

Com o objetivo de solucionar os problemas observados com o uso do critério de parada de Cauchy para o processo *sifting*, Huang *et al.* [20] enunciou o critério S -número, de acordo com o qual, o processo *sifting* finaliza quando o número de extremos e de cruzamentos no zero tornam-se iguais ou diferem em apenas uma unidade e se mantêm os mesmos por S vezes consecutivas. Desse modo, o S -número corresponde ao número de iterações sucessivas em que essas condições são satisfeitas. A escolha de S -números elevados requer que o processo de *sifting* ocorra mais vezes, o que pode comprometer o significado físico das séries encontradas. Sendo assim, o ideal é que o S -número não seja um valor muito elevado. Valores no intervalo $3 \leq S \leq 5$ têm apresentado resultados satisfatórios [20].

B. Support Vector Machines

SVM é caracterizado como um método de aprendizado de máquina supervisionado, pois tem por base um conjunto de exemplos (x_i, y_i) . O problema de aprendizado do SVM pode ser caracterizado da seguinte forma [21]: existe uma relação de dependência não linear não conhecida entre um vetor de entrada multidimensional x e um vetor de saída y , em que a única informação disponível é um conjunto de dados $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$.

Durante a fase de aprendizado, também conhecida como fase de treinamento, SVM soluciona um problema de otimização quadrática e convexa, para o qual as condições de Karush-Kuhn-Tucker (KKT) são necessárias e suficientes para a otimalidade global. A função objetivo desse problema inclui o princípio da minimização do risco estrutural (*Structural Risk Minimization* – SRM), o qual corresponde à minimização do limite superior do erro de generalização, envolvendo tanto a capacidade que a máquina apresenta para realizar previsões, a partir de dados de entrada não pertencentes ao conjunto D , bem como o erro de treinamento. O princípio SRM mostrou ser eficiente quando se trabalha com um número reduzido de dados e o objetivo é definir um modelo que tenha capacidade adequada para representar bem o conjunto de treinamento D [8].

Em um problema de regressão, é necessário definir uma função de dependência entre y e x . Para isso, na fase de treinamento, SVM resolve um problema de otimização quadrática convexa, com a formulação dual conforme mostram as Equações 3-6:

$$\max_{\alpha, \alpha^*} -1/2 \sum_i \sum_j (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_i [\varepsilon (\alpha_i + \alpha_i^*) + y_i (\alpha_i - \alpha_i^*)] \quad (3)$$

$$\text{sujeito a: } \sum_i (\alpha_i - \alpha_i^*) = 0 \quad (4)$$

$$0 \leq \alpha_i \leq C \quad (5)$$

$$0 \leq \alpha_i^* \leq C, \quad (6)$$

em que $i, j = 1, 2, \dots, n$, α_i e α_i^* são as variáveis de decisão duais, ε corresponde ao raio do tubo da função perda de ε -insensitividade de Vapnik [8] e o parâmetro C refere-se ao *trade-off* entre a capacidade do modelo em realizar a previsão para novos dados e a acurácia do treinamento. $K(x_i, y_i)$ corresponde à função *kernel*, que é capaz de levar os dados originais, cuja relação entre a saída e as entradas é não linear, a um espaço de maior dimensão (espaço característico), no qual a relação entre eles é expressa linearmente.

Com a resolução do problema de otimização na fase de treinamento, tem-se a função de regressão não linear estimada, segundo a Equação 7:

$$f(x) = \sum_i (\alpha_i - \alpha_i^*) K(x_i, x_j) + b, \quad (7)$$

em que x é uma nova entrada a partir da qual se quer prever a variável resposta de interesse e b é parâmetro obtido a partir da resolução do problema 3-6 [21].

Por apresentar algumas vantagens com relação às demais funções *kernel*, como a de ser uniparamétrica, o *Gaussian*

Radial Basis Function (RBF Gaussiano) é um tipo de *kernel* bastante utilizado, e está representado na Equação 8:

$$K(x_i, x_j) = \exp[-\gamma \|x_i - x_j\|^2]. \quad (8)$$

C. Particle Swarm Optimization

Com a escolha do *kernel* RBF, SVM passa a depender de três parâmetros reais C , ε e γ . A determinação dos valores desses parâmetros influencia diretamente o desempenho apresentado pelo SVM na previsão. Esse problema referente à decisão de quais valores devem ser usados é conhecido como problema de seleção de modelo. Neste trabalho, será utilizado PSO [22], que simula computacionalmente o comportamento de seres vivos que se movimentam em grupos, como bando de pássaros ou cardume de peixes.

O elemento do PSO é uma partícula, que tem como características a posição corrente, a melhor posição já visitada, a velocidade e o *fitness*. Além disso, o conceito de vizinhança é também importante para caracterização do PSO. Se a vizinhança de uma partícula é formada por todas as demais partículas da nuvem, tem-se a abordagem *gbest*. Caso contrário, se a vizinhança é composta apenas por um subconjunto de partículas, tem-se a abordagem *lbest*. Com base nas informações da vizinhança das partículas, PSO usa duas equações de atualização, as quais direcionam a busca pelo ótimo, uma para a velocidade e outra para a posição. Para o leitor interessado, mais detalhes sobre PSO podem ser vistos em [22].

A decisão de quais valores devem ser adotados para os parâmetros do SVM é realizada através de uma avaliação de sua capacidade de previsão. Isso é feito com um conjunto de dados nomeados de dados de validação, a partir do cálculo de medidas de erro. Sendo assim, PSO utiliza parte dos dados disponíveis para realizar o treinamento ou a aprendizagem do SVM e a outra parte das observações é usada para validar o modelo. Os valores de parâmetros que apresentarem o menor erro – proveniente da comparação entre saídas reais e previstas – são os escolhidos. Esses parâmetros caracterizarão o problema de aprendizado do SVM. As medidas de erro *Mean Square Error* (MSE), *Mean Absolute Percentage Error* (MAPE) e o *Mean Absolute Error* (MAE) podem ser usadas como função de ajuste do PSO.

III. METODOLOGIA EMD-PSO-SVM

A metodologia proposta nesse trabalho baseia-se na aplicação conjunta das técnicas EMD, PSO e SVM, com o propósito de realizar a previsão de vendas dos dois produtos mais vendidos de uma empresa do ramo alimentício para três capitais do Nordeste brasileiro. Inicialmente, foi aplicado o EMD para os dados originais de cada capital, para, assim, obter os IMFs e o resíduo correspondentes a cada série. Isso foi feito por meio do *software* R, utilizando a função *emd* do pacote EMD [23], para a qual devem ser especificados os seguintes argumentos: valores de saída ou observados (y), valores de entrada ou período considerado (x), método de ajuste para a fronteira (*boundary*), forma como os envelopes serão construídos (*sm*) e o critério de parada (*stoprule*). Esses parâmetros estão expressos na Fig. 1.

```
emdOut = emd(y, x, boundary="wave", sm="none",
plot.imf=TRUE, stoprule="type2")
```

Figura 1. Exemplo de uso da função emd do pacote EMD do R.

É importante definir um método de ajuste para a fronteira (boundary) pois, é possível que, ao ser feita uma interpolação, haja oscilações ao final da série que a corrompa como um todo [17]. Algumas condições de contorno estão disponíveis no pacote EMD e podem ser adotadas a fim de que isso seja evitado. Por exemplo, adicionam-se ondas artificiais às extremidades do sinal original, formando as chamadas ondas características [17]. Isso é feito a partir da repetição do modo de oscilação dos valores extremos localizados em torno das fronteiras. Há, também, a condição *wave* para o parâmetro boundary. Esse ajuste gera uma nova série caracterizada por construir uma onda a partir de dois extremos subsequentes [23]. No presente trabalho, essa última condição foi a escolhida.

O modo como os envelopes são construídos corresponde ao parâmetro *sm* na função *emd*. Os envelopes podem ser construídos de três formas: interpolação, *spline smoothing*, *kernel smoothing* e fazendo uso de um polinômio de suavização local. Neste trabalho, os envelopes foram construídos através da interpolação, que corresponde a “none”, ou usando “*splinesmoothing*”. As diferentes formas de construção dos envelopes podem ser vistas mais detalhadamente em [24]. Já o argumento *plot.imf* especifica se os IMFs devem ser (TRUE) ou não (FALSE) visualizados graficamente.

O critério de parada a ser utilizado para a metodologia EMD também precisa ser determinado. Há cinco possibilidades de critérios de finalização do processo disponíveis no R. Foram utilizados os critérios “type2” ou “type5” que representam o critério de parada de Cauchy [17] dado pela Equação 2 e o critério *S*-número, respectivamente.

Após realizar a decomposição da série via EMD, os IMFs e resíduo gerados passam a ser entradas para o PSO-SVM, conforme mostra a Fig. 2.

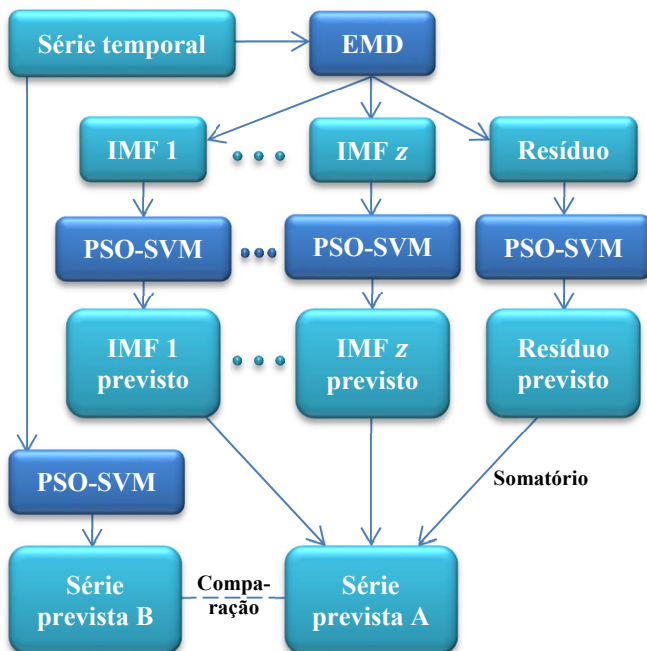


Figura 2. Metodologias EMD-PSO-SVM e PSO-SVM para previsão de série temporal.

A Fig. 2 também ilustra o PSO-SVM, que utiliza como entrada a série temporal original. O resultado do EMD-PSO-SVM, denotado por “série prevista A” pode então ser comparado à saída do PSO-SVM, indicada por “série prevista B” na Fig. 2. O PSO-SVM é aplicado à série original, ou seja, aquela correspondente aos valores realmente observados sem a decomposição pelo EMD. Já com o uso do EMD, o PSO-SVM é usado para prever a série original a partir dos IMFs e do resíduo. Desse modo, foi definido um modelo de previsão com o PSO-SVM para cada IMF e para o resíduo correspondente. Isso porque, após a aplicação do EMD, a previsão para a série real (antes da decomposição) corresponde à soma das previsões para cada componente em que a série foi decomposta conforme Equação 1.

Como já mencionado anteriormente, o desempenho do SVM é bastante dependente do conjunto escolhido de parâmetros (C , ε e γ). Dessa forma, visando otimizar essa escolha, nesse trabalho é utilizada a técnica PSO. Os dados podem ser organizados de acordo com um modelo autorregressivo de primeira ordem, de modo que se considera que $y_t = f(y_{t-1})$, ou seja, o valor da venda na data t é função do valor da própria venda na data $t - 1$. Isso leva à perda de uma observação para as séries originais, para os IMFs e o resíduo.

Após dar entrada dos dados no PSO-SVM, é necessário dividir o conjunto de observações para as três etapas: treinamento, validação e teste. Em seguida, devem ser definidos alguns parâmetros para o PSO: o número de partículas consideradas; o fator de constricção (ϕ) usado com o objetivo de evitar um aumento considerável da velocidade das partículas, os parâmetros c_1 e c_2 , o número máximo de iterações. É preciso, também definir os critérios de parada do PSO e a função objetivo (neste caso, MSE) que guiará o processo de busca pelos melhores valores de parâmetros do SVM. Também devem ser fornecidos os intervalos de valores para cada parâmetro do SVM. Neste trabalho, os critérios de parada adotados para o PSO são: número máximo de iterações, diferença entre valores de MSE provenientes de duas iterações consecutivas menor que uma tolerância pré-especificada e número máximo de iterações com o mesmo ótimo.

Definidos todos os parâmetros requeridos pelo PSO, passa-se a etapa de execução do algoritmo. O PSO-SVM fornece, então, as previsões para as séries consideradas. A série decomposta pelo EMD tem sua previsão definida a partir da soma das previsões dos IMFs e do resíduo, enquanto que a série não decomposta tem sua previsão fornecida de forma direta pelo PSO-SVM. A medida de erro MSE é, então, calculada através da soma das diferenças quadráticas dos valores reais pelos previstos, conforme indica a Equação 9:

$$MSE = \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{n}, \quad (9)$$

onde \hat{y}_i é o valor da previsão fornecido pelo EMD-PSO-SVM ou PSO-SVM.

IV. APLICAÇÃO DE EMD-PSO-SVM NA PREVISÃO DE VENDAS EM UMA INDÚSTRIA DE ALIMENTOS

A. Dados

Os conjuntos de entrada e de saída utilizados no estudo representam o histórico de vendas de iogurtes (em toneladas líquidas) de uma empresa no período de 2005 a 2009, em três capitais do Nordeste do Brasil. Esses dados estão organizados segundo séries mensais de vendas para cada cidade e correspondem à demanda agregada dos dois produtos mais significativos em termos de volume vendido para essa empresa. Os dados estão representados graficamente na Fig. 3. É possível perceber que o nível das vendas na Capital 1 é inferior àquele das Capitais 2 e 3.

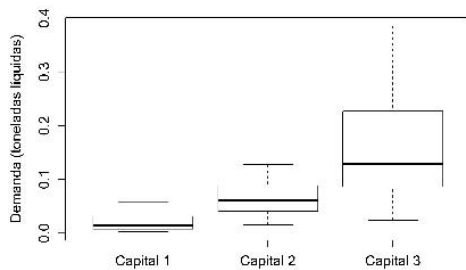


Figura 3. Gráfico boxplot para as capitais.

B. Exemplo de aplicação da metodologia EMD-PSO-SVM

Inicialmente, a aplicação da metodologia EMD-PSO-SVM iniciou-se pelo EMD, cujos argumentos foram especificados conforme descrito na seção anterior. O resultado da decomposição das séries de venda de cada capital é mostrado na Fig. 4. As séries originais, para cada capital estão ilustradas em verde na Fig. 5. Observa-se que elas não oscilam em torno de um valor médio, o que é um indício da presença de não-estacionariedade [25]. Além disso, foi realizado o teste aumentado de Dickey-Fuller [26] para as séries de demanda de cada capital e os valores p resultantes foram $> 0,99$ para a Capital 1, $0,1575$ para a Capital 2 e $0,8001$ para a Capital 3. Portanto, em cada teste, rejeita-se a hipótese alternativa associada à estacionariedade da série.

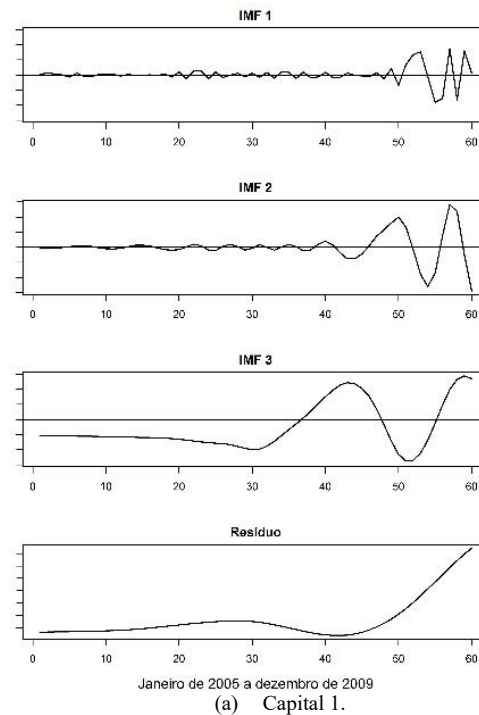
É possível, então, a partir das representações ilustradas na Fig. 4, identificar um comportamento relativamente comum entre os IMFs e o resíduo para cada capital. Para todas as cidades foram obtidos três IMFs e um resíduo. Em geral, os primeiros IMFs das séries apresentam comportamento bastante irregular, correspondendo ao componente da decomposição mais não linear e de maior desordem. Além disso, à medida em que a decomposição é feita, observa-se maior regularidade no comportamento dos IMFs.

Com as séries originais decompostas pelo EMD, os IMFs e o resíduo correspondentes a cada capital foram, pois, utilizados como séries de entrada para o PSO-SVM. Como foi adotado um modelo autorregressivo de primeira ordem, perdeu-se uma observação e, dessa forma, 59 observações foram utilizadas para cada capital.

Os valores das vendas observados no tempo foram divididos nas três fases em que o SVM trabalha: treinamento, validação e

teste. Como todas as séries possuíam o mesmo número de observações, a divisão foi feita com a mesma proporção para todos os casos: aproximadamente 70% para treinamento, 20% para validação e 10% para teste. Sendo assim, foram usados 41 pares de dados na fase de aprendizado da máquina, 12 pares na fase de validação do modelo e 6 pares para observar o comportamento da previsão diante de dados em que a variável resposta não era visível. Além disso, os parâmetros do PSO foram definidos da seguinte forma: número de partículas = 30; número de vizinhos = 8 (*lbest*); número de iterações = 1000; $100 \leq C \leq 1500$; ε definido entre 0,001 e 0,15 da média dos valores de y usados no treinamento; $0,1 \leq \gamma \leq 150$; fator de constricção $\phi = 0,72984$; $c_1 = c_2 = 2,05$; tolerância para dois ótimos consecutivos (critério de parada) = 10^{-9} ; número máximo de iterações com mesmo ótimo = 100. O algoritmo PSO utilizado foi implementado pelos autores [13] em MATLAB e tem interface com o LIBSVM [27].

Após executar o algoritmo PSO-SVM, foram obtidos os modelos para previsão de vendas em cada capital. No caso das séries decompostas pelo EMD, o modelo previsto foi obtido a partir da soma das previsões dadas pelo PSO-SVM para todos os IMFs e o resíduo correspondentes à série específica de cada cidade conforme Equação 1. Já para a série de dados original, a previsão correspondeu àquela já determinada pelo PSO-SVM. O resultado das previsões é mostrado na Fig. 5. A partir da análise gráfica é possível tanto compreender melhor como a demanda de cada cidade varia no tempo como também verificar o quão próximas encontram-se as previsões dos valores reais.



(a) Capital 1.

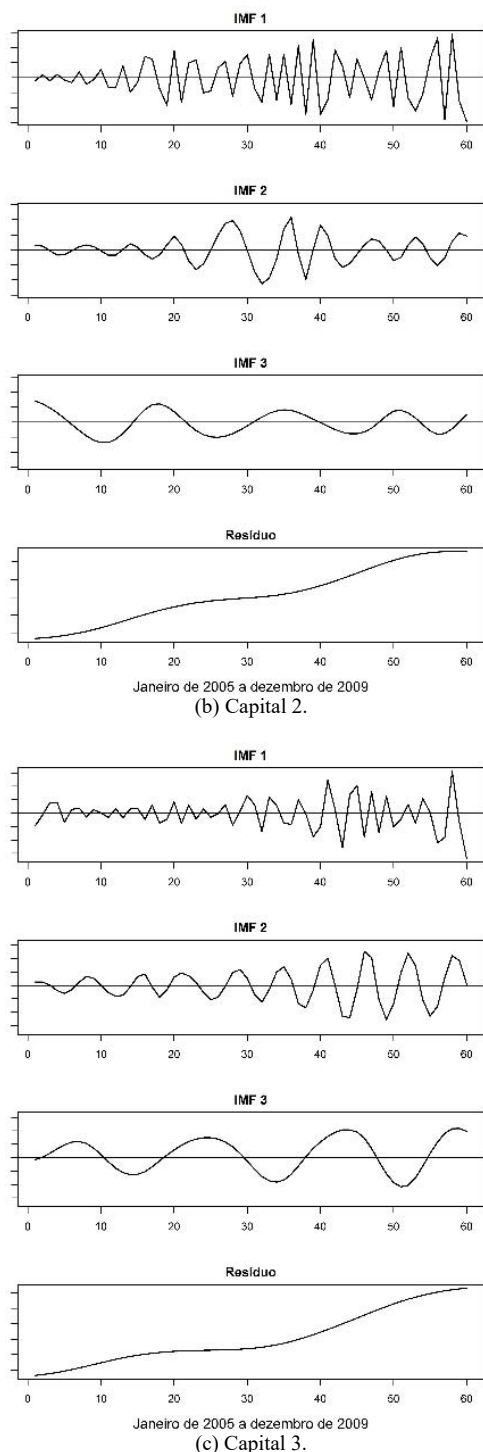


Figura 4. Resultados do EMD.

Em geral, para todas as capitais, verifica-se uma tendência de crescimento ao longo dos anos considerados, o que demonstra uma maior participação dessa empresa no mercado. Percebe-se que, para todas as capitais, as previsões, em geral, acompanham os dados reais e captam o crescimento da demanda. Além disso, vê-se que a série fornecida pelo EMD-PSO-SVM não apresenta grandes distanciamentos em relação à série original. A superioridade do método EMD-PSO-SVM frente ao PSO-SVM fica clara ao se verificar a previsão dada para os valores de demanda finais da Capital 2 (Fig. 5-b). Vale destacar que no caso da Capital 1, entretanto, os períodos finais

dessa cidade não foram devidamente acompanhados por se tratar de um aumento muito grande das vendas e que o SVM (e provavelmente outros métodos quantitativos de previsão de vendas), treinado nos períodos iniciais, dificilmente será capaz de prever um valor tão diferente assim. Contudo, isso não invalida o método EMD-PSO-SVM, já que ele apresentou resultados satisfatórios para a maior parte da série prevista.

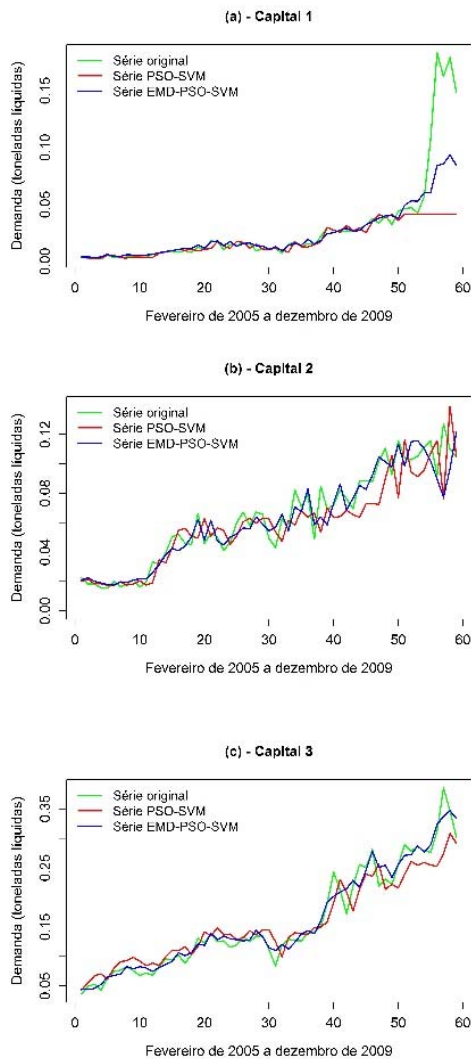


Figura 5. Série original (em verde), série PSO-SVM (em vermelho) e série EMD-PSO-SVM (em azul) para capitais.

Além da análise gráfica, a qualidade das previsões de cada modelo adotado pode ser avaliada a partir da medida de erro MSE calculada nas três fases do SVM. Para a metodologia EMD-PSO-SVM, após obter a série prevista de cada capital por meio da soma dos IMFs e do resíduo correspondentes, calculou-se o MSE associado. O MSE também foi obtido para as previsões do PSO-SVM. Os valores dessa medida de erro estão apresentados na Tabela I para cada etapa do PSO-SVM.

Com base na análise da Tabela I, percebe-se que ao se utilizar o método de decomposição EMD com $sm = \text{"none"}$ e $stoprule = \text{"type2"}$, junto ao PSO e ao SVM, o MSE apresentou ordem de grandeza baixa se comparada àquela dos dados de cada capital. Isso significa que o método pode ser empregado para a realização dessa previsão, sem provocar erros elevados.

Quanto ao MSE relativo ao PSO-SVM, os valores mostraram oscilar mais entre as fases do SVM, com erros pequenos em alguns momentos e altos em outros.

TABELA I

ERROS DE PREVISÃO PARA AS CAPITALIS 1, 2 E 3. METODOLOGIAS: (A) EMD-PSO-SVM COM $sm="none"$ E $stoprule="type2"$, (B) EMD-PSO-SVM COM $sm="splinesmoothing"$ E $stoprule="type5"$ E (C) PSO-SVM. MELHORES VALORES EM NEGRITO.

Cap.	Met.	MSE de treinamento	MSE de validação	MSE de teste	MSE total
1	A	4,249·10 ⁻⁶	2,375·10⁻⁵	4,889·10 ⁻³	5,049·10⁻⁴
	B	4,033·10⁻⁶	1,200·10 ⁻⁴	4,724·10⁻³	5,077·10 ⁻⁴
	C	5,853·10 ⁻⁶	2,907·10 ⁻⁵	1,162·10 ⁻²	1,191·10 ⁻³
2	A	4,363·10⁻⁵	4,893·10 ⁻⁵	5,296·10 ⁻⁴	9,413·10 ⁻⁵
	B	4,468·10 ⁻⁵	4,239·10⁻⁵	9,035·10⁻⁵	4,886·10⁻⁵
	C	1,117·10 ⁻⁴	3,783·10 ⁻⁴	7,252·10 ⁻⁴	2,283·10 ⁻⁴
3	A	1,101·10 ⁻⁴	4,529·10⁻⁴	5,914·10⁻⁴	2,288·10⁻⁴
	B	1,099·10⁻⁴	5,755·10 ⁻⁴	1,254·10 ⁻³	3,210·10 ⁻⁴
	C	3,386·10 ⁻⁴	1,177·10 ⁻³	3,049·10 ⁻³	7,486·10 ⁻⁴

Além disso, percebe-se que o método EMD, de fato, contribuiu para a melhoria da previsão realizada pelo PSO-SVM. Isso é verificado através dos valores do MSE para todas as fases e em todas as capitais, os quais, para o modelo previsto pelo EMD-PSO-SVM mostrou-se estar sempre abaixo daquele previsto apenas pelo PSO-SVM.

Essa diferença entre os erros dos dois modelos em alguns momentos encontra-se na mesma ordem de grandeza enquanto que, em outros, mostra-se em ordens de grandeza diferentes. Contudo, os valores do MSE, em geral, mostraram-se inferiores nos casos em que a aplicação do PSO-SVM foi precedida pela decomposição da série original via EMD.

Também foram realizados experimentos EMD-PSO-SVM com $sm="splinesmoothing"$ e $stoprule="type5"$ na porção EMD da metodologia, que correspondem, respectivamente, à construção dos envelopes via *spline smoothing* e ao critério de parada *S*-número descrito na Seção II-A. Os resultados foram bastante similares à configuração $sm="none"$ e $stoprule="type2"$, como pode ser observado comparando-se os valores de MSE das duas variantes de EMD-PSO-SVM na Tabela I. Ainda, EMD-PSO-SVM com o uso de $sm="splinesmoothing"$ e $stoprule="type5"$ também é associado a erros, em geral, menores quando seu desempenho é comparado ao do PSO-SVM. O desempenho geral das metodologias é também apresentado na Tabela I por meio do MSE total, que corresponde ao cálculo do MSE para toda série, desconsiderando-se as fases do PSO-SVM. Para as Capitais 1 e 3, o desempenho da primeira variante do EMD-PSO-SVM é superior, enquanto os para Capital 2, EMD-PSO-SVM com $sm="splinesmoothing"$ e $stoprule="type5"$ apresentou menor MSE. Já o PSO-SVM teve desempenho inferior para todas as cidades. Isso ratifica o impacto positivo do EMD sobre a capacidade preditiva do PSO-SVM.

V. CONCLUSÃO

Este artigo propôs a aplicação da metodologia EMD-PSO-SVM no contexto de previsão de demanda. Ela mostra-se bastante útil para a análise de séries cujo comportamento é não linear e não estacionário. As técnicas EMD, PSO e SVM foram aplicadas à previsão de vendas de produtos alimentícios a partir de séries temporais. Especificamente, foram utilizadas séries

mensais referentes aos dois produtos mais vendidos de uma empresa em três capitais brasileiras. Executou-se EMD-PSO-SVM e também PSO-SVM com o intuito de avaliar o efeito da decomposição prévia das séries temporais na capacidade preditiva de PSO-SVM.

A partir dos resultados obtidos, a abordagem EMD-PSO-SVM apresentou melhor desempenho quando comparada ao PSO-SVM, já que os valores de MSE associados mostraram-se inferiores para todas as capitais e também para todas as etapas do PSO-SVM. Os resultados encontrados e a versatilidade da metodologia EMD-PSO-SVM sugerem que ela pode ser facilmente adaptada à análise e previsão de séries temporais não-estacionárias em diferentes contextos de aplicação além da indústria alimentícia, por exemplo, nos setores de energia elétrica e de óleo & gás.

AGRADECIMENTOS

O segundo e terceiro autores agradecem ao CNPq pelo auxílio financeiro (processos 307809/2015-3 e 307940/2015-2, respectivamente).

REFERÊNCIAS

- [1] MOREIRA, D. A. *Administração da produção e operações*. 2ed. São Paulo: Cengage Learning, 2011.
- [2] HUANG, S., CHANG, J., HUANG, Q. and CHEN, Y. Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, v. 511, p. 764-775, 2014.
- [3] JIANG, X., ZHANG, L. and CHEN, X. Short-term forecasting of high-speed rail demand: A hybrid approach combining ensemble empirical mode decomposition and gray support vector machine with real-world applications in China. *Transportation Research Part C*, vol. 44, p. 110-127, 2014.
- [4] YU, H. and LIU, H. Improved stock market prediction by combining support vector machine and empirical mode decomposition. In: *Fifth International Symposium on Computational Intelligence and Design*, 2012.
- [5] GHELARDONI, L. GHIO, A. and ANGUIA, D. Energy load forecasting using empirical mode decomposition and support vector regression. *IEEE Transactions on smart grid*, vol. 4, n. 1, p. 549-556, 2013.
- [6] LINS, I. D., ARAUJO, M., MOURA, M. D. C., SILVA, M. A. and DROGUETT, E. L. Prediction of sea surface temperature in the Tropical Atlantic by support vector machines. *Computational Statistics and Data Analysis*, vol. 61, p. 197-198, 2013.
- [7] LIN, G.-F., LIN, H.-Y. and WU, M.-G., Development of a support-vector-machine based model for daily pan evaporation estimation. *Hydrological Process*, vol. 27, p. 3115-3127, 2013.
- [8] VAPNIK, V. N. *The nature of statistical learning theory*. 2ed. New York: Springer-Verlag, 2000.
- [9] HSU, C.-W. and LIN, C.-J. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, v. 13, p. 415-425, 2002.
- [10] ITO, K. and NAKANO, R. Optimizing support vector regression hyperparameters based on cross-validation. *Annals of the International Joint Conference on Neural Networks*, Portland, U.S.A., 2003.
- [11] CHEN, K.-Y. Forecasting systems reliability based on support vector regression with genetic algorithms. *Reliability Engineering and System Safety*, v. 92, p. 423-432, 2007.
- [12] FEI, S.-W., WANG, M.-J., MIAO, Y.-B., TU, J. and LIU, C.-L. Particle swarm optimization-based support vector machine for forecasting dissolved gases content in power transformer oil. *Energy Conversion and Management*, v. 50, p. 1604-1609, 2009.
- [13] LINS, I. D., MOURA, M. C., ZIO, E. and DROGUETT, E. L. A particle swarm-optimized support vector machine for reliability prediction. *Quality and Reliability International*, v. 28, p. 141-158, 2012.

- [14] SOUALHI, A., MEDJAHAR, K. and ZERHOUNI, N. Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression. *IEEE Transactions on Instrumentation and Measurement*, vol. 64, n. 1, p. 52-62, 2015.
- [15] TABRIZI, A., GARIBALDI, L., FASANA, A. and MARCHESIELLO, S. Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Meccanica*, vol. 50, p. 865-874, 2015.
- [16] WU, L. and PENG, C. Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm. *Energies*, vol. 9, n. 4, 261, 19p., 2016.
- [17] HUANG, N.E., SHEN, Z., LONG, S.R., WU, M.L.C., SHIH, H.H., ZHENG, Q.N., YEN, N.C. TUNG, C.C. and LIU, H.H. The empirical mode decomposition method and the Hilbert spectrum for non stationary time series analysis. *Proceedings of the Royal Society A - Mathematical, Physical & Engineering Sciences*, vol. 454A, p. 903 - 995, 1998.
- [18] HUANG, N.E. and WU, Z. A review on Hilbert-Huang transform: method and its applications to geophysical studies. *Reviews of Geophysics*, vol. 46, n. 2, RG2006, 23p, 2008.
- [19] WANG, G., CHEN, X.-Y., QIAO, F.-L., WU, Z. and HUANG, N. E. On intrinsic mode function. *Advance in Adaptive Data Analysis*, vol. 2, n. 3, p. 277-293, 2010.
- [20] HUANG, N. E., WU, M.-L. C., LONG, S.R., SHEN, S. S.P., QU, W., GLOERSEN, P. and FAN, K. L., A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society A - Mathematical, Physical & Engineering Sciences*, vol. 459, p. 2317-2345, 2003.
- [21] KECMAN, V. Support vector machines: an introduction. In: WANG, L. (Ed.). *Support vector machines: theory and applications*. Berlin Heidelberg: Springer-Verlag, 2005, (Studies in Fuzziness and Soft Computing, v. 177). p. 1-47.
- [22] KENNEDY, J. and EBERHART, R. Particle swarm optimization. In: *Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia, 1995.
- [23] KIM, D. and OH, H.-S. EMD: a package for empirical mode decomposition and Hilbert spectrum. *The R Journal*, vol. 1/1, 2009.
- [24] KIM, D., KIM, K.-O. and OH, H.-S. Extending the scope of empirical mode decomposition using smoothing. *EURASIP Journal on Advances in Signal Processing*, vol. 2012:168, 17p., 2012.
- [25] CHATFIELD, C. *The analysis of time series: an introduction*. 6ed. Boca Raton: Chapman & Hall / CRC, 2004.
- [26] BANERJEE, A., DOLADO, J. J., GALBRAITH, J. W. and HENDRY, D. F. *Co-integration, error correction, and the econometric analysis of non-stationary data*. Oxford: Oxford University Press, 1993.
- [27] CHANG, C.-C. and LIN, C.-J., LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, v. 2, n. 3, Article 27, 2011.



Enrique López Droguett é Professor Associado do Departamento de Engenharia Mecânica da Universidade do Chile. Conduz pesquisas em métodos computacionais para gerenciamento de ativos físicos, confiabilidade, manutenção e prognóstico de falhas; liderou estudos envolvendo esses tópicos nos setores de óleo & gás, defesa e redes de distribuição de energia. Dr. López Droguett tem mais de 160 artigos científicos publicados em periódicos e em anais de conferências.



Edgar Braga é Mestre em Engenharia de Produção pela UFPE e é pesquisador colaborador do CEERMA-UFPE. Desenvolve pesquisas nas áreas de logística e de gerenciamento de riscos em cadeias de suprimento.



Ítalla Dayanna da Silva é Engenheira de Produção pela Universidade Federal de Pernambuco (UFPE). Ela desenvolveu estudos relacionados à previsão de demanda e é colaboradora do Centro de Estudos e Ensaios em Risco e Modelagem Ambiental (CEERMA-UFPE).



Márcio das Chagas Moura é Doutor em Engenharia de Produção pela UFPE (2009). É Professor Adjunto do Departamento de Engenharia de Produção da UFPE e pesquisador do CEERMA-UFPE. Seus temas de pesquisa incluem processos estocásticos, soft computing, support vector machines, modelagem e técnicas de simulação aplicadas a risco, confiabilidade e manutenção de sistemas produtivos.



Isis Didier Lins é Doutora em Engenharia de Produção pela UFPE (2013). É Professora Adjunta do Departamento de Engenharia de Produção da UFPE e pesquisadora do CEERMA-UFPE. Seus temas de pesquisa incluem otimização, soft computing, support vector machines, prognóstico de falhas aplicados a risco, confiabilidade e manutenção de sistemas produtivos.