

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Fabiane Barbosa do Nascimento

**Estratégias de Predição de Preços do Mercado Livre de Energia por Redes
Neurais Artificiais e Filtragem Estocástica**

Juiz de Fora

2021

Fabiane Barbosa do Nascimento

**Estratégias de Predição de Preços do Mercado Livre de Energia por Redes
Neurais Artificiais e Filtragem Estocástica**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, da Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito para a obtenção do título de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Leonardo Willer de Oliveira

Coorientador: Prof. Dr. Ivo Chaves da Silva Junior

Juiz de Fora

2021

Fabiane Barbosa do Nascimento

Estratégias de Predição de Preços do Mercado Livre de Energia por Redes Neurais Artificiais e Filtragem Estocástica

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica, da Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito para a obtenção do título de Mestre em Engenharia Elétrica.

Aprovada em 27 de Abril de 2021.

BANCA EXAMINADORA



Prof. Dr. Leonardo Willer de Oliveira - Orientador
Universidade Federal de Juiz de Fora



Prof. Dr. Ivo Chaves da Silva Junior - Coorientador
Universidade Federal de Juiz e Fora



Assinado de forma digital por
VITOR HUGO FERREIRA
vhferreira@d.uff.br.01277679673
Dados: 2021.04.30 11:09:39 -03'00'

Prof. Dr. Vitor Hugo Ferreira - 1º Titular Externo
Universidade Federal Fluminense



Prof. Dr. André Luís M. Marcato - 1º Titular Interno
Universidade Federal de Juiz e Fora



Prof. Dr. Leonardo Rocha Olivi - 2º Titular Interno
Universidade Federal de Juiz e Fora

AGRADECIMENTOS

Agradeço primeiramente a Deus, por jamais me deixar nas horas difíceis e por ter planos para a minha vida que são sempre maiores do que os meus sonhos.

À minha família, pelas palavras de incentivo, por sempre acreditarem em mim e sonharem junto comigo.

Ao professor Leonardo Rocha Olivi, pela grande e positiva influência que efetivamente exerceu em minha vida pessoal e profissional. Pelo apoio incondicional, por sempre estar disposto a me ensinar e ajudar.

Agradeço aos professores orientadores Leonardo Willer de Oliveira e Ivo Chaves da Silva Júnior pela oportunidade, confiança, paciência e apoio oferecido no desenvolvimento desta Dissertação.

Agradeço também, a disponibilidade dos membros da banca por fazerem parte desse momento trazendo contribuições para a melhoria do trabalho desenvolvido.

Agradeço à Denise por permitir que eu iniciasse este sonho, me liberando para a realização das atividades, além do incentivo e apoio constantes dados em minha carreira.

À Universidade Federal de Juiz de Fora e ao Programa de Pós-Graduação de Engenharia Elétrica pelo apoio a este trabalho.

E a todas as outras pessoas que, das mais diferentes maneiras, participaram desta história.

“A persistência é o menor caminho para o êxito.”
Charles Chaplin

RESUMO

O Preço de Liquidação das Diferenças (PLD) é utilizado como base na comercialização de energia elétrica no mercado de curto prazo. O conhecimento dos seus valores futuros e tendências reduz a insegurança na tomada de decisões, permitindo que os agentes de mercado decidam as estratégias mais adequadas e estabeleçam contratações bem-sucedidas, maximizando os lucros e minimizando os riscos de seus empreendimentos. O PLD é, no entanto, influenciado por diversos fatores ligados principalmente às incertezas das demandas e à hidrologia bem como fatores sociais e políticos, o que lhe confere alta volatilidade, sazonalidade, saturação de valores e comportamento estocástico. Desta forma, o presente trabalho propõe uma abordagem investigativa para a predição das tendências futuras do PLD por meio de Redes Neurais Artificiais e Filtragem Estocástica, com a motivação de auxiliar o processo de decisão da compra de energia no mercado de curto prazo. Tal abordagem consiste na combinação de uma rede neural do tipo recorrente com um filtro na saída para atenuar os ruídos inerentes aos dados auxiliando na preservação das tendências dos dados futuros. A rede foi treinada por meio do algoritmo *ADAM* e é aplicada ao mercado brasileiro de energia. A escolha dos dados de entrada da rede foi feita a partir de análises estatísticas de correlação cruzada entre as séries temporais das variáveis de interesse. Os resultados mostram uma boa capacidade de predição com um bom acompanhamento das tendências da variável de interesse ao longo do ano, oferecendo seus principais comportamentos e tendências aos operadores de mercado, auxiliando-os nas tomadas de decisões de compra e venda de energia.

Palavras-chave: Preço de Liquidação das Diferenças, Modelos de Predição, Predição de séries Temporais, Redes Neurais Artificiais, Filtragem Estocástica.

ABSTRACT

The Settlement Price of Differences (PLD) is used as a basis for the sale of electrical energy in the short-term market. The knowledge of its future values and trends mitigates insecurity in decision making, allowing market agents to decide the most appropriate strategies and establish successful contracts, maximizing profits and minimizing the risks of their ventures. PLD is, however, influenced by several factors linked mainly to the uncertainties of demands and hydrology as well as social and political factors, which gives it high volatility, seasonality, saturation of values and stochastic behavior. In this way, the present work proposes an investigative approach for the prediction of future trends of PLD through Artificial Neural Networks and Stochastic Filtration, with the motivation to assist the decision process of purchasing energy in the short-term market. Such an approach consists of combining a recurrent neural network with an output filter to attenuate the noise inherent in the data, helping to preserve future data trends. The network was trained using the ADAM algorithm and is applied to the Brazilian energy market. The choice of network input data was made based on statistical analyzes of cross-correlation between the time series of the variables of interest. The results show a good predictive capacity with a good monitoring of the trends of the variable of interest throughout the year, offering its main behaviors and trends to market operators, assisting them in making energy purchase and sale decisions.

Keywords: Settlement Price of Differences, Forecasting Models, Time Series Forecasting, Artificial Neural Networks, Stochastic Filtering.

LISTA DE ILUSTRAÇÕES

Figura 1.1 – Comparação CMO e PLD.	15
Figura 2.1 – Estrutura do Setor Elétrico Brasileiro.	21
Figura 2.2 – Cronograma de decisão para um sistema hidrotérmico.	23
Figura 2.3 – Regiões do Sistema Interligado Nacional.	24
Figura 2.4 – Topologia de rede recorrente.	28
Figura 3.1 – PLD médio de todos os submercados.	37
Figura 3.2 – Rede com filtragem antes da realimentação.	39
Figura 3.3 – Convergência dos algoritmos de treinamento.	40
Figura 3.4 – Seno contaminado por ruído aleatório Gaussiano.	41
Figura 3.5 – Exponencial negativa multiplicada por seno.	42
Figura 3.6 – Composição de senoides de diferentes frequências.	42
Figura 3.7 – Séries de dados da ONS e suas unidades: CE [MWmed], DM [MWh/h], EA [MWmês], ENA [MWmed], GE [MWmed] e IE [MWmed].	43
Figura 3.8 – Série de dados – CMO e PLD.	44
Figura 3.9 – Função ReLU e sua derivada.	45
Figura 4.1 – Predição para $N_1=N_2=20$ e $n_p=n_q=10$	51
Figura 4.2 – Predição para $N_1=N_2=20$ e $n_p=n_q=20$	51
Figura 4.3 – Predição para $N_1=10$ $N_2=20$ $N_3=15$ e $n_p=n_q=10$	52
Figura 4.4 – Predição para $N_1=10$ $N_2=20$ $N_3=15$ e $n_p=n_q=20$	52
Figura 4.5 – Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=30$	53
Figura 4.6 – Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=30$	53
Figura 4.7 – Predição para $N_1=N_2=20$ e $n_p=n_q=2$	55
Figura 4.8 – Predição para $N_1=10$, $N_2=20$, $N_3=15$ e $n_p=n_q=2$	55
Figura 4.9 – Correlação cruzada do CMO para janelas de dados da série temporal, evidenciando a correlação entre dados da mesma série temporal em tempos diferentes.	57
Figura 4.10–Correlação cruzada: esquerda) CMO e direita) ruído Gaussiano, eviden- ciando a baixa correlação do CMO para janelas em instantes de tempos distantes, de maneira similar a variável aleatória, que não uma possui alta intensidade de estruturas repetidas ou periódicas no longo prazo.	57
Figura 4.11–Correlação das variáveis com o CMO para janelamentos de até 730 dados.	59
Figura 4.12–Correlação das variáveis com o CMO para janelamentos de até 30 dados.	59
Figura 4.13–Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=20$	61
Figura 4.14–Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=20$	62
Figura 4.15–Predição para $N_1=N_2=20$ e $n_p=n_q=20$	62
Figura 4.16–Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=2$	63

Figura 4.17–Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=10$	64
Figura 4.18–Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=2$	66
Figura 4.19–Predição para $N_1=N_2=20$ e $n_p=n_q=10$	68

LISTA DE TABELAS

Tabela 4.1 – Resultados por topologia - 1 passo no futuro	50
Tabela 4.2 – Resultados por topologia - vários passos no futuro (N=53 passos) . . .	54
Tabela 4.3 – Resultados por topologia (CE, DM, EA, GE, CMO e PLD) - vários passos no futuro (1 ano)	60
Tabela 4.4 – Resultados por topologia (CE, DM, EA, GE, IE, ENA, CMO e PLD) - vários passos no futuro (1 ano)	63
Tabela 4.5 – Comparação dos resultados utilizando ou não as variáveis IE e ENA . .	65
Tabela 4.6 – Resultados por topologia (CE, DM, EA, GE, IE, ENA, CMO e PLD) - vários passos no futuro com filtragem (1 ano)	67

LISTA DE ABREVIATURAS E SIGLAS

ACL	Ambiente de Contratação Livre
ACR	Ambiente de Contratação Regulada
AdaGrad	<i>Adaptive Gradient Algorithm</i>
ADAM	<i>Adaptive Moment Estimation</i>
ANEEL	Agência Nacional de Energia Elétrica
ARIMA	<i>Autoregressive integrated moving average</i>
CCEE	Câmara de Comercialização de Energia Elétrica
CE	Carga de Energia
CMO	Custo Marginal da Operação
CMSE	Comitê de Monitoramento do Setor Elétrico
CNPE	Conselho Nacional de Políticas Energéticas
DM	Demanda Máxima
EA	Energia Armazenada
ENA	Energia Natural Afluente
EPE	Empresa de Pesquisa Energética
FIR	<i>Finite Impulse Response</i>
GE	Geração de Energia
IE	Intercâmbio de Energia
IGP-M	Índice Geral de Preços Mercado
IIR	<i>Infinite Impulse Response</i>
LMS	<i>Least Mean Square</i>
MAE	Mercado Atacadista de Energia Elétrica
MME	Ministério de Minas e Energia
ONS	Operador Nacional do Sistema
PLD	Preço de Liquidação das Diferenças

PCH	Pequenas Centrais Hidrelétricas
RE-SEB	Reestruturação do Setor Elétrico Brasileiro
ReLU	<i>Rectified Linear Unit</i>
RMSE	<i>Root Mean Squared Error</i>
RNA	Rede Neural Artificial
TDNN	<i>Time Delay Network</i>

LISTA DE SÍMBOLOS

n_p	Amostras de entrada atrasadas
n_q	Amostras de saídas passadas realimentadas
$data_{norm}$	Função de treinamento normalizada
max_{norm}	Valor máximo permitido para a normalização
min_{norm}	Valor mínimo permitido para a normalização
max_{data}	Valor máximo permitido para a série temporal
min_{data}	Valor mínimo permitido para a série temporal
$data$	série temporal
m_t	Primeiro momento
v_t	Segundo momento
β_1	Taxa de aprendizado do primeiro momento
β_2	Taxa de aprendizado do segundo momento
g_t	Gradiente
\hat{m}_t	Primeiro momento corrigido
\hat{v}_t	Segundo momento corrigido
Θ_t	Parâmetro de ajuste dos pesos
P_i	i-ésimo valor desejado pela rede
Q_i	i-ésimo valor obtido pela rede
N_1	Número de neurônios na primeira camada escondida
N_2	Número de neurônios na segunda camada escondida
N_3	Número de neurônios na terceira camada escondida
n	Quantidade de amostras do conjunto

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos e motivações	17
1.2	Produção Científica	18
1.3	Estrutura da dissertação	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	O Setor Elétrico Brasileiro	19
2.2	A Formação do Preço de Liquidação das Diferenças	21
2.3	Predição do PLD	24
2.4	Redes Neurais Artificiais	25
2.4.1	Redes Recorrentes	27
2.4.2	Métodos de Treinamento	28
2.5	Filtragem	32
2.5.1	Filtros Digitais	32
2.5.2	Filtro de Média Móvel	33
2.5.3	Filtro de Kalman	34
3	METODOLOGIA PROPOSTA	37
3.1	Escolha dos Dados de Entrada	37
3.2	Escolha do Método de Treinamento	39
3.3	Processamento das Variáveis	43
4	RESULTADOS E ANÁLISES	48
4.1	CMO como única variável de entrada	49
4.1.1	Predição de um passo no futuro	49
4.1.2	Predição de vários passos no futuro	54
4.2	Variáveis ONS como variáveis de entrada	58
4.3	Filtragem da saída da Rede Neural	65
4.4	Conclusões sobre as análises dos resultados	68
5	CONCLUSÕES	70
5.1	Propostas de desenvolvimentos futuros	71
	REFERÊNCIAS	72

1 INTRODUÇÃO

A energia elétrica é um bem considerado essencial e indispensável, e é por meio dela que se desenvolvem as principais atividades econômicas de qualquer país. O Brasil desenvolveu seu setor elétrico primariamente por meio de energia hidrelétrica, porém, também possui termelétricas, usinas eólicas e importação da energia de outros países. O setor elétrico brasileiro é estruturado de forma a garantir a segurança do suprimento de energia elétrica, promover a inserção social e a modicidade tarifária (CCEE, 2021f). Em referência aos mecanismos regulatórios e operacionais, este setor passou por diversas modificações ao longo dos anos.

O modelo setorial vigente é liberal e flexível, onde os consumidores livres e especiais podem escolher o fornecedor de energia. As relações comerciais se estabelecem em dois ambientes de contratação: o Ambiente de Contratação Regulado (ACR) e o Ambiente de Contratação Livre (ACL). No ACR, a energia é comprada pelas distribuidoras para os consumidores cativos via leilões e o preço é determinado pela Agência Nacional de Energia Elétrica (ANEEL). Já no ACL, as transações são feitas de forma bilateral entre os consumidores livres, geradores e comercializadores.

Há ainda o mercado de curto prazo, também conhecido como mercado de diferenças, onde se procede a contabilização e liquidação financeira das diferenças apuradas entre os montantes de energia elétrica contratados e verificados (gerados e consumidos) (ANEEL, 2021). Essas diferenças são valoradas por meio do Preço de Liquidação das Diferenças (PLD). A maneira como ele é calculado está intrinsecamente conectada com as características operativas do setor elétrico brasileiro, com seu parque de geração preponderantemente hidrelétrico.

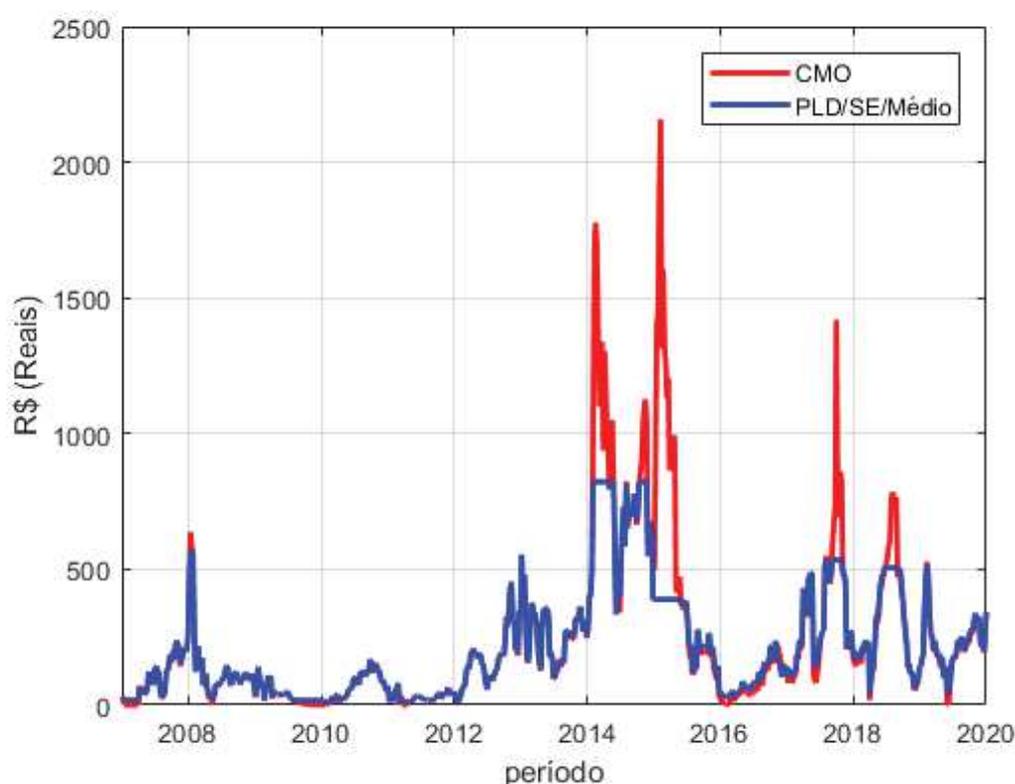
Diante do supracitado, para o cálculo do PLD são utilizados modelos computacionais que objetivam a solução ótima de equilíbrio entre utilizar a energia provida das hidrelétricas ou o armazenamento da água para utilização futura. Estes modelos se baseiam principalmente em parâmetros como condições hidrológicas, demanda de energia, preços de combustível, custo do deficit, entrada de novos projetos e disponibilidade de equipamentos de geração e transmissão para obter o despacho ótimo, estabelecendo a geração hidráulica e térmica para cada submercado no período em estudo (CCEE, 2021d).

Na cadeia de modelos, os estudos são coordenados entre si com diferentes horizontes de análise e distintos graus de detalhamento, representados pelos modelos computacionais NEWAVE, DECOMP E DESSEM, que têm como resultado o Custo Marginal da Operação (CMO) para cada submercado em base mensal, semanal e diária, respectivamente. O CMO é a base para a determinação do PLD. A principal diferença entre eles é que o PLD é limitado por preços máximo e mínimo estabelecidos anualmente pela ANEEL. Além disso, para calcular o PLD, são consideradas restrições de transmissão apenas entre

submercados, ao passo que o CMO leva em conta restrições de transmissão tanto entre os submercados quanto internas aos mesmos. Vale ressaltar que o cálculo é baseado no despacho ex-ante, ou seja, antes da operação real do sistema.

A seguir, a Figura 1.1 representa o comportamento do PLD e do CMO para o submercados SE/CO, no período de janeiro/07 a dezembro/20, na qual é possível verificar de maneira qualitativa a elevada correlação entre o CMO e o PLD, bem como a aplicação dos limites máximos e mínimos do PLD, em cada ano, resultando nas suas saturações.

Figura 1.1 – Comparação CMO e PLD.



Fonte: (ONS, 2020) e (CCEE, 2021d).

De acordo com (CCEE, 2021e), os agentes atuantes nas áreas de geração, distribuição e comercialização de energia elétrica contabilizam 10.615 atualmente. Destes, 1.014 são consumidores livres, que podem negociar energia com qualquer fonte, 7461 consumidores especiais que podem contratar energia advinda de fontes limpas, como solar, eólica e pequenas centrais hidrelétricas (PCH), 391 comercializadores que são os intermediários entre o gerador e o consumidor final, 1.701 geradores e 48 distribuidores.

Ao longo dos anos, a migração para o Mercado Livre aumentou consideravelmente. Cada vez mais o consumidor se interessa por este mercado e quer operar neste modelo contratual, buscado vantagens econômicas. No entanto, a alta volatilidade do PLD, observada na Figura 1.1, é uma das maiores preocupações dos agentes do setor elétrico brasileiro, fundamentalmente no ACL. Os agentes que operam neste ambiente de contratação detêm

diferentes estratégias no contexto da comercialização de energia, visando a maximização dos lucros. Quando ocorrem variações não esperadas no PLD, pode ocorrer um impacto financeiro significativo.

Diante do exposto, é evidenciada a importância de uma melhor compreensão de como o PLD se comporta, de forma a reduzir as suas incertezas nas análises dos agentes do setor, com a motivação de auxiliar na elaboração de estratégias de comercialização de energia, bem como reduzir os riscos envolvidos no processo de tomada de decisão. Uma vez que o PLD é contaminado por ruídos aleatórios, especialmente os de causas naturais hidrológicas, este torna-se um problema estocástico.

Assim, o problema probabilístico relatado centra-se na obtenção de um modelo de predição de resultados futuros. Para isso, faz-se necessário estudar técnicas que atendam a esta especificação. As Redes Neurais Artificiais (RNA) têm grande aplicabilidade, largamente relatada na literatura, no que concerne a predição de séries temporais, sendo desenvolvidas com base em modelos de sistemas de neurônios biológicos. Desta maneira, o presente trabalho propõe a aplicação de RNAs recorrentes, utilizando como entradas:

- i PLD (Preço de Liquidação das Diferenças) (CCEE, 2021b);
- ii Carga de Energia (CE) (ONS, 2021a);
- iii Demanda Máxima (DM) (ONS, 2021c);
- iv Energia Armazenada (EA) (ONS, 2021d);
- v Energia Natural Afluente (ENA) (ONS, 2021e);
- vi Geração de Energia (GE) (ONS, 2021f);
- vii Intercambio de Energia (IE) (ONS, 2021g), e;
- viii Custo Marginal de Operação (CMO) (ONS, 2021b).

Conforme citado anteriormente, o PLD possui grande relação com o CMO, diferindo nos momentos em que os valores atingem os limites estabelecidos pela ANEEL. Nestes momentos específicos de saturação de preços, os modelos de identificação de sistemas, tais como as Redes Neurais Artificiais, ficam sem dados de referência para o aprendizado, uma vez que dados constantes não permitem aos modelos o conhecimento da próxima variação. Assim, optou-se pela predição do CMO, como forma de conhecer a tendência de variação do PLD.

Além disso, para aumentar a investigação de possibilidades na exploração da tendência futura do CMO, este trabalho também propôs a inserção de filtragem dos dados de saída da RNA recorrente. Este procedimento se deu de maneira a realimentá-la

com suas saídas filtradas, para aumentar a acurácia na preservação das tendências dos dados, de modo que as variações decorrentes da alta volatilidade intrínseca dos dados não causassem o acúmulo exacerbado de erros na fase de treinamento da RNA. Assim, o objetivo principal da rede não é a predição exata do preço do CMO em uma determinada semana do futuro, e sim, acurada demonstração da tendência de preços, permitindo um planejamento antecipado por parte dos agentes do setor.

1.1 Objetivos e motivações

Diante do contexto descrito anteriormente, esta pesquisa tem como objetivo a investigação da tendência de variação de Preços no Mercado Livre de Energia, a partir da Predição do Custo Marginal de Operação por meio de Redes Neurais Artificiais Recorrentes com a inserção de filtragem dos dados de saída da RNA. A motivação é desenvolver uma metodologia de predição que auxilie as tomadas de decisões dos operadores do mercado livre de energia nos momentos de firmar contratos de compra e venda.

Nos trabalhos desenvolvidos por (NASCIMENTO, 2017), (NASCIMENTO; OLIVI; LIMA, 2018) e (NASCIMENTO et al., 2019), Redes Neurais Artificiais Recorrentes usando como fonte de conhecimento sobre os dados o próprio histórico (série temporal) de preços do PLD se mostraram eficientes na predição do PLD uma semana à frente por meio do treinamento *Resilient Backpropagation*. Para a predição de períodos maiores, a falta de informações na própria série temporal dificulta a obtenção de boas predições, já que o PLD não possui padrões cíclicos (que favorecem o aprendizado). Além disso, as saturações do PLD causadas pelos limites mínimo e máximo estabelecidos pela ANEEL fazem com que a RNA perca a referência, uma vez que dados constantes não permitem aos modelos o conhecimento da próxima variação, além da integração dos erros e da alta volatilidade dos dados. Esses aspectos são tratados na presente dissertação, através dos objetivos específicos definidos a seguir:

- i Utilizar a previsão do CMO para conhecer a tendência de variação do PLD;
- ii Inserção de mais variáveis como entrada da Rede Neural;
- iii Análise estatística de correlação das variáveis de entrada da Rede Neural;
- iv Análise de novos métodos de treinamento de Redes Neurais Artificiais;
- v Propor um filtro na saída da Rede Neural Recorrente que auxilie na preservação das tendências dos dados, e;
- vi Realizar simulações computacionais para verificação do funcionamento de todas as topologias estudadas e abordadas neste estudo.

1.2 Produção Científica

O trabalho realizado durante o Mestrado discutido nesta dissertação resultou na publicação de um artigo sobre o mesmo assunto.

- NASCIMENTO, F. B. do.; OLIVI, L. R.; LIMA, L. H. L., OLIVEIRA, L. W.; JÚNIOR, I. C. S. *Weekly forecast of the Settlement Price of Differences in the Brazilian short-term market through recurrent artificial neural networks*. Em: **Anais do 14º Simpósio Brasileiro de Automação Inteligente**, 2019, Ouro Preto.

1.3 Estrutura da dissertação

A organização do presente trabalho se dá em cinco capítulos. O primeiro capítulo trata-se desta introdução, apresentando o tema proposto. O segundo capítulo trata de uma revisão bibliográfica da literatura e dos temas abordados no trabalho. O terceiro capítulo descreverá o desenvolvimento técnico utilizado na resolução do problema. O quarto capítulo mostra os resultados e suas implicações no tema proposto. O quinto capítulo conclui os resultados e aponta trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo faz uma revisão técnica acerca dos assuntos que serão tratados neste trabalho. Envolve uma descrição do setor elétrico brasileiro, o foco de aplicação do trabalho, e também uma descrição das técnicas propostas para aplicação ao problema de predição de preços no mercado livre de energia.

2.1 O Setor Elétrico Brasileiro

A atual estrutura do Setor Elétrico Brasileiro teve início com as reformas ocorridas no setor iniciadas na década de 90. Essas reformas foram motivadas basicamente pela ineficiência do modelo institucional existente, que colocava em risco a oferta e expansão do sistema energético.

Até 1995, o modelo do Setor Elétrico era composto basicamente por empresas verticalizadas, predominantemente estatais, onde o financiamento era feito através de recursos públicos. Toda atividade relacionada a energia elétrica era monopolizada, não havia competição, todos os consumidores eram cativos e o mercado era completamente regulado.

Em 07 de julho de 1995 ocorreu uma importante modificação institucional na estrutura do Setor Elétrico, quando foi promulgada a Lei nº 9.074 (BRASIL, 1995) que estabeleceu a criação do Produtor Independente de Energia e do Consumidor Livre de Energia, estabelecendo uma nova formatação da comercialização de energia elétrica no país. Esta Lei inseriu a possibilidade de uma empresa privada produzir e comercializar energia elétrica e estabeleceu os primeiros passos rumo à competição da comercialização de energia elétrica.

Essa necessidade de atrair investimentos privados se tornou tão evidente que em 1996, consultores nacionais e internacionais, liderados pela Coopers&Lybrand foram contratados para a condução do programa de reforma do Setor Elétrico. Os resultados deste trabalho foram consolidados no Projeto Reestruturação do Setor Elétrico Brasileiro (RE-SEB), coordenado pelo Ministério de Minas e Energia (MME). Foi a primeira tentativa de estabelecer um modelo sustentável para a comercialização de energia no país e que formou as bases para o modelo institucional brasileiro existente hoje.

O RE-SEB previa a separação das empresas de energia em áreas específicas: geração, transmissão, distribuição e comercialização de energia, e incentivava a competição nos segmentos de geração e comercialização, mantendo sob regulação do Estado os setores de distribuição e transmissão de energia elétrica (SOUZA, 2008).

Essa necessidade de desverticalização das empresas de energia e incentivo de competição na geração e comercialização apresentou a necessidade de criação de uma

estrutura regulatória. Para atender este anseio houve, em 1997, a constituição da Agência Nacional de Energia Elétrica (ANEEL) por meio do Decreto 2.335 (BRASIL, 1997a) e a criação do Conselho Nacional de Políticas Energéticas com a Lei 9.478 (BRASIL, 1997b).

Em consonância com a ANEEL criou-se, em 1998, o Operador Nacional do Sistema (ONS) que cria e aplica regras e normas para o gerenciamento harmônico da produção e transmissão de eletricidade e o Mercado Atacadista de Energia Elétrica (MAE), um ambiente para a realização das transações de compra e venda de energia elétrica.

O projeto RE-SEB findou-se em 1998 e as principais conclusões do mesmo foram: a necessidade de desverticalizar as empresas de energia elétrica, incentivar a competição na geração e comercialização e manter sob regulação os setores de transmissão e distribuição de energia elétrica. Contudo, a grave crise de abastecimento sofrida em 2001 pelo Setor Elétrico resultou na implementação de um plano de racionamento de energia elétrica bem como em propostas de alterações no modelo proposto, criadas pelo Comitê de Revitalização do Modelo do Setor Elétrico (NASCIMENTO, 2017).

Em 2003, obtiveram-se as premissas para o novo modelo institucional, que foram publicadas por meio das Medidas Provisórias 144 e 145. A Medida Provisória 144 (BRASIL, 2003a) dispunha sobre a comercialização da energia elétrica e, para sua comercialização criou-se uma instituição no sistema interligado dando continuidade às atividades do MAE, a Câmara de Comercialização de Energia Elétrica (CCEE) e a Medida Provisória 145 (BRASIL, 2003b) criava a Empresa de Pesquisa Energética (EPE), entidade responsável pelo planejamento do Setor Elétrico no longo Prazo.

Em 2004, as Medidas Provisórias 144 e 145 foram convertidas nas Leis 10.848 (BRASIL, 2004e) e 10.847 (BRASIL, 2004d), respectivamente. Estas leis são a base institucional e legal do novo modelo do Setor Elétrico Brasileiro. Ainda neste ano, o Comitê de Monitoramento do Setor Elétrico (CMSE) foi constituído pelo Decreto 5.175 (BRASIL, 2004c), houve também a regulamentação do Operador Nacional do Sistema (ONS) por meio do Decreto 5.081 (BRASIL, 2004a).

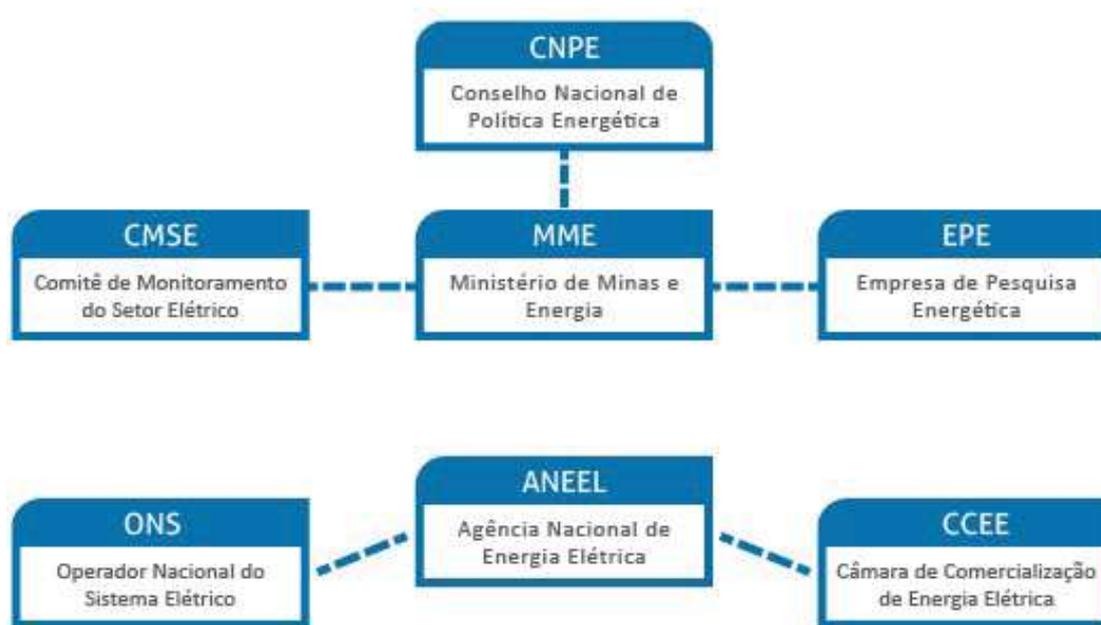
Em relação à comercialização, o Decreto 5.163 (BRASIL, 2004b), em julho de 2004, separou a comercialização em dois ambientes para a celebração de contratos de compra e venda de energia, com lógicas e estruturas distintas: O Ambiente de Contratação Livre (ACL) e o Ambiente de Contratação Regulada (ACR). No ACR as operações de compra e venda acontecem de forma a atender o mercado cativo e o ACL é o segmento do mercado onde acontecem as operações de compra e venda de energia livremente negociadas, firmadas por meio de contratos bilaterais.

As negociações nos dois ambientes acontecem conforme as regras e procedimentos de comercialização da CCEE, regulamentados através de resoluções da ANEEL. Além disso, também foi criado o Mercado de Curto Prazo, conhecido como mercado das diferenças

onde ocorre o ajuste entre os volumes contratados e os volumes medidos de energia.

Um outro ponto importante do Decreto 5.163 é a separação de atribuições e deveres em diferentes instituições. O mapeamento organizacional apresentado na Figura 2.1 integra o modelo institucional vigente, implantado em 2004, o qual é fruto de um aprimoramento originado em 1998 com o projeto RE-SEB.

Figura 2.1 – Estrutura do Setor Elétrico Brasileiro.



Fonte: (CCEE, 2021a).

O Conselho Nacional de Políticas Energéticas (CNPE) é responsável por traçar políticas energéticas nacionais e quem realiza a gestão destas políticas é o Ministério de Minas e Energia (MME). O acompanhamento e avaliação da segurança e continuidade no fornecimento de energia é responsabilidade do Comitê de Monitoramento do Setor Elétrico (CMSE).

A Agência Nacional de Energia Elétrica (ANEEL) regula a fiscalização no setor. A Empresa de Pesquisa Energética (EPE) faz o planejamento energético da expansão da oferta de energia no país, com estudos e pesquisas que ajudam a traçar as políticas e realizar a gestão. A operação física do sistema é feita pelo Operador Nacional do Sistema (ONS) e por fim, a Câmara de Comercialização de Energia Elétrica (CCEE) viabiliza as operações do mercado de energia elétrica, administrando as transações e realizando os leilões.

2.2 A Formação do Preço de Liquidação das Diferenças

No sistema elétrico brasileiro, a decisão sobre o despacho e a transmissão de energia elétrica é feita de forma centralizada pelo Operador Nacional do Sistema (ONS) que

faz a programação da geração com base em otimização de custos. A metodologia para determinação do PLD é operacionalizada por modelos computacionais, que representam algoritmos de programação dinâmica dual estocástica. O NEWAVE é utilizado para programação de médio e longo prazo (até 5 anos) com discretização mensal. O modelo DECOMP com programação de prazo mais curto (2 meses) com discretização semanal e por fim, o modelo DESSEM que é o desdobramento de uma programação a curtíssimo prazo (até 1 semana) com discretização horária.

Esta cadeia de programas tem como função a gestão dos recursos energéticos do país de forma centralizada. Pretende-se minimizar o uso dos recursos energéticos atendendo a dois objetivos conflitantes: Assegurar o atendimento presente e futuro da carga, ou seja, evitar a possibilidade de racionamento e operar ao menor custo total. Sendo assim, a decisão sobre a quantidade de energia que cada usina ofertará, cabe ao ONS que libera o despacho de cada usina com base em uma ordem de mínimo custo até que seja possível atender completamente a demanda.

O sistema de geração de energia elétrica brasileiro é hidrotérmico com grande predominância de hidráulicas. A análise de seu comportamento é probabilística, pois a decisão sobre a utilização da água armazenada nos reservatórios (estoques de energia) está intimamente relacionada com a incerteza das afluições futuras.

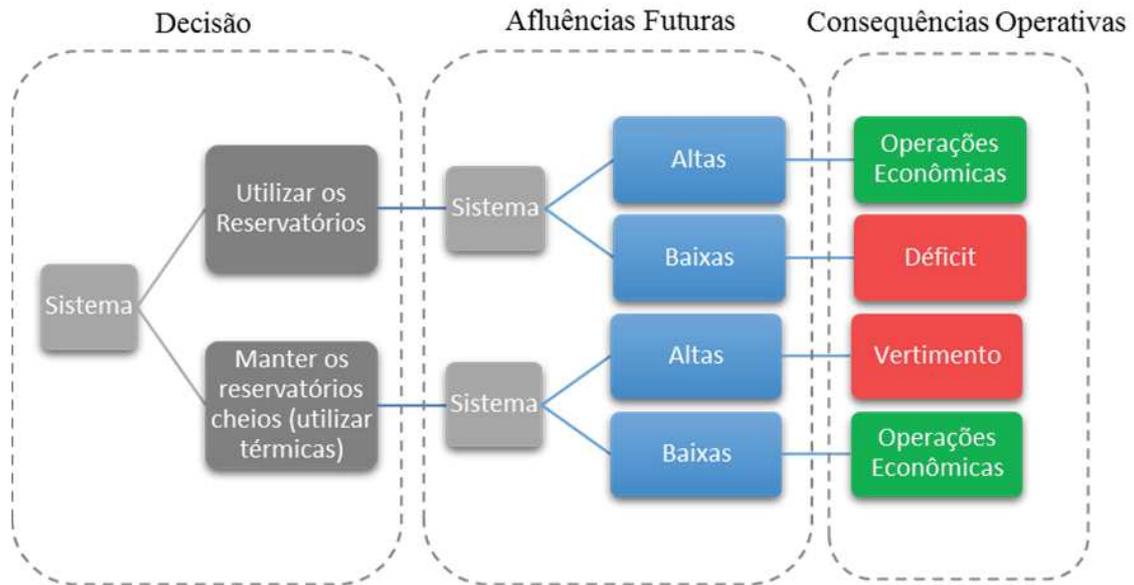
A decisão de minimizar o consumo das termelétricas com despacho intensivo das hidrelétricas implica em menores custos, devido à economia de combustíveis, mas tende a aumentar o custo de operação futuro, em especial se as afluições hidrológicas forem baixas. Por outro lado, despachar de forma acentuada as termelétricas minimiza o consumo da água, preservando os níveis dos reservatórios hidrelétricos. Esta política ameniza o aumento do custo de operação futuro caso a previsão de chuvas seja desfavorável, mas implica no aumento do custo de operação em curto prazo devido à necessidade de despachar mais termelétricas. A Figura 2.2 mostra este processo de decisão e planejamento.

Conforme os autores de (SOUZA, 2013) a energia elétrica no Brasil possui características singulares, as quais pode-se citar:

- Não é possível armazená-la ou estocá-la no atacado para futuro uso;
- Incerteza da demanda;
- Dependente da hidrologia;
- Expansão da geração exige longo prazo.

Diante do exposto, o programa computacional conhece não somente a configuração do sistema, mas também as características passadas da hidrologia. Então ele projeta a variabilidade das hidrologias já verificadas no passado para o período a frente e verifica os

Figura 2.2 – Cronograma de decisão para um sistema hidrotérmico.



Fonte: (NASCIMENTO, 2017).

possíveis cenários hidrológicos futuros. A operação é simulada nesses cenários, objetivando uma tomada de decisão de operação que minimize o custo total. Além da otimização do parque gerador, é considerada a transmissão da energia.

O Brasil é territorialmente muito vasto com um extenso sistema de transmissão conectando as diversas regiões do país, que são divididas em 4 subsistemas: Norte, Nordeste, Sudeste/Centro-Oeste e Sul, conforme evidenciado na Figura 2.3. A interligação entre os subsistemas permite que os custos sejam reduzidos, por meio do intercâmbio de energia. Este intercâmbio também é analisado nas simulações.

Em síntese, os algoritmos de programação levam em consideração uma série de variáveis representativas das atuais condições hidrológicas do sistema, bem como as condições da trajetória de demanda de expansão física do sistema. De acordo com o CCEE (CCEE, 2021c), o NEWAVE tem o objetivo de determinar a estratégia de geração hidráulica e térmica minimizando o valor esperado do custo de operação para todo o período de planejamento. Um dos principais resultados desse modelo são as funções de custo futuro, que traduzem para o modelo de curto prazo o impacto da utilização da água armazenada nos reservatórios. O objetivo do DECOMP é determinar o despacho de geração das usinas hidrelétricas e termelétricas que minimiza o custo de operação ao longo do período de planejamento, dado o conjunto de informações disponíveis (previsões de cargas, vazões, disponibilidades, limites de transmissão entre submercados, função de custo futuro do NEWAVE, etc). Este modelo fornece a Função de Custo Futuro, resultado da estratégia de solução do curto prazo, para o acoplamento com o modelo de curtíssimo prazo - DESSEM. O objetivo do DESSEM é determinar o despacho de geração

Figura 2.3 – Regiões do Sistema Interligado Nacional.



Fonte: (CCEE, 2021f), (NASCIMENTO, 2017).

das usinas hidrelétricas e termelétricas que minimiza o custo de operação ao longo do período de planejamento, bem como o custo marginal de operação para cada período e por submercado, dado o conjunto mais detalhado das informações (previsões de carga, vazões, geração eólica, disponibilidades, limites de transmissão entre subsistemas, função de custo futuro do DECOMP, etc).

Vale ressaltar que até 31 de dezembro de 2020, os valores oficiais do PLD eram calculados com base nos resultados da execução do modelo DECOMP. A partir de 1º de janeiro de 2021, os valores oficiais do PLD são calculados com base nos resultados da execução do modelo DESSEM.

2.3 Predição do PLD

O Preço de Liquidação das Diferenças é uma variável chave do setor elétrico brasileiro. O PLD propicia flexibilidade nas negociações, possibilita ajustes entre a energia contratada e a energia gerada, e é adotado como referência para contratos de longo prazo. (LEITE, 2013)

As incertezas associadas ao planejamento de um sistema com predominância da geração hidrelétrica conferem ao PLD acentuada volatilidade. Esta volatilidade é uma das maiores preocupações dos agentes do setor elétrico, fundamentalmente no ACL. Por conseguinte, a predição dessa variável tem se mostrado um dos grandes desafios enfrentados por pesquisadores nos últimos anos.

Reston Filho ([FILHO, 2014](#)) desenvolveu em seu trabalho um modelo híbrido de predição do preço da energia no mercado de curto prazo. Sua abordagem combina modelos de Redes Neurais Artificiais (RNA) e filtros autorregressivos integrados de médias móveis (ARIMA), aplicando um processo em dois passos: Primeiramente, o autor realiza a predição de variáveis explanatórias e então, os valores futuros destas variáveis são utilizados para a predição dos preços de energia. O modelo desenvolvido é capaz de estimar o PLD até 12 passos (semanas) a frente, aproximadamente 3 meses. Os resultados são bons para os doze passos à frente, mostrando resultados de validação coerentes com os preços futuros reais.

Rodrigues ([RODRIGUES, 2009](#)) propôs uma RNA com dados de entrada disponibilizados pelo ONS e pela CCEE, a saber: preços do mercado spot, armazenamento de água, energia natural afluyente, energia máxima, geração em Itaipu, hidrelétricas (exceto Itaipu), térmica convencional, térmica nuclear, fonte eólica, carga do sistema, intercâmbios (transmissão de energia entre os submercados) e internacional (importação ou exportação de energia para Argentina, Paraguai ou Venezuela). Foram utilizados dados de 6 semanas anteriores ao dia t , tendo como saída da rede o preço previsto para o dia t . Os resultados foram apresentados para os quatro submercados brasileiros.

No geral, as RNAs apresentaram resultados razoáveis frente à heterogeneidade dos dados de entrada. Testes de sensibilidade foram realizados a fim de avaliar o comportamento da rede sem as variáveis carga e energia armazenada e constatou-se resultados com elevada imprecisão, demonstrando a dependência forte de tais entradas. O trabalho de ([COELHO, 2018](#)) também faz uso das Redes Neurais Artificiais de forma a realizar predições do Preço de Liquidação das Diferenças. A autora fez uso de uma Rede Neural Recorrente em 5 cenários distintos, realizando alterações nos dados de entrada: dados originais, dados com remoção de *outliers*, correção monetária dos dados com base no IGP-M, utilização de dados a partir do ano de 2015 (após nova regulamentação de limites máximo e mínimo) e uma combinação das anteriores. As previsões apresentaram resultados bastante distintos para cada cenário gerado. As predições foram feitas para 20 meses e 5 anos a frente, não apresentando resultados com alta confiabilidade.

Redes Neurais Artificiais são modelos computacionais propícios para o aprendizado de padrões e conseqüente predição de dados e comportamentos. São os modelos mais empregados na literatura e, por este motivo, também serão motivo de investigação neste trabalho. Em decorrência disto, faz-se a seguir uma revisão literária à respeito das Redes Neurais, com suas principais características e aplicações.

2.4 Redes Neurais Artificiais

O cérebro é o órgão mais complexo do corpo humano onde as bilhões de células nervosas possuem interconexões. Algumas destas células, os neurônios, possuem ramificações

ou conexões que se estendem desde o cérebro até as partes mais distantes do corpo. O cérebro é especialmente bom para desempenhar funções como reconhecimento de padrões, controle motor, percepção, tomada de decisões, comunicação e predição.

O primeiro registro do modelamento matemático inspirado em um neurônio biológico data de 1943, onde o neurofisiologista Warren McCulloch e o matemático Walter Pitts desenvolveram um trabalho intitulado “*A logical calculus of the ideas immanent in nervous activity*” (MCCULLOCH; PITTS, 1943) que fazia uma analogia entre células vivas e o processo eletrônico, simulando o comportamento do neurônio natural. O modelo consistia basicamente em um neurônio que executava uma função lógica, onde os nós produziam somente resultados binários e as conexões transmitiam exclusivamente zeros e uns.

Em 1949, o psicólogo Donald Hebb, demonstrou que a capacidade da aprendizagem em Redes Neurais biológicas vem da alteração da eficiência sináptica, isto é, a conexão somente é reforçada se tanto as células pré-sinápticas quanto as pós-sinápticas estiverem excitadas. Hebb foi o primeiro a propor uma lei de aprendizagem específica para as sinapses dos neurônios, conhecida como a Lei de Hebb (HEBB, 1949).

O trabalho de desenvolvimento de modelos matemáticos inspirados no neurônio biológico foi continuado por diversos pesquisadores, dando existência a diversos algoritmos de aprendizado e estruturas neuronais. Dentre as linhas de pesquisa que surgiram, destaca-se o modelo *Perceptron*, apresentado por Frank Rosenblatt em seu livro “*Principles of neurodynamics*” (ROSENBLATT, 1961). Nele, os neurônios eram organizados em camadas de entrada e saída, onde os pesos das conexões eram adaptados a fim de se atingir a eficiência sináptica usada no reconhecimento de caracteres.

O *Perceptron* atraiu interesses devido à sua capacidade de reconhecer padrões simples e em 1960 surgiu a rede *Adaline* (*ADaptive LINear Element*) e a *Madaline* (*Many Adaline*), propostas por Widrow & Hoff (WIDROW, 1962). São redes com aprendizado fundamentado na regra Delta, conhecida como algoritmo de aprendizado LMS (*Least Mean Square*). Em 1969, foi constatado por Minsky & Papert (MINSKY; PAPERT, 1969) que um neurônio do tipo *Perceptron* só é capaz de resolver problemas com dados de classes linearmente separáveis. Esta limitação do uso das Redes Neurais deu início ao primeiro grande inverno das Redes Neurais, uma era de poucas novidades, com cortes nos investimentos e baixa atenção ao setor.

Em 1982 houve uma retomada das pesquisas com a publicação dos trabalhos do físico biólogo John Hopfield relatando a utilização de redes recorrentes baseadas em funções de energia (HOPFIELD, 1982). No final dos anos 1980, com o impulso inicial do trabalho de Hopfield e outros pesquisadores, um interesse significativo ressurgiu na área. A retomada definitiva se deve à criação de algoritmos mais robustos e eficientes, ao desenvolvimento de computadores com maior capacidade de processamento e a novas descobertas sobre o sistema nervoso biológico (SILVA, 2010).

Um dos destaques do período, foi a publicação do livro de David E. Rumelhart, Geoffrey E. Hinton e Ronald J. Williams, em 1986 (RUMELHART; WILLIAMS, 1986). Eles introduziram o poderoso método de treinamento chamado *Backpropagation*, que permitia o ajuste de pesos em uma rede multicamadas. A partir desta proposição, foi despertado o interesse em estudos sobre RNAs novamente e de forma definitiva. Deste ponto, até os dias atuais as Redes Neurais Artificiais vêm sendo aplicadas em diversas áreas do conhecimento humano.

Para o número de camadas de uma rede neural, tem-se o Teorema da Aproximação Universal das Redes Neurais Artificiais, derivado do Teorema de Andrey Kolmogorov (1957) e complementado por George Cybenko (1989). Este teorema estabelece que apenas uma única camada escondida com um número finito de neurônios é capaz de aprender e aproximar qualquer função (SILVA, 2010). Portanto, com apenas uma camada é possível reproduzir qualquer sequência de dados por meio de uma rede neural artificial, sendo um importante ponto de partida para o problema em questão apresentado neste trabalho.

No entanto, o problema proposto pode se revelar mais complexo que o esperado. Assim, mais camadas são passíveis de serem utilizadas caso o problema requeira mais que uma mera aproximação de funções. É possível que a captura das nuances do problema necessite de conexões espaciais de ordens maiores, requerendo, portanto, mais camadas escondidas, da mesma maneira que se faz no reconhecimento de imagens por meio de redes profundas (*deep learning*).

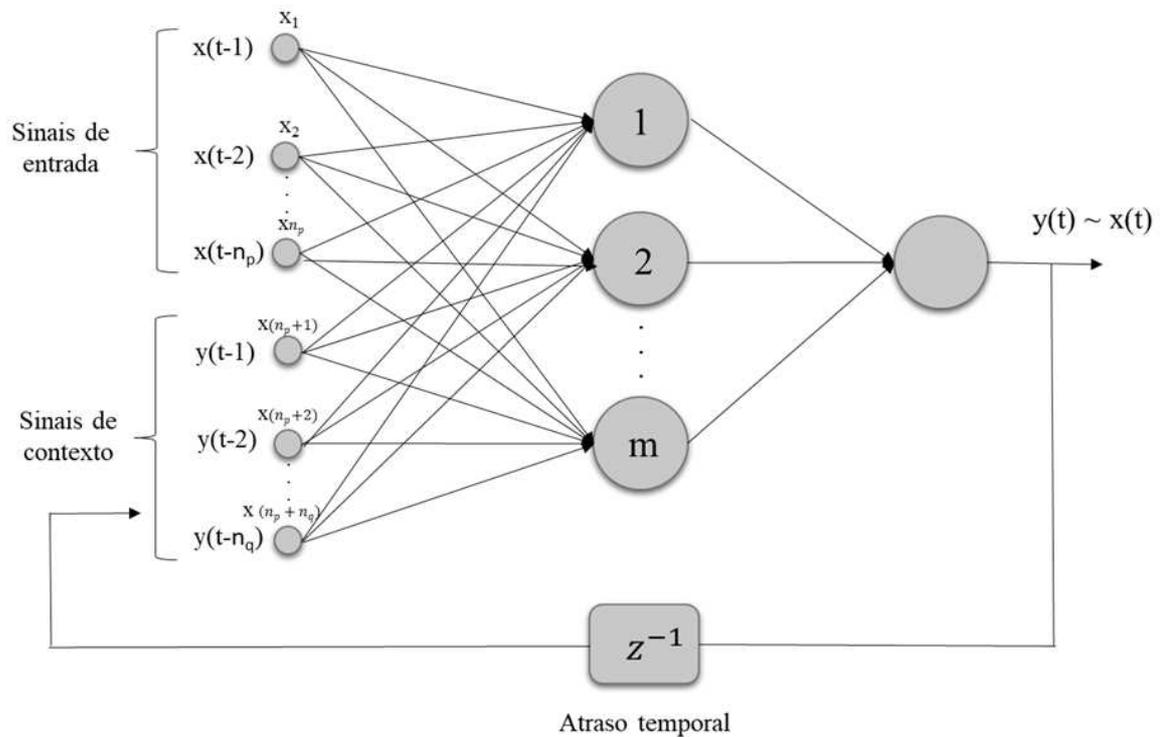
2.4.1 Redes Recorrentes

A capacidade de generalização de uma RNA está intimamente relacionada com a escolha da arquitetura de rede e com o algoritmo de aprendizado utilizado na fase de treinamento da rede. As arquiteturas podem ser divididas em duas principais categorias: redes alimentadas adiante e redes recorrentes (ou redes realimentadas). As redes recorrentes, em especial, são adequadas para o processamento dinâmico de informações, ou seja, em sistemas cujos comportamentos são considerados variantes no tempo ou dependentes dele. Como exemplo de aplicação tem-se a previsão de séries temporais, em que há correlação entre dados subsequentes no tempo (DUDA; HART; STORK, 2000).

Conforme evidenciado na Figura 2.4, essas redes possuem neurônios adicionais atrasados que transmitem informações de camadas posteriores para anteriores ou entre neurônios da mesma camada, criando uma autorregressão que atua como memória interna. Esta autorregressão é conhecida como *Time Delay Neural Network (TDNN)* (SILVA, 2010).

A arquitetura de rede com saídas recorrentes às entradas permite a partir da realimentação de sinais produzidos em instantes anteriores, a recuperação de respostas passadas. Por conseguinte, é possível dizer que esta tipologia possui memória, ou seja, é capaz de lembrar saídas passadas de forma a produzir a resposta atual ou futura. Sendo

Figura 2.4 – Topologia de rede recorrente.



Fonte: (NASCIMENTO, 2017).

assim, a predição dos valores futuros, a partir do instante t , também terá influência dos valores anteriores produzidos pela saída da rede:

$$x(t) = f(x(t-1), x(t-2), \dots, x(t-n_p), y(t-1), y(t-2), \dots, y(t-n_q)) \quad (2.1)$$

em que n_p é a ordem do preditor e indica o número de amostras anteriores necessárias para a estimação de $x(t)$ e o valor n_q expressa a quantidade de saídas passadas que serão utilizadas na estimação de $x(t)$.

Conforme evidenciado por (SILVA, 2010), a configuração de realimentação confere à rede a capacidade de realizar, de maneira implícita, o mapeamento entre entradas e saídas de processos tanto não-lineares como variantes no tempo, tornando-se uma ferramenta bem flexível para aplicações envolvendo identificação de sistemas.

2.4.2 Métodos de Treinamento

As Redes Neurais Artificiais se destacam pela capacidade de aprendizado a partir da apresentação de amostras que representam o comportamento do sistema. Após o aprendizado da relação das entradas e saídas, a rede se torna capaz de produzir uma saída próxima da desejada a partir de quaisquer dados fornecidos às entradas. Portanto, o algoritmo de aprendizado da rede consiste na aplicação de um conjunto de passos com o objetivo de fazer o ajuste dos pesos sinápticos dos neurônios, tendo em vista sintonizar a rede para que as respostas produzidas por ela estejam próximas dos valores desejados.

Há, na literatura, diversos algoritmos usados para o treinamento das redes. Basicamente, eles diferem entre si pelo modo como os pesos sinápticos de um neurônio são modificados. Cada algoritmo apresenta vantagens e desvantagens quanto ao ajuste do peso, a escolha está relacionada com a categoria de problema a ser solucionado.

A lei de aprendizado de Hebb, desenvolvida em 1949, permite o ajuste de pesos e limiar do *Perceptron* visando a classificação de padrões entre duas classes. Resumidamente, se a saída da rede coincide com o esperado, os pesos e limiares são incrementados de forma proporcional ao sinal de entrada, caso a saída da rede seja diferente do valor esperado os pesos sinápticos e limiar são decrementados. O processo descrito se repete para todas as amostras do treinamento até que a saída da rede seja próxima à saída desejada de cada amostra.

Outro método de treinamento popularmente conhecido, foi desenvolvido por Widrow & Hoff, conhecido por regra Delta, algoritmo *LMS* (*Least Mean Square*) ou método do Gradiente Descendente. A ideia desenvolvida com este algoritmo consiste em minimizar a diferença entre a saída desejada e a resposta do combinador linear (composição das entradas ponderadas pelos respectivos pesos adicionada do limiar de ativação). Objetiva-se um vetor de pesos ótimo, tal que o erro quadrático sobre o conjunto de amostras seja o menor possível.

Para problemas que não são linearmente separáveis, é necessária a utilização de uma rede de múltiplas camadas. O algoritmo *Backpropagation* ou algoritmo de retropropagação do erro é bastante conhecido e aplicado no decorrer do processo de treinamento deste tipo de RNA. A primeira fase deste algoritmo é conhecida como *forward*, onde as entradas de uma amostra do conjunto de treinamento são inseridas na rede e seu efeito se propaga através dela desde a entrada, passando camada por camada, até a produção da saída. Logo em seguida, esta saída é comparada aos resultados desejados, e os respectivos sinais de erros são calculados. Sequencialmente na fase *backward*, os pesos das conexões são atualizados, baseados nos sinais de erro. A transmissão dos sinais de erro é feita da camada de saída em direção à camada de entrada. São realizadas sucessivas aplicações das duas fases de forma que os pesos se ajustem com o passar das iterações até que um critério de parada pré estabelecido seja atingido.

Desde seu lançamento, diversas variações do *Backpropagation* foram intensamente desenvolvidas. Tais variações buscam tornar o processo de convergência mais rápido e eficiente. Dentre elas, podem-se citar a inserção do *momentum*, o *resilient-backpropagation*, o *ADAM* (*Adaptive Moment Estimation*) e o método *quasi-Newton* Levenberg-Marquardt.

Uma das variações mais simples do *Backpropagation* é a inserção do termo de *Momentum*, em consonância com o conceito físico. Este parâmetro, que varia entre zero e um, pondera o quanto houve de variação entre duas iterações anteriores e sucessivas das matrizes de peso. A aceleração no processo de convergência acontece pois é levada em

consideração o quão afastada da solução final está a solução atual. Basicamente, quando a solução atual estiver longe da solução final, a variação na direção oposta ao gradiente da função erro quadrático entre duas iterações sucessivas também será grande. Neste cenário adiciona-se um incremento mais significativo na matriz de pesos. Entretanto, quando as variações nas matrizes de peso forem pequenas, a variação do erro quadrático entre duas iterações sucessivas também será, o que significa que a solução está próxima da solução atual. Neste cenário, tem-se uma pequena contribuição do termo *Momentum*.

No *Resilient-Backpropagation* (RIEDMILLER; BRAUN, 1993) apenas o sinal da derivada do gradiente da função de erro é levado em consideração para indicar a direção da atualização dos pesos. Desta forma, é possível eliminar a influência prejudicial do tamanho da derivada parcial na etapa de peso. Assim sendo, ao considerar duas iterações sucessivas, se os sinais do gradiente forem os mesmos, a taxa de aprendizado pode ser incrementada em razão de estar distante de um ponto de mínimo da função de erro. Em contrapartida, deve-se diminuir a taxa de aprendizagem, quando os sinais do gradiente forem diferentes, pois isso implica que o ponto de mínimo da função foi ultrapassado.

O *ADAM* é um algoritmo de treinamento de Redes Neurais Artificiais baseado na utilização dos dois primeiros momentos estatísticos, respectivamente a esperança e a variância, de maneira que os dados preditos converjam para uma média otimizada em termos de espalhamento estatístico, sendo um dos métodos mais indicados para dados estocásticos (P.KINGMA; BA, 2014).

O algoritmo de *Levenberg-Marquardt* (HAGAN; MENHAJ, 1994) é um método gradiente de segunda ordem do tipo *quasi-Newton*. Diferentemente do *Backpropagation*, que é um método de primeira ordem, de descida do gradiente, o algoritmo de *Levenberg-Marquardt* é uma aproximação do método de Newton. Nele, utiliza-se a 2^o derivada (Hessiana) aproximada, esta aproximação é feita a partir do Jacobiano.

De acordo com (DAMETTO, 2018), o método de otimização *ADAM* proposto por (P.KINGMA; BA, 2014) é utilizado na otimização de funções estocásticas, sendo apropriado para problemas com muito ruído. Trata-se de um método adaptativo de primeira ordem, baseado na derivada da função objetivo, ou seja, o gradiente. Ele trabalha com o conceito de Momentos estatísticos para construir a adaptação e atualização dos pesos sinápticos da Rede Neural.

O método *ADAM* é baseado no primeiro e segundo momentos, em que o primeiro momento é dado diretamente pelo gradiente, e o segundo momento pelo quadrado do gradiente. O primeiro momento faz a convergência imediata do peso sináptico, enquanto o segundo momento, que é a medida de espalhamento (variância), acelera e atenua a oscilação da descida do gradiente.

O treinamento do *ADAM* é baseado no *Backpropagation*, configurando uma extensão

do mesmo com a utilização do conceito de momentos. Ele opera basicamente em uma sequência de dois passos. O primeiro passo consiste na propagação do sinal da entrada à saída. As saídas calculadas, tendo como base os pesos sinápticos da rede, são comparadas com os resultados desejados por meio de uma equação de erro quadrático. Caso a amplitude do erro seja alta, significa que as saídas da rede são muito diferentes dos resultados desejados, e então se faz necessário o ajuste dos pesos sinápticos. Então, o sinal de erro é retropropagado modificando os pesos das conexões das camadas internas.

O ajuste dos pesos é feito enquanto o erro de comparação, no caso o erro quadrático médio, for menor que um limiar pré-estabelecido e segue o seguinte encadeamento: na inicialização do algoritmo, as taxas de aprendizado do primeiro e segundo momento β_1 e β_2 são configuradas. As variáveis do primeiro e segundo momentos, respectivamente m_0 e v_0 , são inicializados como nulos na primeira rodada. Semelhantemente ao *Backpropagation*, o *ADAM* faz uma manipulação do gradiente, que é um vetor que aponta na direção de máximo crescimento da função. Logo, é possível obter a região de mínimo crescimento através da direção oposta do gradiente e assim garantir que os pesos obtidos ajustem o erro em seu menor valor.

Com base no cálculo do gradiente g_t e a passagem das iterações t , é possível obter o primeiro e segundo momentos, conforme as Equações 2.2 e 2.3:

$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (2.2)$$

$$v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \quad (2.3)$$

Em seguida, ajustam-se os momentos, de forma a evitar divisões por zero (por meio de um valor bastante pequeno ϵ) bem como reescalar o tamanho do ajuste, conforme as Equações 2.4 e 2.5:

$$\hat{m}_t \leftarrow \frac{m_t}{(1 - \beta_1^t)} \quad (2.4)$$

$$\hat{v}_t \leftarrow \frac{v_t}{(1 - \beta_2^t)} \quad (2.5)$$

Por fim, os parâmetros da rede são calculados e atualizados conforme 2.6:

$$\Theta_t \leftarrow \Theta_{t-1} - \alpha \frac{\hat{m}_t}{(\sqrt{\hat{v}_t} + \epsilon)} \quad (2.6)$$

Após o ajuste, as saídas da rede são calculadas com base nos novos valores de peso. A saída gerada é comparada com a saída desejada e o erro obtido é comparado com um limiar pré-estabelecido. Se porventura, o erro calculado for menor que o limiar, tem-se que a resposta da rede é satisfatória e generaliza bem os dados que se deseja aprender e o algoritmo é encerrado. No entanto, caso a resposta seja negativa é preciso fazer o ajuste

dos pesos sinápticos novamente e o processo se repete até que a condição de erro mínimo (limiar) seja alcançada.

2.5 Filtragem

O termo filtro é usado frequentemente para descrever um objeto físico ou programa de computador utilizado em um conjunto de dados ruidosos com o objetivo de extrair determinadas informações responsáveis pela degradação dos dados (HAYKIN, 1996).

Os filtros podem ser classificados em duas categorias principais que se diferem tanto no funcionamento quanto na montagem física: os filtros analógicos e os digitais. Os filtros analógicos utilizam circuitos eletrônicos analógicos compostos por uma combinação de resistores, capacitores e indutores para produzir o efeito de filtragem desejado. Já os filtros digitais utilizam processamento digital para a realização de cálculos numéricos nas amostras do sinal de entrada. Eles são representados por equações discretas que ao serem aplicadas em um sinal digital de entrada $x(n)$, produzem uma saída $y(n)$ com frequências dentro de uma faixa selecionada.

Os filtros digitais são programáveis, ou seja, tem sua operação determinada por um programa armazenado na memória do processador. Diferentemente dos filtros analógicos que para serem alterados é preciso realizar alterações no *hardware* (circuito eletrônico), os filtros digitais podem facilmente ser alterados. Além disso, os circuitos de filtros analógicos são compostos por componentes ativos, sujeitos a variações de alguns parâmetros como: temperatura e valores devido à construção dos componentes, o que os tornam menos estáveis que os filtros digitais, que não estão sujeitos à essas variações.

2.5.1 Filtros Digitais

Classificam-se os filtros digitais a partir do comprimento da sua sequência de resposta ao impulso como Filtro de Resposta ao Impulso Finita (FIR - *Finite Impulse Response*) e Filtro de Resposta ao Impulso Infinita (IIR - *Infinite Impulse Response*). Em FIR a resposta produzida é sem distorção na fase do sinal e é sempre estável pois não são recursivos, e nos filtros IIR a resposta produzida é defasada em relação ao sinal filtrado.

Nos filtros do tipo FIR, a resposta ao impulso possui duração finita e o cálculo dos valores de saída presentes não sofrem influência dos valores de saída passados, ou seja, é um filtro de natureza não recursiva. A saída de um filtro FIR digital genérico é dado por:

$$y(n) = h_0x(n) + h_1x(n - 1) + \dots + h_Px(n - P) \quad (2.7)$$

em que P é a ordem do filtro, $x(n)$ é o sinal de entrada, $y(n)$ o sinal de saída e h_i os coeficientes do filtro. A equação apresentada anteriormente também pode ser escrita na

forma apresentada a seguir:

$$y(n) = \sum_{i=0}^{P-1} h_i x(n-i) \quad (2.8)$$

Já nos filtros digitais do tipo IIR, a resposta ao impulso possui uma duração infinita e a saída adquirida em um determinado instante depende da entrada atual e das saídas oriundas de instantes anteriores, sendo classificado como um filtro de natureza recursiva. A representação da saída é dada por:

$$y(n) = a_0 x(n) + a_1 x(n-1) + a_2 x(n-2) + \dots - [b_1 y(n-1) + b_2 y(n-2) + \dots] \quad (2.9)$$

em que $x(n)$ é o sinal de entrada, $y(n)$ o sinal de saída e a_i e b_i são os coeficientes de recursão do filtro. A expressão anterior também pode ser representada conforme a seguir:

$$y(n) = \sum_{i=0}^N a_i x(n-i) - \sum_{k=1}^M b_k y(n-k) \quad (2.10)$$

A escolha do tipo de filtro a ser utilizado deve ir de encontro ao que produz um melhor resultado para a solução desejada. Dentre os filtros digitais destacam-se os filtros de frequência seletivas (filtros de *Butterworth*, *Chebyshev*, Elíptico, dentre outros) e os filtros ótimos (Wiener e Kalman).

Os filtros seletivos em frequência, podem ser: passa-baixa, passa-alta, passa-faixa e rejeita-faixa. Um filtro passa-baixa tem a característica de transmitir as baixas frequências de um sinal (banda de passagem) e atenuar suas altas frequências (banda de rejeição). Em contrapartida, os filtros passa-alta permitem a passagem de sinais de alta frequência e reduz a intensidade de sinais de baixa frequência. De forma semelhante, tem-se o funcionamento dos filtros passa-faixa e rejeita-faixa, onde é permitido que uma faixa entre frequências de referências passe livremente ou sejam atenuadas, respectivamente.

2.5.2 Filtro de Média Móvel

O filtro média móvel (MAF - *Moving Average Filter*) é um filtro passa-baixa que possui resposta impulsional finita (FIR). Este tipo de filtro é elementar em processamento digital de sinais, principalmente por ser o filtro digital mais fácil para entender e utilizar. Apesar de sua simplicidade, o filtro média móvel é interessante para reduzir o ruído aleatório pois mantém uma resposta de degrau nítida. Isso o torna o filtro principal para sinais codificados no domínio do tempo (SMITH, 1999).

Como o próprio nome sugere, o filtro média móvel opera realizando o cálculo da média de um número de pontos do sinal de entrada de forma a produzir os pontos no sinal de saída. Esta média vai se alterando, entre os valores do sinal em questão. Na forma de equação, tem-se:

$$y[i] = \frac{1}{M} \sum_{j=0}^{M-1} x[i+j] \quad (2.11)$$

em que x é o sinal de entrada, y é o sinal de saída e M é o número de pontos usados na média móvel. Por exemplo, em um filtro média móvel com janela de cinco intervalos, o sinal de saída do décimo dado é escrito por:

$$y[10] = \frac{x[10] + x[9] + x[8] + x[7] + x[6]}{5} \quad (2.12)$$

Como alternativa, o grupo de pontos do sinal de entrada pode ser escolhido simetricamente em torno do ponto de saída:

$$y[10] = \frac{x[8] + x[9] + x[10] + x[11] + x[12]}{5} \quad (2.13)$$

As equações 2.12 e 2.13 exemplificam o filtro média móvel simples, um dos mais comuns. Além deste, os filtros de média móvel cumulativa e média móvel ponderada são comumente utilizados.

A teoria dos filtros ótimos é desenvolvida com base em um sinal de erro. Este sinal de erro é definido como a diferença entre o sinal desejado e o sinal estimado, a saída do filtro. Vários modelos foram desenvolvidos visando uma otimização do sinal de erro, dentre esses métodos pode-se destacar o método dos mínimos quadrados. Os filtros de Wiener e Kalman são baseados na teoria dos mínimos quadrados, que consiste em minimizar a soma dos erros quadráticos médios em termos estocásticos.

2.5.3 Filtro de Kalman

O filtro de Kalman, foi proposto em 1960 por Rudolf E. Kalman (KALMAN, 1960). O filtro de Wiener era largamente utilizado em várias pesquisas. Contudo, este filtro apresenta limitações em relação aos dados de entrada, que precisam ser estacionários. Sendo assim, o filtro de Wiener não era capaz de solucionar problemas dinâmicos.

No filtro de Kalman, o comportamento de um sistema dinâmico é descrito por um conjunto de variáveis de estados. As variáveis de estados contêm todas as informações necessárias sobre o comportamento do sistema de forma que dados os valores presentes e passados da entrada, pode-se determinar os estados futuros do sistema (KAILATH, 1980).

O advento das comunicações digitais, principalmente o surgimento e evolução dos computadores digitais, provocou uma grande utilização do algoritmo de Kalman, notadamente na área científica. Dentre os fatores que favoreceram a popularidade do filtro de Kalman, pode-se destacar:

- Utilização de estruturas recursivas;
- Manipulação da álgebra matricial simples, que possibilita sua implementação em computador digital;

- Ampla aplicabilidade em várias áreas da engenharia (onde a estimação de estados é encontrada com frequência);
- Adaptabilidade a aplicações não-lineares.

Como mencionado anteriormente, o filtro de Kalman visa estimar os estados de um sistema dinâmico. Ou seja, a partir das medidas do sistema, seu modelo e as estatísticas dos ruídos que atuam tanto nas medidas quanto no modelo do sistema, determina-se uma estimativa dos valores dos estados do sistema no decorrer do tempo.

O Filtro de Kalman é frequentemente implementado no formato de propagação-atualização (projeção-correção). As equações de propagação são responsáveis por projetar a estimativa do estado corrente e da matriz de covariância, já as equações de atualização são responsáveis pela realimentação, ou seja, incorporam uma nova medida para obter uma estimativa atualizada (melhorada). Resumidamente, a etapa de projeção é realizada inicialmente baseando-se na estimativa atual do estado. Em seguida, após receber uma nova medição, a etapa de correção é realizada.

Para a implementação do filtro, a princípio é feita a projeção (a priori) da estimativa do estado \hat{x} , conforme Equação 2.14, e da covariância do erro (matriz P), conforme Equação 2.15. Nesta etapa o vetor estimado de estado \hat{x}_k é atualizado em função do estado anterior do tempo $k - 1$. Também é feita a projeção (a priori) do erro de covariância, que é uma pré-atualização da matriz P envolvendo a soma da matriz de estados do sistema com a matriz Q que é a matriz de covariância do ruído do sistema. A matriz Q é traçada pelo projetista de acordo com o conhecimento do processo.

$$\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_{k-1} \quad (2.14)$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \quad (2.15)$$

Uma vez calculadas as projeções (ou, também conhecida como etapa de predição), é necessário corrigir o valor de \hat{x} segundo a comparação com leitura de um sensor real y relativo aos estados que se deseja medir, de modo que $\hat{x} \rightarrow x$. Além disso, a covariância P também deve ser corrigida, de tal maneira que $P \rightarrow 0$. Para tanto, há o cálculo do ganho ótimo K , que tem o papel de fazer a correção, tanto de \hat{x} quanto de P , denominado por ganho de Kalman. Esta é, portanto, a fase de correção:

$$K_k = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1} \quad (2.16)$$

em que R é a covariância do ruído de medição do sensor y .

Atualização (ou, correção) da estimativa de estados:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(y_k - C\hat{x}_{k|k-1}) \quad (2.17)$$

Atualização (a posteriori) da matriz de covariância:

$$P_{k|k} = (I - K_k C) P_{k|k-1} \quad (2.18)$$

No filtro de Kalman, o sistema x e o sensor y são contaminados por ruídos, e o ganho de Kalman conduz uma estimativa \hat{x} , que é um modelo matemático dado pela esperança de x , diretamente para onde os estados x estão no momento. Como consequência, o espalhamento (covariância) entre x e \hat{x} , dado pela matriz P , tende a zero, uma vez que $\hat{x} \rightarrow x$.

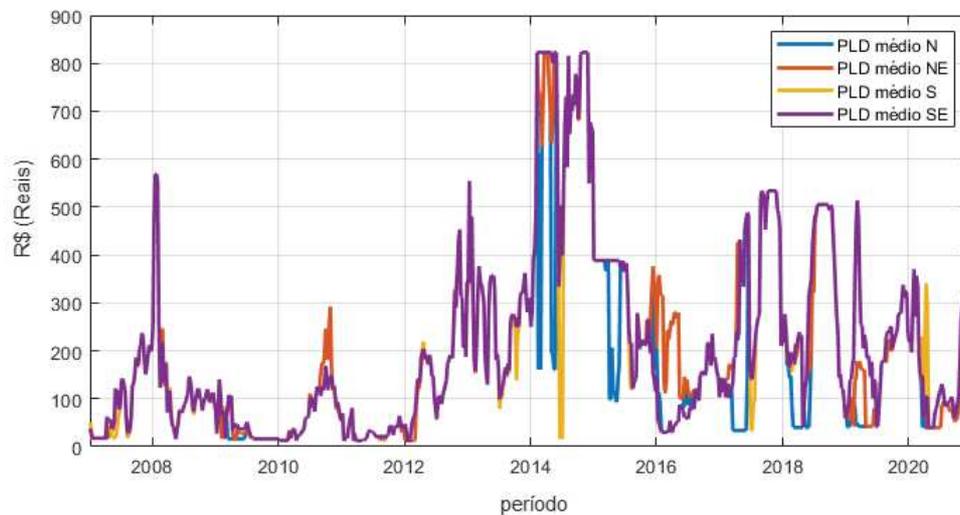
Resumidamente, quando se faz necessária a estimativa de valores e a mensuração destes valores contém incertezas, erros ou variações aleatórias, o filtro de Kalman se mostra como uma ferramenta de extrema importância. Pois, através de um processo matemático iterativo que utiliza um conjunto de equações, juntamente com dados de entrada recebidos pelo sistema, é possível estimar a grandeza desejada.

Neste trabalho, os filtros descritos (passa-baixa FIR, IIR, média móvel e Kalman) serão utilizados para reduzir os ruídos produzidos pela Rede Neural nos dados preditos por sua saída. Como a rede é recorrente, os ruídos são realimentados na rede e se acumulam, fazendo com que a rede divirja com o passar do tempo. Como a proposta é preservar as tendências dos dados ao invés de uma predição perfeita dos valores do CMO/PLD, a filtragem ajuda a eliminar principalmente os ruídos de alta frequência que venham a ser produzidos pela rede. Os filtros serão testados de modo a revelar qual tem melhores condições de ajudar a preservação da tendência de preços ao longo do ano.

3 METODOLOGIA PROPOSTA

Conforme apresentado no Capítulo 2, as Redes Neurais Artificiais têm sido amplamente utilizadas para a predição dos preços no mercado de curto prazo. Neste sentido, o presente trabalho visa verificar um modelo de predição do PLD para o submercado Sudeste/Centro Oeste. Qualquer submercado poderia ter sido escolhido, uma vez que seus comportamentos são semelhantes, conforme pode-se verificar na Figura 3.1. Este capítulo apresenta a metodologia proposta, incluindo a escolha dos dados de entrada, do método de treinamento da RNA, e o processamento dos dados de entrada e de saída.

Figura 3.1 – PLD médio de todos os submercados.



Fonte: (CCEE, 2021b).

3.1 Escolha dos Dados de Entrada

Como mencionado no Capítulo 1, o valor do PLD e do CMO são fortemente semelhantes, sendo a diferença entre eles dada pela saturação imposta pelos preços mínimo e máximo estipulados anualmente pela ANEEL para o PLD. Tendo em vista que essas restrições ao PLD provocam perda de informação dos dados, optou-se por arquitetar uma Rede Neural que predissesse o valor do CMO, para então poder restringi-lo com os limites máximo e mínimo, obtendo-se, assim, uma boa estimativa para a tendência de comportamento do PLD.

Uma primeira abordagem consiste em alimentar as entradas da rede recorrente com os dados do CMO e avaliar se, por não serem saturados, estes dados auxiliam na predição de várias semanas a frente. A seguir, serão investigadas outras variáveis que compõem o cálculo do PLD. A escolha destas variáveis será feita a partir da correlação entre o CMO e

todas as variáveis selecionadas. As variáveis que apresentarem maiores correlações irão compor a entrada da RNA, além das próprias realimentações desta rede.

Desta forma, foram utilizadas como entradas da RNA combinações das variáveis divulgadas no histórico do ONS (ONS, 2020) e CCEE (CCEE, 2021b), sendo elas, Carga de Energia (CE), Demanda Máxima (DM), Energia Armazenada (EA), Energia Natural Afluente (ENA), Geração de Energia (GE), Intercambio de Energia (IE), Preço de Liquidação das Diferenças (PLD), e Custo Marginal de Operação (CMO).

A configuração da entrada da RNA, envolvendo as variáveis anteriores, é apresentada na equação (3.1), com discretização semanal, onde t é o índice da semana relativa à saída da rede (semana que se quer prever); n_p é o número de atrasos temporais das entradas, a ser escolhido por meio da análise de correlação entre o CMO e as outras variáveis da ONS; y é a saída da Rede Neural (que pode ser filtrada ou não); e n_q é o número de atrasos TDNN *Time Delay Neural Network* da saída, em que $n_q \leq n_p$.

$$\text{Entradas RNA} = \begin{bmatrix} CMO(t-1) \\ \vdots \\ CMO(t-n_p) \\ ENA(t-1) \\ \vdots \\ ENA(t-n_p) \\ DM(t-1) \\ \vdots \\ DM(t-n_p) \\ \vdots \\ \text{Variáveis ONS} \\ \vdots \\ y(t-1) \\ \vdots \\ y(t-n_q) \end{bmatrix} \quad (3.1)$$

Por se tratarem de dados estocásticos, contaminados por ruídos aleatórios, propõe-se uma etapa de filtragem na saída da RNA para alisamento dos dados, objetivando uma melhor qualidade da predição devido à redução da influência de ruídos e à redução da volatilidade dos dados de entrada da Rede Neural.

O processo de filtragem ocorre mediante transformações matemáticas nas séries temporais, possibilitando uma limpeza nos componentes de altas frequências da série. Optou-se pela implementação e análise dos filtros Passa-Baixa (FIR e IIR), Média Móvel e Filtro de Kalman com a perspectiva de redução de ruídos e consequente prolongamento

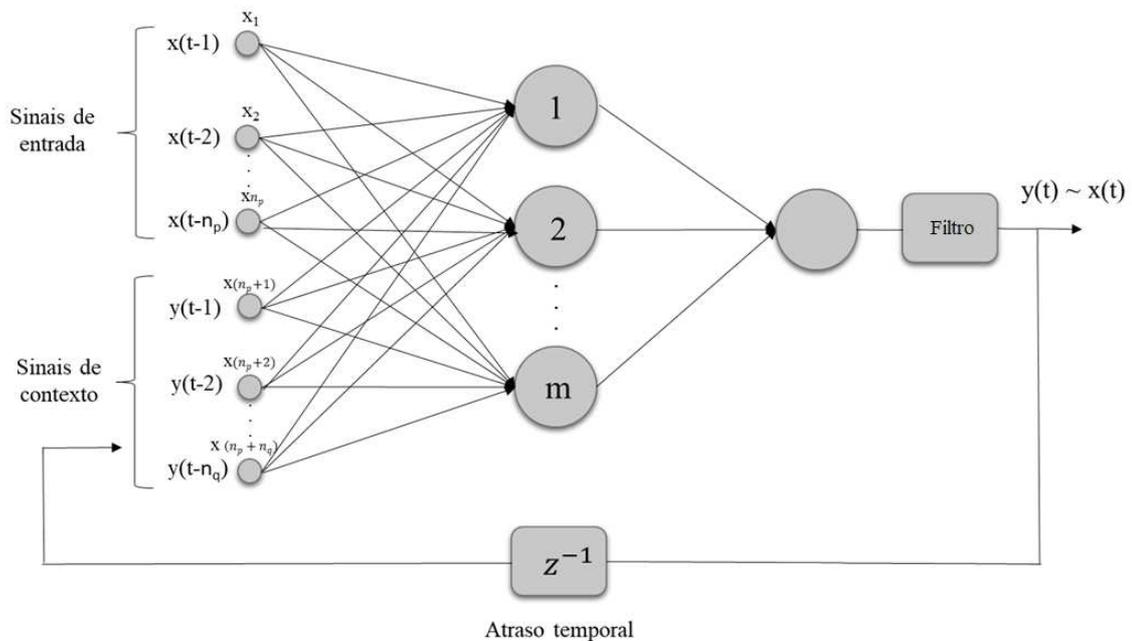
da qualidade de tendências de predição. A Figura 3.2 apresenta a arquitetura de rede proposta com a inserção da etapa de filtragem antes da realimentação dos dados.

3.2 Escolha do Método de Treinamento

Conforme citado no subitem 2.4.2, existem vários métodos de treinamento de uma Rede Neural Artificial, e cada um deles promove um efeito diferente no treinamento e posterior predição da RNA. No sentido de avançar no entendimento dos métodos de treinamento mais comumente utilizados e então escolher um bom otimizador para a RNA proposta, foi feita a predição de uma série temporal previamente conhecida, contaminada por um ruído Gaussiano, utilizando cinco opções de otimizadores, a saber:

- i *Backpropagation* original;
- ii *Backpropagation* com *Momentum*;
- iii *Resilient Backpropagation*;
- iv *Levenberg-Marquardt*, e;
- v *ADAM*.

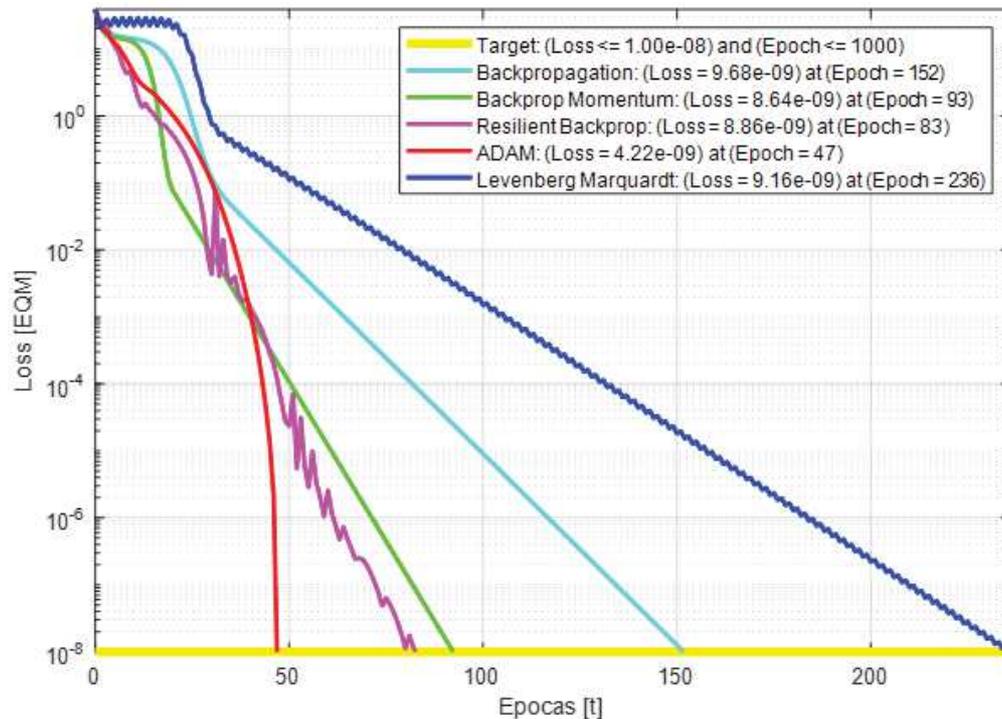
Figura 3.2 – Rede com filtragem antes da realimentação.



A proposta é comparar os otimizadores anteriormente listados utilizando sinais bem conhecidos, para avaliar seus desempenhos, tanto na predição quanto no tempo de convergência na presença de ruído. A Figura 3.3 apresenta os resultados destes otimizadores. Vale ressaltar que todos os otimizadores partiram dos exatos mesmos pesos sinápticos

iniciais, com o mesmo número de neurônios, e todos os demais parâmetros idênticos. Todos os otimizadores foram implementados pela autora, sem o uso de ferramentas prontas.

Figura 3.3 – Convergência dos algoritmos de treinamento.



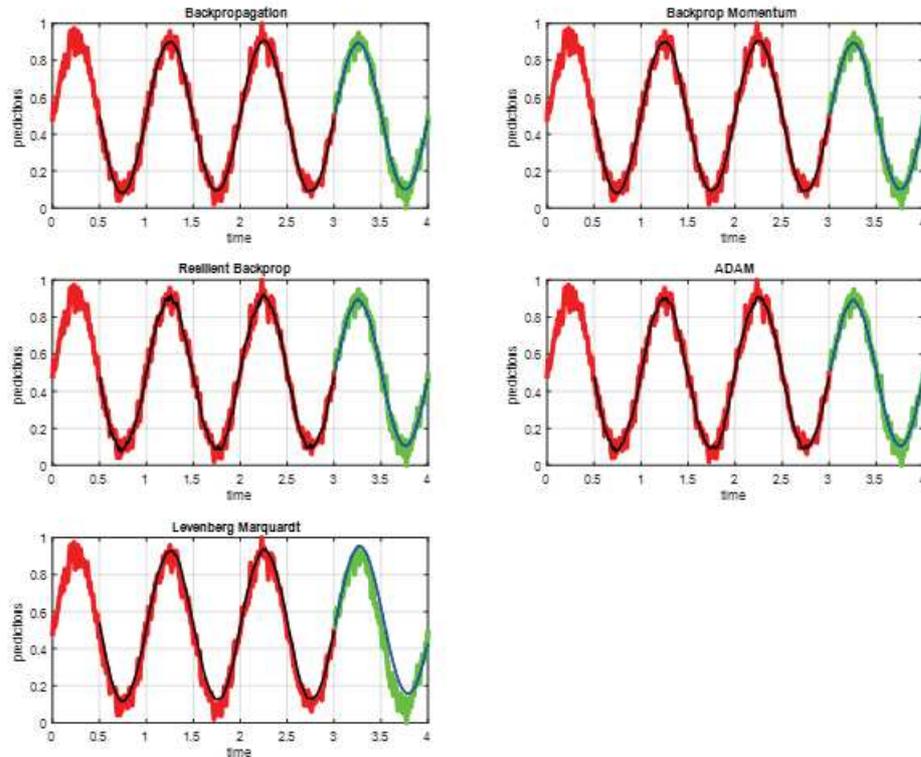
Nota-se, por inspeção visual, que o método *ADAM* (P.KINGMA; BA, 2014) converge rapidamente e, principalmente, sem demonstrar as mesmas oscilações do segundo colocado, o *Resilient Backpropagation*. Desta maneira, é possível confirmar, levando em consideração o afirmado no artigo dos autores, que o método *ADAM* é mais bem adaptado para lidar com séries temporais, especialmente as contaminadas por ruídos aleatórios, sendo um método de otimização estocástica.

O *ADAM* é bem adaptado para ruídos por suas propriedades inerentemente estatísticas de treinamento, uma vez que o *ADAM* utiliza tanto a Esperança Estatística (primeiro momento estatístico), quanto a Variância (segundo momento estatístico), para atualizar os pesos sinápticos da RNA. Isso faz com que esses pesos converjam rapidamente, com o efeito de *Momentum*, para a média dos dados, com baixo desvio padrão, tornando-o um dos métodos mais bem-sucedidos da literatura em treinamento de Redes Neurais Artificiais (P.KINGMA; BA, 2014).

A seguir, nas Figuras 3.4, 3.5 e 3.6, estão apresentados os resultados de treinamento e predição para diferentes funções de teste, bem conhecidas, contaminadas por ruído, utilizadas para avaliar a convergência dos algoritmos de treinamento. Em todas as Figuras, o sinal em “Vermelho” é a série original de treinamento, em “Verde” é a série

original futura (o objetivo a ser alcançado), em “Preto” é o aprendizado da RNA e, finalmente, o sinal em "Azul" refere-se à predição da variável de interesse no futuro pela Rede Neural.

Figura 3.4 – Seno contaminado por ruído aleatório Gaussiano.



Nota-se que, para as funções de teste, que são periódicas, a RNA teve um bom desempenho no aprendizado e na predição. Ainda assim, é possível perceber pequenas diferenças entre os treinamentos e predições, em que os melhores colocados foram os métodos *ADAM* e *Resilient Backpropagation*, tanto qualitativa quanto quantitativamente.

Considerando que o método *ADAM* é mais rápido e mais suave em sua convergência, este será o método adotado a ser aplicado no aprendizado e na predição do CMO para tempos futuros. Esse algoritmo, especializado em séries estocásticas, ou seja, contaminadas por ruídos aleatórios, utiliza uma variação do *Backpropagation*, baseado no cálculo de momentos, que é, de fato, a combinação das vantagens de outros dois algoritmos, o *AdaGrad* e o *RMSProp*. O *ADAM* utiliza taxas de aprendizado individuais adaptativas para o primeiro e o segundo momentos do gradiente, momentos estes estimados via média móvel, e faz também uma correção de viés de inicialização (P.KINGMA; BA, 2014).

O sistema elétrico brasileiro possui suas variáveis fortemente contaminadas por fatores aleatórios, como o clima, chuvas, disponibilidades de geração, demandas repentinas, dentre outras. Conforme desempenho dos otimizadores para dados na presença de ruídos, o

Figura 3.5 – Exponencial negativa multiplicada por seno.

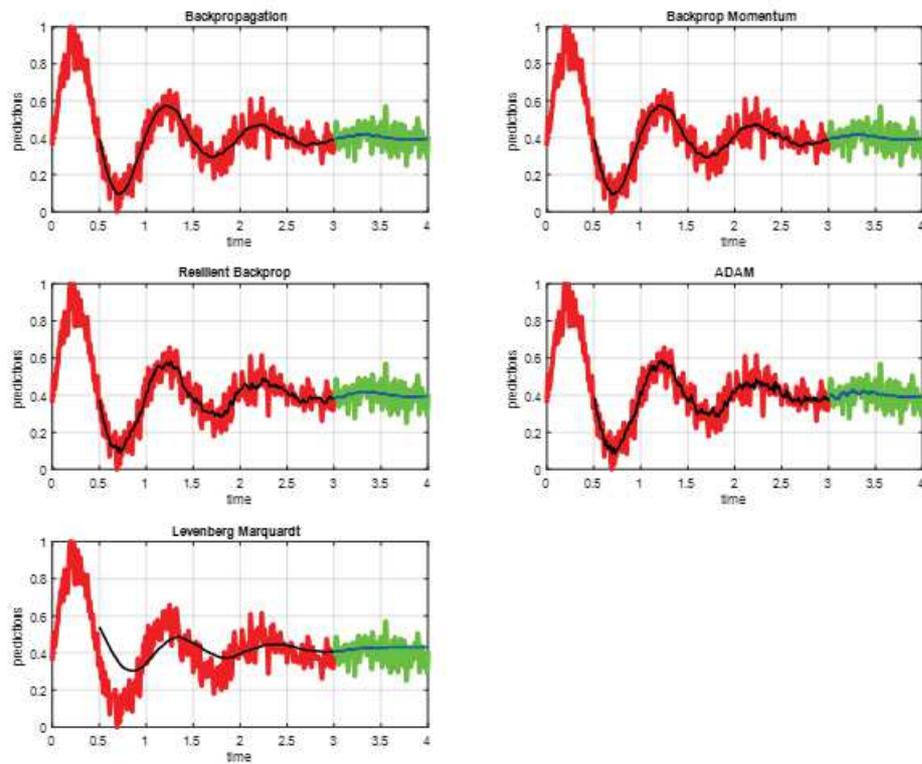
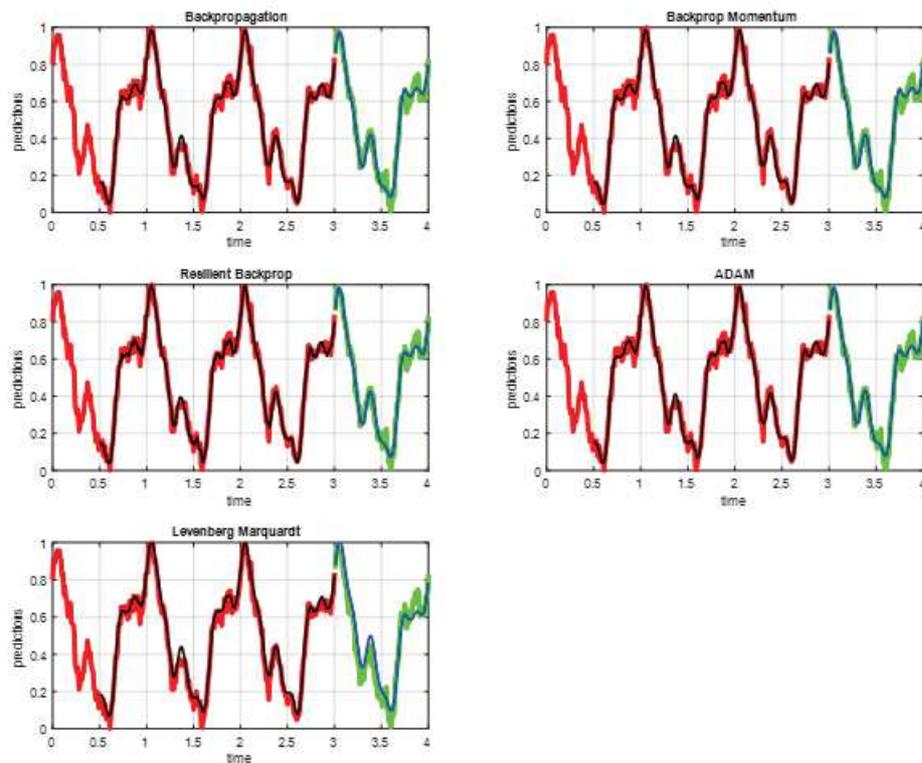


Figura 3.6 – Composição de senoides de diferentes frequências.

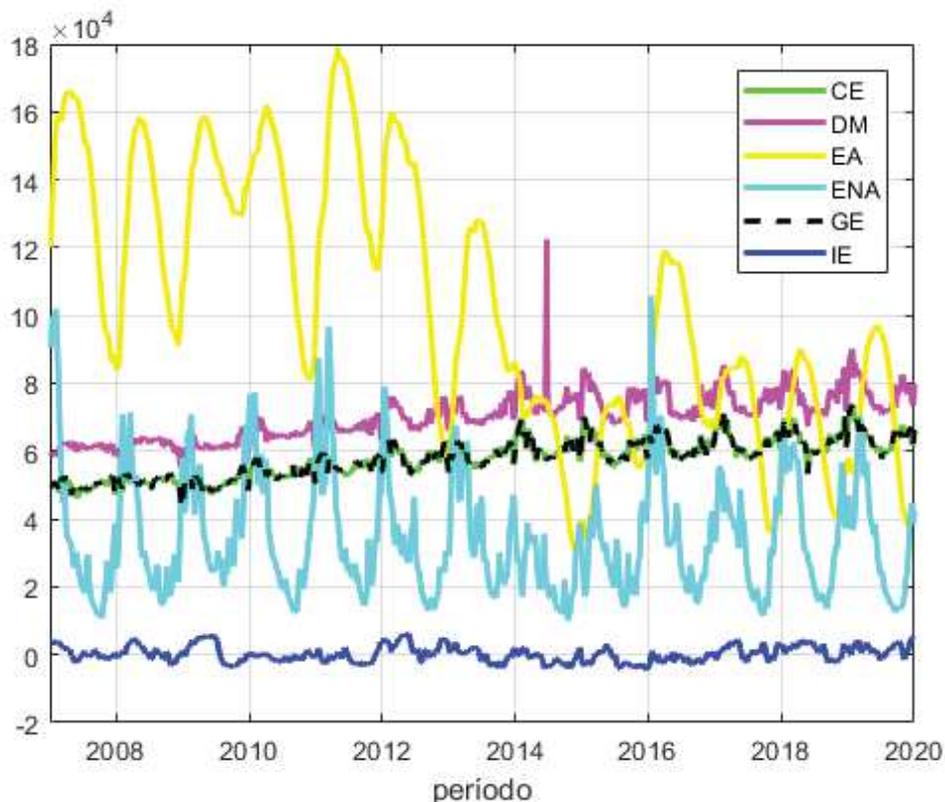


ADAM torna-se a escolha mais bem adaptada para o objeto de pesquisa desta dissertação. Suas vantagens incluem, ainda, um bom resultado com dados ruidosos e gradientes escassos (problema enfrentado principalmente por redes recorrentes), pouca necessidade de ajuste de parâmetros e uma melhor eficiência computacional.

3.3 Processamento das Variáveis

Para desenvolver os modelos de predição dos preços e tendências do PLD, a base de dados utilizada foi obtida na Câmara de Comercialização de Energia Elétrica (CCEE) e do Operador Nacional do Sistema Elétrico (ONS), e abrange da primeira semana operativa de 2007 até a última semana operativa de 2020, totalizando 730 semanas. As variáveis que compõem a base de dados de entrada foram a Energia Armazenada (EA), Demanda Máxima (DM), Carga de Energia (CE), Geração de Energia (GE), Energia Natural Afluente (ENA) e Intercâmbios de Energia (IE), além das próprias séries históricas dos valores do Preço de Liquidação das Diferenças (PLD) e Custo Marginal de Operação (CMO). Os dados contidos na base se referem ao submercado Sudeste/Centro Oeste. O comportamento das variáveis componentes do banco de dados está representado nas Figuras 3.7 e 3.8, que demonstram o comportamento das mesmas para o período de tempo escolhido.

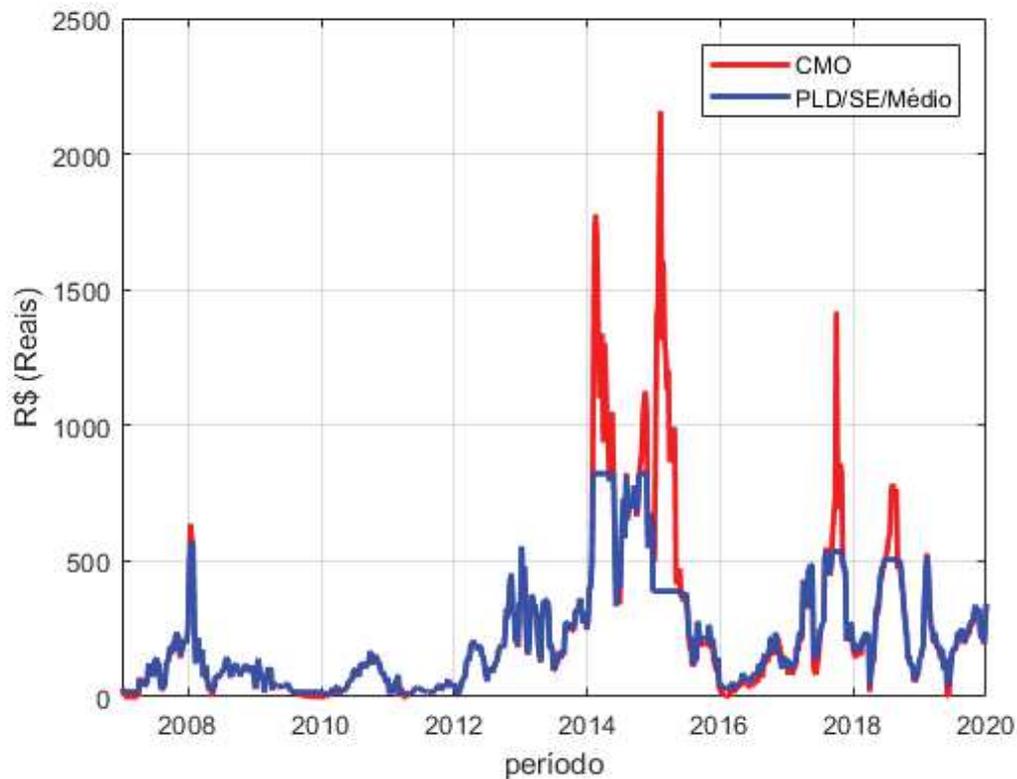
Figura 3.7 – Séries de dados da ONS e suas unidades: CE [MWmed], DM [MWh/h], EA [MWmês], ENA [MWmed], GE [MWmed] e IE [MWmed].



Fonte: (ONS, 2020).

Pode-se observar que as variáveis, sujeitas a sazonalidade das estações, apresentam volatilidade e incertezas. Esta forte volatilidade se deve à influência constante de variáveis exógenas como chuvas, clima, disponibilidades de geração, demandas repentinas, dentre diversas outras. Ainda assim, quando comparadas com o CMO/PLD, possuem um comportamento mais propício a previsões, uma vez que se verifica, por inspeção visual da Figura 3.7 que existem estruturas periódicas nestas variáveis. Estruturas periódicas produzem altas correlações cruzadas, favorecendo algoritmos de aprendizado como Redes Neurais, que capturam essas tendências para prever acuradamente comportamentos futuros.

Figura 3.8 – Série de dados – CMO e PLD.



Fonte: (ONS, 2020) e (CCEE, 2021d).

Primeiramente, as variáveis foram conduzidas a um pré-processamento, objetivando realizar um ajuste de base temporal e compensar dados faltantes para todas. Os poucos dados faltantes foram estimados por meio de interpolação linear entre o dado imediatamente anterior e posterior da série em ordem cronológica. É de suma importância que todas as séries de dados históricos estejam na mesma base temporal e compostas pelo mesmo número de elementos, a fim de que eventuais descompassos temporais entre as variáveis não acarretem erros de previsão.

Após serem processados e completados, os dados são normalizados no intervalo $[0, 1]$. Esta etapa consiste no escalonamento das amostras para a faixa de variação dinâmica

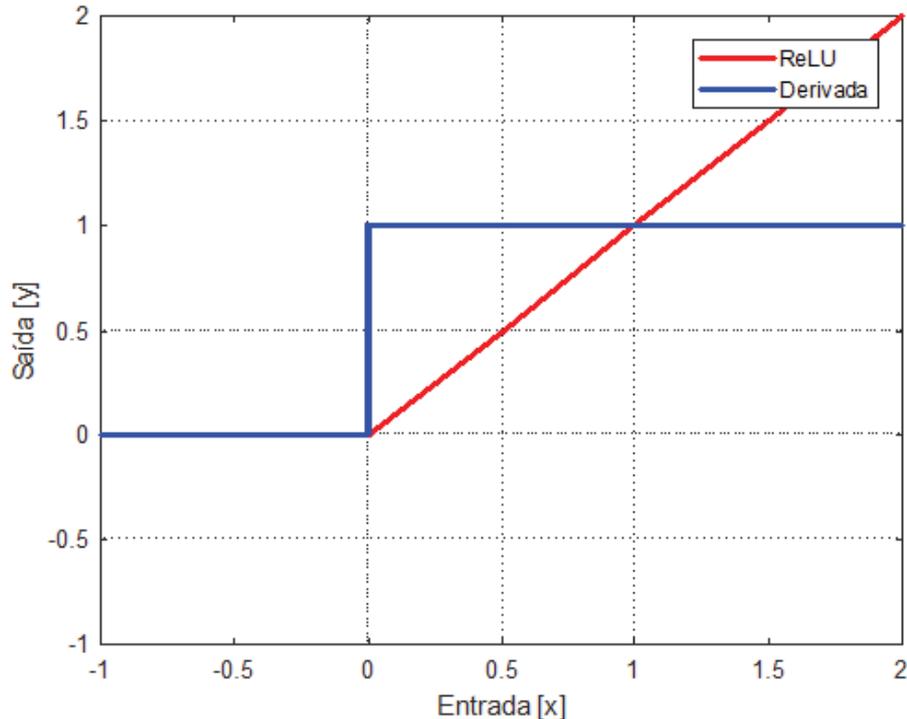
da função de ativação. Desta forma, os valores das entradas são mantidos em um intervalo condizente com os principais teoremas de convergência, e respeita os limites das principais funções de ativação utilizadas em Redes Neurais, evitando a saturação da saída do neurônio quando os potenciais de ativação apresentarem valores elevados.

Diversas funções de ativação foram experimentadas, como a tangente hiperbólica, logística, ReLU, Softplus e exponencial. Após diversos testes, a função de ativação escolhida para a Rede Neural deste trabalho é a ReLU (*Rectified Linear Unit*). Um neurônio com uma função de ativação ReLU assume quaisquer valores reais como sua(s) entrada(s), mas só é ativado quando essas entradas são maiores que 0. A função e sua derivada estão representadas nas Equações 3.2 e 3.3 e representadas graficamente na Figura 3.9.

$$ReLU(x) = \max\{0, x\} \quad (3.2)$$

$$ReLU'(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{c.c} \end{cases} \quad (3.3)$$

Figura 3.9 – Função ReLU e sua derivada.



A ReLU é a função de ativação mais utilizada nos projetos de Redes Neurais modernos (ACADEMY, 2019). Ela mantém a linearidade desejada na parte positiva, e insere a não linearidade na parte negativa, permitindo um treinamento mais rápido e efetivo de arquiteturas mais profundas em extensos e complexos conjuntos de dados,

por não fazer uso de expoentes, realizando apenas operações de comparação, adição e multiplicação. De forma a evitar as regiões de saturação da função ReLU, optou-se por normalizar os dados no intervalo de 0 a 1. A Equação 3.4 demonstra como é feita a normalização dos dados, por meio de uma relação linear:

$$data_{norm} = max_{norm} - \frac{(max_{norm} - min_{norm})(max_{data} - data)}{max_{data} - min_{data}} \quad (3.4)$$

Na Equação 3.4, $data$ é o conjunto de treinamento, no caso, a série temporal representada pelos valores das variáveis disponibilizadas pela CCEE e ONS, os quais se deseja aprender com a rede; $data_{norm}$ é a função de treinamento normalizada; max_{norm} é o valor máximo permitido para a normalização e min_{norm} é o respectivo valor mínimo permitido, então $data_{norm}$ está contido no intervalo $[max_{norm}, min_{norm}]$; equitativamente, max_{data} é o valor máximo permitido para a série temporal e min_{data} o respectivo valor mínimo.

Após a normalização, o conjunto de dados (14 anos de operação do sistema) foi dividido, sendo as 677 primeiras semanas (13 anos) para treinamento (92,74%) e as 53 últimas para validação (7,26%), o que corresponde ao ano de 2020. O conjunto de treinamento é responsável pelo aprendizado da rede e ajuste dos pesos sinápticos. O segundo conjunto é utilizado para avaliar o desempenho de generalização da rede.

Os dados de treinamento, composto pelo conjunto das entradas e saídas desejadas é obtido a partir dos dados normalizados. Tendo em vista a topologia representada na Figura 2.4, no primeiro instante a rede recorrente faz uso dos sinais de entrada (composto por uma combinação de variáveis) atrasadas no tempo em conjunto com os sinais de contexto iguais a zero, uma vez que a saída inexiste no momento. No instante posterior, o primeiro valor produzido pela saída da rede estará disponível e pode realimentar uma entrada da rede, enquanto as demais ainda são zero. Sucessivamente, os próximos instantes das saídas estarão disponíveis e serão introduzidas nas entradas. O conjunto de treinamento é construído com a repetição do processo descrito para todos os instantes posteriores.

Conforme demonstrado na seção 3.2, após a realização de testes com algoritmos de treinamento mais comumente utilizados na literatura, optou-se por implementar o *ADAM*, que é um algoritmo de treinamento com boa adaptação para dados estocásticos, além da sua atrativa velocidade de convergência. O aprendizado da rede é feito de forma supervisionada, mediante a apresentação do conjunto dos dados de entrada e os desejados de saída. De forma iterativa, à medida em que a rede recebe as entradas, os resultados da saída são comparados com as saídas desejadas.

O processo de treinamento é utilizado até que a resposta atenda às condições de parada. Posteriormente, os dados são generalizados e preditos para quantos passos no futuro sejam desejados. A rede autorregressiva, realimentada com a saída filtrada da mesma tem o propósito de generalizar as tendências dos dados-objetivo, no caso, do CMO.

Isto significa que a missão não é prever o preço futuro do CMO com total acurácia, e sim, capturar seu comportamento geral, de tal maneira que um analista possa utilizar esta informação para decidir qual o melhor momento para gerar contratos de compra e venda de energia.

4 RESULTADOS E ANÁLISES

Este capítulo apresenta os resultados obtidos com a aplicação dos métodos propostos para a predição do CMO e posterior análise do comportamento do PLD aplicados à base de dados disponibilizada pela CCEE. Além do desempenho da rede recorrente, será explorada a combinação de filtros à RNA. Para a consecução da pesquisa, todo o desenvolvimento e análise foram conduzidos através do *software* Matlab ([MATHWORKS, 2021](#)).

Toda a implementação dos resultados apresentados, envolvendo a RNA, análises estatísticas e o processamento foi realizada pela autora no âmbito da presente dissertação, sem a utilização de pacotes prontos. O desempenho de predição dos modelos implementados será avaliado retendo alguns dados da amostra para, em seguida, utilizar o modelo para fazer predições para os dados de validação, a fim de ver o quão precisos eles são.

A estatística de erro utilizada tem como base a raiz quadrada do erro quadrático médio (*Root Mean Squared Error*, RMSE), que conforme descrito por ([MEDIUM, 2016](#)), é uma das métricas mais utilizadas para medir a acurácia de modelos preditivos e, de fato, uma das mais utilizadas na literatura em função de sua propriedade estatística, embora existam outras. O fato interessante é ela ter a mesma unidade, no caso, Reais (R\$/MWh), da variável de interesse. A partir dela é possível avaliar o desempenho da Rede Neural, possibilitando a escolha da melhor configuração. Nesta métrica, os dados preditos pela rede, após treinada, são comparados com um conjunto de dados que não está presente no processo de treinamento, gerando os erros de predição a partir da Equação 4.1:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |P_i - Q_i|^2} \quad (4.1)$$

em que n é o número de amostras no conjunto, P_i e Q_i são os i -ésimos valores desejado e obtido pela RNA, respectivamente.

A qualidade do treinamento da Rede Neural Artificial está relacionada diretamente com a definição de determinados parâmetros. Foram realizados diversos testes empíricos para a definição de alguns parâmetros e, a partir destes testes, foram fixados os que apresentaram melhor desempenho para a RNA na capacidade de aprendizagem e predição. É necessário fixar alguns parâmetros para evitar que o universo de busca se torne muito extenso. Em relação ao método de treinamento *ADAM*, foram fixadas as taxas de aprendizado do primeiro e segundo momento, $\beta_1 = 0,5$ e $\beta_2 = 0,95$, bem como o valor da taxa de aprendizado por época $\alpha = 0,01/\sqrt{t}$ e $\epsilon = 1 \times 10^{-8}$.

O critério de parada do treinamento foi limitado em 1×10^{-9} . Ou seja, o algoritmo converge quando o erro calculado em função do RMSE atinge a precisão requerida. Em conjunto com o erro, utilizou-se a contabilização do número de épocas necessárias para a efetivação do treinamento da RNA. Este parâmetro não pode ser pequeno, de forma que a Rede Neural Artificial não consiga aprender adequadamente, e nem elevado

demasiadamente para não provocar *overtraining*, condição em que a rede memoriza os padrões fornecidos, tornando-se incapaz de generalização. Constatou-se que a RNA produz respostas com qualidade quando treinada em até 150 épocas ($t_{max} = 150$).

A seguir apresentam-se os resultados para diferentes situações, em que a inserção de cada uma das várias estratégias será gradual, para que seus efeitos possam ser analisados frente a um histórico evolutivo, partindo das implementações mais simples.

4.1 CMO como única variável de entrada

Esta seção apresenta os resultados obtidos para a predição do CMO para o submercado Sudeste/Centro-Oeste, utilizando apenas o próprio histórico do CMO como entrada da RNA. Serão aplicadas duas metodologias para explorar a capacidade de predição da Rede Neural Recorrente. A primeira metodologia refere-se à abordagem clássica e consiste na utilização da RNA para a predição do futuro a um passo, ou seja, a rede preverá apenas o valor do CMO uma semana à frente. Na segunda abordagem, tem-se a predição de vários passos no futuro, no caso, um ano inteiro à frente, correspondente ao ano de 2020.

Com os parâmetros fixos definidos nas seções anteriores, foram testadas diversas configurações de topologias de rede, objetivando encontrar as topologias que permitam uma melhor generalização dos dados. Os testes foram feitos variando o número de camadas escondidas, bem como o número de neurônios pertencentes a cada camada. Outro parâmetro alterado foi o número de entradas atrasadas no tempo e o número de saídas realimentadas como entrada da rede. A escolha de tais parâmetros influencia o treinamento e a convergência da RNA, podendo causar, inclusive, a divergência da mesma.

4.1.1 Predição de um passo no futuro

Nesta primeira abordagem, foram simuladas 16 configurações diferentes de RNAs, com variações tanto no número de camadas escondidas quanto no número de neurônios por camada além das alterações no número de atrasos das entradas e realimentações da saída, que variaram entre 2, 10, 20 e 30. A razão das escolhas desta faixa de valores para os atrasos em específico foram, inicialmente, por busca empírica com base em trabalhos anteriores (NASCIMENTO, 2017), (NASCIMENTO; OLIVI; LIMA, 2018) e (NASCIMENTO et al., 2019), porém, após a análise estatística apresentada, em breve, na Seção 4.2, este intervalo se consolidou como o melhor, tanto em função da correlação cruzada entre as séries temporais quanto pela reduzida quantidade de dados disponíveis.

A Tabela 4.1 mostra as configurações e resultados das RNAs simuladas, em que N_1 , N_2 e N_3 correspondem ao número de neurônios na primeira, segunda e terceira camada oculta, respectivamente, n_p é o número de atrasos da entrada e n_q é o número de realimentações das saídas.

Tabela 4.1 – Resultados por topologia - 1 passo no futuro

Configurações da rede				Treinamento			Validação		
N_1	N_2	N_3	n_p e n_q	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão
10	x	x	2	6,7161	2,5916	10,9769	4,9243	2,2191	10,1995
			10	6,5275	2,5549	10,8980	4,6885	2,1653	10,0727
			20	6,6016	2,5694	10,9300	4,9591	2,2269	10,2033
			30	6,9219	2,6310	11,0538	5,0530	2,2479	10,2648
20	20	x	2	4,8229	2,1961	10,0945	4,1018	2,0253	9,7458
			10	4,7187	2,1723	10,0341	3,8761	1,9688	9,6070
			20	4,6690	2,1608	10,0030	3,9083	1,9769	9,6270
			30	4,9923	2,2344	10,1628	3,9810	1,9952	9,6693
10	20	15	2	4,7439	2,1780	10,0599	4,1041	2,0259	9,7479
			10	4,6275	2,1512	10,0008	3,9436	1,9858	9,6460
			20	4,6750	2,1622	10,0111	3,9332	1,9832	9,6435
			30	4,8478	2,2018	10,0983	3,9614	1,9903	9,6588
50	50	50	2	4,9478	2,2244	10,1688	4,2112	2,0521	9,7994
			10	4,5841	2,1410	9,9595	3,9270	1,9817	9,6369
			20	5,5472	2,3553	10,4646	4,3924	2,0958	9,9004
			30	10,2907	3,2079	12,2130	6,7672	2,6014	11,0424

Analisando os resultados apresentados na referida tabela é possível perceber que as redes configuradas com 10 e 20 atrasos TDNN nos dados do CMO de entrada (n_p) e 10 e 20 realimentações da saída (n_q) produzem bons resultados em todas as variações dos números de camadas e neurônios por camada, com um destaque maior para as topologias com duas camadas escondidas ($N_1 = N_2 = 20$) e 3 camadas escondidas ($N_1 = 10$, $N_2 = 20$ e $N_3 = 15$) que apresentaram os melhores resultados globais. Os resultados de predição destas topologias são qualitativamente muito próximos e estão representados nas Figuras 4.1 e 4.2 para $N_1 = N_2 = 20$ e nas Figuras 4.3 e 4.4 para $N_1 = 10$, $N_2 = 20$ e $N_3 = 15$.

Ao utilizar os atrasos da entrada e da realimentação da saída nos valores mínimo e máximo, ou seja, $n_p = n_q = 2$ ou 30, a RNA apresenta os maiores erros para quaisquer configurações em relação ao número de camadas escondidas e número de neurônios por camada. A topologia com pior desempenho é representada pelas Figuras 4.5 e 4.6, com 50 neurônios nas camadas intermediárias (N_1 , N_2 e N_3), 30 atrasos da entrada (n_p) e 30 realimentações das saídas (n_q). Nota-se que mesmo no pior resultado de predição, ainda assim a rede foi capaz de acompanhar a tendência dos dados, mostrando que apesar do erro ter sido o menos satisfatório, a rede incorporou o conhecimento da variação dos dados.

Tem-se que para as diferentes quantidades de neurônios nas camadas intermediárias, os melhores resultados de predição, independente dos valores assumidos por n_p e n_q , são produzidos pela rede com 2 camadas escondidas ($N_1 = N_2 = 20$). Ao elevar o número de camadas, a rede com 3 camadas ($N_1 = 10$, $N_2 = 20$ e $N_3 = 15$) produz resultados satisfatórios, mas ligeiramente piores quando comparados aos resultados da rede de duas camadas, evidenciando o comportamento altamente não-linear das Redes Neurais, em que não necessariamente a rede com maior número de neurônios terá melhor desempenho.

Figura 4.1 – Predição para $N_1=N_2=20$ e $n_p=n_q=10$.

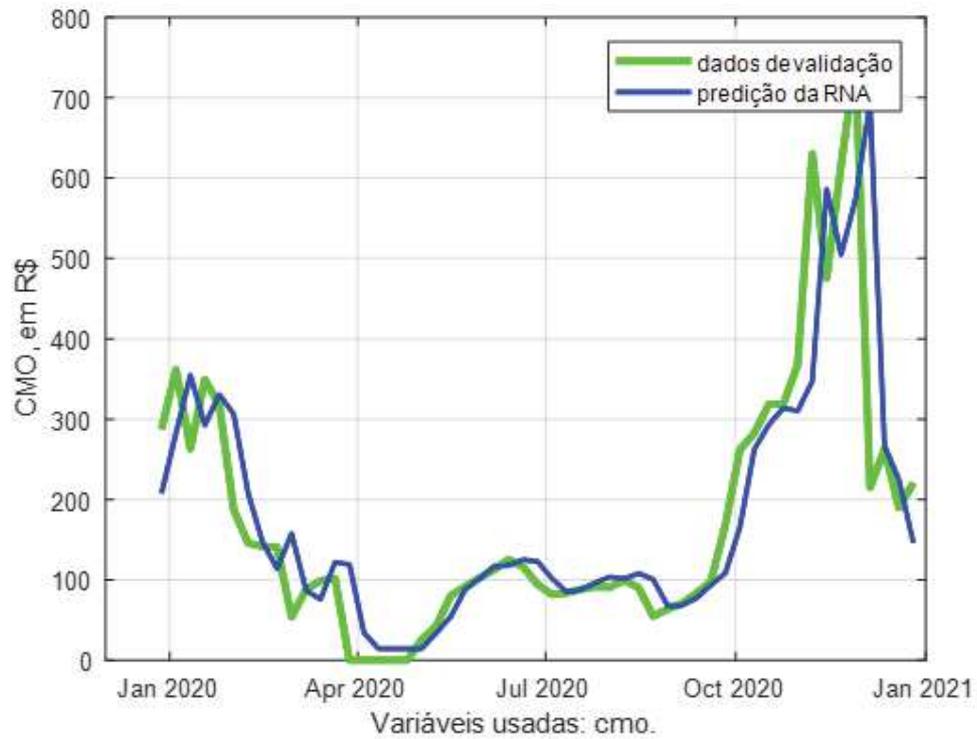


Figura 4.2 – Predição para $N_1=N_2=20$ e $n_p=n_q=20$.

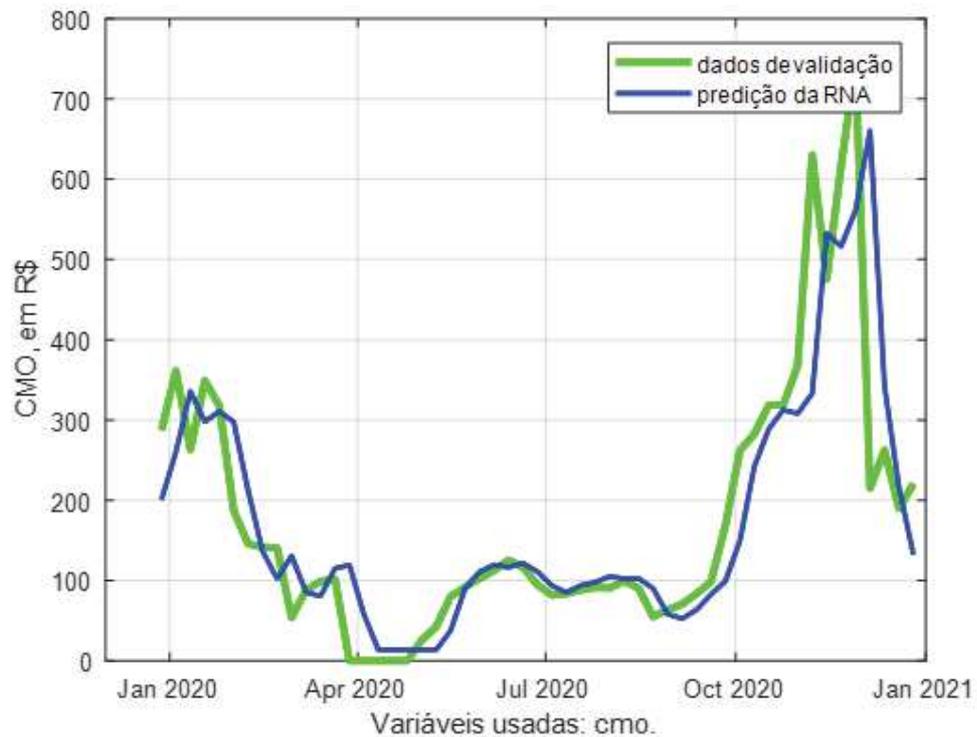


Figura 4.3 – Predição para $N_1=10$ $N_2=20$ $N_3=15$ e $n_p=n_q=10$.

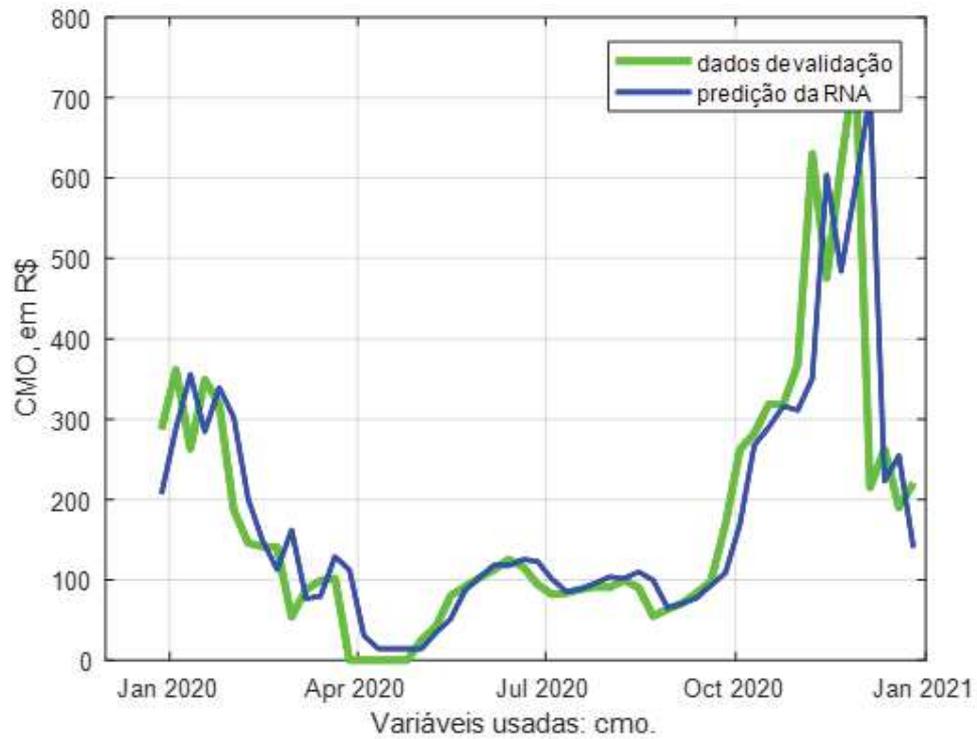


Figura 4.4 – Predição para $N_1=10$ $N_2=20$ $N_3=15$ e $n_p=n_q=20$.

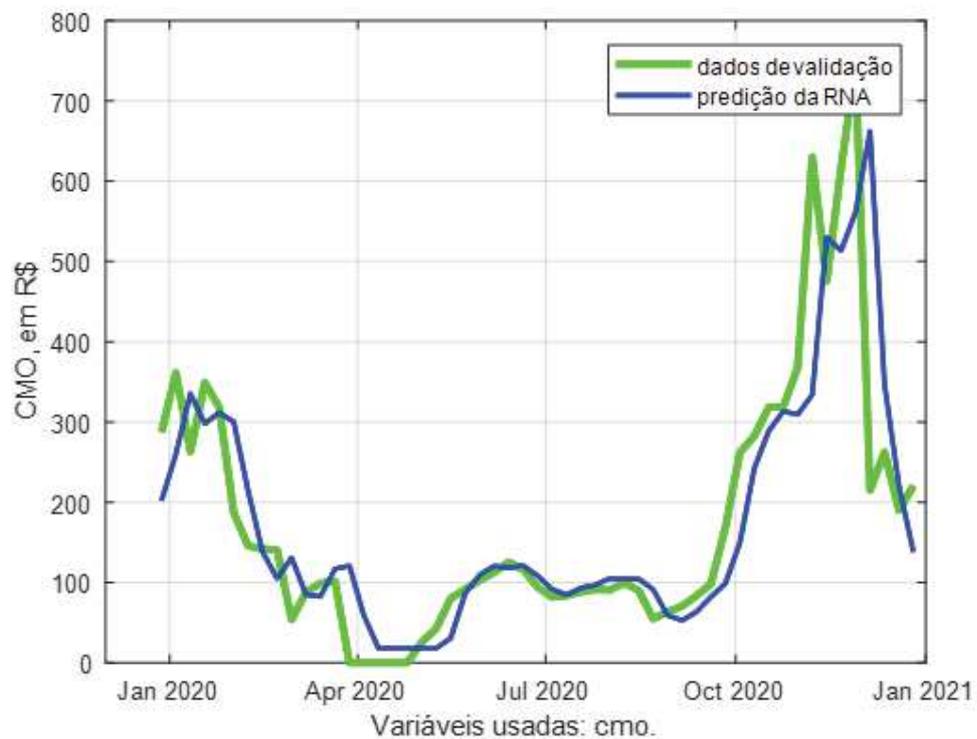


Figura 4.5 – Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=30$.

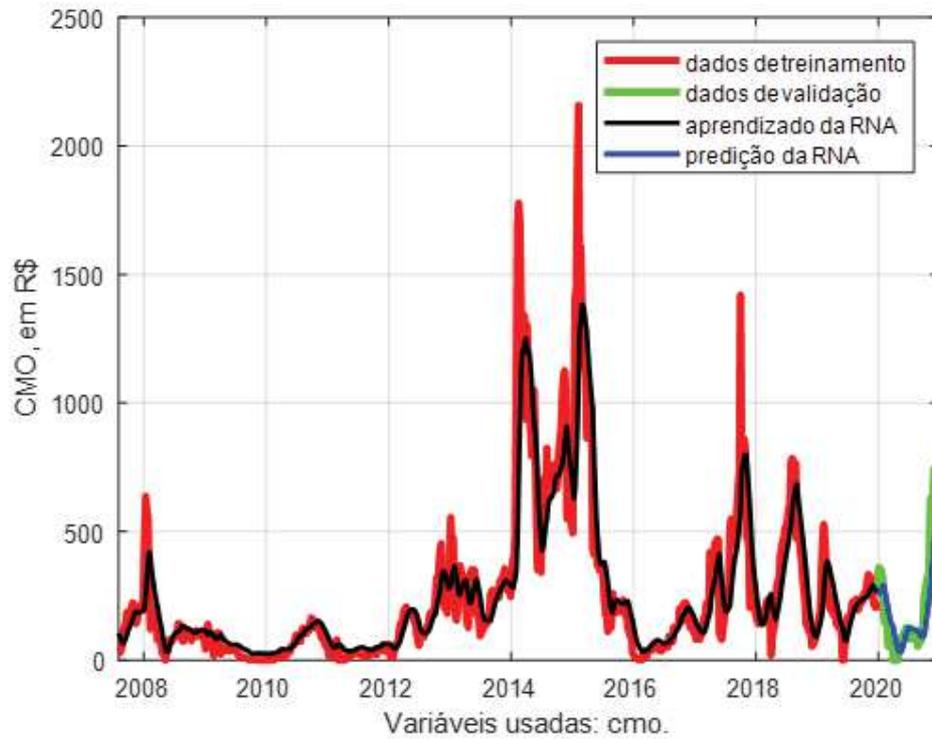
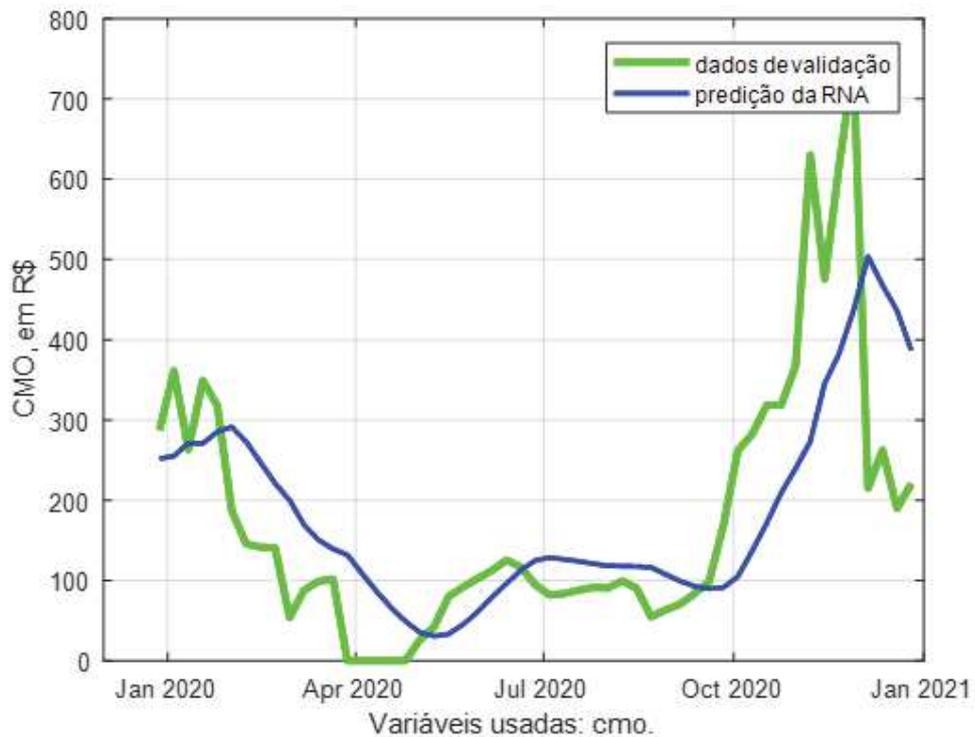


Figura 4.6 – Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=30$.



Na topologia com menor quantidade de camadas ($N_1=10$) os indicadores retornaram os piores resultados, denotando a falta de capacidade de generalização. Contudo, mesmo as piores topologias, apresentam resultados consideráveis, com erros de aproximadamente 2,21 R\$/MWh demonstrando uma boa capacidade de generalização da rede e conseguindo capturar boas tendências dos preços, auxiliando tomadas de decisões no curtíssimo prazo.

4.1.2 Predição de vários passos no futuro

Esta seção apresenta os resultados considerando a predição de vários passos no futuro a cada rodada da rede, no caso, 1 ano completo. Nesta abordagem não estão disponíveis para a rede informações do conhecimento das semanas imediatamente anteriores na etapa de predição, sendo realimentada a rede pela sua própria predição futura.

De maneira a expor a diferença entre ambas as metodologias, tem-se a seguinte situação: a predição do dado futuro do tempo $t + 1$ resulta na saída y_{t+1} cujo resultado verdadeiro é d_{t+1} . Quão menor é o erro entre y_{t+1} e d_{t+1} , melhor a generalização da rede. Para a topologia da seção anterior, que prediz apenas um passo no futuro, a predição do resultado do próximo tempo, $t + 2$, utilizaria como dado realimentado o dado verdadeiro d_{t+1} , uma vez que este está disponível. Já a topologia desta seção realimentará a própria saída da rede, y_{t+1} , considerando que os dados verdadeiros não estão disponíveis.

Os testes foram feitos para as mesmas topologias apresentadas na seção anterior. A Tabela 4.2 apresenta os resultados quantitativos para a predição de até um ano no futuro (53 semanas) e as Figuras 4.7 e 4.8 apresentam os resultados quantitativos da melhor e pior topologia, respectivamente.

Tabela 4.2 – Resultados por topologia - vários passos no futuro (N=53 passos)

Configurações da rede				Treinamento			Validação		
N_1	N_2	N_3	n_p e n_q	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão
10	-	-	2	40,1190	6,3340	16,9981	12,2712	3,5030	12,8151
			10	32,0223	5,6588	16,1098	12,2886	3,5055	12,7913
			20	33,3683	5,7765	16,1887	12,4814	3,5329	12,8728
			30	6,9219	2,6310	11,0538	12,7619	3,5724	12,8793
20	20	-	2	4,5580	2,1350	9,9590	12,3320	3,5117	12,8277
			10	5,2482	2,2909	10,3092	12,5313	3,5400	12,8856
			20	34,8049	5,8996	16,4250	12,4607	3,5300	12,8457
			30	6,0815	2,4661	10,6949	12,6414	3,5555	12,9138
10	20	15	2	44,2286	6,6505	17,4201	12,2794	3,5042	12,8164
			10	44,9028	6,7010	17,2718	12,4960	3,5350	12,8050
			20	42,1993	6,4961	17,1117	12,3015	3,5073	12,8202
			30	40,1005	6,3325	17,0009	12,4708	3,5314	12,8505
50	50	50	2	4,6222	2,1499	9,9921	12,3299	3,5114	12,8336
			10	4,9279	2,2199	10,1529	12,5194	3,5383	12,8690
			20	34,9152	5,9089	16,3336	12,3904	3,5200	12,8430
			30	35,5607	5,9633	16,4486	15,5224	3,5387	12,8833

Figura 4.7 – Predição para $N_1=N_2=20$ e $n_p=n_q=2$.

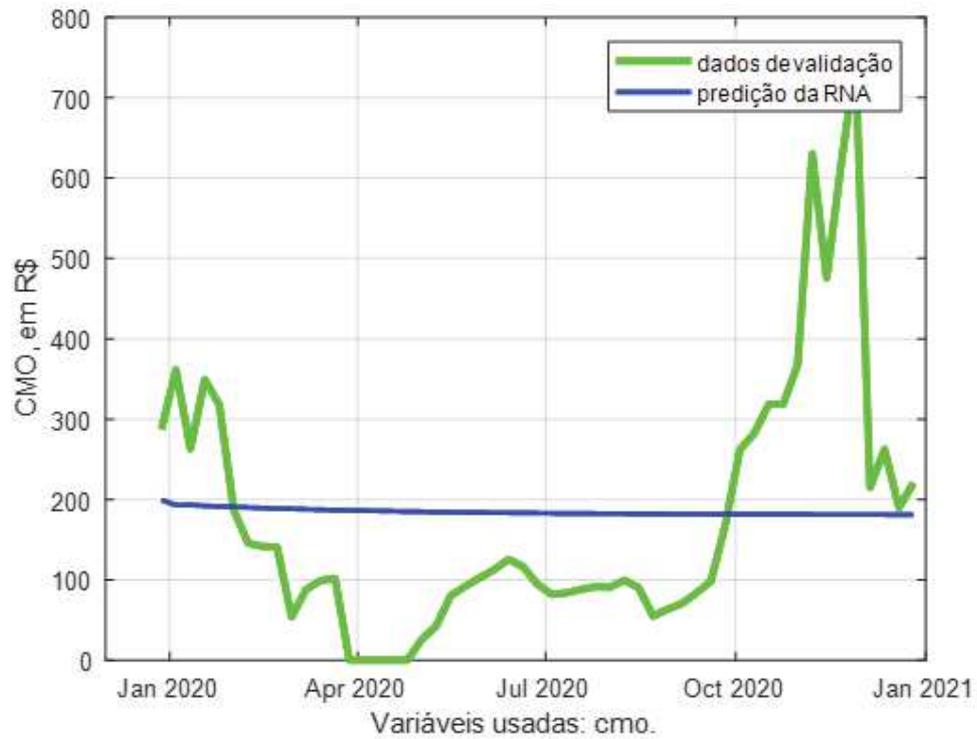
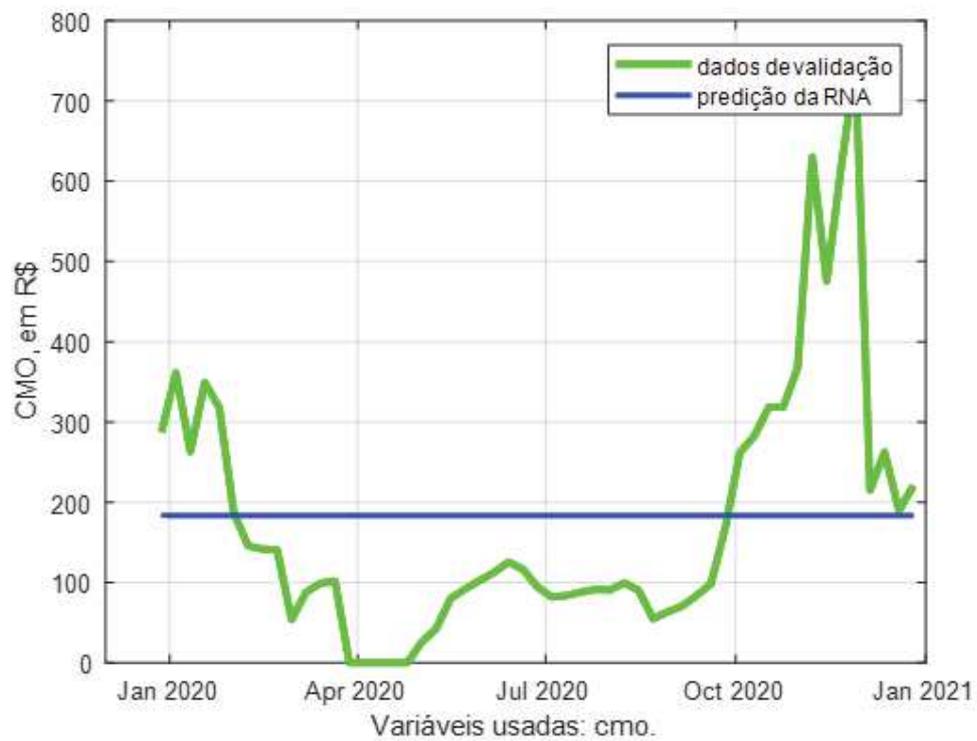


Figura 4.8 – Predição para $N_1=10$, $N_2=20$, $N_3=15$ e $n_p=n_q=2$.



Como pode-se observar, as Figuras 4.7 e 4.8 convergiram para a média dos dados. Esta é uma particularidade do método de treinamento, o *ADAM*, por ser um treinamento estocástico baseado no primeiro e segundo momentos estatísticos, é esperado que minimize ambos na fase de atualização dos pesos sinápticos. O resultado produzido convergiu para próximo da média anual de preços do CMO, o que não é um resultado satisfatório em termos de captura de tendências da variável ao longo do ano.

Estratégia similar foi executada em (NASCIMENTO et al., 2019), porém, com outro método de treinamento, o *Resilient Backpropagation*, e para a variável PLD, que, conforme visto ao longo deste trabalho, é uma versão saturada do CMO. Nesta oportunidade, verificou-se um comportamento diferente imposto pelo método de treinamento, a predição não convergiu para a média e sim, para um comportamento de alta volatilidade. Porém, o resultado de predição para vários passos no futuro mostrou-se também não satisfatório.

A conclusão que se levanta para ambos os resultados é que a própria variável, CMO (e, conseqüentemente, PLD), não possuem informações suficientes dentro de si próprias, para realizar uma predição futura para vários passos. Estas variáveis não possuem estruturas de repetição periódicas ao longo do tempo, e são contaminadas por ruídos aleatórios. Estes fatores, em conjunto, tornam esta variável imprópria para predição de longo prazo se utilizada sem nenhum outro tipo de informação adicional.

Esta conclusão também é suportada pela Figura 4.9, que mostra a correlação cruzada do CMO consigo mesmo, para janelas temporais diferentemente deslocadas no tempo. Nota-se que o mesmo possui alta correlação na diagonal e seu entorno, a situação em que as janelas estão temporalmente pouco distantes, sendo uma possível explicação para o bom desempenho da predição de uma semana a frente. No entanto, se o distanciamento temporal das janelas é grande, a correlação é baixa, o que lembra o comportamento de uma variável aleatória, dada na Figura 4.10. Esta pode ser uma das razões da predição de vários passos não funcionar bem para uma rede cujas entradas sejam alimentadas apenas pelo CMO, uma vez que não exista uma correlação temporal de longo alcance para o CMO. Seriam necessárias, portanto, outras fontes de informação nas entradas da Rede Neural para que a mesma pudesse aprender padrões úteis na predição de longo prazo do CMO.

No entanto, o CMO/PLD é composto por outras variáveis, conforme exposto ao longo dos capítulos, variáveis estas disponíveis no ONS (ONS, 2020). O mecanismo de confecção do CMO/PLD leva em consideração diversas informações disponíveis no histórico do ONS. Portanto, para que informações adicionais úteis sejam inseridas na Rede Neural, de forma que esta possa capturar as tendências e auxiliar especialistas na tomada de decisões contratuais, diversas variáveis da ONS foram consideradas estatisticamente. Correlações entre estas variáveis e o CMO foram calculadas para diversas janelas temporais e testadas como entradas da Rede Neural, com o objetivo de fornecer informações importantes para o aprendizado de padrões.

Figura 4.9 – Correlação cruzada do CMO para janelas de dados da série temporal, evidenciando a correlação entre dados da mesma série temporal em tempos diferentes.

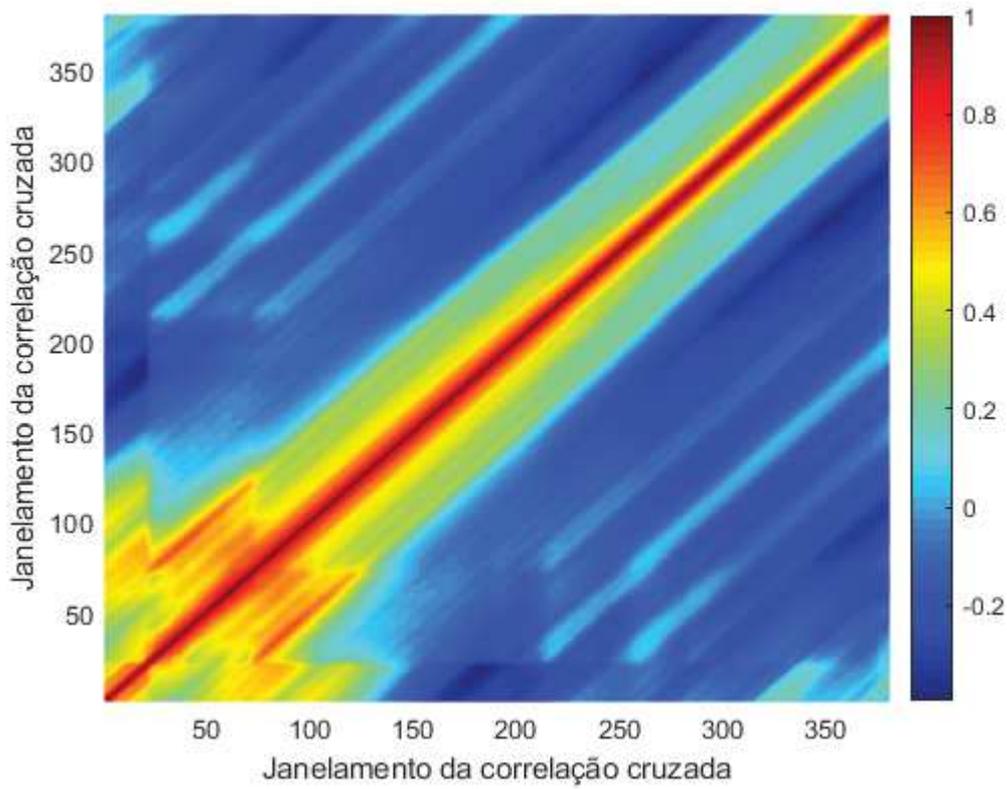
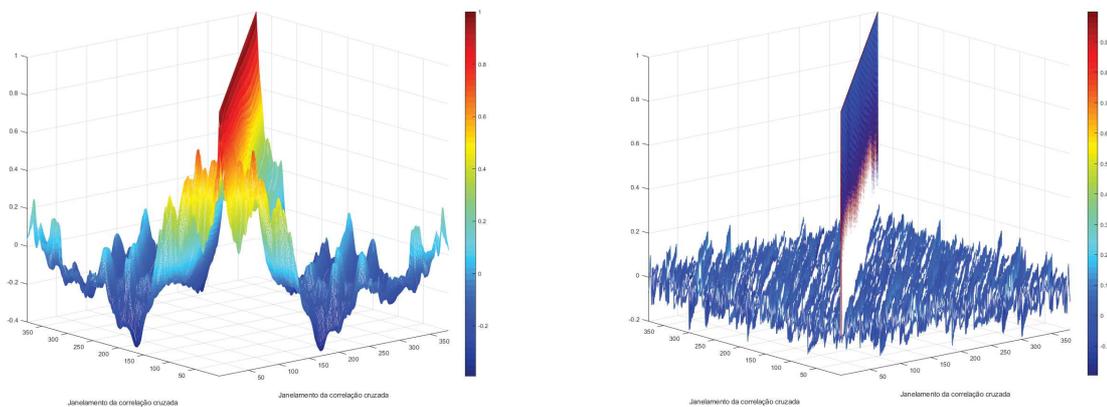


Figura 4.10 – Correlação cruzada: esquerda) CMO e direita) ruído Gaussiano, evidenciando a baixa correlação do CMO para janelas em instantes de tempos distantes, de maneira similar a variável aleatória, que não uma possui alta intensidade de estruturas repetidas ou periódicas no longo prazo.



4.2 Variáveis ONS como variáveis de entrada

Para seleção das variáveis explanatórias a comporem a base de dados de entrada, foi realizada uma análise de correlação que indica a similaridade de cada variável com o CMO. Esta análise foi feita por meio de janelamentos de 2 à 730, que é o número total de dados. Para cada janelamento, calculou-se a correlação entre cada variável e o CMO. A correlação em janelas foi realizada pois, em uma rede com entradas atrasadas no tempo – TDNN (*Time Delay Neural Network*), a predição de valores posteriores a partir do instante t , associadas ao comportamento do processo, é computada em função do conhecimento de seus valores imediatamente anteriores, isto é:

$$x(t) = f(x(t-1), x(t-2), \dots, x(t-n_p)) \quad (4.2)$$

em que n_p é a quantidade de amostras passadas que serão necessárias para fazer a estimativa de do valor $x(t)$.

Estes atrasos de tempo garantem que as amostras anteriores que refletem o comportamento temporal do processo sejam sempre inseridas dentro da rede. Desta forma, a correlação janelada permite a escolha do melhor número de atrasos de tempo nas entradas, ou seja, aquele que indicar a melhor correlação entre as variáveis de maneira a maximizar as chances da rede ter uma boa predição.

A Figura 4.11 apresenta a correlação absoluta das variáveis com o CMO em função do tamanho das janelas. O indicador de correlação varia de -1 a 1, onde 0 representa nenhuma correlação entre as variáveis. Na correlação positiva, à medida que uma variável cresce a outra também cresce e na negativa, à medida que uma cresce a outra decresce.

Nota-se, por inspeção visual, que a melhor variável analisada em termos de correlação é a Energia Armazenada (EA), que em todas as situações de janelamento mantém uma correlação superior a 0,5. É possível observar ainda, que a partir de janelas de tamanho 400, há uma estabilização da correlação. Contudo, em função do número limitado de amostras, é necessário escolher a menor janela possível que mantenha um nível aceitável de correlação, de forma que haja dados suficientes de entrada para que a rede consiga generalizá-los. Para escolher o melhor número de entradas atrasadas no tempo, foi feita uma análise das correlações para janelamentos de até 30 dados, conforme evidenciado na Figura 4.12.

Analisando a Figura 4.12 para as variáveis nos 30 primeiros intervalos, nota-se que existe uma estabilização da correlação para as 30 primeiras janelas. Portanto, a RNA desta dissertação foi implementada com valores de atrasos TDNN das entradas de até 30, pois existe uma queda na correlação dos dados após 30 atrasos (vide Figura 4.11) e, além disso, existe um número limitado de dados restritivo ao crescimento do número de entradas da rede. Testes posteriores serão feitos de modo a selecionar o melhor número possível de atrasos TDNN dentro deste intervalo [1,30]. Além disso, a variável ENA e IE

Figura 4.11 – Correlação das variáveis com o CMO para janelamentos de até 730 dados.

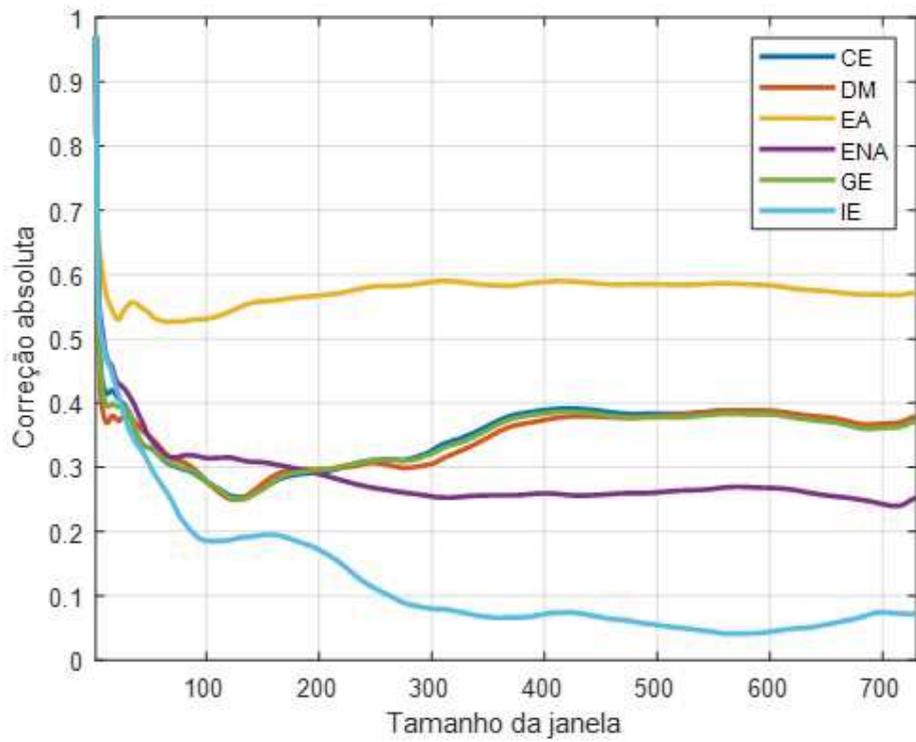
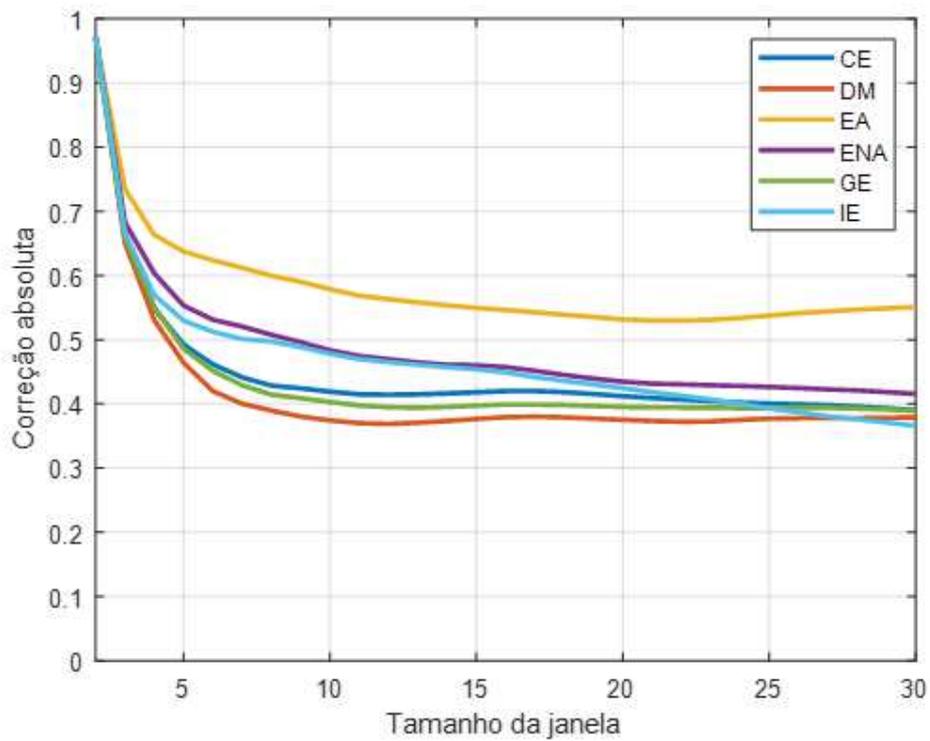


Figura 4.12 – Correlação das variáveis com o CMO para janelamentos de até 30 dados.



também poderão ser consideradas como novas entradas da Rede Neural caso se provem úteis na predição, pois, para os primeiros atrasos elas possuem um comportamento melhor que as outras variáveis, e apenas no longo prazo (janelas grandes) é que perdem qualidade de correlação.

Nesta abordagem serão utilizadas séries de dados de diversas variáveis importantes relacionadas à composição do CMO e, conseqüentemente, do PLD. A predição será feita para vários passos no futuro a cada rodada da rede. Conforme citado na seção 4.1.2, a predição do dado futuro do tempo $t + 1$ resulta na saída y_{t+1} cujo resultado verdadeiro é d_{t+1} . Como a presente dissertação pretende analisar de forma investigativa da variável CMO, a topologia desta seção realimentará a própria saída da rede, y_{t+1} , considerando que os dados verdadeiros d futuros não estão disponíveis. Contudo, adota-se como conhecidas as demais variáveis nos tempos futuros, em razão de não precisar predizê-las também. Desta forma, a predição do resultado do próximo tempo, $t + 2$ para as demais variáveis, utiliza como dado realimentado o dado verdadeiro $t + 1$, uma vez que este está disponível nesta abordagem.

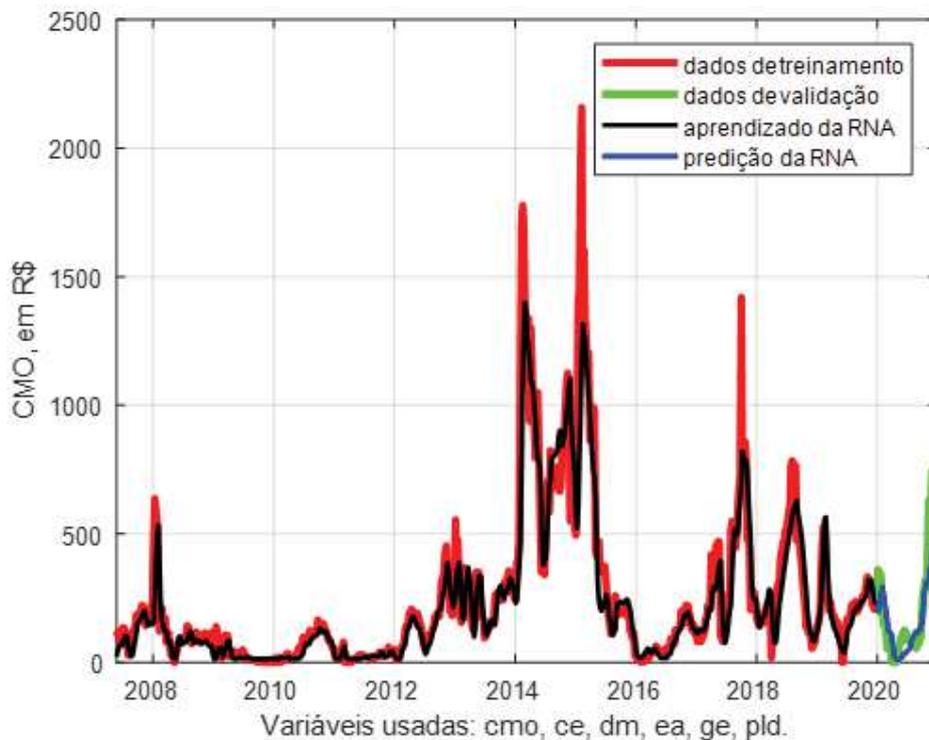
Inicialmente, foram simuladas RNAs tendo como entradas as séries TDNN das variáveis CE, DM, EA, GE, além do próprio CMO e PLD. Ou seja, optou-se por excluir as variáveis que apresentaram pior correlação com o CMO, a saber, ENA e IE. As topologias testadas foram as mesmas analisadas na abordagem anterior, da Seção 4.1, utilizando apenas o CMO como entrada da rede. As mesmas redes foram utilizadas em função de seus parâmetros fixados na seção inicial deste capítulo. Os resultados estão dispostos na Tabela 4.3.

Tabela 4.3 – Resultados por topologia (CE, DM, EA, GE, CMO e PLD) - vários passos no futuro (1 ano)

Configuração da rede				Treinamento			Validação		
N_1	N_2	N_3	n_p e n_q	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão
10	-	-	2	6,9364	2,6337	11,0624	7,8569	2,8030	11,0703
			10	7,5244	2,7431	11,2421	6,7167	2,5917	10,9589
			20	6,4877	2,5471	10,7635	5,3162	2,3057	10,3329
			30	8,0586	2,8388	11,2613	7,0889	2,6625	11,1130
20	20	-	2	4,7808	2,1865	10,0259	7,2070	2,6846	11,1655
			10	4,2412	2,0594	9,7147	6,4533	2,5403	10,9158
			20	5,5246	2,3504	10,4106	5,1836	2,2767	10,3339
			30	4,8630	2,2052	9,8928	7,2484	2,6923	11,1810
10	20	15	2	5,2949	2,3011	10,2931	6,9168	2,6300	11,1021
			10	6,5496	2,5592	10,8294	6,4946	2,5484	10,9327
			20	6,1860	2,4872	10,6643	5,5283	2,3512	10,4684
			30	6,6911	2,5867	10,7931	6,7386	2,5959	10,9707
50	50	50	2	6,2682	2,5036	10,5066	5,7029	2,3881	10,4278
			10	6,2168	2,4934	10,7003	5,8604	2,4208	10,6538
			20	7,1964	2,6826	11,0854	4,9575	2,2265	10,1728
			30	8,6566	3,1075	11,9237	6,9481	2,6359	11,0824

A topologia que apresentou melhor resultado quantitativo de predição é a composta por 3 camadas escondidas com 50 neurônios cada. Embora não seja a configuração com os menores erros de treinamento, nota-se nas Figuras 4.13 e 4.14, as quais apresentam graficamente o desempenho da rede, que a mesma foi capaz de capturar a tendência geral da variação anual do CMO, mostrando os picos e vales nos preços, evidenciando quais são as épocas do ano melhores para eventuais contratos de compra e venda de energia.

Figura 4.13 – Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=20$.



Na sequência de melhores resultados, tem-se a rede com 2 camadas escondidas ($N_1=N_2=20$) e atrasos de entrada n_p e realimentação n_q iguais a 20. Esta rede produz um resultado qualitativo muito semelhante à topologia com melhor desempenho, conforme evidenciado na Figura 4.15. Desta maneira, é possível concluir que duas camadas escondidas possuem a capacidade de capturar as tendências da variável com uma menor complexidade computacional sem perda significativa na qualidade dos dados preditos.

É possível observar nos resultados apresentados na Tabela 4.3 que os menores erros para cada topologia, considerando o número de camadas e número de neurônios por camada, são obtidos quando os atrasos n_p e n_q equivalem a 20. Com menos neurônios a captura de padrões não é tão efetiva, e com mais neurônios a rede torna-se redundante, provocando uma má generalização (excessiva) dos dados. Conforme apresentado na Figura 4.12, manter o número de atrasos em até 30 dados produz resultados de correlação próximos para todas as variáveis. Sendo assim, a RNA foi simulada com todas as variáveis como entrada da rede, com os resultados produzidos dispostos na Tabela 4.4.

Figura 4.14 – Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=20$.

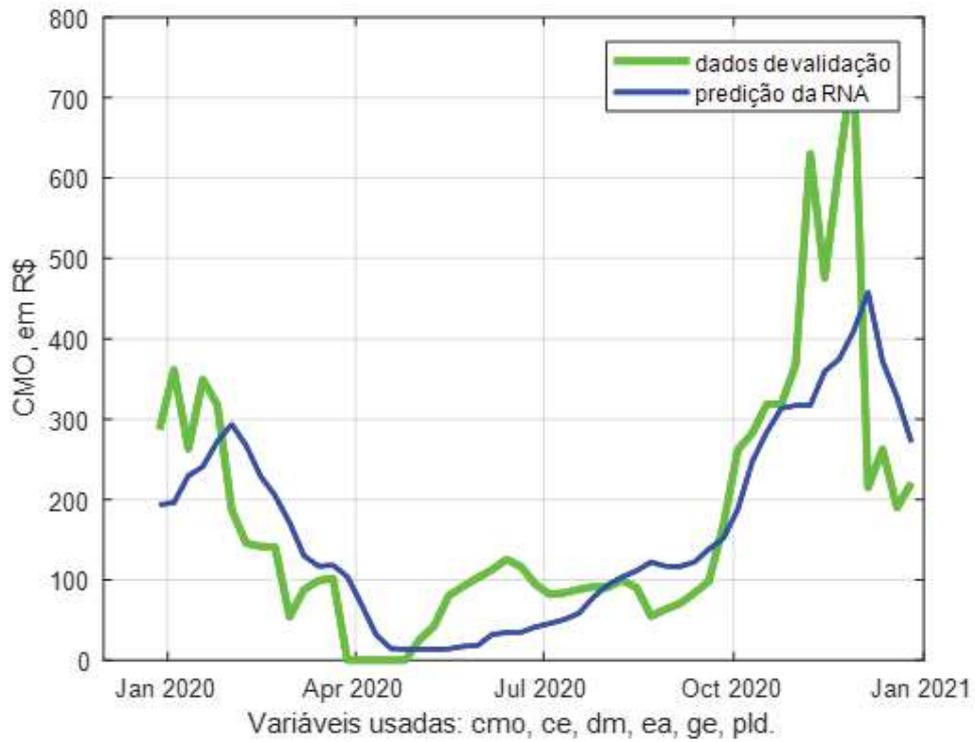


Figura 4.15 – Predição para $N_1=N_2=20$ e $n_p=n_q=20$.

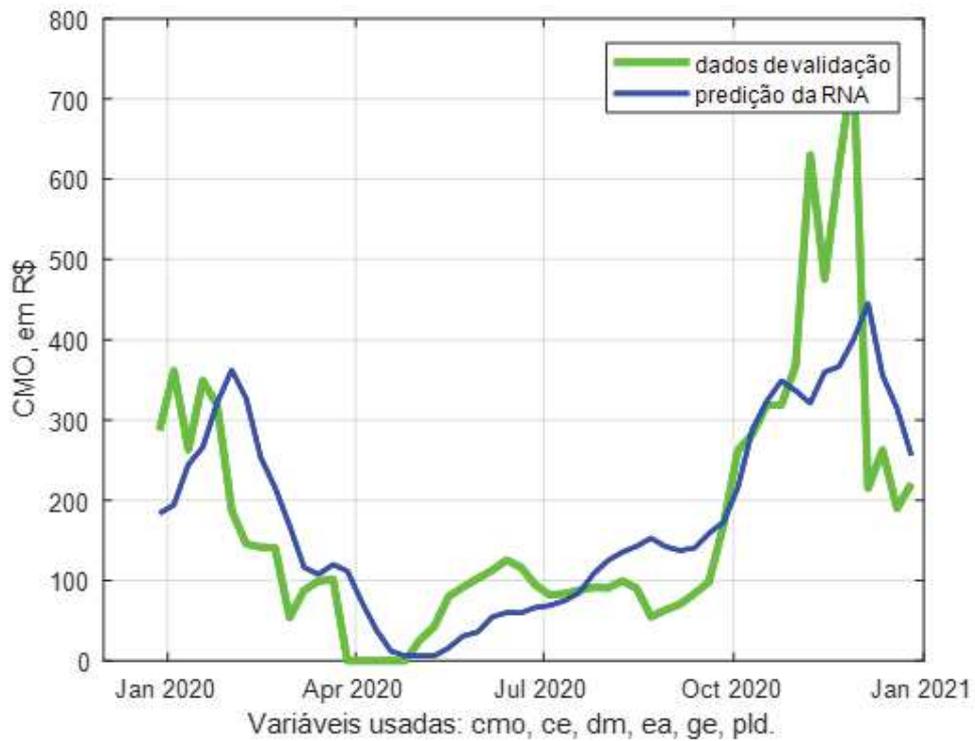


Tabela 4.4 – Resultados por topologia (CE, DM, EA, GE, IE, ENA, CMO e PLD) - vários passos no futuro (1 ano)

Configurações da rede				Treinamento			Validação		
N_1	N_2	N_3	n_p e n_q	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão
10	x	x	2	6,7917	2,6061	11,0053	8,4328	2,9039	10,5715
			10	9,2775	3,0459	11,8477	5,1350	2,2661	10,2821
			20	5,5850	2,3633	10,3906	6,9393	2,6343	11,0963
			30	15,4887	3,9356	13,5273	10,9649	3,3113	12,4565
20	20	x	2	5,2171	2,2841	10,2585	6,1462	2,4791	10,2577
			10	6,9043	2,6276	10,9604	5,0900	2,2561	10,2322
			20	3,4488	1,8571	9,2672	6,4169	2,5322	10,8208
			30	5,9358	2,4364	10,6243	10,6884	3,2693	12,2286
10	20	15	2	4,5675	2,1372	9,9294	5,8822	2,4253	10,6371
			10	4,9867	2,2331	9,9990	5,3268	2,3080	10,3976
			20	4,5105	2,1238	9,8674	6,9800	2,6420	11,0615
			30	6,1091	2,4717	10,7200	10,1956	3,1931	12,1827
50	50	50	2	4,4300	2,1047	9,7030	4,3991	2,0947	9,8723
			10	6,2309	2,4962	10,6137	4,8588	2,2043	10,1557
			20	10,6026	3,2562	12,3035	7,3011	2,7021	11,2558
			30	46,4712	6,8170	17,3084	13,4194	3,6632	13,0911

As duas redes que produziram os melhores resultados também são compostas por três camadas escondidas iguais de 50 neurônios cada. Nesta abordagem, o número de atrasos TDNN das entradas é configurado como 2 e 10, respectivamente. As Figuras 4.16 e 4.17 apresentam os resultados das melhores topologias.

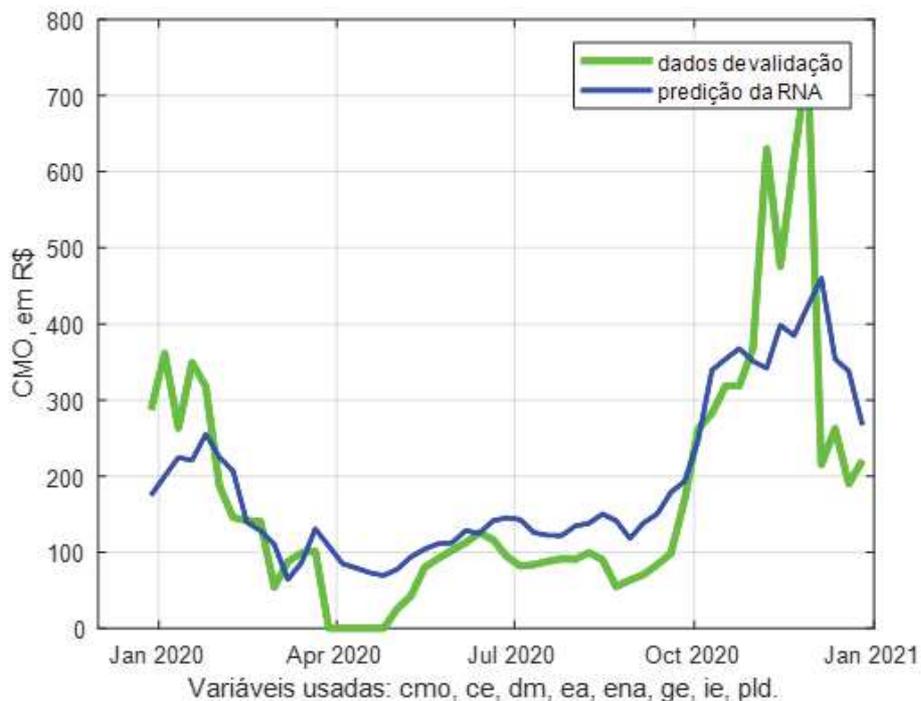
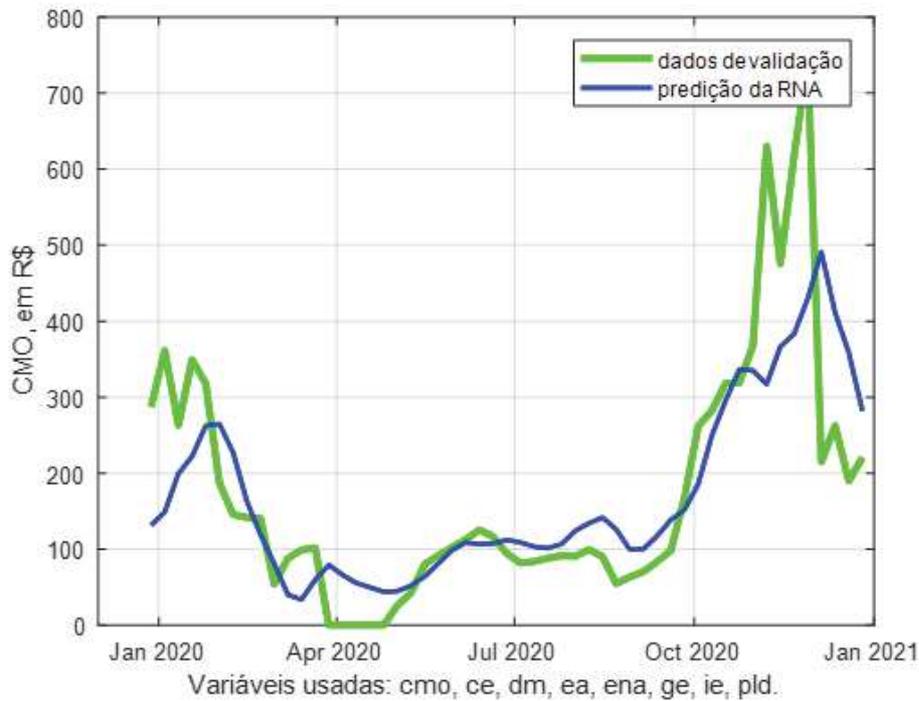
Figura 4.16 – Treinamento e Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=2$.

Figura 4.17 – Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=10$.



Verificou-se que os melhores resultados em cada conjunto de topologia foram obtidos utilizando todas as variáveis disponíveis (vide Figuras 4.11 e 4.12) como dados de entrada da rede, com atrasos n_p e n_q equivalentes a 10. É possível notar no gráfico apresentado pela Figura 4.12 que a partir do tamanho de janela 15, as variáveis IE e ENA têm um declínio na correlação com o CMO. O que justifica os erros de predição apresentarem os piores resultados quando utiliza-se os atrasos equivalentes a 30, independente do número de camadas e neurônios por camada.

Analisando de forma quantitativa os melhores resultados produzidos, tendo em vista a profundidade da rede, ou seja, o número de camadas escondidas, é possível notar que uma única camada é capaz de mapear a função o que vai de encontro ao Teorema da Aproximação Universal, derivado do Teorema de Andrey Kolmogorov (1957), que estabelece a capacidade de aproximação universal de funções para Redes Neurais Artificiais com apenas uma camada escondida e um número finito de neurônios (SILVA, 2010).

No entanto, para o mapeamento dos dados tratados nesta dissertação, o nível de abstração dos dados é grande, e assim os melhores resultados foram atingidos por intermédio de uma topologia com duas e três camadas escondidas. Conforme já citado, os dados são contaminados por ruídos, o que eleva a complexidade do problema, exigindo uma rede com mais capacidade de abstração. Foram realizados testes empíricos com redes com o número de camadas superior a três. Contudo estas não produziram bons resultados, por isso foram omitidas do presente trabalho.

Complementarmente, comparando os resultados produzidos pela rede utilizando ou não as variáveis IE e ENA nota-se que elas possuem desempenho muito semelhante, conforme esperado. Uma vez que a correlação das variáveis com o CMO são muito próximas quando os janelamentos são limitados em até 30. A Tabela 4.5 mostra os melhores resultados de cada topologia, comparando a rede utilizando todas as variáveis como dados de entrada e excluindo as variáveis IE e ENA.

Tabela 4.5 – Comparação dos resultados utilizando ou não as variáveis IE e ENA

Dados de Entrada	N_1	N_2	N_3	n_p e n_q	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Diferença RMSE %
CE, DM, EA, GE, CMO e PLD	10	-	-	20	5,3162	2,3057	10,3329	3,41%
CE, DM, EA, GE, IE, ENA, CMO e PLD				10	5,1350	2,2661	10,2821	
CE, DM, EA, GE, CMO e PLD	20	20	-	20	5,1836	2,2767	10,3339	1,81%
CE, DM, EA, GE, IE, ENA, CMO e PLD				10	5,0900	2,2561	10,2322	
CE, DM, EA, GE, CMO e PLD	10	20	15	20	5,5283	2,3512	10,4684	3,64%
CE, DM, EA, GE, IE, ENA, CMO e PLD				10	5,3268	2,3080	10,3976	
CE, DM, EA, GE, CMO e PLD	50	50	50	20	4,9575	2,2265	10,1728	11,26%
CE, DM, EA, GE, IE, ENA, CMO e PLD				2	4,3991	2,0947	9,8723	

De maneira qualitativa, comparando as Figuras 4.14 e 4.15 com as Figuras 4.16 e 4.17 é possível verificar que a rede utilizando todas as entradas se destaca frente à rede que não utiliza as variáveis IE e ENA, sendo capaz de generalizar a tendência de dados de maneira mais fiel. A adição de informações de IE e ENA é benéfica.

4.3 Filtragem da saída da Rede Neural

Esta última abordagem objetiva analisar o desempenho do uso combinado das Redes Neurais com a aplicação de filtros em sua saída. Conforme mostrado nas seções anteriores, o comportamento dos dados preditos podem ter embutidos componentes com altas variações que, ao serem realimentadas podem amplificar os efeitos dos erros. Ao realimentar uma saída filtrada, pretende-se diminuir estas variações na predição, tornando-a mais suave.

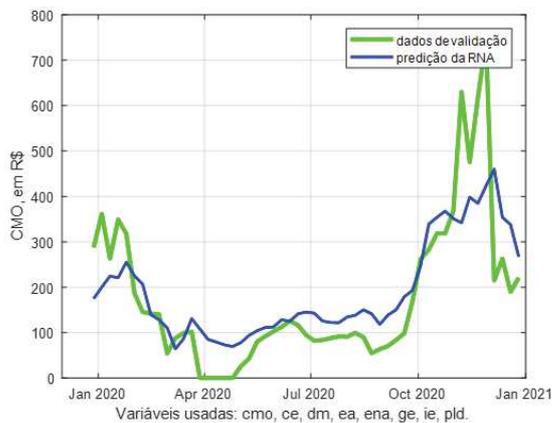
Serão testadas as mesmas topologias de rede analisadas na seção 4.2 considerando todas as variáveis na entrada da rede e incorporando 4 tipos de filtragem na saída da RNA: Passa-baixa FIR, Passa-baixa IIR, Média-Móvel e Filtro de Kalman. Os parâmetros dos filtros foram configurados empiricamente, de modo que os que estão sendo utilizados são os que produziram os melhores resultados.

Ao aplicar os diferentes filtros na saída da rede, obtiveram-se resultados superiores aos encontrados na seção 4.2. Sendo que os melhores e piores resultados são produzidos pelas mesmas topologias encontradas anteriormente. Entretanto, com a aplicação da filtragem o erro quantitativo produzido pela rede é menor ao produzido sem a filtragem. A Tabela 4.6 apresenta os resultados obtidos considerando todos os filtros aplicados e com $n_p = n_q = 2$ e 10. Vale ressaltar que a rede foi testada com 20 e 30 atrasos TDNN nos dados das entradas. Entretanto, os piores resultados foram produzidos pela rede com 30 atrasos, seguido da rede configurada com 20 atrasos.

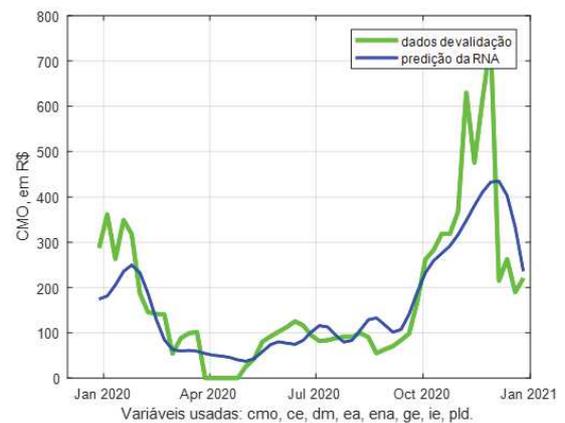
Em relação ao número de atrasos da rede, de forma similar à rede sem filtro, os melhores resultados são produzidos com redes de 10 atrasos. No que diz respeito à profundidade da rede e número de neurônios, tem-se uma melhor eficácia na rede com três camadas escondidas com 50 neurônios cada, seguido da rede com duas camadas com 20 neurônios. A diferença percentual entre o erro quadrático médio produzido pelo melhor resultado da rede com três e duas camadas é de 8,16%. Sendo que a diferença é pouco significativa frente à elevação do desempenho computacional com a inserção de mais camadas. A Figura 4.18 apresenta de forma qualitativa uma comparação dos resultados obtidos com a rede filtrada e sem a aplicação do filtro para o melhor resultado obtido.

Figura 4.18 – Predição para $N_1=N_2=N_3=50$ e $n_p=n_q=2$.

(a) Rede sem filtro.



(b) Rede com filtro Passa Baixa IIR.



Nota-se que as predições foram melhores com a aplicação do filtro na saída da rede, como dado pelo erro quadrático médio inferior. A predição é mais suave e representa melhor a dinâmica dos dados. O erro médio da predição reduziu 13,34% e a correlação entre os valores preditos e reais aumentou significativamente. Desta forma, o processo de uso da filtragem na saída da rede colaborou para a redução do erro nas predições.

Com relação aos filtros, nota-se um desempenho superior para o filtro de média-móvel, que apresentou melhores resultados em 75% topologias testadas. A Figura 4.19

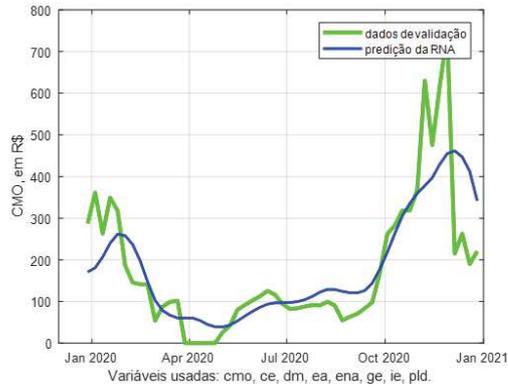
Tabela 4.6 – Resultados por topologia (CE, DM, EA, GE, IE, ENA, CMO e PLD) - vários passos no futuro com filtragem (1 ano)

Configuração da rede					Treinamento			Validação		
N_1	N_2	N_3	n_p e n_q	Filtro	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão	Erro Quadrático Médio (R\$) ²	Erro Médio (R\$)	Desvio Padrão
10	-	-	2	Sem Filtro	6,7917	2,6061	11,0053	8,4328	2,9039	10,5715
				P-B FIR	6,2909	2,5082	10,7968	7,6942	2,7738	10,3729
				P-B IIR	5,0955	2,2574	10,2139	7,4600	2,7313	10,3293
				Média Móvel	7,5359	2,7452	11,2965	7,2075	2,6847	10,1051
				Kalman	4,5049	2,1225	9,9333	8,4584	2,9083	10,6416
			10	Sem Filtro	9,2775	3,0459	11,8477	5,1350	2,2661	10,2821
				P-B FIR	9,0186	3,0031	11,7664	4,8956	2,2126	10,1555
				P-B IIR	9,1560	3,0259	11,7774	4,6579	2,1582	10,0604
				Média Móvel	6,9835	2,6426	11,0223	4,2809	2,0690	9,8197
				Kalman	4,5590	2,1352	9,9038	4,9676	2,2288	10,2246
20	20	-	2	Sem Filtro	5,2171	2,2841	10,2585	6,1462	2,4791	10,2577
				P-B FIR	4,3604	2,0882	9,7894	6,0708	2,4639	10,0896
				P-B IIR	3,3368	1,8267	9,0095	4,1445	2,0358	9,5326
				Média Móvel	3,8963	1,9739	9,5053	5,2708	2,2958	9,7617
				Kalman	4,5147	2,1248	9,9236	6,2962	2,5092	10,4085
			10	Sem Filtro	6,9043	2,6276	10,9604	5,0900	2,2561	10,2322
				P-B FIR	6,1209	2,4741	10,5998	4,7152	2,1715	10,0532
				P-B IIR	3,4973	3,0818	11,8541	4,6842	2,1643	10,0410
				Média Móvel	6,3512	2,5202	10,7533	4,1504	2,0372	9,7736
				Kalman	4,3554	2,0870	9,8024	4,9859	2,2329	10,2162
10	20	15	2	Sem Filtro	4,5675	2,1372	9,9294	5,8822	2,4253	10,6371
				P-B FIR	3,7822	1,9448	9,4188	5,1149	2,2616	10,2946
				P-B IIR	4,5714	2,1381	9,8299	4,3735	2,0913	9,8797
				Média Móvel	4,2973	2,0730	9,7550	4,6559	2,1578	10,0492
				Kalman	4,2156	2,0532	9,7392	5,9377	2,4367	10,6790
			10	Sem Filtro	4,9867	2,2331	9,9990	5,3268	2,3080	10,3976
				P-B FIR	7,3927	2,7190	11,1379	5,0175	2,2400	10,2479
				P-B IIR	7,4814	2,7353	11,1723	4,8768	2,2083	10,1494
				Média Móvel	7,5211	2,7425	11,1720	4,4923	2,1195	9,9706
				Kalman	4,4690	2,1140	9,8516	5,2062	2,2817	10,3293
50	50	50	2	Sem Filtro	4,4300	2,1047	9,7030	4,3991	2,0947	9,8723
				P-B FIR	3,7646	1,9402	9,2286	4,2218	2,0547	9,7844
				P-B IIR	5,1286	2,2647	9,8733	3,8121	1,9524	9,5126
				Média Móvel	3,8103	1,9520	9,2946	3,9635	1,9908	9,6343
				Kalman	4,7633	2,1825	9,9462	4,5539	2,1340	9,9902
			10	Sem Filtro	6,2309	2,4962	10,6137	4,8588	2,2043	10,1557
				P-B FIR	5,3023	2,3027	10,2408	4,5492	2,1329	9,9849
				P-B IIR	7,5574	2,7491	11,2259	4,4225	2,1029	9,9271
				Média Móvel	5,7025	2,3880	10,3990	3,9578	1,9894	9,6507
				Kalman	4,5997	2,1447	9,9129	4,8790	2,2088	10,1520

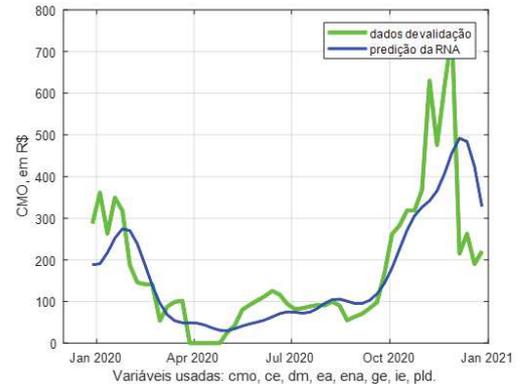
apresenta uma comparação da predição produzida pela rede configurada com $N_1=N_2=20$ e $n_p=n_q=10$.

Figura 4.19 – Predição para $N_1=N_2=20$ e $n_p=n_q=10$.

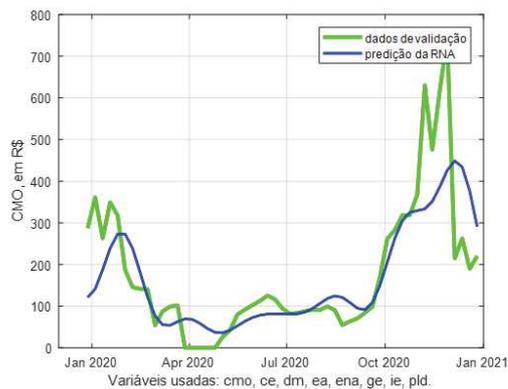
(a) Filtro Média Móvel.



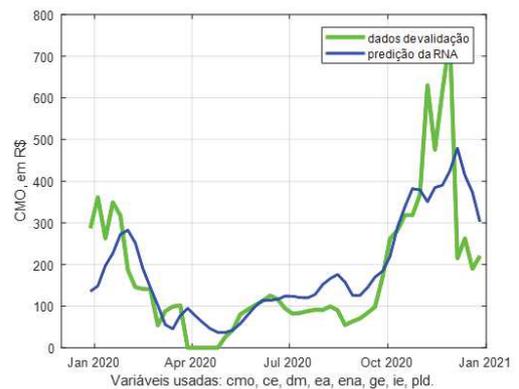
(b) Passa-Baixa IIR.



(c) Passa-Baixa FIR.



(d) Filtro de Kalman.



4.4 Conclusões sobre as análises dos resultados

A análise dos erros foi realizada de forma a investigar a influência de determinadas variáveis na predição do CMO, bem como avaliar o desempenho da rede recorrente com a aplicação de filtros em sua saída para a atenuação dos ruídos inerentes aos dados. A medida de acuracidade RMSE foi aplicada para que fosse possível mensurar as diferenças entre os dados preditos e os dados originais.

Dentre as abordagens analisadas, a análise de correlação das variáveis dos dados de entrada da rede com a variável a ser predita mostrou-se eficiente e acertada, permitindo um direcionamento na escolha dos melhores atrasos TDNN para a RNA. A inclusão de mais variáveis como dados de entrada melhorou o desempenho de predição.

Ressalta-se também que a inserção do filtro corretivo na saída da rede permitiu que a rede predisse valores do CMO com maior acurácia. Embora houvessem diferenças

entre os filtros aplicados, nota-se que as predições proporcionaram um ganho de eficiência nas predições, reduzindo as variações de alta frequência, minimizando a volatilidade dos dados e produzindo uma saída que acompanha eficientemente as tendências dos dados, proporcionando uma boa ferramenta de análise de longo prazo.

5 CONCLUSÕES

Este trabalho mostrou uma abordagem baseada em Redes Neurais Artificiais aliada a filtragem estocástica para a obtenção da predição de uma variável importante no setor elétrico brasileiro, o Custo Marginal de Operação (CMO), associado à formação do Preço de Liquidação das Diferenças (PLD). O objetivo foi investigar uma maneira de se obter a predição de médio-longo prazo, de 1 ano a frente, semana a semana, para oferecer aos analistas da área de compra e venda de energia elétrica uma ferramenta que mostre a principal tendência dos preços da energia, facilitando o processo decisório no fechamento de contratos.

Este trabalho contou com uma metodologia exploratória advinda de estratégias na literatura, bem como a utilização de técnicas extras, de filtragem, em concomitância com a Rede Neural proposta. Primeiramente, este trabalho, utilizando a literatura, definiu a abordagem de aprendizado a ser utilizada, uma Rede Neural Artificial. Em seguida, diversos testes exploraram diferentes métodos de treinamento, com a escolha final pelo método *ADAM* (P.KINGMA; BA, 2014), por se tratar de um método de treinamento estatístico, excelente para dados estocásticos, caso da variável de interesse, o CMO.

A primeira prática consistiu em comparar a Rede Neural escolhida com os resultados de (NASCIMENTO et al., 2019), que faz a predição do PLD semanalmente. Em ambos os trabalhos, concluiu-se que para predições de médio-longo prazo (vários passos no futuro), a série temporal CMO/PLD não é suficiente, uma vez que se comporta de maneira parecida com uma variável aleatória em termos de correlação cruzada, não possuindo estruturas periódicas e, portanto, não oferecendo informação suficiente de aprendizado para a Rede Neural.

Assim, buscou-se a adição de novas informações para as entradas da Rede Neural por meio de variáveis explanatórias do Operador Nacional do Sistema Elétrico (ONS). Assim, diversas variáveis foram selecionadas e comparadas com a variável de interesse por meio de correlação. Este teste foi importante para determinar quais variáveis seriam mais importantes e, principalmente, a quantidade de entradas TDNN seriam escolhidas, prezando sempre pela alta correlação entre as variáveis de entrada e a variável predita.

Assim, testes da Rede Neural mostraram que a rede consegue, com estas novas variáveis de entrada, aprender os padrões de comportamento da variável de interesse. No entanto, como o CMO é uma variável com alta volatilidade, estes mesmo comportamento, ao ser realimentado na Rede Neural Recorrente, acumulava erros com o passar do tempo, oferecendo uma saída ruidosa.

Desta forma, adicionou-se uma abordagem de filtragem das saídas de Rede Neural Recorrente, de maneira a mitigar o problema da alta volatilidade dos dados preditos, melhorando a predição da rede, obtendo, portanto o bom acompanhamento das tendências

da variável de interesse ao longo do ano. Exploraram-se os principais filtros disponíveis para a tarefa de eliminar contaminações e componentes de alta frequência do sinal. O resultado é uma predição suave, que acompanha a variável de interesse, oferecendo os principais momentos e comportamentos, auxiliando a tomada de decisões de compra e venda de energia.

5.1 Propostas de desenvolvimentos futuros

Seguindo a linha de pesquisa desenvolvida nesta dissertação e tendo em vista os resultados obtidos, algumas aplicações podem ser feitas com o intuito de aperfeiçoar o sistema apresentado. Os tópicos a seguir apresentam perspectivas no desenvolvimento de trabalhos futuros:

- i Explorar um modelo híbrido, onde as variáveis de entrada da rede também sejam preditas para o caso de n passos no futuro;
- ii Investigar a performance da rede implementada em outros mercados de energia, de forma a criar uma abordagem com maior capacidade de generalização;
- iii Realizar uma análise de sensibilidade de forma a avaliar o desempenho do modelo assumindo as variáveis de entrada não preditas como sendo as do último ano do histórico;
- iv Implementar a integração do sistema proposto com outras técnicas de inteligência artificial, como a lógica *fuzzy*, de tal maneira a capturar nuances dos dados, aumentando as informações disponíveis para a Rede Neural, tal qual diversas abordagens tem sido utilizadas nos mercados de ações (ZAVADZKI et al., 2020).

REFERÊNCIAS

- ACADEMY, D. S. *Deep Learning Book*. 2019. Disponível em <<http://www.deeplearningbook.com.br>>. Acessado em 09 mar. 2021.
- ANEEL, A. N. de E. E. *Mercado de Curto Prazo*. 2021. Disponível em <https://www.aneel.gov.br/home?p_p_id=101&p_p_lifecycle=0&p_p_state=maximized&p_p_mode=view&_101_struts_action=%2Fasset_publisher%2Fview_content&_101_returnToFullPageURL=%2F&_101_assetEntryId=15050759&_101_type=content&_101_groupId=656835&_101_urlTitle=mercado-de-curto-prazo-mcp&inheritRedirect=true>. Acessado em 25 fev. 2021.
- BRASIL. *Lei nº 9.074, de 7 de julho de 1995*. 1995. Disponível em <http://www.planalto.gov.br/ccivil_03/leis/19074cons.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Decreto 2.335, de 6 de outubro de 1997*. 1997. Disponível em <http://www.planalto.gov.br/ccivil_03/decreto/d2335.HTM>. Acessado em 28 fev. 2021.
- BRASIL. *Lei 9.478, de 6 de agosto de 1997*. 1997. Disponível em <http://www.planalto.gov.br/ccivil_03/leis/19478.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Medida Provisória 144, de 11 de dezembro de 2003*. 2003. Disponível em <http://www.planalto.gov.br/ccivil_03/mpv/Antigas_2003/144.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Medida Provisória 145, de 11 de dezembro de 2003*. 2003. Disponível em <http://www.planalto.gov.br/ccivil_03/mpv/Antigas_2003/145.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Decreto 5.081, de 14 de maio de 2004*. 2004. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/decreto/d5081.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Decreto 5.163, de 30 de julho de 2004*. 2004. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/decreto/d5163.HTM>. Acessado em 28 fev. 2021.
- BRASIL. *Decreto 5.175, de 9 de agosto de 2004*. 2004. Disponível em <http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/Decreto/D5175.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Lei 10.847, de 15 de março de 2004*. 2004. Disponível em <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.847.htm>. Acessado em 28 fev. 2021.
- BRASIL. *Lei 10.848, de 15 de março de 2004*. 2004. Disponível em <http://www.planalto.gov.br/ccivil_03/_Ato2004-2006/2004/Lei/L10.848.htm>. Acessado em 28 fev. 2021.
- CCEE, C. de Comercialização de E. E. *Com quem relaciona*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/pages_publico/onde-atuamos/com_quem_se_relaciona?_afLoop=100800964117829&_adf.ctrl-state=130o6afo7p_1#!%40%40%3F_>

afrLoop%3D100800964117829%26_adf.ctrl-state%3D130o6afo7p_5>. Acessado em 28 fev. 2021.

CCEE, C. de Comercialização de E. E. *Deck de Preços*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/pages_publico/o-que-fazemos/como_ccee_atua/precos/deck_de_precos?_afrLoop=23998935792650&_adf.ctrl-state=1iiq7eugc_14#!%40%40%3F_afrLoop%3D23998935792650%26_adf.ctrl-state%3D1iiq7eugc_18>. Acessado em 26 fev. 2021.

CCEE, C. de Comercialização de E. E. *Metodologia*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/pages_publico/o-que-fazemos/como_ccee_atua/precos/metodologia_de_precos?_afrLoop=104858829549211&_adf.ctrl-state=19bttq07jq_1#!%40%40%3F_afrLoop%3D104858829549211%26_adf.ctrl-state%3D19bttq07jq_5>. Acessado em 02 mar. 2021.

CCEE, C. de Comercialização de E. E. *Preços*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/oquefazemos_menu_lateral/precos?_afrLoop=208972491694133&_adf.ctrl-state=yklide4en_1#!%40%40%3F_afrLoop%3D208972491694133%26_adf.ctrl-state%3Dyklide4en_5>. Acessado em 26 fev. 2021.

CCEE, C. de Comercialização de E. E. *Quem participa*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/pages_publico/quem-participa?_adf.ctrl-state=6dms8ozk5_18&_afrLoop=26108519444505#!>. Acessado em 27 fev. 2021.

CCEE, C. de Comercialização de E. E. *Setor Elétrico*. 2021. Disponível em <https://www.ccee.org.br/portal/faces/pages_publico/onde-atuamos/setor_eletrico?_adf.ctrl-state=wdsdnrt8i_1&_afrLoop=125894233318809#!%40%40%3F_afrLoop%3D125894233318809%26_adf.ctrl-state%3Dwdsdnrt8i_5>. Acessado em 25 fev. 2021.

COELHO, I. *Estudo sobre a aplicação de Redes Neurais Recorrentes para previsão da geração eólica e do Preço de Liquidação das Diferenças*. [S.l.]: Monografia, UFSC, 2018.

DAMETTO, R. C. *Estudo da aplicação de redes neurais artificiais para predição de séries temporais financeiras*. [S.l.]: Dissertação, Unesp, 2018.

DUDA, R. O.; HART, P. E.; STORK, D. G. *Pattern Classification*. 2. ed. [S.l.]: Wiley-Interscience, 2000. v. 63.

FILHO, J. C. R. *Previsão de multi-passos a frente do preço de energia elétrica de curto prazo no mercado brasileiro*. Tese (Doutorado) — Universidade Federal do Pará, 2014.

HAGAN, M. T.; MENHAJ, M. B. *Training feedforward networks with the Marquardt algorithm*. *IEEE transactions on Neural Networks*, v. 5, n. 6, p. 989-993, 1994.

HAYKIN, S. *Adaptive Filter Theory*. [S.l.]: 2nd ed. Englewood Cliffs, N.J.: Prentice Hall, 1996.

HEBB, D. O. *The Organization Of Behavior A Neuropsychological Theory*. [S.l.]: John Wiley And Sons Inc New York, Chapman And Hall Limited London, 1949.

HOPFIELD, J. J. *Neural networks and physical systems with emergent collective computational abilities*. *Proceedings of the national academy of sciences*, v. 79, n. 8, p. 2554-2558, 1982.

- KAILATH, T. *Linear Systems*. [S.l.]: Englewood Cliffs, NJ: Prentice Hall, 1980.
- KALMAN, R. E. *A New Approach to Linear Filtering and Prediction Problems*. *Transaction of the ASME – Journal of Basic Engineering*, pp. 35-45, 1960.
- LEITE, A. L. da S. *Causas da volatilidade do preço spot de eletricidade no Brasil*. [S.l.]: Ensaio FEE, v. 34, p. 647-668-668, 2013.
- MATHWORKS. *Matlab*. 2021. Disponível em <<https://www.mathworks.com/products/matlab.html>>. Acessado em 02 mar. 2021.
- MCCULLOCH, W. S.; PITTS, W. *A logical calculus of the ideas immanent in nervous activity*. *The bulletin of mathematical biophysics*. v. 5, n. 4, p. 115-133, 1943.
- MEDIUM. *Human in a Machine World*. 2016. Disponível em <<https://medium.com/human-in-a-machine-world/mae-and-rmse-which-metric-is-better-e60ac3bde13d>>. Acessado em 13 mar. 2021.
- MINSKY, M. L.; PAPERT, S. A. *Perceptrons: an introduction to computational geometry*. [S.l.]: MIT Press, Cambridge, Massachusetts, 1969.
- NASCIMENTO, F. B. do. *Redes Neurais Artificiais aplicadas à predição do Preço de Liquidação das Diferenças no Mercado de Energia*. [S.l.]: Monografia, UFJF, 2017.
- NASCIMENTO, F. B. do; OLIVI, L. R.; LIMA, L. H. L. *Multi-step forecasting of the Brazilian Energy Market Price using Autoregressive Neural Networks*. *VII Simpósio Brasileiro de Sistemas Elétricos (SBSE2018)*, Niterói-RJ, 2018.
- NASCIMENTO, F. B. do et al. *Weekly forecast of the Settlement Price of Differences in the Brazilian short-term market through recurrent artificial neural networks*. *14º Simposio Brasileiro de Automação Inteligente, Ouro Preto*, 2019.
- ONS, O. N. do S. E. *Histórico da Operação*. 2020. Disponível em <<http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao>>. Acessado em 08 jun. 2020.
- ONS, O. N. do S. E. *Carga de Energia*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/carga_energia.aspx>. Acessado em 07 abr. 2021.
- ONS, O. N. do S. E. *Custo Marginal de Operação*. 2021. Disponível em <<http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/cmo.aspx>>. Acessado em 07 abr. 2021.
- ONS, O. N. do S. E. *Demanda Máxima*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/demanda_maxima.aspx>. Acessado em 07 abr. 2021.
- ONS, O. N. do S. E. *Energia Armazenada*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/energia_armazenada.aspx>. Acessado em 07 abr. 2021.
- ONS, O. N. do S. E. *Energia Natural Afluenta*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/energia_afluenta_subsistema.aspx>. Acessado em 07 abr. 2021.

ONS, O. N. do S. E. *Geração de Energia*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/geracao_energia.aspx>. Acessado em 07 abr. 2021.

ONS, O. N. do S. E. *Intercâmbios de Energia*. 2021. Disponível em <http://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/intercambios_energia.aspx>. Acessado em 07 abr. 2021.

P.KINGMA, D.; BA, J. *Adam: A method for stochastic optimization*. *arXiv preprint arXiv:1412.6980*, 2014.

RIEDMILLER, M.; BRAUN, H. *A direct adaptive method for faster backpropagation learning: The RPROP algorithm*. *IEEE international conference on neural networks*. p. 586-591., 1993.

RODRIGUES, A. L. *Redes Neurais Artificiais aplicadas na previsão de preços do mercado spot de energia elétrica*. [S.l.]: Dissertação, USP, 2009.

ROSENBLATT, F. *Principles of neurodynamics. perceptrons and the theory of brain mechanisms*. [S.l.]: Cornell Aeronautical Lab Inc Buffalo NY, 1961.

RUMELHART, G. E. H. D. E.; WILLIAMS, R. J. *Parallel Distributed Processing*. [S.l.]: MT Press, Vol.1, 1986.

SILVA, D. H. S. e. R. A. F. Ivan Nunes da. *Redes Neurais Artificiais: para engenharia e ciências aplicadas*. [S.l.]: Editora Artliber, 2010.

SMITH, S. W. *The Scientist and Engineer's Guide to Digital Signal Processing*. [S.l.]: 2nd ed. California Technical Publishing San Diego, California, 1999.

SOUZA, A. J. S. de. *Análise e Previsão da volatilidade do preço de liquidação das diferenças no mercado brasileiro utilizando o modelo GARCH*. [S.l.]: Dissertação, UFBA, 2013.

SOUZA, F. C. *Dinâmica Da Gestão De Riscos No Ambiente De Contratação Regulada No Setor Elétrico Brasileiro*. Tese (Doutorado) — Universidade Federal Do Rio De Janeiro, 2008.

WIDROW, B. *Generalization and information storage in network of Adaline 'neurons'. Self-organizing systems-1962*, 1962.

ZAVADZKI, S. et al. *Computational Intelligence Techniques Used for Stock Market Prediction: A Systematic Review*. *IEEE Latin America Transactions*, v. 18, n. 04, p. 744–755, 2020.