

# Prediction of Brazilian Electric Energy Price Using Recurrent Artificial Neural Networks and Correction Filter

Vítor Giudice Batista de Araujo Porto\* e Leonardo Rocha Olivi\*

\*Departamento de Energia Elétrica, Faculdade de Engenharia,  
Universidade Federal de Juiz de Fora, Juiz de Fora, MG, Brasil

(e-mails: [giudice.vitor@engenharia.ufjf.br](mailto:giudice.vitor@engenharia.ufjf.br), [leonardo.olivi@ufjf.edu.br](mailto:leonardo.olivi@ufjf.edu.br))

**Abstract:** The Settlement Price of Differences (SPD) is a variable used to determinate the value to be settled in the Chamber of Electric Energy Commercialization (CCEE) relative to the difference between the amount of energy that has been supplied to a company and the amount that had been contracted. This variable changes on a weekly basis and is calculated based on statistical and mathematical models of optimization, and therefore its behavior is highly non-linear. This paper predicts the SPD value one week ahead of time using an LSTM recurrent neural network architecture and a correction filter, avoiding the most common pitfalls of using recurrent networks in time series forecasting. The results show how dealing with the right input variables leads to a trustworthy prediction of the SPD.

**Resumo:** O Preço de Liquidação das Diferenças (PLD) é uma variável utilizada para determinar o valor a ser cobrado pelos volumes de energia que serão liquidados na CCEE (Câmara de Comercialização de Energia Elétrica), e é atualizado semanalmente. Seu cálculo é baseado em modelos estatísticos e matemáticos de otimização, e, portanto, apresenta um comportamento altamente não-linear. Este trabalho propõe, por meio de uma arquitetura recorrente de redes neurais artificiais LSTM e um filtro corretivo, a predição do preço do PLD uma semana à frente, buscando obter as melhores variáveis de entrada, a fim de contornar problemas recorrentes que aparecerem com o uso de redes recursivas em séries temporais. O resultado mostra como a obtenção das variáveis corretas acarretam em uma predição confiável do PLD.

**Keywords:** Prediction; Filtering; Neural Networks; LSTM; Forecasting; Electrical Energy Price.

**Palavras-chaves:** Predição; Filtragem; Redes Neurais; LSTM; Previsão; Preço da Energia Elétrica.

## 1. INTRODUÇÃO

Na comercialização de energia elétrica no mercado livre, empresas e grandes negócios usualmente compram energia em contratos de longa duração. No entanto, estas situações de longo prazo podem mudar com as discrepâncias nas necessidades energéticas das empresas em decorrência das mudanças que ocorrem com o passar do tempo. Assim, as empresas podem acabar por fazer uma contratação que termina por ser maior ou, mais comumente, menor do que aquilo que de fato consumiram, e o mesmo pode ocorrer com os agentes fornecedores de energia (CCEE, 2017).

Dessa forma, é preciso quitar essa diferença entre a quantia contratada e o que foi consumido/fornecido. Essa operação ocorre no Mercado de Curto Prazo (MCP), e está sujeita ao Preço de Liquidação das Diferenças (PLD), que dita quanto custará a energia a ser liquidada. O PLD é uma variável obtida por otimização e é atualizado semanalmente.

O valor do PLD é baseado no Custo Marginal de Operação (CMO), que leva em consideração a oferta de energia em um período de estudo e a demanda energética em um mesmo período. De acordo com a Câmara de Comercialização de Energia Elétrica (CCEE) (ABRACEEL, 2017; ONS, a), o Custo Marginal de Operação (CMO) é baseado, entre outros, nos valores previstos de demanda de energia, condições hidrológicas, e nas disponibilidades declaradas de geração,

utilizando os métodos NEWAVE e DECOMP para produzir um resultado semanal para o CMO para cada submercado. A partir desse resultado é obtido o PLD, sendo este também limitado por um preço máximo e um preço mínimo (ONS, b).

A energia elétrica brasileira é, atualmente, composta majoritariamente pela geração hidráulica, que corresponde a cerca de 65% da produção da matriz elétrica do país. Sendo assim, a capacidade brasileira de produção de energia sofre direta influência de fatores naturais, dependendo de estações e fatores climáticos, acarretando em uma alta volubilidade do PLD, conforme Figura 1 (ABRACEEL, 2017; EPE).

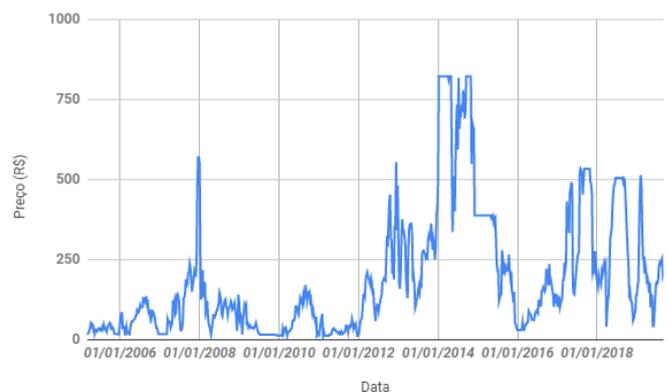


Figura 1 – Alta Volubilidade do Preço de Liquidação das Diferenças para o submercado Sudeste/Centro-Oeste.

Visto isso, é de suma importância que se possa prever com confiabilidade o comportamento desse preço, a fim que se possa obter lucros ao se comercializar energia no Mercado de Curto Prazo.

Este trabalho se propõe a apresentar uma abordagem diferente para a implementação de redes neurais artificiais recorrentes para a predição do PLD uma semana a frente, utilizando os dados disponibilizados pelo Operador Nacional do Sistema (ONS). O presente trabalho se divide em 5 Seções de maneira que, na Seção 2 apresentam-se as bases teóricas que levaram a construção da arquitetura proposta, na Seção 3 é exposto o processo de elaboração do trabalho, na Seção 4 são discutidos os resultados obtidos, e na Seção 5 são apresentadas as conclusões obtidas a partir dos resultados.

## 2. FUNDAMENTAÇÃO TEÓRICA

No ramo do aprendizado de máquinas, redes neurais artificiais (RNAs) são modelos computacionais capazes de reconhecer padrões altamente não-lineares com facilidade. Por padrão, os modelos tradicionais de redes neurais artificiais sugeriram com o intuito de emular neurônios biológicos, criando uma rede de diversos neurônios artificiais ligados entre si, que mapeiam um conjunto de entradas até uma ou mais saídas (Haykin, 2001; Silva, 2010), conforme a Figura 2.

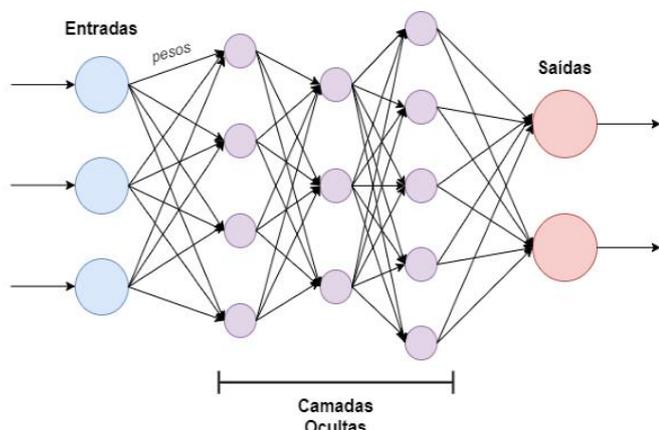


Figura 2 – Descrição da arquitetura básica de uma rede neural artificial tipo Perceptron Multicamadas.

Essa ligação entre os neurônios se dá a partir de pesos sinápticos, que tem como objetivo a ponderação do quanto as entradas anteriores de um determinado neurônio influenciam sobre este, que será ativado ou não a partir de uma função de ativação. Se ativado, um neurônio passa sua informação adiante, caso contrário, mantém-se inativo, e não contribui para a próxima camada de neurônios (Haykin, 2001).

Essa arquitetura tradicional, chamada de Perceptron Multicamadas (do inglês *Multilayer Perceptron*, MLP) se mostrou bastante eficiente para reconhecer padrões, lineares ou não, dado que alimentada com um número grande (em relação às arquiteturas tradicionais) de amostras de dados (Richard et al., 2000). Devido a sua eficiência, outras arquiteturas baseadas no MLP se desenvolveram. Em

especial, esse trabalho trata da arquitetura *Long Short Term Memory* (LSTM).

A arquitetura LSTM é considerada recorrente pois utiliza saídas antecedentes no tempo como entradas para sua próxima predição. Isso permite que informações anteriores interfiram no resultado atual da rede. Esse tipo de arquitetura é particularmente eficaz no reconhecimento de voz, de escrita, e na predição de séries temporais estocásticas.

Dentre as arquiteturas recorrentes, a LSTM tem se provado eficaz, pois seu conjunto de sub-redes (*gates*) dentro de um neurônio (também denominado por célula) permite selecionar quais informações são úteis e quais não devem interferir na predição atual da rede (Haykin, 2008), conforme a Figura 3.

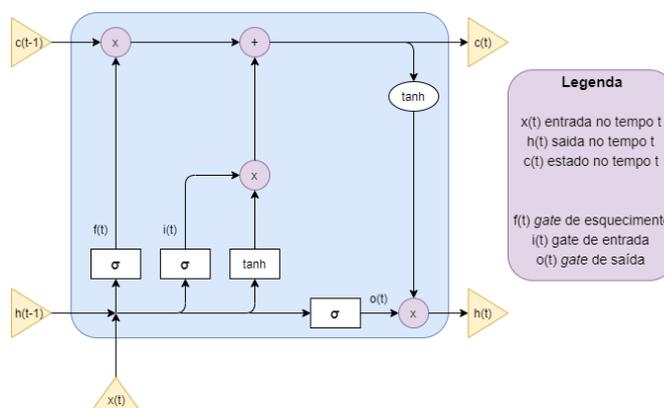


Figura 3 – Arquitetura básica de uma célula *Long Short Term Memory* (LSTM) utilizada e obtida a partir da literatura.

A arquitetura LSTM deste trabalho, foi escolhida em face dos trabalhos precusores na literatura, que mostraram maiores taxas de sucessos na predição ao utilizarem redes recorrentes, ou modelos que se valem de autorregressão.

Os autores de (Reston et al, 2014) propuseram uma metodologia de predição do PLD por meio de uma rede neural artificial tipo Perceptron multicamadas simples, que utiliza o próprio PLD como entrada, bem como outras variáveis de interesse fornecidas pelo Operador Nacional do Sistema (ONS, b), como a carga do sistema e a energia armazenada nos reservatórios, filtradas por um modelo de médias móveis integradas autorregressivas (ARIMA).

Os autores de (Nascimento et al., 2018) utilizaram redes neurais recorrentes tipo Elman para inserir o comportamento autorregressivo no processo de predição do modelo computacional neuronal. As redes recorrentes mostraram-se mais aptas a extrair dados de séries temporais contaminadas por ruídos estocásticos, uma vez que a modelagem autorregressiva proporciona a chance dos pesos se adaptarem a essas mudanças. Em (Conejo et al., 2005), os autores combinaram o poder de decomposição bidimensional de Wavelets, no tempo e na frequência/escala, além da filtragem autorregressiva dos dados por modelos ARIMA para predições um dia à frente do preço de eletricidade espanhol.

Por meio de um filtro de médias móveis autorregressivas com entradas exógenas (ARMAX) das saídas de predição de uma máquina de vetores de suporte (*Support Vector Machine*,

SVM), os autores de (Yan et al., 2014) propuseram a predição dos preços do mercado interconectado Pennsylvania – New Jersey – Maryland (PJM).

Os autores de (Chang et al., 2018) utilizaram uma rede neural tipo LSTM treinada por meio do algoritmo ADAM (Kingma et al., 2015) para a predição de preços do mercado de energia australiano. Os resultados dos autores sofreram de atraso na saída, caracterizando um modelo persistente (Flovik, 2018), denotando que não havia informação suficiente nos dados para o aprendizado, levando a rede a imitar o comportamento da saída imediatamente anterior.

Levando em consideração que a maioria dos trabalhos insere a autorregressão dos dados em algum momento do algoritmo, devido à natureza do processo estocástico envolvido, este trabalho propõe à utilização de uma rede recorrente tipo LSTM, que possui a capacidade de adaptação mediante ao grau de importância dos dados ao longo do tempo, bem como à filtragem da saída por meio de uma correção ponderada pela combinação de momentos anteriores da rede.

### 3. METODOLOGIA

Como os valores do PLD e do CMO se assemelham fortemente (ONS, a; ONS, b), com a maior diferença sendo os limites superior e inferior do PLD, optou-se por arquitetar uma rede neural artificial que predissesse o CMO, para, a partir deste, se obter o PLD, tendo em vista que as restrições de limite superior e inferior do PLD provocam a perda da informação pela saturação, enquanto que o CMO não oferece esta deficiência de informação em seus dados. Uma vez conseguidos os valores previstos de CMO, demonstrando nos resultados deste trabalho, basta limitá-los com os preços mínimo e máximo fornecidos pela Agência Nacional de Energia Elétrica (ANEEL) para obter o valor do PLD, que é a variável de interesse.

A fim de prever, então, o CMO com acurácia, foi proposta uma rede multi-camada composta por camadas LSTM, e uma última camada com um único neurônio tradicional, que servirá para produzir o valor final da rede, que posteriormente será filtrado.

Para a função de perdas (*Loss*) utilizada como objetivo para treinar a rede neural, foi escolhida a seguinte função:

$$Loss(t) = (Y_{true}(t) - Y_{true}(t - 1)) - (Y_{pred}(t) - Y_{pred}(t - 1)) \quad (1)$$

em que *Loss(t)* representa a função de perdas para a predição da semana *t*, *Y<sub>true</sub>* indica o valor verdadeiro da variável CMO, e *Y<sub>pred</sub>* o valor predito pela rede neural.

A função de perdas (1) atua como um penalizador de tendências, uma vez que faz com que predições que não conseguem determinar com exatidão o valor do PLD, mas conseguem prever o padrão de crescimento ou decréscimo que os valores farão, não sejam penalizadas.

Como entradas para a rede neural, foram utilizados as séries históricas de Energia Natural Afluente por Subsistema (ENA), Demanda Máxima (DM), Energia Armazenada (EA)

e Geração de Energia (GE), todos disponíveis no histórico de operação da ONS (ONS, b).

Optou-se por não utilizar os valores anteriores do PLD como entrada, pois a adição de atrasos temporais da saída como entrada faz com que a rede obtenha um comportamento de modelo persistente, ou seja, faz com que a saída tenda a apenas copiar o último valor registrado, conforme a Equação 2, sendo este o mesmo problema enfrentado pelos autores de (Chang et al, 2018), evidente em seus resultados.

$$f(t + 1) = f(t) \quad (2)$$

O comportamento da Equação 2, onde *f(t)* é uma série temporal qualquer, provoca, na predição de um tempo a frente, que o histórico da saída se assemelhe ao histórico da entrada deslocado de uma unidade temporal.

A escolha das variáveis de entrada supracitadas, a saber, ENA, DM, EA e GE, se baseou na correlação dentre todos os dados disponíveis no histórico da ONS e o histórico do CMO, todos comparados entre si. As variáveis escolhidas demonstraram os maiores valores de correlação, todos acima de 0,9, conforme o exemplo mostrado na Figura 4.

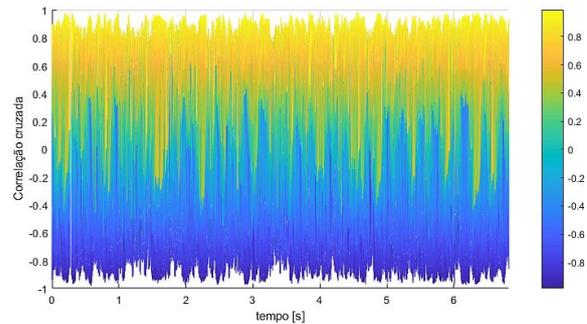


Figura 4 – Correlação do CMO com o dado DM usado como entrada da rede, mostrando correlação acima de 0,9.

Foram obtidos os históricos do CMO e de cada uma das variáveis de entrada para 690 semanas, da semana iniciada no dia 29/07/2006 à semana do dia 12/10/2019, tendo em vista que, de acordo com a CCEE (CCEE, 2017), para a determinação do PLD são consideradas semanas que se iniciam aos sábados. De posse desses dados, a série do CMO foi separada entre dados de treino (*training data*), para treinamento *offline* da rede, e dados de validação (*evaluation data*) utilizados para avaliar o desempenho da mesma.

A partir do histórico de saída obtido da rede já treinada, é feito um tratamento dos dados por meio de filtragem corretiva, de forma que o resultado o qual será considerado como o valor do CMO da semana seguinte fica de acordo com a seguinte equação:

$$Y_{out}(t) = Y_{true}(t - 1) + (Y_{pred}(t) - Y_{pred}(t - 1)) \quad (3)$$

Em que (3) é um filtro corretivo baseado no conhecimento de *Y<sub>true</sub>(t)*, que é o valor real do CMO no tempo *t*, e *Y<sub>pred</sub>(t)* é a saída literal imediata da rede neural para a semana *t*. Este filtro captura a variação da tendência da variável de interesse.

Esse tratamento se deve ao fato de que a arquitetura proposta se mostrou eficaz, conforme a literatura exposta, em prever as tendências de variações da variável de interesse. Ou seja,

por vezes não era capaz de prever seu valor com alto grau de exatidão, porém, tinha a capacidade de previsão das tendências da variável. Assim, o valor predito estará próximo ao valor anterior, evitando oscilações abruptas indesejadas, porém, tendendo ao valor de fato da variável de interesse.

Por fim, com o intuito de quantizar a eficácia do modelo proposto, os resultados finais obtidos para o CMO são comparados com os resultados de um modelo persistente, tratado nesse trabalho como linha de base comparativa. De acordo com os autores de (Flovik, 2018), uma série temporal, que pode ser aproximada como um modelo de passeio aleatório, ao ser utilizada como entradas com atrasos temporais de si mesma em uma rede neural, resulta na repetição da saída imediatamente anterior dos dados de treinamento, Equação 2.

Como consequência, esta repetição das saídas cria um efeito de atraso no tempo, em que a rede apenas repete o valor imediatamente anterior, comportamento denominado como modelo persistente. Esse fenômeno de persistência pode ser observado, por exemplo, no trabalho dos autores de (Chang et al., 2018), em que um resultado semelhante a de um modelo persistente pode ser observado como resultado apresentado. E (Flovik, 2018) argumenta que este não é um bom resultado porque não há previsão de fato, apenas repetição de valores conhecidos. Portanto, um dos objetivos deste trabalho é mitigar o comportamento persistente.

Dessa forma, neste trabalho, o modelo persistente será utilizado como base de comparação ao proposto. Assim, são comparadas as médias das diferenças de preço entre o modelo persistente e os valores verdadeiros, bem como entre o modelo proposto e os mesmos valores verdadeiros, valores estes dados em Reais (moeda). Assim, é possível obter um parâmetro de melhoria entre os modelos, como se segue:

$$Melhoria = 1 - \frac{\sum_1^n Y_{proposto_t}}{\sum_1^n Y_{persistente_t}} \quad (4)$$

Em que  $Y_{proposto_t}$  é o valor de saída do modelo proposto para a semana  $t$ ,  $Y_{persistente_t}$  é o valor de saída do modelo persistente para a mesma semana  $t$ , e  $n$  é o número total de semanas observadas.

#### 4. RESULTADOS

Para mitigar os problemas de desaparecimento ou explosão de gradientes (*vanishing gradient* e *exploding gradient*), fez-se uma normalização das variáveis de entrada, utilizando o *min-max normalization* entre 0 e 1, a fim de se utilizar uma função de ativação retificadora (ReLU), a qual oferece os benefícios de uma função de ativação linear, porém, introduz a não linearidade necessária para a boa convergência dos métodos. Quanto aos otimizadores, optou-se por utilizar o ADAM (Kingma et al., 2015), que, de acordo com os autores de (Chang et al., 2018), se mostrou mais rápido e efetivo ao ser usado com redes LSTM.

Após a calibração dos parâmetros disponíveis (número de camadas, quantidade de células por camada e número de atrasos temporais das entradas), o melhor resultado foi

observado em uma rede com 3 camadas, a primeira com 6 células LSTM, a segunda com 8 células LSTM, e a última com 1 neurônio Perceptron tradicional. Essa calibração se deu por experimentação, visto que na literatura geral ainda não há um consenso quanto ao método mais prático ou efetivo para se determinar tais parâmetros

A fim de se evitar o *overfitting* ou mínimos locais, foi adicionado o método *dropout* entre cada uma das camadas supracitadas, com uma chance de 15% de ocorrência. Além disso, após a fase de calibração, observou-se que a configuração mais vantajosa para as entradas foi com 16 atrasos temporais, conforme a Equação 5, com  $N = 16$  como o número de atrasos temporais.

$$Entradas = \begin{bmatrix} ENA(t) \\ ENA(t-1) \\ \vdots \\ ENA(t-N) \\ DM(t) \\ DM(t-1) \\ \vdots \\ DM(t-N) \\ EA(t) \\ EA(t-1) \\ \vdots \\ EA(t-N) \\ GE(t) \\ GE(t-1) \\ \vdots \\ GE(t-N) \end{bmatrix} \quad (5)$$

A rede neural passou pelo processo de treinamento ADAM por 360 épocas. Porém, para evitar o *overfitting*, apenas quando a perda (*Loss*) dos dados de avaliação melhorava em relação à última época registrada, os resultados dessa nova época eram salvos em seu lugar. Os resultados do treinamento e da validação podem ser vistos na Figura 5.

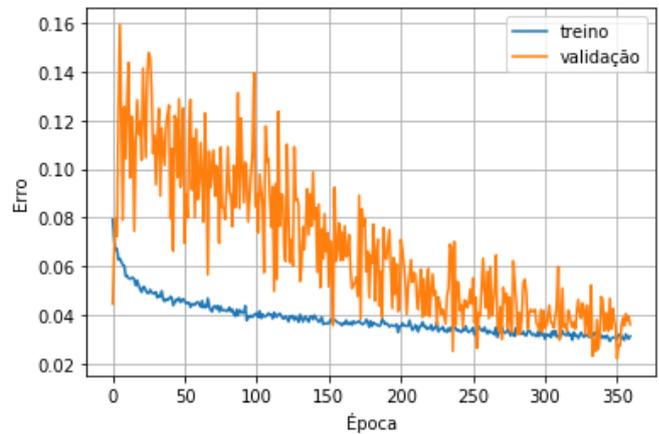


Figura 5 – Histórico do comparativo de perda (*Loss*) dos dados de treino e de validação, onde ambos convergem.

Como é possível observar na convergência da Figura 5, o treinamento em 360 épocas foi uma boa escolha, visto que não houve *overfitting* dos dados, e, ao final, a variação da perda dos dados de validação não era mais significativa frente aos dados de treinamento, sendo este o momento de parada.

A Figura 6 exibe os valores preditos para os dados de treinamento. Como é possível observar, a arquitetura da rede recorrente LSTM conseguiu identificar a tendência geral dos dados, apesar de, por vezes, não ser capaz de prever o valor exato dos mesmos, especialmente nos picos. Este comportamento é esperado tanto pelos relatos da literatura bem como pelos métodos aplicados para evitar o *overfitting* da rede, como o *dropout*, e é consequência direta da Figura 5, onde o treinamento se encerra antes de que os dados de treinamentos fiquem extremamente bem adaptados, porém, os dados de validação não.

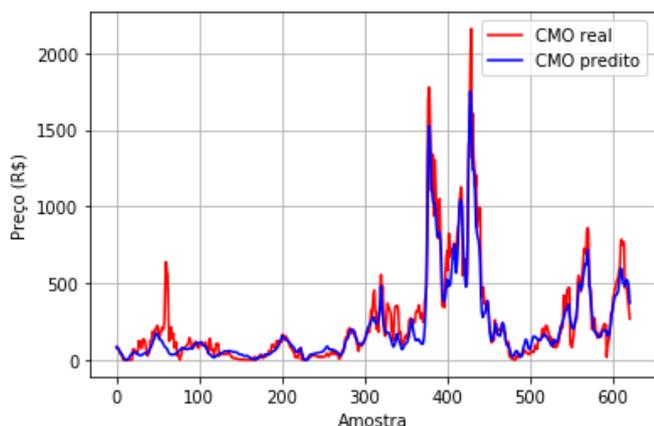


Figura 6 – Comparação dos dados reais e preditos do CMO para os dados de treino.

A partir dos valores exibidos anteriormente, utilizou-se a Equação 3 para filtrar as saídas da rede e obter os resultados definitivos para os valores preditos para os dados de treino. Com esses valores é possível quantizar o quão melhor foi a predição da arquitetura LSTM em relação ao modelo persistente, a partir da Equação 4, na Tabela 1.

**Tabela 1. Melhoria do modelo proposto em relação ao modelo persistente nos dados de treino**

Modelo	Média das Diferenças/Erros (R\$)	Melhoria
Persistente	42,93	11,3%
Proposto	38,08	

Como pode ser visto, a predição pela arquitetura LSTM proposta, com correção via Equação 3, para os dados de entrada selecionados, supera o modelo persistente em 11,3%, o que mostra a eficácia do modelo para os dados. Então, é obtido o histórico de saída da rede para os dados de avaliação, que não foram utilizados para treiná-la, conforme se observa na Figura 7.

É possível perceber na Figura 7 que, novamente, a rede foi capaz de acompanhar a tendência para os dados de validação, inclusive sem atrasos característicos do modelo persistente. Ainda assim, como esse histórico de saída não prediz os valores exatos do CMO, e sim sua tendência, utiliza-se a Equação 3 para se obter os resultados definitivos para as 52 semanas em questão.

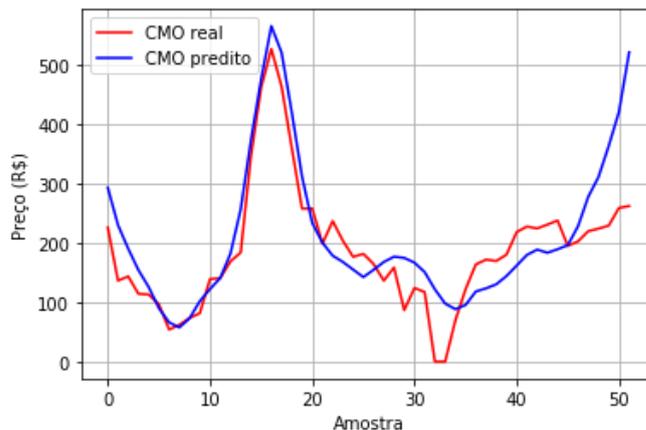


Figura 7 – Comparação dos dados reais e preditos para os dados escolhidos para validação.

Com o sucesso apresentado na Figura 7, fez-se então a predição dos dados. A Figura 8 mostra, após o tratamento via a filtragem (apresentada na Equação 3), como foi possível obter valores significativamente melhores que os do modelo persistente de comparação, (Chang et al., 2018). Tem-se resultados satisfatórios para grande parte da série de dados, conforme pode-se verificar ao longo das predições, especialmente nas primeiras 27 semanas (primeiros 6 meses). Os dados quantitativos referentes às métricas são apresentados na Tabela 2, com relação à média das diferenças.

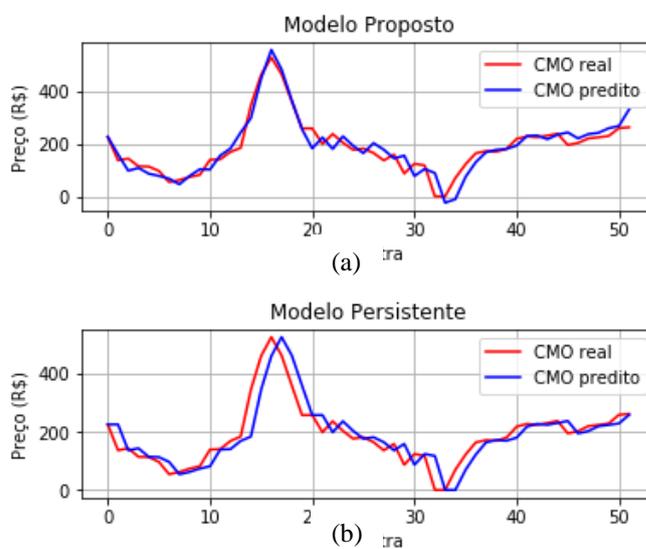


Figura 8 – Comparação dos dados reais e preditos com relação (a) aos dados de validação tratados pelo modelo proposto e (b) para o modelo persistente da literatura.

**Tabela 2. Melhoria do modelo proposto em relação ao modelo persistente nos dados de validação**

Modelo	Média das Diferenças/Erros (R\$)	Melhoria
Persistente	33,72	29,1%
Proposto	23,92	

Os dados mostram que para as 52 semanas (aproximadamente 1 ano) reservadas para a predição dos valores, a estratégia proposta por este trabalho foi capaz de superar o modelo persistente encontrado na literatura, a base comparativa deste trabalho, em 29,1%, e principalmente sem os atrasos, sendo uma melhoria significativa para a predição do PLD, para a quantidade de dados utilizada.

É preciso observar, porém, que, de acordo com o observado nos dados de treinamento, a arquitetura apresentada não é capaz de prever exatamente todas as orientações de mudança dos valores com perfeição, o que é um fato esperado. A partir disso, e em acórdância com a teoria do comportamento das perdas de validação e treinamento de redes neurais (Richard, 2000) (vide Figura 9), é esperado que, para uma série de validação grande o suficiente, a sua melhoria em relação ao modelo persistente diminua, uma vez que a persistência eventualmente aparece pontualmente em predições futuras, como é possível observar nas semanas finais da Figura 8 do modelo proposto.

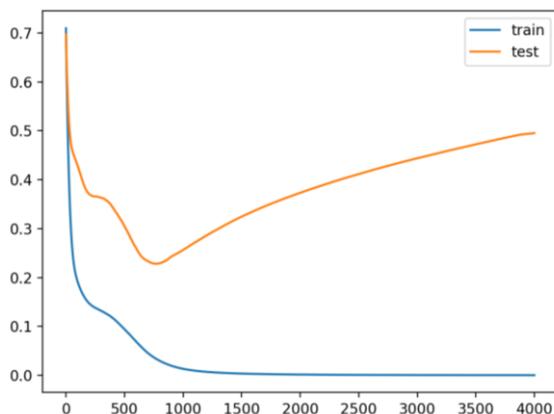


Figura 9 – Comportamento da curva de perdas (*Loss*) dos dados de treino e de validação, conforme o tamanho da série aumenta.

## 5. CONCLUSÕES

O presente trabalho apresentou uma nova estratégia para a abordagem da predição do PLD por meio de redes recursivas tipo LSTM, com a adição de um termo de filtragem das saídas da rede neural, atuando como um filtro estocástico. Além disso, a arquitetura não utiliza atrasos temporais da variável predita, e sim, as variáveis que dão sua origem, obtidas por meio de correlação entre ambas, o que pôde atenuar significativamente os atrasos característicos de modelos persistentes da literatura.

Das variáveis disponíveis no ONS, as que resultaram em maior correlação com o CMO foram as escolhidas. É possível concluir que essa abordagem se mostrou eficaz, dado que, comparado com o modelo persistente, tido aqui como linha de base comparativa, foi possível quantizar uma significativa melhora, de quase 30%, deste trabalho com relação ao que se tinha na literatura.

## AGRADECIMENTOS

Os autores agradecem à Universidade Federal de Juiz de Fora, órgão de apoio e fomento do projeto.

## REFERÊNCIAS

- ABRACEEL (2017), "Cartilha do Mercado Livre de Energia", Associação Brasileira dos Comercializadores de Energia. Disponível: <http://www.abraceel.com.br/>. Acesso: 08/06/2020.
- CCEE (2017), "Ambiente Livre e Ambiente Regulado", Câmara de Comercialização de Energia Elétrica. [Online]. Disponível: <http://www.ccee.org.br/>. Acesso: 08/06/2020.
- Chang, Z., Zhang, Y. e Chen, W. (2018), "Effective Adam-Optimized LSTM Neural Network for Electricity Price Forecasting", *IEEE 9th International Conference on Software Engineering and Service Science*.
- Conejo, A., Plazas, M., Espinola, R. e Molina, A. (2005), "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models", *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 1035-1042.
- EPE, "Matriz Energética e Elétrica", Empresa de Pesquisa Energética, [Online]. Disponível: <https://www.epe.gov.br/>. Acesso: 08/06/2020.
- Flovik, V. (2018), "How (not) to use Machine Learning for time series forecasting: Avoiding the pitfalls". [Online]. Disponível: <http://towardsdatascience.com/how-not-to-use-machine-learning-for-time-series-forecasting-avoiding-the-pitfalls-19f9d7adf424>. Acesso: 08/06/2020.
- Haykin, S. (2001), *Redes Neurais Princípios e Prática*, Bookman.
- Haykin, S. (2008), *Neural Networks and Learning Machines*, 3a edição, Prentice Hall.
- J. Reston Filho, C. Affonso e R. Oliveira. (2014), "Energy price prediction multi-step ahead using hybrid model in the Brazilian market", *Electric Power Systems Research*, vol. 17, pp. 115-122.
- Kingma, D. e Ba, J. (2015), "ADAM: a Method for Stochastic Optimization". *International Conference on Learning Representations*.
- Nascimento, F., Olivi, L. e Lima, L. (2018), "Multi-step forecasting of the Brazilian energy market price using autoregressive neural networks", *Simposio Brasileiro de Sistemas Eletricos (SBSE)*, Niteroi.
- ONSa, "Operador Nacional do Sistema Elétrico", Operador Nacional do Sistema Elétrico, [Online]. Disponível: <http://www.ons.org.br/>. Acesso: 08/06/2020.
- ONSb, "Histórico da Operação Nacional", Operador Nacional do Sistema Elétrico, [Online]. Disponível: <http://www.ons.org.br/>. Acesso: 08/06/2020.
- Richard O. Duda, Peter E. Hart e David G. Stork. (2000), *Pattern Classification*, 2a edição, Wiley-Interscience.
- Silva, I.N. (2010), *Redes Neurais Artificiais para Engenharia e Ciências Aplicadas*. Curso Prático, ArtLiber.
- Yan, X. e Chowdhury, N. (2014), "Mid-term electricity market clearing price forecasting utilizing hybrid support vector machine and auto-regressive moving average with external input", *International Journal of Electrical Power & Energy Systems*, vol. 63, pp. 64-70.