XX ENGEMA
Encontro Internacional sobre Gestão Empresarial e Meio Ambiente

ISSN: 2359-1048
Dezembro 2018

Previsões de séries temporais utilizando um modelo hibrido ARIMA-KNN baseado na decomposição DWT aplicado ao mercado energético espanhol

JOÃO GUILHERME ANGSTMANN
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

ROBERTO IVO DA ROCHA LIMA FILHO

THIAGO MACHADO LEITÃO
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Previsões de séries temporais utilizando um modelo hibrido ARIMA-KNN baseado na decomposição DWT aplicado ao mercado energético espanhol

Primeiro Autor
Filiacao
Endereco da Instituicao
e-mail

Segundo Autor
Filiacao
Endereco da Instituicao
e-mail

RESUMO

O presente trabalho propõe a utilização de um modelo de previsão de series temporais aplicado ao setor energético espanhol baseado na decomposição discreta de Wavelet relacionado com o método *Autoregressive Integrated Moving Average* (ARIMA) e *k-nearest neighbors* (KNN), com dados obtidos pelo Operador Del Mercado Ibérico de 2013 á 2015. Os modelos apresentados demonstraram um ganho significativo de acurácia quando comparado com o método proposto por Box-Jenkins (1976) em todos os parâmetros analisados. Portanto, conclui-se que nosso modelo perfaz em uma alternativa satisfatório ao método tradicional e abre a oportunidade de testar modelos semelhantes em outras áreas do conhecimento.

PALAVRAS CHAVE. Previsão em Energia. Modelos Hbridos. Espanha.

EN - OR in Energy

1. Introdução

Diante das grandes mudanças do cenário econômico mundial, um dos setores de maior presença, no âmbito da micro e macro economia é o energético. Fundamentado nas camadas de geração, transmissão e distribuição de energia elétrica, os impactos das decisões deste mercado afetam drástica e continuamente toda a sociedade, em seus diferentes níveis sociais. Com isto, faz-se necessário compreender melhor como funciona a estrutura deste fenômeno da economia e aperfeiçoar nossa capacidade de realizar previsões.

Cientes desta carência, economistas têm se dedicado ao desenvolvimento de modelos econométricos cada vez mais complexos e acurados capazes de gerar previsões confiáveis, visando embasar politicas de investimento públicas e privadas no setor. Garantindo menores custos ao consumidor final, mais eficiência e segurança ao sistema.

O mercado energético da Espanha, é um exemplo prático de como modelos econométricos podem ser importantes, pois este se baseia em um modelo parecido com a concorrência perfeita, com negociações diárias equilibrando oferta e demanda que determinam preços e quantidades a serem transacionados. Deste modo, uma possível diretriz estratégica para as empresas pertencentes a este mercado seria ter noções sobre o comportamento previsto destas variáveis ao médio-longo prazo.

Assim, o estudo de previsão estatístico do preço e quantidade de energia diários pode possuir fundamental importância para o setor energético, porém, pela complexidade destas variáveis, um modelo de regressão puro pode trazer resultados insuficientemente precisos. Como uma tentativa de obter resultados melhores, este trabalho propõe a aplicação de um modelo híbrido, baseado na decomposição discreta de Wavelet relacionado com o método *Autoregressive Integrated Moving Average* (ARIMA) e *k-nearest neighbors* (KNN), a fim de captar o comportamento temporal deste conjunto de dados e auxiliar, de uma forma simples, a obter previsões para o comportamento futuro do mercado.

O restante deste artigo está organizado da seguinte forma. Na seção 2 trazemos uma breve revisão da literatura. Na seção 3 nós descrevemos os dados. O modelo é apresentado na seção 4. Por fim, as analises dos resultados e as conclusões nas seções 5 e 6, respectivamente.

2. Revisão da Literatura

Table 1: Literatura tipica de previsões utilizando modelos híbridos de ARIMA no setor de energia

Literatura Tipica	Modelos de previsão	Main results
Tan et al. [2010]	Wavelet+ARIMA + GARCH	introduz a transformada wavelet associada com um hibrido ARIMA-GARCH aplicado ao mercado energético.
Liu et al. [2012]	ARIMA-ANN; ARIMA+KALMAN	O autor traz uma comparação entre dois modelos aplicados á previsão numérica do clima trazendo mais estabilidade ao sistema O filtro de Kalman é utilizado para encontrar o melhor ajuste para previsão da velocidade do vento.
Babu e Reddy [2014]	ARIMA-ANN	Insere filtro de médias moveis para extrair o resíduo não linear do processo ARIMA, resultando em um modelo mais acurado que os obtidos até então.
Khandelwal et al. [2015]	Wavelet+ARIMA+ANN	O autor utiliza a transformada wavelet para tratar o dado bruto, contudo a fim de extrair o resíduo não linear do processo ARIMA o autor utiliza diferença aritmética simples entre o ajuste e a base de dados originais.

O mercado de energia elétrica é essencial para a sociedade tal como a conhecemos hoje. Os impactos das decisões deste mercado afetam desde contas de luz da população até negociações e decisões de investimento de grandes empresas diariamente. Tal fato coloca o mercado energético no foco de estudos econômicos, tendo tido atenção especial dos economistas, que vêm desenvolvendo inúmeras ferramentas de previsão que ajudam os agentes deste mercado tomarem decisões mais assertivas.

Um dos primeiros modelos neste sentido foi o modelo auto-regressivo integrado de média móvel (ARIMA), proposto por Box et al. [1976]. A enfase deste método é a analise probabilística,

ou estocástica, de series temporais permitindo que este seja aplicado em diversas áreas da economia e engenharia Gujarati [2004]. Contudo, este modelo apresenta dificuldade de realizar ajustes não lineares.

Por sua vez métodos baseados em *machine learning*, como Redes Neurais Artificias (RNA), apresentam bons resultados em ajustes não lineares mas não conseguem ajustar bem as duas componentes Zhang [2003].

Neste contexto, surgiram os modelos híbridos que focam na combinação de dois ou mais métodos para atingirem melhores resultados como os modelos de Tan et al. [2010], Liu et al. [2012] e Babu e Reddy [2014]. Estes modelos vem progredindo continuamente com adições de novas etapas ou processos como a adição da transformação discreta de Wavelet proposta por Khandelwal et al. [2015] para separar as componentes das series temporais.

Em adição a estes estudos nós decidimos introduzir uma nova variante deste hibrido substituindo o algoritmo RNA por um não-paramétrico capaz de ajustar adequadamente a função sem que seja necessário fazer fortes suposições sobre o seu comportamento, de modo que o modelo seja livre para aprender qualquer forma funcional do conjunto de treinamento Tom M. Mitchell e Thomas [1997]. Sendo escolhido o método K-Nearest Neighbor devido sua a simplicidade e extensa bibliográfica.

2.1. Mercado Ibérico de Eletricidade

O mercado Energético Ibérico surgiu em 1998 com a finalidade de tornar mais livre a comercialização da energia na península ibérica, em termos de um melhor preço e serviço. Sua gestão é feita pelo *Operador Del Mercado Iberico* (OMIE) e movimenta valor superior a 10 bilhões de euros, correspondendo a mais de 80% do consumo elétrico da região.

Na península ibérica existem dois tipos de mercados principais, o diário e o intradiario. Os preços do mercado diário são estipulados diariamente ás 12 horas para as vinte e quatro horas do dia seguinte. De modo que o preço é obtido através do cruzamento da oferta e demanda, seguindo o modelo marginalista adotado pela União Europeia, com no algoritmo aprovado para todos os mercados europeus (EUPHEMIA).

Neste mecanismo, compradores e vendedores realizam contato com este agente, que define os preços e conduz a transação. Os vendedores apresentam ofertas para os vinte e quatro períodos horários, com as seguintes características obrigatórias: período horário referente, unidade de produção, preço mínimo e quantidade. Outras complementares que podem ser incluídas são condição de indivisibilidade, gradiente de carga, ingressos mínimos e parada programada del Mercado Ibrico de Energa Polo Espaol [2017].

Por sua vez, o mercado intradiario segue a mesma logica do mercado diário e permite que os interessados modifiquem suas posições ao longo do dia, participando de seis sessões de leilões que ocorrem entre 4 da manhã e meio dia del Mercado Ibrico de Energa Polo Espaol [2017].

3. Dados

Para elaboração deste artigo consultamos a base publica de dados do portal *Operador Del Marcado Iberico* referente ao preço médio diário (Pt), intradiário (Rt) e o consumo difio (Qt) de energia elétrica transacionada no Mercado de Eletricidade da Península Ibérica nos períodos de jan. de 2013 a dez. de 2015. Perfazendo 978 observações de cada variável, onde Pt e Rt são expressos em Euros por MWh e Qt em MWh, ambas em função do tempo em dias.

A Fig. 1 exemplifica a dispersão dos preços diários ao longo do tempo. Nota-se que os valores variam de 16,76 á 73,06 Euros por MWh e sugerem um comportamento estacionário. O mesmo tipo de análise pode ser aferida para os preços intradiários.

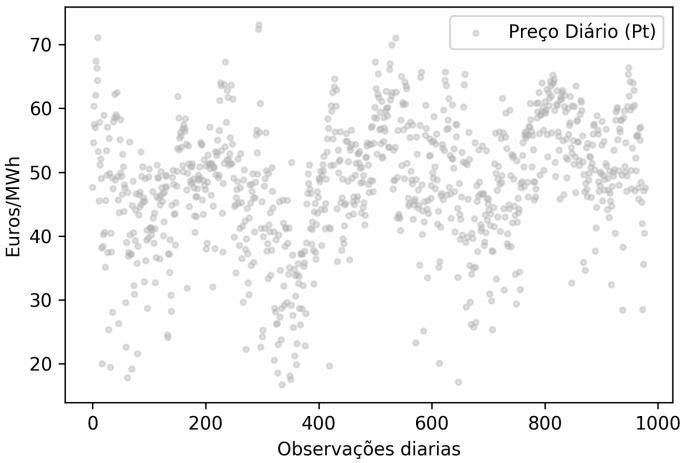

Figure 1: Dispersão dos Preços Diários

Por sua vez, a Fig. 2 traduz o comportamento do consumo diário em MWh ao longo do período analisado, observa-se que o consumo varia de 350.294 á 628.460 MWh e também aparenta um comportamento estacionário.

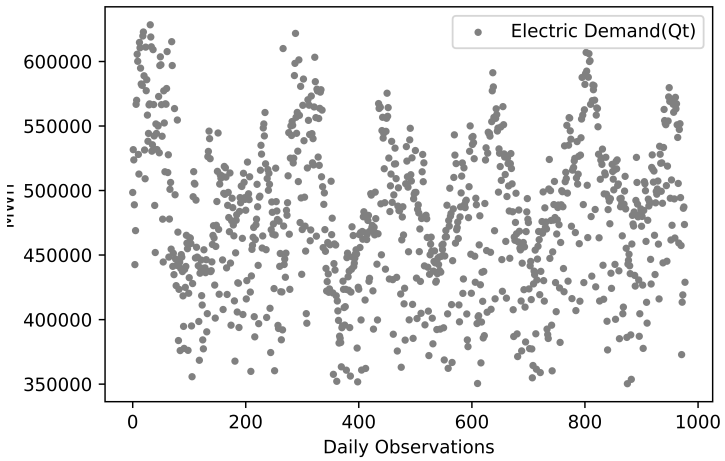

Figure 2: Dispersão do Consumo Diário

De modo a testar as hipóteses de estacionalidade levantadas pela observação ocular das Fig. 1 e 2, fora aplicado o teste de Dickey-Fuller aumentado descrito na próxima seção.

3.1. Teste de Dickey-Fuller Aumentado

A metodologia Box e jenkis é particularmente acurada para previsão de series temporais estacionarias, visto que a média, variância e correlação de variáveis podem ser aproximadas a partir de uma única base de dados representativa. Desta forma, é usual, no campo da econometria aplicada, adotar o teste de Dickey-Fuller, de modo a mensurar a propriedade estacionaria da série temporal e optar pelo modelo de ajuste mais indicado Enders [2014].

O teste de Dickey-Fuller testa a hipótese de raiz unitária em series temporais, onde se assume que os valores estocástico de seus diferentes termos são correlacionados entre si. O teste

faz esse ajuste adicionando um valor residual da variável dependente ΔY_t de modo que a equação possa ser descrita como:

$$\Delta y_t = \alpha + \beta t + \gamma_1 y_{t-1} + \delta_1 \Delta y_{t-1} + ... + \delta_{\rho-1} \Delta y_{t-\rho+1} + \varepsilon_t \quad (1)$$

Ao adicionar atrasos de ordem ρ a formulação ADF permite processos autorregressivos de ordens superiores. Em sequência o teste permite aplicar o teste de raiz única com hipótese nula $\gamma = 0$ contra a hipótese alternativa de $\gamma < 0$, onde o primeiro caso rejeita a hipótese nula.

Aplicando os testes concluiu-se que as séries analisadas não são estacionarias.

4. Método De previsão

Nesta seção os autores apresentam o modelo hibrido proposto, que utiliza Transformação Discreta de Wavelet (DWT) associada com regressão ARIMA e *K-Nearest Keighbors*. Um esboço do seu funcionamento pode ser observado na Fig. 3.

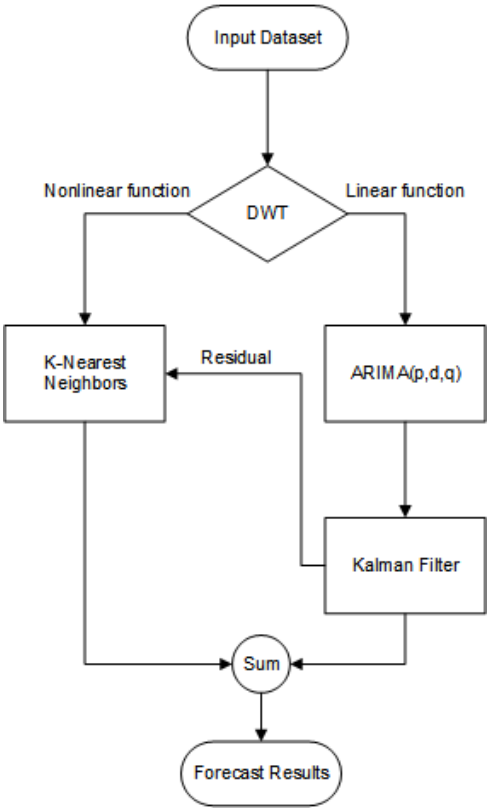

Figure 3: Fluxograma do modelo hibrido proposto

4.1. Transformação Discreta de Wavelet (DWT)

A Transformação Discreta de Wavelet consiste em uma função ortogonal que pode ser aplicada a um grupo infinito de dados com a finalidade de trata-los e revelar informações escondidas que não estão disponíveis no dado bruto Lei et al. [2013]. Podendo ser formulada como:

$$DWT(m,k) = \frac{1}{a} \sum_{n=0}^{N-1} s(n) g(\frac{k-b}{a}) \quad (2)$$

Onde s(n) é a função original, $a = a_0^m$, $b = n a_0^m$, N o número de amostras. Por sua vez a equação g() é *mother wavelet*. Neste artigo, utilizaremos um filtro simples. Logo, $a_0 = 1$.

Ao final deste processo teremos duas funções, chamadas de aproximação e diferença. Aplicando a transformada inversa, teremos duas funções uma linear e outra não-linear, que juntas correspondem a função original (Fig. 4).

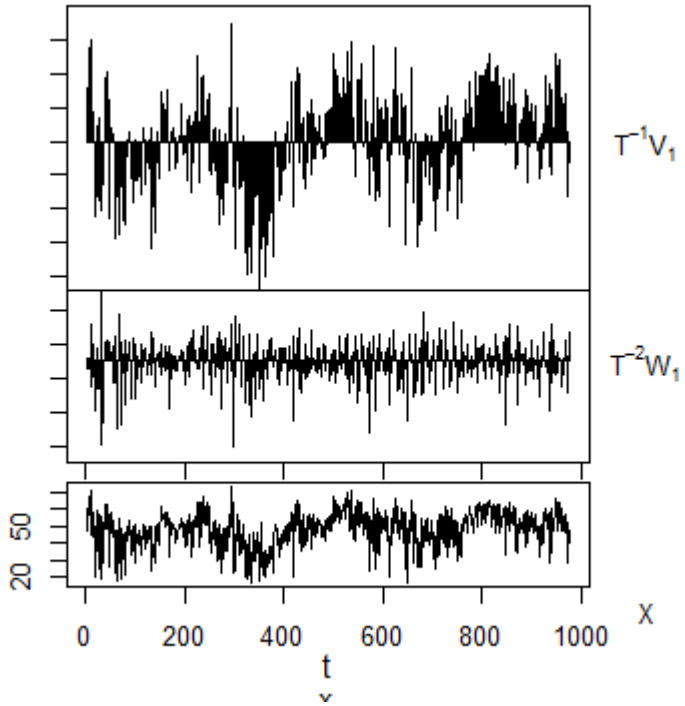

Figure 4: Decomposição DWT para a série Pt

4.2. Modelo auto-regressivo integrado de média móvel (ARIMA)

Obtido a componente linear da função, aplicamos o modelo auto-regressivo integrado de média móvel (ARIMA), proposto por Box et al. [1976]

Sua modelagem se baseia no princípio que os valores futuros de séries temporais podem ser definidos com base em uma função linear das observações passadas somado com um determinado valor estocśtico, e seus resultados expressos em ARIMA(p,d,q), onde *p* representa o número de termos autorregressivos, *d* o número de vezes que a série precisou ser diferenciada para atingir a estacionalidade e *q* o número de médias móveis Gujarati [2004]. Analisando as propriedades de autocorrelação (ACF) e autocorrelação parcial (PACF) da série é possível identificar qual o modelo ARIMA mais apropriado para explicar o comportamento dos dados, que neste caso corresponde aos preços e demanda no mercado energético espanhol. Na Figura 5, podemos observar ambos, ACF e PACF, para as os preços diários.

De modo a selecionar os melhores modelos de ARIMA(p,d,q) foram aplicados os critérios de *Akaike information criterion* (AIC) e *Bayesian information criterion* (BIC), ambos parâmetros são fundamentados na teoria da informação que estabelece o quão bem um fenômeno é explicado por um modelo em comparação com outros modelos possíveis. Tal analise é realizada com base na diferença entre os números de variáveis que expliquem o fenômeno e a Função de verossimilhança, desta forma ambos os critérios penalizam modelos com muitas variáveis sendo preferível o modelo que possuir os menores valores de AIC e BIC Enders [2014]. Abaixo, nota-se os melhores modelos encontrados.

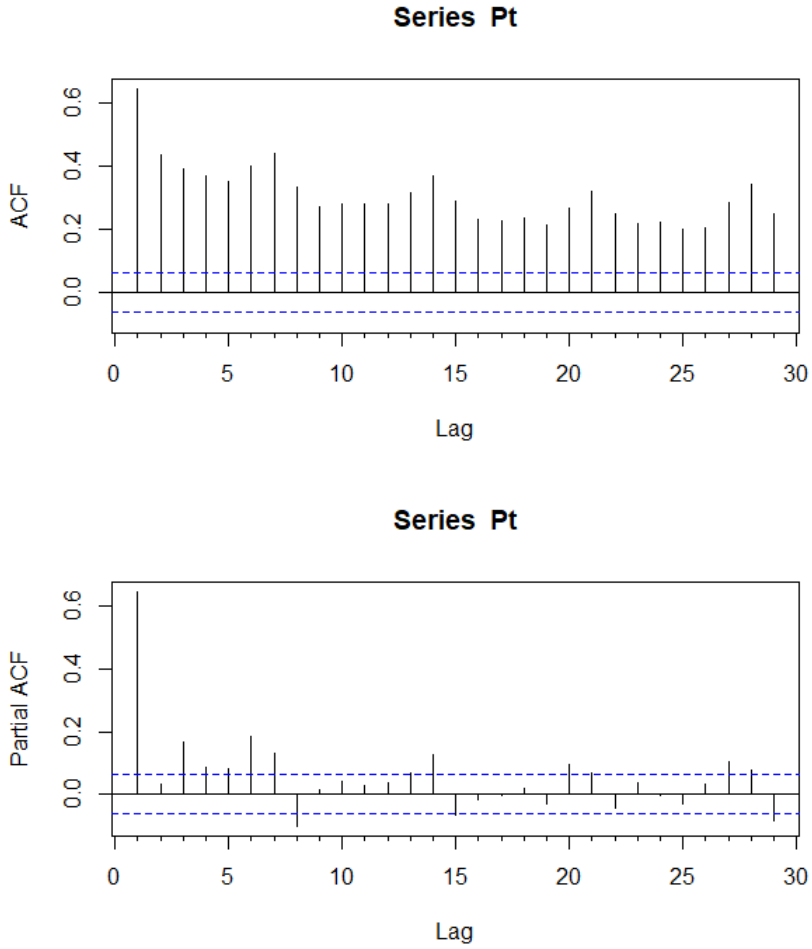

Figure 5: ACF e PACF para a série PT

Table 2: Resumo dos melhores modelos ARIMA com seus respectivos valores de AIC e BIC

Descrição	Modelo	AIC	BIC
Pt	ARIMA (11,0,0)	5407.922	5471.434
Rt	ARIMA (11,0,0)	5411.662	5475.174
Qt	ARIMA (3,0,0)	22930.744	22955.171

4.3. Filtro de Kalman

Visando melhorar a performance do modelo, aplicamos o filtro de Kalman para extrair o resíduo entre o ajuste realizado pelo método a ARIMA e a serie original. Entre os seus benefícios, estão sua capacidade de realizar ajustes exatos para uma amostragem finita e a função probabilidade de um processo ARMA Gaussiano Hamilton [1994]. Tais características, nos permitem maximizar a performance do modelo.

4.4. K-Nearest Neighbor Learning

Finalmente, aplica-se o método *K-Nearest Neightbor* (KNN) para ajustar a parcela não-linear obtida na seção 4.1. e o resíduo da seção anterior. KNN é um método não-paramétrico baseado em instâncias, que assume que todas as instancias correspondem a pontos n-dimensionais no espaço $\mathbb{R}^n$, sendo seus vizinhos definidos em relação a sua distância euclidiana para com uma dada instância conhecida Tom M. Mitchell e Thomas [1997].

Uma das vantagens deste método consiste em sua característica não paramétrica, ou seja, não é necessŕio fazer fortes suposições sobre o real formato da função Altman [1992], o que lhe torna um método simples e de fácil implementação. Ainda, podemos refinar este modelo, atribuindo pesos á contribuição de cada k vizinhos, de acordo com sua distância ao ponto de analise. Neste artigo, atribuímos pesos á cada ponto, sendo esta pontuação inversamente proporcional ao quadrado da distância entre o conjunto de treinamento e o valor desconhecido.

4.5. Avaliação do Erro

No contexto das técnicas de previsão as medidas de precisão são de fundamental importância para avaliação da eficiência da modelagem.

Com isto em mente, foi adotado as duas principais avalizações para series temporais: Erro Médio Percentual Absoluto (EMPA) e Raiz do Quadrado Médio do erro (RQME) e

$$EMPA = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|(100\%) \tag{3}$$

$$RQME = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2} \tag{4}$$

$$\tag{5}$$

5. Resultados da previsão

De modo a demonstrar a eficiência do modelo proposto realizamos modelagens adicionais com outros métodos de previsão cujo as eficiências são reconhecidas pela literatura. Para efeito de comparação foram utilizadas as mesmas bases de dados.

A Tabela 3 traz um compilado sobre o desempenho dos ajustes por modalidade. Nota-se que o modelo proposto apresentou os melhores resultados em quase todos os parâmetros analisados ficando com EMPA abaixo de 2,7%. Apresentando uma grande evolução em comparação com os resultados do ARIMA puro, especialmente no que diz respeito ao EMPA. Por outro lado, o método KNN puro se mostrou bastante competitivo, ficando bem próximo dos resultados do modelo proposto mas ainda assim é observável uma redução média de 20,7% no erro RQME para os preços diários e intradiários. Tal melhoria pode ser bastante significativa em um mercado tão competitivo quanto o energético.

Table 3: Comparativo do desempenho dos ajustes

Método	PT		RT		QT	
	MSE	MAPE	MSE	MAPE	MSE	MAPE
ARIMA	7,307152	12,098556	7,307154	12,09856	41279,901196	7,036143
KNN	3.51788	2.86181	3.88427	2.973668	**21286.8507**	1.77335
Proposto	**2.868**	**2,3757**	**3.3320**	**2.6206**	21828.6277	**1.7327**

Comparamos também o resultado do nosso modelo proposto com alguns outros modelos publicados nos últimos anos Tabela 4. Desta forma, observa-se, com a devida cautela, que nosso modelo é competitivo para este cenário.

Table 4: Resumo de resultados de modelos de previsões aplicados no mercado de energia nos últimos anos

Method	From	Reported MAPE
Wavelet fuzzy neural network	Hanmandlu e Chauhan [2011]	1.51%
Neuro fuzzy	Hanmandlu e Chauhan [2011]	2.83%
Adaptive neuro-fuzzy inference system	Hanmandlu e Chauhan [2011]	3.00%
Wavelet transform + Evolutionary algorithm	Amjady e Keynia [2009]	2.04%
Wavelet + particle swarm optimization	Bashir e El-Hawary [2009]	3.96%
SOM neural networks	Lpez et al. [2012]	2.18%

Source:Lpez et al. [2012]

A Fig. 6 demonstra graficamente o comportamento da serie original e dos ajustes para o preço diário. Nota-se que o ajuste corresponde bem com os dados reais em todos o período analisado incluindo eventuais picos de energia, o que colabora com o observado na análise teórica dos erros.

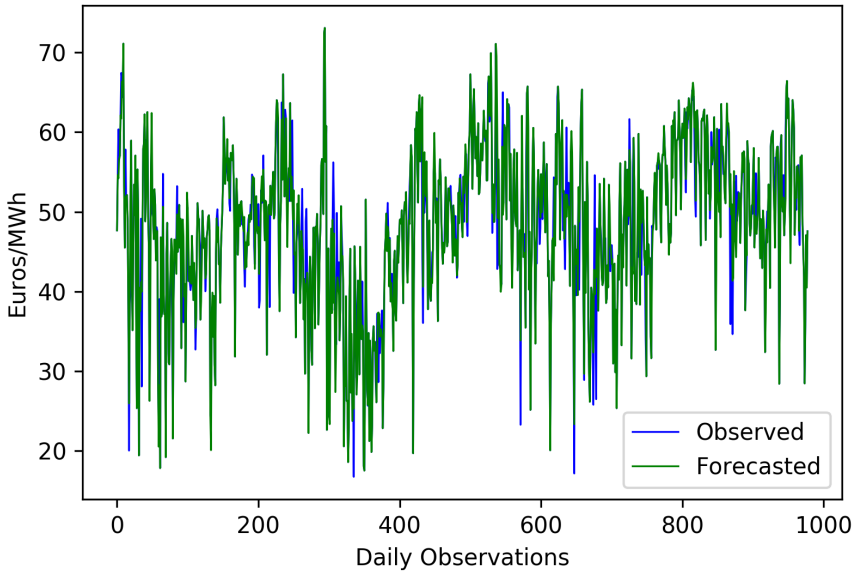

Figure 6: Gráfico comparativo entre o ajuste proposto e os dados originais para o preço diário

6. Conclusão

Durante este trabalho foi possível analisar as particularidades do mercado de energia espanhol, caracterizado como sistema de competição perfeita e encontrar modelos que explicam satisfatoriamente o comportamento dos preços e da demanda, com estes dados seria possível realizar previsões sobre o desempenho futuro do mercado e embasar alocação de recursos e politicas de preços por parte dos agentes garantindo um sistema mais eficiente.

Além disso, pontua-se que o modelo proposta é generalista e pode ser aplicado em outras áreas do conhecimento. A titulo de exemplo, metodologia semelhante poderia ser utilizada para prever a velocidade do vento em áreas de produção eólica em dado intervalo de tempo e aumentar a confiabilidade desta matriz energética como em Tan et al. [2010] ou previsão dos índices pluviométricos em áreas de produção hidroelétricas. Minimizando o risco de investimentos no setor e contribuindo para o desenvolvimento de energia sustentável à preços acessíveis.

Por fim, espera-se que este projeto possa colaborar com o desenvolvimento de novas metodologias e que possa, futuramente, ser aplicado no mercado energético brasileiro e contribuir para a eficiência do setor.

7. Agradecimentos

Nós agradecemos a Fundação de Amparo a Pesquisa do Rio de Janeiro (FAPERJ) pela ajuda financeira dada ao projeto.

References

Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185. ISSN 00031305. URL http://www.jstor.org/stable/2685209.

Amjady, N. e Keynia, F. (2009). Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy*, 34(1):46–57. URL https://ideas.repec.org/a/eee/energy/v34y2009i1p46-57.html.

Babu, C. N. e Reddy, B. E. (2014). A moving-average filter based hybrid arimaann model for forecasting time series data. *Applied Soft Computing*, 23:27 – 38. ISSN 1568-4946. URL http://www.sciencedirect.com/science/article/pii/S1568494614002555.

Bashir, Z. A. e El-Hawary, M. E. (2009). Applying wavelets to short-term load forecasting using pso-based neural networks. *IEEE Transactions on Power Systems*, 24(1):20–27. ISSN 0885-8950.

Box, G. E., Jenkins, G. M., Reinsel, G. C., e Ljung, G. M. (1976). *Time series analysis: forecasting and control*. John Wiley & Sons.

del Mercado Ibrico de Energa Polo Espaol, O. (2017). About our market. URL http://www.omie.es/en/home/markets-and-products/about-our-market.

Enders, W. (2014). *Applied Econometric Time Series (Wiley Series in Probability and Statistics)*. Wiley, fourth edition.

Gujarati, D. N. (2004). *Basic Econometrics*. Tata McGraw Hill, fourth edition.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, first edition.

Hanmandlu, M. e Chauhan, B. K. (2011). Load forecasting using hybrid models. *IEEE Transactions on Power Systems*, 26(1):20–29. ISSN 0885-8950.

Khandelwal, I., Adhikari, R., e Verma, G. (2015). Time series forecasting using hybrid arima and ann models based on dwt decomposition. *Procedia Computer Science*, 48:173 – 179. ISSN 1877-0509. URL http://www.sciencedirect.com/science/article/pii/S1877050915006766. International Conference on Computer, Communication and Convergence (ICCC 2015).

Lei, L., Wang, C., e Liu, X. (2013). Discrete Wavelet Transform Decomposition Level Determination Exploiting Sparseness Measurement. URL https://doi.org/10.5281/zenodo.1087700.

Liu, H., qi Tian, H., e fei Li, Y. (2012). Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction. *Applied Energy*, 98:415 – 424. ISSN 0306-2619. URL http://www.sciencedirect.com/science/article/pii/S0306261912002875.

Lpez, M., Valero, S., Senabre, C., Aparicio, J., e Gabaldon, A. (2012). Application of som neural networks to short-term load forecasting: The spanish electricity market case study. *Electric Power Systems Research*, 91:18 – 27. ISSN 0378-7796. URL http://www.sciencedirect.com/science/article/pii/S0378779612001095.

Tan, Z., Zhang, J., Wang, J., e Xu, J. (2010). Day-ahead electricity price forecasting using wavelet transform combined with arima and garch models. *Applied Energy*, 87(11): 3606 – 3610. ISSN 0306-2619. URL http://www.sciencedirect.com/science/article/pii/S0306261910001807.

Tom M. Mitchell, T. M. e Thomas, M. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math, first edition.

Zhang, G. (2003). Time series forecasting using a hybrid arima and neural network model. *Neurocomputing*, 50:159 – 175. ISSN 0925-2312. URL http://www.sciencedirect.com/science/article/pii/S0925231201007020.