

**UNIVERSIDADE NOVE DE JULHO – UNINOVE**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**

**FRANCISCO ELÂNIO BEZERRA**

**ABORDAGEM PARA SELEÇÃO DE ATRIBUTOS E DETERMINAÇÃO DE**  
**ATRASOS DE TEMPO EM MODELOS AUTORREGRESSIVOS DE PREVISÃO**  
**BASEADOS EM REDES NEURAIS ARTIFICIAIS**

**São Paulo**  
**2020**

**FRANCISCO ELÂNIO BEZERRA**

**ABORDAGEM PARA SELEÇÃO DE ATRIBUTOS E DETERMINAÇÃO DE  
ATRASOS DE TEMPO EM MODELOS AUTORREGRESSIVOS DE PREVISÃO  
BASEADOS EM REDES NEURAIAS ARTIFICIAIS**

Tese apresentada à Universidade Nove de  
Julho – UNINOVE para obtenção do título  
de Doutor em Engenharia de Produção.

Prof. Fabio Henrique Pereira, Dr. -  
Orientador

**São Paulo  
2020**

**PARECER DA COMISSÃO EXAMINADORA DE DEFESA DE**  
**TESE DE**

Francisco Elânio Bezerra

Título da Tese: Abordagem para Seleção de Atributos e Determinação de Atrasos de Tempo em Modelos Autorregressivos de Previsão Baseados em Redes Neurais Artificiais.

A Comissão examinadora, composta pelos professores abaixo, considera o(a) candidato(a) Francisco Elânio Bezerra Aprovado.

São Paulo, 27 de novembro de 2020.

Prof(a). Dr(a).Fabio Henrique Pereira (UNINOVE / PPGEP) - orientador



---

Prof(a). Dr(a).Julio Carlos Teixeira (UFABC / PPGENE) – Membro Externo



---

Prof(a). Dr(a).Silvio Giuseppe Di Santo (USP / PEA) - Membro Externo



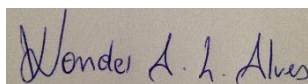
---

Prof(a). Dr(a).Cleber Gustavo Dias (UNINOVE / PPGI) - Membro Externo



---

Prof(a). Dr(a).Wonder Alexandre Luz Alves (UNINOVE / PPGI) - Membro Externo



Bezerra, Francisco Elânio.

Abordagem para seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão baseados em redes neurais artificiais. / Francisco Elânio Bezerra. 2020.

217 f.

Tese (Doutorado) – Universidade Nove de Julho - UNINOVE, São Paulo, 2020

Orientador (a): Prof. Dr. Fabio Henrique Pereira.

1. Modelo autorregressivo. 2. Rede neural multicamadas. 3. Seleção de atributos. 4. PCA. 5. Transformada wavelet.

I. Pereira, Fabio Henrique. II. Título

CDU 658.5

Dedico este trabalho para minha família, meu pai Wilson, minha mãe Iranda, minha esposa Kelly e meus irmãos Marcelo e Camila.

## AGRADECIMENTOS

Primeiramente, quero agradecer a Deus por diversas oportunidades de conhecer e aprender inúmeras coisas que jamais pensei que teria oportunidade de conhecer e participar. Também quero agradecer a oportunidade de enxergar que existe um mundo lá “fora”, ao qual podemos desbravar.

Agradeço a toda minha família pelo apoio durante toda minha vida acadêmica, em especial, à minha mãe que, na sua simplicidade, me aconselha, encoraja, apoia, acredita nos meus sonhos e comemora a cada conquista. Muito obrigado por tudo Dona Iranda.

Quero agradecer meu pai, que sem saber, me ajudava a descontrair um pouco jogando bilhar e “batendo papo” quando estava tentando escrever algum parágrafo e não conseguia.

Também quero agradecer a minha esposa, que me acompanha desde o início desta jornada. Muito obrigado pela sua paciência, apoio e compreensão.

Agradeço ao Professor Fábio pela paciência, orientações e ajuda no desenvolvimento dos artigos e da tese. Espero que possamos publicar mais artigos no futuro.

Finalmente, agradeço a Universidade Nove de Julho pela bolsa e a oportunidade em desenvolver este trabalho.

## EPÍGRAFE

*“Deus é poderoso para fazer infinitamente mais do que pedimos ou pensamos.”*

*Efésios 3:20*

## RESUMO

Em modelos autorregressivos o valor atual de uma série temporal é regredido em valores passados da mesma série, e valores passados e presentes de uma ou mais variáveis exógenas, com diversas aplicações na previsão de valores futuros, tais como: previsão de eletricidade, monitoramento da condição de equipamentos, mercado financeiro, previsão do clima, entre outros. Apesar de existirem diversas abordagens na literatura propondo melhorar a precisão da previsão, há uma lacuna de pesquisa no que diz respeito a seleção do número de valores passados da série temporal de interesse e das variáveis exógenas a serem utilizadas no modelo. Além disso, as abordagens usuais na literatura adotam o mesmo valor de atraso para todas as variáveis envolvidas, o que pode não representar adequadamente a relação subjacente do processo em estudo. Assim, o objetivo desta tese é propor uma abordagem para seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão, baseados em redes neurais artificiais, análise de componentes principais e transformada *wavelet*. Para alcançar o objetivo proposto, diferentes abordagens foram testadas, desde efeito de atrasos de tempo na variável de resposta, efeito entre atrasos de tempo e taxa de amostragem, seleção das variáveis e aproximações mais relevantes e transformada *wavelet* mais adequada para construir aproximações para criar atrasos de tempo para um modelo autorregressivo de previsão. Os resultados mostraram que a seleção de atributos e o uso da transformada *wavelet* para criar atrasos de tempo, com uma versão esparsa e simplificada dos dados, melhora o entendimento entre as variáveis de entrada e saída e, conseqüentemente, a precisão da previsão em modelos autorregressivos. As abordagens utilizadas neste trabalho envolveram aplicações no setor elétrico de energia, tais como a análise de concentração de gás dissolvido em óleo em um transformador de potência, energia elétrica de uma distribuidora de eletricidade e energia elétrica de uma usina geradora de eletricidade. Os resultados destas abordagens resultaram na publicação em um congresso nacional e revistas científicas internacionais.

**Palavras-chave:** Modelo autorregressivo, Rede neural multicamadas, Seleção de atributos, PCA, Transformada *wavelet*.

## ABSTRACT

*In autoregressive models, the current value of time series is regressed to past values of the same series, and past and present values of one or more exogenous variables, with several applications in the prediction of future values, such as: electricity prediction, condition monitoring equipment, financial market, weather prediction, among others. Although, there are different approaches in the literature proposing improve the accuracy of prediction, there is a research gap regarding the selection of the number of past values of the time series of interest and exogenous variables used in the model. Moreover, the usual approaches in the literature adopt the same delay value for all the variables involved, which may not adequately represent the underlying relationship of the process under study. Thus, the objective of this thesis is to propose an approach for selecting attributes and determining time delays in autoregressive prediction models, based on artificial neural networks, principal component analysis and wavelet. To achieve the objective proposed in this thesis, different approaches were tested, from the effect of time delays on the response variable, the effect between time delays and sampling rate, selection of the most relevant variables and approximations created by type of wavelet transform for build approximations to create time delays for an autoregressive forecasting model. The results have showed the selection of attributes and the use of wavelet transform to create time delays, with a sparse and simplified version of the data, improves the understanding between the input and output variables and, consequently, the prediction accuracy in autoregressive models. These approaches have involved applications in the electric energy sector, such as, oil dissolved gas in power transformer, electric energy from an electricity distributor company, electric energy from an electricity power generator. The results of these approaches have resulted in publication in a national congress and international scientific journals.*

**Keywords:** *Autoregressive model, Time series, Attribute selection, Artificial neural network, Principal Component Analysis, Wavelet transform.*

## LISTA DE FIGURAS

Figura 1 – Representação de um modelo univariado .....	36
Figura 2 – Representação de um modelo multivariado .....	37
Figura 3 - Arquitetura da Rede Multicamadas .....	39
Figura 4 – Gráfico com 50 observações. a) Duas variáveis $X1$ , $X2$ . b) Duas componentes principais $Z1$ , $Z2$ das variáveis $X1$ , $X2$ .....	45
Figura 5 – A função wavelet de Haar para vários valores do parâmetro de escala $a$ e parâmetro de translação $b$ . Se $a < 1$ , a função wavelet é contraída, e se $a > 1$ , a wavelet é expandida.	48
Figura 6 – Diagrama em bloco da análise do filtro .....	52
Figura 7 – Banco de filtro de nível 3 .....	53
Figura 8 – Representação da DWT no domínio da frequência.....	53
Figura 9 – DWT de uma matriz em termos de filtros.....	54
Figura 10 – Representação da matriz $A = A0$ na transformada <i>wavelet</i> bidimensional.....	55
Figura 11 – Família de <i>Wavelet Daubechies</i> .....	56
Figura 12 – <i>Wavelets</i> : a) <i>Coif</i> e b) <i>Sym</i> .....	57
Figura 13 – Objetivo geral e específico das abordagens utilizadas na tese .....	61
Figura 14 – Abordagem para previsão de demanda de energia elétrica.....	62
Figura 15 – Estrutura de previsão de eletricidade das classes de consumo.....	64
Figura 16 – Diagrama de abordagem proposto .....	66
Figura 17 – Fluxograma do modelo autorregressivo de previsão não-linear proposto (NAR-DWT).....	67
Figura 18 – Modelo para seleção e taxa de contribuição da concentração e previsão de gases .....	68
Figura 19 – Abordagem aplicada nas concentrações de gás em ordem cronológica reversa ...	71
Figura 20 – Arquitetura da rede MLP proposta para previsão de cada concentração de gás ...	72
Figura 21 – Previsão de energia elétrica por classe de consumo a) Comercial, b) Industrial, c) Residencial .....	75
Figura 22 – Previsão de energia elétrica a) Instante $t$ , b) Variáveis atrasadas em $t-5$ , $t-10$ e $t-20$ .....	77
Figura 23 – Previsão de energia elétrica em GWh .....	79
Figura 24 – Melhor Influência temporal para cada concentração de gás .....	83
Figura 25 – Valores reais e previstos da concentração de gás por dois dias .....	84
Figura 26 – Estrutura de uma rede MLP com duas camadas ocultas .....	110
Figura 27 – Abordagem para previsão de demanda de energia elétrica.....	112
Figura 28 – Arquitetura da RNA empregada no modelo.....	113
Figura 29 – Modelo de previsão da demanda de energia utilizando <i>Simulink</i> .....	114
Figura 30 – Curvas de demanda para as classes de consumo: Comercial.....	115
Figura 31 – Curvas de demanda para as classes de consumo: industrial .....	115
Figura 32 – Curvas de demanda para as classes de consumo: residencial .....	116
Figura 33 – Curvas de demanda para as classes de consumo: comercial, industrial e residencial.....	116
Figura 34 – Rede MLP com uma estrutura de duas camadas ocultas .....	130

Figura 35 – Estrutura de previsão de eletricidade das classes de consumo.....	135
Figura 36 – Previsão de eletricidade com variáveis de entrada no tempo $t$ .....	138
Figura 37 – Previsão de eletricidade com variáveis de entrada no tempo $t-5$ , $t-10$ e $t-20$ .....	138
Figura 38 – Nível de consumo de eletricidade e situação do distribuidor de eletricidade no ano .....	139
Figura 39 – Uma rede MLP com duas camadas ocultas .....	153
Figura 40 – Diagrama de abordagem proposto .....	157
Figura 41 – Previsão de carga de energia sem seleção de variáveis e com 6 variáveis que mais contribuem no sistema .....	165
Figura 42 – Arquitetura do circuito aberto (a) e ciclo fechado (b) da rede neural autorregressiva não-linear.....	180
Figura 43 – Fluxograma do modelo de previsão autorregressiva não-linear proposto (NAR- DWT).....	182
Figura 44 – Previsão versus valores-alvo da concentração de hidrogênio no óleo .....	186
Figura 45 – Previsão em várias etapas da concentração de gás de monóxido de carbono.....	187
Figura 46 – Modelo para seleção e taxa de contribuição da concentração e previsão de gases .....	204
Figura 47 – Concentração de gás $C_2H_4$ decomposta pelo wavelet $db2$ a $db20$ .....	207
Figura 48 – Valores reais e previstos da concentração de gás por dois dias .....	209

## LISTA DE TABELAS

Tabela 1 – Arquitetura proposta da rede MLP com algoritmo backpropagation .....	64
Tabela 2 – Representação do sinal decomposto pela wavelet para cada atraso de tempo nas variáveis de entrada .....	69
Tabela 3 – Representação da decomposição da concentração de gás CO <sub>2</sub> .....	70
Tabela 4 – Contribuição e status dos artigos produzidos com a tese .....	88
Tabela 5 – Erro MAPE das previsões dos artigos 1 e 2 .....	74
Tabela 6 – Previsão com a melhor técnica de série temporal escolhida pelo critério de ajuste por ano .....	76
Tabela 7 – Desempenho das previsões e erro e tempo de treinamento .....	76
Tabela 8 – Previsão de energia elétrica (GWh) para todas classes de consumo com a seleção do melhor atraso de tempo.....	77
Tabela 9 – Erro médio MAPE(%) utilizando técnicas de séries temporais.....	78
Tabela 10 – Taxa de contribuição e erro MAPE das variáveis de entrada.....	80
Tabela 11 – Desempenho das funções da wavelet e o efeito desse fator na precisão do modelo proposto. ....	81
Tabela 12 – Erro de previsão NAR–DWT para concentrações de gás dissolvido em óleo. ....	81
Tabela 13 – Efeito do parâmetro d no desempenho do processo de treinamento para gás CO <sub>2</sub> . .....	82
Tabela 14 – Nível de correlação de casa atraso de tempo para concentração de gás.....	83
Tabela 15 – Erro das previsões para cada concentração de gás .....	85
Tabela 16 – Comparação de abordagens utilizadas para previsão em modelos autorregressivos .....	86
Tabela 17 – Melhor ajuste através do critério de desempenho AIC.....	117
Tabela 18 – Comparação da eficiência na previsão pelo índice de desempenho MAPE.....	117
Tabela 19 – Melhor ajuste escolhido por critério de ajuste, técnica selecionada e previsão de classe de consumo (GWh) por ano .....	136
Tabela 20 – Desempenho das previsões e erro e tempo de treinamento .....	137
Tabela 21 – Melhores entradas de previsão e atraso de tempo para cada classe de consumo	137
Tabela 22 – Técnicas utilizadas e comparação de previsões.....	140
Tabela 23 – Resultados do teste KMO e Bartlett .....	158
Tabela 24 – Variabilidade dos dados explicada pelos principais componentes (%). ....	160
Tabela 25 – Coeficientes variáveis nos quatro primeiros componentes principais.....	161
Tabela 26 – Resultados da taxa de contribuição das variáveis.....	164
Tabela 27 – Comparação das medidas de desempenho.....	166
Tabela 28 – Técnica para seleção de variáveis e previsão .....	167
Tabela 29 – Efeito do parâmetro d no desempenho do processo de treinamento para gás CO <sub>2</sub> . .....	184
Tabela 30 – Desempenho das funções da wavelet e o efeito desse fator na precisão do modelo proposto. ....	185
Tabela 31 – Erro de previsão NAR–DWT para concentrações de gás dissolvido em óleo. ..	185
Tabela 32 – Erro de previsão NAR – DWT para razões de concentração de gases.....	186

Tabela 33 – Erro de previsão para o gás etileno por diferentes modelos.....	187
Tabela 34 – Descrição da falha para concentração de gás .....	200
Tabela 35 – Diagnóstico de falhas pelo método da razão de <i>Dornenburg</i> .....	200
Tabela 36 – Classificação de falhas usando códigos de razão IEC.....	201
Tabela 37 – Teste de esfericidade KMO e Bartlett .....	207
Tabela 38 – Ordem de importância, taxa de cada ordem de <i>wavelet</i> para concentração de gás .....	208
Tabela 39 – Atraso no nível de correlação .....	208
Tabela 40 – Valores previstos com e sem a seleção do melhor tempo de atraso .....	209
Tabela 41 – Comparação das concentrações previstas de gás.....	211

## LISTA DE ABREVIATURAS

- ACF – Função de correlação automática
- ACO – Otimização por colônia de formigas (ACO)
- ANEEL – Agência Nacional de Energia Elétrica
- ANFIS – Sistema de inferência neuro *fuzzy* adaptativo
- AR – Modelo autorregressivo
- ASO – Operação de soma algébrica
- AT – Alta tensão
- BR – Regularização Bayesiana
- CV – Validação cruzada
- CWT – Transformada *wavelet* contínua
- DBN – Rede de crença profunda
- dBN – *wavelets Daubechies* de ordem N
- DP – Grau de polimerização
- DWT – Transformada discreta de *wavelet*
- ENN – Redes neurais por *ensemble*
- EQM – Erro quadrático médio
- FAC – Funções de autocorrelação
- FIS – Sistema de inferência *fuzzy*
- FOA – Algoritmo *Fly fruit*
- GA – Algoritmo genético
- GE – *General Electric*
- GMDH – Método de Grupo de Tratamento de Dados
- GRNN – Rede neural de regressão geral
- GWO – Otimização de *Gray Wolf*
- IA – Inteligência artificial (IA),
- IMFs – Funções de modo intrínseco
- KMO – Kaiser-Meyer-Olkin
- LARS – Método de regressão dos mínimos ângulos
- LASSO – Método de *shrinkage* e seleção de atributos para modelos de regressão linear
- LS-SVM – Máquina de vetores de suporte de mínimos quadrados
- LSTM – Memória longa de curto prazo

MAD – Erro absoluto médio

MAPE – Erro percentual absoluto médio

MLP – *Perceptron* multicamadas

mRMR – Seleção de recurso de redundância mínima

NAR – rede neural autorregressiva não linear

NARX – *Nonlinear autoregressive exogenous model*

ONS – Operador Nacional do Sistema Elétrico

PCA – Análise de componentes principais

PDs – Descargas parciais

PSO – Otimização de enxame de partículas

RMSE – Raiz do erro quadrático médio

RNA – Rede neural artificial

RNN – Rede neural recorrente

RP – Propagação reversa resiliente

SA – *Simulated annealing* (SA)

SARIMAX – Média móvel sazonal autorregressiva integrada com variáveis exógenas

SAT – Seleção do atraso de tempo

SAX – Aproximação simbólica agregada

SVM – Máquina de vetores de suporte

SWT – Transformada *wavelet* estacionária

TS – *Tabu search* (TS)

VAR – Autorregressão vetorial

WEC – Coeficiente de energia *wavelet*

WIPSO – Otimização do enxame de partículas melhoradas com peso

WT – Transformada *wavelet*

## LISTA DE SÍMBOLOS

- $C_{a,b}$  – Coeficientes da expansão de uma função em termos de uma base de funções *wavelets*  
 $I_p$  – Matrix identidade de  $p \times p$  variáveis aleatórias  
 $|R|$  – Determinante da matriz de correlação  
 $c_{0,k}$  – Coeficientes de aproximação de uma transformada *wavelet* discreta no nível 0  
 $d_{0,k}$  – Coeficientes de detalhes de uma transformada *wavelet* discreta no nível 0  
 $f_n$  – Faixa de frequência dos filtros passa-baixa e passa-alta  
 $X^2$  – Força de relacionamento entre as variáveis  
 $\rho_{jk}$  – Coeficiente de correlação parcial entre  $X_j$  e  $X_k$   
 $r_{jk}$  – Coeficiente de correlação simples entre as variáveis  $X_j$  e  $X_k$   
 $\psi_{a,b}(t)$  – Função *wavelet* contínua  
 $\psi_{jk}(x)$  – Função *wavelet* discreta  
 $\emptyset$  – Multiplicador de Lagrange  
 $F$  – Horizonte de previsão  
 $\mathbb{R}$  – Conjunto dos números reais  
 $X$  – Conjunto de números reais  
 $G, H$  – Filtros passa-baixa e passa-alta  
 $S$  – Matriz de covariância substituta da matriz  $\Sigma$   
 $V$  – Subespaço vetorial gerado por translação e dilatações da função *wavelet*  
 $W$  – Subespaço vetorial gerado por translação e dilatações da função escaladora  
 $Y[n]$  – Saídas dos coeficientes de detalhes do filtro passa-alta e dos coeficientes de aproximação do filtro passa-baixa  
 $Z$  – Componentes principais  
 $a, b$  – Parâmetros de dilatação e translação de uma função *wavelet*  
 $cov(x, y)$  – Covariância entre as variáveis aleatórias  $x$  e  $y$   
 $dim(V)$  – Dimensão do espaço vetorial  $V$   
 $g, h$  – Coeficientes dos filtros passa-baixa e passa-alta  
 $i, j, n, p$  – Números inteiros  
 $k$  –  $k$ -ésimo passo de um processo iterativo  
 $m$  – Quantidade de componentes principais  
 $u, X$  – Vetores  
 $v$  – Distribuição qui-quadrado com graus de liberdade.

$\Sigma$  – Matriz de covariância

$\alpha$  – Autovetor da matriz de covariância  $\Sigma$

$\lambda$  – Autovalor da matriz de covariância  $\Sigma$

$\phi$  – Função escaladora

## SUMÁRIO

1. INTRODUÇÃO .....	20
1.1. Considerações gerais .....	20
1.2. Motivação.....	22
1.3. Objetivo da pesquisa .....	24
1.3.1. Objetivo geral .....	24
1.3.2. Objetivos específicos .....	24
1.4. Principais contribuições da tese .....	25
1.5. Estrutura da tese .....	26
2. MODELOS AUTORREGRESSIVOS DE PREVISÃO UTILIZADOS EM PROBLEMAS DO SETOR ELÉTRICO - ESTADO DA ARTE .....	27
2.1. Modelos de previsão de eletricidade, carga de energia e preço de eletricidade.....	27
2.2. Modelos de previsão de concentração de gás dissolvido em óleo em transformadores de potência.....	30
2.3. Outros modelos de previsão dos setores elétricos .....	32
3. APRESENTAÇÃO TEÓRICA .....	35
3.1. Modelos de previsão com técnicas estatísticas .....	35
3.2. Modelos de previsão com redes neurais.....	38
3.2.1.1. Rede Neural <i>Perceptron</i> Multicamadas .....	38
3.2.1.2. Rede neural autorregressiva não linear .....	39
3.2.1.3. Rede neural recorrente .....	40
3.3. Técnicas para seleção de atributos .....	41
3.3.1. Coeficientes da correlação de Pearson, de Kendall, de Spearman .....	42
3.3.2. Análise de componentes principais.....	42
3.4. Introdução sobre a transformada <i>wavelet</i> .....	47
3.4.1. Transformada de <i>wavelet</i> discreta.....	50
3.4.2. O banco de filtros da transformada <i>wavelet</i> discreta .....	52
3.4.2.1. <i>Wavelet Daubechies, Coifman e Symmlet</i> .....	55
3.4. Teste Kaiser-Meyer-Olkin e Bartlett.....	57
3.5. Indicadores de desempenho .....	58
4. MODELO PROPOSTO E CONTRIBUIÇÃO DA TESE .....	60
4.4. Modelo proposto da tese .....	60
4.4.3. Modelo autorregressivo para previsão de demanda de eletricidade .....	61
4.4.4. Abordagem para seleção de atributos (seleção de variáveis de entrada) .....	64
4.4.5. Uso de <i>wavelet</i> mais adequada para construir aproximações suaves aos dados de entrada.....	66
4.4.6. Desenvolvimento de um novo modelo autorregressivo de previsão baseado em <i>wavelet</i> , PCA e seleção de atraso de tempo .....	68
4.5. Desenvolvimento da produção bibliográfica da tese.....	87
5. RESULTADOS DOS MODELOS AUTORREGRESSIVOS DE PREVISÃO .....	74
5.1. Resultados do primeiro e segundo modelo autorregressivo de previsão.....	74
5.2. Resultados do terceiro modelo autorregressivo de previsão .....	78

5.3. Resultados do quarto e quinto modelo autorregressivo de previsão .....	81
6. CONSIDERAÇÕES FINAIS .....	88
6.1. Conclusões e contribuições .....	89
6.2. Limitações do trabalho .....	92
6.3. Desenvolvimentos futuros.....	93
REFERÊNCIAS BIBLIOGRÁFICAS .....	94
APÊNDICES .....	103
APÊNDICE A (Artigo 1) – Uma abordagem computacional para previsão de demanda de energia por classe de consumo para as distribuidoras de energia elétrica no Brasil.	103
APÊNDICE B (Artigo 2) – Uma abordagem computacional para previsão e minimização dos custos de eletricidade no mercado de curto prazo dos distribuidores de eletricidade no Brasil. ....	122
APÊNDICE C (Artigo 3) – Uma abordagem baseada em PCA para seleção e taxa de contribuição para previsão de carga de energia elétrica. ....	145
APÊNDICE D (Artigo 4) – Modelos de redes neurais autorregressivas não lineares para previsão de concentrações de gases dissolvidos em óleo em transformadores .....	174
APÊNDICE E (Artigo 5) – Transformada do tipo <i>wavelet</i> para otimizar a ordem de um modelo de rede neural autorregressiva para prever a concentração de gás dissolvido no óleo do transformador de potência .....	194
REFERÊNCIAS .....	213

# 1. INTRODUÇÃO

Este capítulo apresenta as considerações gerais desta tese, descrevendo o que é série temporal, modelo de previsão subjetivo, univariado, multivariado e métodos de seleção de variáveis. Também apresenta, de forma resumida, o uso de técnicas utilizadas por alguns trabalhos para desenvolver modelos autorregressivos de previsão no setor elétrico de energia.

## 1.1. Considerações gerais

A evolução dos dados na era da informação desempenha um papel significativo, pois os dados fornecem uma base para controle de máquinas, funcionamento do sistema industrial, mercado econômico, gestão ambiental, setor elétrico de energia, prevenção de emergência e controle avançado em vários sistemas complexos, etc. Para lidar com essa evolução, são necessárias abordagens e ferramentas capazes de prever e auxiliar um gestor em relação às medidas a serem adotadas para a tomada de decisão (JÚNIOR, 2015; YANG et al., 2018; BAI, et al., 2019).

A previsão de uma série temporal utilizada para esse fim, consiste na análise da coleção de observações feitas sequencialmente, cujo processo observacional e consequente quantificação numérica, produz uma sequência de dados distribuídos no tempo (ROITMAN, 2001; CHATFIELD, 2003; MORETTIN e TOLOI, 2006; BOX e JENKINS, 2015; YANG et al., 2018; MILLS, 2019).

Geralmente, o estudo de uma série temporal tem os seguintes objetivos: caracterização, que visa procurar determinar propriedades fundamentais da série; modelagem, que descreve o comportamento da série e; por último, a previsão, que consiste em antever, de forma precisa, o comportamento da série à partir do modelo elaborado (ROITMAN, 2001).

Em relação aos métodos de previsão, podem ser classificados em previsões subjetivas, univariadas e multivariadas. As previsões subjetivas são feitas usando julgamento, intuição. Já as previsões univariadas, chamadas de métodos ingênuos ou de projeção, referem-se a previsão de valores futuros de determinada série e são baseadas inteiramente em valores passados. Por outro lado, em previsões multivariadas, a resposta depende dos valores de uma ou mais séries, denominadas variáveis preditivas ou explicativas. Por exemplo, as previsões de vendas que podem depender de estoques ou índices econômicos (ROITMAN, 2001; CHATFIELD, 2009; YANG et al., 2018).

Nos modelos de previsão, destacam-se, em particular, os modelos autorregressivos, nos quais o valor atual da série temporal é regredido em valores passados da mesma série e, possivelmente, valores passados e presentes de uma ou mais variáveis exógenas (variável que é determinada fora do modelo), com diversas aplicações na previsão de valores futuros de variáveis (ADHIKARI e AGRAWAL, 2013; BOX e JENKINS, 2015; CHATFIELD, 2003).

Um problema encontrado em modelos autorregressivos de previsão, sejam univariados ou multivariados, está relacionado à definição de o quanto do passado deve ser analisado para prever o futuro, características óbvias não lineares, não estacionárias e complicadas na série temporal. Além disso, outra questão importante nesses modelos está relacionada à seleção das principais variáveis explicativas para fins de previsão (ABEDINIA et al., 2017; YANG et al., 2018; LIANG et al., 2019; BAI, et al., 2019).

O capítulo 2 mostra diversas abordagens para melhorar a precisão dos modelos autorregressivos de previsão, por meio da seleção de variáveis mais relevantes, como os trabalhos de Esfetang e Kazemzadeh (2018), Prasojo et al. (2018), Amjady e Keynia (2009), Abedinia et al. (2017), Lang et al. (2018), Yang et al. (2019), BAI, et al., 2019 e Liang et al. (2019).

Também é possível notar que alguns trabalhos sobre modelos autorregressivos de previsão, decompueram o sinal de entrada, criando diferentes séries temporais para os dados de entrada antes de realizar a previsão, tais como os trabalhos de Frizzo Stefenon et al. (2020), Zheng et al. (2018), Pereira et al. (2018) e Liu et al. (2019), Diaconescu (2008).

Outros modelos autorregressivos procuram desenvolver o modelo de previsão utilizando o recurso do atraso de tempo nas variáveis de entrada para prever com  $F$  passos à frente, tais como os trabalhos de Diaconescu (2008), Hong (2010), Nengbao et al. (2014), Lahouar e Slama (2015), Lang et al. (2018) e Pereira et al. (2018).

Existem ainda, modelos autorregressivos que procuram utilizar a técnica *wavelet* ou um modelo *fuzzy* para melhorar a precisão da previsão sem o recurso da seleção de variáveis ou do atraso de tempo nas variáveis de entrada, tais como os trabalhos de Dong et al (2015), Bayram e Şeker (2017), Noori et al. (2017) e Prasojo et al. (2018).

Os métodos de seleção de variáveis são fundamentais, pois podem identificar um subconjunto de variáveis explicativas que contêm as informações mais relevantes no conjunto de dados completo e, portanto, influenciar a precisão da previsão (YANG et al., 2019; ABEDINIA et al., 2017; LI et al., 2019). Entretanto, percebe-se que apesar da preocupação em selecionar variáveis fundamentais e da preocupação em utilizar recurso de atraso de tempo nas variáveis de entrada para o modelo autorregressivo de previsão, apenas o trabalho de Lang et

al. (2018) se concentrou na seleção do melhor atraso de tempo para cada variável de entrada antes de desenvolver o modelo autorregressivo de previsão.

Assim, esta tese de doutorado, construída por meio de artigos científicos, estuda os efeitos da seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão baseados em Redes Neurais Artificiais com aplicações em três bases de dados no setor de energia elétrica: carga de energia elétrica (geração e distribuição de energia) e concentração de gás dissolvido em óleo de um transformador de potência.

Na sequência da introdução, serão apresentadas as motivações, o objetivo geral e objetivos específicos, as principais contribuições e a estrutura da tese.

## **1.2. Motivação**

Com o advento de sensores menores e baratos, dispositivos e aplicativos inteligentes, a taxa de geração de dados aumenta rapidamente, fazendo que exista grande desafio em tomada decisões, permitindo oportunidades para análises e prognósticos preditivos aprimorados visando reduzir custos, aumentar a produtividade e as receitas (KITCHIN, 2014; TANG et al., 2014; ZHANG 2017; GHAVAMI, 2020).

Um dos desafios associados ao grande volume de dados é a dificuldade em representar os dados de séries temporais, por causa da sua alta dimensionalidade, e os ruídos, geralmente causados pela fonte de dados e imperfeição das tecnologias utilizadas para sua coleta (FU, 2011; NASEHI e POURGHASSEM, 2011; TANG et al., 2014; AGHABOZORGI et al., 2015). Além disso, múltiplas variáveis impactam o desenvolvimento do modelo autorregressivo, dificultando a previsão de uma determinada variável (BAI, et al., 2019).

Sendo assim, em problema de previsão da série temporal, deve ser considerado a característica de sequência da série temporal, bem como o efeito mútuo das variáveis relacionadas (BAI, et al., 2019).

Duas técnicas comumente utilizadas para resolver esse problema são a redução de dimensionalidade e a classificação ou seleção de variáveis. Enquanto a redução de dimensionalidade busca remover recursos ruidosos e redundantes (Yen e Leong, 2006; Tang et al., 2014), as técnicas de seleção de variáveis são utilizadas para reduzir os atributos das variáveis, obter um subconjunto de recursos relevantes para atingir os objetivos de classificação e agrupamento, além de trazer melhoria do desempenho dos algoritmos de aprendizado de máquina em termos de redução do tempo para, assim, construir o modelo de previsão e

aumentar a precisão (KUMAR e MINZ, 2014; TANG et al., 2014; VELMURUGAN e ANURADHA, 2016; SINGH et al., 2016).

No setor elétrico, por exemplo, as empresas distribuidoras e geradoras de energia precisam lidar com grande volume de dados coletados por meio de medidores inteligentes que oferecem oportunidades para as empresas acompanharem o comportamento de consumo de eletricidade dos clientes por meio de perfil de carga (WANG et al., 2016).

De forma semelhante, a coleta de dados de sensores também permite monitorar o estado de funcionamento de diferentes equipamentos como, por exemplo, os transformadores de potência, por meio de análise de concentrações de gases. Também é possível estudar o comportamento e a capacidade de carga e descarga de energia elétrica em baterias, estudar o comportamento do consumo de eletricidade da indústria, comércio, residência, entre outras oportunidades (WANG et al., 2016).

Além disso, as informações futuras também são muito importantes, pois pode orientar a operação antecipada para ajuste de um sistema, adaptação ambiental e prevenção de acidentes. Portanto, a previsão confiável dos dados no domínio do tempo torna-se uma questão extremamente urgente (BAI, et al., 2019).

Apesar de haver modelos autorregressivos híbridos cada vez mais inteligentes para melhorar a precisão das previsões, seja em problemas de preço de eletricidade, de carga de eletricidade, de concentração de gás dissolvido em óleo, ajuste de um sistema, prevenção de acidentes, mercado financeiro, entre outras, poucos trabalhos visam selecionar características de entrada ideais para os modelos de previsão (Yang et al., 2019).

Percebe-se, portanto, que existe uma lacuna na literatura em relação à seleção das variáveis mais importantes, além da escolha do atraso de tempo ideal para as variáveis de entrada no modelo autorregressivo de previsão.

Nos diversos trabalhos que se preocupam em melhorar a precisão das previsões, é adotado o mesmo valor de atraso de tempo para todas as variáveis envolvidas, o que pode não representar adequadamente a relação subjacente do processo em estudo, isso significa determinar o número de valores passados da série temporal de interesse e das variáveis exógenas a serem usadas no modelo de previsão. Além disso, existe a dificuldade em selecionar algumas variáveis importantes entre dezenas ou centenas de variáveis disponíveis.

Outra questão pertinente aos modelos de previsão desenvolvidos está relacionada ao uso de diferentes tipos de *wavelet* utilizadas nos modelos autorregressivos. Nesse sentido, tanto em Pereira et al. (2018), Liu et al. (2019) e Zheng et al. (2018), os autores utilizaram diferentes *wavelets* para suavizar o sinal de entrada, seja para usar em rede neural autorregressiva não

linear (NAR) (Pereira et al., 2018), seja para ser utilizado em abordagem híbrida, máquina de vetores de suporte de mínimos quadrados (LS-SVM) e um algoritmo de otimização de enxame de partículas (PSO) (Zheng et al., 2018), seja para ser utilizada em abordagem com LS-SVM e algoritmo de competição imperialista. Contudo, a transformada do tipo *wavelet*, utilizada naqueles trabalhos, não são empregadas com objetivo de criar sinais de entrada com diferentes atrasos de tempo.

As lacunas expostas nesta seção e na revisão da literatura evidenciam a importância do estudo na seleção de atributos e seus efeitos em modelos autorregressivos de previsão, além da seleção da definição da ordem das variáveis de entrada em um modelo autorregressivo de previsão.

Assim, a contribuição deste trabalho, se dá com o desenvolvimento de um novo modelo autorregressivo de previsão, sendo que para isso, foi necessário testar diferentes abordagens para alcançar o objetivo geral desta tese. O desenvolvimento de algumas abordagens, com intuito de criar e testar abordagens para entender melhor o processo subjacente entre as variáveis de entrada e saída do modelo, resultaram em publicações em revistas internacionais.

### **1.3. Objetivo da pesquisa**

De forma a delimitar o estudo que se pretendeu realizar sobre uso de modelos autorregressivos de previsão são apresentados a seguir os objetivos desta tese.

#### **1.3.1. Objetivo geral**

Desenvolver uma abordagem baseada na transformada *wavelet* e na análise de componentes principais para seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão fundamentados em redes neurais artificiais.

#### **1.3.2. Objetivos específicos**

- a) Estudar o efeito de atrasos de tempo na variável de resposta em um modelo autorregressivo de previsão.
- b) Avaliar o efeito entre atrasos de tempo e taxa de amostragem em modelos autorregressivos de previsão.

- c) Selecionar e mostrar a taxa de contribuição das variáveis de entrada mais relevantes para um modelo autorregressivo de previsão.
- d) Estudar o tipo de *wavelet* mais adequada para construir aproximações com uma versão esparsa e simplificada dos dados de entrada, e testar em um modelo autorregressivo de previsão.

#### 1.4. Principais contribuições da tese

A principal contribuição deste trabalho é a utilização da *wavelet Daubechies* para criar séries temporais com diferentes atrasos de tempo e usá-las como entrada para um modelo autorregressivo de previsão. A característica principal desta abordagem é criar séries temporais com diferentes atrasos de tempo, selecionar e mostrar o nível de contribuição de cada atraso de tempo para cada variável de entrada. Essa abordagem é então aplicada ao problema de previsão de concentração de gases dissolvidos em óleo em transformador de potência.

Para responder o objetivo geral desta tese, neste caso, desenvolvimento de uma abordagem baseada na transformada *wavelet* e na análise de componentes principais para seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão fundamentados em redes neurais artificiais, foi necessário testar diferentes abordagens, as quais deram sustentação para a ideia principal da tese.

Foram testadas abordagens para verificar o efeito de atrasos de tempo na variável de resposta de um modelo, avaliar o efeito entre atrasos de tempo e taxa de amostragem, seleção de variáveis e aproximações mais relevantes para o modelo e, por fim, estudar o tipo de *wavelet* mais adequada para construir aproximações com objetivo de criar atrasos de tempo para um modelo autorregressivo de previsão

Por fim, são apresentados modelos autorregressivos com dados reais, coletados mensalmente da Agência Nacional de Energia Elétrica (ANEEL), dados diários, coletados do Operador Nacional do Sistema Elétrico (ONS) e um conjunto de dados históricos diários de gás dissolvido em óleo, coletado de um transformador equipado com sensor da *GE Kelman-Transfix* (*GE – General Electric*, São Paulo, Brasil) e *GE Intellix BMT 330* (*GE – General Electric*, São Paulo, Brasil).

Em todas as aplicações, a abordagem demonstrou uma contribuição prática significativa, apresentando resultados competitivos com o estado da arte, como pode ser observado no capítulo de resultados, bem como nos artigos científicos resultantes desta tese. Destaca-se, no entanto, as contribuições teóricas do método proposto. Tradicionalmente, uma

das etapas do processo de predição de um modelo autorregressivo é a transformação da série temporal em uma tabela atributo-valor que consiste em associar  $d$  observações do passado (atributos) para prever uma observação no futuro (valor) para criar um problema de aprendizado supervisionado (BROWNLEE, 2018).

Se um atraso de tempo igual a três for adotado,  $d = 3$ , por exemplo, o modelo terá três valores de entrada e um valor de saída para cada variável. Ao aplicar a transformada *wavelet* aos dados históricos da variável, as  $d$  entradas são reunidas em uma única representação, independente do atraso de tempo considerado, resultando em modelo com um número muito menor de variáveis preditivas. Adicionalmente, ao criar representações *wavelet* do sinal original para diversos valores de  $d$  e selecionar os mais representativos com base na PCA é possível determinar o valor ideal de  $d$  para cada variável de entrada de forma individual.

Por fim, as abordagens propostas nesta tese, resultaram na publicação de um artigo científico em um congresso e em revistas científicas internacionais, demonstrando a importância de cada abordagem desenvolvida.

## **1.5. Estrutura da tese**

O restante da tese está organizado da seguinte forma: capítulo 2 apresenta modelos autorregressivos de previsão utilizados em problemas do setor elétrico - estado da arte, capítulo 3, mostra a teoria relacionada sobre técnicas de seleção de atributos e seus efeitos em modelos autorregressivos. O capítulo 4 descreve os procedimentos metodológicos utilizados nesta para construir os modelos autorregressivos de previsão. O capítulo 5 mostra os resultados de cada modelo autorregressivo, bem como a produção acadêmica da tese, composta por cinco artigos científicos, publicados, aceitos ou em análise em periódicos internacionais (ver apêndices), os quais evidenciam os respectivos objetivos, métodos de pesquisa, resultados e conclusões. Por fim, o capítulo 6 mostra as considerações finais da tese, revelando o alcance dos objetivos deste trabalho por meio da produção dos artigos científicos apresentados nos apêndices, além de apresentar as limitações e possíveis pesquisas futuras.

## **2. MODELOS AUTORREGRESSIVOS DE PREVISÃO UTILIZADOS EM PROBLEMAS DO SETOR ELÉTRICO - ESTADO DA ARTE**

Este capítulo apresenta a revisão da literatura sobre modelos autorregressivos de previsão desenvolvidos e aplicados no setor elétrico de energia.

Os modelos desenvolvidos nesta tese utilizam dados do setor elétrico de energia, tais como concentração de gás dissolvido em óleo em transformador de potência, dados de receita de fornecimento de energia, número de unidades consumidoras, tarifa de fornecimento de energia e consumo de energia elétrica de uma distribuidora de eletricidade, dados de geração de energia elétrica, como energia armazenada, demanda instantânea, demanda horária máxima, afluente natural de energia bruta e armazenável, carga de energia, troca de energia, volumes hidrológicos, nível de quantidade, fluxo efluente e afluente.

Os modelos autorregressivos desenvolvidos nesta tese refletem os interesses do grupo de pesquisa e, por esta razão, a revisão da literatura se concentra nos modelos autorregressivos de previsão aplicados no setor elétrico de energia.

No entanto, apesar desta tese mostrar modelos autorregressivos de previsão com dados do setor elétrico de energia, as abordagens podem ser aplicadas em qualquer setor de interesse.

### **2.1. Modelos de previsão de eletricidade, carga de energia e preço de eletricidade**

Em um modelo autorregressivo, uma série temporal é uma maneira de capturar autocorrelação e sazonalidade em uma sequência de observações ordenadas ao longo do tempo (MORETIN e TOLOI, 2004; BRACHER e HELD, 2018).

A análise de série temporal possui basicamente três objetivos: identificar a natureza do fenômeno representado, procurando um padrão de comportamento, como a tendência, a existência de variação sazonal, valores discrepantes, alterações estruturais etc. Também tem como objetivo usar a variação em série para explicar a variação em outra série, e, finalmente, prever a evolução da variável da série temporal a partir de um modelo matemático que descreva o comportamento das observações (JÚNIOR, 2016).

A literatura mostra diversos trabalhos que procuram melhorar a precisão de modelos autorregressivos de previsões aplicados a problemas do setor elétrico, mercado financeiro, operação para ajuste de um sistema, adaptação ambiental, prevenção de acidentes, etc. Muitos desses trabalhos utilizam técnicas que podem ser categorizadas em dois grupos: 1) abordagens

estatísticas, como análise de regressão, análise de série temporal; e 2) técnicas baseadas em inteligência artificial (IA), como: *perceptron* multicamadas (MLP), rede neural com aprendizado profundo, lógica *Fuzzy*, máquina de vetores de suporte (SVM), máquina de vetores de suporte de mínimos quadrados (LS-SVM), floresta aleatória, entre outras (HONG, 2010).

Como exemplos, Diaconescu (2008), Amjady e Keynia (2009) e Hong (2010) que propuseram modelos de previsão de consumo de eletricidade, preço de eletricidade e carga de eletricidade no curto prazo, respectivamente.

No trabalho de Diaconescu (2008), foi proposto um modelo multivariado de previsão, criando séries temporais caóticas utilizando quatro séries temporais: logística, *Mackey-Glass*, *fractal Weirstrass* e índice BET nas variáveis de entrada (consumo de eletricidade e temperatura). Os autores testaram as variáveis de entrada no instante de tempo  $t$ ,  $t-1$  e  $t-2$  em uma rede neural autorregressiva não-linear com variáveis exógenas (NARX do inglês *Nonlinear autoregressive exogenous model*) e observaram que a previsão de uma série temporal será aprimorada analisando simultaneamente as séries temporais relacionadas. Assim, o consumo de eletricidade é mais bem previsto quando os últimos dias de consumo de eletricidade e temperatura são aplicados simultaneamente como entrada para a rede NARX.

Já em Amjady e Keynia (2009), os autores propuseram uma nova estratégia de previsão de preços nos mercados de eletricidade, na qual a técnica de seleção de variáveis proposta possui duas etapas; no primeiro estágio, entradas candidatas irrelevantes são removidas por uma versão modificada do algoritmo de alívio. Entre o subconjunto escolhido do primeiro estágio, a análise de correlação cruzada filtra os candidatos redundantes no segundo estágio.

As entradas candidatas selecionadas no trabalho de Amjady e Keynia (2009) têm relevância máxima com a saída (preço da próxima hora) e redundância mínima. Esse conjunto eficiente de entradas é usado para o mecanismo de previsão, que é composto por uma cadeia de previsores baseados em uma rede neural em cascata. Na cadeia, cada previsor é alocado para prever o preço de 1 hora do dia seguinte.

Já em Hong (2010) foi proposta uma estrutura de previsão integrada utilizando uma rede neural MLP com três abordagens: a primeira, utilizando apenas a tendência linear e a temperatura como entradas; a segunda abordagem, utilizando hora, dia da semana, mês, tendência linear e a temperatura como entradas e; na terceira abordagem, utilizando hora, dia da semana, mês, tendência linear, temperatura, temperaturas das três horas anteriores, temperatura ponderada das 24 horas anteriores como entradas para a rede neural.

Em Nengbao et al. (2014), Lahouar e Slama (2015), Wang et al. (2016) e Abedinia et al. (2017), foram propostos modelos de previsão de carga de energia.

Nengbao et al. (2014) propuseram dois modelos autorregressivos, um modelo de série temporal média móvel sazonal autorregressiva integrada com variáveis exógenas (SARIMAX) e um modelo de rede neural artificial (RNA). Para isso, utilizaram variáveis *dummy* (variáveis categóricas): variáveis *dummy* de hora em hora, hora do dia, dia da semana, mês do ano, Ramadã, feriados e temperatura como variáveis exógenas, atrasadas em  $t-1$ ,  $t-2$ ,  $t-6$ ,  $t-12$  e  $t-24$ . Um método chamado “pré-clareamento” foi usado para determinar o efeito retardado de temperatura na carga de eletricidade. Após isso, as temperaturas foram utilizadas como entrada em uma rede neural MLP com algoritmo *backpropagation*.

Enquanto em Lahouar e Slama (2015), foi proposto um modelo autorregressivo de previsão de carga de energia de um dia à frente com resolução de uma hora. Os autores usaram uma técnica chamada floresta aleatória de regressão, tendo como entradas: estação do ano, temperatura, tipo de dia, carga horária e carga do dia, pico de carga de manhã e à noite com atraso em  $t-1$ .

No modelo desenvolvido por Wang et al. (2016) a aproximação simbólica agregada (SAX) é realizada para cada cliente para reduzir a escala do conjunto de dados, enquanto o modelo de Markov baseado no tempo é aplicado para modelar a dinâmica do consumo de eletricidade, transformando o grande conjunto de dados de curvas de carga em várias matrizes de transição de estado. Posteriormente, é realizado um agrupamento por busca rápida e descoberta de picos de densidade para obter a dinâmica típica do comportamento de consumo, com a diferença entre quaisquer dois padrões de consumo, e classificar os clientes em vários clusters.

Abedinia et al. (2017) e Lang et al. (2018) propuseram um modelo de previsão de carga e preço dos mercados de eletricidade. No trabalho de Abedinia et al. (2017) é empregado um novo método de seleção de variáveis, em que um filtro seleciona um subconjunto mínimo dos recursos mais informativos, considerando a relevância, redundância e interação das entradas candidatas (valores de carga e temperatura defasadas) de forma coordenada. Enquanto a parte do *wrapper* ajusta as configurações do filtro, uma rede neural MLP com algoritmo *backpropagation* faz as previsões de carga e preço dos mercados de eletricidade no curto prazo.

Já no trabalho de Lang et al. (2018) é proposto um modelo que combina vantagens da rede neural com pesos aleatórios e máquina de vetores de suporte, ele pode capturar características mais dinâmicas dos dados multivariados da série temporal por meio de amostras de ponderação. A informação mútua é usada para determinar os atrasos de tempo, entre  $x_t$  e  $x_{t-d}$  com diferentes  $d$  (atraso de tempo), em que o primeiro valor mínimo local de informação mútua, corresponde ao atraso de tempo para a série temporal.

Em Yang et al. (2019) e Liang et al. (2019), é proposto um modelo de previsão de carga de eletricidade com base na ideia de selecionar recursos de entrada ideais. Yang et al. (2019) desenvolveram um modelo híbrido que combina função de correlação automática (ACF) e máquina de vetores de suporte de mínimos quadrados (LS-SVM). A ACF é aplicada para selecionar as variáveis de entrada informativa entre onze variáveis, além da seleção da janela de tempo de atraso de oito dias anteriores como entrada. Por fim, LS-SVM é utilizado para previsão, e seus parâmetros são otimizados por algoritmo de otimização de *Gray Wolf* (GWO) e validação cruzada (CV).

Enquanto em Liang et al. (2018), uma série de carga original é decomposta em uma quantidade de funções de modo intrínseco (IMFs) e um resíduo com frequência diferente, de modo a enfraquecer a volatilidade da série influenciada por fatores complicados. Em seguida, a seleção de recurso de redundância mínima (mRMR) é empregada para obter a melhor característica definida através da análise de correlação entre cada IMF e as características, incluindo tipos diurnos, temperatura, condições de meteorologia e assim por diante. Finalmente, um algoritmo de otimização *fly fruit* (FOA) é utilizado para otimizar o fator de suavização em uma rede neural de regressão geral (GRNN).

## **2.2. Modelos de previsão de concentração de gás dissolvido em óleo em transformadores de potência**

Em relação ao problema de concentração de gás dissolvido em óleo em transformadores, existem diversas técnicas que são utilizadas para previsão da concentração de gás e da relação entre gases.

No trabalho de Noori et al. (2017) e Prasojo et al. (2018) foi desenvolvido um modelo de previsão de concentração de gás dissolvido em óleo em transformadores de potência com sistema de inferência *fuzzy* (FIS). Enquanto Noori et al. (2017) propuseram um sistema FIS para detectar a decomposição do papel do isolamento de transformadores, na qual utilizaram o modelo *Mamdani* como motor de inferência *fuzzy*, funções triangulares como as funções de pertinência e o método *Zadeh* para a implicação (AND) e operações de agregação (OR), que, respectivamente, estão relacionadas às operações “*Min*” e “*Max*”, e, de acordo com as diretrizes IEC 60599, 221 regras difusas “*IF THEN*” foram configuradas para o FIS usando o operador “*AND*”.

No modelo de previsão desenvolvido por Prasojo et al. (2018), o grau de polimerização (DP) do transformador foi previsto usando o modelo de sistema de inferência neuro *fuzzy*

adaptativo (ANFIS) baseado no uso de características dielétricas e gases dissolvidos como entrada, na qual a melhor combinação de variáveis de entrada e função de pertinência foi selecionada para construir o modelo ANFIS ideal, que foi, então, comparado e avaliado. Os parâmetros das funções de pertinências foram definidos por treinamento, usando algoritmo híbrido, que é a combinação de *backpropagation* e mínimos quadrados.

Já no trabalho de Lin et al. (2018), um método autorregressivo de previsão com base na tendência da concentração de gás no óleo do transformador ( $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  e  $C_2H_6$ ) foi proposto. Para isso, um modelo de memória longa de curto prazo (LSTM) é estabelecido para prever a futura concentração de gás. Em seguida, a precisão e os fatores de influência do modelo LSTM são analisados com exemplos. O modelo de rede de crença profunda (DBN) é usado para estabelecer a operação do transformador usando as informações na biblioteca de casos de falha do transformador.

Por outro lado, Zheng et al. (2018), Pereira et al. (2018) e Liu et al. (2019) propuseram um modelo híbrido de previsão de concentração de gás no óleo do transformador, utilizando diferentes tipos de *wavelet*.

Em Pereira et al. (2018), os autores utilizaram as *wavelet* (*Coiflets*, *Symlets* e *Daubechies*) e uma rede neural NAR para prever valores futuros de sete tipos de gás dissolvido em óleo e as razões IEC e Rogers ( $CH_4 / H_2$ ,  $C_2H_2 / C_2H_4$ ,  $C_2H_4 / C_2H_6$ ). Nessa abordagem, a transformação *wavelet* discreta cria uma versão esparsa e simplificada com boas propriedades de aproximação dos dados de entrada e a rede neural NAR é treinada e validada de acordo com uma abordagem de validação cruzada *k-fold*.

Já em Zheng et al. (2018) e Liu et al. (2019) foi proposta uma abordagem utilizando *wavelet* e a máquina de vetor de suporte de mínimos quadrados (LS-SVM). No trabalho de Zheng et al. (2018) os autores testaram a *wavelet* (*Morlet*, *Marr* e *DOG*) com a máquina de vetor de suporte de mínimos quadrados (LS-SVM) e um algoritmo de otimização por enxame de partículas (PSO) com mutação para otimizar os parâmetros de regressão *wavelet* LS-SVM, tendo como entrada ( $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ). Enquanto Liu et al. (2019) utilizaram LS-SVM com um algoritmo de competição imperialista para otimizar os hiperparâmetros da regressão *wavelet* LS-SVM.

### 2.3. Outros modelos de previsão dos setores elétricos

Este subcapítulo mostra outros problemas no setor elétrico, tais como: identificação de descarga parcial, estado de energia de bateria de íons de lítio, falhas dentro de um motor elétrico, geração de energia em parques eólicos, carga e descarga de baterias, entre outros.

Descargas parciais (PDs) são descargas elétricas que ocorrem dentro ou fora do isolamento de um sistema de alta tensão (AT) sob estresse elétrico (KREUGER, 1989; MAS'UD et al., 2016).

A análise de falhas de isolamento elétrico revela que esse tipo de falha é a causa raiz de mais de 60% dos danos a equipamentos de alta tensão (RAYMOND et al., 2015). Assim, o reconhecimento de PD tem sido um tópico de interesse para distinguir diferentes fontes de falha de PD em sistemas de isolamento de aparelhos de energia (MAS'UD et al., 2014; RAYMOND et al., 2015).

Em Mas'ud et al (2014) por exemplo, foi desenvolvido uma técnica para classificar padrões de PD com base no aprendizado de redes neurais por ensemble (ENN). A técnica ENN é baseada no treinamento de uma série de modelos de redes neurais com parâmetros estatísticos de padrões de descarga parcial e na combinação de suas previsões. Combinando as saídas das redes neurais constituintes por meio de uma unidade de agregação usando a estratégia de média ponderada dinamicamente, ao qual fornece uma avaliação final dos padrões de descarga parcial em relação a uma gama de tipos de falha de descarga parcial.

O estado de energia em baterias é um outro grande problema do setor elétrico, principalmente por causa dos carros elétricos. Isso vem fazendo com que diversos autores proponham novos métodos de previsão para o estado de saúde das baterias (DONG et al., 2015).

Diante deste problema, um modelo de previsão baseado em rede neural *wavelet* com três camadas e estimador de filtro de partículas para a estimativa do estado de energia de bateria de íons de lítio foi proposto por Dong et al (2015). O modelo de bateria baseado em rede neural *wavelet* é usado para simular todas as características elétricas dinâmicas das baterias. Enquanto um estimador de filtro de partículas é aplicado para estimar o estado de energia da célula da bateria para suprimir os ruídos de medição de corrente e tensão.

Em Esfetang e Kazemzadeh (2018) é proposta uma nova técnica híbrida para previsão de geração de energia em parques eólicos. Em que, o número efetivo de parâmetros de entrada é determinado pela análise de correlação de dados. Em seguida, o tipo do número efetivo de parâmetros é selecionado a partir dos parâmetros atuais via rede neural. A transformada *wavelet* do tipo *Daubechie* é usada para filtrar e decompor os dados de entrada relacionados à energia

eólica; enquanto a rede neural com função de base radial é utilizada como um preditor preliminar.

A previsão final do trabalho de Esfetang e Kazemzadeh (2018) é combinada com três redes neurais MLP com algoritmos de aprendizagem: regularização *Bayesiana* (BR), propagação reversa resiliente (RP) e *backpropagation*. Por fim, a técnica heurística com otimização do enxame de partículas melhoradas com peso (WIPSO) é usada para melhorar a precisão das previsões e escapar do mínimo local para aperfeiçoar os pesos das redes neurais.

Os estudos de prognóstico e diagnóstico em motores de indução é outro aspecto importante para proporcionar a continuidade operacional do processo industrial (BAYRAM e ŞEKER, 2017; PURUSHOTTAM e RAJIV, 2019). Assim, Bayram e Şeker (2017) e Purushottam e Rajiv. (2019) propuseram um modelo de previsão para detectar falhas em um motor de indução.

No trabalho de Bayram e Şeker (2017) para identificar falhas dentro de um motor elétrico, os autores coletaram dados de componentes de baixa frequência, incluindo harmônicos mais altos de frequência rotacional e componentes de alta frequência, que indicam operações defeituosas do motor.

Após detectar todas as frequências de falha ocultas nos dados artificiais, a abordagem é então aplicada aos dados experimentais para ver se ela pode ser aceita como uma ferramenta de detecção de falha potencial adequada para motores elétricos saudáveis. Por fim, a operação de soma algébrica (ASO), baseada na natureza redundante da transformada *wavelet* estacionária (SWT) é introduzida para detecção de falha preditiva.

Já no trabalho de Purushottam e Rajiv. (2019) uma nova metodologia para detecção de falhas elétricas e mecânicas em motores de indução é proposta. Para isso, os autores transformam o sinal de entrada do domínio do tempo (vibração radial, axial, tangencial e corrente trifásica) em sinais de frequência de tempo usando transformada *wavelet* contínua (CWT). Em seguida, os recursos apropriados são selecionados usando o modelo de *wrapper*. Esses recursos são enviados à SVM para detectar se ocorreu um defeito. A detecção de falha é realizada para velocidade usando várias *wavelets*.

Frizzo Stefenon et al. (2020) propuseram um método híbrido para previsão de falhas em isoladores de energia elétrica. Para isso, utilizaram o Coeficiente de Energia *Wavelet* (WEC) para extração de características dos dados, que consiste de um procedimento de decomposição *wavelet* ortogonal, em que os coeficientes de aproximação são divididos para obter um vetor de coeficientes de aproximação e coeficientes de detalhe em uma escala mais grosseira. Após isso,

o novo vetor de coeficientes de aproximação é dividido e o vetor de coeficientes, também é decomposto em duas partes como na divisão do vetor de aproximação.

Com isso, uma árvore binária completa é produzida na árvore unidimensional e a melhor árvore é a função de análise de pacote *wavelet*, que calcula a subárvore ótima de uma árvore inicial para um critério de tipo de entropia. Por fim, um Método de Grupo de Tratamento de Dados (GMDH) para predição de séries temporais é empregado, na qual consideram vários subconjuntos de componentes da função base, chamados modelos parciais. Os coeficientes desses modelos são estimados pelo método dos mínimos quadrados.

### 3. APRESENTAÇÃO TEÓRICA

Este capítulo apresenta modelos de previsão com técnicas estatísticas, modelos de previsão com redes neurais, tais como: rede neural *perceptron* multicamadas, rede neural autorregressiva não linear e rede neural recorrente. Também apresenta técnicas para seleção de atributos (seleção de variáveis mais importantes e seleção de atraso de tempo mais importante), tais como: coeficientes da correlação de Pearson, Kendall, Spearman, análise de componentes principais. Por fim, este capítulo apresenta conceitos sobre a transformada *wavelet* discreta e suas aplicações, teste Kaiser-Meyer-Olkin, Bartlett e os indicadores de desempenho, utilizados para medir a precisão das previsões.

#### 3.1. Modelos de previsão com técnicas estatísticas

Série temporal é uma sequência de dados distribuídos no tempo, em que uma grande classe de fenômenos (naturais, econômicos, engenharia, etc.), são produzidas a partir de observações e, conseqüentemente, quantificação numérica (SOUZA, 1981).

Uma série temporal exhibe uma ampla variedade de características, e a apreciação delas é essencial para entender suas propriedades e sua evolução (ROITMAN, 2001; CHATFIELD, 2003; MORETTIN e TOLOI, 2006; BOX e JENKINS, 2015; MILLS, 2019).

A forma mais simples de demonstrar uma série temporal é aquela que interpreta  $X_t$ ;  $t = 1, 2, \dots, N$ , como sendo um conjunto de observações discretas observadas em tempos equidistantes, apresentando uma dependência serial entre as mesmas (SOUZA, 1981).

Por conseqüência, importante uso e objetivo das séries temporais são valores não observados, previsão de valores futuros (WEI, 2006, MILLS, 2019). Prever valores de uma série temporal observada é um problema importante em diversas áreas, incluindo economia, planejamento de produção, previsão de vendas, previsão de energia, controle de estoque, entre outras (CHATFIELD, 2003).

As previsões são declarações condicionais sobre o futuro com base em premissas específicas e, basicamente, existem dois aspectos importantes na previsão: o fornecimento da previsão para um valor futuro da série e o fornecimento de erro relacionado a determinada previsão (CHATFIELD, 2003; MILLS, 2019).

No estudo da análise de séries temporais  $X_t$ , e seja  $(X_{t-1}, X_{t-2}, \dots, X_{t-d})$ , realizações passadas do processo  $X_t$ , conhecido também como série histórica ou dados, em que  $d$  é o

horizonte de atraso. Existe interesse na determinação das relações de dependência temporal da série  $X_t$  através de uma análise estatística de sua série histórica  $X_{t-1}, X_{t-2}, \dots, X_{t-d}$  (SOUZA, 1981).

Após determinar as relações de dependência em  $X_t$ , é possível determinar os prováveis valores que assumirão as variáveis aleatórias futuras  $X_{t+1}, X_{t+2}, \dots, X_{t+p}$ ; sendo  $p$  o horizonte de previsão. Para a previsão de  $X_{t+j}$  feita no instante  $t$  será usado a notação:  $X_t(j)$ ;  $j = 1, 2, \dots, p$  (SOUZA, 1981).

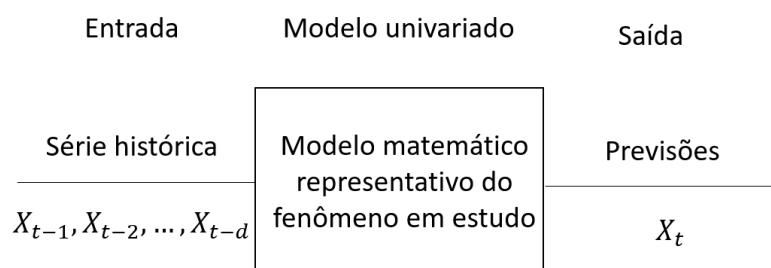
Existe uma grande variedade de procedimentos de previsão, contudo, as previsões não são generalizadas, e o analista deve estar preparado para modificá-las, conforme necessário, a qualquer informação externa, pois nenhum método pode ser utilizado de forma universal (CHATFIELD, 2003).

Segundo Roitman (2001), existem vários métodos estatísticos para modelar as séries temporais com o objetivo de previsão, tais como: Modelos Autorregressivos-Integrados-Médias Móveis (Sazonal Multiplicativo ou ARIMA, *Airline*), Modelos de Alisamento Exponencial (Sazonal e Winters – Aditivo e Multiplicativo), Modelos de Tendência Linear (com erros autorregressivos, com termos sazonais), entre outros.

De acordo com Roitman (2001) e Chatfield (2009), os métodos de previsão podem ser classificados em previsões subjetivas, univariadas e multivariadas. As previsões subjetivas são feitas usando julgamento, intuição, conhecimento comercial e qualquer outra informação relevante.

As previsões univariadas (Figura 1), chamados de métodos ingênuos ou de projeção, são aquelas previsões que se baseiam somente na informação referente à série temporal em estudo no estabelecimento do modelo de previsão, como por exemplo, as previsões de vendas futuras de determinado produto são baseadas inteiramente em vendas passadas.

**Figura 1 – Representação de um modelo univariado**



Fonte: Adaptado de Souza (1981)

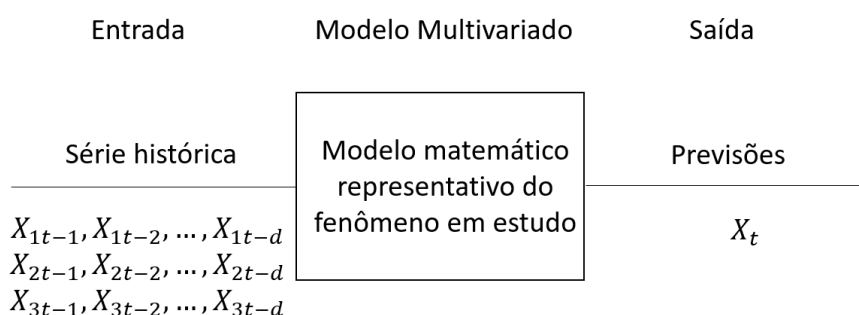
Um simples exemplo de previsão univariada é o modelo autorregressivo (AR), também conhecido como processo de Markov, possui funções de autocorrelação (FAC), que é uma maneira de medir e explicar a associação entre observações em uma série de tempo. Por exemplo, se o modelo é AR(1), o valor atual é baseado no valor anterior, se é AR(2) o valor atual é baseado nos dois valores anteriores, e assim por diante (MORETTIN E TOLOI, 2006; ADHIKARI E AGRAWAL, 2013).

Por outro lado, em previsões multivariadas (Figura 2), utiliza-se modelo de regressão linear múltipla, a variável de resposta depende dos valores de uma ou mais séries, denominadas variáveis preditivas ou explicativas. Por exemplo, as previsões de vendas, que podem depender de estoques ou índices econômicos (ROITMAN, 2001; CHATFIELD, 2009).

Neste caso, o estudo de  $X_t$ , depende de séries históricas, neste exemplo, três variáveis históricas e seus respectivos atrasos de tempo, chamados de  $X_{1t}$ ,  $X_{2t}$  e  $X_{3t}$ ,  $(X_{1t-1}, X_{1t-2}, \dots, X_{1t-d}; X_{2t-1}, X_{2t-2}, \dots, X_{2t-d}; X_{3t-1}, X_{3t-2}, \dots, X_{3t-d})$ . Assim, existe interesse na determinação das relações de dependência temporal da série  $X_t$  através de uma análise estatística de sua série histórica (SOUZA, 1981).

Após determinar as relações de dependência em  $X_t$ , é possível determinar os prováveis valores que assumirão as variáveis aleatórias futuras  $X_{t+1}, X_{t+2}, \dots, X_{t+p}$ ; sendo que os valores futuros são determinados por um modelo de função de transferência obtido pela análise das séries de entrada e saída (SOUZA, 1981).

**Figura 2 – Representação de um modelo multivariado**



**Fonte: Adaptado de Souza (1981)**

A autorregressão vetorial (VAR) é técnica utilizada em series temporais multivariadas, seu modelo utiliza processo estocástico para capturar as interdependências lineares entre várias séries temporais (CHATFIELD, 2003, WEI, 2006, MILLS, 2019).

O modelo VAR oferece a ideia de regressão automática univariada para regressões temporais de séries temporais, em que os valores defasados de todas as séries exibidas como regressivas são mostrados como regressores. Neste modelo, existe regressão de um vetor de variáveis de séries temporais em vetores atrasados dessas variáveis (CHATFIELD, 2003, WEI, 2006, MILLS, 2019).

O uso de técnicas estatísticas tradicionais de séries temporais requer experiência e habilidade consideráveis para selecionar o tipo apropriado de modelo para determinado conjunto de dados. Um problema nessas abordagens é definir tendências e sazonalidades de forma satisfatória; além disso, é importante garantir o tipo certo de tendência (linear ou não linear, local ou global), pois esses modelos exigem a especificação do modelo de série temporal assumido (LEWIS, 2017).

### **3.2. Modelos de previsão com redes neurais**

Os modelos que utilizam redes neurais não precisam especificar a natureza exata do relacionamento entre a entrada e a saída, (linear, não linear, sazonalidade, tendência), já que as camadas ocultas podem aproximar funções de decisão extremamente complexas, eliminando a necessidade de pré- especificação da geração de dados (HAYKIN, 1999; ROITMAN, 2001; LEWIS, 2017). Contudo, a rede neural precisa ser dinâmica e ter memória, ou seja, entradas atrasadas no tempo ou laços de realimentação (recorrentes) (ROITMAN, 2001).

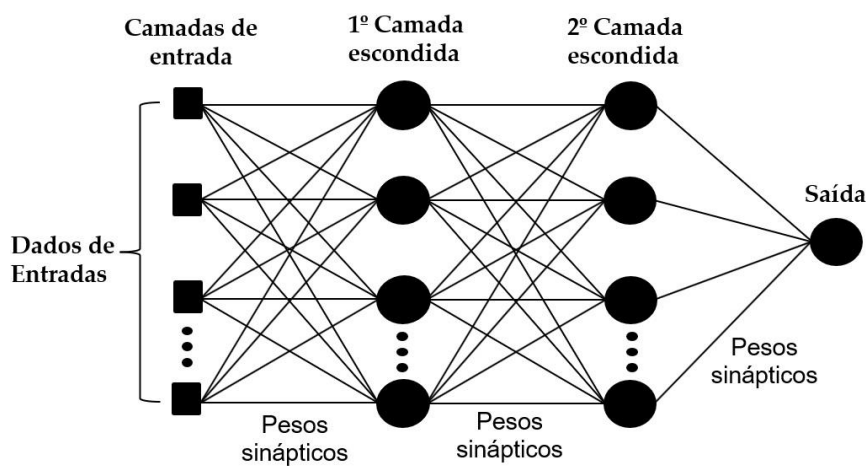
Nos últimos anos, as redes neurais artificiais (RNAs) foram aplicadas em diversas áreas e constituem ferramenta de inteligência artificial (IA) incorporada na modelagem da inteligência humana, que trabalha em paralelo e cujo conhecimento é adquirido pela aprendizagem, as sinapses ponderadas são usadas para o conhecimento e sua principal característica é armazenar o conhecimento experimental a ser utilizado (HAYKIN, 1999).

#### **3.2.1.1. Rede Neural *Perceptron* Multicamadas**

Um *perceptron* multicamada é uma rede neural artificial que pode trabalhar com várias superfícies de decisão não lineares, usa várias regras de aprendizado e tem alto grau de conectividade através dos pesos sinápticos (HAYKIN, 1999). A Figura 3 mostra a estrutura de uma rede MLP, que consiste em uma camada de entrada, duas camadas ocultas e uma camada de saída.

Neste tipo de rede neural o sinal de entrada se propaga em uma direção única, camada por camada. A camada de entrada receberá dados de entrada da rede, a primeira camada oculta é alimentada a partir da camada de entrada que, por sua vez, é constituída pelos nós de origem, as saídas resultantes da primeira camada oculta são aplicadas como entrada para a segunda camada oculta. Por fim, a camada de saída tem, então, a função de processar as informações (GUNAY, 2016).

**Figura 3 - Arquitetura da Rede Multicamadas**



Uma propriedade importante das redes neurais é sua capacidade de aprender com o ambiente e, por meio do treino, melhorar seu desempenho. Para isso, utiliza regra de aprendizagem supervisionada, que fornece um conjunto de exemplos do comportamento da rede. Nesse tipo de aprendizagem, há um professor com conhecimentos sobre o ambiente, que fornece para a rede neural a resposta desejada (HAYKIN, 2008; HAGAN et al., 2014).

Uma aprendizagem supervisionada que vem sendo muito utilizada em rede MLP é o algoritmo *backpropagation* que busca, a cada iteração, processar todos os exemplos de treino e compara o valor predito na saída da rede ao valor real. Esse algoritmo pode trabalhar com várias superfícies de decisão não lineares e é adequado quando as instâncias são representadas por muitos atributos (HAYKIN, 2008 e MITCHELL, 1997).

### **3.2.1.2. Rede neural autorregressiva não linear**

Na maioria dos casos, as aplicações de séries temporais são caracterizadas por grandes variações e períodos transitórios, dificultando a modelagem de séries temporais usando um

modelo linear. Uma rede neural autorregressiva não linear (NAR), pode ser sugerida para resolver este problema de modelagem de séries temporais (RUIZ et al., 2016).

A rede NAR permite a previsão de valores futuros de uma série temporal, apoiada em seu histórico de valores, utilizando um mecanismo de realimentação, no qual um valor previsto pode servir como entrada para novas previsões. A rede é criada e treinada em malha aberta, utilizando os valores reais alvo como *feedback*, após o treinamento, a rede é convertida em um circuito fechado e os valores previstos são usados para fornecer novas entradas de *feedback* para a rede (PEREIRA et al., 2018).

A topologia de uma rede NAR (Equação 1) usada para prever o valor de  $X_t$ , depende de atrasos de tempo desta mesma variável, neste caso,  $X_{t-1}, X_{t-2}, \dots, X_{t-d}$ , são chamados de atrasos de *feedback*,  $f$  é uma função não linear, onde os valores futuros dependem apenas dos valores *anteriores* do sinal de saída. Ao usar uma rede NAR, a rede de loop fechado é usada para executar uma previsão de várias etapas (Equação 2) (RUIZ et al., 2016).

$$X_t = f(X_{t-1} + X_{t-2} + \dots + X_{t-d}), \quad (1)$$

$$X_{t+p} = f(X_{t-1} + X_{t-2} + \dots + X_{t-d}), \quad (2)$$

### 3.2.1.3. Rede neural recorrente

Uma classe alternativa de modelos para sistemas dinâmicos de tempo discretos controlados não lineares determinísticos é a de modelos baseados em redes neurais no espaço de estados, chamadas de recorrentes (TIUMENTSEV E EGORCHEV, 2019).

A rede neural recorrente assume muitas formas diferentes, cada uma realçando uma forma específica de realimentação global. No entanto, todas incorporam um *perceptron* de múltiplas camadas estático ou partes dele e exploram a capacidade de mapeamento não linear do *perceptron* de múltiplas camadas (HAYKIN, 1999).

Recentemente, as redes de *deep learning* surgiram com grande sucesso, usando abordagem eficaz com a capacidade de aprender grande quantidade de dados de maneira não supervisionada e fazer previsões com maior precisão. A ideia principal da rede *deep learning* é resolver problemas intuitivos, permitindo que os computadores aprendam a experiência e, assim, entendam hierarquias de conceitos mais complexos a partir de conceitos mais simples.

A experiência adquirida por esse tipo de rede significa que não há necessidade de criar estruturas de dados, mesmo em problemas com alto grau de complexidade (GOODFELLOW et al., 2016).

A rede neural *long short-term memory* (LSTM) é considerada um tipo de rede neural recorrente (RNN), também chamadas de rede neurais empilhadas, e foi desenvolvida por Hochreiter e Schmidhuber em 1997. Essa rede é capaz de capturar dependências de longo prazo em sequência e foram utilizadas com sucesso em muitos problemas, incluindo controle adaptativo, identificação do sistema, no reconhecimento de fala, entre outros. Esse modelo é organizado em células dotadas de um estado interno que pode ser passado para outras células adjacentes, que incluem várias operações (NIELSEN, 2017; LEWIS, 2017).

### 3.3. Técnicas para seleção de atributos

Na estatística, correlação é a análise bivariada que mede os pontos fortes da associação entre duas variáveis (ZAID, 2015; MIRKIN, 2011; COHEN et al., 2003). Uma correlação varia entre +1 e -1, sendo que, quando positiva, indica até que ponto essas variáveis aumentam ou diminuem em paralelo; enquanto a correlação negativa indica até que ponto a variável aumenta, à medida que a outra diminui (ZAID, 2015; COHEN et al., 2003).

Quando o valor do coeficiente de correlação está em torno de  $\pm 1$ , é considerado um perfeito grau de associação entre as duas variáveis. À medida que o valor do coeficiente de correlação for para 0, a relação entre as duas variáveis será mais fraca (ZAID, 2015; COHEN et al., 2003).

Caso haja relação entre as variáveis, ela pode ser usada de várias maneiras, das quais dois tipos de aplicação são tipicamente discerníveis, a primeira para previsão de valores de uma variável em relação às outras, a segunda, adicionando a relação com o conhecimento do domínio, interpretando e explicando-o em termos do conhecimento existente (MIRKIN, 2011).

Em estatística, geralmente, são medidos três tipos de correlações: correlação de Pearson, correlação de Kendall e correlação de Spearman. No entanto, existem outras técnicas que são utilizadas para verificar a relação entre duas ou mais variáveis, como: métodos baseados em filtros, métodos *wrapper*, métodos incorporados, método de regressão dos mínimos ângulos (LARS), método de *shrinkage* e seleção de atributos para modelos de regressão linear (LASSO), PCA, *Wavelet*, LR, MARS, floresta aleatória, efeito recente, recursos por especialistas baseado em um conjunto de regras *if-then*, recursos de informações mútuas, técnicas heurísticas, tais como: algoritmo genético (GA), *Simulated annealing* (SA),

Otimização por colônia de formigas (ACO), *tabu search* (TS) e otimização por enxame de partículas (PSO), entre outras.

### 3.3.1. Coeficientes da correlação de Pearson, de Kendall, de Spearman

O coeficiente da correlação de Pearson foi criado pelo matemático Karl Pearson em 1905. Basicamente, é um índice adimensional com valores situados entre -1 e 1, que reflete a intensidade de uma relação linear entre dois conjuntos de dados. Se o valor se aproxima de +1, a força da associação entre as duas variáveis aumenta, mostrando que estão na mesma direção. Por outro lado, se aproxima de -1, as duas variáveis estão se movendo na direção oposta (PEARSON, 1905; COHEN et al., 2003; GUTTMAN e JAYALATH, 2020).

Coeficiente de correlação tau de Kendall foi desenvolvida por Maurice George Kendall em 1938. É considerado um teste não paramétrico que mede a força da dependência entre duas variáveis que, basicamente, verifica a semelhança entre as ordens dos dados quando classificados por cada uma das quantidades.

O coeficiente de correlação de postos de Spearman, ou  $r_s$  de Spearman, foi criado pelo estatístico Charles Spearman em 1904. Também é um teste não paramétrico usado para medir o grau de associação entre duas variáveis, e não assume nenhuma suposição sobre a distribuição dos dados.

### 3.3.2. Análise de componentes principais

A análise de componentes principais (PCA) teve sua primeira aparição nos Estados Unidos, como um exercício de álgebra, quando Silvester (1889) considerou o problema da redução de uma matriz quadrada em sua decomposição de valor singular. Uma década depois, nas mãos de Pearson (1901), PCA ressurgiu em um contexto biológico, quando ele reformulou a análise de regressão linear em uma nova forma de modo a evitar a relação assimétrica usual entre variáveis dependentes e independentes em regressões lineares.

PCA é um método estatístico que usa transformação ortogonal para converter um conjunto de observações de variáveis correlacionadas em um conjunto de valores de variáveis lineares não correlacionadas, ou explicar a matriz de variância-covariância de um conjunto de variáveis  $m$ -dimensionais por meio de algumas combinações lineares dessas variáveis (PEARSON, 1901, WEI, 2006).

Esse método é usado para reduzir a dimensionalidade de grandes conjuntos de dados de maneira não supervisionada, transformando amplo conjunto de variáveis em um modelo menor, que preserva o máximo de informações, preservando o máximo de 'variabilidade' (ou seja, informações estatísticas) possível. Isso significa que preservar o máximo de variabilidade possível, e encontrar novas variáveis que são funções lineares daquelas no conjunto de dados original, que sucessivamente maximizam a variância (WEI, 2006; SCHIMIT e PEREIRA, 2018).

Basicamente, PCA segue as seguintes etapas para redução da dimensão de um conjunto de dados e/ou redução de variáveis. A primeira etapa é a padronização dos dados, na qual os dados são escalados de tal forma que todas variáveis e seus valores estejam dentro de uma faixa semelhante. A segunda etapa é a computação da matriz de covariância, que expressa a correlação entre as diferentes variáveis no conjunto de dados (JOLLIFFE, 2002).

A terceira etapa está relacionada ao cálculo dos autovetores da matriz de covariância para identificar os principais componentes. Os autovetores da matriz de covariância são as direções dos eixos onde há mais variância dos dados, também chamados de componentes principais. Já os autovalores são simplesmente os coeficientes ligados aos autovetores, que dão a quantidade de variância transportada em cada componente principal (JOLLIFFE, 2002).

A quarta etapa está relacionada a computação dos principais componentes, o vetor de características, utilizado para isso, é simplesmente uma matriz que tem como colunas os autovetores dos componentes que decidimos manter. Por fim, a última etapa reformula os dados ao longo dos eixos dos componentes principais (JOLLIFFE, 2002).

### **3.3.2.1. Definição e derivação das componentes principais**

Sabendo que  $X$  é um vetor de  $p$  variáveis aleatórias, que as variâncias das  $p$  variáveis aleatórias e a estrutura das covariâncias ou correlações entre as  $p$  variáveis sejam de interesse. Existe interesse em procurar variáveis derivadas com ( $< p$ ) que preservem a maioria das informações fornecidas por essas variâncias e correlações ou covariâncias (JOLLIFFE, 2002).

Embora o PCA não ignore as covariâncias e correlações, ele se concentra nas variações. Sendo assim, o primeiro passo é procurar uma função linear  $\alpha_1'X$  (Equação 3) dos elementos de  $X$  com máxima variância, onde  $\alpha_1$  é um vetor de  $p$  constantes  $\alpha_{11}, \alpha_{12}, \dots, \alpha_{1p}$ , e  $'$  denota transposição, de modo que:

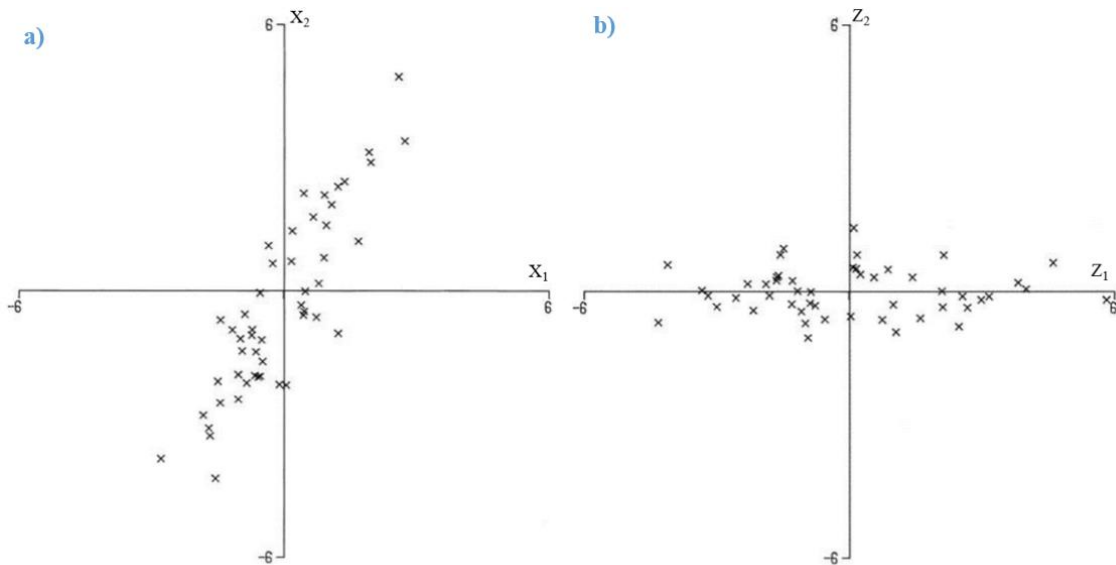
$$\alpha'_1 X = \alpha_{11} X_1 + \alpha_{12} X_2 + \dots + \alpha_{1p} X_p = \sum_{j=1}^p \alpha_{1j} X_j \quad (3)$$

Em seguida, procure uma função linear  $\alpha'_2 X$ , não correlacionada com  $\alpha'_1 X$  tendo variância máxima, e assim por diante, de modo que  $k$ -ésimo estágio, uma função linear  $\alpha'_k X$  seja encontrada, de tal forma que tenha variância máxima sujeita a não estar correlacionada com  $\alpha'_1 X, \alpha'_2 X, \dots, \alpha'_{k-1} X$ . A  $k$ -ésima variável derivada,  $\alpha'_k X$ , é a  $k$ -ésima componente principal. Até  $p$  componentes principais serem encontrados, mas espera-se, em geral, que a maior parte da variação em  $X$  seja contabilizada por  $m$  componentes principais, em que  $m < p$ .

A Figura 4 mostra um gráfico de 50 observações em duas variáveis altamente correlacionadas  $X_1, X_2$  e as duas componentes principais  $Z_1$  e  $Z_2$ . É possível analisar que existe uma variação considerável em ambas as variáveis, embora um pouco mais na direção de  $X_2$  do que  $X_1$ . Se transformarmos as componentes principais em  $Z_1$  e  $Z_2$ , é possível verificar que há uma variação maior na direção de  $Z_1$ , do que em qualquer uma das variáveis originais, mas pouca variação na direção de  $Z_2$ .

De forma mais geral, se um conjunto de  $p (> 2)$  variáveis tiver correlações substanciais entre elas, os primeiros componentes principais serão responsáveis pela maior parte da variação nas variáveis originais. Por outro lado, os últimos componentes principais identificam direções nas quais há pouca variação; isto é, eles identificam relações lineares quase constantes entre as variáveis originais.

**Figura 4 – Gráfico com 50 observações. a) Duas variáveis  $X_1, X_2$ . b) Duas componentes principais  $Z_1, Z_2$  das variáveis  $X_1, X_2$**



**Fonte: Jolliffe, 2002**

Com a definição das componentes principais é necessário encontrá-los. Assim, considere, um vetor de variáveis aleatórias  $X$  tem uma matriz de covariância conhecida  $\Sigma$ . Esta é a matriz cujo  $(i, j)$ -ésimo elemento é a covariância (conhecida) entre os  $i$ -ésimo e  $j$ -ésimo elementos de  $X$  quando  $i = j$ , e a variância do  $j$ -ésimo elemento de  $X$  quando  $i = j$ . Em um caso mais realista, onde  $\Sigma$  é desconhecido, segue substituindo  $\Sigma$  por uma matriz de covariância de amostra  $S$ . Acontece que para  $k = 1, 2, \dots, N$ , o  $k$ -ésimo componente principal é dado por  $Z_k = \alpha_k' X$ , onde  $\alpha_k$  é um autovetor de  $\Sigma$  correspondendo ao seu  $k$ -ésimo maior autovalor  $\lambda_k$ . Além disso, se  $\alpha_k$  for escolhido para ter comprimento unitário ( $\alpha_k' \alpha_k = 1$ ), então  $var(Z_k) = \lambda_k$ , onde  $var(Z_k)$  denota a variância de  $Z_k$ .

Para derivar a forma das componentes principais, considere primeiro  $\alpha_1' X$ ; o vetor  $\alpha_1$  maximiza  $var(\alpha_1' X) = \alpha_1' \Sigma \alpha_1$ . É claro que, do jeito que está, o máximo não será alcançado para  $\alpha_1$  finito, então uma restrição de normalização deve ser imposta. A restrição usada na derivação é  $\alpha_1' \alpha_1 = 1$ , ou seja, a soma dos quadrados dos elementos de  $\alpha_1$  é igual a 1. Outras restrições, por exemplo  $Max_j |\alpha_{1j}| = 1$ , pode ser mais útil em outras circunstâncias e pode ser facilmente substituído. No entanto, o uso de outras restrições além de  $\alpha_1' \alpha_1 = constante$  na derivação leva a um problema de otimização mais difícil e irá produzir um conjunto de variáveis derivadas diferentes dos componentes principais.

Para maximizar  $\alpha_1' \Sigma \alpha_1$  sujeito a  $\alpha_1' \alpha_1 = 1$ , a abordagem padrão é usar a técnica dos multiplicadores de Lagrange. Maximizar  $\alpha_1' \Sigma \alpha_1 - \lambda(\alpha_1' \alpha_1 - 1)$ , em que  $\lambda$  é um multiplicador de Lagrange. A diferenciação em relação  $\alpha_1$  é dada por  $\Sigma \alpha_1 - \lambda \alpha_1 = 0$ , ou  $(\Sigma - \lambda I_p) \alpha_1 = 0$ ,  $I_p$  é a matriz identidade ( $p \times p$ ). Assim,  $\lambda$  é um autovalor de  $\Sigma$  e  $\alpha_1$  é o autovetor correspondente. Para decidir qual dos  $p$  autovetores dá  $\alpha_1' X$  com variância máxima, observe que a quantidade a ser maximizada é  $\alpha_1' \Sigma \alpha_1 = \alpha_1' \lambda \alpha_1 = \lambda \alpha_1' \alpha_1 = \lambda$ . Então,  $\lambda$  deve ser o maior possível. Assim,  $\alpha_1$  é o autovetor correspondente ao maior autovalor de  $\Sigma$ , e  $var = \alpha_1' X = \alpha_1' \Sigma \alpha_1 = \lambda_1$ , o maior autovalor.

Em geral, o  $k$ -ésimo componente principal de  $X$  é  $\alpha_k' X$  e  $var(\alpha_k' X) = \lambda_k$ , onde  $\lambda_k$  é  $k$  maior autovalor de  $\Sigma$ , e  $\alpha_k$  é o autovetor correspondente.

O segundo componente principal,  $\alpha_2' X$ , maximiza  $\alpha_2' \Sigma \alpha_2$  sujeito a não ser correlacionado com  $\alpha_1' X$ , ou equivalentemente sujeito a  $cov(\alpha_1' X, \alpha_2' X) = 0$ , onde  $cov(x, y)$  denota a covariância entre as variáveis aleatórias  $x$  e  $y$ . Mas  $cov(\alpha_1' X, \alpha_2' X) = \alpha_1' \Sigma \alpha_2 = \alpha_2' \Sigma \alpha_1 = \alpha_2' \lambda_1 \alpha_1 = \lambda_1 \alpha_2' \alpha_1 = \lambda_1 \alpha_1' \alpha_2$ . Assim, qualquer uma das equações  $\alpha_1' \Sigma \alpha_2 = 0$ ,  $\alpha_2' \Sigma \alpha_1 = 0$ ,  $\alpha_1' \alpha_2 = 0$  e  $\alpha_2' \alpha_1 = 0$  poderia ser usado para especificar correlação zero entre  $\alpha_1' X$  e  $\alpha_2' X$ .

Escolhendo o último deles (de forma arbitrária), e observando que uma restrição de normalização é novamente necessária, a quantidade a ser maximizada é  $\alpha_2' \Sigma \alpha_2 = \lambda(\alpha_2' \alpha_2 - 1) - \phi \alpha_2' \alpha_1$ , onde  $\lambda, \phi$  são multiplicadores de Lagrange. A diferenciação em relação a  $\alpha_2$  gera  $\Sigma \alpha_2 - \lambda \alpha_2 - \phi \alpha_1 = 0$  e a multiplicação da equação anterior por  $\alpha_1$  gera  $\alpha_1' \Sigma \alpha_2 - \lambda \alpha_1' \alpha_2 - \phi \alpha_1' \alpha_1 = 0$ , na qual os dois primeiros termos são zero e  $\alpha_1' \alpha_1 = 1$ , se reduz a  $\phi = 0$ .

Portanto,  $\Sigma \alpha_2 - \lambda \alpha_2 = 0$ , ou equivalentemente  $(\Sigma - \lambda I_p) \alpha_2 = 0$ , então  $\lambda$  é mais uma vez um autovalor de  $\Sigma$ , e  $\alpha_2$  o autovetor correspondente. Novamente,  $\lambda = \alpha_2' \Sigma \alpha_2$ , então  $\lambda$  deve ser o maior possível, assumindo que  $\Sigma$  não tem autovalores repetidos, portanto,  $\lambda$  não pode ser igual a  $\lambda_1$ . Se assim for,  $\alpha_2 = \alpha_1$ , violando a restrição  $\alpha_1' \alpha_2 = 0$ . Portanto,  $\lambda$  é o segundo maior autovalor de  $\Sigma$ , e  $\alpha_2$  é o autovetor correspondente.

Também pode ser mostrado que para o terceiro, quarto, ...,  $Z$ -ésimo componentes principais, os vetores dos coeficientes  $\alpha_3, \alpha_4, \dots, \alpha_p$  são os autovetores de  $\Sigma$  correspondentes a  $\lambda_3, \lambda_4, \dots, \lambda_p$ , o terceiro e o quarto maiores, e o menor autovalor, respectivamente. Além disso,  $var(\alpha_k' X = \lambda_k)$  para  $k = 1, 2, \dots, N$ .

Mais informações sobre os autovalores e autovetores, bem como outras formas de representação da análise de componentes principais, podem ser encontradas em Jolliffe, 2002.

### 3.4. Introdução sobre a transformada *wavelet*

A palavra *wavelet*, também conhecida como “*Ondaleta*”, teve origem da palavra francesa “*Ondalette*”, que significa onda pequena. A decomposição de uma função com o uso de *wavelets* é conhecida como transformada *wavelet* (WT) (ADDISON, 2002; GUTIERREZ, 2002; OLKKONEN, 2011).

Durante a última década ou mais, a análise de transformada *wavelet* emergiu como uma nova ferramenta, permitindo a decomposição de sinais no domínio do tempo-frequência para análise de dados (ADDISON, 2002; DA SILVEIRA e KOZAKEVICIUS).

A transformada *wavelet* evoluiu com o tempo, em 1909, por exemplo, Alfred Haar propôs a análise escalonada, chamada de base de Haar, que é uma família de *wavelets* ortogonais de suporte compacto. Denis Gabor criou uma família de *wavelets* não ortogonais de suporte ilimitado com gaussianas transladadas em 1946.

Em 1984, Jean Morlet e Alex Grossmann propuseram o conceito e uso de *wavelets* para modelagem de sinais sísmicos. Ao passo que Stéphane Mallat e Yves Meyer, entre 1987 e 1993, criaram a análise de multiresolução baseada em transformada *wavelet* discreta. Em 1988, Ingrid Daubechies e Ronald Coifman criaram as famílias ortogonais de suporte compacto por métodos iterativos. Por outro lado, Albert Cohen, Ingrid Daubechies e Jean Feauveau propuseram *wavelets* biortogonais em 1992 (DE OLIVEIRA, 2007).

As funções da *wavelet*  $\psi_{a,b}(t)$ , vista na Equação 4 são obtidas por intermédio de translações, que são introduzidas por meio de um parâmetro  $b$ , que indica que a função  $\psi(t)$  foi transladada no eixo  $t$  de uma distância equivalente a  $b$  e dilatações, que são introduzidos por um fator de escalamento  $a$  de uma função chamada *wavelet* mãe  $\psi(t)$  (PEREIRA, 2007; OLKKONEN, 2011; SUNDARARAJAN, 2015).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (4)$$

Em que:  $b \in \mathbb{R}$  é um parâmetro de translação, enquanto  $a \in \mathbb{R}^+$  ( $a \neq 0$ ) é um parâmetro de dilatação ou escala. O fator  $a^{-1/2}$  é uma constante de normalização de modo que o valor fornecido por meio da integrabilidade quadrada de  $\psi_{a,b}$  é a mesma para todas as escalas  $a$ . Nota-se que o parâmetro de escala  $a$  na Equação 4 rege as dilatações da variável independente

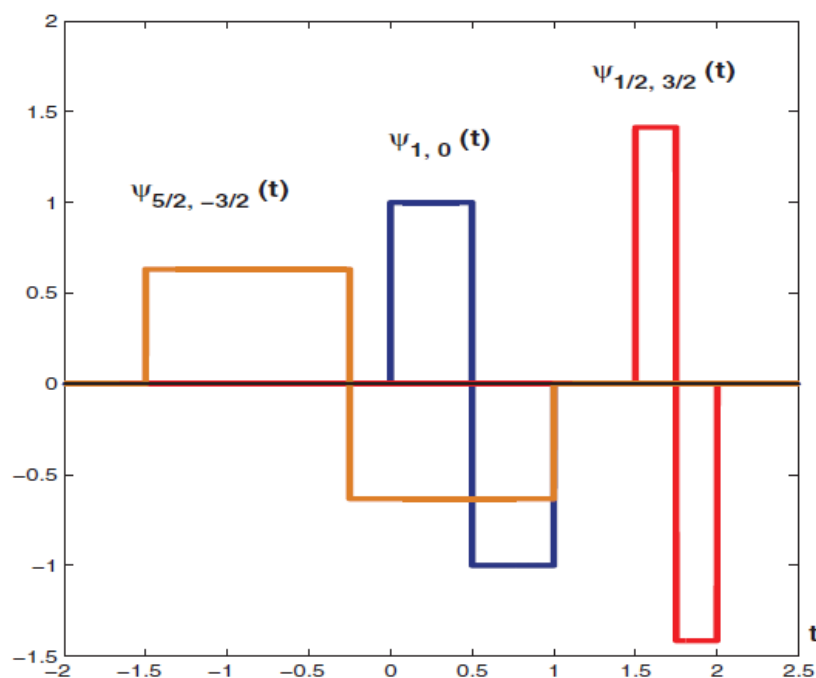
$(t - b)$ . Da mesma forma, o fator  $a^{-1/2}$  rege a dilatação nos valores tomados por  $\psi$  (Figura 5). (OLKKONEN, 2011).

O termo  $\frac{1}{\sqrt{a}}$  é um fator de normalização que garante que a energia de  $\psi_{a,b}(t)$  seja independente de  $a$  e de  $b$ , ou seja, preserva a norma da função no espaço  $L^2(\mathbb{R})$  (OLKKONEN, 2011)

As funções definidas na Equação 4 formam uma base para  $L^2(\mathbb{R})$ , fazendo com que uma função  $f(t)$  nesse espaço, possa ser aproximada por uma combinação das *wavelets* usando coeficientes  $C_{a,b}$  dados pela Equação 5 (PEREIRA, 2007).

$$C_{a,b} = \langle f, \psi_{a,b} \rangle. \quad (5)$$

**Figura 5 – A função *wavelet* de Haar para vários valores do parâmetro de escala  $a$  e parâmetro de translação  $b$ . Se  $a < 1$ , a função *wavelet* é contraída, e se  $a > 1$ , a *wavelet* é expandida.**



Fonte: Olkkonen, 2011

As *wavelets* estão sendo utilizadas em diversas áreas, como por exemplo; mecânica de fluídos, computação numérica, análise de imagens, processamento de sinais, séries temporais, sistemas de controle, fenômenos biológicos, medicina, psicologia, entre outras (ADDISON, 2002; PERCIVAL e WALDEN, 2006; GUTIERREZ, 2002).

Um problema fundamental com os dados na forma de séries temporais, por exemplo, é como processá-los a fim de extrair informações corretas e significativas. Além disso, muitas séries temporais, como séries econômicas, financeiras, séries relacionadas a eletricidade, entre outras, exibem comportamentos não estacionários, tais como mudanças nas tendências. Assim, a transformada *wavelet* surge como uma ferramenta muito útil para analisar séries temporais do tipo não estacionária (GUTIERREZ, 2002; OLKKONEN, 2011).

Além disso, a WT pode ser usada para análise de correlação e coerência entre duas séries temporais, análise espectral cruzada e *wavelets* de verificação com base no espaço (RHIF et al., 2019).

### 3.4.1. Transformada de *wavelet* discreta

Uma família de *wavelets* discretas é obtida introduzindo parâmetros de dilatação  $a$  e fatores de translação  $b$  discretos, ou seja, é necessário discretizar os parâmetros  $a$  e  $b$  da Equação 6.

Aqui pode-se observado que  $\psi_{jk}(x)$  (Equação 3) é obtido de  $\psi(x)$  por uma dilatação binária  $2^{-j}$  e uma translação “diádica”  $k2^j$ . Sendo que  $a = 2^j$  e  $b = k2^j$ , em que  $j, k \in \mathbb{Z}$  como é usual, as *wavelets* discretas adquirem a seguinte forma:

$$\psi_{jk}(x) = 2^{-\frac{j}{2}}\psi(2^{-j}x - k). \quad (6)$$

Uma observação importante é que a translação depende do passo de dilatação  $a$ . Isso acontece por que as *wavelets* permitem uma cobertura do plano tempo-frequência com diferentes resoluções para baixas e altas frequências. Na verdade, a transformada utiliza *wavelets* mais comprimidas de curta duração na análise em frequências mais altas, o que permite a visualização do sinal analisado em uma escala com melhor resolução temporal. Já na frequência mais baixa, o uso de *wavelets* mais dilatadas permite visualizar características globais (PEREIRA, 2008).

A transformada discreta de *wavelet* (DWT) é fundamentada na decomposição do espaço  $V_n$  em subespaços mutuamente ortogonais de acordo com a Equação 7.

$$V_n = V_l \oplus W_l \oplus W_{l+1} \oplus \dots \oplus W_{n+1} \quad (7)$$

Desta forma, a função  $u \in V_n$  pode ser ampliada de acordo com a Equação 8.

$$u = \sum_k c_{n,k} \phi_{nk} \quad (8)$$

Ou neste caso, ainda pode ser usado a Equação 9.

$$u = \sum_k c_{l,k} \phi_{lk} + \sum_{j=n+1}^l \sum_k d_{j,k} \psi_{jk} \quad (9)$$

Com  $\phi_{nk}(x) = 2^{-\frac{n}{2}} \phi(2^{-n}x - k)$  e  $\psi$  como em  $\psi_{jk}(x) = 2^{-\frac{j}{2}} \psi(2^{-j}x - k)$ .

De acordo com a análise em multiresolução, a variável  $u$ , como pode ser observado na Equação 10, também pode ser decomposta em  $V_{n+1}$  e  $W_{n+1}$  da seguinte forma:

$$\begin{aligned} u &= \sum_k \langle u, \phi_{n+1,j} \rangle \phi_{n+1,j} + \sum_k \langle u, \psi_{n+1,j} \rangle \psi_{n+1,j} \\ &= \sum_j c_{n+1,j} \phi_{n+1,j} + \sum_j d_{n+1,j} \psi_{n+1,j} \end{aligned} \quad (10)$$

Em que os novos coeficientes  $c_{n+1,j}$  e  $d_{n+1,j}$  são obtidos usando a ortogonalidade das funções  $\phi, \psi$ . Uma observação importante é que os coeficientes  $c_{n+1,j}$  mostram uma representação do sinal  $u$  na base gerada por  $\phi_{n+1,j}$  para o espaço  $V_{n+1}$ . Assim, esta representação é considerada ótima se  $\|P_{n+1}u\| \leq \|u - v\|, \forall v \in V_{n+1}$ , sendo  $P_{n+1}u = \sum_j c_{n+1,j} \phi_{n+1,j}$ .

Como  $u \in V_n$  e  $u = \sum_j c_{n,j} \phi_{n,j}$  tem-se a Equação 11.

$$\begin{aligned} u - P_{n+1}u &= \sum_j c_{n,j} \phi_{n,j} - \sum_j c_{n+1,j} \phi_{n+1,j} \\ &= P_n u - P_{n+1}u, \end{aligned} \quad (11)$$

E como  $V_n = V_{n+1} \oplus W_{n+1}$ , segue que  $u - P_{n+1}u = Q_{n+1}u$ , em que  $Q_{n+1}$  é a projeção de  $u$  sobre  $W_{n+1}$ . Assim, a consequência imediata é  $u - P_{n+1}u \perp P_{n+1}u$ , em que  $P_{n+1}u$  é a projeção de  $u$  sobre o espaço  $V_{n+1}$ .

A DWT resulta em um sinal em  $V_n$ , representado pelos coeficientes da expansão na base  $\phi_{n,j}$  é decomposto em novos coeficientes que representam esse sinal em  $V_{n+1}$  e  $W_{n+1}$ . A Equação 12 mostra os novos coeficientes, em que  $l = \dim(V_{n+1})$ ,  $m = \dim(W_{n+1})$  e  $l + m = k = \dim(V_n)$ .

$$\begin{aligned}
 u &= (c_{n,1}, c_{n,2}, \dots, c_{n,k}) \\
 &= (c_{n+1,1}, c_{n+1,2}, \dots, c_{n+1,l}, d_{n+1,1}, d_{n+1,2}, \dots, d_{n+1,m})
 \end{aligned}
 \tag{12}$$

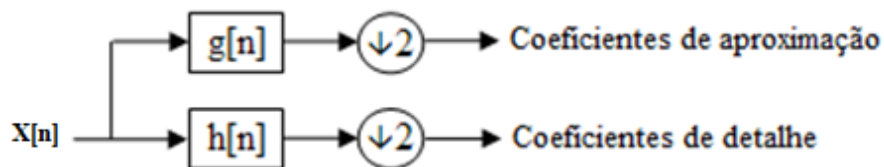
### 3.4.2. O banco de filtros da transformada *wavelet* discreta

O DWT de um sinal  $X_t$ ;  $t = 1, 2, \dots, N$  é calculado passando-o por uma série de filtros. Primeiro, as amostras são passadas por um filtro passa-baixa  $g$  e um filtro passa-alta  $h$  com resposta de impulso, resultado em uma convolução dos dois, como pode ser observado na Equação 13.

$$Y[n] = (X_t * g)[n] = \sum_{k=-\infty}^{\infty} X_t[k]g[n-k]
 \tag{13}$$

As saídas mostram os coeficientes de detalhes do filtro passa-alta e os coeficientes de aproximação a partir do filtro passa-baixa, como pode ser observado na Figura 6.

Figura 6 – Diagrama em bloco da análise do filtro



No entanto, como metade das frequências do sinal foram removidas, metade das amostras podem ser descartadas de acordo com a regra de *Nyquist*. A saída do filtro passa-baixa  $g$ , no diagrama acima, é então subamostrado por 2, e processado novamente, através de um novo filtro passa-baixa  $g$ , e um filtro passa-alta  $h$  com metade da frequência de corte do anterior, resultando nas Equações 14 e 15.

$$Y[\text{passa - baixa}] = (X_t * g) \downarrow 2
 \tag{14}$$

$$Y[\text{passa - alta}] = (X_t * h) \downarrow 2
 \tag{15}$$

Na árvore da Figura 7, conhecida como banco de filtros, esta decomposição é repetida para aumentar ainda mais a resolução de frequência, enquanto os coeficientes de aproximação decompostos com filtros passa-alta e baixa são amostrados. Isso é representado como uma árvore binária com “nós” que representam um subespaço com uma localização de tempo-frequência diferente.

Em cada nível no diagrama acima, o sinal é decomposto em frequências baixas e altas. Devido ao processo de decomposição, o sinal de entrada deve ser um múltiplo de  $2^n$  onde  $n$  é o número de níveis.

A Figura 8 mostra uma representação da DWT no domínio da frequência. Para um sinal com 32 amostras, faixa de frequência  $f_n$  e 3 níveis de decomposição, serão produzidas 4 escalas de saída.

Figura 7 – Banco de filtro de nível 3

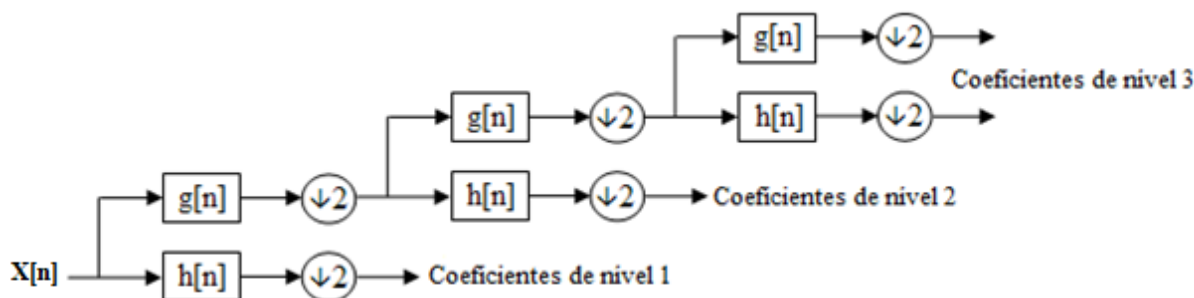
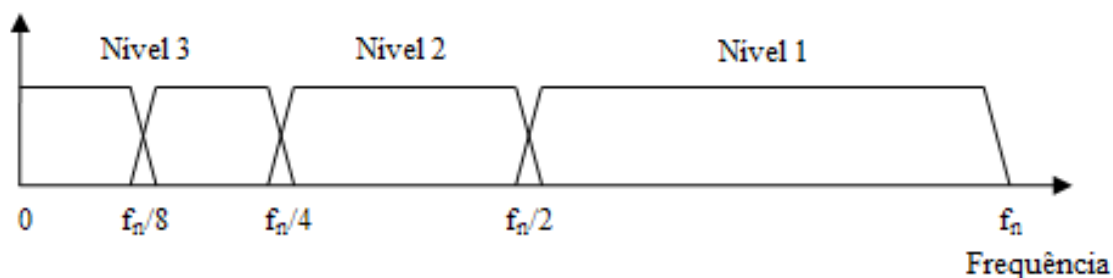


Figura 8 – Representação da DWT no domínio da frequência



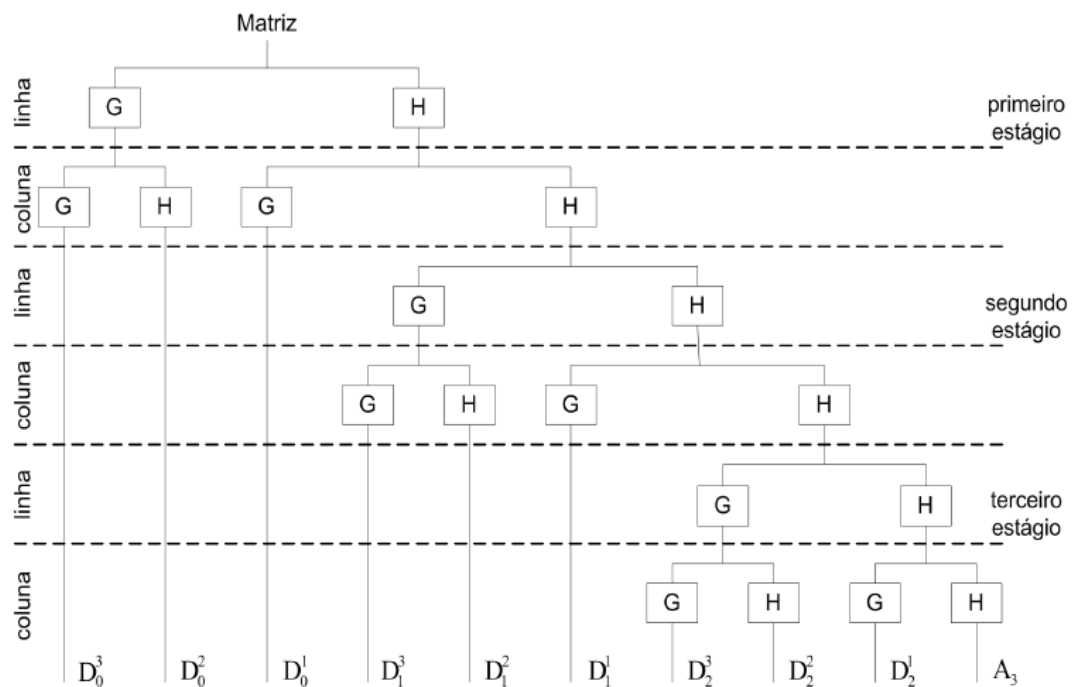
A transformada de *wavelet* discreta, também pode ser representada matricialmente. Neste caso, é definido a matriz  $R$ , que é pré-multiplicada pelo sinal discreto contendo os coeficientes dos filtros a serem utilizados. Usando como exemplo filtros de comprimento 4, a matriz ficará de acordo com a Equação 16.

$$R_{N \times N} = \begin{bmatrix} h_1 & h_2 & h_3 & h_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_1 & h_2 & h_3 & h_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ h_3 & h_4 & 0 & 0 & 0 & 0 & h_1 & h_2 \\ g_1 & g_2 & g_3 & g_4 & 0 & 0 & 0 & 0 \\ 0 & 0 & g_1 & g_2 & g_3 & g_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 \\ g_3 & g_4 & 0 & 0 & 0 & 0 & g_1 & g_2 \end{bmatrix} \quad (16)$$

Sendo  $N$  o número de amostras no sinal, as primeiras  $N/2$  linhas de  $R$  são formadas pelos coeficientes dos filtros passa-baixa  $h$  e as linhas restantes pelos filtros passa alta  $g$ . O fator de decimação, desloca os coeficientes de uma linha para outra como pode ser observado na Equação 13.

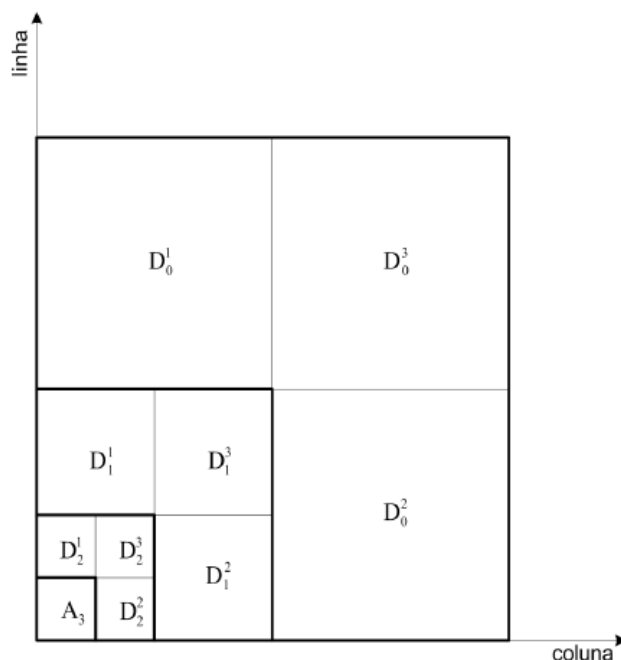
As Figuras 9 e 10 ilustram a transformada *wavelet* discreta por meio dos filtros passa-baixa e passa-alta. Enquanto a Figura 8 mostra a matriz com os respectivos filtros  $G$  e  $H$ , a Figura 10 mostra que para cada nível  $i$  são gerados uma aproximação  $A_{i+1}$ , e três classes de coeficientes de detalhes  $D_i^1, D_i^2, D_i^3$ .

Figura 9 – DWT de uma matriz em termos de filtros



Fonte: Pereira, 2008

Figura 10 – Representação da matriz  $A = A_0$  na transformada *wavelet* bidimensional



Fonte: Pereira, 2008

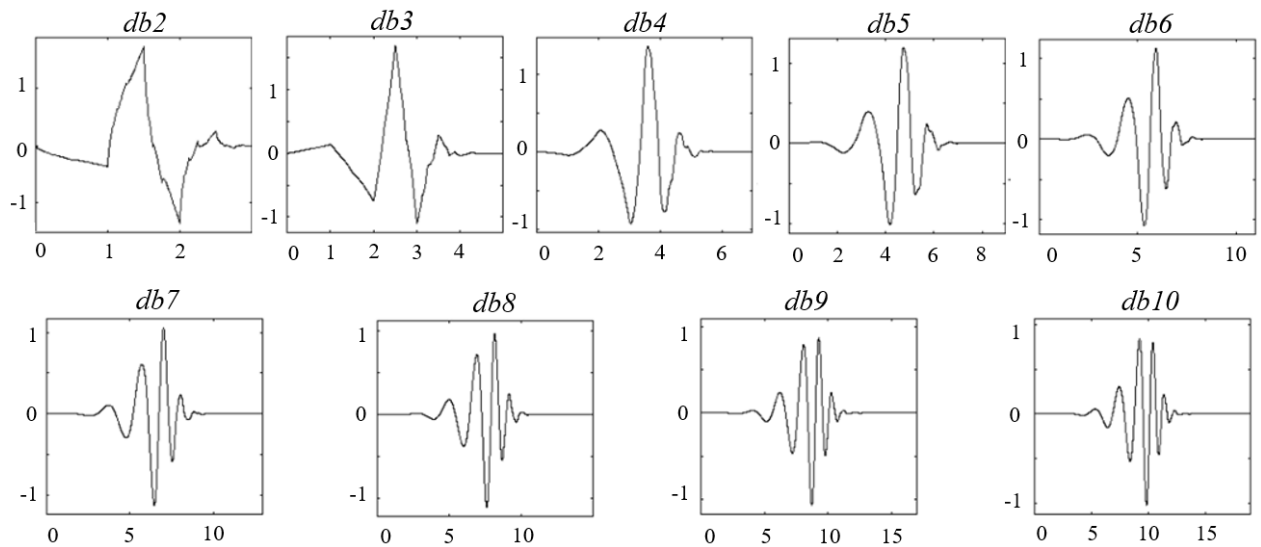
### 3.4.2.1. Wavelet Daubechies, Coifman e Symmlet

A família de *wavelets* biortogonais exibe a propriedade de fase linear, necessária para a reconstrução do sinal. Essa *wavelet* tem suporte compacto e é simétrica. Além disso, são definidas como pares de bases mutuamente ortogonais, mais nenhum desses pares é ortogonal.

As *wavelets* ortogonais de *Daubechies* da Figura 11, “*dbN*”, são perfeitamente compactas no domínio do tempo, mas no domínio da frequência, têm alto grau de superposição espectral entre as escalas. Apesar de não possuírem fase linear, um erro no sinal de entrada não cresce com a transformação, e a estabilidade numérica computacional é assegurada. Além disso, têm a capacidade de aproximar com precisão funções constantes e lineares de forma relativamente simples (TANG e WU, 2013; GOGOLEWSKI, 2020).

A regularidade das *wavelets* de *Daubechies* aumentam linearmente com  $N$ , porém a um custo que aumenta com o comprimento do suporte. O suporte das *wavelets*  $\Psi(t) = dbN$  é o intervalo fechado dado por  $[1-N, N]$ . Na Figura 11 é possível ver as formas de ondas correspondentes às *wavelets* de *Daubechies*  $dbN$  (MALLAT, 2009; DE OLIVEIRA, 2007).

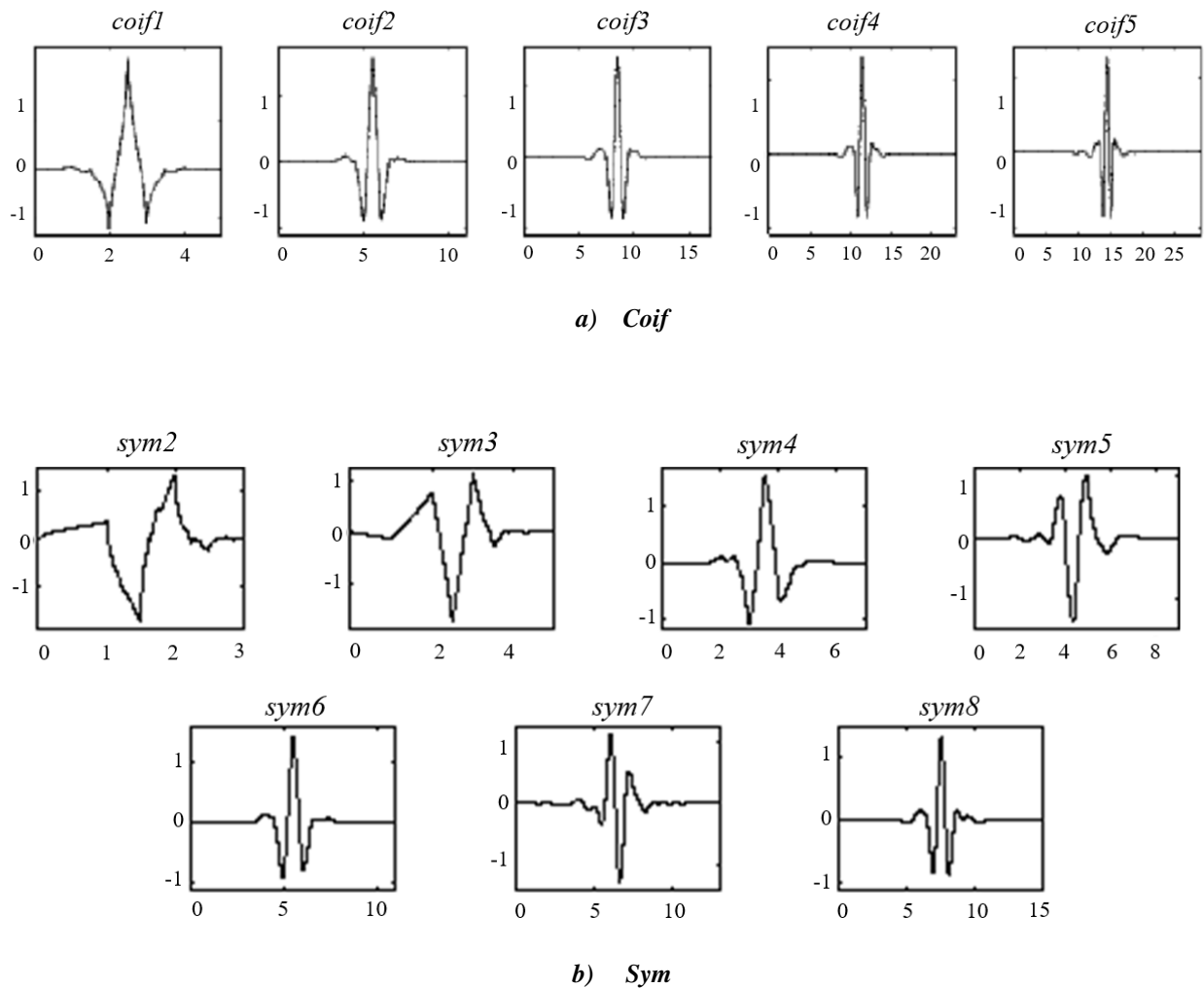
**Figura 11 – Família de Wavelet Daubechies**



As *wavelets* de *Daubechies* são muito assimétricas, para obter uma *wavelet* simétrica ou antissimétrica pode ser utilizado as *wavelet Symmlets*, que são mais simétricas, as quais foram projetadas para garantir momentos nulos tanto na função de escala, como na *wavelet*-mãe.

Por outro lado, as *wavelets* de *Coifman* (Figura 12b) são semelhantes às *wavelet* de *Daubechies*, pois possuem número máximo de momentos de fuga (*vanishing moments*), que limita a capacidade das *wavelets* de representar comportamento polinomial ou informações em um sinal, como por exemplo, *db2* tem um momento de fuga, *db4* tem dois, *db6* tem três, etc. No entanto, nas *wavelets Coifman*, os momentos de finalização são igualmente distribuídos entre a função de escala (Figura 12b).

**Figura 12 – Wavelets: a) Coif e b) Sym**



### 3.4. Teste Kaiser-Meyer-Olkin e Bartlett

O critério ou teste Kaiser-Meyer-Olkin (KMO) foi desenvolvido em 1977, é utilizado para identificar se o modelo de análise fatorial em uso está adequadamente ajustado aos dados, testando a consistência geral dos dados. O método verifica se a matriz de correlação inversa está próxima da matriz diagonal, que consiste em comparar os valores dos coeficientes de correlação linear observados com os valores dos coeficientes de correlação parciais.

O teste KMO retorna valores entre 0 e 1, quanto mais próximo de 1 indica que as amostragens são adequadas, geralmente indicam que uma análise fatorial pode ser útil com seus dados, enquanto mais próximo de 0 indica que há grandes correlações parciais em comparação à soma de correlações, mostrando correlações generalizadas, gerando problemas para a análise fatorial e os resultados da análise fatorial provavelmente não serão muito úteis.

A Equação 17 mostra a medida de adequação  $KMO$ , onde  $r_{jk}$  é o coeficiente de correlação simples entre as variáveis  $X_j$  e  $X_k$  e  $p_{jk}$  é o coeficiente de correlação parcial entre  $X_j$  e  $X_k$ , dados os outros  $X_s$ .

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 \sum \sum_{j \neq k} p_{jk}^2} \quad (17)$$

Já o teste de esfericidade de Bartlett é uma técnica criada por Maurice Stevenson Bartlett em 1937, ele indica a força do relacionamento entre variáveis. A técnica testa a hipótese de que a matriz de correlação é matriz de identidade (matriz na qual todos os elementos diagonais são 1 e todos os elementos diagonais externos são 0, o que indicaria que suas variáveis não são relacionadas e, portanto, inadequadas para a detecção da estrutura. Pequenos valores (abaixo de 0,05) de significância indicam que a análise fatorial pode ser útil com seus dados.

As Equações 18 e 19 mostram como Maurice Stevenson Bartlett desenvolveu o teste de esfericidade de Bartlett, onde  $n$  = tamanho da amostra,  $p$  = número de variáveis,  $|R|$  é a determinante da matriz de correlação e  $v$  é uma distribuição qui-quadrado com graus de liberdade.

$$\chi^2 = - \left[ (n - 1) - \frac{2p + 5}{6} \right] \ln|R| \quad (18)$$

$$v = \frac{p(p - 1)}{2} \quad (19)$$

Os testes KMO e Bartlett foram utilizados com auxílio do *software* da *IBM SPSS Statistics*. Neste caso, não foram utilizadas as equações 14, 15 e 16 para calcular e testar a consistência geral dos dados, pois o *software* da *IBM SPSS Statistics*, calcula  $KMO$  e  $\chi^2$ , inserindo os dados. Esta etapa será explicada no capítulo de Metodologia.

### 3.5. Indicadores de desempenho

Os indicadores de desempenho são utilizados para medir a acurácia de determinada previsão, medindo o quão perto as previsões estão dos dados reais. Quanto menor o valor do erro, mais eficiente é a previsão, conseqüentemente, mais próximo a previsão está do valor real.

Quanto maior o erro na previsão, maior será a imprecisão na previsão, e mais longe a previsão estará dos valores reais.

No erro quadrático médio (EQM) é calculada a somatória da diferença média quadrática entre o valor real e a previsão (resposta) dada pela rede, em que  $X_t$  representa o valor real,  $Y_t$  representa o valor previsto e  $n$  número de períodos, conforme pode ser observado na Equação 20:

$$EQM = \frac{1}{n} \sum_{t=1}^n (X_t - Y_t)^2 \quad (20)$$

O erro absoluto médio (MAD) mede o desvio absoluto do ajuste em relação à média. Depende da soma dos desvios entre  $X_t$  e  $Y_t$  de  $n$  períodos, dividida pelo número de períodos, conforme pode ser observado pela Equação 21.

$$MAD = \frac{1}{n} \sum_{t=1}^n |X_t - Y_t| \quad (21)$$

O erro percentual absoluto médio (MAPE) depende da diferença entre valor real  $X_t$  e valor previsto  $Y_t$ , dividido pelo valor real  $X_t$ , dividido por  $n$  períodos, como pode ser visto na Equação 22.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{X_t - Y_t}{X_t} \right| \quad (22)$$

A raiz do erro quadrático médio (RMSE), vista na Equação 23 é um indicador que interpreta o valor como a medida do desvio médio entre o valor real observado ( $X_t$ ) e o valor previsto ( $Y_t$ ).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (X_t - Y_t)^2} \quad (23)$$

## 4. MODELO PROPOSTO E CONTRIBUIÇÃO DA TESE

Neste capítulo, são apresentados os modelos desenvolvidos para construção desta tese, passando pela construção dos modelos autorregressivos univariados e multivariados de previsão, desde a seleção de variáveis de entrada, até a seleção de atraso de tempo para cada variável de entrada.

Na sequência do trabalho, é exposto o modelo proposto da tese, constituída por cinco abordagens, que se transformaram em um modelo autorregressivo de previsão e serviram como auxílio para construir um modelo que procura selecionar o melhor atraso de tempo para cada variável de entrada.

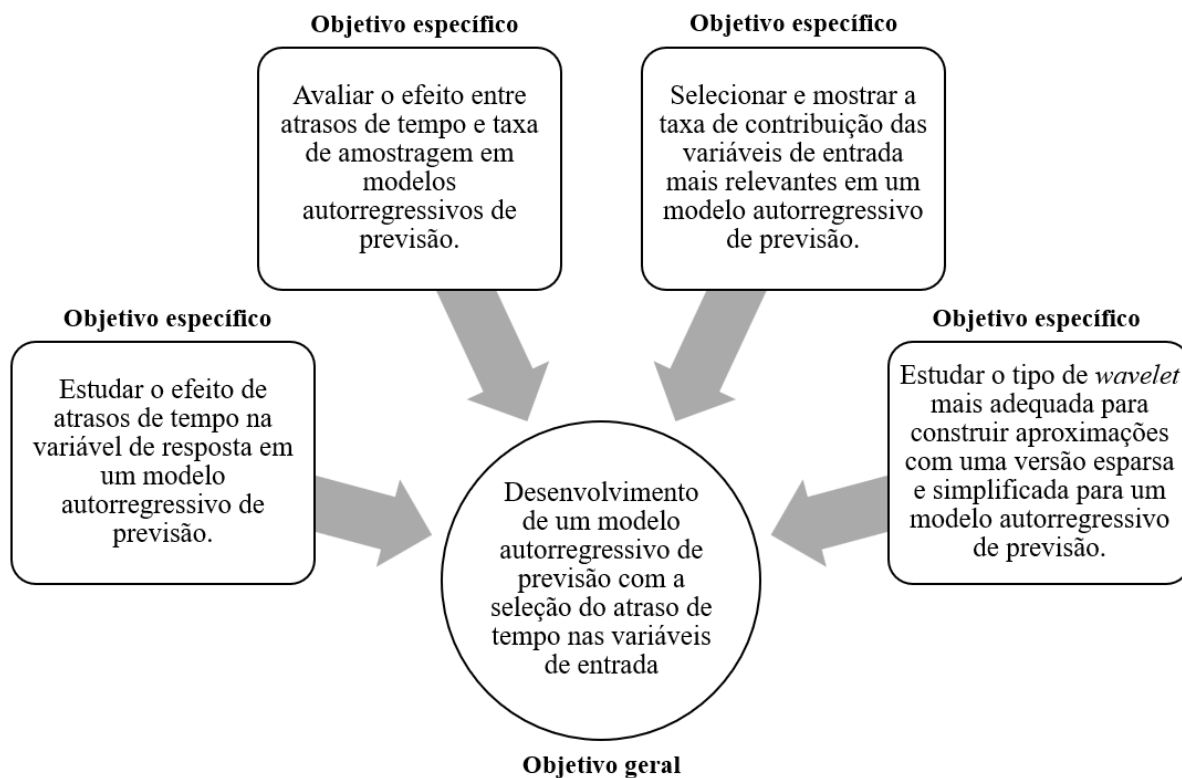
Além disso, este capítulo apresenta o desenvolvimento da produção bibliográfica da tese, que descreve a publicação de um artigo científico em um congresso nacional e quatro artigos científicos publicados e/ou aceitos em revistas internacionais.

### 4.4. Modelo proposto da tese

Para alcançar o objetivo proposto nesta tese, neste caso, desenvolvimento de um modelo autorregressivo de previsão, baseado na seleção de atributos, seleção da melhor variável de entrada, bem como a seleção do melhor atraso de tempo para cada variável de entrada, utilizando *wavelet*, PCA e rede neural artificial, foi necessário testar diferentes abordagens, na qual cada abordagem busca responder os objetivos específicos descritos no subcapítulo 1.3.2, além de propor o aprofundamento no desenvolvimento dos modelos autorregressivos de previsão.

A busca pela resposta de cada objetivo específico, gerou o desenvolvimento de modelos autorregressivos univariados e multivariados. Por fim, o desenvolvimento desses modelos, gerou o desenvolvimento de um modelo autorregressivo multivariado capaz de selecionar o atraso de tempo ótimo para cada variável de entrada, como pode ser observado na Figura 13.

**Figura 13 – Objetivo geral e específico das abordagens utilizadas na tese**



O desenvolvimento dos modelos autorregressivos teve como objetivo testar atrasos de tempo nas variáveis de entrada, testar a seleção e contribuição das variáveis, testar uma representação esparsa, simplificada e sem ruídos dos dados de entrada. Com isso, foi percebido que era possível desenvolver um novo modelo autorregressivo de previsão, capaz de criar uma representação esparsa, simplificada e com diferentes séries de tempo dos dados originais. A partir disso, o modelo é capaz de selecionar o atraso de tempo ótimo para cada variável de entrada.

A seguir são descritas as abordagens utilizadas em cada etapa desta tese, mostrando a construção dos modelos autorregressivos de previsão.

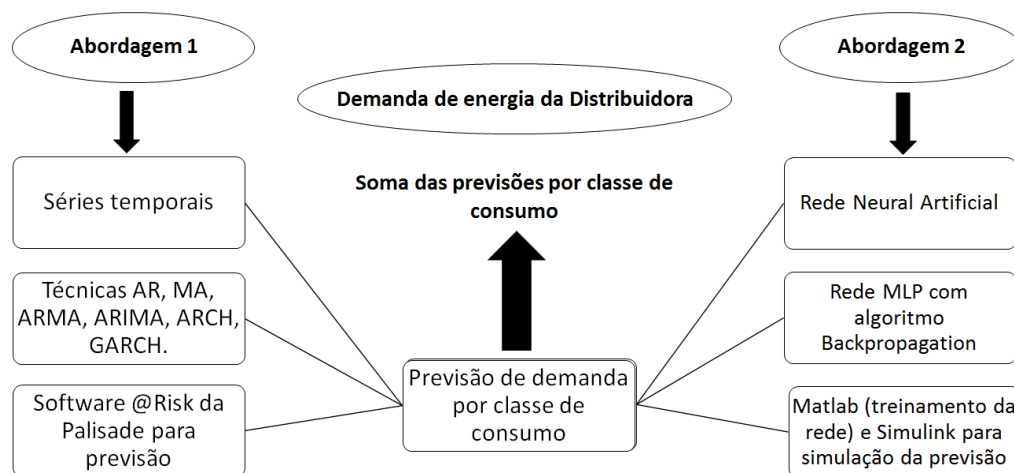
#### **4.4.3. Modelo autorregressivo para previsão de demanda de eletricidade**

Para responder os dois primeiros objetivos específicos: estudar o efeito de atrasos de tempo na variável de resposta de um modelo autorregressivo de previsão e avaliar o efeito entre atrasos de tempo e taxa de amostragem em modelos autorregressivos de previsão, foi proposto duas abordagens.

Nestas duas abordagens foram coletados dados mensais, de janeiro de 2003 a dezembro de 2016, de receita de fornecimento de energia, tarifa média de fornecimento de energia e número de unidades consumidoras para prever o consumo de eletricidade de cada classe consumidora de uma distribuidora de energia elétrica.

Para responder à pergunta sobre o efeito de atrasos de tempo na variável de resposta de um modelo autorregressivo de previsão, foi proposta a abordagem da Figura 14, na qual o trabalho foi desenvolvido com duas abordagens para previsão de demanda por classe de consumo, na qual a primeira utiliza técnicas de séries temporais e a segunda rede neural artificial.

**Figura 14 – Abordagem para previsão de demanda de energia elétrica**



Na abordagem 1, os dados de cada classe de consumo são transformados em estacionários pois, assim, a média, a variância e a autocorrelação não mudam ao longo do tempo, retirando a tendência dos dados das três classes de consumo (comércio, indústria e residência).

A previsão da demanda de energia elétrica das três classes de consumo utilizando técnicas de séries temporais (AR, MA, ARMA e ARCH) foi desenvolvida de forma univariada, ou seja, foi utilizado o consumo de eletricidade de cada classe de consumo para prever a mesma variável de consumo.

Nessa abordagem, foi utilizado o *software @Risk da Palisade* que é um *plugin* do *Excel* utilizado para simulação, análise de riscos; trabalha com simulação de Monte Carlo, mostra diversos gráficos estatísticos, faz cálculos de correlações de variáveis, entre outras. As

previsões foram selecionadas pelos Critérios AIC e BIC, que escolhem a melhor técnica de série temporal.

Já na abordagem 2, é desenvolvido um modelo autorregressivo multivariado com auxílio de uma rede MLP com algoritmo *backpropagation*, tendo dados históricos de receita de fornecimento de energia, número de unidades consumidoras e tarifa média de fornecimento de energia com uma camada de entrada, oculta e saída. Nesse caso, a rede neural foi treinada com a função de ativação tangente sigmoideal e 70% dos dados foram utilizados para treinamento e 30% para teste.

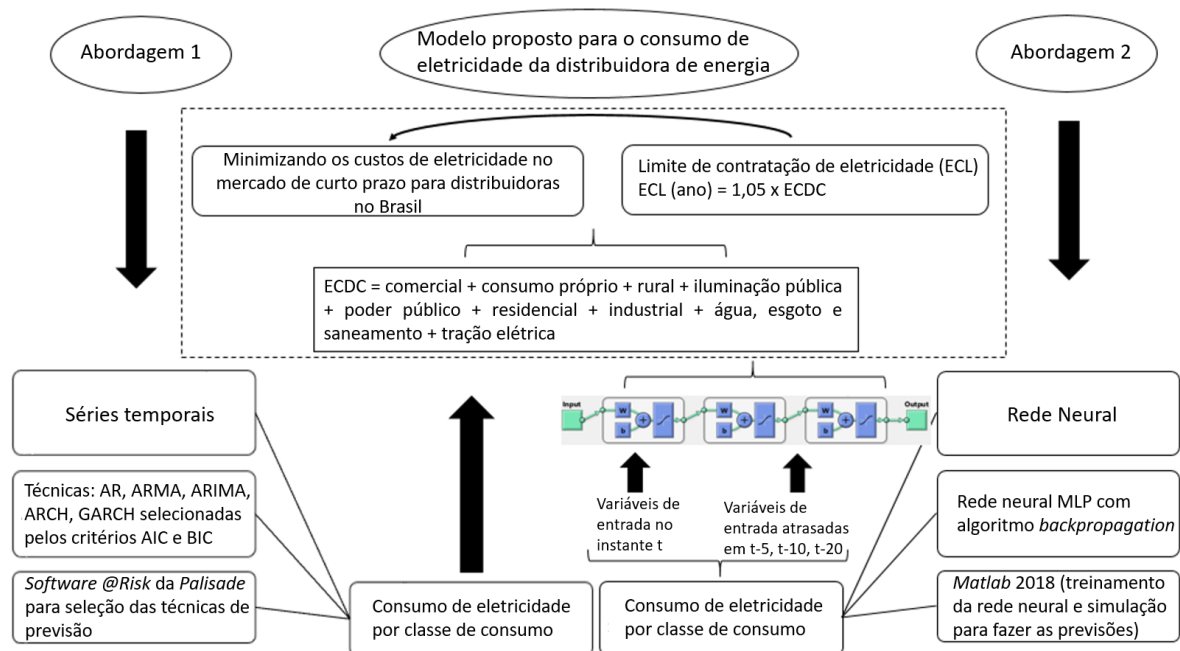
Com objetivo de prever a demanda de energia elétrica com um mês de antecedência, as variáveis de entrada (receita de fornecimento de energia elétrica, número de unidades consumidoras e tarifa média de fornecimento de energia) foram utilizadas no instante  $t$  e atrasadas em  $t-1$  e  $t-2$ . A eficiência das previsões é realizada por meio do índice do desempenho MAPE e 12 meses referente a 2016 são deixados para fazer um teste do modelo autorregressivo de previsão.

Enquanto a primeira abordagem procurou estudar o efeito de atrasos de tempo na variável de resposta de um modelo autorregressivo de previsão, o segundo modelo procurou avaliar o efeito de atrasos de tempo e taxa de amostragem em modelos autorregressivos de previsão. Para isso, foi proposto a mesma abordagem utilizada na abordagem 1 e 2 do primeiro modelo autorregressivo. No entanto, o interesse agora, é em avaliar o efeito maiores atrasos de tempo e taxa de amostragem em um modelo autorregressivo de previsão.

As mesmas variáveis de entrada utilizadas no primeiro modelo autorregressivo são atrasadas em  $t-5$ ,  $t-10$  e  $t-20$ . Neste caso, para  $t-5$  são utilizadas as variáveis no instante de tempo  $t-1$ ,  $t-2$ ,  $t-3$ ,  $t-4$  e  $t-5$ , e assim por diante para  $t-10$  e  $t-20$ . Assim, o modelo passa a ter mais entradas, visto que para três variáveis de entrada, o modelo terá 15 variáveis de entrada com as variáveis no instante  $t-5$ , para  $t-10$ , terá 30 variáveis de entrada, para  $t-20$ , terá 60 variáveis de entrada. Por fim, as previsões são feitas para nove classes de consumo de uma distribuidora da Região Sudeste do Brasil.

O objetivo principal desta abordagem (Figura 15) é verificar o impacto da regressão das variáveis de entrada em um modelo autorregressivo de previsão e, conseqüentemente, oferecer precisão nas previsões, gerando condições para a distribuidora de energia diminuir seus prejuízos com multas por sub ou sobrecontratação de energia elétrica.

**Figura 15 – Estrutura de previsão de eletricidade das classes de consumo**



A Tabela 1 mostra a configuração rede MLP com algoritmo *backpropagation* treinada com diferentes atrasos de tempo  $t-5$ ,  $t-10$  e  $t-20$ , alterando a quantidade de neurônios na camada oculta. Nesse caso, a seleção do melhor atraso de tempo é feita por meio do erro MAPE médio das previsões feitas de 11/2018 a 10/2019.

**Tabela 1 – Arquitetura proposta da rede MLP com algoritmo *backpropagation***

Atraso das variáveis de entrada	Quantidade de neurônios					
$t-5$	5	10	20	30	40	50
$t-10$	5	10	20	30	40	50
$t-20$	5	10	20	30	40	50

#### 4.4.4. Abordagem para seleção de atributos (seleção de variáveis de entrada)

Ao desenvolver dois modelos autorregressivos de previsão para responder os dois primeiros objetivos específicos, surgiram dois questionamentos em relação a esses modelos, o primeiro está relacionado a quantidade de variáveis utilizadas em um modelo autorregressivo de previsão, o segundo está relacionado a seleção do melhor atraso de tempo para cada variável de entrada.

Assim, surgiu a ideia de desenvolver um modelo para selecionar e mostrar o nível de contribuição de cada variável de entrada em um modelo autorregressivo de previsão.

Para isso, foram coletadas 27 variáveis diárias de 01/01/2001 a 31/08/2019 no site do Operador Nacional do Sistema Elétrico (ONS), como: energia armazenada (GWh), demanda instantânea (MW), demanda horária máxima (MWh/h), afluyente natural de energia bruta (MWmed), afluyente natural de energia armazenável (MWmed), energia carga (MWmed), troca de energia (MWmed) e dados hidrológicos de cinco grandes hidrelétricas que fornecem eletricidade para o sudeste do Brasil, Itaipu, Ilha Solteira, Marimbondo, Três Marias e Furnas, que, em 2016, 2017 e 2018, geraram 43,15%, 39,12% e 37,18% do total da energia hidrelétrica para a Região Sudeste (ONS, 2019).

As variáveis utilizadas em relação às usinas foram vazão afluyente ( $m^3/s$ ), que se refere ao fluxo que atinge a usina hidrelétrica ou a estrutura hidráulica, vazão efluente ( $m^3/s$ ), que representa o fluxo que sai da usina hidrelétrica ou da estrutura hidráulica, nível de quantidade (m), que são os pontos em direção à fonte, na direção oposta à corrente, a montante e, finalmente, o volume líquido em %, que representa o volume do reservatório entre o valor máximo normal de operação e o nível operacional mínimo normal.

O diagrama para selecionar e mostrar o nível de contribuição de cada variável de entrada pode ser observado na Figura 16. Antes de utilizar o PCA para selecionar e mostrar a taxa de contribuição das variáveis de entrada, os dados das 27 variáveis são normalizados entre 0 e 1. Então, as técnicas KMO e Bartlett foram utilizadas para verificar adequação e estrutura dos dados e, com isso, avaliar a adequabilidade da análise fatorial. Para isso, o software *IBM SPSS Statistics 22* foi utilizado para realizar os testes em uma matriz com 27 variáveis e 6815 dados.

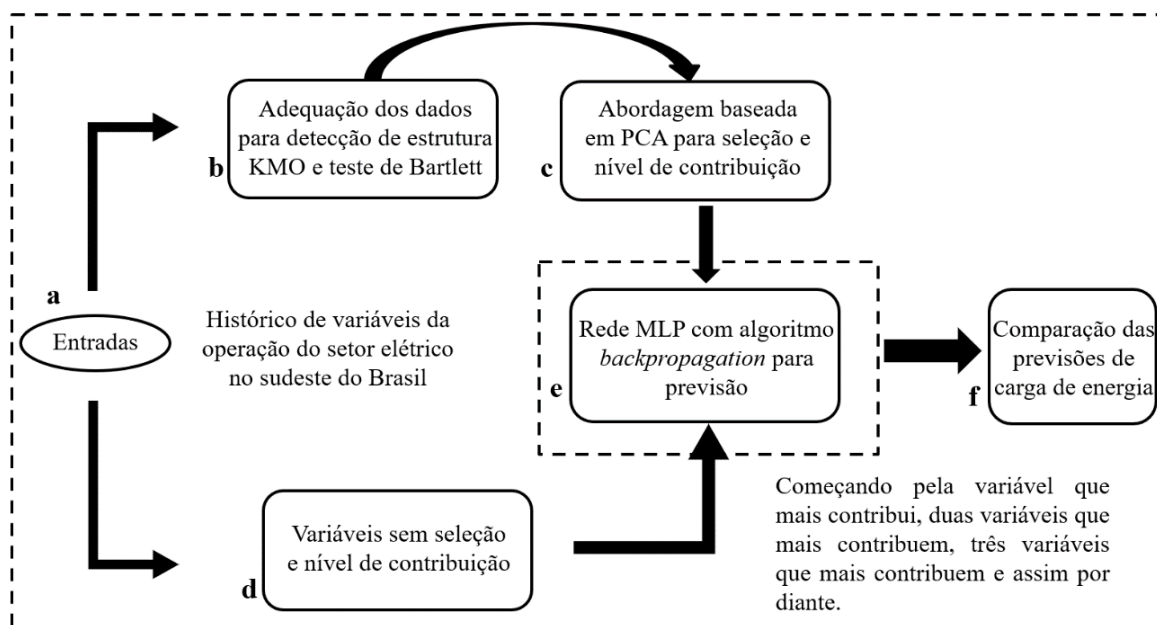
Para seleção e taxa de contribuição das variáveis de entrada, foi utilizado a abordagem descrita no subcapítulo 3.3.2, chamada de seleção de recursos pela Análise de componentes principais e utilizada no trabalho de Schimit e Pereira (2018). Assim, as variáveis são classificadas em ordem crescente por taxa de contribuição.

Finalmente, uma rede MLP com algoritmo *backpropagation* e função de ativação tangente sigmoide foi usada para treinar 26 redes no *Matlab 2018a*, separando 70% dos dados para treinamento, 15% para validação e 15% para testes. A rede foi configurada com uma camada de entrada, duas camadas ocultas (10 neurônios), pois, segundo Jurado et al. (2015), com mais neurônios, o custo computacional aumenta significativamente e o desempenho não melhora.

Inicialmente, a previsão foi realizada com 26 variáveis de entrada, deixando a variável de carga energética GWh como saída e, posteriormente, de acordo com a taxa de contribuição

das variáveis, vários testes de previsão foram realizados, classificando os modelos que possuem melhores e piores previsões de acordo com o menor erro MAPE, MAE, RMSE e MAD.

Figura 16 – Diagrama de abordagem proposto



#### 4.4.5. Uso de *wavelet* mais adequada para construir aproximações suaves aos dados de entrada

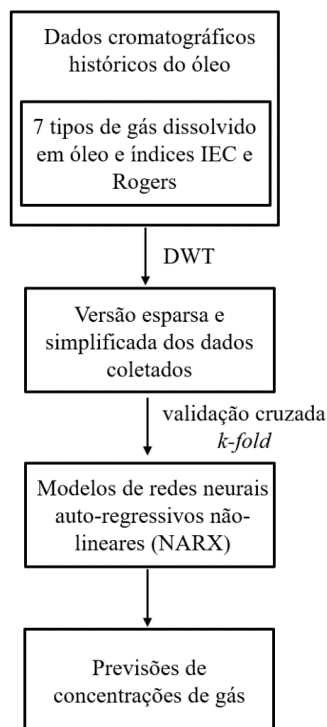
Como dito anteriormente, os trabalhos apresentados na literatura buscam melhorar a precisão dos modelos autorregressivos de previsão. Para desenvolver um modelo autorregressivo de previsão com alta precisão e comparar com a literatura, foi utilizada uma abordagem com a transformada discreta de *wavelet* (*Daubechies*, *Symmlets* e *Coifman*), descritas nos subcapítulos 3.4.1 e uma rede neural NAR. Além disso, as *wavelets* foram testadas para verificar qual *wavelet* apresenta dados de entrada com boas aproximações para desenvolver o modelo autorregressivo de previsão.

Basicamente, a ideia deste modelo é criar uma versão esparsa e simplificada dos dados de entrada, testando diferentes níveis de decomposição do sinal de entrada utilizando *sym2*, *sym3*, *sym4*, *db1*, *db3*, *db5*, *coif1*, *coif3* e *coif5*.

A Figura 17 mostra a abordagem utilizando a rede neural NAR e DWT para prever valores futuros de sete tipos de gás dissolvido em óleo, as razões IEC, que determina os níveis

de concentração de gás em transformadores e o método Rogers, que utiliza as seguintes concentrações de gases  $\text{CH}_4/\text{H}_2$ ,  $\text{C}_2\text{H}_2/\text{C}_2\text{H}_4$ ,  $\text{C}_2\text{H}_4/\text{C}_2\text{H}_6$ .

**Figura 17 – Fluxograma do modelo autorregressivo de previsão não-linear proposto (NAR-DWT)**



Os dados de concentrações de gases de um transformador de potência foram coletados diariamente, por meio de dois sensores, *GE Kelman – Transfix* e *GE Intellix BMT 330*. Após a coleta, o conjunto de dados é avaliado usando a transformação de *wavelet* discreta *Daubechies* (*db1*, *db3* e *db5*), *Symlets* (*sym3*, *sym4* e *sym5*), *Coiflets* (*coif1*, *coif3* e *coif5*) para criar uma versão esparsa e simplificada com boas propriedades de aproximação.

O modelo autorregressivo univariado composto por uma rede neural NAR é treinada utilizando o valor da própria concentração de gás como entrada para fazer a previsão para a mesma concentração de gás.

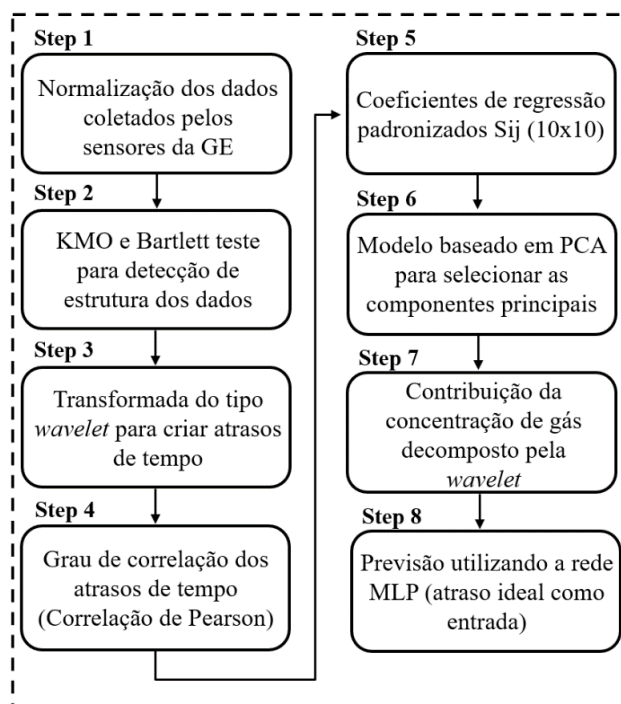
O modelo foi validado com a abordagem de validação cruzada *k-fold*, na qual os dados foram divididos em dez subconjuntos e o treinamento repetido dez vezes usando um dos dez subconjuntos de cada vez para teste e validação, enquanto os outros nove subconjuntos formavam um treinamento conjunto único. A estimativa de erro é calculada como média em todas as dez tentativas.

#### 4.4.6. Desenvolvimento de um novo modelo autorregressivo de previsão baseado em *wavelet*, PCA e seleção de atraso de tempo

Com as abordagens criadas anteriormente para cada modelo autorregressivo de previsão, foi possível testar a ideia principal desta tese, e responder o objetivo geral, que é desenvolver uma abordagem baseada na transformada *wavelet* e na análise de componentes principais para seleção de atributos (seleção de atraso de tempo) e determinação de atrasos de tempo em modelos autorregressivos de previsão fundamentados em redes neurais artificiais.

Na abordagem da Figura 18 é proposto um novo modelo para criar e selecionar diferentes series temporais com atraso de tempo e, por fim, criar um modelo autorregressivo multivariado com os melhores atrasos de tempo em uma rede neural.

Figura 18 – Modelo para seleção e taxa de contribuição da concentração e previsão de gases



Utilizando os mesmos dados da abordagem anterior, neste caso, dados de concentrações de gases de um transformador de potência, os dados de entrada são normalizados entre 0 e 1, em seguida, são utilizados os testes KMO e Bartlett para avaliar a adequabilidade da análise fatorial dos dados da mesma forma que foi utilizado na abordagem da Figura 17.

Após isso, a transformada *wavelet* é empregada de duas formas, primeiro os dados de cada concentração de gás são decompostos mantendo o nível de decomposição em 1, alterando

a *wavelet* de  $db2$ ,  $db4$ , ... para  $db20$ , a fim de criar aproximações suaves e com atrasos de tempo, neste caso,  $t-2$ ,  $t-4$  até  $t-20$ . A Tabela 2 ilustra a forma como foram inseridos os dados para verificar qual atraso de tempo tem mais impacto nas cinco concentrações de gases  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $H_2$  e  $CH_4$ .

**Tabela 2 – Representação do sinal decomposto pela *wavelet* para cada atraso de tempo nas variáveis de entrada**

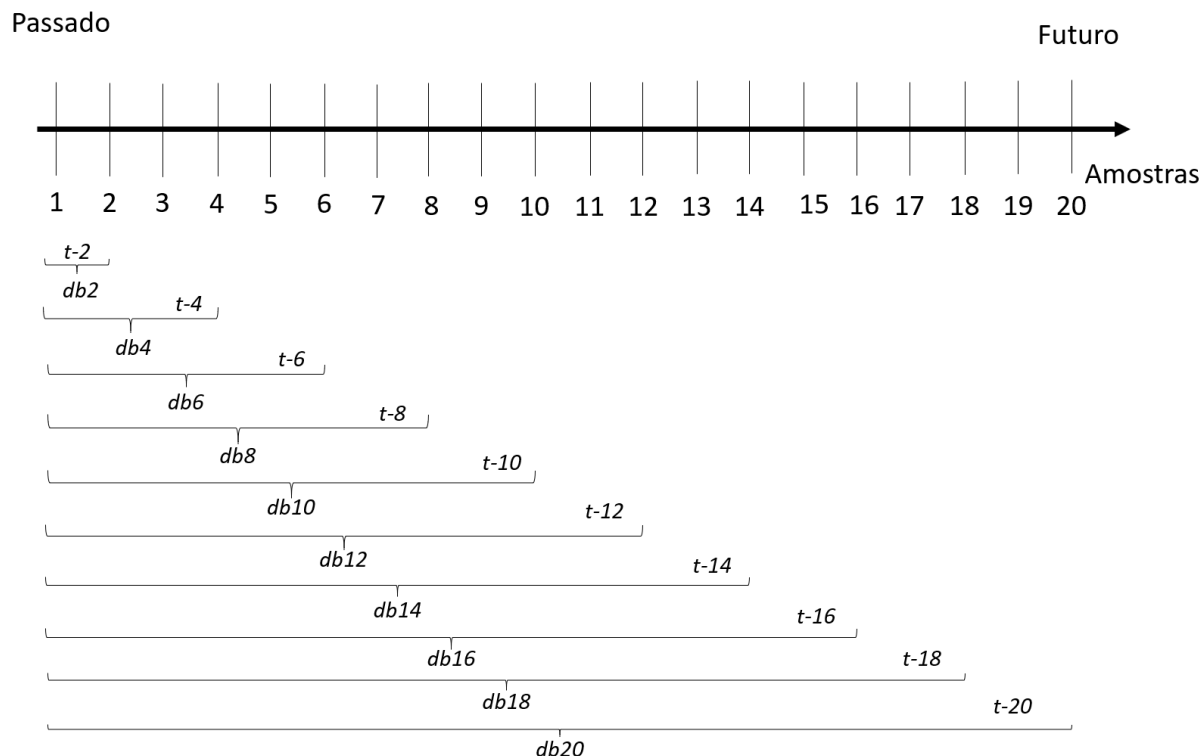
<i>Wavelet</i> decomposta	Tipo de concentração de gás				
	$C_2H_2$	$C_2H_4$	$C_2H_6$	$H_2$	$CH_4$
$C_2H_2(db2)$	$t-2$	$t-2$	$t-2$	$t-2$	$t-2$
$C_2H_2(db4)$	$t-4$	$t-4$	$t-4$	$t-4$	$t-4$
⋮	⋮	⋮	⋮	⋮	⋮
$C_2H_2(db20)$	$t-20$	$t-20$	$t-20$	$t-20$	$t-20$
$C_2H_4(db2)$	$t-2$	$t-2$	$t-2$	$t-2$	$t-2$
$C_2H_4(db4)$	$t-4$	$t-4$	$t-4$	$t-4$	$t-4$
⋮	⋮	⋮	⋮	⋮	⋮
$C_2H_4(db20)$	$t-20$	$t-20$	$t-20$	$t-20$	$t-20$
$C_2H_6(db2)$	$t-2$	$t-2$	$t-2$	$t-2$	$t-2$
$C_2H_6(db4)$	$t-4$	$t-4$	$t-4$	$t-4$	$t-4$
⋮	⋮	⋮	⋮	⋮	⋮
$C_2H_6(db20)$	$t-20$	$t-20$	$t-20$	$t-20$	$t-20$
$H_2(db2)$	$t-2$	$t-2$	$t-2$	$t-2$	$t-2$
$H_2(db4)$	$t-4$	$t-4$	$t-4$	$t-4$	$t-4$
⋮	⋮	⋮	⋮	⋮	⋮
$H_2(db20)$	$t-20$	$t-20$	$t-20$	$t-20$	$t-20$
$CH_4(db2)$	$t-2$	$t-2$	$t-2$	$t-2$	$t-2$
$CH_4(db4)$	$t-4$	$t-4$	$t-4$	$t-4$	$t-4$
⋮	⋮	⋮	⋮	⋮	⋮
$CH_4(db20)$	$t-20$	$t-20$	$t-20$	$t-20$	$t-20$

Com isso, cada concentração de gás, passa a ter dez variáveis de entrada com diferente comprimento, como pode ser observado na Tabela 3 para a concentração de gás  $CO_2$ . Isso acontece devido a teoria explicada no subcapítulo 3.4.2.

Além disso, a transformada *wavelet* é aplicada nas concentrações de gás em ordem cronológica reversa, de modo que cada amostra da aproximação seja criada com valores passados do sinal original para criar séries temporais. A Figura 19 mostra a abordagem utilizada com a *wavelet Daubechies*  $db2$  a  $db20$  para criar séries temporais com atrasos de tempo.



**Figura 19 – Abordagem aplicada nas concentrações de gás em ordem cronológica reversa**



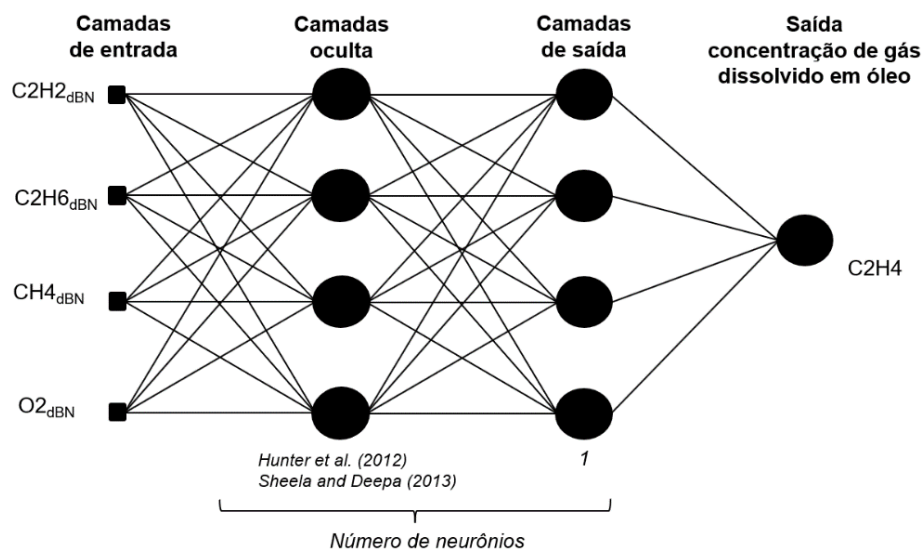
Em seguida, o PCA e correlação de Pearson são aplicados para selecionar quais das aproximações criadas na Tabela 4 mais contribuem para o modelo autorregressivo multivariado de previsão. Posteriormente, o PCA é aplicado na Tabela 4 de acordo com a abordagem descrita no subcapítulo 3.3.2 para determinar a taxa de contribuição de cada atraso de tempo.

Assim, a seleção de variáveis, empregada no terceiro modelo autorregressivo (artigo 3), e a seleção de atraso de tempo empregada no modelo autorregressivo 5 (artigo 5) utilizaram os seguintes passos: cálculo da matriz de covariância de um vetor de características; construção de uma matriz com os primeiros  $q$  principais autovetores; cálculo das projeções de cada ponto no subespaço PCA; definição da taxa de contribuição de cada variável original na projeção como uma soma ponderada do produto interno entre a variável e cada componente principal.

A correlação de Pearson, também é aplicada nos dados da Tabela 4 utilizando o *software IBM SPSS Statistics* para determinar a taxa de contribuição de cada atraso de tempo.

Com o maior nível de contribuição de cada concentração de gás, é empregada uma rede MLP com algoritmo *backpropagation*, como pode ser observado na Figura 20, com uma camada de entrada, uma camada oculta, uma camada de saída. A quantidade de neurônios na camada oculta foi utilizada seguindo duas abordagens, a primeira, desenvolvida por Hunter et al. (2012), e a segunda, desenvolvida por Sheela e Deepa (2013).

**Figura 20 – Arquitetura da rede MLP proposta para previsão de cada concentração de gás**



O quadro 1 mostra o modelo estrutural completo para o desenvolvimento da tese. É possível observar a proposta, composta por meio de artigos científicos, na qual é possível observar, a coleta de dados, horizonte de previsão, normalização dos dados, técnica para suavização dos dados, técnica para seleção atributos (seleção da variável de entrada e seleção de atraso de tempo), atraso das variáveis de entrada, técnicas para previsão e medida de precisão da previsão para cada abordagem.

Quadro 1 - Modelo estrutural para desenvolvimento da tese

	Abordagem 1 – Apêndice A	Abordagem 2 – Apêndice B	Abordagem 3 – Apêndice C	Abordagem 4 – Apêndice D	Abordagem 5 – Apêndice E
Proposta	Previsão para três classes de consumo de uma distribuidora de energia	Previsão e minimização do custo da eletricidade de uma distribuidora	Taxa de contribuição das variáveis para entendimento do processo que gerou os dados	Previsão de concentrações de gás dissolvido em óleo (transformador de potência)	Otimização da ordem de um modelo autorregressivo
Coleta de dados	Agência Nacional de Energia Elétrica (ANEEL)	Agência Nacional de Energia Elétrica (ANEEL)	Operador Nacional do Sistema Elétrico (ONS)	Dados de DGC - Sensor da GE Kelman- <i>Transfix</i> e GE <i>Intellix</i> BMT 330	Dados de DGC - Sensor da GE Kelman- <i>Transfix</i> e GE <i>Intellix</i> BMT 330
Horizonte de previsão	Mensal	Mensal	Diário	Diário	Diário
Normalização dos dados	Min-Max	Min-Max	Min-Max	Min-Max	Min-Max
Técnica para suavização dos dados	Não foi aplicado	Não foi aplicado	Não foi aplicado	Transformada do tipo Wavelet ( <i>sym</i> , <i>db</i> , <i>coif</i> )	Transformada do tipo Wavelet ( <i>db</i> )
Técnica para seleção e classificação de atributos	Não foi aplicado	Feita pelo menor erro MAPE das previsões mensais	Abordagem utilizando PCA	Não foi aplicado	PCA e correlação de Pearson
Atraso das variáveis de entrada	Variáveis de entrada no instante $t$ , $t-1$ , $t-2$	Variáveis de entrada no instante $t$ , $t-5$ , $t-10$ e $t-20$	Variáveis de entrada no instante $t$	Variáveis de entrada no instante $t-2$ a $t-10$	Wavelet para criar atrasos de $t-2$ a $t-20$
Técnicas para previsão	Rede MLP com algoritmo backpropagation	Rede MLP com algoritmo backpropagation e técnicas de séries temporais	Rede Multicamadas com algoritmo backpropagation	Rede neural autorregressiva com Entradas exógenas (NARX)	Rede Multicamadas com algoritmo backpropagation
Medida de precisão da previsão	MAPE	MAPE	MAPE, RMSE, MSE e MAD	Erro máximo e médio (%)	MAPE

## 5. RESULTADOS DOS MODELOS AUTORREGRESSIVOS DE PREVISÃO

Este capítulo mostra os resultados obtidos de todas abordagens utilizadas nesta tese, desde o impacto de atrasos de tempo, uso de seleção de recursos (variáveis mais importantes, seleção de atrasos de tempo) e seus impactos em modelos autorregressivos de previsão. Uso de *wavelet* para criar uma versão esparsa, simplificada com diferentes séries temporais e seus impactos em modelos autorregressivos de previsão. Além de mostrar a eficiência destas abordagens, por meio de índice de desempenho MAPE.

### 5.1. Resultados do primeiro e segundo modelo autorregressivo de previsão

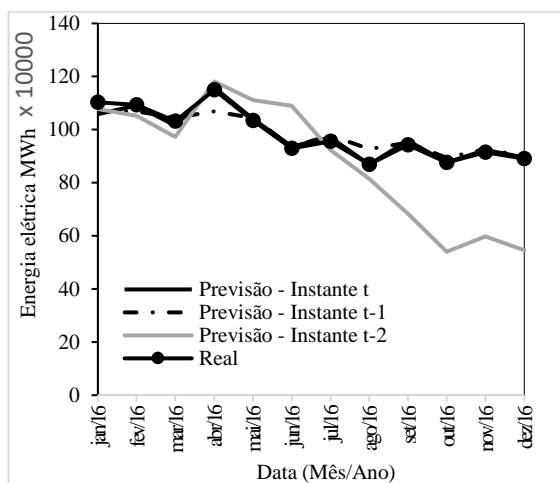
A Tabela 4 e a Figura 21 mostram os resultados das abordagens utilizadas no primeiro modelo autorregressivo. A Tabela 5 mostra o erro MAPE médio para 12 meses referente a 2016 para as três classes de consumo, comercial, industrial e residencial, na qual é possível observar que os menores erros MAPEs são 1,04% para a classe comercial, 9,34% para a classe de consumo industrial e 1,02% para a classe de consumo residencial.

**Tabela 4 – Erro MAPE das previsões do primeiro modelo autorregressivo (artigo 1)**

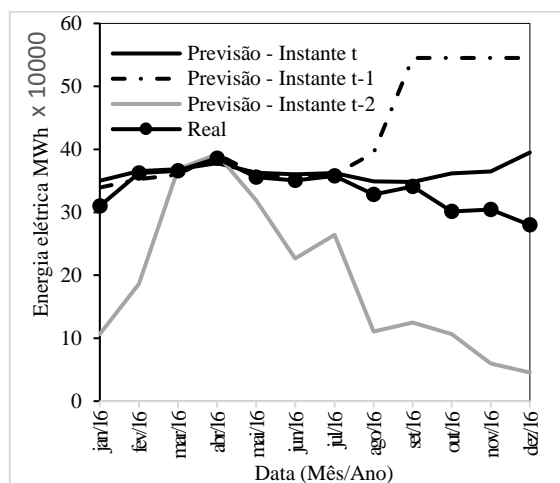
Classes de consumo	Atrasos de tempo na variável de entrada		
	Instante $t$	Instante $t-1$	Instante $t-2$
Comercial	1,04	2,46	15,66
Industrial	9,34	29,44	45,56
Residencial	1,02	1,9	7,21

A Figura 21 mostra os resultados da previsão para as classes de consumo comercial, industrial e residencial com as variáveis de entrada no instante  $t$ ,  $t-1$  e  $t-2$ . Na qual é possível observar que a previsão com as variáveis de entrada no instante  $t$  gera previsões mais assertivas quando comparado com as variáveis de entrada no instante  $t-1$  e  $t-2$ .

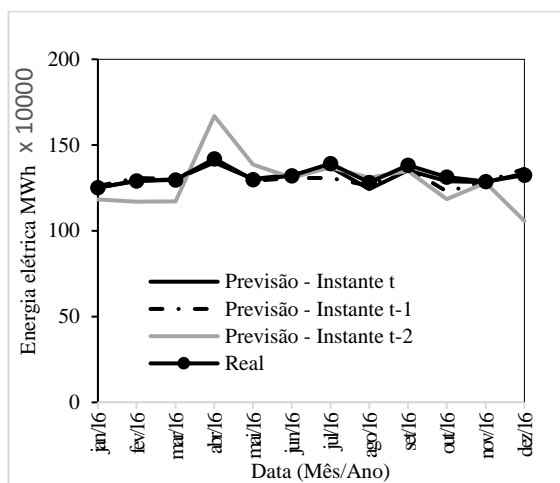
**Figura 21 – Previsão de energia elétrica por classe de consumo a) Comercial, b) Industrial, c) Residencial**



(a)



(b)



(c)

As Tabelas 6, 7 e 8 e a Figura 22 mostram os resultados da abordagem utilizada no segundo artigo científico, na qual apresenta as previsões mensais de 11/2018 a 10/2019 com as variáveis de entrada no instante  $t$  e com as variáveis atrasadas.

As Tabelas 6, 7 e 8 mostram os resultados das melhores previsões com as técnicas de séries temporais, o desempenho das previsões utilizando a rede neural MLP para cada configuração de neurônio e as previsões utilizando o melhor atraso de tempo para cada classe de consumo.

A Tabela 5 revela o melhor ajuste para selecionar a melhor técnica para realizar as previsões com as técnicas de séries temporais, neste caso, a melhor técnica é média móvel (MA) de ordem 1, 2 e a autorregressiva (AR) de ordem 2.

Tabela 5 – Previsão com a melhor técnica de série temporal escolhida pelo critério de ajuste por ano

Critério de ajuste	Classes de consumo	A	B	C	D	E	F	G	H	I	Total por classe
AIC	Valor	4514	2984	3092	3971	3761,4	4652,9	-419,8	3842	3112	-
	Técnica selecionada	MA1	AR2	MA2	MA1	MA1	MA2	MA1	MA2	MA1	-
	Eletricidade por ano	9817	38,3	10,3	865,3	1194,3	16176	3010,5	325,7	137,9	31575
BIC	Valor	4514	2997	3099	3971	3761,4	4653,5	-419,8	3853	3112	-
	Técnica selecionada	MA1	AR2	MA1	MA1	MA1	MA1	MA1	MA1	MA1	-
	Eletricidade por ano	9817	38,3	9,8	865,3	1194,3	16235	3010,5	315,5	137,9	31624

A Tabela 6 revela os resultados do desempenho da rede neural MLP em função da alteração dos neurônios na camada oculta. Neste caso, os melhores resultados para cada classe de consumo são 5 neurônios para as classes de consumo comercial, consumo próprio, rural e tração elétrica, 10 neurônios para as classes de consumo iluminação pública, 20 neurônios para as classes de consumo poder público e industrial, 30 neurônios para a classe de consumo água, esgoto e saneamento, 40 neurônios para a classe de consumo residencial. Ainda é possível observar o erro MSE e o tempo de treinamento de cada rede neural.

Tabela 6 – Desempenho das previsões e erro e tempo de treinamento

Classe / Numero de neurônios	05 05 1	10 10 1	20 20 1	30 30 1	40 40 1	50 50 1	Performance	
							MSE	Tempo
Comercial	<b>0,16</b>	1,67	5,5	1,04	3,01	4,67	1,71E-04	1s
Consumo próprio	<b>0,28</b>	1,3	0,87	1,45	2,84	1,52	5,49E-05	2s
Rural	<b>1,45</b>	6,38	3,03	22,88	141,5	183,1	1,69E-04	2s
Iluminação pública	0,52	<b>0,31</b>	0,72	0,4	0,53	0,7	1,60E-04	3s
Poder público	0,67	0,43	<b>0,17</b>	0,37	0,7	1,62	8,76E-07	11s
Residencial	0,48	0,36	0,73	1,08	<b>0,24</b>	2,56	3,41E-04	45s
Industrial	2,33	1,4	<b>1</b>	1,31	1,75	4,63	1,64E-05	11s
Água, esgoto e saneamento	3,48	2,57	3,33	<b>2,22</b>	3,16	3,05	3,08E-06	14s
Tração elétrica	<b>0,48</b>	0,55	1	0,69	1,59	2,5	2,14E-04	2s

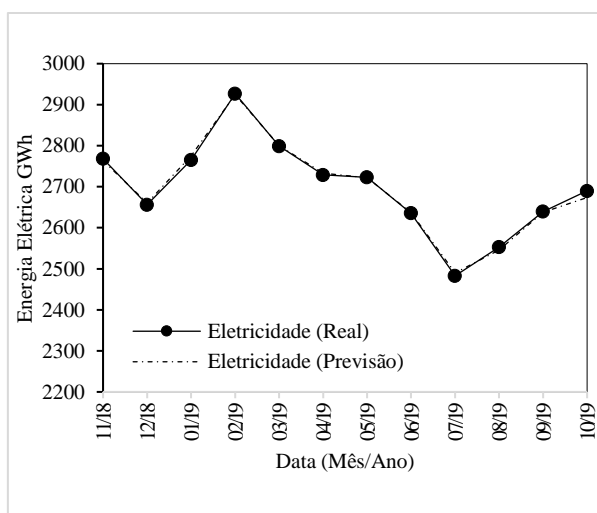
Sob a perspectiva da escolha do melhor atraso de tempo para cada classe de consumo, a Tabela 7 mostra as previsões para todas classes de consumo em função do melhor atraso de tempo, revelando que para a classe de consumo, foi utilizado um determinado atraso de tempo. A classe de consumo C, D e H utilizaram o atraso de tempo  $t-5$ , a classe de consumo B e I utilizaram as variáveis de entrada com atraso em  $t-10$ , as classes de consumo A, E, F e G utilizaram as variáveis em  $t-20$ .

**Tabela 7 – Previsão de energia elétrica (GWh) para todas classes de consumo com a seleção do melhor atraso de tempo**

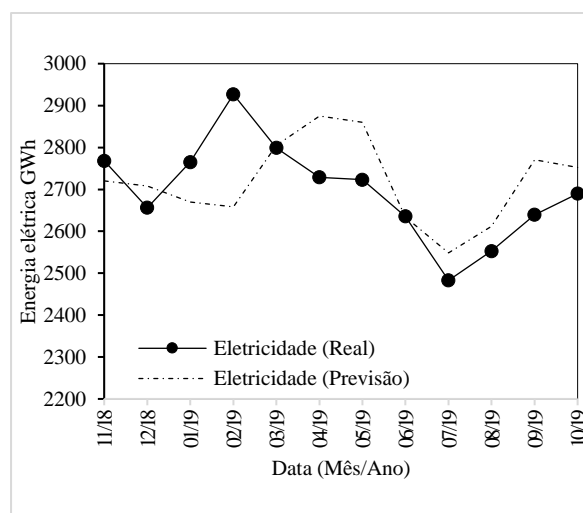
Classe de consumo	Atraso de tempo	Data (Mês/Ano)											
		11/18	12/18	01/19	02/19	03/19	04/19	05/19	06/19	07/19	08/19	09/19	10/19
A	$t-20$	840	903	1009	927	923	1023	914	851	791	766	822	828
B	$t-10$	3	4	4	3	3	3	3	3	3	3	3	3
C	$t-5$	1	1	3	1	1	1	1	1	1	1	1	1
D	$t-5$	56	56	59	65	69	59	56	61	59	57	63	57
E	$t-20$	95	112	103	107	91	109	106	103	102	105	111	121
F	$t-20$	1398	1319	1196	1249	1424	1378	1475	1305	1284	1376	1463	1439
G	$t-20$	287	273	259	271	261	268	265	268	268	265	269	264
H	$t-5$	28	28	28	28	28	28	28	28	28	28	28	28
I	$t-10$	12	12	10	7	6	7	12	13	13	11	12	12
Soma	-	2721	2708	2670	2658	2805	2875	2860	2634	2549	2611	2771	2753
Real	-	2768	2656	2765	2927	2799	2729	2723	2636	2483	2553	2640	2690

A Figura 22 mostra os resultados das previsões com as variáveis no instante  $t$  e  $t-5$ ,  $t-10$  e  $t-20$ . Os menores erros MAPE's são obtidos com as variáveis no instante  $t$ . No entanto, com a escolha do atraso de tempo para classe de consumo, os resultados apresentam um erro MAPE anual de 3,27%. O que faz com que a distribuidora de energia elétrica esteja sobrecontratada, ao invés de subcontratada, caso a previsão utilize as variáveis de entrada no instante  $t$ . Isso obrigaria a distribuidora a recontratar energia elétrica e a pagar pelo erro na subcontratação de energia elétrica.

**Figura 22 – Previsão de energia elétrica a) Instante  $t$ , b) Variáveis atrasadas em  $t-5$ ,  $t-10$  e  $t-20$**



(a)



(b)

A Tabela 8 mostra os resultados referente as previsões para as nove classes de consumo da distribuidora de energia elétrica utilizando as técnicas de séries temporais. É possível notar que o erro médio MAPE para 12 meses varia entre 2,27% e 19,92%.

**Tabela 8 – Erro médio MAPE(%) utilizando técnicas de séries temporais**

<b>Classe de consumo</b>	<b>MAPE(%)</b>
Comercial	8,08
Consumo próprio	5,28
Rural	19,92
Iluminação pública	8,91
Poder público	7,29
Residencial	2,27
Industrial	5,47
Água, esgoto e saneamento	7,73
Tração elétrica	9,58

## **5.2. Resultados do terceiro modelo autorregressivo de previsão**

A Figura 23 e a Tabela 10 mostram os resultados da abordagem utilizada no terceiro artigo científico. Na Figura 23 é possível observar os três melhores resultados com 7, 9 e 11 variáveis de entrada. Com estas variáveis de entrada foi possível obter erro médio MAPE para 31 dias de 0,001738%, 0,002566% e 0,002320%, respectivamente. As sete variáveis de entrada representam mais de 89% da taxa de contribuição das variáveis no modelo autorregressivo de previsão.

Já a Tabela 9, mostra a codificação para cada variável de entrada, taxa de contribuição, quantidade de variáveis utilizadas no modelo autorregressivo de previsão e o erro MAPE em função das variáveis de entrada no modelo autorregressivo de previsão. É possível observar que energia armazenada (GWh), demanda instantânea (MW), demanda horária máxima (MWh/h), afluente natural de energia bruta (MWmed), afluente natural de energia armazenável (MWmed) são as cinco variáveis mais importantes com as seguintes contribuições: 9,21, 4,53, 4,46, 4,23 e 3,83, respectivamente.

Figura 23 – Previsão de energia elétrica em GWh

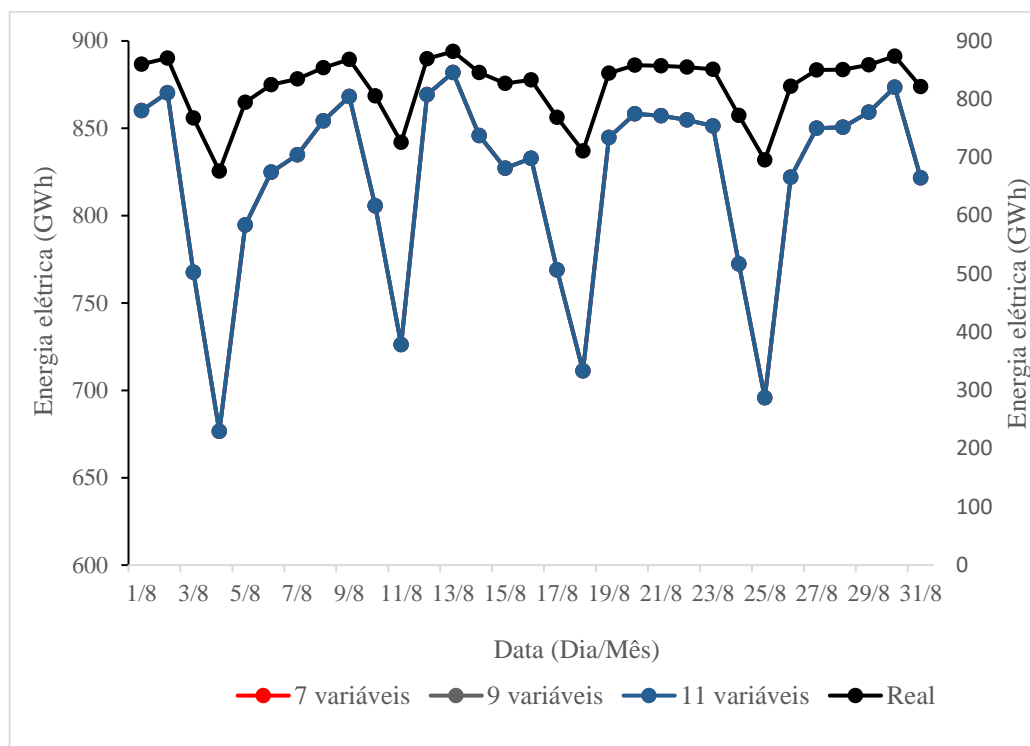


Tabela 9 – Taxa de contribuição e erro MAPE das variáveis de entrada

Codificação das variáveis	Variáveis de entrada	Taxa de contribuição	Qtd. de variáveis utilizadas no modelo de previsão	MAPE(%)
A	Energia armazenada (GWh)	9,2165		
B	Demanda instantânea (MW)	4,5369	A,B	6,1960
C	Demanda horária máxima (MWh/h)	4,4680	A,B,C	2,5544
D	Afluentes Natural de Energia Bruta (MWmed)	4,2380	A,B,C,D	2,1304
E	Afluentes Natural de Energia Armazenável (MWmed)	3,8349	A,B,C,D,E	2,6239
F	Carga de energia (MWmed)	3,7813	A,B,C,D,E,F	2,6718
G	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Itaipu	1,3326	A,B,C,D,E,F,G	0,0017
H	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Itaipu	1,3314	A,B,C,D,E,F,G,H	0,0053
I	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Ilha Solteira	0,5742	A,B,C,D,E,F,G,H,I	0,0026
J	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Ilha Solteira	0,5687	A,B,C,D,E,F,G,H,I,J	0,0064
K	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Marimbondo	0,1900	A,B,C,D,E,F,G,H,I,J,K	0,0023
L	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Marimbondo	0,1858	A,B,C,D,E,F,G,H,I,J,K,L	0,0039
M	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Furnas	0,0934	A,B,C,D,E,F,G,H,I,J,K,L,M	0,0055
-	Carga de energia (GWh)	0,0908	-	0,0034
N	Dados hidrológicos / nível de quantidade (m) – Furnas	0,0890	A,B,C,D,E,F,G,H,I,J,K,L,M,N	0,0028
O	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Furnas	0,0850	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O	0,0035
P	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Três Marias	0,0705	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P	0,0073
Q	Troca de Energia (MWmed)	0,0696	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q	0,0020
R	Dados hidrológicos / nível da quantidade (m) – Três Marias	0,0658	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R	0,0057
S	Dados hidrológicos / fluxo de efluentes m <sup>3</sup> /s – Três Marias	0,0628	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S	0,0086
T	Dados hidrológicos / nível da quantidade (m) – Marimbondo	0,0513	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T	0,0034
U	Dados hidrológicos / nível da quantidade (m) – Ilha Solteira	0,0380	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U	0,0039
V	Dados hidrológicos / nível da quantidade (m) – Itaipu	0,0256	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V	0,0142
W	Dados / volumes hidrológicos – Ilha Solteira	0,0077	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W	0,0028
X	Dados / volumes hidrológicos – Furnas	0,0076	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X	0,0065
Y	Dados / volumes hidrológicos – Marimbondo	0,0070	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y	0,0066
Z	Dados / volumes hidrológicos – Três Marias	0,0067	A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z	0,0080

### 5.3. Resultados do quarto e quinto modelo autorregressivo de previsão

A Tabela 10 mostra o desempenho das funções *wavelet* (*sym*, *db* e *coif*) e o efeito desse fator na precisão do modelo. Na qual é possível observar o erro MSE, R, erro médio e erro máximo para cada função *wavelet*. Os resultados mostram que a *wavelet coif5* obteve menor erro médio e máximo, alcançando 0,04% e 0,28%, respectivamente.

**Tabela 10 – Desempenho das funções da *wavelet* e o efeito desse fator na precisão do modelo proposto.**

<i>Wavelet</i>	<i>mse</i>	R	<i>avg_err</i> (%)	<i>max_err</i> (%)
<i>sym2</i>	4E-07	0,99	0,06	0,34
<i>sym3</i>	1E-06	0,99	0,08	0,50
<i>sym4</i>	7E-07	0,99	0,08	0,52
<i>db1</i>	7E-07	0,99	0,15	1,02
<i>db3</i>	1E-06	0,99	0,08	0,52
<i>db5</i>	2E-07	0,99	0,03	0,19
<i>coif1</i>	1E-06	0,99	0,10	0,60
<i>coif3</i>	4E-07	0,99	0,06	0,35
<i>coif5</i>	3E-07	0,99	0,04	0,28

A Tabela 11 mostra o erro médio e máximo da previsão para cada concentração de gás. Na qual é possível observar que esta abordagem utilizando funções *wavelet* para criar uma versão esparsa e simplificada dos dados de entrada, melhora o modelo autorregressivo de previsão.

**Tabela 11 – Erro de previsão NAR–DWT para concentrações de gás dissolvido em óleo.**

Tipo de Gás	<i>Erro médio</i> (%)	<i>Erro máximo</i> (%)
H <sub>2</sub>	0,460	6,030
CO	0,080	0,790
CO <sub>2</sub>	0,060	0,340
CH <sub>4</sub>	0,100	0,890
C <sub>2</sub> H <sub>6</sub>	0,290	2,340
C <sub>2</sub> H <sub>4</sub>	0,320	3,110
C <sub>2</sub> H <sub>2</sub>	0,330	2,370

A Tabela 13 mostra o efeito do parâmetro de atraso de tempo “d” no desempenho do processo de treinamento, avaliado usando o erro quadrado médio (MSE) e o coeficiente de

determinação R, que é a medida de qualidade do ajuste para regressão linear entre o alvo e as previsões. Nesse caso, foi definido “d” de 2 a 10, e o erro foi calculado como média em todas as 10 tentativas, de acordo com a abordagem de validação cruzada. A Tabela 12 mostra os resultados dessa avaliação usando a concentração de gás CO<sub>2</sub> como exemplo. O modelo NAR–DWT apresenta um ajuste muito preciso, independentemente do valor de “d”.

**Tabela 12 – Efeito do parâmetro d no desempenho do processo de treinamento para gás CO<sub>2</sub>.**

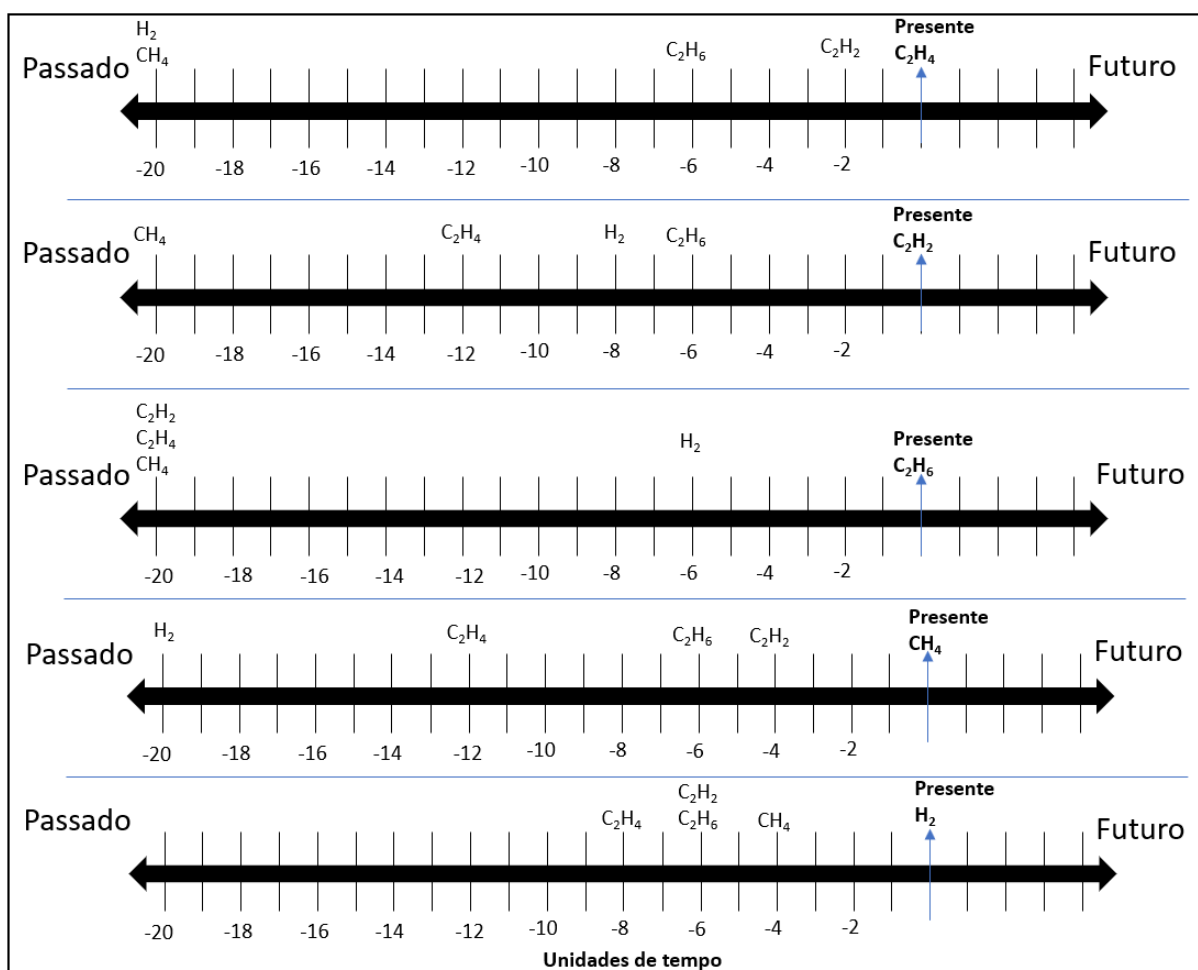
<i>d</i>	<i>MSE</i>	<b>R</b>
2	0,0000007	1
3	0,0000001	1
4	0,0000007	1
5	0,0000006	1
6	0,0000005	1
7	0,0000006	1
8	0,0000008	1
9	0,0000008	1
10	0,0000008	1

A Tabela 13 e a Figura 24 ilustram os resultados do quinto modelo autorregressivo de previsão. Os resultados mostram a influência temporal em cada concentração de gás no transformador de potência, bem como o nível de contribuição de cada atraso de tempo nas cinco concentrações de gases. Na concentração de gás H<sub>2</sub>, CH<sub>4</sub> leva 20 unidades de tempo para impactar em C<sub>2</sub>H<sub>4</sub>, enquanto C<sub>2</sub>H<sub>6</sub> e C<sub>2</sub>H<sub>2</sub> levam 6 e 2 unidades de tempo para influenciar C<sub>2</sub>H<sub>4</sub>. Ao passo que CH<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>, H<sub>2</sub> e C<sub>2</sub>H<sub>6</sub> levam 10, 12, 20 e 6 unidades de tempo respectivamente para influenciar em C<sub>2</sub>H<sub>2</sub>. Na concentração de gás C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, CH<sub>4</sub> e H<sub>2</sub> levam 20, 20, 20 e 6 unidades de tempo. Na concentração de gás CH<sub>4</sub>, H<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>6</sub> e C<sub>2</sub>H<sub>2</sub> levam 2, 12, 4 e 4 unidades de tempo. Por fim, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub> e CH<sub>4</sub> levam 8, 6, 6 e 4 unidades de tempo para impactar em H<sub>2</sub>.

Tabela 134 – Nível de correlação de casa atraso de tempo para concentração de gás

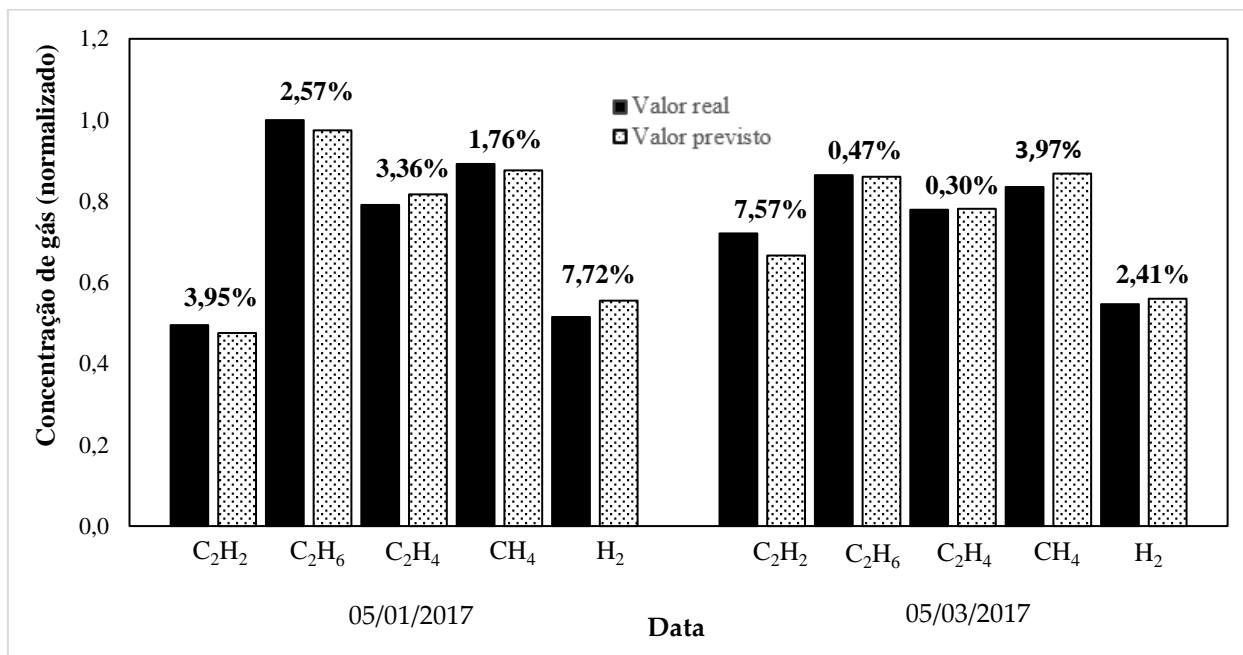
Concentração de Gás (atraso)	C2H2	C2H6	C2H4	H2	CH4
C <sub>2</sub> H <sub>2</sub> (db2)	0,01254	0,00116	<b>0,01904</b>	0,00029	0,00608
C <sub>2</sub> H <sub>2</sub> (db4)	0,00449	0,00116	0,01796	0,00109	<b>0,00689</b>
C <sub>2</sub> H <sub>2</sub> (db6)	0,00000	0,00032	0,01232	<b>0,00548</b>	0,00490
C <sub>2</sub> H <sub>2</sub> (db20)	<b>0,01769</b>	<b>0,00563</b>	0,00000	0,00137	0,00029
C <sub>2</sub> H <sub>6</sub> (db4)	0,01103	0,01440	0,03349	0,01061	<b>0,00314</b>
C <sub>2</sub> H <sub>6</sub> (db6)	<b>0,02190</b>	<b>0,02045</b>	<b>0,04203</b>	<b>0,01232</b>	0,00281
C <sub>2</sub> H <sub>4</sub> (db8)	0,00922	0,00017	0,00073	<b>0,00292</b>	0,00026
C <sub>2</sub> H <sub>4</sub> (db10)	0,00865	0,00003	0,00130	0,00044	0,00044
C <sub>2</sub> H <sub>4</sub> (db12)	<b>0,01082</b>	0,00010	<b>0,00240</b>	0,00000	<b>0,00185</b>
C <sub>2</sub> H <sub>4</sub> (db20)	0,00706	<b>0,00144</b>	0,00130	0,00036	0,00010
H <sub>2</sub> (db2)	0,00410	0,00130	0,00410	<b>0,00504</b>	0,00102
H <sub>2</sub> (db6)	0,00068	<b>0,00348</b>	0,00384	0,00026	0,00109
H <sub>2</sub> (db8)	0,00130	0,00250	0,00240	0,00048	0,00017
H <sub>2</sub> (db20)	<b>0,01300</b>	0,00130	<b>0,00608</b>	0,00176	<b>0,00270</b>
CH <sub>4</sub> (db4)	0,00005	0,00281	0,00036	<b>0,00336</b>	0,00044
CH <sub>4</sub> (db10)	<b>0,00922</b>	0,00058	0,00032	0,00023	0,00144
CH <sub>4</sub> (db14)	0,00044	0,00000	0,00006	0,00012	<b>0,00281</b>
CH <sub>4</sub> (db20)	0,00336	<b>0,00336</b>	<b>0,00212</b>	0,00160	0,00020

Figura 24 – Melhor Influência temporal para cada concentração de gás



A Figura 25 mostra dois dias previstos (01/05/2017 e 03/05/2017), comparando os valores reais para as cinco concentrações de gás. Para a concentração de gás  $C_2H_4$  é possível observar que o erro MAPE chegou a 3,36% para o dia 01/05/2017 e 0,30% para o dia 03/05/2017.

Figura 25 – Valores reais e previstos da concentração de gás por dois dias



Já a Tabela 14, mostra o erro MAPE médio para cada concentração de gás utilizando 8 e 15 neurônios na camada oculta da rede neural MLP. Nota-se que para as cinco concentrações de gás, a seleção do atraso de tempo (*SAT*), apresenta erro MAPE médio menor, quando comparado com o uso do mesmo atraso de tempo para todas as variáveis de entrada.

Tabela 54 – Erro das previsões para cada concentração de gás

Concentração de gás	Erro MAPE médio (%)		
	Atraso de tempo	8 neurônios	15 neurônios
C <sub>2</sub> H <sub>2</sub>	SAT	58,05	<b>5,76</b>
	<i>t</i> -2	24,32	95,10
	<i>t</i> -4	15,42	88,29
C <sub>2</sub> H <sub>6</sub>	SAT	7,65	<b>1,53</b>
	<i>t</i> -2	8,81	12,29
	<i>t</i> -4	13,00	8,49
C <sub>2</sub> H <sub>4</sub>	SAT	7,89	<b>1,83</b>
	<i>t</i> -2	4,31	1,91
	<i>t</i> -4	4,64	22,78
CH <sub>4</sub>	SAT	<b>2,87</b>	14,41
	<i>t</i> -2	3,32	7,62
	<i>t</i> -4	3,86	15,44
H <sub>2</sub>	SAT	<b>5,07</b>	6,17
	<i>t</i> -2	8,12	6,21
	<i>t</i> -4	8,63	16,15

A Tabela 15 mostra a abordagem, horizonte de previsão, tipo de previsão e erro MAPE de alguns trabalhos, comparando com as abordagens utilizadas nos cinco modelos autorregressivos de previsão, publicados em revistas científicas internacionais.

É possível observar que o menor erro MAPE dos modelos, foi apresentado com a seleção das variáveis mais importantes para o modelo autorregressivo de previsão, alcançando 0,00174 de erro, mostrando a importância da seleção de variáveis em modelos autorregressivos de previsão.

Por outro lado, a abordagem utilizando *wavelet* para criar uma versão esparsa e simplificada dos dados, apresentou resultados interessantes, quando comparado com modelos apresentados na literatura, alcançando um erro MAPE de 0,32.

Os resultados do quinto artigo, também apresentaram resultados interessantes na previsão de concentração de gás dissolvido em óleo em transformador de potência, alcançando erro MAPE de 1,525, além de oferecer melhor entendimento da relação temporal entre as concentrações de gases.

Tabela 15 – Comparação de abordagens utilizadas para previsão em modelos autorregressivos

Autor	Abordagem	Horizonte de previsão	Tipo de previsão	MAPE
Jurado et al. (2015)	Raciocínio Indutivo Fuzzy, Floresta Aleatória e Redes Neurais	Hora	Previsão de energia em edifícios	4,490
Ghadimi et al. (2019)	Redes neurais ridgelet, Elman e seleção de recursos	Semana		1,300
Bantugon e Gallano (2016)	Rede neural e Holter Winter	Dia		5,640
Lang et al. (2019)	Rede neural ponderada com pesos aleatórios e núcleos	Dia	Previsão de eletricidade	1,320
Wang e Hong (2016)	Efeito recente e Big data	Hora		3,390
Lahouar e Slama (2015)	Floresta aleatória e seleção de entrada de especialistas	Hora		<2,3
Lin et al., 2018	LSTM_DBN Network	Dia		0,260
Liu et al., 2019	Wavelet Least SVM and Imperialist Competition Algorithm	Dia	Concentração de gás dissolvido em óleo	0,168
Zhang et al., 2018	Abordagem usando LSSVM	Dia		1,690
Lin et al. (2018)	KPCA-FFOA-GRNN	Dia		3,270
Artigo 1	Rede neural MLP – variável de entrada no tempo t	Mês	Previsão de eletricidade (Comercial, Residencial e Industrial)	1,489
				1,337
Artigo 2	Rede neural MLP – variáveis de entrada atrasada	Mês	Previsão de eletricidade (todas classes de consumo de eletricidade)	7,452
				<0,8
				<9,2
Artigo 3	Rede neural MLP – variáveis de entrada atrasada	Ano	Previsão de eletricidade	0,015
				0,750
Artigo 3	RNA e PCA para seleção e taxa de contribuição das variáveis de entrada	Dia	Previsão de eletricidade	0,00174
Artigo 4	NAR-DWT	Dia	Concentração de gás dissolvido em óleo	0,320
Artigo 5	Wavelet e rede neural MLP	Dia		1,525

#### **5.4. Desenvolvimento da produção bibliográfica da tese**

A partir da resposta de cada objetivo específico e geral, foi desenvolvido e testado diferentes abordagens em modelos autorregressivos de previsão, na qual cada etapa avalia um aspecto do modelo. Os resultados destas abordagens resultaram na publicação de cinco artigos científicos.

A Tabela 16 mostra a contribuição de cada abordagem, status e ano da publicação ou aceite do artigo científico. Dos cinco artigos submetidos, apenas um foi publicado em congresso nacional (Seminário de Distribuição de Energia Elétrica – SENDI, 2018), o artigo 2 foi aceito na revista *Engineering Review*, o artigo 3 foi aceito na revista *International Journal of Energy Sector Management*, o artigo 4 foi publicado na revista *Energies* e o artigo 5 foi publicado na revista *Sensors*.

Tabela 16 – Contribuição e status dos artigos produzidos com a tese

Artigo	Título	Contribuição	Status de publicação	Ano
Artigo 1: Apêndice A	Uma abordagem computacional para previsão de demanda de energia por classe de consumo para as distribuidoras de energia elétrica no Brasil	Modelo de previsão com rede MLP para três classes de consumo de uma distribuidora de energia.	Apresentado no seminário nacional de distribuição de energia elétrica (SENDI)	2018
Artigo 2: Apêndice B	Uma abordagem computacional para previsão e minimização dos custos de eletricidade no mercado de curto prazo dos distribuidores de eletricidade no Brasil	Estudo do impacto do atraso das variáveis de entrada para prever eletricidade no mercado de energia	Aceito pela revista <i>Engineering Review</i>	2020
Artigo 3: Apêndice C	Uma abordagem de seleção e classificação de variável baseada em PCA para previsão de carga de energia elétrica	Calcula a taxa de contribuição das variáveis de entrada e fornece uma melhor compreensão do processo subjacente que gerou os dados.	Aceito pela revista <i>International Journal of Energy Sector Management</i>	2019
Artigo 4: Apêndice D	Modelos de redes neurais autorregressivas não lineares para previsão de concentrações de gases dissolvidos em óleo de transformadores	Previsão de alta precisão com um modelo híbrido rede <i>NAR</i> e <i>wavelet</i> discreta para prever as concentrações de gás dissolvido em óleo do transformador	Publicado pela revista <i>Energies</i>	2018
Artigo 5: Apêndice E	Transformada do tipo <i>wavelet</i> para otimizar a ordem de um modelo de rede neural autorregressiva para prever a concentração de gás dissolvido no óleo do transformador de potência	Abordagem baseada em <i>wavelet</i> para determinar o atraso ideal para cada variável externa e para a variável de interesse em um modelo de previsão autorregressiva com variáveis exógenas. Redução do número de variáveis de entrada e ampliação do conhecimento da relação temporal entre os gases.	Publicado pela revista <i>Sensors</i>	2020

## 6. CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as considerações finais da tese, constituída pela conclusão e contribuições, que descreve a contribuição dos cinco modelos desenvolvidos para construção desta tese. Do mesmo jeito que apresenta as limitações dos modelos, bem como desenvolvimentos futuros.

### 6.1. Conclusões e contribuições

Esta pesquisa teve como objetivo principal estudar seleção de atributos e determinação de atrasos de tempo em modelos autorregressivos de previsão baseados em redes neurais artificiais.

Os objetivos da tese foram atingidos por meio do desenvolvimento e encadeamento de cinco artigos científicos, sendo que alguns foram publicados, outros aceitos em periódicos de relevância internacional. Neles foram discutidos desde a revisão da literatura até a validação das propostas.

A discussão apresentada nesta tese mostrou que a literatura mostra diversos trabalhos que se preocupam em desenvolver modelos para descobrir padrões em dados e realizar previsões com alta precisão. No entanto, pouca atenção é dada para a questão da seleção de atributos, seleção de variáveis de entrada e seleção do atraso de tempo da série temporal de interesse.

Na abordagem utilizada para responder o primeiro objetivo específico (artigo 1 – apêndice A), por exemplo, apesar dos critérios de seleção de modelos AIC e BIC selecionarem o modelo MA(1) dentre os modelos AR(1), AR(2), ARCH, ARMA e MA(2), as previsões realizadas com o modelo MA(1) alcançaram erros MAPEs entre 3,61% a 33,36% para os 12 meses (Tabela 6).

Enquanto o modelo com a rede neural MLP com algoritmo *backpropagation* utilizando as variáveis de entrada no instante  $t$ , apresenta resultados nas previsões com erros MAPEs inferiores as técnicas de séries temporais, conforme pode ser observado na Figura 21a, Figura 21b r Figura 21c.

A abordagem computacional utilizada com a rede MLP se mostrou eficiente com as entradas no instante  $t$ , porém, em alguns casos, variáveis de entrada nos instantes  $t-1$ ,  $t-2$  com atraso de tempo na demanda não foi possível obter resultados tão interessantes quanto com as

variáveis de entrada no instante  $t$ . Isso acontece porque a RNA associa os dados das três variáveis ao resultado, que nesse caso, é a energia elétrica. Quando há grandes alterações na energia elétrica ou nas variáveis de entrada, a RNA perde o poder de generalização e, por isso, não consegue aprender o comportamento daqueles dados (MEHDI e BIJARI, 2010).

Por outro lado, na abordagem desenvolvida para responder o segundo objetivo específico (artigo 2 – apêndice B), a abordagem utilizada neste trabalho mostrou que o atraso das variáveis de entrada (receita de fornecimento de energia, número de unidades consumidoras e tarifa média de fornecimento) no instante de tempo  $t-5$ ,  $t-10$  e  $t-20$  são úteis para prever o consumo anual de eletricidade da distribuidora de energia e auxiliar no processo de comercialização de energia elétrica no Brasil. Ou seja, o número de amostras para cada atraso de tempo (taxa de amostragem), altera a precisão da previsão para cada classe de consumo, o que obriga a estudar e, conseqüentemente, treinar uma rede neural MLP para cada classe de consumo.

Apesar das 12 previsões mensais feitas com as variáveis atrasadas em  $t-5$ ,  $t-10$  e  $t-20$  nas variáveis de entrada apresentarem erros maiores que as variáveis de entrada no tempo  $t$ , a soma das 12 previsões mensais apresentaram melhores resultados no consumo anual, pois, apesar de a distribuidora de energia estar sobrecontratada, o consumo anual da distribuidora estaria dentro do limite de 33987,29 GWh (Figura 22 e Tabela 9), o que não aconteceria com as variáveis de entrada no instante  $t$ .

Uma observação importante nessa abordagem é que, quando as variáveis de entrada estão atrasadas, será necessário encontrar o atraso mais adequado para as entradas de cada classe de consumo, além de treinar e selecionar a melhor configuração de neurônios para a cada classe de consumo. Ou seja, encontrar a taxa de amostragem adequada para cada classe de consumo, pois como são consumos independentes, possuem características distintas de consumo de eletricidade (Tabela 7 e 8).

Na abordagem utilizada para responder o terceiro objetivo específico (artigo 3 – apêndice C) foi proposto um modelo baseado em PCA para definir a taxa de contribuição para cada variável e, em seguida, selecionar as principais variáveis como entrada para a rede neural para previsão de carga de energia elétrica na Região Sudeste do Brasil.

Embora a previsão de carga de energia seja uma área para a qual bancos de dados com dezenas ou centenas de variáveis estão disponíveis, a abordagem pôde selecionar apenas sete dentre 26 variáveis que contribuem com mais de 89% para o modelo, o que gerou um modelo autorregressivo com alta precisão, atingindo erros de previsão MAPE, RMSE, MSE e MAE, de 1,74E-03%, 4,48E-04, 2,01E-07 e 1,39E-02, respectivamente, para os 31 dias de agosto de

2019. Esses resultados estão de acordo com o que Chatfield (2009) descreve em seu trabalho, pois o número mais sensível de variáveis explicativas é no máximo seis ou sete.

Além de melhorar a precisão da previsão e fornecer um preditor mais rápido, a abordagem baseada em PCA proposta neste trabalho para calcular a taxa de contribuição das variáveis de entrada, fornece melhor compreensão do processo subjacente que gerou os dados, o que é fundamental para a realidade brasileira devido ao clima acentuado e variações econômicas.

Na abordagem utilizada para responder o quarto objetivo específico (artigo 4 – apêndice D), foi proposto um estudo combinando um modelo de rede neural autorregressiva não linear com a transformada de *wavelet* discreta para prever as concentrações de gás dissolvido em óleo do transformador de potência.

O modelo apresentou baixa sensibilidade as variações nos dados de entrada para treinamento da rede NAR, um problema comum em abordagens baseadas em redes neurais (CHENG e YU, 2018). As funções da *wavelet* criaram versões simplificadas e esparsas dos dados originais, e o parâmetro de atraso de tempo  $d$ , geraram a robustez necessária para previsão de alta precisão. Além disso, a adoção da abordagem de validação cruzada aumentou a confiabilidade do modelo, levando a previsões de qualidade.

Além disso, o modelo foi capaz de prever múltiplos valores futuros das concentrações de gás, no caso do gás etileno, o erro máximo de previsão do modelo NAR–DWT foi cerca de 70% menor que os outros modelos testados, o que permite a emissão de alertas em caso de possíveis falhas, fornece estimativas de intervalo confiáveis para o tempo médio ocorrência de anormalidade. O resultado é coerente com a literatura, que afirma que os modelos de rede autorregressivas não lineares são menos sensíveis a dependências de longo prazo, além de apresentar melhores capacidades de generalização e aprendizado (BANOOKH e BARAKATI, 2012).

Por fim, a abordagem utilizada para responder o objetivo geral da tese (artigo 5 – apêndice E), foi apresentada uma abordagem baseada na transformação do tipo *wavelet*, PCA, correlação de Pearson e uma rede neural MLP.

A aplicação da *wavelet* cria aproximações esparsas das concentrações de gases com diferentes atrasos de tempo e diferentes comprimentos (séries temporais). Então, as técnicas PCA e correlação de Pearson são utilizadas para determinar o atraso de tempo ideal para cada variável de entrada no modelo de regressão automática, enquanto os atrasos ideais são utilizados na rede MLP para fazer as previsões para as concentrações de gases.

O modelo de previsão tem alta precisão (Figura 25 e Tabela 14), pois se concentra na tendência dos sinais de entrada das aproximações sem ruído calculadas pela transformação *wavelet* e considera o atraso de tempo ótimo para cada variável.

Além de simplesmente avaliar a relação de diferentes gases no quinto modelo autorregressivo, a proposta possibilita identificar a relação entre um gás específico e as aproximações atrasadas de todos os outros gases. Esse resultado é fundamental para os sistemas de monitoramento, pois indica que variações em diferentes gases levam tempos diferentes para serem refletidas nas concentrações de outros gases sob observação. Como exemplo, observou-se que as variações de  $H_2$  e  $CH_4$  levam 20 unidades de tempo para impactar em  $C_2H_4$ , enquanto  $C_2H_6$  e  $C_2H_2$  levam seis e duas unidades de tempo, respectivamente, para influenciar  $C_2H_4$  (Figura 24).

Na abordagem utilizada para responder o objetivo geral desta tese, foi possível observar que o desenvolvimento de um modelo autorregressivo, criado a partir de *wavelets Daubechies* para criar diferentes atrasos de tempo, e de uma abordagem para seleção do atraso de tempo utilizando PCA, melhora a precisão das previsões e mostra a relação temporal entre as variáveis de entrada.

Assim, a abordagem amplia a discussão da relação temporal entre as variáveis de entrada (concentrações de gases) e os modelos autorregressivos, proporcionando expansão do conhecimento da relação temporal entre os atrasos de tempo em cada variável, nesse caso, para os gases subjacentes ao processo de degradação do óleo isolante e do material isolante sólido.

## 6.2. Limitações do trabalho

As principais limitações desta pesquisa, apresentada no formato de artigos científicos, são aqui destacadas.

Os modelos desenvolvidos nesta tese foram aplicados apenas em uma base de dados específica, por exemplo, para o problema de previsão de energia elétrica para a distribuidora de eletricidade foi desenvolvido um modelo para o artigo 1 e um modelo para o artigo 2. No entanto, as abordagens adotadas no artigo 3, 4 e 5, não foram testadas na base de dados da distribuidora de eletricidade e na base de dados da usina geradora de eletricidade.

A abordagem utilizada no artigo 5 poderia ser empregada nos modelos autorregressivos dos artigos 1, 2, 3 e 4, criando séries temporais com atraso de tempo com auxílio das *wavelets Daubechies* e, posteriormente, selecionar as séries temporais com atraso de tempo que mais contribuem para o modelo.

Assim, as abordagens utilizadas nesta tese poderiam ser aplicadas em outras bases de dados, inclusive em base de dados que não faz parte do setor elétrico de energia.

Uma outra limitação deste trabalho, envolve o uso de diferentes *wavelets*. No artigo 5, por exemplo, não foi testado diferentes *wavelets* para criar atrasos de tempo.

### 6.3. Desenvolvimentos futuros

Como trabalho futuro, existem dois caminhos a serem seguidos, o primeiro, está relacionado ao uso da *wavelet* no modelo autorregressivo de previsão; ao utilizar essa abordagem, o comprimento do sinal de entrada tem a metade do comprimento do sinal de saída, o que dificulta o uso do comprimento total do sinal de saída no modelo de previsão.

O segundo caminho também está relacionado ao uso de *wavelet* em modelos autorregressivos. Na abordagem dos modelos autorregressivos publicados no artigo 4 e 5, foram utilizadas *wavelets* de primeira geração, que são translações e dilatações de uma função única. Ao passo que a transformada *wavelet* de segunda geração é uma alternativa da transformada *wavelet* discreta rápida, que usa um esquema de “levantamento” para predição espacial linear, não linear e operadores para implementar a transformada *wavelet*, além de ser computacionalmente mais rápida do que as *wavelets* de primeira geração.

Será que ao utilizar *wavelets* de segunda geração é possível construir séries temporais com atraso de tempo com melhores aproximações do que as *wavelets* de primeira geração? Qual o impacto das *wavelets* de segunda geração nos modelos autorregressivos de previsão?

O segundo caminho a ser seguido é uma continuação deste trabalho, e a resposta para esta pergunta, poderá ser respondida em um eventual pós-doutorado.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABEDINIA, O., AMJADY, N., ZAREIPOUR H. "A New Feature Selection Technique for Load and Price Forecast of Electrical Power Systems," in **IEEE Transactions on Power Systems**, vol. 32, no. 1, pp. 62-74, Jan. 2017.

ADHIKARI, R., AGRAWAL, R. **An Introductory Study on Time series Modeling and Forecasting**. 2013. 10.13140/2.1.2771.8084.

AGHABOZORGI, S.R., SHIRKHORSHIDI, A.S., WAH, T.Y. (2015). Time-series clustering - A decade review. **Inf. Syst.**, 53, 16-38.

ALBERT, C., DAUBECHIES, I., FEAUVEAU, J. C. Biorthogonal bases of compactly supported *wavelets*. **Communications on Pure and Applied Mathematics**, 45:485-500,1992.

AMJADY, Nima & KEYNIA, Farshid. (2009). Day-ahead price forecasting of electricity markets by a new feature selection algorithm and cascaded neural network technique. *Energy Conversion and Management*. 50. 2976-2982. 10.1016/j.enconman.2009.07.016

ARDALANI-FARSA, Muhammad; ZOLFAGHARI, Saeed. (2010). Chaotic time series prediction with residual analysis method using hybrid Elman–NARX neural networks. **Neurocomputing**. 73. 2540-2553. 10.1016/j.neucom.2010.06.004.

BAI, Yuting; JIN, Xuebo, WANG, Xiaoyi, SU, Tingli, KONG, Jianlei; LU, Yutian, "Compound Autoregressive Network for Forecast of Multivariate Time Series", **Complexity**, vol. 2019, Article ID 9107167, 11 p, 2019. <https://doi.org/10.1155/2019/9107167>

BANOOKH, A.; BARAKATI, S.M. Optimal design of double folded stub microstrip filter by neural network modelling and particle swarm optimization. **J. Microw. Optoelectron. Electromagn. Appl.** 2012, 11, 204–213. [CrossRef].

BARTLETT, M. S. "Properties of sufficiency and statistical tests". Proceedings of the Royal Statistical Society - Serie A, 1937, 60:268-282, 1937.

BOX, G.E.P., JENKINS, G.M., **Time Series Analysis: Forecasting and Control**, New Jersey: HoldenDay, 2015. 756p.

BROWNLEE, J. **Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs and LSTMs in Python**; Machine Learning Mastery: New York, NY, USA, 2018.

CERNY, C.A., and KAISER, H.F. “A study of a measure of sampling adequacy for factor-analytic correlation matrices”. **Multivariate Behavioral Research**, 1977. 12(1), 43-47.

CHATFIELD, Chris. **The analysis of time series an introduction. 5nd ed. Boca Raton, London, New York, Washington, D.C.:** Chapman & Hall/CRC, 2003.

CHENG, L.; YU, T. Dissolved Gas Analysis Principle Based Intelligent Approaches to Fault Diagnosis and Decision Making of Large Oil-Immersed Power Transformers: **A Survey. Preprints** 2018, 2018040109 (doi:10.20944/preprints201804.0109.v1).

CHRISTENSEN, Ronald. **Advanced Linear Modeling: multivariate, time series, and spatial data; nonparametric regression and response surface maximization.** 2nd ed. Springer-Verlag. New York: 2001.

COHEN, jacob.; COHEN, Patricia.; WEST, stephen G.; Aiken, Leona S. **Applied multiple regression/correlation analysis for the behavioral sciences.** 3rd ed. Lawrence Erlbaum Associates, Publishers: London, 2003.

DAUBECHIES, I. Orthonormal bases of compactly supported *wavelets*, comm. **Pure Appl. Math.** 41 (1988), no. 7,909-996, DOI 10.1002/cpa.3160410705. Mr951745.

DAUBECHIES, I. Orthonormal bases of compactly supported *wavelets*. **Commun. Pure Appl. Math.**, 41, pp. 909-996, 1988.

DAUBECHIES, I.; SWELDENS, W. Factoring *wavelet* transforms into lifting steps. **The Journal of Fourier Analysis and Applications**, 4, 247–269, 1998.

DE OLIVEIRA, H. M. (2007). **Análise de Sinais para Engenheiros: Uma abordagem via WAVELETS.** Sociedade Brasileira de Telecomunicações, 1. ed. Rio de Janeiro, RJ, 2007.

DIACONESCU, E. “The use of NARX neural networks to predict chaotic timeseries”, **WSEAS Transactions on Computer Research** 2008, 3 (3), 182–191.

DONG, Guangzhong.; ZHANG, Xu.; ZHANG, Chenbin.; ZONGHAI, Chen. (2015). A method for state estimation of lithium-ion batteries based on neural network model. *Energy*. 90, 879-888. 10.1016/j.energy. 2015.07.120.

FARIA JUNIOR, H.; COSTA, J.; OLIVAS, J. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. **Renewable and Sustainable Energy Reviews** 2015. 46. 10.1016/j.rser.2015.02.052.

FRIZZO STEFENON, Stéfano & DAL MOLIN RIBEIRO, Matheus & NIED, Ademir & MARIANI, Viviana & COELHO, Leandro & MENEGAT DA ROCHA, Diovana & GREBOGi, Rafael & RUANO, Antonio. (2020). *Wavelet* Group Method of Data Handling for Fault Prediction in Electrical Power Insulators. *International Journal of Electrical Power & Energy Systems*. 123. 106269. 10.1016/j.ijepes.2020.106269.

FU, Tak-Chung. (2011). A Review on Time Series Data Mining. **Engineering Applications of Artificial Intelligence**. 24. 164-181. 10.1016/j.engappai.2010.09.007.

GABOR, D. Theory of communication, **Journal of the institute of electrical engineers**, v. 93, n.26, pp 429-457, 1946.

GANGSAR, Purushottam & TIWARI, Rajiv. (2019). Diagnostics of mechanical and electrical faults in induction motors using *wavelet*-based features of vibration and current through support vector machine algorithms for various operating conditions. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*. 41. 10.1007/s40430-019-1574-5.

GHAVAMI, P. K. **Big data Analytics Methods: Analytics Techniques in Data Mining, Deep Learning and Natural Language Processing**: 2nd edition. Boston/Berlin: Walter de Gruyter Inc., 2020.

GOGOLEWSKI, D. Influence of the edge effect on the *wavelet* analysis process. **Measurement** 2020, 152, 107314.

GOODFELLOW, Ian., COURVILLE, Aaron . **Deep Learning: Adaptive Computation and Machine Learning series**. Massachusetts: The MIT Press, 2016.

GROSSMANN, A.; MORLET, J. "Decomposition of Hardy Functions into Square Integrable *Wavelets* of Constant Shape". *SIAM Journal on Mathematical Analysis*. 15 (4): 723–736, 1984. doi:10.1137/0515056.

GUNAY, M.E.: Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: case of Turkey. **Energy Policy**, 90 (2016), 92–101.

GUTTMAN, Irwin.; JAYALATH, Kalanka. P. **Statistics and probability with applications for engineers and scientists using Minitab, R AND JMP**. 2nd.ed. John Wiley & Sons, Inc.: EUA, 2020.

HAGAN, M. A.; DEMETH, H. B.; BEALE, M. H.; JESUS, O. D. **Neural Network Design**. Oklahoma. 2<sup>o</sup> Edition, 2014.

HAYKIN, S.: **Neural networks and learning machines**. 3rd. Ed. Pearson: Prentice Hall, Toronto, 2008.

HAYKIN, S.: **Neural Networks: A Comprehensive Foundation**. Pearson, Toronto, 1999.

HOCHREITER, S., SCHMIDHUBER, J. Long-term memory, **Neural Computation**. Vol. 9, No 8, p. 1735-1780, 1997.

HONG, T. “**Short term electric load forecasting**”. North Carolina State University (2010).

HUNTER, D.; Yu, H.; PUKISH III, M. S.; KOLBUSZ, J.; WILAMOWSKI, B. M. “Selection of proper neural network sizes and architectures: a comparative study,” **IEEE Transactions on Industrial Informatics** 2012, vol. 8, no. 2, pp. 228–240, 2012.

JOLLIFFE, I.T. *Principal Component Analysis*.: Springer Verlag, 2002.

JENSEN, A., LA COUR-HARBO, A. *Ripples in Mathematics - The Discrete Wavelet Transform*. 2001, Springer. 246 p.

KENDALL, M. G. **A NEW MEASURE OF RANK CORRELATION**. *Biometrika*. 30 (1-2): 81–93, 1938. ISSN 0006-3444. doi:10.1093/biomet/30.1-2.81.

KHASHEI, M., BIJARI, M. 2010. An artificial neural network (p,d,q) model for timeseries forecasting. **Expert Syst. Appl.** 37, 1 (January 2010), 479–489. DOI:<https://doi.org/10.1016/j.eswa.2009.05.044>.

KITCHIN, R. The Data Revolution: **Big Data, Open Data, Data Infrastructures and Their Consequences**. 1st ed. Sage Publications Ltd, California, USA, 2014.

KREUGER, F.H. Partial Discharge Detection in High-Voltage Equipment; Butterworth-Heinemann: Woburn, MA, USA, 1989.

KUMAR, V., MINZ, S. Feature selection: A literature review. **Smart Computing Review**, 2014. 4. 211-229. 10.1145/2740070.2626320.

LAHOUAR, J. B. H. Slama. “Day-ahead load forecast using random forest and expert input selection”. **Energy Conversion Management**, 103 (2015), 1040–1051.

LANG, K., ZHANG, M., YUAN, Y. & YUE, X. (2018). Short-term load forecasting based on multivariate time series prediction and weighted neural network with random weights and kernels. *Cluster Computing*, 10.1007/s10586-017-1685-7.

LEWIS, N. D. **Neural networks for time series forecasting with R: Intuitive Step by Step Blueprint for Beginners**. Quantity purchases, 2017.

LIANG, Y., NIU, D., HONG, W. C. “Short term load forecasting based on feature extraction and improved general regression neural network model”. **Energy**, 2018. 166, 653–663.

LIN, J.; LEI, S.; YINGJIE, Y.; GEHAO, S.; DA, X.; XIUCHEN, J. Prediction Method for Power Transformer Running State Based on LSTM\_DBN Network. **Energies**, MDPI, Open Access Journal 2018, vol. 11(7), pages 1-14, July.

LIU, J.; ZHENG, H.; ZHANG, Y.; LI, X.; FANG, J.; LIU, Y.; LIAO, C.; LI, Y.; ZHAO, J. Dissolved Gases Forecasting Based on *Wavelet* Least Squares Support Vector Regression and Imperialist Competition Algorithm for Assessing Incipient Faults of Transformer Polymer Insulation. **Polymers (Basel)** 2019, 11(1):85. Published 2019 Jan 8. doi:10.3390/polym11010085.

MALLAT, Stéphane. **A Wavelet Tour of Signal Processing: The Sparse way**, 2nd Edition, “in *Computer Music Journal*”, vol. 31, no. 3, pp. 83-85, Sept. 2009.

MAS'UD, Abdullahi & STEWART, Blair & MCMEEKIN, Scott. (2014). Application of an ensemble neural network for classifying partial discharge patterns. *Electric Power Systems Research*. 110. 154–162. 10.1016/j.epsr.2014.01.010.

MENEZES JR, Jose Maria; BARRETO, Guilherme. (2008). Long-term time series prediction with the NARX network: An empirical evaluation. **Neurocomputing**. 71. 3335-3343. 10.1016/j.neucom.2008.01.030.

MILLS, T. C. **Applied time series analysis: A practical guide to modelling and forecasting**. 1st. ed. Academic Press. Loughborough, United Kingdom, 2019.

MIRKIN, B. **Core concepts in data analysis: summarization, correlation and visualization**. Undergraduate topics in computer science. Springer-Verlag, London, United Kingdom, 2011.

MITCHELL. T. M. **Machine learning**. Ithaca: McGraw-Hill Science, 1997.

MORETTIN, P.A., DE CASTRO TOLOI, C.M. **Análise de séries temporais**. Edgard Blucher, 2006.

NASEHI, S., POURGHASSEM, H. "A novel effective feature selection algorithm based on S-PCA and *wavelet* transform features in EEG signal classification," 2011 **IEEE 3rd International Conference on Communication Software and Networks**, Xi'an, 2011, pp. 114-117.

NENGBAO, Liu; BABUSHKIN, Vahan; AFSHARI, Afshin. (2014). Short-Term Forecasting of Temperature Driven Electricity Load Using Time Series and Neural Network Model. **Journal of Clean Energy Technologies**. 2. 10.7763/JOCET.2014.V2.149.

NIELSEN, Michael. **Deep learning: chapter 6**. Disponível em:<<http://neuralnetworksanddeeplearning.com/chap6.html>>. Acesso em: 23 de fev. 2017.

NOORI, M.; EFFATNEJAD, R.; HAJIHOSSEINI, P. (2017). Using Dissolved Gas Analysis Results to Detect and Isolate the Internal Faults of Power Transformers by Applying a *Fuzzy*

Logic Method. **IET Generation, Transmission & Distribution** 2017. 11. 10.1049/iet-gtd.2017.0028.

NOURANI ESFETANG, Naser & KAZEMZADEH, Rasool, 2018. "A novel hybrid technique for prediction of electric power generation in wind farms based on WIPSO, neural network and *wavelet* transform," *Energy*, Elsevier, vol. 149(C), pages 662-674.

ONS. "**Operador Nacional do Sistema Elétrico. Resultados da Operação**: "Histórico da operação". Disponível em in <<http://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>>. Acesso em: novembro. 2019.

PEARSON, Karl. (1901), "On lines and planes of closest fit to systems of points in space". **Philosophical Magazine**, 2 (6), 559–572.

PEARSON, Karl. **On the General Theory of Skew Correlation and Non-linear Regression**. London: Dulau&Co, 1905.

PEREIRA, F.H.; BEZERRA, F.E.; JUNIOR, S.; SANTOS, J.; CHABU, I.; SOUZA, G.F.M.; MICERINO, F.; NABETA, S.I. Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations. **Energies** 2018, *11*, 1691.

PRASOJO, Rahman A. Power Transformer Insulation Assessment Based on Oil-Paper Measurement Data Using SVM-Classifer. **Preprints** 2018, 2018060002 (doi: 10.20944/preprints201806.0002.v1).

RAYMOND, Wong Jee Keen & ILLIAS, Hazlee & BAKAR, Ab & MOKHLIS, Hazlie. (2015). Partial Discharge Classifications: Review of Recent Progress. **Measurement**. 68. 10.1016/j.measurement.2015.02.032.

RUIZ, L.G.B.; CUÉLLAR, M.P.; CALVO-FLORES, M.D.; JIMÉNEZ, M.D.C.P. An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. **Energies** 2016, *9*, 684.

RHIF, M.; ABBES, A.; FARAH, I.; MARTINEZ, B.; SANG, Y.-F. (2019). *Wavelet* Transform Application for/in Non-Stationary Time-Series Analysis: A Review. **Applied Sciences** 2019. *9*. 1347. 10.3390/app9071345.

ROITMAN, V. DE L. **Um modelo computacional de redes neurais para predição do índice de desemprego aberto**. 2001. 145p tese (Doutorado em Ciências em Engenharia Civil) - Universidade Federal do Rio de Janeiro. Rio de Janeiro - RJ, 2001.

SCHIMIT, P. H., and PEREIRA, F. H. (2018), “Disease spreading in complex networks: A numerical study with Principal Component Analysis”. **Expert Systems with Applications**, 97, 41-50.

SHEELA, K.G.; DEEPA, S.N.; Review on methods to fix number of hidden neurons in neural networks. **Math Prob Eng** 2013 (Article ID 425740, 11 pages).

SINGH, D. A., BALAMURUGAN, S.A., LEAVLINE, E.J. (2016). **Literature Review on Feature Selection Methods for High-Dimensional Data**. *International Journal of Computer Applications*, 136, 9-17. (2016). **Literature Review on Feature Selection Methods for High-Dimensional Data**. International Journal of Computer Applications. 136. 10.5120/ijca2016908317.

SOUZA, Reinaldo. Castro. Metodologias para a análise e previsão de séries temporais univariadas e multivariadas. **Brazilian Review of Econometrics**. Vol. 1 No. 2 (1981).

SPEARMAN, C. “General intelligence” objectively determined and measured. **American Journal of Psychology**, New York, n. 15, p. 201-293, 1904.

TANG, J., ALELYANI, S., LIU, H. Feature selection for classification: A review. In *Data Classification: Algorithms and Applications*. CRC Press. 2014. p. 37-64 <https://doi.org/10.1201/b17320>.

TANG, W. H., WU, Q. H. (2013). **Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence**. Springer Publishing Company, Incorporated.

TIUMENTSEV, Y. V., EGORCHEV, M. V., **Neural network modelling and identification of dynamical systems**. 1st edition. Academic Press is an imprint of Elsevier: 2019.

VELMURUGAN, T., ANURADHA, C. Performance evaluation of feature selection algorithms in educational data mining. **Perform. Eval.** 5(02) 2016.

WANG, Y., CHEN, Q., KANG, C., and XIA, Q., "Clustering of Electricity Consumption Behavior Dynamics Toward *Big data* Applications," in **IEEE Transactions on Smart Grid**, vol. 7, no. 5, pp. 2437-2447, Sept. 2016.

WEI, W. W. S. **Time Series Analysis: Univariate and Multivariate Methods**. 2nd edition. Pearson, 2006.

YANG, Haimin; PAN, Zhisong; TAO, Qing and QIU, Junyang 2018, Online learning for vector autoregressive moving-average time series prediction, **Neurocomputing**, vol. 315, pp. 9-17, doi: 10.1016/j.neucom.2018.04.011.

YANG, A., LI, W., YANG, X. (2019), "Short-term electricity load forecasting based on feature selection and least squares support vector machines". **Knowledge-Based Systems**, 163, 159–173.

YEN, G. G. AND W. F. LEONG. "Fault classification on vibration data with *wavelet*-based feature selection scheme." **31st Annual Conference of IEEE Industrial Electronics Society, 2005. IECON 2005**. (2005): 6 pp.

ZAID, M. A. **Correlation and Regression Analysis, Textbook: Organisation of Islamic cooperation statistical economic and social research and training centre for Islamic countries**, Oran, Ankara, Turkey, 2015.

ZHANG, A. **Data Analytics: Practical Guide to Leveraging the Power of Algorithms, Data Science, Data Mining, Statistics, Big Data, and Predictive Analysis to Improve Business, Work, and Life**, CreateSpace Independent Publishing Platform, 2017.

ZHENG, H.; ZHANGA, Y.; LIUA, J.; WEIA, H.; ZHAOD, J.; LIAO, R. A novel model based on *wavelet* LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. **Electric Power Systems Research 2018**. 155. 10.1016/j.epsr.2017.10.010.

## APÊNDICES

### APÊNDICE A (Artigo 1) – Uma abordagem computacional para previsão de demanda de energia por classe de consumo para as distribuidoras de energia elétrica no Brasil.

Seminário Nacional de Distribuição de Energia Elétrica (SENDI) – Seminário.

*Presented:* 20 a 23 de novembro de 2018.

*Corresponding author:* Me. Francisco Elânio Bezerra.

*Corresponding Author's Institution:* Universidade Nove de Julho.

*First Author:* Elânio Bezerra; Fabio Henrique Pereira

*Order of Authors:* Francisco Elânio Bezerra, Fabio Henrique Pereira

*Order of Authors:* Francisco Elânio Bezerra<sup>1</sup>, Cleber Gustavo Dias<sup>2</sup>, Fabio Henrique Pereira<sup>1,2\*</sup>

**Resumo:** As distribuidoras de energia no Brasil compram eletricidade por meio de leilões. Caso haja ineficiência na contratação, a Câmara de Comercialização de Energia Elétrica, que opera a comercialização de energia, aplica penalidades. Assim, este artigo propõe uma abordagem computacional para previsão de eletricidade utilizando unidades consumidoras, receita e tarifa média de fornecimento de energia como variáveis de entrada na rede neural artificial com algoritmo *backpropagation*. Além disso, técnicas de séries temporais selecionadas pelos critérios *Bayesiano* e *Akaike* são empregadas para prever eletricidade, o que é importante, pois serve de suporte na compra da energia em leilões e no processo de liquidação de diferenças no mercado de curto prazo, bem como para gerenciamento energético, atendimento ao cliente e faturamento das distribuidoras. Os resultados mostram que a rede multicamada com algoritmo *backpropagation* é capaz de aprender o comportamento dos dados que influenciam o consumo elétrico por classe de consumo. Este trabalho pode ser utilizado para acompanhar a evolução no consumo de eletricidade e, conseqüentemente, ajudar as distribuidoras no processo de contratação de energia em leilões, pode reduzir as perdas por multas devido à sub e sobrecontratação de energia e reduzir os custos da distribuidora.

**Palavras-chave:** comercialização de energia, rede neural artificial, previsão de eletricidade, rede multicamadas, séries temporais.

*A computational approach to energy demand forecasting for consumer class for electricity distributors in Brazil*

**Abstract:** *Energy distributors in Brazil buy electricity through auctions. In case of inefficiency in the contracting, the Electric Energy Trading Chamber, which operates the energy trading, imposes penalties. Thus, this article proposes a computational approach to electricity forecasting using number of consumers, revenue and average energy supply tariff as input variables in an artificial neural network with a backpropagation algorithm. In addition, time series techniques selected by the Bayesian and Akaike criteria have used to forecast electricity. This electricity forecast is important, as it supports the purchase of electricity at auctions and the settlement of differences in the short-term market and also for energy management, customer service and billing of distributors. The results show that the multilayer network with backpropagation algorithm is able to learn the behaviour of the data that influence electricity consumption by consumption class. This work can be used to monitor the evolution in electricity consumption and, consequently, help distributors in the process of contracting energy in auctions, can reduce the loss of fines due to under and over-contracting of energy and reduce the costs of the energy distributor.*

**Keywords:** *energy trading, artificial neural network, electricity forecasting, multilayer network, time series.*

## 1. INTRODUÇÃO

A evolução tecnológica e o crescimento populacional dos últimos anos contribuíram para que o índice de expansão do consumo de energia elétrica fosse superior ao Produto Interno Bruto (PIB), sendo necessário gerar mais energia elétrica (EPE, 2014). Com isso, o investimento nessa área é essencial para suprir a demanda dos consumidores, aumentar a oferta de energia, reduzir o custo da comercialização e trazer crescimento para o Brasil.

Desde a criação do novo modelo de comercialização de eletricidade no país, estabelecido pelo Decreto N° 5.163 em 2004, as distribuidoras são obrigadas a comprar a energia por meio de leilões realizados no Ambiente de Contratação Regulado (ACR) para atender seus consumidores (classe de consumo), tais como: comercial, consumo próprio, iluminação pública, industrial, poder público, residencial, rural, serviço público (ANEEL, 2016).

Esse novo modelo também obriga as distribuidoras a atender 100% da sua demanda de tal forma que, caso haja mais consumo do que o contratado de energia, a distribuidora será

penalizada. Em contrapartida, não é permitido contratar energia acima de 5% de sua demanda anual, pois, acima desse percentual, a distribuidora também sofrerá penalidades aplicadas pela ANEEL (ANEEL, 2016).

Mesmo com um modelo de comercialização diferente do Brasil, a introdução da reestruturação da indústria de energia elétrica na Espanha fez com que a previsão de preço da eletricidade se tornasse atividade importante para todos os participantes do mercado de eletricidade (PINDORIYA et al., 2008).

A previsão de preços no mercado de curto prazo (*spot*) oferece condições para os fornecedores de energia construírem estratégias de licitação, apresenta condições para os consumidores maximizarem a utilização da energia comprada no mercado *pool* e ainda pode oferecer condições para usar a capacidade de autoprodução para se proteger contra preços elevados (PINDORIYA et al., 2008).

Com isso, a previsão de demanda de energia elétrica para as distribuidoras se tornou fundamental. Além de servir como base para resolver vários problemas de tomada de decisão, tais como: planejamento de preços, tarifas, operação das redes de distribuição, planejamento de produção, gerenciamento de carga, atendimento ao cliente e faturamento (YALCINOZ E EMINOGLU, 2005; BANTUGON E GALLANO, 2016).

Isso faz com que o monitoramento e a previsão de fornecimento em tempo real sejam relevantes e vitais para empresas que disponibilizam energia elétrica, pois têm implicações no planejamento e otimização de energia. Além disso, pode servir de base para analisar o consumo em relação a outras variáveis, como o número de consumidores que estão aumentando ou diminuindo da classe de consumo comercial, industrial, residencial, entre outras (DEB et al., 2017).

Diante desse cenário, técnicas de inteligência computacional vêm sendo utilizadas para auxiliar as distribuidoras no processo de previsão de demanda e, conseqüentemente, auxiliar as distribuidoras na compra de energia elétrica.

Ramos e Susteras (2006) descreveram que o uso de técnicas de inteligência computacional pode servir como apoio para as distribuidoras e geradoras desenvolverem suas estratégias de comercialização e auxiliarem no processo de compra e venda de energia elétrica nos leilões. Para isso, propuseram um modelo de otimização utilizando algoritmo genético que simula o comportamento das distribuidoras e calcula o prejuízo ou eventuais ganhos correspondentes a cada indivíduo da geração.

Campos (2008) desenvolveu um trabalho de previsão de demanda utilizando técnicas de inteligência computacional: ARIMA, NARIMA, Rede *Neuro-Fuzzy* (RNF) e Rede Neural

Artificial (RNA) para a Inglaterra e a cidade de Minas Gerais no Brasil, em seguida, comparou a eficiência na previsão por meio dos índices de desempenho *Mean Percentual Error* (MPE), *Mean Absolute Percentual Error* (MAPE) e *Root Mean Square Error* (RMSE) para as classes de consumo da distribuidora CEMIG-Distribuição: residencial, comercial, industrial, rural, poder público, iluminação pública, serviço público e consumo próprio.

Swaroop e Hussein (2012) utilizaram rede neural artificial (RNA) para previsão da carga de energia elétrica e assim, determinar o planejamento e operação do sistema de potência da região Al Batinah em Oman. No trabalho, foram utilizados os dados mensais históricos de 2007, 2008 e 2009 de carga, temperatura e umidade para preverem a carga elétrica de 2010.

Um sistema híbrido utilizando RNA e ARIMA foi abordado por Khashei e Bijari (2010) para produzir um modelo mais preciso na previsão, em que, primeiro, um modelo de média móvel autorregressivo integrado (ARIMA) é usado para gerar os dados necessários das séries temporais em estudo, em seguida, a RNA é usada para modelar os dados gerados pelo modelo ARIMA e utilizado para prever o valor futuro das séries temporais.

O uso de RNA e ARIMA também foi utilizado por Pati e Shukla (2014) para prever o número de erros de um aplicativo de software, com intuito de ajudar gerentes de softwares a tomarem decisões sobre alocação de recursos e investimentos, na qual os resultados mostraram eficiência na previsão utilizando modelo híbrido.

No trabalho de Wang e Meng (2012), as técnicas ARIMA e RNA são utilizadas para fazer a previsão de consumo de energia elétrica da província de Hebei na China, onde foram utilizados dados históricos de 1980 a 2008, na qual a eficiência das previsões foram comparadas por meio do erro absoluto médio (MAE), erro quadrático médio da raiz (RMSE) e erro médio percentual absoluto (MAPE).

Em trabalho mais recente, Zhuang et al. (2016) desenvolveram a previsão de demanda de energia elétrica em curto prazo com RNA, Rede Neuro *Fuzzy* (RNF), Rede Neural *Wavelet* (RNW), em que foi considerada demanda por hora durando 28 dias. Os resultados das três técnicas foram comparados pelo erro percentual absoluto médio (MAPE), sendo que a técnica RNW se mostrou mais eficiente.

No trabalho de Deb et al. (2017) foi apresentado estudo sobre a abrangência de algumas técnicas de previsão de séries temporais, tais como: ARIMA, RNA, *Fuzzy*, *Neurofuzzy*, *Wavelet*, Algoritmos evolucionários, algoritmos híbridos, que funcionam com duas ou mais técnicas para prever o consumo de energia em prédios. Os resultados mostraram que o uso de abordagens híbridas é mais eficiente, pois, caso uma técnica não possa prever com eficiência todos intervalos desejados, outra técnica pode suprir a deficiência.

Bantugon e Gallano (2016) comparam o desempenho do método *Holt-Winters* e da RNA na previsão de demanda de energia elétrica a curto e longo prazo para o estado das Filipinas, nos Estados Unidos, em que os resultados foram comparados por meio dos indicadores de desempenho MAPE, erro médio absoluto (MAE) e o (RMSE). Os resultados mostraram que o método *Holt-Winters* fizeram previsões melhores que a RNA, chegando a 9% para a previsão de curto prazo e a 3% na previsão de longo prazo.

Khashei e Bijari (2010) descreveram que o uso de RNAs para reconhecimento de padrões, classificação, agrupamento e previsão de séries temporais vêm sendo utilizadas com alto grau de precisão em diversas aéreas.

Dessa forma, percebe-se que diversas técnicas de inteligência computacional são usadas para previsão de demanda de energia elétrica, tendo como variáveis demanda, clima, preço, entre outras.

Por isso, o presente trabalho propõe o uso de inteligência computacional (rede MLP com algoritmo *backpropagation*) e técnicas de séries temporais para previsão de demanda por classe de consumo (residencial, industrial e comercial) para auxiliar as distribuidoras do Brasil na compra e conseqüentemente, na diminuição dos riscos de penalidades.

Este trabalho está organizado da seguinte forma: a seção 1 mostra alguns trabalhos relacionados às técnicas utilizadas na previsão de demanda de energia elétrica, a seção 2 apresenta as regras de comercialização de energia elétrica no Brasil, fundamentação teórica sobre redes neurais artificiais, séries temporais e critérios para seleção de modelos, a seção 3 mostra a metodologia abordada no trabalho, a seção 4 resultados e, por último, apresenta a conclusão.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. REGRAS DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA

Para comprar energia elétrica no ACR, as distribuidoras são obrigadas a participar de leilões para três e cinco anos à frente (A-3 e A-5), denominados leilões de energia nova, em que os contratos, na modalidade, podem ser feitos com prazo mínimo de quinze e máximo de trinta e cinco anos. Existe a possibilidade de comprar energia nos leilões de energia existentes para um ano à frente, chamado A-1 e nos leilões de ajuste com entrega imediata, cujos contratos podem ser, no mínimo, três e, no máximo, quinze anos. (ANEEL, 2016c).

Os procedimentos para regulação tarifária (PRORET) foram aprovados pela Resolução Normativa n° 435/2011 e são utilizados para regulamentar os processos tarifários; estão organizados em 12 módulos, subdivididos em submódulos. O submódulo 4.3 mostra a sobrecontratação de energia e exposição ao mercado de curto prazo, detalhando os mecanismos envolvidos para verificar a exposição da distribuidora (ANEEL, 2016).

A contabilização da energia é feita pela Câmara de Comercialização de Energia Elétrica (CCEE) como forma de liquidar as diferenças entre contratação e consumo de energia elétrica, sendo dependente do balanço energético da distribuidora (BED) que é composto pela diferença entre consumo total (TRC) e contratos realizados pelas distribuidoras (PCL), todos em megawatt-hora (MWh), como pode ser observado na Equação 24 (ANEEL, 2016a).

$$BED = -TRC + PCL \quad (24)$$

Nas regras de comercialização de energia: consolidação de resultados da CCEE, é descrita a forma como é feita a contabilização do volume de energia negociado no Mercado de Curto Prazo, que depende do produto do balanço energético BED em MWh pelo preço do PLD em reais, calculados de forma semanal por patamar de carga para cada submercado (região do país), como pode ser observado na Equação 25 (ANEEL, 2016a).

$$MCP = \pm BED \times PLD \quad (25)$$

A consolidação do resultado total mensal no mercado de curto prazo (*TM MCP*) ou exposição financeira da distribuidora no *MCP*, vista na Equação 26 e descrita nas regras de comercialização: consolidação de resultados da CCEE demonstra que o resultado *TM MCP* que é calculado mensalmente para cada distribuidora e depende do resultado da somatória no *MCP*,

na qual o resultado  $TM\ MCP$  pode ser positivo, quando a distribuidora terá mais energia contratada do que consumo, ou negativa, quando há mais consumo do que contrato de energia. Os índices “r” e “s” representam patamar de carga e semana respectivamente (ANEEL, 2016a).

$$TM\ MCP = \pm \sum_{r,s} (MCP) \quad (16)$$

A CCEE contabiliza o montante de energia elétrica contratada, consumida e vendida no mercado de curto prazo para cada distribuidora e faz a apuração da energia vendida e comprada no MCP (VMCP e CMCP) é feita por mês, como forma de obter o excedente de contratação da distribuidora no ano. Quando o volume de energia vendido no MCP (VMCP), for maior que CMCP (VMCP > CMCP), há sobrecontratação da distribuidora, quando o volume de energia comprado em contratos for maior do que o volume vendido no MCP (CMCP > VMCP), há exposição da distribuidora (ANEEL, 2016a).

A energia exigida no ano  $Eexano$  pela distribuidora é contabilizada somando a energia exigida por mês e o limite da sobrecontratação no ano  $LSOBREano$  é de 5% da energia anual exigida, conforme mostra a Equação 27 (ANEEL, 2016a).

$$LSOBREano = 1,05 \times Eexano \quad (27)$$

Quando a energia contratada pela distribuidora for insuficiente para atender seu mercado, poderá ser penalizada pela subcontratação, compra de energia no curto prazo e pelo erro na contratação, cujo custo real da energia não repassada depende do volume contratado, PLD e do valor anual de referência.

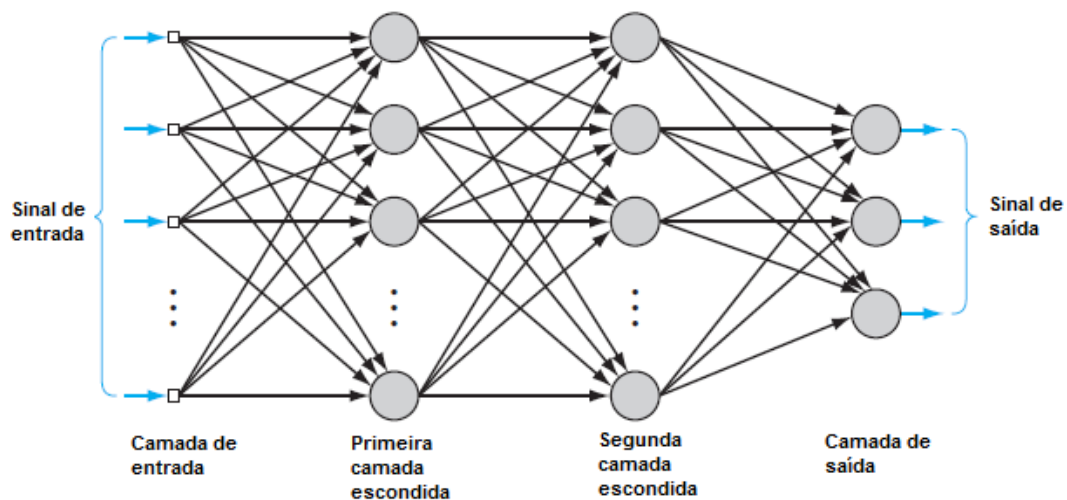
## 2.2. REDES NEURAS ARTIFICIAIS

Rede neural é um processador distribuído que trabalha de forma paralela, seu conhecimento é adquirido pela rede por meio de aprendizagem na qual os pesos sinápticos são utilizados para conhecimento; sua principal característica é armazenar conhecimento experimental para ser utilizado (HAYKIN, 2008).

O *perceptron* multicamadas (MLP) é um tipo de rede neural artificial que pode trabalhar com várias superfícies de decisões não lineares, utilizando várias regras de aprendizagem e possui alto grau de conectividade por meio dos pesos sinápticos da rede (HAYKIN, 2008).

A Figura 26 mostra a estrutura de uma rede MLP com camada de entrada, duas camadas ocultas e uma camada de saída. A camada de entrada recebe os dados de entrada da rede, a primeira camada oculta é alimentada a partir da camada de entrada constituída pelos nós de origem, as saídas resultantes da primeira camada oculta são aplicadas à próxima camada oculta e a camada de saída tem a função de processar as informações. O sinal de entrada da rede MLP propaga-se em direção única, camada por camada. (HAYKIN, 2008).

**Figura 26 – Estrutura de uma rede MLP com duas camadas ocultas**



Fonte: HAYKIN, 2008, p. 124

### 2.3. TÉCNICAS DE SÉRIES TEMPORAIS E CRITÉRIO PARA SELEÇÃO DE MODELOS

Uma série temporal consiste em qualquer conjunto de observações organizadas no tempo, em que a característica essencial é que as observações próximas são dependentes e a análise da dependência requer o uso de algumas técnicas, para isso, há necessidade de desenvolver modelos estocásticos e dinâmicos para dados de séries temporais (MORETTIN e TOLOI, 2006; BOX e JENKINS, 2015).

Característica essencial da série temporal é que as observações próximas são dependentes e a análise da dependência requer uso de algumas técnicas, para isso, há necessidade de desenvolver modelos estocásticos e dinâmicos para dados de séries temporais (BOX e JENKINS, 2015).

Os modelos estocásticos são representados por modelos estacionários e não estacionários; nos modelos estacionários, é assumido que o processo permanece em equilíbrio estatístico com propriedades probabilísticas que não mudam ao longo do tempo, variando sobre

um nível médio constante fixo e com variância constante, enquanto nos modelos não estacionários, os dados se alteram ao longo do tempo (BOX e JENKINS, 2015).

A previsão está relacionada à antecipação de algum evento e abrange áreas de negócios, indústria, governo, economia, ciências ambientais, medicina, ciências sociais, política, finanças, entre outras. Os problemas de previsões podem ser classificados em curto prazo, em que envolvem previsões apenas de alguns períodos de tempo (dias, semanas, meses) no futuro, previsões em médio prazo, que se estendem a alguns meses, e previsões de longo prazo, que estão relacionadas a previsões acima de um ano (MONTGOMERY et al., 2015).

Existem diversos métodos utilizados para fazer previsões, alguns trabalham com tendência, sazonalidade, função de autocorrelação (FAC), que é uma maneira de medir e explicar a associação entre observações em uma série de tempo. Existem métodos que procuram explorar a estrutura de autocorrelação dos resíduos de previsão do período atual com aqueles ocorridos em períodos anteriores, entre outras características (MONTGOMERY et al., 2015).

As técnicas de séries temporais são escolhidas por meio de critérios de seleção de modelos que procuram selecionar modelos que estão mais próximos da realidade.

O Critério de Informação Akaike (AIC) foi criado pelo estatístico japonês Hirotugu Akaike (1973), fundado sobre a teoria da informação, funciona minimizando a distância de Kullback-Leibler (K-L) que é a medida da distância entre o modelo verdadeiro e o modelo estatístico para seleção dos modelos. De acordo com Morettin e Toloi (2006), quanto menor o valor de AIC, melhor será o ajuste do modelo.

A Equação 28 mostra o cálculo para determinar o valor de AIC, que depende do estimador de máxima verossimilhança  $\hat{\sigma}_p^2$  (compatível com a informação disponível), e do número de parâmetros  $p$ .

$$AIC = -2\log\hat{\sigma}_p^2 + 2p \quad (28)$$

Em 1978, Gideon Schwarz desenvolveu o critério de informação Bayesiana (BIC), que maximiza a probabilidade de identificar o verdadeiro modelo dentre os modelos avaliados e funciona como AIC, quanto menor seu valor, melhor será o ajuste do modelo. A Equação 29 mostra o desenvolvimento desse modelo, em que  $\hat{\sigma}_k^2$  é o estimador de máxima verossimilhança, “ $p$ ” o número de parâmetros a serem estimados, e “ $m$ ” é o número de observações da amostra:

$$BIC = -2\log\hat{\sigma}_k^2 + p \log m \quad (29)$$

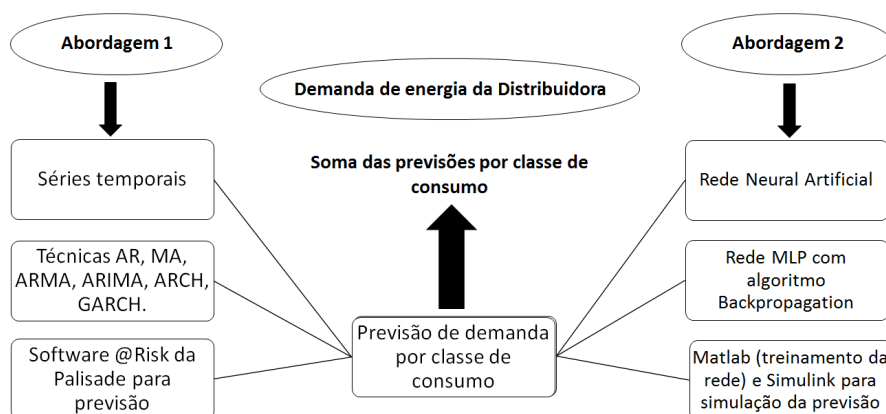
### 3. DESENVOLVIMENTO

#### 3.1. ABORDAGEM COMPUTACIONAL PARA PREVISÃO DE DEMANDA

Com dados históricos de receita de fornecimento de energia (RFE), número de unidades consumidoras (UC) e tarifa média de fornecimento de energia (TMF) em uma rede MLP com algoritmo *backpropagation* são feitas previsões da demanda por classe de consumo (comércio, indústria e residência). Sendo que a eficiência das previsões com RNA e das técnicas de séries temporais para as classes de consumo residencial, industrial e comercial são feitas por meio do índice de desempenho MAPE.

Assim, este trabalho segue duas abordagens para previsão de demanda por classe de consumo, como pode ser observado na Figura 27, na qual a primeira utiliza técnicas de séries temporais e a segunda rede neural artificial.

**Figura 27 – Abordagem para previsão de demanda de energia elétrica**



A abordagem 1 consiste em prever a demanda de energia elétrica das três classes de consumo utilizando técnicas de séries temporais (AR, MA, ARMA e ARCH), que por meio do critério AIC e BIC, escolhem a melhor técnica de série temporal, já que estas técnicas procuram por modelos que estão mais próximos da realidade.

As técnicas de séries temporais aplicadas por meio do *software @Risk da Palisade* que é um *plugin* do *Excel*, utilizado para simulação, análise de riscos, trabalha com simulação de Monte Carlo, mostra diversos gráficos estatísticos, faz cálculos de correlações de variáveis, entre outros.

Dessa forma, antes de fazer a previsão da demanda de energia para 12 meses, referente a 2016, para as três classes de consumo utilizando dados históricos de janeiro de 2003 a dezembro de 2015, os dados de cada classe de consumo são transformados em estacionários,

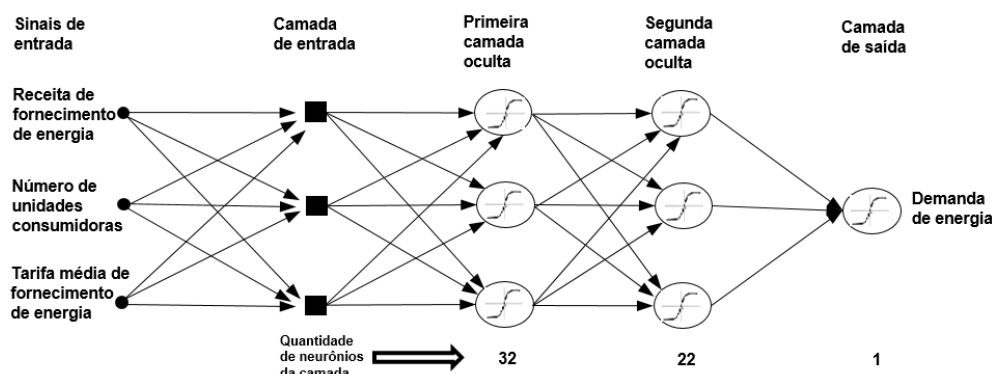
pois, assim, a média, a variância e a autocorrelação não mudam ao longo do tempo, retirando a tendência dos dados das três classes de consumo.

Já na abordagem 2, são utilizados dados históricos de receita de fornecimento de energia elétrica, número de unidades consumidoras e tarifa média de fornecimento de energia para prever a demanda por classe de consumo nos instantes  $t$ ,  $t-1$ ,  $t-2$ .

Nessa abordagem, é aplicada uma rede MLP com algoritmo *backpropagation*, na qual foram utilizadas 156 observações (meses) de janeiro de 2003 a dezembro de 2015 para treinamento da rede, deixando 12 meses referentes a 2016 para comparação com os valores reais de demanda, sendo que a eficiência na previsão é realizada por meio do índice do desempenho MAPE.

A arquitetura da rede proposta no trabalho pode ser observada na Figura 28, na qual é possível notar que a rede possui três camadas. A camada de entrada foi constituída pelas variáveis de receita de fornecimento de energia, número de unidades consumidoras e tarifa média de fornecimento, já as camadas ocultas foram constituídas pela primeira camada oculta que é alimentada pelas variáveis de entrada e possui 32 neurônios e pela segunda camada oculta que é alimentada pela saída da primeira camada oculta e possui 22 neurônios. Por fim, a camada de saída possui um neurônio e tem o papel de processar as informações de saída da segunda camada oculta.

Figura 28 – Arquitetura da RNA empregada no modelo

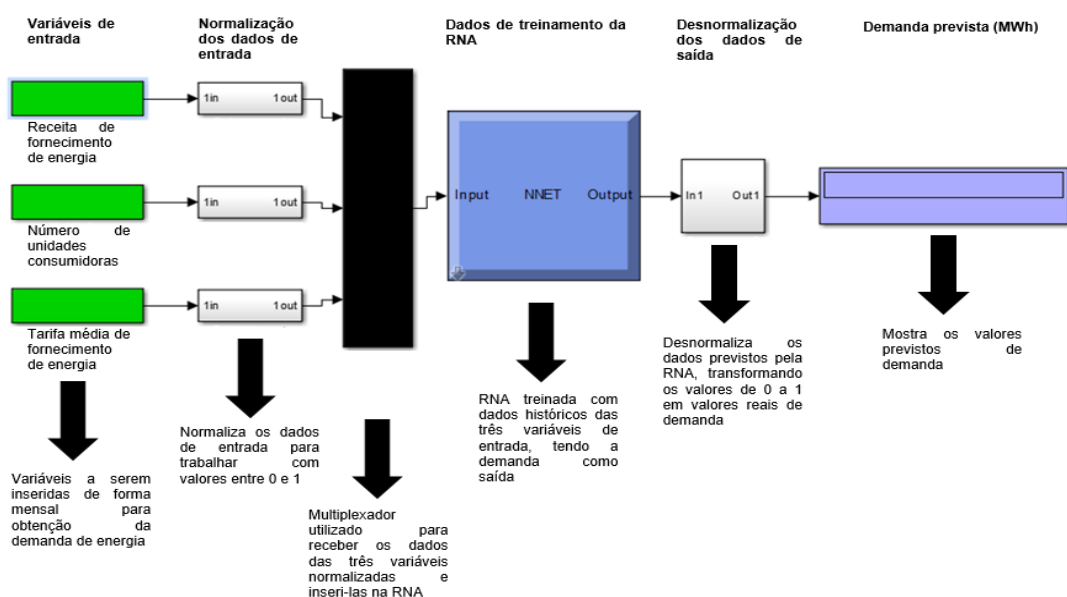


A fim de facilitar a variação das entradas para obter a saída, foi utilizado o *Simulink*, que é um ambiente utilizado para construção e simulação de sistemas dinâmicos. Pelo comando `gensim (net,-1)`, foi criada uma RNA customizada para cada classe de consumo, de tal forma que com cada RNA treinada para cada classe de consumo foi construído o modelo no *Simulink* mostrado na Figura 29.

É possível observar que as variáveis de entrada foram inseridas de forma mensal no modelo e antes de enviar os dados para a RNA foi necessário normalizar os dados de entrada pela rotina  $[pn, \text{minp}, \text{maxp}] = \text{premmx}(p)$ , em que a função `premmx` recebe o conjunto utilizado para o treinamento ( $p$ ) como parâmetro e retorna o conjunto treinamento normalizado ( $pn$ ), valor mínimo, representado pela rotina `minp` e valor máximo representado por `maxp`.

Foi necessário desnormalizar os dados e inserir um visor para demonstrar os valores previstos para as três classes de consumo.

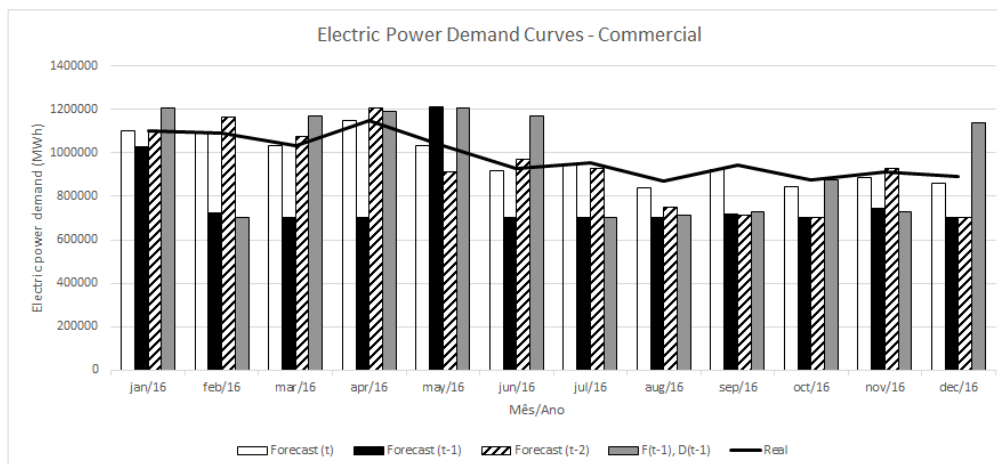
**Figura 29 – Modelo de previsão da demanda de energia utilizando *Simulink***



#### 4. RESULTADOS

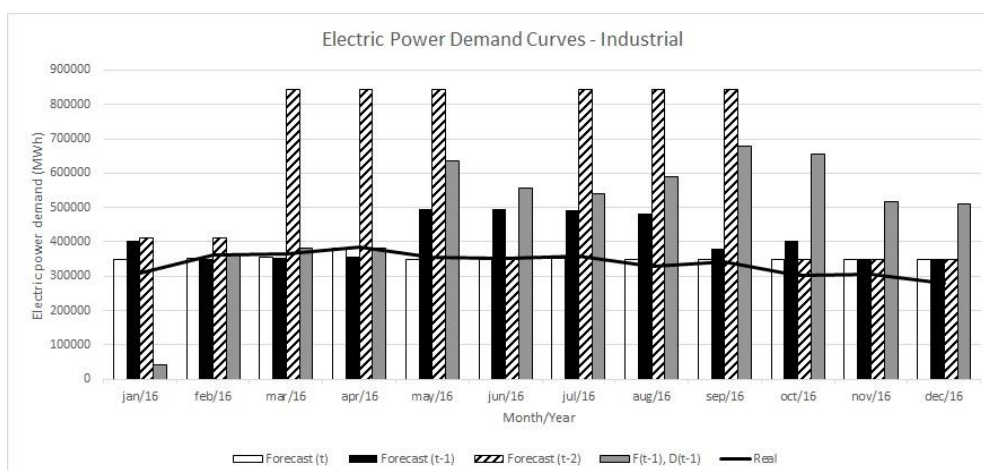
É possível observar que a RNA conseguiu aprender o comportamento dos dados de entrada no instante  $t$  da classe de consumo comercial, obtendo índice de desempenho MAPE máximo de 3,62%. Com relação ao aprendizado no instante  $t-1$ ,  $t-2$  com atraso na demanda, nota-se que, em alguns meses, a previsão teve índice de desempenho MAPE máximo de 10%, como é o caso do mês de janeiro no instante  $t-1$ , janeiro, fevereiro, março, abril, junho, julho, novembro no instante  $t-2$  e janeiro, abril, outubro no instante  $t-1$  com atraso na demanda como pode ser verificado na Figura 30.

**Figura 30 – Curvas de demanda para as classes de consumo: Comercial**



As previsões da classe de consumo industrial no instante  $t$  obtiveram bons resultados, apenas os meses de janeiro, outubro, novembro e dezembro obtiveram índice de desempenho MAPE acima de 10%, chegando no máximo a 25,16%. No instante  $t-1$ , os meses de fevereiro, março e abril obtiveram índice MAPE de 3,36%, 3,49% e 7,61%. Já no instante  $t-2$  a rede não conseguiu aprender o comportamento dos dados, apenas para o mês de junho, na qual foi obtido índice MAPE de 0,10%. Por último, a previsão no instante  $t-1$  com atraso na demanda, obteve índice MAPE de 0,26%, 4,07% e 1,12% para os meses de fevereiro, março e abril como pode ser verificado na Figura 31.

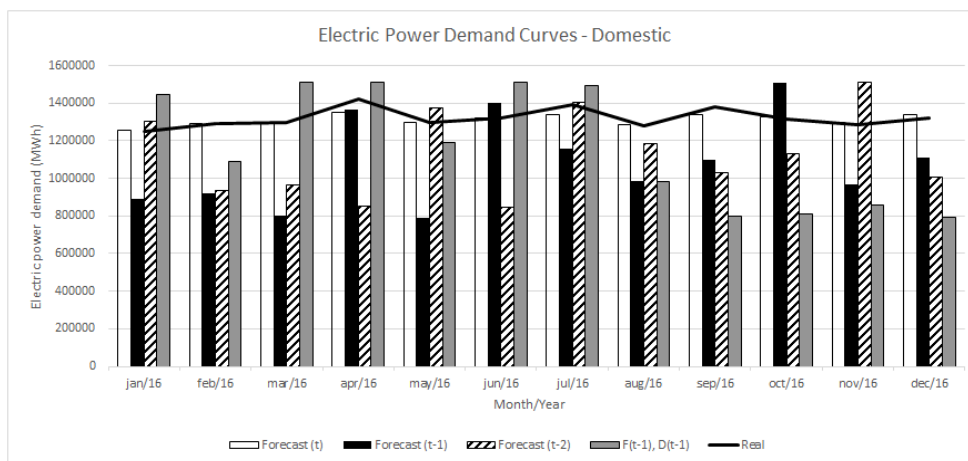
**Figura 31 – Curvas de demanda para as classes de consumo: industrial**



A previsão da classe de consumo residencial utilizando as variáveis de entrada no instante  $t$ , obtiveram índice de desempenho MAPE máximo de 4,97%. Enquanto as previsões no instante  $t-1$  obtiveram para os meses de abril e junho, 4,06% e 5,94%, respectivamente, ao

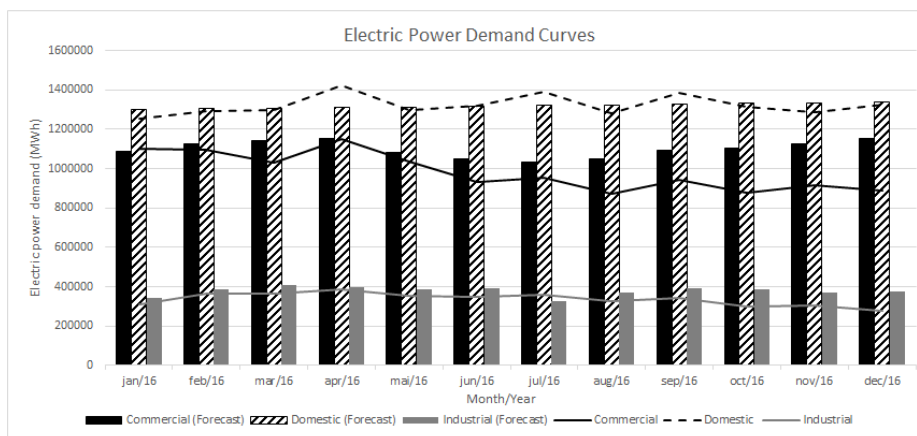
passo que as previsões que obtiveram as melhores previsões foram os meses de janeiro, maio, julho e agosto, pois obtiveram índice de desempenho MAPE abaixo de 8%. Já as previsões no instante  $t-1$ , obtiveram apenas duas previsões abaixo de 7,5% como pode ser observado na Figura 32.

**Figura 32 – Curvas de demanda para as classes de consumo: residencial**



Uma comparação entre a previsão e consumo real das três classes de consumo pode ser observada na Figura 33, na qual se nota que as técnicas MA(1) e MA(2) obtiveram melhores previsões pelo critério de desempenho Akaike.

**Figura 33 – Curvas de demanda para as classes de consumo: comercial, industrial e residencial**



Basicamente, os critério de seleção de modelos AIC e BIC escolheram as técnicas MA(1) e MA(2) como melhores para previsão da demanda das três classes de consumo. As classes de consumo comercial e industrial tiveram desempenho AIC e BIC iguais para a técnica MA(1), já a classe de consumo residencial teve melhor desempenho pelo critério BIC e foi escolhida a técnica MA(2), como pode ser observado na Tabela 17.

Tabela 17 – Melhor ajuste através do critério de desempenho AIC

Técnicas/Classes de consumo	Residencial		Comercial		Industrial	
	AIC	BIC	AIC	BIC	AIC	BIC
AR1	497,54	488,57	3699,63	3708,60	3673,74	3682,71
AR2	518,88	506,97	3691,69	3703,59	3664,75	3676,65
ARCH	464,69	455,72	13092,05	13101,02	11915,00	11923,97
ARMA	510,19	498,28	3695,40	3707,31	3665,47	3677,38
MA1	518,16	518,16	3682,28	3682,28	3655,93	3655,93
MA2	525,56	513,65	3688,85	3700,75	3662,87	3674,78

A Tabela 18 mostra o índice de desempenho MAPE das previsões com as duas técnicas, RNA e série temporal, é possível observar o poder de generalização e previsão da RNA com as variáveis no instante t obteve índices de desempenho MAPE melhores que as técnicas MA(1) e MA(2). Porém, para alguns meses o índice de desempenho MAPE está abaixo de 4,5% para a classe de consumo comercial, abaixo de 7,9% para todos os meses da classe de consumo residencial e abaixo de 8,5% para quatro meses da classe de consumo industrial.

Tabela 18 – Comparação da eficiência na previsão pelo índice de desempenho MAPE

Mês/Ano	Comercial		Residencial		Industrial	
	RNA	MA(1)	RNA	MA(2)	RNA	MA(1)
Jan/16	0,016506	1,52	0,25	3,80	12,98	11,30
Feb/16	0,008606	2,80	0,10	0,94	2,68	6,66
Mar/16	0,056031	10,78	0,09	0,68	3,10	10,76
Apr/16	0,159306	0,38	4,97	7,84	0,99	3,61
Mai/16	0,069837	4,30	0,11	1,14	1,50	8,38
Jun/16	1,012171	12,80	0,04	0,31	0,10	11,06
Jul/16	0,741018	7,84	3,74	5,18	2,11	8,09
Aug/16	3,622724	20,71	0,47	3,28	6,72	12,85
Sep/16	2,318754	15,97	3,26	4,09	2,68	14,12
Oct/16	3,471141	25,68	1,21	1,23	16,28	27,33
Nov/16	3,118378	23,12	0,63	3,58	15,12	21,36
Dec/16	3,273225	29,39	1,17	1,00	25,16	33,36

#### 4. CONCLUSÕES

Este trabalho apresentou uma abordagem computacional para previsão de demanda de energia elétrica em médio prazo por classe de consumo; para isso, foi utilizado RNA e técnicas de séries temporais, demonstrando a eficiência da previsão por meio do índice de desempenho MAPE.

Com critérios para seleção de modelos AIC e BIC, foi possível observar que não houve muita diferença entre eles, provavelmente, porque a quantidade de amostras não foi alterada. Isso interfere no desempenho do critério de seleção de modelos, já que, para quantidade grande

de amostras, o modelo BIC é mais eficiente que o AIC como pode ser observado nos trabalhos de Emiliano et al. (2010) e Ding et al. (2017).

Com relação ao uso da RNA, os resultados apresentados mostraram que utilizando receita de fornecimento de energia, número de unidades consumidoras e tarifa média de fornecimento em uma rede MLP com algoritmo *backpropagation*, é possível prever demanda de energia elétrica por classe de consumo. A abordagem computacional utilizada com a rede MLP se mostrou eficiente com as entradas no instante  $t$ , porém, como apontado por Mehdi e Bijari (2010) em alguns casos, variáveis de entrada no instante  $t-1$  e  $t-2$  com atraso de tempo na demanda, não foi possível obter resultados tão bons quanto as variáveis no instante  $t$ . Isso acontece porque a RNA associa os dados das três variáveis ao resultado, que é a demanda de energia. Quando há grandes alterações na demanda ou nas variáveis de entrada, a RNA perde o poder de generalização e, por isso, não consegue aprender o comportamento daqueles dados.

Apesar de os resultados mostrarem previsões com RNA superiores às técnicas de séries temporais, o ideal é utilizar modelos híbridos, pois apenas um modelo único pode não conseguir gerar resultados satisfatórios, uma vez que modelos híbridos procuram evitar incertezas, gerando resultados com menor variação ou erro de generalização, como apontado por Khashei e Bijari (2010).

Ainda sobre os modelos híbridos, Deb et al. (2017) descrevem que esses modelos podem aproveitar o melhor que cada técnica oferece, como a técnica ARIMA que determina a periodicidade e linearidade dos dados e o algoritmo evolutivos que determinam os resíduos.

A abordagem empregada neste trabalho se mostrou eficiente e pode ser utilizada pelas distribuidoras no Brasil para melhorar a eficiência na contratação de energia e, conseqüentemente, diminuir prejuízos com subcontratação e sobrecontratação de energia, já que a diferença entre contratação e consumo tende a diminuir. Assim sendo, a consolidação do resultado total mensal no mercado de curto prazo (TM MCP), realizado pela CCEE mensalmente para aplicar as multas, também tende a diminuir.

Esta metodologia pode ser empregada para qualquer distribuidora de energia elétrica no Brasil, desde que a RNA seja treinada com os dados históricos da receita, número de unidades consumidoras e tarifa média de fornecimento de energia da distribuidora em questão e pode auxiliar as distribuidoras no mercado de curto prazo, visto que, caso as distribuidoras usem esta metodologia para previsão de todas as classes de consumo em questão, poderão saber o consumo mensal e terão condições de avaliar se precisam comprar ou vender energia elétrica.

## REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control.**, Boston, v.19, n.6, p.716-723, Dec. 1974.

BANTUGON, M. J. T.; GALLANO, R. J. C., Short- and long-term electricity load forecasting using classical and neural network-based approach: A case study for the Philippines, 2016 **IEEE Region 10 Conference (TENCON)**, Singapore, 2016, p.3822-3825.

BOX, G.E.P., JENKINS, G.M., **Time Series Analysis: Forecasting and Control**, New Jersey: Holden Day, 2015. 756 p.

BRASIL. ANEEL. Resolução Normativa n.º 703 de 28 de março de 2016. **Define a estrutura dos Procedimentos de Regulação Tarifária – PRORET, que consolida os componentes financeiros das tarifas de distribuição – Sobrecontratação de energia e exposição ao mercado de curto prazo – ANEEL.** Disponível em: <[http://www2.aneel.gov.br/cedoc/aren2016703\\_Proret\\_Submod\\_4\\_3\\_V0.pdf](http://www2.aneel.gov.br/cedoc/aren2016703_Proret_Submod_4_3_V0.pdf)>. Acesso em: 02 fev. 2017.

DEB, C., ZHANG, F., YANG, J., LEE, S., SHAH, K.: A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, 74 (2017), 902-924.

CAMPOS, R. Júnior. **Previsão de séries temporais com aplicações a séries de consumo de energia elétrica.** 2008. 110 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Minas Gerais, Minas Gerais, 2008.

EMILIANO, P. C.; VIVANCO, M. J. F.; MENEZES, F. S.; VEIGA, E. P. **Critérios de informação de akaike versus bayesiano: análise comparativa.** In: Sinape, 2010, São Pedro. Anais do Sinape, 2010.

DING, J., TAROKH, V., YANG, Yuhong. "**Bridging AIC and BIC: A New Criterion for Autoregression**", to appear in *IEEE Transactions on Information Theory*, 2017, available at <https://arxiv.org/abs/1508.02473>. (Early version was presented at Tsinghua University as invited talk).

HAYKIN, S. **Neural Networks: A Comprehensive Foundation.** Toronto: Pearson, 2008.

KHASHEI, M., BIJARI, Mehdi. “An artificial neural network (p, d, q) model for timeseries forecasting.” **Expert Syst. Appl.** 37 (2010): 479-489.

MONTGOMERY, D. C., JENNINGS, C. L., KULAHCI, M. **Introduction to Time Series Analysis and Forecasting**, New York. 2° Edition, John Wiley & Sons Inc., 2015.

MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2.ed. São Paulo: Edgar Blücher, 2006.

PALISADE CORPORATION – **Guide to Using @Risk: Risk Analysis and Simulation Add-In for Microsoft® Excel** – version 4.5, February, 2004 – site: [www.palisade.com](http://www.palisade.com)

PATI, J., SHUKLA, K. A comparison of ARIMA, Neural Network and a Hybrid Technique for Debian Bug Number Prediction, in '**Computer and Communication Technology (ICCCT)**, 2014 International Conference on', pp. 47-53, (2014).

PINDORIYA, N. M.; SINGH, S. N.; SINGH, S. K. An adaptive *wavelet* neural network-based energy price forecasting in electricity markets. **IEEE Transactions on Power Systems**, v. 23, n. 3, p. 1423–1432, 2008. Disponível em: <<http://ieeexplore.ieee.org.ez345.periodicos.capes.gov.br/stamp/stamp.jsp?arnumber=4494757>>. Acesso em 15 fev. 2017.

RAMOS, D. S.; SUSTERAS, G. L. Utilização de Algoritmos Genéticos para Previsão da Contratação de Energia pelas Distribuidoras. **Revista IEEE América Latina**, v. 1, p. 35, 2006.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CCEE(a), 2012 version. **Energy Balance**. Available at: <<http://www.ccee.org.br>>. Access in: 2017 January 20th.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CCEE(b), 2016 version. **Results Consolidation**. Available at: <<http://www.ccee.org.br>>. Access in 2017 January 20th.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CCEE(c), 2016 version. **Contracts**. Available at: <<http://www.ccee.org.br>>. Access in: 2017 January 20th.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CCEE(d), 2016 version. **Price of the difference liquidation**. Availabe at: <<http://www.ccee.org.br>>. Access in: 2017 January 20th.

SCHWARZ, G. **Estimating the dimensional of a model.** *Annals of Statistics*, Hayward, v.6, n.2, p.461-464, Mar. 1978.

SWAROOP, R.; HUSSEIN, A. ABDULQADER. Load forecasting for power system planning and operation using artificial neural network at al batinah region oman. **Journal of Engineering Science and Technology**. Malaysia. V. 7, N.4, p.498-504, august, 2012.

WANG, X., MENG, M. A hybrid neural network and ARIMA model for energy consumption forecasting. **J Comput** 2012; 7:1184–90.

YALCINOZ, T., EMINOGLU, U. Short term and medium-term power distribution load forecasting by neural networks. **Energy Conversion and Management**, Turkey, v.46, p.1393-1405, junho, 2005.

ZHUANG, Linlin., LIU, Hai., ZHU, Jimin., WANG, Shulin., SONG, Yong. Comparison of forecasting methods for power system short-term load forecasting based on neural networks, **IEEE International Conference on Information and Automation (ICIA)**, China, p. 114-119, august, 2016.

**APÊNDICE B (Artigo 2) – Uma abordagem computacional para previsão e minimização dos custos de eletricidade no mercado de curto prazo dos distribuidores de eletricidade no Brasil.**

*Engineering Review – Open Access.*

Qualis: A3

*Manuscript ID: A computational approach to forecasting and minimizing electricity costs in the short-term market for electricity distributors in Brazil.*

*Corresponding author:* Dr. Fabio Henrique Pereira.

*Corresponding Author's Institution:* Universidade Nove de Julho.

*First Author:* Elânio Bezerra; Fabio Henrique Pereira.

*Order of Authors:* Francisco Elânio Bezerra<sup>1</sup>, Luís Carlos Santos Júnior<sup>2</sup>, Cleber Gustavo Dias<sup>2</sup>, Fabio Henrique Pereira<sup>1,2\*</sup>.

**Resumo:** No Brasil, as distribuidoras de eletricidade devem comprar energia no mercado por meio de leilões para os próximos anos; quando houver ineficiência na contratação, a câmara de Comercialização de Energia Elétrica (CCEE) poderá aplicar penalidades. Este artigo propõe uma abordagem computacional para a previsão de eletricidade para nove classes de consumo da empresa de eletricidade no Sudeste do Brasil. Assim, foram treinadas 81 redes *perceptron* de multicamadas (MLP) com o algoritmo *Levenberg-Marquardt* usando dados de receita de fornecimento de energia (SI), número consumidores (NC) e tarifa média de fornecimento de energia (ST). Com isso, 54 redes neurais foram treinadas com variáveis no instante  $t$  e 27 treinadas com variáveis atrasadas. Além disso, técnicas de séries temporais selecionadas pelo critério de informação de Akaike (AIC) e critério de informação bayesiano (BIC) foram usadas para selecionar a melhor técnica de série temporal para prever o consumo de eletricidade. Os resultados mostraram que é necessário encontrar o número de neurônios e o atraso de tempo da variável de entrada ideal para cada classe de consumo, pois cada classe possui uma característica diferente de consumo de eletricidade. As técnicas de séries temporais e a rede MLP com variáveis de entrada no instante  $t$  mostraram que essa técnica seria subcontratada em -5,26, -793,67 e -745,39 GWh no ano, gerando penalidade para essa empresa distribuidora de eletricidade. Por outro lado, as variáveis atrasadas apontariam para a sobrecontratação de +244,70 GWh, o que é permitido pelo mercado brasileiro de eletricidade. Essa abordagem mostrou-se mais eficaz, com o MAPE menor que 1% ao ano, reduzindo perdas com multas.

**Palavras-chave:** empresa distribuidora de eletricidade, comercialização de energia elétrica, previsão de eletricidade, rede neural multicamada.

### **A computational approach to forecasting and minimizing electricity costs in the short-term market for electricity distributors in Brazil**

***Abstract:** In Brazil, electricity distributors must buy electric energy at the electricity market in auctions to years ahead, when there is inefficiency in the contracting of electric energy, the chamber of Commercialization of Electric Energy (CCEE), can apply penalties. This paper proposes a computational approach to forecasting electricity to 9 classes of the consumption of the electricity company in southeast of Brazil. So, we have trained 81 multi-layer perceptron (MLP) network with Levenberg-Marquardt algorithm using energy supply income data (SI), number of consumptions (NC) and average energy supply tariff (ST), 54 were trained with variables at instant  $t$  and 27 trained with delayed variables. Also, time series techniques selected by the Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to select the best time-series technique to make prediction of electricity consumption. The results showed that it's necessary to find the number of neurons and time delay of the ideal input variable for each consumption class, because each class has a different electricity consumption characteristic. The time series techniques and the MLP network with input variables at time  $t$  showed this technique would be subcontracted at -5.26, -793.67 and -745.39 GWh in the year, generating a penalty for this electricity distributor company. On the other hand, the lagging variables would point to an overcontracting of +244.70 GWh, which is allowed by the Brazilian electricity market. This approach turned out to be more effective, with MAPE lower than 1% per year, reducing losses with penalties.*

**Keywords:** electricity distributor company, commercialization of electric energy, forecasting electricity, multilayer neural network.

## **1. INTRODUÇÃO**

A criação do novo modelo comercial de energia elétrica no Brasil, pelo Decreto nº 5.163, de 2004, obriga a compra de energia elétrica por cada empresa de distribuição (EDC) por leilões realizados no Ambiente de Contratação Regulada (RCE) para atender clientes pertencentes a

diversas classes de consumo, como comercial, consumo próprio, iluminação pública, indústria, autoridades públicas, serviço doméstico, rural e serviço público, entre outras (ANEEL, 2016).

O novo modelo também obriga cada EDC a se comprometer a fornecer eletricidade a seus consumos e distribuir eletricidade a 100% da carga de eletricidade; se houver mais consumo do que a eletricidade contratada, a EDC será penalizado. Por outro lado, não é permitido contratar eletricidade acima de 5% da demanda anual de carga, porque, acima desse valor, a EDC também sofrerá com as penalidades aplicadas pela agência nacional de energia elétrica (ANEEL, 2016).

Cada país adota um sistema diferente de comercialização de energia elétrica. Na China, por exemplo, todo o setor elétrico é controlado por empresas estatais: não há concorrência nos serviços de geração, transmissão e distribuição de energia elétrica. O marketing é feito pelo mercado atacadista. Apenas a EDC pode comprar eletricidade no mercado de curto prazo e pode realizar contratos mensais ou anuais envolvendo a geração (SER, 2016).

O mercado de eletricidade na Alemanha funciona como mercado atacadista que trabalha em dois ambientes: curto prazo, cuja eletricidade no curto prazo é comercializada para cobrir o risco de variações de preços ou por negociação direta entre os agentes do mercado (SER, 2016).

O estado da Califórnia, nos Estados Unidos, trabalha com mercado atacadista de energia, em que o Operador de Sistema Independente da EDC Califórnia (CAISO) administra sozinho cerca de 80% da rede elétrica, sendo os outros 20% de responsabilidade de empresas públicas. Grandes quantidades de eletricidade são comercializadas no mercado para o dia seguinte ou para a próxima hora, cujos preços para compra e venda de eletricidade para o dia seguinte são determinados de hora em hora (SER, 2016).

Nesse cenário, a previsão de eletricidade fornece condições para a implementação eficiente de políticas de energia, aumentando as condições de demanda e resolvendo vários problemas de tomada de decisão, como: planejamento de preços, tarifas, rede de distribuição operações, planejamento de produção de energia, gerenciamento de tarifas, atendimento ao cliente e cobranças (BANTUGON e GALLANO, 2016; KANKAL e UZLU, 2017).

Portanto, a previsão de eletricidade tornou-se vital para quem trabalha no mercado de eletricidade, principalmente para a EDC no Brasil, que precisa reportar anualmente o consumo de eletricidade à Agência Nacional de Energia Elétrica (ANEEL).

A previsão de preços nos mercados de curto e longo prazo é necessária porque, pela previsão, os envolvidos na geração, distribuição e comercialização de energia podem definir suas estratégias de licitação, aumentar as receitas e reduzir possíveis perdas de mercado no curto prazo (CATALAO et al., 2007; SARADA e BAPIRAJU, 2014).

Como cada país possui diferentes tipos de usina de geração de energia elétrica, diferentes fatores afetam a demanda por eletricidade. No entanto, é necessário adotar um método especial de previsão, pois nenhum método pode ser generalizado para todos os padrões de demanda (ALMESHAI EI e SOLTAN, 2011), porque a previsão de carga de energia é problema não linear e apresenta alto nível de volatilidade e, por esse motivo, é necessário o uso inteligente de métodos e ferramentas apropriadas de previsão (HONG, 2010; LAHOUAR, 2015).

Embora vários estudos apontem para a importância do preço médio da energia, demografia, clima, atividades sociais, situação econômica do país, fatores sazonais e energéticos para influenciar o consumo de eletricidade (KANKAL e UZLU, 2017; MOHAMED e BODGER, 2005; ZOR et al., 2017), é complicado prever a eletricidade com precisão devido a esses fatores (ZOR et al., 2017). Portanto, previsões de curto prazo usam variáveis de calendário, dados climáticos e comportamento do cliente, enquanto previsões de longo prazo consideram variações populacionais, desenvolvimento econômico, construção industrial e desenvolvimento de tecnologia (FALLAH et al., 2019).

Diante desse problema, algumas técnicas usaram Neural Artificial (BANTUGON e GALLANO, 2016; ALMESHAI EI e SOLTAN, 2011; CATALAO et al., 2007; SARADA e BAPIRAJU, 2014; SWAROOP e HUSSEIN, 2012; GUNAY, 2016; HONG, 2010; NENGBAO et al., 2014), técnicas de séries temporais (BANTUGON e GALLANO, 2016; GHADIMI et al., 2018; Nengbao et al., 2014), Meta-heurística (GHADIMI et al., 2018), Induzida Difusa Raciocínio (JURADO et al., 2015), algoritmo genético (CATALAO et al., 2007), *Autoencoders Stacked Denoising* (Wang et al., 2016), *Deep Neural Networks* Marino et al. (2016) e outros têm sido usados nos últimos anos para prever eletricidade, especialmente redes neurais.

O objetivo deste artigo é prever eletricidade por classe de consumo usando SI, NC e ST como variáveis de entrada nos instantes de tempo  $t$ ,  $t-5$ ,  $t-10$  e  $t-20$  em uma rede MLP com algoritmo *backpropagation*.

A eficiência da previsão é testada por meio do erro percentual médio absoluto (MAPE) para todas as classes de consumo de uma distribuidora de energia elétrica no Sudeste do Brasil e comparada aos resultados de outros estudos com precisão significativa.

O restante do artigo está organizado da seguinte forma: subseção 1.1, trabalhos relacionados; subseção 1.2, novidade e contribuição real deste artigo; seção 2, regras de comercialização de energia elétrica no Brasil; seção 3, técnicas empregadas na previsão; seção 4, metodologia, seção 5, os resultados; as considerações finais e discussão estão contidos na seção 6.

## 1.1 TRABALHO RELACIONADO

Em Swaroop e Hussein (2012), a previsão de demanda de energia elétrica foi proposta para determinar o planejamento e operação do sistema de energia da região da região de Al Batinah, em Omã. Eles usaram redes neurais artificiais (RNA), usando dados mensais históricos de 2007, 2008 e 2009 para consumo, temperatura e umidade, para prever o consumo para 2010.

Em Almeshaii e Soltan (2011) foi demonstrado um método para integrar diferentes modelos de previsão para analisar o padrão de consumo elétrico e prever a demanda de carga. Foram utilizados dados históricos de três anos e o consumo do Kuwait foi previsto usando uma média móvel por sete dias, que teve erro médio de 30,55 MW com ASM e previsões de 0,0384 por 30 dias com erro médio de 174,47 MW.

Em Kouhi et al. (2014), os autores desenvolveram a previsão de carga para a empresa de energia PJM e o mercado britânico de energia, usando dados históricos de carga com atraso de uma hora e um dia em RNA com a seleção de características caóticas inteligentes, baseadas na reconstrução teorema do espaço de fase e, portanto, previsão de carga para os seguintes dias do ano de 2006: 11 a 17 de fevereiro, 8 a 14 de junho, 9 a 15 de agosto e 6 a 12 de novembro.

O trabalho desenvolvido em LAZO et al. (2012) foi baseado na otimização da aquisição de energia no Mercado de Curto Prazo (STM) nos leilões de um ano à frente (A-1) e no ajuste, utilizando algoritmo genético, buscando minimizar o custo relacionados à contratação de energia elétrica, às multas por subcontratação e ao preço por diferenças (PDS). A contratação ideal foi calculada usando simulação de Monte Carlo para gerar vários cenários de consumo pelo período de cinco anos.

Enquanto em Bantugon e Gallano (2016) comparamos o desenvolvimento do método Holt-Winters e da RNA na previsão de curto e longo prazo da demanda de energia elétrica para o Estado das Filipinas, nos Estados Unidos, onde os resultados foram comparados pelos indicadores de desenvolvimento MAPE, erro médio absoluto (AME) e erro médio quadrático (RMSE).

Em Wang et al. (2016) as técnicas ARIMA e ANN são usadas para prever o consumo de eletricidade da província de Hebei, na China, onde foram utilizados dados históricos de 1980 a 2008, nos quais a eficiência na previsão foi comparada AME, RMSE e MAPE.

Um estudo sobre o escopo de algumas técnicas de previsão de séries temporais foi realizado em Deb et al. (2017), como ARIMA, ANN, *fuzzy*, neural, *wavelet*, algoritmos evolutivos e algoritmos híbridos que funcionam com duas ou mais técnicas para prever o

consumo de energia em edifícios. Os resultados mostraram que o uso de abordagens híbridas é mais eficiente porque, se uma técnica não fornece tudo de maneira eficiente, outra técnica pode suprir essa deficiência.

Já em Gunay (2016), foi desenvolvido um método de previsão de eletricidade para a Turquia que utilizou dados históricos 1975-2013: população, Produto Interno Bruto (PIB) per capita, porcentagem de inflação, porcentagem de desemprego, temperatura média no verão e a temperatura média no inverno usando regressão linear múltipla e rede multicamada (MLP) e comparou os resultados de 2007 a 2013 a outros resultados da literatura.

Uma maneira de prever por minuto e hora a carga de uma residência foi desenvolvida em (MARINO et al., 2016). Para isso, eles usaram redes neurais profundas com duas arquiteturas do algoritmo LSTM (*long short-term memory*), padrão LSTM e uma arquitetura direta de sequência a sequência (S2S). Há um conjunto de amostras de dezembro de 2006 a novembro de 2010 por minuto e hora, em que os três primeiros anos foram usados para treinar o modelo e o último ano usado para testar. Os resultados mostraram boas previsões para as primeiras 60 horas com rede LSTM padrão, no entanto, para previsão por minuto.

Um modelo de RNA também foi utilizado em Khashei e Bijari (2010) para reconhecimento de padrões, classificação, agrupamento e previsão de séries temporais, produzindo resultados com alto grau de precisão em diversas áreas.

Como se pode ver, existem muitos estudos sobre previsão ou preço da eletricidade e eles usaram técnicas diferentes para resolver o problema.

Como dito anteriormente, cada país possui diferentes tipos de usina elétrica, diferentes fatores afetam a demanda por eletricidade (ALMESHAI EI e SOLTAN, 2011). Como o Brasil tem uma particularidade, tem enfrentado uma crise econômica e política nos últimos anos, isso afetou ou afetará o consumo de eletricidade. É necessário adotar um método especial de previsão, pois nenhum método pode ser generalizado para todos os padrões de demanda (ALMESHAI EI e SOLTAN, 2011).

O objetivo e a novidade deste artigo é usar a segunda condição mencionada por Hong e Fan (2016) e usar variáveis históricas SI, NC, ST nos instantes de tempo  $t$ ,  $t-5$ ,  $t-10$  e  $t-20$  em uma rede MLP para prever a eletricidade mensal por classe de consumo e auxiliar a EDC no Brasil diante dos problemas mencionados anteriormente.

## 2. REGRAS DE COMERCIALIZAÇÃO DE ENERGIA ELÉTRICA NO BRASIL

Para adquirir energia elétrica no ambiente de contratação regulada (ACR), a EDC é obrigada a participar de leilões nos próximos três a cinco anos (A-3, A-5), chamados de novos leilões de energia, em que os contratos dessa categoria podem ser com prazo mínimo de quinze e prazo máximo de trinta e cinco anos. Existe a possibilidade de poder de compra nos leilões de energia existentes por um ano, chamado A-1, e nos leilões de ajuste com entrega imediata, onde os contratos podem ser de no mínimo três e no máximo quinze anos (CEPB(C), 2016).

Os procedimentos de regulamentação tarifária (PRORET) são utilizados para regular a tarifa processada, foram aprovados pela Resolução Normativa nº 435/2011, estão organizados em 12 módulos, subdivididos em submódulos, onde o submódulo 4.3 mostra a supercontratação de energia e exposição a mercado de curto prazo, detalhando os mecanismos envolvidos na verificação da exposição do EDC (ANEEL, 2016).

A contabilização da energia é feita pelo Conselho de Comercialização de Energia Elétrica (CEPB), como forma de extinguir as diferenças entre a contratação e o consumo de energia elétrica. Depende do balanço energético da distribuidora (EBD), que é composto pela diferença entre o consumo total (TRC) e os contratos assinados pelas distribuidoras (PCL), todos em Mega-Watt-Hora (MWh), conforme observado em Equação 30 (CEPB(A), 2016).

$$EBD = -TRC + PCL \quad (30)$$

Nas regras de comercialização de energia são descritas a forma de contabilização do volume de energia negociado no mercado de curto prazo, que depende do produto do balanço energético EBD em MWh com o preço do PDS em Real, calculado semanalmente com as cobranças padrão de cada sub mercado (região do país), como podemos observar na Equação 31 (CEPB(A), 2016; CEPB(D), 2016).

$$STM = \pm EBD \times PDS \quad (31)$$

A consolidação do resultado mensal total no mercado de curto prazo, TM (STM), ou exposição financeira do EDC no STM, é mostrada na Equação 32 e descrita nas regras de comercialização: consolidação dos resultados do CEPB, mostra que o TM (STM), calculado mensalmente para cada EDC, depende do resultado da soma no STM, onde o resultado do TM (STM) pode ser positivo, quando o EDC tiver mais energia contratada do que consumida ou

negativa, quando houver mais consumo do que contratos de energia. Os indicadores “r” e “s” são padrões de carga e semanais, respectivamente (CEPB(A), 2016).

$$TM(STM) = \pm \sum_{r,s}^n(STM) \quad (32)$$

O CEPB contabiliza a quantidade de energia elétrica contratada, consumida e vendida no STM para cada EDC, e a apuração da energia comprada e vendida no STM (VSTM e CSTM) é feita mensalmente, sendo superada na contratação energia elétrica da EDC no ano. Quando o volume de energia vendida no SSTM é maior que o VSTM (SSTM > VSTM), ocorre contratação excessiva do EDC; quando o volume de energia comprado em contratos é maior que o volume vendido no STM (VSTM > SSTM), existe exposição do EDC (CEPB(B), 2016).

A energia demandada no ano pela EDC, ED (ano), é contabilizada pela soma da energia demandada mensalmente, e o limite acima da contratação no ano, LAC (ano), é calculado como 5% da energia demandada anual, como exibido na Equação 33 (CEPB(B), 2016).

$$ECL(ano) = 1,05x ED(ano) \quad (33)$$

Quando a energia contratada pela EDC não é suficiente para atender seu mercado, a EDC pode ser penalizada pela subcontratação, compra de energia no curto prazo e pelo erro na contratação de energia, em que o custo real da energia não repassado depende do volume contratado, do PDS e do valor de referência anual.

### 3. TÉCNICAS EMPREGADAS NA PREVISÃO

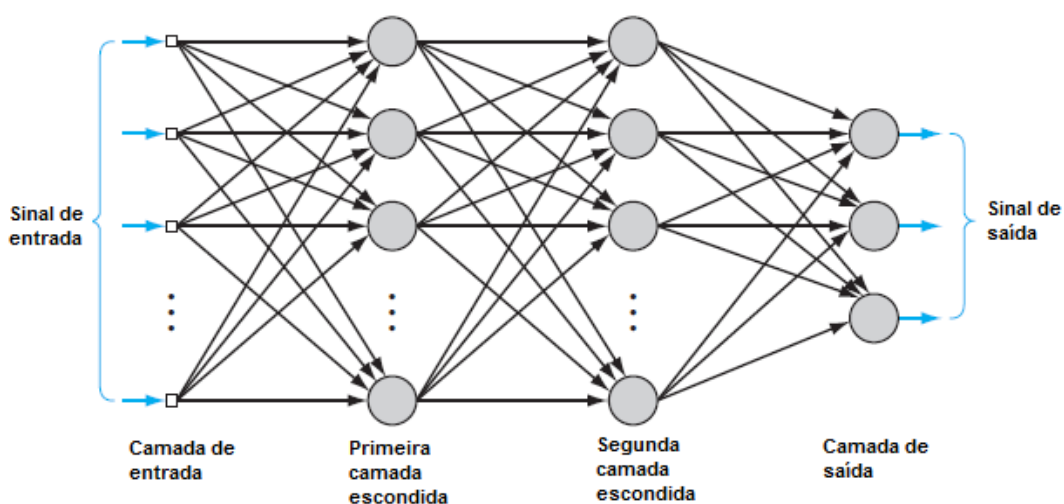
#### 3.1 REDES NEURAS ARTIFICIAIS

Nos últimos anos, as redes neurais artificiais (RNAs) foram aplicadas em diversas áreas e constituem uma ferramenta de Inteligência Artificial (IA) incorporada à modelagem da inteligência humana, que trabalha em paralelo e cujo conhecimento é adquirido por aprendizagem, onde sinapses ponderadas são usadas para o conhecimento. Sua principal característica é armazenar o conhecimento experimental a ser utilizado (HAYKIN, 1999).

Um *perceptron* multicamada (MLP) é uma rede neural artificial que pode trabalhar com várias superfícies de decisão não lineares, usa várias regras de aprendizado e tem alto grau de conectividade pelos pesos sinápticos (HAYKIN, 1999).

A Figura 34 mostra a estrutura de uma rede MLP, que consiste em uma camada de entrada, duas camadas ocultas e uma camada de saída. O sinal de entrada se propaga em direção única, camada por camada. A camada de entrada receberá dados de entrada da rede, a primeira camada oculta é alimentada a partir da camada de entrada, que por sua vez é constituída pelos nós de origem, as saídas resultantes da primeira camada oculta são aplicadas como entrada para a segunda camada oculta. A camada de saída tem, então, a função de processar as informações (GUNAY, 2016).

Figura 34 – Rede MLP com uma estrutura de duas camadas ocultas



Fonte: Haykin, 2008, p. 124

Uma função logística é amplamente usada para determinar os valores de mudança de cada peso, ela define a saída de um neurônio em termos de  $v$ : quando  $v$  for maior que ou igual a 0,  $\varphi(v)$  será igual a 1; se " $v$ " for menor que 0,  $\varphi(v)$  será 0 e " $v$ " representa a soma das entradas mais o limite de ativação. O sinal de saída é representado pela soma do produto dos sinais de entrada, pelo peso sináptico, mais o viés.

Para minimizar o erro de treinar a rede, um algoritmo *backpropagation* pode ser implementado, ele funciona em duas fases: a fase de *feed-forward*, na qual os pesos sinápticos são fixos e o sinal de entrada é propagado pela rede, camada por camada, até a saída e a fase reversa, na qual um sinal de erro é produzido, comparando a saída da rede à resposta desejada. Isso se propagou de volta pela rede, camada por camada, resultando em ajustes contínuos dos pesos sinápticos (HAYKIN, 1999).

### 3.2 TÉCNICAS DE SÉRIES TEMPORAIS E CRITÉRIO PARA SELEÇÃO DE MODELOS

A série temporal consiste em qualquer conjunto de observações organizadas no tempo. A característica essencial da série temporal é que as observações próximas umas das outras são dependentes. A análise dessa dependência requer o uso de técnicas específicas, para isso, é necessário o desenvolvimento de modelos (MONTGOMERY et al., 2015; MORETTIN e TOLOI, 2006).

Modelos estocásticos incluem modelos estacionários e não estacionários. Em modelos estacionários, supõe-se que o processo permaneça em equilíbrio estatístico com propriedades probabilísticas que não mudam ao longo do tempo, variando em torno de um nível médio constante fixo e com variação constante, enquanto em modelos não estacionários essas características dos dados mudam ao longo do tempo (MONTGOMERY et al., 2015).

A previsão refere-se à antecipação de algum evento e é usada em áreas como negócios, indústria, governo, economia, ciências ambientais, medicina, ciências sociais, política e finanças, entre outras. Os problemas de previsão podem ser classificados em curto prazo, que envolvem previsões de apenas alguns períodos de tempo (dias, semanas, meses) no futuro, previsões a médio prazo, que se estendem a alguns meses, e previsões a longo prazo, relacionados com previsões ao longo de um ano (MONTGOMERY et al., 2015).

Existem vários métodos usados para fazer previsões: alguns funcionam com tendência, sazonalidade, bem como a função de autocorrelação (ACF), que é uma maneira de medir e explicar a associação entre observações na série temporal. Existem métodos que buscam explorar a estrutura de autocorrelação dos resíduos de previsão do período atual com os ocorridos em períodos anteriores, entre outras características (MONTGOMERY et al., 2015).

As técnicas de séries temporais são escolhidas pelos critérios de seleção de modelos que buscam identificar modelos mais próximos da realidade.

O Critério de Informação de Akaike (AIC) foi criado pelo estatístico japonês Hirotugu Akaike em 1973 Akaike (1974), com base na teoria da informação. Ele funciona minimizando a distância *Kullback-Leibler* (K-L), que é a medida da distância entre o modelo verdadeiro e um modelo selecionado. É usado para a seleção de um modelo. De acordo com Morettin e ToloI (2006), quanto menor o valor da AIC, melhor o ajuste do modelo.

A Equação 34 ilustra o cálculo da AIC, que depende do estimador de máxima verossimilhança (compatível com as informações disponíveis) e  $p$ , o número de parâmetros.

$$AIC = -2\sigma_p^2 + 2p \quad (34)$$

Em 1978, foi desenvolvido Schwarz (1978) o critério de informação bayesiano (BIC), que maximiza a probabilidade de identificar o modelo verdadeiro entre os modelos avaliados e funciona de maneira semelhante à AIC: quanto menor o valor, melhor o ajuste do modelo. A Equação 35 mostra o desenvolvimento desse modelo, usando o estimador de máxima verossimilhança, no qual “p” é o número de parâmetros a serem estimados m é o número de pontos da amostra:

$$BIC = -2\log\sigma_k^2 + p\log(m) \quad (35)$$

Os melhores modelos são selecionados pelos valores mais baixos da AIC e BIC. Esses critérios adicionam uma penalidade que aumenta com o número de regressores. Como a penalidade é maior para o BIC, esse critério tende a favorecer modelos mais parcimoniosos que o AIC. A ideia é usar esses critérios para encontrar um modelo parcimonioso que descreva o processo dinâmico da variável (VERBEEK, 2009).

#### 4. METODOLOGIA

Esta seção apresenta a proposta metodológica para as distribuidoras de energia elétrica diminuïrem o impacto com multas aplicadas no mercado brasileiro de eletricidade, conforme descrito na seção 2.

A primeira etapa foi coletar os dados históricos de renda do fornecimento de energia (SI), número de consumos (NC) e tarifa média de fornecimento de energia (ST) da empresa de eletricidade do Sudeste do Brasil de janeiro de 2003 a outubro de 2019 na ANEEL como se pode ver em <http://relatorios.aneel.gov.br>.

Conforme relatado em Almeshaiel e Soltan (2011), nos países em desenvolvimento, a demanda por eletricidade cresce com taxa dinâmica e alta. Portanto, cada país é indiferente a fatores que afetam a demanda por eletricidade.

Como pôde ser observado na seção 1, existem alguns artigos que apontam para a importância de seleção de variáveis no modelo de previsão de eletricidade; no entanto, é difícil prever com precisão devido a esses fatores (KANKAL e UZLU, 2017; MOHAMED e BODGER, 2005; ZOR et al., 2017).

Por esse motivo, foram utilizadas três variáveis SI, NC e ST, para prever nove classes de consumo de uma distribuidora de energia, tais como: comercial (A), consumo próprio (B), iluminação pública (C), indústria (D), poder público (E), serviço doméstico (F), rural (G) e público (água, esgoto e saneamento (H), tração elétrica (I)) e cada classe de consumo possui 202 dados.

As variáveis climáticas não foram incorporadas ao modelo, uma vez que são incorporadas à demanda por eletricidade (CATALAO et al., 2007; VORONIN et al., 2014) e o preço médio da energia está correlacionado ao consumo de eletricidade (MOHAMED e BODGER, 2005).

Na estrutura de previsão de eletricidade das classes de consumo (Figura 35), existem duas abordagens para fazer previsões para o consumo de eletricidade da EDC. Primeiro, na abordagem 1, as previsões foram realizadas usando séries temporais para cada classe de consumo. Os dados são transformados em estacionários usando o *@Risk do software Palisade* e, em seguida, os critérios AIC e BIC mais baixos escolhem as melhores técnicas de séries temporais (AR, ARMA, ARIMA, ARCH, GARCH), que farão as previsões conforme descrito na seção 3.

A melhor previsão de cada classe de consumo é resumida nas nove classes de consumo para formar o consumo mensal de eletricidade da empresa distribuidora (ECDC).

Na abordagem 2, é utilizada uma rede MLP com algoritmo Levenberg-Marquardt e função de ativação tangente sigmoide para treinar 194 dados para cada classe de consumo no Matlab 2018a, separando 70% dos dados para treinamento, 15% para validação, 15% para testes e 12 meses para testar as previsões, que foram usadas de duas maneiras. Primeiro, os dados de janeiro de 2003 a agosto de 2008 para treinar a rede MLP com variáveis na renda média de fornecimento de energia ( $SI$ ), número de unidades consumidoras ( $NC$ ), tarifa média de fornecimento de energia ( $ST$ ) e  $d(t)$ , eletricidade consumo por classe de consumo (Equação 36).

$$d(t) = f(SI_t, NC_t, ST_t, \dots, SI_{tn}, NC_{tn}, ST_{tn}) \quad (36)$$

Este artigo utilizou a mesma arquitetura de rede MLP proposta por Osman et al. (2009), treinando de seis maneiras, uma camada de entrada (um neurônio), uma camada de saída (um neurônio) e, como não existe regra universal para determinar o número de neurônios na camada oculta, a única maneira foi verificar o desempenho de diferentes modelos usando diferentes

números de neurônios (SARADA e BAPIRAJU, 2014; SWAROOP e HUSSEIN, 2012; OSMAN et al., 2009; NENGBAO et al., 2014). Portanto, escolhemos 5, 10, 20, 30, 40 e 50 neurônios para experimentos com variáveis de entrada no instante de tempo  $t$  e 10 neurônios com variáveis de entrada atrasadas.

Foi escolhido o menor erro médio MAPE de 12 previsões e, em seguida, a melhor previsão de cada classe de consumo somada às outras classes de consumo para formar o ECDC mensal.

A segunda maneira tem as mesmas variáveis de entrada (SI, NC e ST); nesse momento, foi utilizada a abordagem de previsão de eletricidade desenvolvida por (VORONIN et al., 2014). Entretanto, em vez de usar um modelo híbrido (ARMA, GARCH e ANN) em  $t-1$  e  $t-2$  para prever o instante de tempo  $t$ , foram utilizadas as variáveis SI, UC e TMF para prever classes de consumo do EDC usando variáveis de entrada no instante de tempo  $t-5$ ,  $t-10$  e  $t-20$  para oferecer um mês de previsão (Equações 37, 38 e 39).

$$d(t) = f(SI_{t-1}, NC_{t-1}, ST_{t-1}, \dots, SI_{t-5}, NC_{t-5}, ST_{t-5}) \quad (37)$$

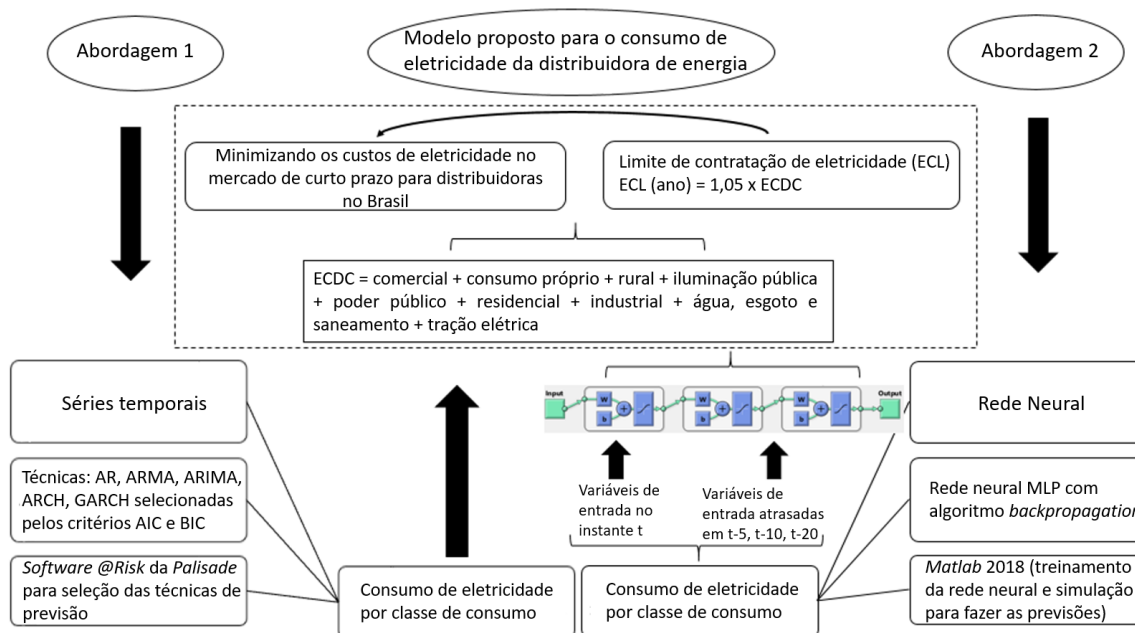
$$d(t) = f(SI_{t-1}, NC_{t-1}, ST_{t-1}, \dots, SI_{t-10}, NC_{t-10}, ST_{t-10}) \quad (38)$$

$$d(t) = f(SI_{t-1}, NC_{t-1}, ST_{t-1}, \dots, SI_{t-20}, NC_{t-20}, ST_{t-20}) \quad (39)$$

Mantendo a camada oculta com dez neurônios, as variáveis de entrada foram atrasadas em  $t-5$ ,  $t-10$  e  $t-20$ . Em seguida, foi escolhido o menor erro MAPE médio das 12 previsões. A melhor previsão de cada classe de consumo é somada às outras classes de consumo para formar o ECDC mensal.

Observe-se que, antes de enviar os dados para a RNA, foi necessário padronizar os dados de entrada para trabalhar com valores entre 0 e 1. Com esses valores normalizados, a rede treinada recebeu os dados de entrada e generalizou a partir de parâmetros que foram descritos anteriormente e finalmente, a desnormalização é necessária para retornar os valores de demanda esperados

**Figura 35 – Estrutura de previsão de eletricidade das classes de consumo**



## 5. RESULTADOS E DISCUSSÃO

A Tabela 19 mostra o melhor ajuste dos critérios AIC e BIC, a técnica de séries temporais selecionada, a previsão mensal de eletricidade por classe de consumo por 12 meses. Os critérios AIC e BIC não mudaram substancialmente na escolha da técnica de séries temporais. Basicamente, selecionou-se a técnica MA para todas as classes de consumo, exceto a classe de consumo B.

O número de amostras afeta a escolha do melhor modelo técnico e o BIC é mais eficiente que o AIC quando há número considerável de amostras (mais de 500) (DING et al., 2018). No entanto, como há apenas 202 amostras em cada classe de consumo, provavelmente não houve muita diferença na seleção de técnicas de séries temporais.

Embora VORONIN et al. (2014) tenham usado as previsões dos modelos ARMA e GARCH como entrada para uma RNA, os resultados mostraram MAPE <10,30%, enquanto apenas o uso de técnicas de séries temporais selecionadas pelos critérios AIC e BIC gerou MAPEs <7,9%.

O desempenho obtido nas previsões da rede MLP para cada classe de consumo pode ser visto na Tabela 20. A classe comercial atingiu 0,16% de erro MAPE, consumo próprio 0,28%, rural, 1,45%, iluminação pública 0,31%, poder público 0,17%, residencial 0,24%, industrial 1%, água, esgoto e saneamento 2,22%, tração elétrica 0,48%. Além de mostrar o desempenho

do treinamento da rede MLP pelo erro MSE e do tempo de treinamento para obter a melhor previsão por classe de consumo.

Devido à estrutura de dados de cada classe de consumo, a rede MLP apresentou erros MAPE com número diferente de neurônios, por exemplo, comercial, consumo próprio, rural, tração elétrica apresentaram melhores resultados com cinco neurônios, poder público e classes de consumo industrial apresentaram erros menores com 20 neurônios, iluminação pública com 10 neurônios, doméstico 40, enquanto a classe de consumo de água, esgoto e saneamento apresentou melhores resultados com 30 neurônios.

Não foi possível obter os menores erros MAPE com o mesmo número de neurônios para cada classe de consumo, conforme mostrado na Tabela 20, pois cada classe de consumo tem série temporal diferente.

As melhores previsões obtidas com os melhores atrasos de tempo nas variáveis de entrada para cada classe de consumo podem ser observadas na Tabela 21, na qual as classes A, E, F, G geraram melhores resultados em  $t-20$ , B e I no tempo  $t-10$ , C e H no tempo  $t-5$ .

**Tabela 19 – Melhor ajuste escolhido por critério de ajuste, técnica selecionada e previsão de classe de consumo (GWh) por ano**

Critério de ajuste	Classes de consumo	A	B	C	D	E	F	G	H	I	Total por classe
AIC	Valor	4514,1	2984,2	3092,4	3971,1	3761,4	4652,9	-419,8	3841,6	3111,8	-
	Técnica selecionada	MA1	AR2	MA2	MA1	MA1	MA2	MA1	MA2	MA1	-
	Eletricidade por ano	9816,5	38,3	10,3	865,3	1194,3	16176,4	3010,5	325,7	137,9	31575,2
BIC	Valor	4514,1	2997	3098,6	3971,1	3761,4	4653,5	-419,8	3852,8	3111,8	-
	Técnica selecionada	MA1	AR2	MA1	MA1	MA1	MA1	MA1	MA1	MA1	-
	Eletricidade por ano	9816,5	38,3	9,8	865,3	1194,3	16235,3	3010,5	315,5	137,9	31623,5

Tabela 20 – Desempenho das previsões e erro e tempo de treinamento

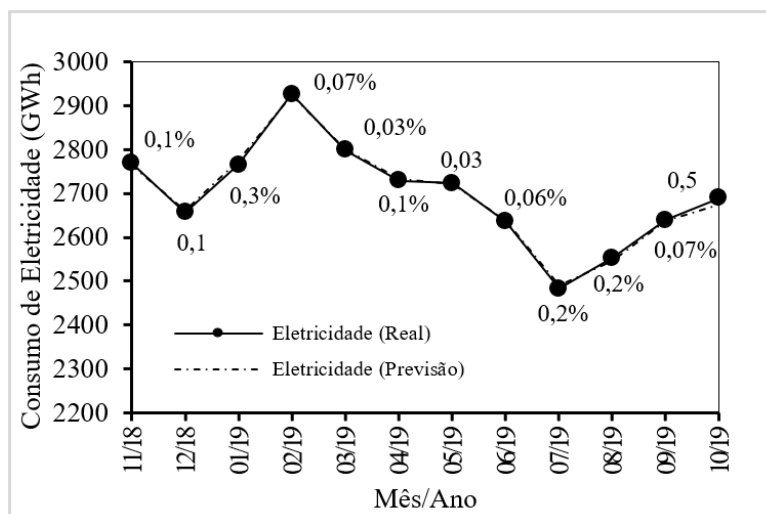
Classe / Numero de neurônios	05 05 1	10 10 1	20 20 1	30 30 1	40 40 1	50 50 1	Performance	
							MSE	Tempo
Comercial	<b>0,16</b>	1,67	5,5	1,04	3,01	4,67	1,71E-04	1s
Consumo próprio	<b>0,28</b>	1,3	0,87	1,45	2,84	1,52	5,49E-05	2s
Rural	<b>1,45</b>	6,38	3,03	22,88	141,5	183,1	1,69E-04	2s
Iluminação pública	0,52	<b>0,31</b>	0,72	0,4	0,53	0,7	1,60E-04	3s
Autoridades públicas	0,67	0,43	<b>0,17</b>	0,37	0,7	1,62	8,76E-07	11s
Doméstico	0,48	0,36	0,73	1,08	<b>0,24</b>	2,56	3,41E-04	45s
Industrial	2,33	1,4	<b>1</b>	1,31	1,75	4,63	1,64E-05	11s
Água, esgoto e saneamento	3,48	2,57	3,33	<b>2,22</b>	3,16	3,05	3,08E-06	14s
Tração elétrica	<b>0,48</b>	0,55	1	0,69	1,59	2,5	2,14E-04	2s

Tabela 21 – Melhores entradas de previsão e atraso de tempo para cada classe de consumo

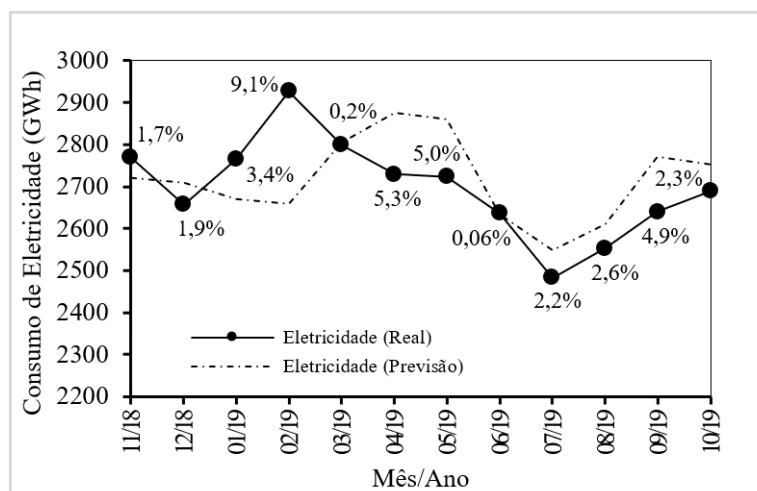
Consumo Classe	Atraso de tempo	Dia (Mês/Ano)											
		nov/18	dez/18	jan/19	fev/19	mar/19	abr/19	mai/19	jun/19	jul/19	ago/19	set/19	out/19
A	<i>t-20</i>	840	903	1009	927	923	1023	914	851	791	766	822	828
B	<i>t-10</i>	3	4	4	3	3	3	3	3	3	3	3	3
C	<i>t-5</i>	1	1	3	1	1	1	1	1	1	1	1	1
D	<i>t-5</i>	56	56	59	65	69	59	56	61	59	57	63	57
E	<i>t-20</i>	95	112	103	107	91	109	106	103	102	105	111	121
F	<i>t-20</i>	1398	1319	1196	1249	1424	1378	1475	1305	1284	1376	1463	1439
G	<i>t-20</i>	287	273	259	271	261	268	265	268	268	265	269	264
H	<i>t-5</i>	28	28	28	28	28	28	28	28	28	28	28	28
I	<i>t-10</i>	12	12	10	7	6	7	12	13	13	11	12	12
Soma	-	2721	2708	2670	2658	2805	2875	2860	2634	2549	2611	2771	2753
Real	-	2768	2656	2765	2927	2799	2729	2723	2636	2483	2553	2640	2690

A soma das previsões feitas com a rede MLP e as variáveis de entrada no instante de tempo  $t$  por classe de consumo gera o consumo mensal do EDC, portanto, é possível calcular o erro MAPE entre o consumo real e previsto de eletricidade mensal de 12 meses (Figuras 36 e 37).

**Figura 36 – Previsão de eletricidade com variáveis de entrada no tempo  $t$**



**Figura 37 – Previsão de eletricidade com variáveis de entrada no tempo  $t-5$ ,  $t-10$  e  $t-20$**



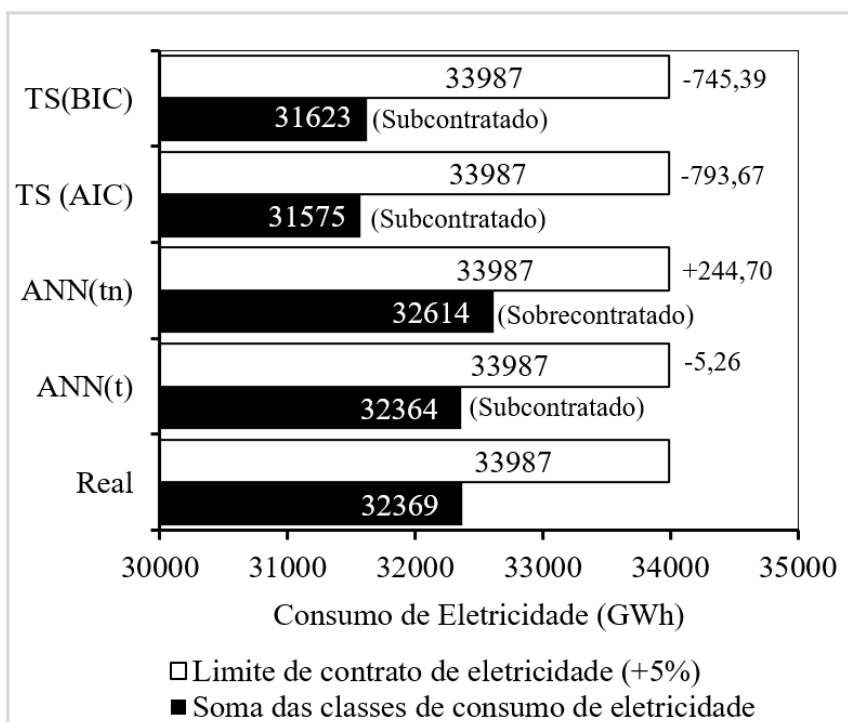
Embora as previsões mensais com variáveis de entrada no tempo  $t-5$ ,  $t-10$  e  $t-20$  mostrem erros MAPE maiores do que as variáveis no tempo  $t$  (Figuras 36 e 37) as previsões com variáveis de entrada no tempo  $t$  não oferecem condições para determinar o que acontecerá no futuro. Atrasando as variáveis de entrada, é possível fazer previsões para o próximo mês, o que é mais interessante para os distribuidores de eletricidade.

Outra questão importante é que o consumo anual é verificado pela CCEE/ANEEL conforme descrito na seção 2. Se o EDC for subcontratado ou sobrecontratado (5% acima do consumo anual), sofrerá penalidades pela má gestão da energia elétrica.

A partir do consumo mensal real do EDC, temos o consumo anual de eletricidade, aplicando a Equação 38, o EDC tem o limite de eletricidade que pode ser contratado, nesse caso, 33987 GWh.

A soma das 12 previsões mensais de eletricidade gera o consumo anual da concessionária. Os erros MAPE menores nas previsões mensais com as variáveis no tempo  $t$  (Figura 22), o EDC estaria subcontratado em -5.2605 GWh. A mesma situação ocorreria com previsões usando técnicas de séries temporais selecionadas pelos critérios AIC e BIC, -793,67 GWh e -745, GWh, respectivamente (Figura 36). Enquanto nas previsões com variáveis atrasadas (Figura 23), apesar de apresentar erros maiores que as variáveis de entrada no tempo  $t$ , apresentam-se melhores resultados no consumo anual, pois, apesar de sobrecontratado, o EDC estaria dentro do limite de 33987,29 GWh (f).

**Figura 38 – Nível de consumo de eletricidade e situação do distribuidor de eletricidade no ano**



Embora a literatura mostre vários estudos de previsão de eletricidade e preços usando o PIB do país, preço da eletricidade, carga de pico de energia, produto interno bruto, população, cargas defasadas, variáveis climáticas, padrões de consumo, variáveis dummy, variável exógena, indicadores econômicos sociais, entre outros (BANTUGON e GALLANO, 2016; MOHAMED e BODGER, 2005; GHADIMI et al., 2018; OSMAN et al., 2009; JURADO et al., 2015; NENGBAO et al., 2014; GUNAY, 2016; CATALAO et al., 2007; ALMESHAI EI e

SOLTAN, 2011), dados históricos SI, NC e ST de nove classes de consumo foram suficientes para capturar a distribuição de energia do comportamento do consumo de eletricidade, fazendo boas previsões com variáveis de entrada em tempo  $t$  e  $t-5$ ,  $t-10$  e  $t-20$  (Figuras 25 e 26, Tabela 8 e Tabela 9).

Por fim, a Tabela 22 mostra alguns trabalhos que utilizaram técnicas para previsão de carga energia elétrica, comparando com o modelo proposto neste trabalho. É possível observar que a abordagem utilizada gerou resultados competitivos, quando comparado com os resultados de literatura.

**Tabela 22 – Técnicas utilizadas e comparação de previsões**

<b>Autores</b>	<b>Técnica para fazer previsão</b>	<b>Horizonte de previsão</b>	<b>MAPE (%)</b>
Hong (2010)	Rede neural artificial	Ano	4,35
Catalan et al. (2008)	Rede <i>Levenberg-Marquardt</i> MLP	Semana	<13,7
Jurado et al. (2015)	Raciocínio Indutivo Difuso	Hora	4,49
Nengbao et al., (2014)	Séries temporais (SARIMAX)	Dia	<1,6
	MLP <i>Levenberg-Marquardt</i>	Semana	<3,6
Wang and Hong (2016)	Efeito recente para previsão de carga elétrica	Hora	3,39
Bantugon and Gallano (2016)	Rede neural e <i>Holter Winter</i>	Dia	5,64
Ghadimi et al. (2019)	Técnica de seleção de recursos e meta-heurística aprimorada	Semana	1,3
	Séries temporais (AIC)		<7,8
	Séries temporais (BIC)		<7,6
	MLP – variável de entrada no tempo $t$	Mês	<0,8%
	MLP – variáveis de entrada atrasada		<9,2%
Esse Artigo	Séries temporais (AIC)		2,45
	Séries temporais (BIC)		2,3
	MLP – variável de entrada no tempo $t$	Ano	0,015
	MLP – variáveis de entrada atrasada		0,75

## 6. CONCLUSÃO

As variáveis SI, NC e ST atrasadas em  $t-5$ ,  $t-10$  e  $t-20$  foram úteis, tendo em vista a variação do consumo em diferentes classes de consumo, apresentando resultados interessantes, atingindo o consumo de 32614 GWh no ano, sendo 0,756% acima do consumo anual da distribuidora, sobrecontratada em +244,70 GWh (Figura 27).

Os resultados mostraram que é necessário encontrar a configuração adequada do número de neurônios para cada classe de consumo nas experiências realizadas com as variáveis no tempo  $t$ . Por outro lado, quando as variáveis de entrada são atrasadas, é necessário encontrar o

atraso mais adequado para cada classe de consumo, pois são consumos independentes e, portanto, possuem características distintas de consumo de eletricidade e isso faz com que esta seja a desvantagem dessa abordagem, pois o EDC terá que treinar redes neurais diferentes para obter eficiência na previsão.

Até o final de 2013, as classes de consumo residencial e comercial de eletricidade se estabilizaram, enquanto o consumo das indústrias só diminuiu desde 2003. Provavelmente, por causa da situação econômica nacional, em que os empreendedores não investiram no país.

Como trabalhos futuros, poderiam ser usados quantidade maior de dados para cada classe de consumo e testar posteriormente, como  $t-30$ ,  $t-40$ ,  $t-50$  para verificar se a rede neural é capaz de capturar o comportamento do consumo de eletricidade de cada classe de consumo com mais precisão.

Outra questão interessante é usar a variável de saída (consumo de eletricidade) como variável de entrada atrasada e verificar se o modelo melhora ou piora.

## REFERÊNCIAS

ANEEL: Normative Resolution No. 703, from March 28, 2016. **Structure of the Taxes Regulation Procedures – PRORET.** Available at: <[http://www2.aneel.gov.br/cedoc/aren2016703\\_Proret\\_Submod\\_4\\_3\\_V0.pdf](http://www2.aneel.gov.br/cedoc/aren2016703_Proret_Submod_4_3_V0.pdf)>. Accessed Feb. 2, 2016.

SRE - REPORT IV: **Conceptual aspects of economic regulation, tariff model and training facility of rates in 25 countries studied.** Available at <<https://www.cpfl.com.br/energiassustentaveis/inovacao/projetos/Documents/PB302/aspectos-conceituais.pdf>>. Accessed Dec. 11, 2016.

HONG, T., FAN, S.: Probabilistic electric load forecasting: A tutorial review. **Int. J. Forecasting**, 32 (2016), 3, 914–938.

BANTUGON, M. J. T., GALLANO, R. J. C.: Short- and long-term electricity load forecasting using classical and neural network-based approach: A case study for the Philippines, **IEEE Region 10 Conference (TENCON)**, Singapore, 2016, 3822-3825.

ALMESHAEI, E., SOLTAN, H.: A methodology for Electric Power Load Forecasting. **Alexandria Engineering Journal** 50 (2011), 137–144.

CATALAO, J., MARIANO, S., MENDES, V., FERREIRA, L.: Short-term electricity prices forecasting in a competitive market: A neural network approach, **Electric Power Systems Research**, 77 (2007), 1297-1304.

SARADA, K., BAPIRAJU, V.: Comparison of day-ahead price forecasting in energy market using Neural Network and Genetic Algorithm, **Int. Conf. on Smart Electric Grid (ISEG)**, Guntur, 2014, 1-5.

KANKAL, M., UZLU, E.: Neural network approach with teaching-learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey. **Neural Computing and Applications**, 28 (2017), 1, 737–747.

SWAROOP, R.; HUSSEIN, A.: Load forecasting for power system planning and operation using artificial neural network at al batinah region oman. **Journal of Engineering Science and Technology**, 7 (2012), 4, 498-504.

KOUHI, S., KEYNIA, F., RAVADANEGH, S.: A new short-term load forecast method based on neuro-evolutionary algorithm and chaotic feature selection. **Int. J. Electr. Power Energy Syst.**, 62 (2014), 862–867.

LAZO, J. G. L., FIGUEIREDO, K., LIMA, D. A.: Analysis of an intelligent module of electric power contracting in short term for distributors. **Automation & Control (Rio de Janeiro)**, 23 (2012), 6, 711-725.

WANG, L., ZHANG, Z., CHEN, J.: Short-Term Electricity Price Forecasting with Stacked Denoising Autoencoders. **IEEE Transactions on Power Systems**. 32 (2016), 2673-2681.

DEB, C., ZHANG, F., YANG, J., LEE, S., SHAH, K.: A review on time series forecasting techniques for building energy consumption. **Renewable and Sustainable Energy Reviews**, 74 (2017), 902-924.

GUNAY, M.E.: Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: case of Turkey. **Energy Policy**, 90 (2016), 92–101.

MARINO, D., AMARASINGHE, K., MANIC, M.: Building Energy Load Forecasting using Deep Neural Networks. 42nd **Annual Conference of the IEEE Industrial Electronics Society**, Florence, 2016, 23-26.

KHASHEI, M., BIJARI, M.: An artificial neural network (p, d, q) model for time-series forecasting. **Expert Syst. Appl.**, 37 (2010), 479-489.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CEPB(c), (2016) version. **Contracts**. Available at: <<http://www.CEPB.org.br>>. Accessed on June 20, 2016.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CEPB(a), (2012) version. **Energy Balance**. Available at: <<http://www.CEPB.org.br>>. Accessed on June 20, 2016.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CEPB(d), (2016) version. **Price of the difference liquidation**. Available at: <<http://www.CEPB.org.br>>. Accessed on June 20, 2016.

ELECTRIC ENERGY COMMERCIALIZATION RULES – CEPB(b), (2016) version. **Results Consolidation**. Available at: <<http://www.CEPB.org.br>>. Accessed on June 20, 2016.

HAYKIN, S.: **Neural Networks: A Comprehensive Foundation**. Pearson, Toronto, 1999.

BOX, G., JENKINS, G.: **Time Series Analysis: Forecasting and Control**, Holden-Day, New Jersey, 2015.

MORETTIN, P. A., TOLOI, C. M. C.: **Time Series Analysis (in Portuguese)**. 2nd ed. Edgar Blücher, São Paulo, 2006.

MONTGOMERY, D. C., JENNINGS, C. L., KULAHCI, M.: **Introduction to Time Series Analysis and Forecasting**, 2nd ed. John Wiley & Sons Inc, New York, 2015.

AKAIKE, H.: A new look at the statistical model identification. **IEEE Transactions on Automatic Control** 19 (1974), 6, 716-723.

SCHWARZ, G.: Estimating the dimension of a model. **Annals of Statistics**, 6 (1978), 2, 461-464.

VERBEEK, M.: **A Guide to Modern Econometrics**. 2nd edition, John Wiley & Sons Inc., 31, 285–8, 2004.

OSMAN, Z., AWAD, M., MAHMOUD, T.: Neural network-based approach for short-term load forecasting. **IEEE/PES Power Systems Conference and Exposition**, Seattle, 2009, 15-18.

- DING, J., TAROKH, V., YANG, Y.: Bridging AIC and BIC: A New Criterion for Autoregression, **IEEE Transactions on Information Theory**, 64 (2018), 6, 4024 - 4043.
- JURADO, S., NEBOT, A., MUGICA, F. AVELLANA, N. (2015), “Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques”. **Energy**, 86, 276–291, 10.1016/j.energy.2015.04.039.
- GHADIMI, N., AKBARIMAJD, A., SHAYEGHI, H. ABEDINIA, O. (2018). “Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting”. **Energy**. 161. 10.1016/j.energy.2018.07.088.
- HONG, T. (2010), “**Short term electric load forecasting**”. North Carolina State University (2010).
- Fallah, S.N.; Ganjkhani, M.; Shamshirband, S.; Chau, K.-W. Computational Intelligence on Short-Term Load Forecasting: A Methodological Overview. **Energies** 2019, 12, 393.
- NENGBAO, Liu; BABUSHKIN, Vahan; AFSHARI, Afshin. (2014). Short-Term Forecasting of Temperature Driven Electricity Load Using Time Series and Neural Network Model. **Journal of Clean Energy Technologies**. 2. 10.7763/JOCET.2014.V2.149.
- LAHOUAR, J. B. H. Slama (2015), “Day-ahead load forecast using random forest and expert input selection”. **Energy Conversion Management**, 103, 1040–1051.
- MOHAMED, Zahra; BODGER, Patrick. (2005). Forecasting electricity consumption in New Zealand using economic and demographic variables. **Energy**. 30. 1833-1843. 10.1016/j.energy.2004.08.012.
- VORONIN, Sergey; PARTANEN, Jarmo; KAURANNE, Tuomo. (2014). A hybrid electricity price forecasting model for the Nordic electricity spot market. **International Transactions on Electrical Energy Systems**. 24. 10.1002/etep.1734.
- ZOR, Kasim; TIMUR, Oguzhan; TEKE, Ahmet. (2017). A state-of-the-art review of artificial intelligence techniques for short-term electric load forecasting. **Conference: 2017 6th International Youth Conference on Energy (IYCE)**. 1-7. 10.1109/IYCE.2017.8003734.

### **APÊNDICE C (Artigo 3) – Uma abordagem baseada em PCA para seleção e taxa de contribuição para previsão de carga de energia elétrica.**

*International Journal of Energy Sector Management – Open Access Hybrid.*

Qualis: A3

*Manuscript ID: IJESM-10-2019-0031*

*Title: A PCA-based variable ranking and selection approach for electric energy load forecasting.*

*Corresponding author:* Dr. Fabio Henrique Pereira.

*Corresponding Author's Institution:* Universidade Nove de Julho.

*First Author:* Elânio Bezerra; Fabio Henrique Pereira.

*Order of Authors:* Francisco Elânio Bezerra, Fabio Henrique Pereira.

*Order of Authors:* Francisco Elânio Bezerra<sup>1</sup>, Flávio Grassi<sup>2</sup>, Cleber Gustavo Dias<sup>2</sup>, Fabio Henrique Pereira.<sup>1,2\*</sup>

**Resumo:** Este artigo propõe uma abordagem baseada na Análise de Componentes Principais (PCA) para definir uma taxa de contribuição para cada variável e, em seguida, selecionar as principais variáveis como entrada para uma rede neural visando previsão de carga de energia na região sudeste do Brasil. **Projeto, metodologia, abordagem** – A abordagem proposta define uma taxa de contribuição de cada variável como a soma ponderada do produto interno entre a variável e cada componente principal. Portanto, a taxa de contribuição é usada para selecionar os recursos mais importantes de 27 variáveis e 6815 dados de eletricidade para um modelo de previsão com rede MLP com algoritmo *backpropagation*. Vários testes, começando pela variável mais significativa como entrada e adicionando a próxima variável mais significativa e assim por diante, são realizados para prever a carga de energia (GWh). Os testes de esfericidade KMO e Bartlett foram utilizados para verificar a consistência geral dos dados para análise fatorial. **Conclusões** – Embora a previsão de carga de energia seja área para a qual bancos de dados com dezenas ou centenas de variáveis estão disponíveis, a abordagem pôde selecionar apenas seis variáveis que contribuem com mais de 85% para o modelo. Embora as taxas de contribuição das variáveis das usinas, mais a troca de energia adicionada, tenham apenas 14,14% de contribuição, a variável energia armazenada tem a taxa de contribuição de 26,31%, sendo fundamental para a precisão da previsão. **Originalidade, valor** – Além de melhorar a precisão da previsão e fornecer um modelo preditor mais rápido, a abordagem baseada em PCA, proposta para calcular a taxa de contribuição das variáveis de entrada fornece melhor

entendimento do processo subjacente que gerou os dados, fundamental para a realidade brasileira devido às acentuadas variações climáticas e econômicas.

**Palavras-chave:** previsão de carga de energia, PCA, seleção de variáveis, classificação variável.

***A PCA-based variable ranking and selection approach for electric energy load forecasting***

**Abstract:** *This paper proposes an approach based upon the Principal Components Analysis (PCA) to define a contribution rate for each variable and then select the main variables as input to a neural network for energy load forecasting in the region south-eastern Brazil.*

**Design/methodology/approach** – *The proposed approach defines a contribution rate of each variable as a weighted sum of the inner product between the variable and each principal component. So, the contribution rate is used for selecting the most important features of 27 variables and 6815 electricity data for a MLP network backpropagation prediction model. Several tests, starting from the most significant variable as input, and adding the next most significant variable and so on, are accomplished to predict energy load (GWh). The KMO and Bartlett sphericity test were used to verify the overall consistency of the data for factor analysis.*

**Findings** – *Although energy load forecasting is an area for which databases with tens or hundreds of variables are available, the approach could select only 6 variables that contribute more than 85% for the model. While the contribution rates of the variables of the plants, plus energy exchange added, have only 14.14% of contribution, the variable the stored energy has a contribution rate of 26.31% being fundamental for the prediction accuracy. **Originality/value** – *Besides improving the forecasting accuracy and providing a faster predictor, the proposed PCA-based approach for calculating the contribution rate of input variables providing a better understanding of the underlying process that generated the data, which is fundamental to the Brazilian reality due to the accentuated climatic and economic variations.**

**Keywords:** *energy load forecasting, PCA, features selection, variable ranking.*

## 1. INTRODUÇÃO

O setor elétrico está prestes a sofrer um processo de colapso tecnológico, em que uns dos vetores mais altos dessa ruptura é o potencial de expansão de distribuição dos recursos

energéticos, impondo a tendência à descentralização e acelerando a difusão de distribuição de sistemas de micro e minigeração. Paralelamente, as políticas de resposta à demanda estão se tornando cada vez mais relevantes, permitindo que a demanda por energia se torne mais flexível (ANEEL, 2016).

Cerca de 5,6 bilhões de pessoas consomem eletricidade no mundo. Assim, a geração, distribuição e o consumo de energia desempenham papel crítico na sustentação do planeta e o principal desafio está na integração inteligente das ações de todos os usuários conectados à rede (JURADO et al., 2015; TAO HONG, 2014).

Segundo Keles et al. (2016), as trocas de eletricidade foram cada vez mais estabelecidas nos países europeus. O volume de energia do mercado *spot* de câmbio de energia na Europa cresceu de 311 TWh em 2011 para 346 TWh em 2013 nas três áreas da Alemanha/Áustria, Suíça e França.

No Brasil, as escolhas e realizações da política energética se comparam a alguns dos desafios energéticos mais urgentes do mundo, onde quase 45% da demanda de energia primária é atendida por energia renovável, tornando o setor de energia do Brasil um dos menos responsáveis por emissões de carbono do mundo. A demanda total de energia primária dobrou no Brasil desde 1990, liderada por forte crescimento no consumo de eletricidade e no aumento da demanda por combustíveis para transporte, apoiada pelo crescimento econômico e pelo crescimento da classe média (IEA, 2019). Por exemplo, em 2003, o país consumiu 267,09 TWh de energia, enquanto o consumo foi de 345,38 TWh, 341,98 TWh, 330,11 TWh, 313,65 TWh e 312,57 TWh de energia elétrica em 2014, 2015, 2016, 2017 e 2018, respectivamente, aumento de quase 30% de 2003 a 2014 (AIE, 2019).

Diante desse cenário, para a operação eficaz dos sistemas de energia é essencial a previsão precisa da carga de energia para diferentes horizontes temporais. Existem quatro horizontes: 1) curtíssimo prazo – previsão de algumas horas; 2) curto prazo – previsões de um dia a semanas; 3) médio prazo – previsões de até um ano e 4) longo prazo – previsões acima de um ano (LAHOUAR e SLAMA, 2015; GHADIMI et al., 2018; HONG e FAN, 2016). Seguindo esse raciocínio, tem-se dois problemas típicos de previsão: a curto prazo, no qual é necessário prever as cargas de tempo para os próximos dias; e a longo prazo: no qual é necessário prever os picos sazonais para os próximos anos (TAO HONG, 2014). No entanto, nos últimos anos, a previsão de curto prazo tem sido a mais pesquisada (HONG e FAN, 2016; YANG et al., 2019).

Como a previsão de carga de energia é problema não linear e tem alto nível de volatilidade, é necessário usar métodos inteligentes e ferramentas de previsão apropriadas (GHADIMI et al., 2018; LAHOUAR et al., 2015). Muitas técnicas têm sido usadas para prever

a carga nos últimos anos, sendo as mais populares as redes neurais, análise de regressão e análise de séries temporais (NOWOTARSKI et al., 2016).

No entanto, apesar dos bons resultados apresentados pelos métodos de previsão de carga de energia na literatura acadêmica, essas técnicas sofrem com a falta de eficiência nas técnicas de seleção de características para modelar as não linearidades e as variáveis de interação dos processos de previsão. Assim, um passo essencial para muitas técnicas é determinar quais variáveis explicativas usar. Além disso, há pouca pesquisa sobre a seleção de recursos ideais de entrada (seleção de recurso) quando um modelo de previsão é estabelecido (HONG, 2010; ABEDINIA et al., 2017; YANG et al. 2019; SPEISER et al. 2019).

Diante desse cenário, este trabalho propõe uma nova abordagem para calcular a taxa de contribuição para as variáveis de entrada na previsão de carga de energia a curto prazo. O artigo está organizado da seguinte forma: a seção 1 discute o problema e a importância da previsão de carga de energia a curto prazo. A seção 2 fornece uma revisão da literatura e a seção 3 apresenta a base teórica. O método usado neste trabalho é abordado na seção 4 e os resultados estão na seção 5. Finalmente, a seção 6 apresenta nossas conclusões.

## **2. TRABALHO RELACIONADO**

Recentemente, várias técnicas foram propostas para resolver o problema de previsão de carga de energia. Apesar dos modelos híbridos cada vez mais inteligentes projetados para essa tarefa, ainda são escassas as abordagens para selecionar as principais variáveis explicativas para fins de previsão (ABEDINIA et al., 2017; YANG et al. 2019). Os métodos de seleção de variáveis são fundamentais porque podem identificar um subconjunto de variáveis explicativas que contêm as informações mais relevantes no conjunto de dados completo e, portanto, podem influenciar a precisão da previsão (YANG et al., 2019; PERES e FOGLIATTO, 2018; ABEDINIA et al., 2017; SHAFI et al., 2019; LI et al., 2019).

Peña e Poncela (2006) compararam modelos para redução de dimensão em séries temporais e testes de dimensão de estrutura dinâmica. Eles consideraram séries temporais estacionárias e não estacionárias e discutiram os principais componentes, análises canônicas, modelos de componentes escalares, modelos de classificação reduzida e modelos de fatores. Eles concluíram que a análise de correlação canônica entre os valores atuais e passados das séries temporais seja ferramenta muito poderosa para a redução de dimensão. Além disso, essa abordagem permite a visão unificada de muitos procedimentos propostos para redução de tamanho, como a inclusão de componentes principais, a análise canônica de Box e Tiao (1977),

os modelos de classificação reduzida de Reinsel et al. (1998) e os modelos de componentes escalares de Tiao e Tsay (1989).

Analisando esse problema, algumas técnicas foram propostas para selecionar as variáveis que mais influenciam o modelo de previsão. Ghosh et al. (2014) propuseram um método baseado na otimização estocástica para identificar o subconjunto ideal de variáveis medidas para monitoramento de processos, uma vez que a incorporação de variáveis medidas, que não fornecem informações adicionais sobre falhas, prejudica o monitoramento de desempenho de processos estatísticos.

Lahouar e Slama (2015) propuseram um plano para prever a carga de curto prazo no qual uma técnica especializada de seleção de recursos foi usada com base em um conjunto de regras *if-then*. Os valores de carga dos dois dias anteriores no mesmo horário, os picos da manhã e da noite do dia anterior, o número do mês, o tipo de dia e a temperatura máxima e mínima do dia foram utilizados como entradas. A ideia era prever as próximas 24 horas de carga usando o aprendizado de máquina (floresta aleatória), caracterizado por imunidade das variações de parâmetros e validação interna cruzada.

Ghadimi et al. (2018) desenvolveram estudo sobre previsão de carga de curto prazo com uma hora e um dia de antecedência usando variáveis históricas de carga, temperatura e umidade com estratégia de previsão híbrida, incluindo a técnica de seleção em que o sinal de carga foi filtrado pela técnica de seleção de recursos para selecionar candidatos apropriados como entradas para o mecanismo de previsão e, em seguida, o mecanismo de previsão, em dois estágios, foi implementado com base nas redes neurais *Ridgelet* e *Elman*. Uma rede *perceptron* multicamada (MLP) com o algoritmo *Levenberg-Marquardt* foi implementada com e sem seleção de recursos e, com isso, foi possível atingir menor erro médio percentual absoluto (MAPE) com a técnica de seleção de recursos.

Bannor Boateng e Acheampong (2019) desenvolveram trabalhos de previsão de demanda de energia evidenciados internacionalmente na Austrália, China, França, Índia e EUA, nos quais foram utilizados oito parâmetros de entrada para modelar a demanda de energia, desenvolvimento financeiro, crescimento econômico, industrialização, população, abertura comercial, urbanização, preço da energia e investimento estrangeiro direto. Um MLP com algoritmo *backpropagation* foi utilizado para o treinamento dos dados e uma análise de sensibilidade foi realizada para determinar as variáveis mais influentes no sistema.

Hong (2010) propôs um mecanismo de seleção de variáveis acoplado à regressão linear aplicada, redes neurais artificiais (RNA) ou regressão nebulosa para previsão de curto prazo.

Os resultados mostraram que, para cada uma das três técnicas, o mecanismo proposto foi capaz de reduzir os erros de previsão.

Speiser et al. (2019) compararam métodos de seleção *random forest* para modelagem de previsão de classificação. Eles usaram 311 conjuntos de dados de classificação disponíveis on-line gratuitamente e avaliaram as taxas de erro de previsão, o número de variáveis, o tempo computacional e a área sob a curva de operação do receptor para muitos métodos de seleção *random forest*.

Wang et al. (2016) adotaram uma abordagem de *big data* para seleção de variáveis, em que o algoritmo permite a seleção de muitas variáveis de temperatura defasadas e médias para aumentar a precisão da previsão.

Li (2016) e Li (2019) propuseram um agrupamento multivariado de séries temporais com base na análise de componentes principais padrão, inspirados no princípio *k-means*. Assim, são incluídos dois estágios principais, construindo o espaço de coordenação da projeção para um cluster e reatribuindo membros de séries temporais multivariadas para cada cluster. Nessa abordagem, são definidos dois parâmetros, o número (K) de clusters e a menor dimensão reduzida (p). Em experimentos multivariados de conjuntos de dados de séries temporais, o número K de clusters foi definido de maneira uniforme para o número correspondente de classes, enquanto a dimensão foi reduzida pela criação de autovetores retidos da matriz de covariância padrão para cada cluster.

Li et al. (2019) propuseram uma representação de recursos e mensuração de similaridade com base na sequência de covariância para séries temporais multivariadas, na qual a covariância entre duas variáveis foi aplicada para analisar a extração de características para cada série temporal multivariada. Assim, uma sequência de covariância pode ser construída para representar a característica multivariada de séries temporais. Além disso, a análise de componentes principais (PCA) foi usada para extrair as características da sequência de covariância derivadas de um conjunto de dados de séries temporais multivariadas.

Liang et al. (2019) propuseram um modelo híbrido que combina decomposição do modo empírico (EMD), regressão mínima e máxima de redundância (mRMR) e uma rede neural de regressão geral (GRNN) com algoritmo de otimização *fly fruit* (FOA) para previsão de carga. Liang et al. (2019) decompuseram uma série de carga original em várias funções intrínsecas (FMI) e um resíduo com frequências diferentes reduziu a volatilidade da série influenciada por fatores complicados. O mRMR foi então empregado para obter o melhor conjunto de recursos pela análise de correlação entre cada FMI e os recursos, incluindo tipos de dias, temperatura e condições climáticas. Finalmente, FOA foi aplicado para otimizar o fator de suavização no

GRNN, onde a carga máxima esperada foi obtida pela soma dos resultados esperados para todos os FMIs.

Chinnathambi et al. (2018) investigaram três tipos de técnicas de seleção, como importância relativa, usando regressão linear (LR), *splines* de regressão adaptativa multivariada (MARS) e floresta aleatória (RF) para reduzir o erro de previsão do preço à vista por hora na eletricidade ibérica mercados. Essas técnicas foram investigadas por vários períodos, como uma semana, duas semanas, três semanas e um mês, 45 dias, 60 dias, 75 dias e 90 dias. Além disso, eles também foram testados para um conjunto de dados de dias da semana e fins de semana, com duração de um mês, dois meses, três meses e seis meses. Finalmente, um modelo ARIMA de dois estágios foi testado para um conjunto de dados com e sem variáveis explicativas para entender a influência da previsão residual. Os resultados mostraram que a técnica ARIMA com a máquina de vetores de suporte obteve melhores resultados para um conjunto de dados menor.

Lang et al. (2018) abordam o problema de previsão de séries temporais multivariadas, no qual a carga elétrica e os dados de temperatura correspondentes foram analisadas como séries temporais correlativas e reconstruídos para o espaço de fase multivariado. Informações mútuas foram usadas para escolher o atraso temporal para cada série temporal escalar, enquanto redes neurais e máquinas de vetores de suporte foram usadas para previsão.

Keles et al. (2016) propuseram uma metodologia de previsão estendida para os preços *spot* da eletricidade no dia seguinte. Eles aplicaram redes neurais artificiais, feitas usando diferentes algoritmos de cluster, focadas na seleção e preparação dos principais dados de entrada que têm impacto perceptível nos preços da eletricidade. Seu trabalho analisou as características dos preços, removeu os componentes sazonais e de tendência, normalizou os dados, determinou o atraso de tempo nas variáveis de entrada usando o método de informação mútua, determinou a abordagem para a previsão multiperódica, selecionou um subconjunto de dados de entrada disponíveis usando um algoritmo *nearest-neighbour* e otimizou a rede neural.

Platon et al. (2015) propuseram um estudo para prever o consumo horário de eletricidade de um edifício institucional. Usando dados de medição de uma instalação institucional canadense e informações climáticas preliminares, o PCA foi usado para reduzir o número de entradas sem diminuir a precisão do modelo, identificando variáveis significativas que contêm a maior parte da variabilidade geral presente no conjunto de dados e, para previsão horária, foram utilizados técnicas de inteligência, RNA e raciocínio baseado em casos (RBC).

Por último, mas não menos importante, Jahangoshai Rezaee et al. (2019) propuseram um trabalho de previsão de energia a curto prazo para uma usina em Isfahan Providence, no centro do Irã. Para esse fim, eles coletaram dados horários de 2014 a 2015, atingindo 2.374

amostras nas quais foram usadas como entrada temperatura, umidade, pressão do ar, velocidade do vento, direção do vento em um sistema de inferência *neuro-fuzzy* adaptável (ANFIS) com algoritmos meta heurísticos (PSO, GA, DE, ACO). Embora os autores não utilizassem técnicas de seleção variável obtiveram resultados interessantes, RMSE de 1,5706%, média de 0,015456, DST de 1,5724 e R2 de 0,96572.

Como vimos, há uma lacuna nessa área de seleção de recursos. Para preenchê-la, algumas técnicas foram propostas, como sequência de covariância e PCA para séries temporais multivariadas, usadas por Li et al. (2019); Regressão de redundância mínima e máxima (mRMR), usada por Liang et al. (2019); regressão linear (RL), *splines* de regressão adaptativa multivariada (MARS) e floresta aleatória (FR), testadas por Chinnathambi et al. (2018); informação mútua, usada por Lang et al. (2018) e Keles et al. (2016); e t-SNE, usado por Shafi et al. (2019) e Kim e Cho (2019). Assim, algumas técnicas foram propostas para determinar as variáveis de entrada mais importantes para melhorar a precisão da técnica de previsão.

No entanto, como não há unanimidade no uso de determinada técnica para seleção de recurso ou variável e taxa de contribuição, a literatura mostra que há alguma preocupação em selecionar as variáveis que serão usadas no modelo de previsão. Sendo assim, este trabalho propõe uma nova abordagem usando o PCA para seleção e taxa de contribuição das variáveis de entrada em um modelo de previsão de carga de energia de curto prazo na Região Sudeste do Brasil.

### **3. QUADRO TEÓRICO**

#### **3.1 REDE NEURAL ARTIFICIAL**

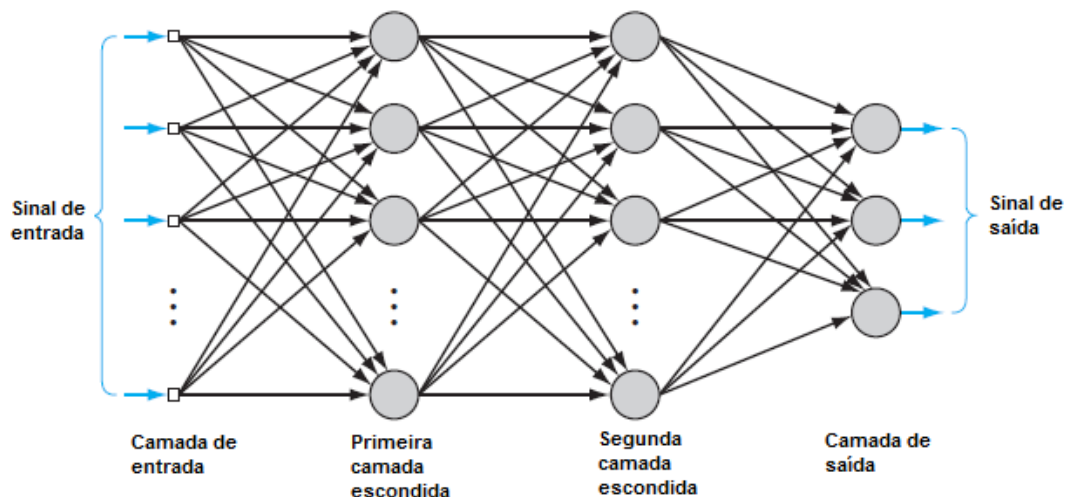
Nos últimos anos, as redes neurais artificiais (RNAs) constituem ferramenta de inteligência artificial (IA) incorporada na modelagem de inteligência humana, que trabalha em paralelo e cujo conhecimento é adquirido pela aprendizagem, onde sinapses ponderadas são usadas para o conhecimento. Sua principal característica é armazenar conhecimento experimental para uso (HAYKIN, 2008).

A MLP é rede neural artificial que pode trabalhar com várias superfícies de decisão não lineares, usa várias regras de aprendizado e possui alto grau de conectividade pelos pesos sinápticos (HAYKIN, 2008).

A Figura 39 mostra a estrutura de uma rede MLP, que consiste em uma camada de entrada, duas camadas ocultas e uma camada de saída. O sinal de entrada se propaga em direção única, camada por camada. A camada de entrada receberá dados de entrada da rede e a primeira

camada oculta é alimentada a partir da camada de entrada, que, por sua vez, é constituída pelos nós de origem; as saídas resultantes da primeira camada oculta são aplicadas à camada oculta. A camada de saída tem então a função de processar as informações (HAYKIN, 2008).

**Figura 39 – Uma rede MLP com duas camadas ocultas**



Fonte: Haykin, 2008, p. 124

Uma função amplamente usada para determinar os valores de mudança para cada peso é a função logística, que define a saída de um neurônio em termos de “ $v$ ”. Quando “ $v$ ” é maior ou igual a 0,  $\varphi(v)$  será igual a 1; se “ $v$ ” for menor que 0,  $\varphi(v)$  será 0. O “ $v$ ” representa a soma das entradas mais o limite de ativação. O sinal de saída é representado pela soma dos produtos dos sinais de entrada e o peso das sinapses mais o viés.

Para minimizar erros no treinamento da rede, um algoritmo de *Levenberg-Marquardt* pode ser implementado, ele funciona em duas fases. A fase de avanço, na qual os pesos sinápticos são fixos e o sinal de entrada é propagado pela rede, camada por camada, até a saída e, a fase de retorno na qual um sinal de erro é produzido pela comparação de uma saída de rede com a resposta desejada. O sinal é propagado de volta pela rede, camada por camada, resultando em ajustes contínuos dos pesos sinápticos (HAYKIN, 2008).

### 3.2 ANÁLISE DE COMPONENTES PRINCIPAIS

O PCA foi inventado em 1901 por Karl Pearson. É um método estatístico que usa transformação ortogonal para converter um conjunto de observações de variáveis correlacionadas em um conjunto de valores de variáveis lineares não correlacionadas, chamadas componentes principais (PEARSON, 1901). Esse método é usado para reduzir a

dimensionalidade de grandes conjuntos de dados de maneira não supervisionada, transformando um amplo conjunto de variáveis em outro menor que preserva o máximo de informações possível.

Considere “k” variáveis aleatórias e medições (Y) de “n” amostras, o PCA encontra alguns componentes que descrevem a maior parte da variação nos dados usando “k” das variáveis aleatórias  $Y_1, Y_2, \dots, Y_k$  para produzir os componentes  $Z_1, Z_2, \dots, Z_k$  que não estão correlacionados. Os componentes podem ser pedidos com base na quantidade de variação explicada por cada componente,  $var(Z_1) \geq var(Z_2) \geq \dots \geq var(Z_k)$ . Portanto, o primeiro componente significativo,  $Z_1$ , explica a quantidade máxima de variação, o segundo componente significativo,  $Z_2$ , explica a segunda maior quantidade de variação, que é menor que a quantidade explicada por  $Z_1$ , e assim por diante para outros componentes (ALKARKHI, 2019).

Algebricamente, os principais componentes são representados por uma combinação linear de todas as variáveis  $Y_1, Y_2, \dots, Y_k$  e os coeficientes dos componentes principais  $a_{ij}$ , como pode ser observado na Equação 40:

$$\begin{aligned} Z_1 &= a_{11}Y_1 + a_{12}Y_2 + a_{13}Y_3 + \dots + a_{1k}Y_k \\ Z_2 &= a_{21}Y_1 + a_{22}Y_2 + a_{23}Y_3 + \dots + a_{2k}Y_k \\ &\vdots \\ Z_3 &= a_{k1}Y_1 + a_{k2}Y_2 + a_{k3}Y_3 + \dots + a_{kk}Y_k. \end{aligned} \tag{40}$$

Como  $Z_1$  não está correlacionado a  $Z_2$ , temos  $cov(Z_1, Z_2) = 0$ .

Para calcular os componentes principais, é necessário calcular a matriz de covariância, na qual os valores diagonais representam a variação, enquanto os valores fora da diagonal representam a covariância das variáveis, entre diferentes variáveis. Em seguida, é necessário calcular os autovalores,  $\lambda_i$ , da matriz de covariância, nos quais o número de autovalores é igual ao número de variáveis e pode ser organizado na ordem  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq 0$ . Os valores próprios representam as variações dos componentes principais,  $Z_i$ , o que significa o valor próprio  $\lambda_1$  representa a variação dos primeiros componentes significativos  $Z_1$ ,  $\lambda_2$  representa a variação do segundo componente significativo  $Z_2$  e assim por diante (ALKARKHI, 2019).

Após o cálculo dos autovalores, os autovetores associados a cada autovalor devem ser calculados. Se os autovalores forem distintos, os autovetores correspondentes serão ortogonais. Por fim, é necessário selecionar os primeiros componentes que representam a maior parte da variação para representar os dados.

A soma de todos os valores próprios é igual à soma dos elementos diagonais da matriz de covariância, que é igual à soma das variações das variáveis originais (Equação 41). Portanto,  $S_{ii}$  representa a variância das variáveis, que é apresentada na diagonal da matriz de covariância, e os coeficientes do componente principal  $a_{ij}$ , medem a importância da variável para o componente principal, independentemente das demais variáveis.

$$\sum_{i=1}^k S_{ii} = \lambda_1 + \lambda_2 + \dots + \lambda_k \quad (41)$$

### 3.3 TESTE DE KAISER-MEYER-OLKIN E BARTLETT

O critério ou teste KMO (1977) é para identificar se o modelo de análise fatorial em uso está adequadamente ajustado aos dados, testando a consistência geral dos dados. O método verifica se a matriz de correlação inversa está próxima à matriz diagonal, que consiste em comparar os valores dos coeficientes de correlação linear observados aos valores dos coeficientes de correlação parciais. A Equação 42 mostra a medida de adequação, onde  $r_{jk}$  é o coeficiente de correlação simples entre as variáveis  $X_j$ ,  $X_k$  e  $p_{jk}$  é o coeficiente de correlação parcial entre  $X_j$  e  $X_k$ , dados os outros  $X_s$ .

$$KMO = \frac{\sum \sum_{j \neq k} r_{jk}^2}{\sum \sum_{j \neq k} r_{jk}^2 \sum \sum_{j \neq k} p_{jk}^2} \quad (42)$$

O teste KMO retorna valores entre 0 e 1, quanto mais próximo de 1 indica que as amostragens são adequadas, geralmente indicam que a análise fatorial pode ser útil para trabalhar com dados, quanto mais próximo de 0 indica que há grandes correlações parciais em comparação com a soma de correlações, existem correlações generalizadas, gerando problemas para a análise fatorial e os resultados da análise fatorial provavelmente não serão muito úteis.

#### 3.3.1 Teste de Bartlett

O teste de esfericidade de Bartlett, criado por Maurice Stevenson Bartlett em 1937, indica a força do relacionamento entre variáveis. A técnica testa a hipótese de que a matriz de correlação é matriz de identidade (matriz na qual todos os elementos diagonais são 1 e todos os

elementos diagonais externos são 0, o que indicaria que suas variáveis não são relacionadas e, portanto, inadequadas para a detecção da estrutura. Pequenos valores (abaixo de 0,05) do nível de significância indicam que a análise fatorial pode ser útil com seus dados.

As Equações 43 e 44 mostram como Maurice Stevenson Bartlett desenvolveu o teste de esfericidade de Bartlett, onde:  $n$  = tamanho da amostra,  $p$  = número de variáveis,  $|R|$  = determinante da matriz de correlação e  $v$  é uma distribuição qui-quadrado com graus de liberdade.

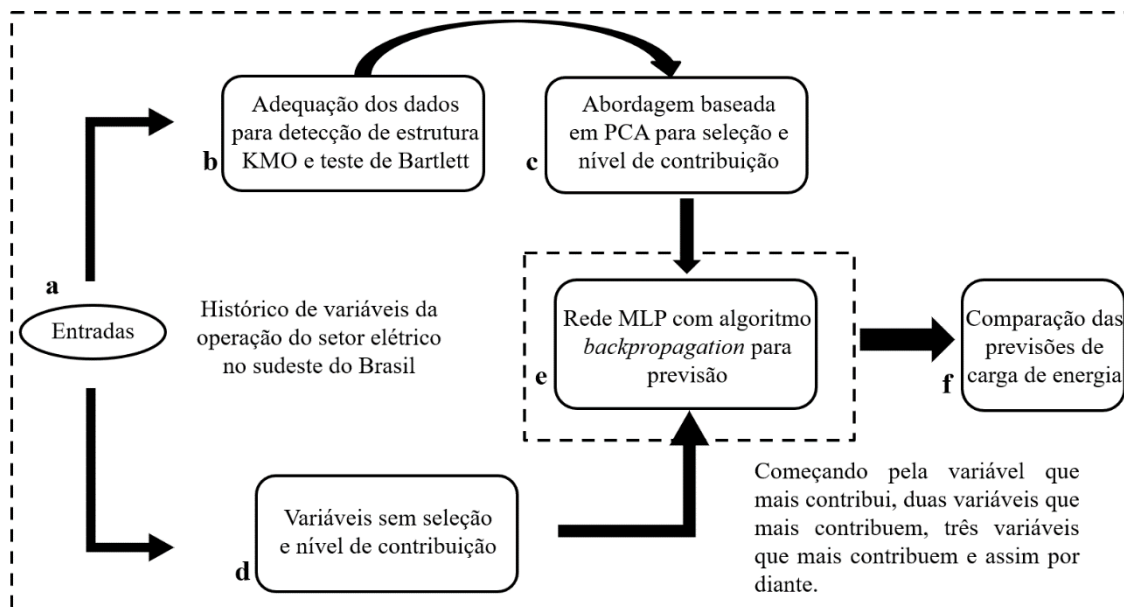
$$X^2 = - \left[ (n - 1) - \frac{2p + 5}{6} \right] \ln|R| \quad (43)$$

$$v = \frac{p(p - 1)}{2} \quad (44)$$

#### 4. METODOLOGIA

A Figura 40 mostra a abordagem proposta para a taxa de seleção e contribuição das variáveis para a previsão da carga de energia em GWh. Consiste na coleta de dados, detecção de estrutura para verificação e adequação dos dados usando KMO e Bartlett, uso de PCA para seleção e taxa de contribuição de variáveis no modelo, uso de rede MLP com algoritmo *backpropagation* para fazer previsões com as variáveis que possuem a maior e menor taxa de contribuição, comparação de previsões para verificar as variáveis de entrada que geram os menores erros.

Figura 40 – Diagrama de abordagem proposto



#### 4.1 COLETA DE DADOS, KMO E TESTE DE BARTLETT

O bloco a mostra como as variáveis foram coletadas para a previsão de carga de energia GWh na Região Sudeste do Brasil. Atualmente, foram coletadas 27 variáveis, tais como: energia armazenada (GWh), demanda instantânea (MW), demanda horária máxima (MWh/h), afluyente natural de energia bruta (MWmed), afluyente natural de energia armazenável (MWmed), energia carga (MWmed), troca de energia (MWmed) e dados hidrológicos de cinco grandes hidrelétricas que fornecem eletricidade para o Sudeste do Brasil, Itaipu, Ilha Solteira, Marimbondo, Três Marias e Furnas, que em 2016, 2017 e 2018 geraram 43,15%, 39,12% e 37,18% do total da energia hidrelétrica para a região sudeste do Brasil (ONS, 2019).

As variáveis utilizadas em relação às usinas foram vazão afluyente ( $m^3/s$ ), que se refere ao fluxo que atinge uma usina hidrelétrica ou a uma estrutura hidráulica, vazão efluente ( $m^3/s$ ), que representa o fluxo que sai da usina hidrelétrica ou da estrutura hidráulica, nível de quantidade (m), que são os pontos em direção à fonte, na direção oposta à corrente, a montante e, finalmente, o volume líquido em porcentagem, que representa o volume do reservatório entre o valor máximo normal de operação e o nível operacional mínimo normal.

As variáveis foram coletadas no site do governo do Operador Nacional do Sistema (disponível em <http://ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>). Com as informações coletadas diariamente de 01/01/2001 a 31/08/2019.

Vários trabalhos, como: Lahouar e Slama (2015), Ghadimi et al. (2018), Jahangoshai Rezaee et al. (2019), Dudek (2016), Yang et al. (2019), Jurado e col. (2015), Wang et al. (2016), Nowotarski et al. (2016), Liang et al. (2019), Lang et al. (2018), Lahouar (2015), entre outros, utilizam a carga de energia elétrica como variável de resposta e, como no mercado elétrico, o consumo de energia dos clientes é variável muito importante, pois varia a todo instante, portanto, essa variável foi escolhida como resposta (saída para o modelo de previsão para testar a contribuição de variáveis como entrada para a rede neural).

O bloco b refere-se ao uso das técnicas KMO e Bartlett, pois, como mencionado anteriormente, esses testes indicam a adequação dos dados para a detecção da estrutura. Para isso, o software *IBM SPSS Statistics 22* foi utilizado para realizar os dois testes aplicados em uma matriz com 27 variáveis e 6815 dados para cada variável.

A análise dos resultados mostra os valores gerados pelas técnicas KMO e Bartlett. A Tabela 23 mostra que o KMO atingiu 0,844, indicando que a taxa de variação de variáveis pode ser causada por fatores subjacentes, mostrando que a análise fatorial é útil para os dados (CERNY e KAISER, 1977), enquanto o teste de esfericidade de Bartlett mostrou significância zero, indicando que a análise fatorial é útil com os dados.

**Tabela 23 – Resultados do teste KMO e Bartlett**

<b>Técnica</b>	<b>Teste</b>	<b>Valor</b>
Kaiser-Meyer-Olkin	Medida de adequação da amostra	0,844
Teste de esfericidade de Bartlett	Approx. Qui-square	423456,3
	Df	351
	Sig.	0

## **4.2 CALCULANDO UMA TAXA DE CONTRIBUIÇÃO E SELECIONANDO AS PRINCIPAIS VARIÁVEIS DE ENTRADA USANDO O PCA**

O PCA uma análise multivariada simples baseada na decomposição de autovalores da matriz de covariância de dados e o objetivo é configurar uma imagem de menor dimensão dos dados para revelar a estrutura interna que melhor explica a variância (SCHIMIT e PEREIRA, 2018). Essa técnica é frequentemente usada quando o sistema possui muitas variáveis de entrada e é necessário encontrar as mais influentes para a saída (JOLLIFFE, 2002).

O uso do PCA neste estudo segue os mesmos parâmetros adotados no estudo de Schimit e Pereira (2018), mas os autores utilizaram o PCA para analisar quais parâmetros topológicos são mais relevantes para o sucesso ou insucesso de uma doença.

Como esse problema possui 26 variáveis de entrada para uma saída, o PCA prova ser técnica interessante para resolver parte do problema. A Figura 29 do bloco c mostra como foi desenvolvida a abordagem para a taxa de seleção e contribuição de variáveis usando PCA no *Matlab* 2018a.

O procedimento proposto possui as seguintes etapas:

- 1) computa uma matriz  $S$  com coeficientes de regressão padronizados para cada variável de entrada;
- 2) executa o PCA sobre o conjunto de dados  $S$ ;
- 3) classifica os componentes principais em ordem crescente por valor próprio. Nesse caso, foram selecionados 4 componentes principais que representam mais de 99% dos dados;
- 4) cria a matriz de projeção  $P$ ;
- 5) calcula a matriz  $S'$  como a representação de  $S$  no espaço do componente principal, representado por  $S' = S_{k \times d} P_{d \times p}$ ;
- 6) calcula uma taxa de contribuição  $C$  de cada variável de entrada  $X_j$ ,  $j = 1, 2, \dots, d$ :  $C_{(j)} = \sum_{t=1}^p (X_j^T (P_{ji} X_j) \lambda_i)$ ; e
- 7) classifica as variáveis de entrada em ordem crescente por índice de contribuição.

Após a aplicação do PCA, a Tabela 24 mostra a variabilidade dos dados explicada por componentes (%) e o valor acumulado, demonstrando que o primeiro, segundo terceiro e quarto componente, que explicam 57,91%, 33,94%, 5, 98% e 1,31% dos dados, enquanto a Tabela 25 mostra os quatro componentes principais que representam mais de 99% da variabilidade dos dados das 27 variáveis, enquanto 23 componentes explicam menos de 1% dos dados.

Tabela 24 – Variabilidade dos dados explicada pelos principais componentes (%).

<b>Componente</b>	<b>Por componente</b>	<b>Cumulativo</b>	<b>Componente</b>	<b>Por componente</b>	<b>Cumulativo</b>
1	57,91	57,91	14	0,00	99,99
2	33,94	91,85	15	0,00	100,00
3	5,98	97,83	16	0,00	100,00
4	1,31	99,14	17	0,00	100,00
5	0,36	99,49	18	0,00	100,00
6	0,20	99,70	19	0,00	100,00
7	0,13	99,83	20	0,00	100,00
8	0,06	99,89	21	0,00	100,00
9	0,04	99,93	22	0,00	100,00
10	0,02	99,95	23	0,00	100,00
11	0,02	99,97	24	0,00	100,00
12	0,01	99,98	25	0,00	100,00
13	0,01	99,98	26	0,00	100,00
14	0,00	99,99	27	0,00	100,00

Tabela 25 – Coeficientes variáveis nos quatro primeiros componentes principais

Variáveis	Comp. 1	Comp. 2	Comp. 3	Comp. 4
Energia armazenada (GWh)	0,9407	-0,3324	0,0339	0,0349
Demanda instantânea (MW)	-0,0002	0,0359	0,5776	0,0463
Demanda horária máxima (MWh / h)	-0,0020	0,0407	0,5803	0,0452
Afluente Natural de Energia Bruta (MWmed)	0,2658	0,7215	-0,0625	-0,3311
Afluente Natural de Energia Armazenável (MWmed)	0,1933	0,5935	-0,0487	0,5255
Carga de energia (MWmed)	-0,0125	0,0646	0,5614	0,0113
Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Itaipu	0,0392	0,0646	0,0384	-0,5209
Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Itaipu	0,0380	0,0592	0,0602	-0,5154
Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Ilha Solteira	0,0354	0,0182	-0,0079	-0,1366
Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Ilha Solteira	0,0331	0,0119	0,0027	-0,1425
Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Marimbondo	0,0161	0,0173	-0,0074	-0,0303
Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Marimbondo	0,0147	0,0057	0,0031	-0,0488
Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Furnas	0,0092	0,0196	-0,0152	0,0022
Carga de energia (GWh)	0,0001	-0,0001	-0,0001	0,0000
Dados hidrológicos / nível de quantidade (m) – Furnas	0,0057	-0,0016	0,0046	-0,0213
Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Furnas	0,0086	0,0192	-0,0189	0,0161
Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Três Marias	0,0275	0,0184	-0,0295	0,1692
Troca de Energia (MWmed)	0,0002	-0,0001	0,0000	0,0002
Dados hidrológicos / nível da quantidade (m) – Três Marias	0,0059	0,0015	-0,0086	-0,0103
Dados hidrológicos / fluxo de desfluentes m <sup>3</sup> /s – Três Marias	0,0001	0,0000	0,0000	0,0000
Dados hidrológicos / nível da quantidade (m) – Marimbondo	0,0000	0,0000	0,0000	0,0001
Dados hidrológicos / nível da quantidade (m) – Ilha Solteira	0,0000	0,0000	0,0000	-0,0001
Dados hidrológicos / nível da quantidade (m) – Itaipu	0,0006	-0,0001	-0,0001	0,0010

### 4.3 TREINAMENTO E TESTE DA REDE MLP

Uma rede MLP com algoritmo de treinamento *Levenberg-Marquardt* e função de ativação tangente sigmoide foi usada para treinar as 26 redes no *Matlab 2018a*, separando 70% dos dados para treinamento, 15% para validação e 15% para testes. A rede foi configurada com uma camada de entrada, duas camadas ocultas (10 neurônios), pois, segundo Jurado e col. (2015), com mais neurônios, o custo computacional aumenta significativamente e o desempenho não melhora, além de uma camada de saída (um neurônio).

Essa rede foi escolhida porque tal tipo de rede neural não precisa especificar a forma funcional entre a entrada e uma variável de saída, pois a rede tenta aprender os padrões dos dados históricos para mapear as variáveis para determinar a saída (KUSTER et al., 2017; HONG e FAN, 2016).

Inicialmente, a previsão foi realizada com 26 variáveis de entrada, deixando a variável de carga energética GWh como saída e, posteriormente, de acordo com a taxa de contribuição das variáveis, vários testes de previsão foram realizados. No entanto, como declarado anteriormente na seção 2, é essencial selecionar recursos de entrada ideais removendo dados irrelevantes para facilitar análises futuras e melhorar a precisão da previsão de carga de energia (SHAFI et al., 2019; ABEDINIA et al., 2017; YANG et al., 2019).

### 4.4 RESULTADOS E IMPACTO DAS PRINCIPAIS VARIÁVEIS NAS PREVISÕES

No bloco f, são feitas previsões com 27 variáveis que apresentam os níveis e previsões de maior e menor contribuição. Previsão com a variável que mais contribui, nesse caso, energia armazenada (GWh), depois previsão com energia armazenada (GWh) e demanda instantânea (MW), energia armazenada (GWh), demanda Instantânea (MW) e demanda horária máxima (MWh/h), energia armazenada (GWh), demanda instantânea (MW), demanda horária máxima (MWh/h) e afluyente natural de energia bruta (MWmed), ou seja, adicionando a taxa de contribuição de cada variável como entrada à rede neural. Isso permite que se veja quais variáveis geram os menores erros MAPE, RMSE, MSE e MAD nos 31 dias de agosto de 2019.

## 5. RESULTADOS E DISCUSSÃO

A Tabela 26 mostra o resultado obtido com a nova abordagem utilizada com o PCA para seleção e taxa de contribuição das variáveis. É possível observar que as seis primeiras variáveis representam mais de 85% da importância dos dados.

A energia armazenada (GWh) provou ser a variável mais importante no modelo, pois representa 26,31% de importância e, embora existam várias tecnologias de armazenamento de energia, como armazenamento de energia da bateria, ar comprimido, bateria de fluxo, energia do volante, energia do hidrogênio. A capacidade de armazenamento de energia instalada na Europa e na Alemanha é quase inteiramente responsável por usinas hidrelétricas bombeadas (PUNYS et al., 2013; STENZEL e LINSSEN, 2016; MELIKOGLU, 2017).

Esse tipo de energia é tão importante que a Alemanha estuda, além de usinas hidrelétricas bombeadas convencionais, usando diferenças de altura devido à topografia natural, outras abordagens para usinas hidrelétricas bombeadas não convencionais, denominadas hidrelétricas bombeadas subterrâneas. (PICKARD, 2012; STENZEL e LINSSEN, 2016).

Mais de 75% da eletricidade no Brasil é gerada por usinas hidrelétricas e a Região Sudeste do Brasil consome 83,4% da energia gerada por essas usinas (ONS, 2019). As usinas hidrelétricas podem armazenar energia suficiente por longos períodos e, nos mercados não regulamentados de eletricidade, as unidades de armazenamento de energia podem ser consideradas como ativos do sistema usados pelo operador do sistema para ajudar a manter a segurança energética ou reduzir os custos operacionais (MELIKOGLU, 2017). Provavelmente, por esse motivo, essa variável é a mais importante nesse modelo.

Em relação à segunda variável mais importante (demanda instantânea) e a terceira variável mais importante (demanda horária máxima) selecionada, Hong e Fan (2016) descrevem que elas fazem parte de um dos cinco novos problemas apontados por Hong e Fan (2016) para previsão de eletricidade. Isso pode justificar o motivo pelo qual essas variáveis são importantes.

Além disso, pode-se notar que as variáveis das cinco usinas apresentam menor taxa de contribuição, representando 14,14%. No entanto, observe-se que a data hidrológica (afluente e efluente) tem taxa de contribuição superior à quantidade e volumes de água para cada usina.

Tabela 26 – Resultados da taxa de contribuição das variáveis

Ordem	Nome Variável	Taxa contribuição (normalizada)
1	Energia armazenada (GWh)	9,2165
2	Demanda instantânea (MW)	4,5369
3	Demanda horária máxima (MWh/h)	4,4680
4	Afluente Natural de Energia Bruta (MWmed)	4,2380
5	Afluente Natural de Energia Armazenável (MWmed)	3,8349
6	Carga de energia (MWmed)	3,7813
7	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Itaipu	1,3326
8	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Itaipu	1,3314
9	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Ilha Solteira	0,5742
10	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Ilha Solteira	0,5687
11	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Marimbondo	0,1900
12	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Marimbondo	0,1858
13	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Furnas	0,0934
14	Carga de energia (GWh)	0,0908
15	dados hidrológicos / nível de quantidade (m) – Furnas	0,0890
16	Dados hidrológicos / fluxo de efluentes (m <sup>3</sup> /s) – Furnas	0,0850
17	Dados hidrológicos / fluxo afluente (m <sup>3</sup> /s) – Três Marias	0,0705
18	Troca de Energia (MWmed)	0,0696
19	dados hidrológicos / nível da quantidade (m) – Três Marias	0,0658
20	Dados hidrológicos / fluxo de efluentes m <sup>3</sup> /s – Três Marias	0,0628
21	dados hidrológicos / nível da quantidade (m) – Marimbondo	0,0513
22	dados hidrológicos / nível da quantidade (m) – Ilha Solteira	0,0380
23	dados hidrológicos / nível da quantidade (m) – Itaipu	0,0256
24	Dados / volumes hidrológicos – Ilha Solteira	0,0077
25	Dados / volumes hidrológicos – Furnas	0,0076
26	Dados / volumes hidrológicos – Marimbondo	0,0070
27	Dados / volumes hidrológicos – Três Marias	0,0067
	<b>Total</b>	<b>35,0224</b>

A Tabela 27 mostra a medida de desempenho MAPE, RMSE, MSE e MAD das 31 previsões para agosto de 2019 por taxa de contribuição das variáveis. É possível observar que a linha marcada em negrito e itálico mostra que, com sete variáveis, seis entradas e uma saída (carga de energia GWh) geram-se melhores resultados quando comparada aos outros casos. Isso significa que apenas energia armazenada (GWh), demanda instantânea (MW), demanda horária máxima (MWh/h), afluente natural de energia bruta (MWmed), afluente natural de energia armazenável (MWmed) e carga de energia (MWmed) são suficientes para prever energia de carga (GWh).

Com apenas uma variável como entrada (energia armazenada (GWh), o erro de previsão atinge 6,2% do MAPE, enquanto com 26 variáveis como entrada, o erro atinge 8,04E-03 do MAPE, o que foi apontado por Platon et al. (2015), que afirmam que poucas variáveis no modelo não conseguem explicar a saída satisfatoriamente, enquanto muitas variáveis ou variáveis que não contribuem significativamente podem causar problemas de convergência, baixa precisão e maior esforço computacional do modelo.

A questão de reduzir o número de variáveis no modelo não é importante apenas porque é possível usar uma técnica para reduzir variáveis, mas também para os requisitos de memória para armazenamento dos dados e para a possibilidade de melhor compreensão dos dados (SIQUEIRA et al., 2018). Como observamos nas Tabelas 26 e 27, nem todas as variáveis ajudam a melhorar o modelo.

Apesar de fazer boas previsões sem taxa de contribuição das variáveis que mais contribuem no sistema, é possível observar na Figura 41 que o erro é menor na previsão com as variáveis que contribuem com mais de 85%, nesse caso, as seis primeiras variáveis que mais contribuem para o modelo de previsão. O erro médio MAPE e RMSE para as previsões de 31 dias de agosto é de 8,04E-03% e 2,17E-03 sem seleção de variável, enquanto as previsões com as variáveis mais contribuintes atingidas são 1,74E-03% e 4,48E-04.

**Figura 41 – Previsão de carga de energia sem seleção de variáveis e com 6 variáveis que mais contribuem no sistema**

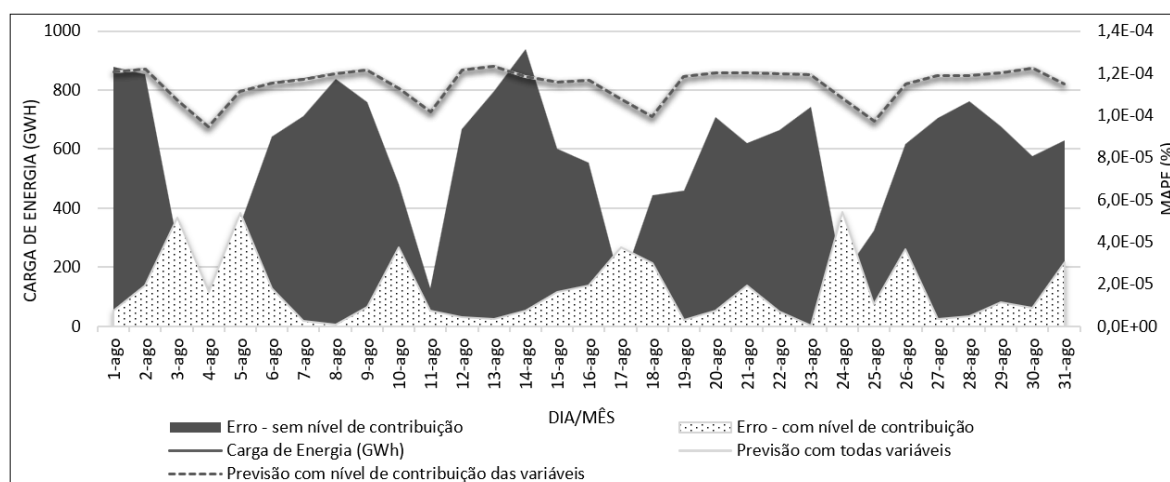


Tabela 27 – Comparação das medidas de desempenho

Número de variáveis de entrada	MAPE	RMSE	MSE	MAD	Média	Min	Max	Desvio Padrão	Variância
Real (GWh)	-	-	-	-	820,39	676,00	882,00	55,42	3071,11
1	6,20E+00	1,60E+00	2,55E+00	4,95E+01	807,29	782,09	850,83	26,12	682,12
2	2,55E+00	6,41E-01	4,11E-01	1,99E+01	831,00	733,67	879,07	40,16	1612,90
3	2,13E+00	5,37E-01	2,88E-01	1,66E+01	826,04	731,99	874,17	39,04	1524,00
4	2,62E+00	6,59E-01	4,34E-01	2,04E+01	839,31	726,50	883,91	41,63	1733,09
5	2,67E+00	6,72E-01	4,51E-01	2,08E+01	836,84	729,08	888,48	40,56	1645,42
6	<b>1,74E-03</b>	<b>4,48E-04</b>	<b>2,01E-07</b>	<b>1,39E-02</b>	<b>820,37</b>	<b>676,44</b>	<b>881,94</b>	<b>55,40</b>	<b>3069,39</b>
7	5,34E-03	1,44E-03	2,08E-06	4,47E-02	820,35	676,40	881,84	55,37	3066,07
8	2,57E-03	6,66E-04	4,44E-07	2,07E-02	820,38	676,48	881,89	55,39	3068,39
9	6,38E-03	1,68E-03	2,84E-06	5,22E-02	820,33	676,45	881,94	55,39	3068,03
10	2,32E-03	6,22E-04	3,87E-07	1,93E-02	820,38	676,43	881,88	55,39	3067,52
11	3,89E-03	9,89E-04	9,79E-07	3,07E-02	820,41	676,50	881,96	55,38	3066,64
12	5,53E-03	1,45E-03	2,11E-06	4,50E-02	820,34	676,44	881,93	55,40	3068,85
13	3,41E-03	8,85E-04	7,83E-07	2,74E-02	820,36	676,41	881,88	55,40	3069,03
14	2,85E-03	7,37E-04	5,43E-07	2,28E-02	820,38	676,45	881,89	55,40	3069,57
15	3,53E-03	8,97E-04	8,05E-07	2,78E-02	820,37	676,40	881,88	55,41	3070,00
16	7,27E-03	1,84E-03	3,38E-06	5,70E-02	820,37	676,21	881,91	55,39	3067,60
17	1,98E-03	5,20E-04	2,70E-07	1,61E-02	820,38	676,43	881,94	55,39	3068,52
18	5,74E-03	1,52E-03	2,31E-06	4,71E-02	820,34	676,44	881,85	55,39	3067,81
19	8,56E-03	2,29E-03	5,23E-06	7,09E-02	820,31	676,38	881,89	55,38	3066,87
20	3,42E-03	8,72E-04	7,60E-07	2,70E-02	820,38	676,39	881,92	55,41	3070,68
21	3,86E-03	9,81E-04	9,62E-07	3,04E-02	820,37	676,41	881,91	55,42	3070,86
22	1,42E-02	3,81E-03	1,45E-05	1,18E-01	820,28	676,45	881,93	55,33	3061,89
23	2,83E-03	7,44E-04	5,53E-07	2,31E-02	820,37	676,48	881,87	55,39	3067,84
24	6,52E-03	1,69E-03	2,84E-06	5,23E-02	820,39	676,39	881,97	55,44	3074,06
25	6,62E-03	1,73E-03	2,99E-06	5,36E-02	820,36	676,48	881,98	55,40	3069,68
26	8,04E-03	2,17E-03	4,73E-06	6,74E-02	820,32	676,44	881,84	55,36	3064,29

A Tabela 28 compara alguns estudos à abordagem utilizada neste trabalho. Em Yang et al. (2019) por exemplo, é utilizada a seleção de recursos de autocorrelação e os mínimos quadrados suportam os parâmetros de otimização da máquina vetorial pelo algoritmo do lobo cinza e pela validação cruzada alcançada na qual obteve 0,77% do MAPE. O trabalho de Liang et al. (2019), usando redundância mínima relevância máxima em que tinha um horizonte de previsão por hora, obteve erro MAPE de 0,8093.

Enquanto o artigo de Keles et al. (2016), utilizando o Método de Informação Mútua, alcançou 4,52 do RMSE. No entanto, o uso do PCA para taxa de seleção e contribuição e uma rede MLP simples com algoritmo *backpropagation* para previsão de carga de energia alcançaram 1,74E-03% MAPE e 4,48E-04 RMSE.

**Tabela 28 – Técnica para seleção de variáveis e previsão**

<b>Autores</b>	<b>Técnica para seleção de variáveis</b>	<b>Horizonte de previsão</b>	<b>MAPE (%)</b>	<b>RMSE</b>
Platon et al (2015)	Raciocínio baseado em casos, redes neurais artificiais e análise de componentes principais	Hora	-	<7,3
Hong (2010)	Rede neural artificial	Ano	4.35	-
Jurado et al. (2015)	Raciocínio Indutivo Difuso	Hora	4.49	-
Keles et al. (2015)	Método de informação mútua	Mês	-	4.52
Lang et al. (2019)	Rede neural ponderada com pesos aleatórios e núcleos	Dia	1.32	12.85
Wang e Hong (2016)	Efeito recente para previsão de carga elétrica	Hora	3.39	-
Nowotarski et al. (2016)	Previsões de carga irmã	Hora	2.14	-
Liang et al. (2019)	Redundância mínima relevância máxima	Hora	0.8093	9.58
Lahouar e Slama (2015)	Floresta aleatória	Hora	< 2.3	-
Yang et al. (2019)	A seleção de recursos de autocorrelação e os mínimos quadrados suportam os parâmetros de otimização da máquina vetorial pelo algoritmo de lobo cinza e validação cruzada	Os dados de carregamento são amostrados a cada meia hora	0.77	-
Bantugon e Gallano (2016)	Rede neural e Holter Winter	Dia	5.64	2209.8
Ghadimi et al. (2019)	Técnica de seleção de recursos e meta-heurística melhorada	Semana	1.3	-
Este artigo	PCA para taxa de seleção e contribuição	Dia	1.74E-03	4.48E-04
	Sem taxa de seleção		8.04E-03	2.17E-03

Além da seleção de variáveis e taxa de contribuição, este estudo também treinou dados com apenas as variáveis consideradas mais importantes. Isso poderia ter sido aplicado no trabalho Bannor Boateng e Acheampong (2019), pois a seleção de variáveis pode identificar um subconjunto das informações mais relevantes no sistema e, portanto, influenciar a precisão

da previsão de energia (YANG et al., 2019; PERES e FOGLIATTO, 2018; ABEDINIA et al., 2017; SHAFI et al., 2019; LI et al., 2019).

## 6. OBSERVAÇÕES FINAIS

Este artigo propõe uma abordagem baseada na Análise de Componentes Principais (PCA) para definir uma taxa de contribuição para cada variável e, em seguida, selecionar as principais variáveis como entrada para uma rede neural para previsão de carga de energia na Região Sudeste do Brasil. A previsão da carga de energia foi feita a partir de variáveis com maior taxa de contribuição, atingindo os erros de previsão MAPE, RMSE, MSE e MDE, de 1,74E-03%, 4,48E-04, 2,01E-07 e 1,39E-02, respectivamente, para os 31 dias de agosto de 2019.

Devido a essa precisão de previsão, essa abordagem pode ser usada para mercados de eletricidade desregulados, uma vez que a quantidade de energia que a concessionária tem para comprar ou vender no mercado em tempo real a preços desfavoráveis depende do erro de previsão e ajuda as usinas como contratos para compra e venda de eletricidade (DUDEK, 2016; ALAM, 2019). Além disso, a abordagem pode ser útil para usinas hidrelétricas, pois elas precisam ajustar contratos de compra e venda de eletricidade para melhorar a maximização do lucro (DUDEK, 2016; ALAM, 2019).

Além de melhorar a precisão da previsão e fornecer um preditor mais rápido, a abordagem baseada em PCA proposta para calcular a taxa de contribuição das variáveis de entrada fornece melhor compreensão do processo subjacente que gerou os dados, o que é fundamental para a realidade brasileira, devido ao clima acentuado e variações econômicas.

Embora a previsão de carga de energia seja uma área para a qual bancos de dados com dezenas ou centenas de variáveis estão disponíveis, a abordagem pode selecionar apenas seis variáveis que contribuem com mais de 85% para o modelo. Embora as taxas de contribuição das variáveis das usinas, mais a troca de energia adicionada, tenham apenas 14,14% de contribuição, a variável energia armazenada tem taxa de contribuição de 26,31%, sendo fundamental para a precisão da previsão. Assim, as usinas podem prestar mais atenção a essa variável, uma vez que é importante nas hidrelétricas, como apontado por Punys et al. (2013) e Stenzel e Linssen (2016).

Como trabalho futuro, essa metodologia pode ser usada em diferentes tipos de dados para comparar o número de variáveis, tempo e área sob a curva operacional do receptor, como em Speiser et al. (2019). Outra ideia são as variáveis de entrada de atraso, de acordo com o

trabalho de Alam (2019), mas usando as seis variáveis selecionadas com essa abordagem para ver quais atrasos geram menos erros e fornecem às usinas e distribuidores de energia taxa de confiança mais alta para gerar e distribuir eletricidade.

## REFERÊNCIAS

ABEDINIA, O., AMJADY, N., ZAREIPOUR, H. (2016), “A new feature selection technique for load and price forecast of electrical power systems”. **IEEE Transactions on Power Systems**, 32. 1–1. 10.1109/TPWRS.2016.2556620.

ALKARKHI, A. F. M. (2019), “**Principal Components Analysis**”. Chapter 8.

ALAM, M. (2019), “Day-Ahead Electricity Price Forecasting and Scheduling of Energy Storage in LMP Market”. **IEEE Access**. PP. 1-1. 10.1109/ACCESS.2019.2952451.

ANEEL (2016). Normative Resolution No. 703, from March 28, 2016, “**Defines the structure of the tax regulation procedures – PRORET, which consolidates the financial components of the distribution fares – Overcontracting of energy and exposure to the short-term market**”. Available at: <[http://www2.aneel.gov.br/cedoc/aren2016703\\_Proret\\_Submod\\_4\\_3\\_V0.pdf](http://www2.aneel.gov.br/cedoc/aren2016703_Proret_Submod_4_3_V0.pdf)>. Accessed Jun. 2, 2019.

BANNOR, B., ACHEAMPONG, A. (2019), “Deploying artificial neural networks for modeling energy demand: international evidence”. **International Journal of Energy Sector Management**. ahead-of-print. 10.1108/IJESM-06-2019-0008.

BARTLETT, M. S. (1937), “Properties of sufficiency and statistical tests”. **Proceedings of the Royal Statistical Society - Serie A**, 60:268-282, 1937.

BOX, G., TIAO, G. (1977), “**A canonical analysis of multiple time series**”, *Biometrika*, 64, 355–365.

CERNY, C.A., KAISER, H.F. (1977), “**A study of a measure of sampling adequacy for factor-analytic correlation matrices**”. *Multivariate Behavioral Research*, 12(1), 43-47.

CHINNATHAMBI, Radhakrishnan., CAMPION, Mitch., NAIR, Arun., PRAKASH, Ranganathan. (2018), “Investigation of Price-Feature Selection Algorithms for the Day-Ahead Electricity Markets”. 1-6. 10.1109/EPEC.2018.8598282.

- DUDEK, G. (2016), "Pattern-based local linear regression models for short-term load forecasting". **Electric Power Systems Research**. <http://dx.doi.org/10.1016/j.epsr.2015.09.001>.
- GHADIMI, N., AKBARIMAJD, A., SHAYEGHI, H., ABEDINIA, O. (2018). "Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting". **Energy**. 161. 10.1016/j.energy.2018.07.088.
- GHOSH, K., RAMTEKE, M., SRINIVASAN, R. (2014), "Optimal variable selection for effective statistical process monitoring". **Computers & Chemical Engineering**, 60, 260–276.
- HAYKIN, S. (2008), "**Neural networks: A comprehensive foundation**". Toronto: Pearson.
- HONG, S. F. (2016), "Probabilistic electric load forecasting: A tutorial review". **International Journal of Forecasting**, 32(3), 914–938.
- HONG, T. (2010), "**Short term electric load forecasting**". North Carolina State University (2010).
- IEA. International Energy Agency. "**Global Engagement**". Available in <<https://www.iea.org/countries/Brazil/>>. Accessed in: May. 2019.
- JAHANGOSHAI REZAEI, M., DADKHAH, M., FALAHINIA, M. (2019), "Integrating neuro-fuzzy system and evolutionary optimization algorithms for short-term power generation forecasting". **International Journal of Energy Sector Management**. 10.1108/IJESM-09-2018-0015.
- JOLLIFFE, I. (2002), "**Principal component analysis**". Second edition. Springer.
- JURADO, S., NEBOT, A., MUGICA, F., AVELLANA, N. (2015), "Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques". **Energy**, 86, 276–291, 10.1016/j.energy.2015.04.039.
- KUSTER, C., REZGUI, Y., MOURSHED, M. (2017), "Electrical load forecasting models: A critical systematic review". **Sustainable Cities and Society**. 35. 10.1016/j.scs.2017.08.009.
- KELES, D., SCALLE, J., PARASCHIV, F., FICHTNER, W. (2016), "Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks" **Applied Energy**, Elsevier, vol. 162(C), pages 218-230.

- KIM, J. -Y., CHO, S. -B. (2019), “Electric energy consumption prediction by deep learning with state explainable autoencoder”. **Energies**, 12, 739.
- LAHOUAR, J. B. H. Slama (2015), “Day-ahead load forecast using random forest and expert input selection”. **Energy Conversion Management**, 103, 1040–1051.
- LANG, K., ZHANG, M., YUAN, Y., YUE, X. (2018), “Short-term load forecasting based on multivariate time series prediction and weighted neural network with random weights and kernels”. **Cluster Computing**, 10.1007/s10586-017-1685-7.
- LI, H. (2016), “Accurate and efficient classification based on common principal components analysis for multivariate time series”. **Neurocomputing**, 171, 744–753.
- LI, H. (2019), “Multivariate time series clustering based on common principal component analysis”. *Neurocomputing*. 10.1016/j.neucom.2019.03.060.
- LI, H., LIN, C., WAN, X., LI, Z. (2019), “Feature representation and similarity measure based on covariance sequence for multivariate time series”. **IEEE Access**, PP. 1-1. 10.1109/ACCESS.2019.2915602.
- LIANG, Y., NIU, D., HONG, W. C. (2019), “Short term load forecasting based on feature extraction and improved general regression neural network model”. **Energy**, 166, 653–663.
- MELIKOGLU, M. (2017), “Pumped hydroelectric energy storage: Analysing global development and assessing potential applications in Turkey based on Vision 2023 hydroelectricity wind and solar energy targets”. **Renewable and Sustainable Energy Reviews**, 72, 146-153.
- NOWOTARSKI, Liu B., WERON, R., HONG, T. (2016), “Improving short term load forecast accuracy via combining sister forecasts”. *Energy*, 98, 40–49.
- ONS. "**Operador Nacional do Sistema Elétrico. Resultados da Operação: “Histórico da Operação”**". Available in <<http://www.ons.org.br/paginas/resultados-da-operacao/historico-da-operacao>>. Accessed in: November. 2019.
- PEARSON, K. (1901), “On lines and planes of closest fit to systems of points in space”. **Philosophical Magazine**, 2 (6), 559–572.

PEÑA, D., PONCELA, P. (2006), “**Dimension reduction in multivariate time series**”. In: Balakrishnan N., Sarabia J. M., Castillo E. (eds) *Advances in distribution theory, order statistics, and inference. Statistics for industry and technology*. Birkhäuser Boston.

PERES, F., FOGLIATTO, F. (2018), “Variable selection methods in multivariate statistical process control: a systematic literature review”. **Computers in Industry and Engineering**, 115, 603–619.

PICKARD, W. F. (2012), “The history, present state, and future prospects of underground pumped hydro for massive energy storage”. **Proc IEEE**, 100, 473–83.

PLATON, Radu., RAISSI, Dehkordi., Vahid., MARTEL, Jacques. (2015), “Hourly Prediction of a Building’s Electricity Consumption Using Case-based Reasoning, Artificial Neural Networks and Principal Component Analysis”. **Energy and Buildings**. 92. 10-18. 10.1016/j.enbuild.2015.01.047.

PUNYS, P., BAUBLYS, R., KASIULIS, E., VAISVILA, A, PELIKAN, B., Steller J et al. (2013), “Assessment of renewable electricity generation by pumped storage power plants in EU Member States”. **Renewable and Sustainable Energy Reviews**, 26, 190–200.

REINSEL, G. C., VELU, R. P. (1998), “**Multivariate reduced-rank regression**”. Springer, New York.

SCHIMIT, P. H., PEREIRA, F. H. (2018), “Disease spreading in complex networks: A numerical study with Principal Component Analysis”. **Expert Systems with Applications**, 97, 41-50.

SHAFI, I., JAVAID, N., NAZ, A., AMIR, Y., ISHAQ, I., NASEEM, K. (2019), “Feature selection and extraction along with electricity price forecasting using *big data* analytics”. In: Barolli, L., Xhafa, F., Javaid, N. & Enokido, T. (eds) *Innovative mobile and internet services in ubiquitous computing. IMIS 2018. Advances in Intelligent Systems and Computing*, vol 773. Springer, Cham.

SIQUEIRA, Hugo., BOCCATO, Levy., LUNA, Ivette., ATTUX, Romis., LYRA, Christiano. (2018), “Performance Analysis of Unorganized machines in streamflow forecasting of Brazilian plants”. **Applied Soft Computing**. 68. 10.1016/j.asoc.2018.04.007.

SPEISER, J. L., MILLER, M. E., TOOZE, J., IP, E. (2019), “A comparison of random forest variable selection methods for classification prediction modelling”. **Expert Systems with Applications**, doi: <https://doi.org/10.1016/j.eswa.2019.05.028>.

STENZEL, Peter., LINSSEN, Jochen. (2016), “Concept and potential of pumped hydro storage in federal waterways”. **Applied Energy**. 162. 486-493. 10.1016/j.apenergy.2015.10.033.

TAO, H. (2014), “Energy Forecasting: Past, Present, and Future. Foresight”: The **International Journal of Applied Forecasting**, 32, 43–48.

TIAO, G. C., TSAY, R. S. (1989), “**Model specification in multivariate time series**”. Journal of the Royal Statistical Society, Series B, 51, 157–213.

Wang, P., LIU, B., HONG, T. (2016), “Electric load forecasting with recency effect: a *big data* approach”. **International Journal of Forecasting**, 32 (3), 585–597.

YANG, A., LI, W., YANG, X. (2019), “Short-term electricity load forecasting based on feature selection and least squares support vector machines”. **Knowledge-Based Systems**, 163, 159–173.

## **APÊNDICE D (Artigo 4) – Modelos de redes neurais autorregressivas não lineares para previsão de concentrações de gases dissolvidos em óleo em transformadores**

*Energies – Open Access*

Qualis: A2

*Manuscript ID: Energies 2018, 11, 1691; doi:10.3390/en11071691*

*Title: Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations.*

*Corresponding author:* Dr. Fabio Henrique Pereira

*Corresponding Author's Institution:* Universidade Nove de Julho, Universidade de São Paulo.

*First Author:* Fabio Henrique Pereira.

*Order of Authors:* Fabio Henrique Pereira.

*Order of Authors:* Fabio Henrique Pereira<sup>1,2,3,\*</sup>, Francisco Elânio Bezerra<sup>2</sup>, Shigueru Junior<sup>3</sup>, Josemir Santos<sup>3</sup>, Ivan Chabu<sup>3</sup>, Gilberto Francisco Martha de Souza<sup>3</sup>, Silvio Ikuyo Nabeta<sup>3</sup> e Fábio Micerino.<sup>4</sup>

**Resumo:** Os transformadores são uma das partes mais importantes no sistema de energia, especialmente em instalações-chave, em que devem ser monitorados de perto e continuamente. Nesse contexto, métodos baseados nas razões de gás dissolvido permitem associar valores de concentrações de gás à ocorrência de algumas falhas, como descargas parciais e falhas térmicas. Portanto, a previsão acurada das concentrações de gás dissolvido em óleo é ferramenta valiosa para monitorar a condição do transformador e desenvolver o sistema de diagnóstico de falhas. Este estudo propõe um modelo de rede neural autorregressiva não linear com a transformada *wavelet* discreta para prever as concentrações de gás dissolvido em óleo do transformador. O ajuste de dados e a capacidade de previsão acurada do modelo proposto são avaliados em exemplo do mundo real, mostrando resultados melhores em relação aos modelos de previsão atuais e às técnicas comuns de séries temporais.

**Palavras-chave:** gás dissolvido em óleo; rede neural autorregressiva não linear; sistema de diagnóstico de falhas.

*Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations*

**Abstract:** *Transformers are one of the most important part in a power system and, especially in key-facilities, they should be closely and continuously monitored. In this context, methods based on the dissolved gas ratios allow to associate values of gas concentrations with the occurrence of some faults, such as partial discharges and thermal faults. So, an accurate prediction of oil-dissolved gas concentrations is a valuable tool to monitor the transformer condition and to develop a fault diagnosis system. This study proposes a nonlinear autoregressive neural network model coupled with the discrete wavelet transform for predicting transformer oil-dissolved gas concentrations. The data fitting and accurate prediction ability of the proposed model is evaluated in a real-world example, showing better results in relation to current prediction models and common time series techniques.*

**Keywords:** *oil-dissolved gas; nonlinear autoregressive neural network; fault diagnose system.*

## 1. INTRODUÇÃO

Como o transformador é uma das unidades mais importantes do sistema elétrico, é natural que sejam feitos esforços para preservar sua integridade e aumentar sua disponibilidade (GODINA et al., 2015). Para esses fins, as políticas e procedimentos de manutenção são planejadas e aplicadas visando garantir a menor interrupção de tais equipamentos (GODINA et al., 2015; PENG et al., 2018). De fato, qualquer falha nesse equipamento pode afetar toda a rede, comprometer outros elementos dela e gerar impactos econômicos significativos (GODINA et al., 2015; WANG et al., 2018).

Especialmente no caso de transformadores a óleo, as operações de manutenção devem ser realizadas com cuidado adicional, para minimizar o potencial problema de inflamabilidade do material de isolamento térmico (CHENG e YU, 2018). Devido a sua complexidade e importância, os problemas do grau de envelhecimento do isolamento do papel têm sido objeto de estudo em muitos trabalhos recentes (WANG et al., 2018; CHENG e YU, 2018). Vários outros testes de itens de isolamento têm sido parte importante dos sistemas de diagnóstico de falhas dos transformadores, com ênfase na análise cromatográfica de gás dissolvido em óleo, ou seja, análise de gás dissolvido (DGA) (CHENG e YU, 2018; FAIZ e SOLEIMANI, 2017).

Nesse contexto, métodos baseados nas razões de gás dissolvido permitem que valores de concentrações de gás sejam associados à ocorrência de algumas falhas, como descargas parciais e falhas térmicas. As falhas dos transformadores de energia, geralmente, levam à

degradação dos materiais isolantes, o que resulta na liberação de certos gases que são dissolvidos no óleo.

A partir de certo nível de concentração, esses gases atuam como isolantes térmicos, provocando superaquecimento do equipamento e, simultaneamente, diminuindo o vigor dielétrico do óleo, o que pode causar falha no isolamento elétrico. Por outro lado, o superaquecimento do óleo aumenta os níveis de alguns gases como metano e etileno, por exemplo. Portanto, a previsão acurada das concentrações de gás dissolvido em óleo é ferramenta valiosa para monitorar a condição do transformador e desenvolver um sistema de diagnóstico de falhas (LIN et al., 2018).

Nos últimos anos, a análise de gases dissolvidos no óleo dos transformadores, com base nas diretrizes da Comissão Eletrotécnica Internacional (IEC) IEC 60599 (1978) e do Instituto de Engenheiros Elétricos e Eletrônicos (IEEE) IEEE (1991), tornou-se prática generalizada seguida por todas as empresas que trabalham com energia elétrica. Nesse contexto, merece destaque o uso de técnicas de inteligência artificial combinadas à análise cromatográfica de gases dissolvidos em óleo (DGA). Em geral, abordagens inteligentes foram propostas para contornar as limitações de métodos puramente tradicionais baseados em DGA, com ênfase em abordagens de redes neurais artificiais, incluindo rede neural de regressão generalizada (LIN et al. 2018), máquina de vetores de suporte (SVM) (BACHA et al., 2012; YONGBO e JUAN, 2017; ZHANG et al., 2017), sistemas especialistas (EPS) (WANG et al. 2000) e sistemas difusos (ISLAM et al., 2000).

Uma pesquisa recente de Cheng e Yu (2018) mostra que o uso dessas técnicas produziu resultados promissores no desenvolvimento de sistemas de diagnóstico de falhas de alta precisão. No entanto, resultados igualmente recentes indicam que essas técnicas ainda apresentam limitações na previsão de gases dissolvidos em óleo, deixando espaço para melhorias adicionais.

Em relação aos sistemas especialistas, por exemplo, uma simulação precisa da experiência, habilidade e processo de raciocínio dos especialistas depende fortemente da qualidade da base de conhecimento estabelecida, que é uma das principais limitações dessa abordagem na maioria dos casos. Em geral, a base de conhecimento dificilmente considera todos os casos possíveis, o que leva a erros na identificação de sintomas de falhas não presentes no banco de dados (CHENG e YU, 2018; CHEN et al., 2012).

Em relação às redes neurais, a ideia básica é mapear uma relação de entrada e saída altamente não linear e, a partir dessa relação, produzir a conclusão de diagnóstico sobre a falha (CHENG e YU, 2018). Apesar dos resultados satisfatórios, essa abordagem tradicional não é

capaz de prever valores de vários passos à frente e, principalmente, o desempenho é influenciado pelos dados de entrada e é limitado por amostras e parâmetros de treinamento.

Em Lin et al. (2018), por exemplo, os autores propõem o uso da análise de componentes principais (PCA) para melhorar a precisão da previsão, selecionando as entradas mais representativas para o treinamento em rede. De fato, o uso de PCA nesse contexto é generalizado (LIN et al., 2018; DIAS e PEREIRA, 2018; SCHIMIT e PEREIRA, 2018). Além disso, a aplicação usual de modelos não recorrentes, como a Rede Neural de Regressão Generalizada (GRNN), realiza apenas mapeamento não linear entre entradas e saídas e não é apropriada para estimativas futuras de valores de gás dissolvido em óleo.

Outra dificuldade importante nas abordagens de redes neurais é o ajuste dos parâmetros de treinamento. Os modelos GRNN, por exemplo, são fortemente dependentes de parâmetros suavizados. Lin et al. (2018) superaram essa limitação aplicando um método de otimização apenas para selecionar um valor adequado do fator suavizado. Essa abordagem tem a limitação de precisar recalcular o fator suave sempre que os dados são atualizados. Esses autores também aplicam a análise de componentes principais para reduzir a influência dos dados de entrada no modelo.

Apesar de produzir resultados relativamente precisos, com boa capacidade de generalização e pouco ajuste excessivo, o SVM em problemas de regressão também é fortemente influenciado pela qualidade dos vetores de entrada (ZHANG et al., 2017).

Por outro lado, apesar de abordar algumas questões relacionadas à incerteza de dados, o uso da lógica nebulosa aumenta a dependência do conhecimento especializado para criar um conjunto de regras nebulosas, que descrevem os relacionamentos entre variáveis de entrada e saída. Muitos autores propõem a criação de um conjunto de regras baseadas em normas (IEC 599, IEEE), que claramente não é eficiente, uma vez que combinações de diferentes proporções de gases contempladas pela norma podem não ocorrer na prática, levando a sério problema de indecisão ou não decisão no diagnóstico.

Considerando essa situação, propomos aqui a combinação do modelo de rede neural autorregressivo não linear à transformada de *wavelet* discreta, para prever concentrações de gás dissolvido e razão de concentração de gás no óleo do transformador. O objetivo dessa abordagem é criar um modelo que seja invariável em relação ao parâmetro de atraso de tempo e permita a previsão menos sensível às dependências de longo prazo, além de apresentar melhores capacidades de generalização e aprendizado, resultando em previsão de alta precisão com várias etapas à frente das concentrações de gás no óleo.

A transformação *wavelet* localiza recursos nos dados de entrada e concentra seus recursos em alguns coeficientes de *wavelet* sem afetar a qualidade dos dados (JENSEN e LA COUR-HARBO, 2001). Como resultado, temos um conjunto de dados de entrada de complexidade simplificada que leva à previsão de alta precisão, sem nenhuma intervenção humana. Portanto, a hipótese é que o uso de funções de *wavelets* para criar versões esparsas dos dados iniciais pode aumentar a precisão da previsão do modelo, aumentando a confiança nas previsões com várias etapas à frente e reduzindo o efeito do parâmetro de atraso em diagnóstico de falha de alta precisão.

Como o modelo proposto pode prever valores futuros das concentrações de gás dissolvido em óleo, ele é aplicado em um sistema de monitoramento de condição do transformador, em conjunto a técnicas de confiabilidade, para fornecer diagnóstico precoce de possíveis falhas e estimar a vida útil restante do transformador em dados históricos e eventos acumulados durante determinado período.

Deve-se notar que o modelo proposto não se destina a produzir conclusão sobre o diagnóstico de falha do transformador, mas previsão de alta precisão de concentrações e proporções de gás em momentos futuros, permitindo ao operador antecipar possíveis falhas e prosseguir com as medidas de proteção recomendadas. Assim, as principais contribuições do artigo são as seguintes.

O desenvolvimento de um modelo de previsão de gases baseado na combinação de funções *wavelet* com uma rede autorregressiva não linear, insensível ao parâmetro de atraso e ao tipo de função *wavelet*; previsões de alta precisão de valores futuros das concentrações e da proporção de gases dissolvidos em óleo, contribuindo para aumentar a confiabilidade do sistema de monitoramento para antecipar possíveis falhas.

Um sistema de alerta que monitora valores futuros de concentrações de gases e permite antecipar situações anormais e executar medidas de proteção apropriadas.

O ajuste de dados e a capacidade de previsão precisa do modelo proposto são avaliados em um exemplo do mundo real, mostrando melhores resultados em relação a vários modelos de previsão atuais e técnicas comuns de séries temporais.

## 2. TEORIA RELACIONADA

### 2.1 REDE NEURAL ARTIFICIAL

A Rede Neural Artificial (RNA) é método matemático que visa simular o cérebro humano no processo de aquisição de conhecimento, com aplicações bem-sucedidas no mapeamento não linear entre variáveis de entrada e saída, reconhecimento e classificação de padrões, otimização, apenas para citar alguns (HAYKIN, 1999; BANOOKH e BARAKATI, 2012).

Como uma espécie de modelo de biologia do cérebro humano, a RNA configura alguns componentes que definem propriedades essenciais do neurônio biológico, como sinapses, pesos neurais e funções de transferência (HAYKIN, 1999).

Os neurônios artificiais são os elementos de processamento, organizados em sucessivas camadas interconectadas que recebem as informações (variáveis de entrada) e as propagam em direção à camada de saída. A informação recebida é ponderada por um peso sináptico que determina se o próximo neurônio será ativado pela função de ativação, geralmente a função sigmoideal em problemas não lineares.

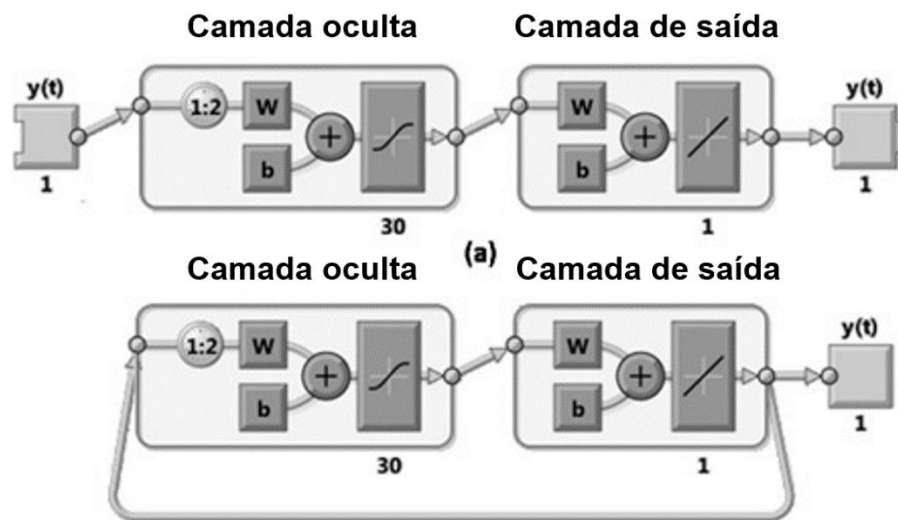
O aprendizado em rede ocorre quando os pesos são ajustados ao longo das camadas, de acordo com a relação entre as entradas e as saídas desejadas. Um dos modelos mais básicos é o *Perceptron* Multicamadas (MLP), amplamente utilizado na aproximação de funções não lineares que descrevem relacionamentos complexos entre variáveis independentes e dependentes em muitas aplicações.

#### 2.1. Rede neural autorregressiva não linear

A rede neural autorregressiva não linear é um tipo de RNA apropriada para estimar valores futuros da variável de entrada. A rede NAR permite a previsão de valores futuros de uma série temporal, apoiada em seu histórico, por meio de mecanismo de realimentação, no qual um valor previsto pode servir como entrada para novas previsões em momentos mais avançados (ISLAM e MORIMOTO, 2017; CASWELL, 2014).

A rede é criada e treinada em circuito aberto, usando os valores-alvo reais como *feedback* e, assim, garantindo maior precisão no treinamento. Após o treinamento, a rede é convertida em circuito fechado e os valores previstos são usados para fornecer novas entradas de *feedback* à rede. A arquitetura do *loop* aberto e fechado é ilustrada na Figura 42.

Figura 42 – Arquitetura do circuito aberto (a) e ciclo fechado (b) da rede neural autorregressiva não-linear



Matematicamente, o modelo prevê valores futuros da série temporal  $y(t)$ , com base em seus valores históricos  $y_{t-1}, y_{t-2}, \dots, y_{t-d}$ , (modelo NAR) utilizando adicionalmente uma série temporal externa  $x_{t-1}, x_{t-2}, \dots, x_{t-d}$  (modelo NARX), em que “d” é o parâmetro de atraso de tempo.

O treinamento em rede é realizado, geralmente, pelo algoritmo *backpropagation* e utiliza o método de descida mais íngreme para minimizar o erro ao quadrado entre os valores reais e os previstos.

## 2.2. TRANSFORMADA DISCRETA DE WAVELET

A transformada *wavelet* é uma técnica de processamento de sinal, amplamente aplicada e desenvolvida em muitas áreas diferentes (PEREIRA et al., 2015; WANG et al., 2018). Devido a seu sucesso em muitas aplicações, vários trabalhos tentam motivar e explicar as ideias básicas por trás das *wavelets* (SWELDENS, 1994). Nesse contexto, destacamos a técnica de levantamento que busca explorar correlações nos dados para construir uma aproximação esparsa no domínio espacial, o que o torna mais eficiente do que as abordagens baseadas no domínio da frequência (SWELDENS, 1994).

Matematicamente, a abordagem de elevação recebe um sinal não aleatório  $S_j$ , de comprimento  $2j$ , como ilustrado na Equação 45, e constrói a aproximação desse sinal escrevendo cada amostra ímpar como a média de duas amostras adjacentes. A precisão da

aproximação é definida por um conjunto de coeficientes de detalhes que são calculados como a diferença entre uma amostra ímpar e sua aproximação computada, conforme expresso na Equações 45 e 46.

$$S^j = \{s_1^j, s_2^j, s_3^j, \dots, s_j^j\} \quad (45)$$

$$d_k^{j-1} = \{s_{2k+1}^j - (s_{2k}^j + s_{2k+2}^j)/2\} \quad (46)$$

Uma operação adicional, chamada atualização, é realizada para preservar o valor médio do sinal original, de acordo com Equação 47.

$$s_k^{j-1} = \{s_{2k}^j + (d_{k-1}^{j-1} + d_k^{j-1})/4\} \quad (47)$$

### 3. O MODELO DE PREVISÃO PROPOSTO

A presença de pequena concentração de gases dissolvidos em óleo no transformador é consequência natural da operação normal do equipamento, devido ao campo elétrico, umidade e oxidação (LIN et al., 2018). No entanto, o aumento na concentração desses gases, que incluem hidrogênio (H<sub>2</sub>), monóxido de carbono (CO), dióxido de carbono (CO<sub>2</sub>), metano (CH<sub>4</sub>), acetileno (C<sub>2</sub>H<sub>2</sub>), etileno (C<sub>2</sub>H<sub>4</sub>) e etano (C<sub>2</sub>H<sub>6</sub>), pode estar relacionado à ocorrência de falhas e anormalidades. A elevação nas concentrações de metano e etileno, por exemplo, pode indicar a ocorrência de alguma falha térmica no transformador, enquanto variações no hidrogênio e acetileno são indicações de falhas elétricas (IEEE, 1991).

Assim sendo, a variação das concentrações de gás ao longo do tempo é questão crítica na análise de diagnóstico de falhas do transformador. Além disso, como alguns desses gases têm forte correlação em situação de falha, muitas taxas de concentração de gases também devem ser consideradas.

Portanto, o modelo de previsão proposto (NAR–DWT) é aplicado para prever valores futuros dos sete tipos de gás dissolvido em óleo e as razões IEC e Rogers (CH<sub>4</sub>/H<sub>2</sub>, C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>4</sub>/C<sub>2</sub>H<sub>6</sub>), de acordo com os seguintes passos:

Etapa 1: um conjunto de dados históricos de gás dissolvido em óleo é coletado de um transformador equipado com um *GE Kelman-Transfix* (GE – General Electric, São Paulo, Brasil) e *GE Intellix BMT 330* (GE – General Electric, São Paulo, Brasil).

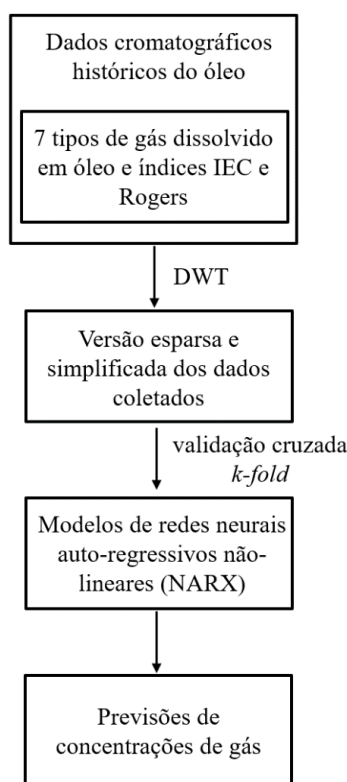
Etapa 2: o conjunto de dados coletados é avaliado usando a transformação de *wavelet* discreta para criar uma versão esparsa e simplificada com boas propriedades de aproximação.

Etapa 3: os modelos de redes neurais autorregressivas não lineares são treinados e validados de acordo com a abordagem de validação cruzada *k-fold*.

Etapa 4: um modelo de rede neural é criado para prever as concentrações e proporções de gases dissolvidos em óleo.

Um fluxograma do modelo de previsão proposto é apresentado na Figura 43.

**Figura 43 – Fluxograma do modelo de previsão autorregressiva não-linear proposto (NAR-DWT)**



### 3.1. Previsão de concentrações de gás no óleo

O modelo autorregressivo não linear foi aplicado e a acurácia da previsão é avaliada usando o desempenho do erro quadrático médio entre o valor-alvo e os valores previstos.

Uma função de treinamento foi aplicada com base na regularização bayesiana, divisão de dados aleatórios com 80% para treinamento e 20% para testes. Também aplicamos validação cruzada *k-fold*, na qual os dados foram divididos em 10 subconjuntos e o treinamento repetido 10 vezes, usando um dos 10 subconjuntos de cada vez para teste e validação, enquanto os outros

9 subconjuntos formavam um treinamento conjunto único. A estimativa de erro é calculada como média em todas as 10 tentativas.

As *wavelets Symlets* e *Daubechies* são aplicadas para criar uma versão esparsa e simplificada para os dados de concentrações e proporções de gases, respectivamente.

Para permitir uma comparação em relação à precisão e validade da previsão, adotamos os mesmos critérios de avaliação de (GODINA et al., 2015):

a) o erro percentual relativo entre os valores alvo e preditos (*avg\_err*), Equação 48;

$$avg\_err = \frac{1}{N} \sum_{i=1}^N \left| \frac{\tilde{x}_i - x_i}{x_i} \right| \times 100 \quad (48)$$

b) o erro relativo máximo (*max\_err*), Equação 49;

$$max\_err = \max \left| \frac{\tilde{x}_i - x_i}{x_i} \right| \quad (49)$$

Em que  $N$  é o número de amostras de dados,  $x_i$  e  $\tilde{x}$  são o destino e o valor previsto, respectivamente.

O modelo proposto é avaliado no exemplo do mundo real, usando um conjunto de dados de concentração de gás dissolvido em óleo de um transformador em uma subestação de 13,8-230 kV, 190 MVA localizada no Brasil. O dispositivo foi equipado com um *GE Kelman-Transfix*, que é equipado com tecnologia de detecção fotoacústica para medir as concentrações de gás. O conjunto de dados é composto por sete meses de observações diárias, realizadas no período de novembro de 2016 a maio de 2017, correspondentes a 176 amostras dos gases  $H_2$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  e  $C_2H_6$ .

#### 4. RESULTADOS NUMÉRICOS

Primeiramente, avaliamos o efeito do parâmetro de atraso de tempo “d” no desempenho do processo de treinamento, avaliado usando o erro quadrado médio (MSE) e o coeficiente de determinação R, que é a medida de qualidade do ajuste para regressão linear entre o alvo e as previsões. Nesse caso, definimos “d” de 2 a 10, com a etapa 1, e o erro foi calculado como

média em todas as 10 tentativas, de acordo com a abordagem de validação cruzada descrita acima.

A Tabela 29 ilustra os resultados dessa avaliação usando a concentração de CO<sub>2</sub> como exemplo. O modelo NAR–DWT apresenta um ajuste muito preciso e pequeno valor, independentemente do valor de “d”. Os resultados para os outros gases são semelhantes.

**Tabela 29 – Efeito do parâmetro  $d$  no desempenho do processo de treinamento para gás CO<sub>2</sub>.**

$d$	$MSE$	$R$
2	0,0000007	1
3	0,0000001	1
4	0,0000007	1
5	0,0000006	1
6	0,0000005	1
7	0,0000006	1
8	0,0000008	1
9	0,0000008	1
10	0,0000008	1

Também testamos o desempenho das funções da *wavelet* e o efeito desse fator na precisão do modelo proposto. Para isso, selecionamos algumas funções usuais de três famílias tradicionais de *wavelets* *Daubechies* (*db*), *Symlets* (*sym*) e *Coiflets* (*coif*), e avaliamos os resultados referentes a  $MSE$ ,  $R$ ,  $max\_err$  e  $avg\_err$ . Os resultados correspondentes estão destacados na Tabela 32.

Os resultados da previsão das concentrações de gás dissolvido em óleo e as razões de gás são apresentados nas Tabela 30 e Tabela 31, respectivamente. É possível ver que o modelo proposto possui grande precisão em relação ao  $max\_err$  e  $avg\_err$ . Além disso, é apresentada na Figura 24 a ilustração da plotagem e erro de saída e de destino do gás H<sub>2</sub>, corroborando o bom grau de ajuste do modelo NAR–DWT.

A partir da Figura 44, é possível verificar a alta precisão na previsão do modelo proposto. O gráfico de saída e destino ilustra claramente que o modelo pode reproduzir com precisão o comportamento oscilatório dos dados de entrada, apresentando erro de destino de saída inferior a 0,01. Mesmo nos pontos em que a variação de gás é maior, como nos instantes de 40s e 128s, o modelo pode acompanhar a variação apesar de produzir erro médio um pouco maior nesse caso.

Tabela 30 – Desempenho das funções da wavelet e o efeito desse fator na precisão do modelo proposto.

<i>Wavelet</i>	<i>mse</i>	<b>R</b>	<i>avg_err (%)</i>	<i>max_err (%)</i>
<i>sym2</i>	4,E-07	0,99	0,06	0,34
<i>sym3</i>	1,E-06	0,99	0,08	0,50
<i>sym4</i>	7,E-07	0,99	0,08	0,52
<i>db1</i>	7,E-07	0,99	0,15	1,02
<i>db3</i>	1,E-06	0,99	0,08	0,52
<i>db5</i>	2,E-07	0,99	0,03	0,19
<i>coif1</i>	1,E-06	0,99	0,10	0,60
<i>coif3</i>	4E-07	0,99	0,06	0,35
<i>coif5</i>	3E-07	0,99	0,04	0,28

Tabela 31 – Erro de previsão NAR–DWT para concentrações de gás dissolvido em óleo.

<b>Tipo de Gás</b>	<i>avg_err (%)</i>	<i>max_err (%)</i>
H <sub>2</sub>	0,460	6,030
CO	0,080	0,790
CO <sub>2</sub>	0,060	0,340
CH <sub>4</sub>	0,100	0,890
C <sub>2</sub> H <sub>6</sub>	0,290	2,340
C <sub>2</sub> H <sub>4</sub>	0,320	3,110
C <sub>2</sub> H <sub>2</sub>	0,330	2,370

Os resultados gerados pelo método proposto foram comparados a importantes métodos atuais de previsão da literatura: KPCA-FFOA-GRNN, FFOA-GRNN, KPCA-GRNN, GRNN, BPNN de Lin et al. (2018) e SVM de Yongbo e Juan (2017).

Algumas técnicas de séries temporais também foram usadas para comparar os resultados dos valores de previsão das concentrações de gases dissolvidos no óleo e suas proporções. Foram utilizados os seguintes modelos estatísticos: modelo de média móvel autorregressiva (ARMA), modelos de média móvel integrada autorregressiva (ARIMA), modelo de média móvel integrada autorregressiva (SARIMA) sazonal, modelo autorregressivo para heterocedasticidade condicional (ARCH) e heteroscedasticidade condicional autorregressiva generalizada (GARCH).

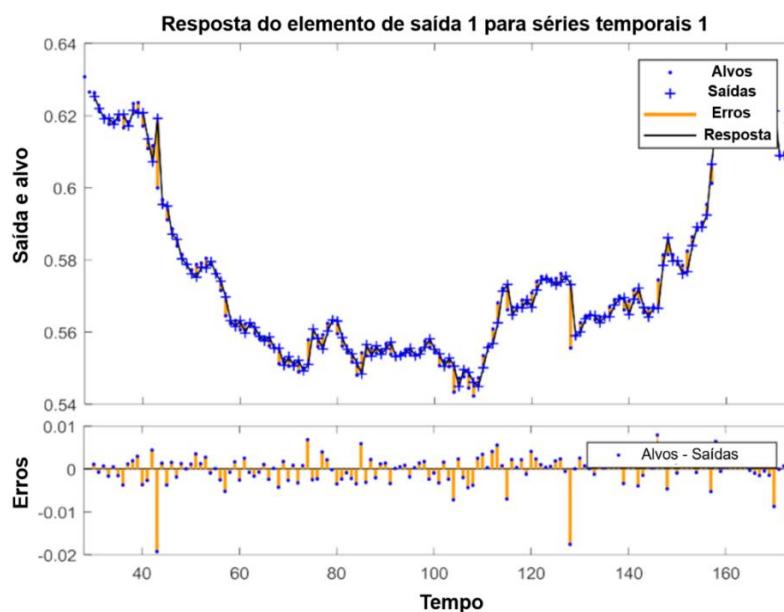
Tabela 32 – Erro de previsão NAR – DWT para razões de concentração de gases.

Tipo de Gás	<i>avg_err</i> (%)	<i>max_err</i> (%)
CH <sub>4</sub> /H <sub>2</sub>	0,58	3,35
C <sub>2</sub> H <sub>2</sub> /C <sub>2</sub> H <sub>4</sub>	2,02	10,48
C <sub>2</sub> H <sub>4</sub> /C <sub>2</sub> H <sub>6</sub>	0,66	9,82

A seleção do modelo de previsão mais adequado entre todas as opções testadas foi realizada utilizando os critérios de informação de Akaike (*AIC*) e Bayesiano (*BIC*), que indicaram o modelo mais ajustado aos dados por análise de qualidade relativa dos modelos estatísticos.

Essa comparação é ilustrada na Tabela 33 para a previsão da concentração de gás etileno. Para este teste, usamos 141 amostras para treinamento e 35 amostras para teste.

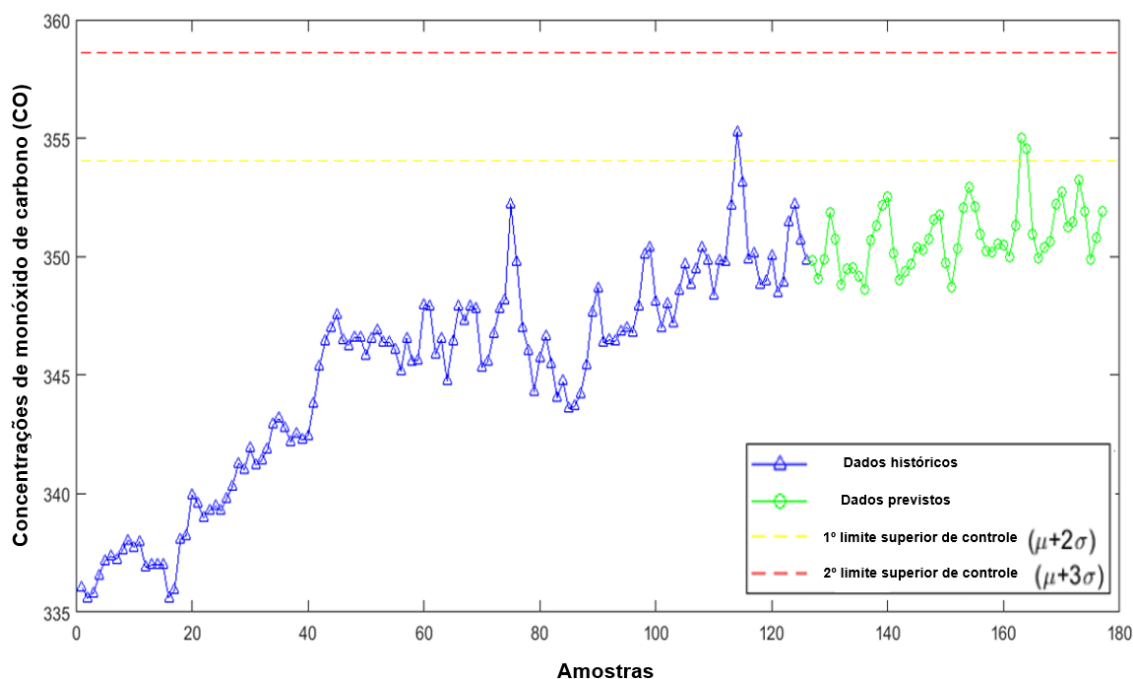
Figura 44 – Previsão versus valores-alvo da concentração de hidrogênio no óleo



**Tabela 33 – Erro de previsão para o gás etileno por diferentes modelos.**

Modelo de Previsão	<i>avg_err</i> (%)	<i>max_err</i> (%)
NAR-DWT	0,32	3,11
KPCA-FFOA-GRNN	3,27	11,34
FFOA-GRNN	5,04	12,81
KPCA-GRNN	7,09	15,14
GRNN	7,93	14,06
BPNN	8,72	19,52
SVM	4,21	10,65
GM	6,69	15,77
AIC	9,15	22,33
BIC	18,4	40,87

O modelo NAR-DWT foi aplicado a um sistema de monitoramento de transformadores, com base na avaliação das condições do equipamento a partir da análise dos gases dissolvidos em óleo. O modelo foi usado para prever a concentração de cada gás por um horizonte de tempo de 50 dias à frente, como mostra a Figura 45. Os resultados de cada previsão foram avaliados em relação aos valores-limite definidos pelas diretrizes de normas adotadas no Brasil, relacionadas às ideias estatísticas de controle de qualidade, conforme destacado na Figura 34.

**Figura 45 – Previsão em várias etapas da concentração de gás de monóxido de carbono**

## 5. DISCUSSÃO

Esta seção apresenta a discussão sobre a aplicação do modelo proposto em exemplo do mundo real, em relação à precisão do ajuste de dados para prever concentrações de gás dissolvido em óleo no transformador.

Em relação à avaliação do desempenho das funções da *wavelet* e do parâmetro de atraso de tempo “d” no processo de treinamento, e o efeito desses fatores na precisão do modelo proposto, a robustez do modelo é clara. A combinação de rede neural autorregressiva não linear e transformação de *wavelet* permite produzir previsões de alta precisão de concentrações e proporções de gases dissolvidos em óleo.

A baixa influência da função *wavelet* na precisão da previsão indica que o modelo proposto é menos sensível a variações nos dados de entrada para treinamento, problema que é comum em abordagens baseadas em redes neurais (CHENG e YU, 2018). Além disso, a adoção da abordagem de validação cruzada aumenta a confiabilidade do modelo e leva a melhores previsões de qualidade.

Os resultados do exemplo real mostraram que o modelo proposto apresentou boas e acuradas previsões, superiores às obtidas pelos métodos testados atualmente, como Rede Neural de Regressão Generalizada (GRNN), Máquina de Vetor de Suporte (SVM), Rede Neural de Propagação Reversa (BPNN) e técnicas usuais de séries temporais. Especificamente, em relação ao gás etileno, o erro máximo de previsão do modelo NAR–DWT foi cerca de 70% menor que os outros modelos testados. O desempenho superior apresentado foi independente do parâmetro de atraso, “d”, utilizado no modelo.

Esse resultado é coerente com a literatura que afirma que os modelos de rede autorregressivas não-lineares são menos sensíveis a dependências de longo prazo, além de apresentar melhores capacidades de generalização e aprendizado (BANOOKH e BARAKATI, 2012). Além disso, o aumento adicional no desempenho é devido à aplicação da transformada de *wavelet* discreta para criar versões simplificadas e esparsas dos dados originais.

A capacidade de prever com acurácia os valores futuros das concentrações de gases permite aumentar a confiabilidade do sistema de monitoramento e antecipar possíveis falhas no sistema. Nesse sentido, a previsão de aumento na concentração de gás acima do valor limite emite alerta indicando a necessidade de verificação imediata do status de operação do transformador. Além disso, várias previsões de concentração de gás são usadas para estimar o tempo médio, em dias, para a condição de alerta, bem como os limites inferior e superior do

intervalo de confiança de 95% para essa média. Essas informações são usadas para calcular a vida útil restante do equipamento.

## 6. CONCLUSÕES

Este estudo propõe a combinação do modelo de rede neural autorregressiva não linear com a transformada de *wavelet* discreta para prever as concentrações de gás dissolvido em óleo do transformador de potência.

O modelo apresenta baixa sensibilidade em relação aos parâmetros de treinamento, às funções da *wavelet* e ao parâmetro de atraso de tempo “d”, que confere robustez necessária para a previsão de alta precisão das concentrações de gás dissolvido em óleo e suas proporções. Como o modelo é capaz de prever múltiplos valores futuros das concentrações de gás, ele foi aplicado no contexto do sistema de análise de diagnóstico de falhas do transformador, permitindo a emissão de alertas em caso de possíveis falhas e fornecendo estimativas de intervalo confiáveis em relação ao tempo médio para ocorrência de anormalidade.

Uma desvantagem do modelo é a necessidade de reciclagem, à medida que novas medidas são incorporadas ao banco de dados, como é comum em todos os modelos baseados em redes neurais artificiais. Trabalhos futuros podem incluir a investigação do desempenho de tal abordagem para prever as causas de fenômeno de falha a partir de banco de dados de ocorrências que pode ser executado de maneira confiável apenas quando o banco de dados de falhas contém número suficiente de casos.

**Contribuições dos autores:** F.H.P. projetou o algoritmo, depurou o código, testou o exemplo e escreveu o manuscrito. S.J. projetou o algoritmo. F.E.B. realizou os testes com modelos estatísticos. J.S., I.C. e S.I.N. avaliou o código e revisou o manuscrito. G.F.M.d.S. fez a aplicação do modelo na análise de diagnóstico de falhas do transformador.

**Recursos:** Esta pesquisa foi financiada pela EDP Energias do Brasil, no âmbito do Programa de Pesquisa e Desenvolvimento Tecnológico da Agência Nacional de Energia Elétrica (ANEEL), número de concessão PD-0673-0050 / 2013.

**Conflitos de interesse:** Os autores declaram não haver conflito de interesse.

### Nomenclatura

DGA            Análise de gás dissolvido

ANN	Rede neural artificial
EPS	Sistema especialista
IA	Inteligência artificial
SVM	Máquina de vetor de suporte
MLP	<i>Perceptron</i> multicamada
BP	<i>Backpropagation</i>
BPNN	Rede neural de regressão generalizada
GRNN	Aprendizado de máquina ML
IEC	Comissão Eletrotécnica Internacional
IEEE	Instituto de Engenheiros Elétricos e Eletrônicos
D	Parâmetro de atraso de tempo
H <sub>2</sub>	Gás hidrogênio
CO	Gás monóxido de carbono
CO <sub>2</sub>	Dióxido de carbono
CH <sub>4</sub>	Gás metano
C <sub>2</sub> H <sub>2</sub>	Gás acetileno
C <sub>2</sub> H <sub>4</sub>	Gás etileno
C <sub>2</sub> H <sub>6</sub>	Gás etano
NAR	Rede neural autorregressiva não linear
NARX	Rede neural autorregressiva não linear com uma série temporal externa
DWT	Transformada de <i>wavelet</i> discreta
MSE	Erro quadrático médio
R	Coefficiente de determinação
KPCA	Análise de Componentes Principais do Kernel
FFOA	Algoritmo de otimização <i>fruit fly</i>
Db	<i>Daubechies wavelets</i>
Sym	<i>Symlets wavelets</i>
Coif	<i>Coiflets</i>
ARMA	Modelo de média móvel autorregressiva
ARIMA	Modelo de média móvel autorregressiva integrada
SARIMA	Modelo de média móvel integrada autorregressiva sazonal
ARCH	Modelo autorregressivo para heterocedasticidade condicional
GARCH	Heterocedasticidade condicional autorregressiva generalizada

## REFERÊNCIAS

GODINA, R.; RODRIGUES, E.M.G.; MATIAS, J.C.O.; CATALÃO, J.P.S. Effect of Loads and Other Key Factors on Oil-Transformer Ageing: Sustainability Benefits and Challenges. **Energies** 2015, 8, 12147–12186. [CrossRef]

WANG, X.; TANG, C.; HUANG, B.; HAO, J.; CHEN, G. Review of Research Progress on the Electrical Properties and Modification of Mineral Insulating Oils Used in Power Transformers. **Energies** 2018, 11, 487. [CrossRef]

PENG, L.; FU, Q.; ZHAO, Y.; QIAN, Y.; CHEN, T.; FAN, S. A Non-Destructive Optical Method for the DP Measurement of Paper Insulation Based on the Free Fibers in Transformer Oil. **Energies** 2018, 11, 716. [CrossRef]

CHENG, L.; YU, T. Dissolved Gas Analysis Principle-Based Intelligent Approaches to Fault Diagnosis and Decision Making for Large Oil-Immersed Power Transformers: A Survey. **Energies** 2018, 11, 913. [CrossRef]

BACHA, K.; SOUAHLIA, S.; GOSSA, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. **Electr. Power Syst.** 2012, 83, 73–79. [CrossRef]

CHEN, W.G.; CHEN, X.; PENG, S.Y.; LI, J. Canonical correlation between partial discharges and gas formation in transformer oil paper insulation. **Energies** 2012, 5, 1081–1097. [CrossRef]

XIANG, C.M.; ZHOU, Q.; LI, J.; HUANG, Q.D.; SONG, H.Y.; ZHANG, Z.T. Comparison of dissolved gases in mineral and vegetable insulating oils under typical electrical and thermal faults. **Energies** 2016, 9, 312. [CrossRef]

HAROLDO, D.F., Jr.; JOÃO, G.S.C.; JOSE, L.M.O. A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. **Renew. Sustain. Energy Rev.** 2015, 46, 201–209.

FAIZ, J.; SOLEIMANI, M. Dissolved gas analysis evaluation in electric power transformers using conventional methods a review. **IEEE Trans. Dielectr. Electr. Insul.** 2017, 24, 1239–1248. [CrossRef]

LIN, J.; SHENG, G.; YAN, Y.; DAI, J.; JIANG, X. Prediction of Dissolved Gas Concentrations in Transformer Oil Based on the KPCA-FFOA-GRNN Model. **Energies** 2018, 11, 1–13. [CrossRef]

IEC 60599: **Interpretation of the Analysis of Gases in Transformers and Other Oil-Filled Electrical Equipment in Service.** 1978. Available online: <https://webstore.iec.ch/publication/23317> (accessed on 12 January 2018).

IEEE Std C57.104-1991. **Guide For Loading Mineral-Oil-Immersed Transformers and Step Voltage Regulators.** 2011. Available online: <https://ieeexplore.ieee.org/document/6166928> (accessed on 12 January 2018).

YONGBO, T.; JUAN, F.A. Prediction method Dissolved Gas Content in Transformer Oil Based on KTA-SVM. **Control Eng. China** 2017, 24, 2263–2267.

ZHANG, S.; BAI, Y.; WU, G.; YAO, Q. The forecasting model for time series of transformer DGA data based on WNN-GNN-SVM combined algorithm. In **Proceedings of the 2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE)**, Xi'an, China, 14–17 May 2017; pp. 292–295.

WANG, Z.; LIU, Y.; GRIFFIN, P.J. A combined ANN and expert system tool for transformer fault diagnosis. **IEEE Trans. Power Deliv.** 2000, 13, 1224–1229. [CrossRef]

ISLAM, S.M.; WU, T.; LEDWICH, G. A novel *fuzzy* logic approach to transformer fault diagnosis. **IEEE Trans. Dielectr. Electr. Insul.** 2000, 7, 177–186. [CrossRef]

DIAS, C.G.; PEREIRA, F.H. Broken Rotor Bars Detection in Induction Motors Running at Very Low Slip Using a Hall Effect Sensor. **IEEE Sens. J.** 2018, 18, 4602–4613. [CrossRef]

SCHIMIT, P.H.T.; PEREIRA, F.H. Disease spreading in complex networks: A numerical study with Principal Component Analysis. **Expert System Application.** 2018, 97, 41–50. [CrossRef]

JENSEN, A.; LA COUR-HARBO, A. **Ripples in Mathematics: The Discrete Wavelet Transform**; Springer: Berlin, Germany, 2001.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**, 2nd ed.; Pearson Prentice Hall: Hamilton, ON, Canada, 1999.

SUGANTHI, S.; MURUGESAN, K.; RAGHAVAN, S. ANN model of RF MEMS lateral SPDT switches for millimeter wave applications. **J. Microw. Optoelectron. Electro-Magn. Appl.** 2012, 11, 130–143. [CrossRef]

BANOOKH, A.; BARAKATI, S.M. Optimal design of double folded stub microstrip filter by neural network modelling and particle swarm optimization. **J. Microw. Optoelectron. Electromagn. Appl.** 2012, 11, 204–213. [CrossRef]

ISLAM, M.P.; MORIMOTO, T. Non-linear autoregressive neural network approach for inside air temperature prediction of a pillar cooler. **Int. J. Green Energy** 2017, 14, 141–149. [CrossRef]

CASWELL, J.M. A Nonlinear Autoregressive Approach to Statistical Prediction of Disturbance Storm Time Geomagnetic Fluctuations Using Solar Data. **J. Signal Inf. Process.** 2014, 5, 42–53. [CrossRef]

PEREIRA, F.H.; GRASSI, F.; NABETA, S.I. A surrogate based two-level genetic algorithm optimization through *wavelet* transform. **IEEE Trans. Magn.** 2015, 51, 1–4. [CrossRef]

LI, H.; YANG, R.; WANG, C.; HE, C. Investigation on Planetary Bearing Weak Fault Diagnosis Based on a Fault Model and Improved *Wavelet* Ridge. **Energies** 2018, 11, 1286. [CrossRef]

WANG, X.; XU, J.; ZHAO, Y. *Wavelet* Based Denoising for the Estimation of the State of Charge for Lithium-Ion Batteries. **Energies** 2018, 11, 1144. [CrossRef]

SWELDENS, W. The lifting scheme: A custom-design construction of biorthogonal *wavelets*. **Appl. Comput. Harmon. Anal.** 1994, 3, 186–200. [CrossRef]

LIU, H.; SHI, J. Applying ARMA-GARCH approaches to forecasting short-term electricity prices. **Energy Econ.** 2013, 37, 152–166. [CrossRef]

**APÊNDICE E (Artigo 5) – Transformada do tipo *wavelet* para otimizar a ordem de um modelo de rede neural autorregressiva para prever a concentração de gás dissolvido no óleo do transformador de potência**

*Sensors – Open Access Journal.*

Qualis: A1.

*Manuscript ID: 762361.*

*Title: Wavelet-like transform to optimize the order of an autoregressive neural network model to predict the dissolved gas concentration in power transformer oil from sensor data.*

*Corresponding author: Dr. Fabio Henrique Pereira*

*Corresponding Author's Institution: Universidade Nove de Julho, Universidade de São Paulo.*

*First Author: Elânio Bezerra; Fabio Henrique Pereira.*

*Order of Authors: Francisco Elânio Bezerra, Fabio Henrique Pereira.*

*Order of Authors: Francisco Elânio Bezerra<sup>1</sup>, Fernando André Zumer Garcia<sup>2</sup>, Silvio Ikuyo Nabeta<sup>3</sup>, Gilberto Francisco Martha de Souza<sup>3</sup>, Ivan Eduardo Chabu<sup>3</sup>, Josemir Coelho Santos<sup>3</sup>, Shigueru Nagao Junior<sup>3</sup>, Fabio Henrique Pereira.<sup>1,2,3\*</sup>*

**Resumo:** A Análise de Gás Dissolvido (DGA) é um dos métodos mais importantes para analisar falhas em transformadores de potência. Em geral, DGA é aplicada em sistemas de monitoramento baseados no modelo autorregressivo, o valor atual da série temporal é regredido em valores anteriores da mesma série e nos valores presente e passado de algumas variáveis exógenas. A principal dificuldade é decidir a ordem do modelo autorregressivo; o que significa determinar o número de valores precedentes a serem usados. Este estudo propõe uma transformação tipo *Wavelet* para otimizar a ordem das variáveis em rede neural autorregressiva não linear para prever a concentração de gases (DGC). *Wavelets Daubechies* de diferentes comprimentos são usadas para criar representações com diferentes atrasos de tempo de dez DGC, que são submetidas a procedimento baseado na Análise de Componentes Principais (PCA) e na correlação de Pearson para descobrir a ordem do DGC. As representações com atrasos ideais para cada DGC são aplicadas como entrada em rede *Multi-Layer Perceptron* (MLP) com algoritmo *backpropagation* para prever cada concentração de gás. Essa abordagem produz melhores resultados que escolher o mesmo atraso de tempo para todas as entradas. As previsões atingiram erro MAPE médio de 5,763% para C<sub>2</sub>H<sub>2</sub>, 1,525%, 1,831%, 2,869%, 5,069% para C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, CH<sub>4</sub>, H<sub>2</sub>, respectivamente.

**Palavras-chave:** análise de gás dissolvido, transformadores de potência, transformada do tipo *Wavelet*, modelo autorregressivo.

### ***Wavelet-like transform to optimize the order of an autoregressive neural network model to predict the dissolved gas concentration in power transformer oil from sensor data***

**Abstract:** *Dissolved Gas Analysis (DGA) is one of the most important methods to analyse fault in power transformers. In general, DGA is applied in monitoring systems based upon autoregressive model, the current value of a time series is regressed on past values of the same series and, present and past values of some exogenous variables. The main difficulty is to decide the order of the autoregressive model; it means determining the number of past values to be used. This study proposes a wavelet-like transform to optimize the order of the variables in a nonlinear autoregressive neural network to predict the gas concentration (DGC) from sensor data. Daubechies wavelets of different lengths are used to create representations with different time delays of ten DGC, which are then subjected to a procedure based on Principal Components Analysis (PCA) and Pearson's correlation to find out the order of DGC. The representations with optimal time delays for each DGC are applied as input in Multi-Layer Perceptron (MLP) network with backpropagation algorithm to predict the gas at the present time. This approach produces better results than choosing the same time delay for all inputs, as usual. The forecasts reached an average MAPE error of 5.763% for C<sub>2</sub>H<sub>2</sub>, 1.525%, 1.831%, 2.869%, 5.069% for C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, CH<sub>4</sub>, H<sub>2</sub> respectively.*

**Keywords:** *dissolved gas analysis; power transformers, wavelet-like transform, autoregressive model.*

## **1. INTRODUÇÃO**

O transformador é um dos dispositivos mais importantes no processo de distribuição de eletricidade e a distribuição confiável de energia depende, em grande parte, da operação sem falhas desse equipamento. A falha do transformador durante a operação pode trazer perda significativa de receita para a concessionária, possíveis danos ambientais, riscos de explosão e incêndio e custos caros de reparo ou substituição (TANG e WU, 2013; CHAKRAVORTI et al., 2013). Caso esses dispositivos falhem, a expectativa de vida operacional e a confiabilidade

podem mudar ao longo dos anos e a eletricidade para os consumidores pode ser interrompida. Portanto, a análise da condição e manutenção do transformador é extremamente importante para garantir confiabilidade confiável da eletricidade. (TANG e WU, 2013; DONG et al., 2019; FARIA JUNIOR et al., 2015; RAHMAN et al., 2018).

Quando o transformador está em operação normal, o óleo isolante e o material isolante sólido deterioram-se gradualmente e pequena quantidade de gás será decomposta, incluindo principalmente hidrogênio ( $H_2$ ), metano ( $CH_4$ ), acetileno ( $C_2H_2$ ), etileno ( $C_2H_4$ ), etano ( $C_2H_6$ ), monóxido de carbono (CO) e dióxido de carbono ( $CO_2$ ). Na condição de falha interna do transformador, a velocidade de emergência desses gases é acelerada (JUN et al., 2018).

Uma das ferramentas mais importantes para o monitoramento da condição do transformador de potência e o diagnóstico de falhas internas é o teste de cromatografia em fase gasosa do óleo do transformador, conhecido como análise de gás dissolvido (DGA) (CHENG e YU, 2018; EL-HOSHY et al., 2019; LI et al., 2016; NOORI et al., 2017).

Vários estudos abordaram a criação de sistemas de monitoramento de condição de transformadores de potência baseados em DGA. Assim, muitas técnicas para prever a concentração de gases foram propostas, como *Wavelet Least Squares*, Máquina de vetores de suporte, Rede Neural, *Deep Learning*, Monocamada WSe2 dopada por PT-PT baseada no método dos primeiros princípios, modelo difuso, rede neural *Long short-term memory* (LSTM), apenas para citar alguns. Em geral, técnicas de inteligência artificial têm sido amplamente utilizadas para desenvolver ferramentas de diagnóstico mais precisas com base nos dados da DGA (BACHA et al., 2012; LI et al., 2016; LIN et al., 2011; LIU et al., 2019; NOORI et al., 2017; RAHMAN et al., 2018; PEREIRA et al., 2018).

Liu et al. (2019), por exemplo, criaram abordagem para diagnosticar falhas de transformadores com base na taxa de gás e na máquina de vetores de suporte (SVM). Primeiro, com base no banco de dados IEC TC 10, as taxas ideais de gás dissolvido são obtidas pelo algoritmo genético projetado para seleção simultânea da taxa DGA e otimização dos parâmetros SVM. Naquele trabalho, três métodos tradicionais são usados: SVM DGA, DGA da rede neural com algoritmo *backpropagation* (BPNN), critérios IEC e proporções de gás IEC de três chaves com SVM e rede neural de propagação traseira são empregadas para comparar a precisão. A técnica SVM também serviu de base para as abordagens em Zheng et al. (2018) e Liu et al. (2019).

Liu et al. (2019) utilizaram a máquina de vetores de suporte de mínimos quadrados (LS-SVM) para previsão de gases dissolvidos ( $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  e  $C_2H_6$ ) e para avaliar falhas incipientes do isolamento de polímeros de transformadores. Enquanto, em Zheng et al. (2018),

foi proposta uma nova abordagem para combinar a regressão técnica de *wavelet* com o LS-SVM para previsão de gases dissolvidos em transformadores de potência imersos em óleo.

No trabalho de Noori et al. (2017), os autores usaram um sistema de inferência difusa (FIS) para determinar concentrações absolutas de óleo dissolvidos do transformador, gases combustíveis totais dissolvidos, gases combustíveis totais, proporções de alguns gases entre si e taxas de gás crescentes para detectar a decomposição de papéis de isolamento de transformadores.

Uma abordagem semelhante foi proposta em Rahman et al. (2018), na qual um sistema de inferência neuro difusa adaptativa (ANFIS) foi empregado para estimar a taxa de degradação do isolamento do transformador com as variáveis de entrada  $H_2$  (hidrogênio),  $CH_4$  (metano),  $N_2$  (nitrogênio),  $O_2$  (oxigênio),  $CO$  (monóxido de carbono),  $CO_2$  (dióxido de carbono),  $C_2H_6$  (etano),  $C_2H_4$  (etileno),  $C_2H_2$  (acetileno) e TDCG (gás combustível total dissolvido).

Em geral, esses numerosos estudos têm utilizado técnicas de inteligência artificial como regressão para prever a concentração de gás ou falhas no transformador de potência. Mais especificamente, o uso de modelos de previsão em conexão com a transformada *wavelet* foi abordado em alguns trabalhos recentes para melhorar a previsão (ZHENG et al., 2018; LIU et al., 2019; PEREIRA et al., 2018). Apesar dos resultados satisfatórios, essas abordagens podem não ser as mais eficientes na previsão de valores futuros da variável de interesse, especialmente para previsão em várias etapas à frente.

Vários estudos empíricos mostram que o aprendizado de dependências temporais de longo prazo pode ser difícil para algoritmos de descida de gradiente, que são mais eficazes, convergem mais rapidamente e generalizam melhor em modelos de redes neurais autorregressivas não lineares do que em outras redes neurais (PEREIRA et al., 2018; DIACONESCU, 2008; LIN et al., 1996; GAO e MENG, 2005; LIN et al., 1997; SIEGELMANN et al., 1997).

Modelos autorregressivos baseados em redes neurais especificam que a variável de saída depende, de maneira não linear, de seus próprios valores passados e de um termo estocástico imperfeitamente previsível. Assim, a previsão de valores futuros da variável de saída pode ser realizada a partir de seus valores passados e presentes. Além disso, o modelo de previsão também pode considerar valores presentes e passados de uma ou mais variáveis externas auxiliares, resultando em modelo autorregressivo não linear com variáveis exógenas.

Por outro lado, Pereira et al. (2018) propuseram um combinação de um modelo de rede neural autorregressiva não linear com a transformada *wavelet* discreta, resultando em previsão de alta precisão com várias etapas à frente de concentrações de gás no óleo. Os autores

investigaram o uso de diferentes funções da *wavelet* e diferentes atrasos no modelo de regressão automática, mas não avaliaram como diferentes atrasos nas séries externas podem influenciar os valores das séries de interesse.

De fato, a definição dos atrasos ideais de entrada e saída é uma das principais limitações do modelo autorregressivo. Em geral, em modelos multidimensionais com “n” variáveis externas, adiamentos de variáveis iguais são adotados. Isso significa que a previsão do valor de saída no tempo  $t+1$ ,  $y(t+1)$  é realizada usando as saídas passadas  $y(t)$ ,  $y(t-1)$ , ...,  $y(t-d_y)$  e as observações passadas  $u_i(t)$ ,  $u_i(t-1)$ , ...,  $u_i(t-d_u)$  das variáveis externas  $u_i$  como entradas,  $i=1, \dots, n$ . Além disso, a adoção de muitas entradas pode aumentar a complexidade do modelo de previsão e reduzir sua precisão. Assim, persistem algumas dificuldades e limitações, apesar dos avanços, motivando pesquisas para novos modelos a serem conduzidos.

A investigação do uso de diferentes atrasos em séries externas que influenciam a saída não parece ter recebido a atenção necessária, principalmente considerando que existe forte correlação entre as concentrações de diferentes gases e falhas nos transformadores. Este trabalho busca contribuir para superar essa limitação, propondo uma transformação do tipo *wavelet* para otimizar a ordem dos fatores em modelo de rede neural autorregressivo, com algumas variáveis exógenas, para prever a concentração de gás dissolvido no óleo do transformador de potência.

O principal objetivo deste trabalho é determinar o atraso ideal para cada entrada e saída, para criar um modelo autorregressivo com número reduzido de entradas e com precisão competitiva em relação à literatura. A hipótese é que as aproximações do tipo *wavelet* das variáveis externas e da variável de saída incorporam a memória temporal do modelo autorregressivo. Além disso, a seleção da melhor aproximação para cada variável determina o atraso ideal para cada entrada e reduz o tamanho do modelo, pois cada amostra da aproximação é calculada considerando uma janela de tempo da série.

Conseqüentemente, as contribuições da abordagem proposta podem ser declaradas da seguinte forma:

1. desenvolvimento de abordagem baseada em transformação do tipo *wavelet* que determine o atraso ideal para cada variável externa e para a variável de interesse em modelo de previsão autorregressivo com variáveis exógenas;
2. um modelo de previsão com alta precisão, pois se concentra na tendência dos sinais de entrada a partir das aproximações sem ruído calculadas pela transformação de *wavelet*;
3. ampliação do conhecimento da relação temporal entre os gases subjacentes ao processo de degradação do óleo isolante e do material isolante sólido;

4. redução do número de variáveis de entrada no modelo de regressão automática ao usar as aproximações resultantes de transformações do tipo *wavelets* de diferentes comprimentos, que já consideram o atraso de tempo determinado para cada variável.

O restante deste artigo está organizado da seguinte forma: teoria relacionada na seção 2, análise de gás dissolvido em óleo na seção 2.1, transformação do tipo *wavelet* discreta na seção 2.2 e modelo exógeno autorregressivo não linear na seção 2.3. Na seção 3 serão apresentados materiais e métodos, resultados na seção 4, discussão na seção 5 e, finalmente, a conclusão na seção 6.

## **2. TEORIA RELACIONADA**

### **2.1 Análise de gás dissolvido em óleo**

Os transformadores de potência são dos dispositivos mais importantes para o sistema elétrico, e mais de 90% dos transformadores são imersos em óleo (CHEN et al., 2019). O gás dissolvido em óleo (DGO) é parâmetro simples usado para monitorar transformadores de potência energizados e avaliar a condição dos transformadores de potência (PTE) (CIGRE, 2011; BUSTAMANTE et al., 2019; ABU BAKAR et al., 2014).

Devido às tensões térmicas e elétricas sofridas pelo isolamento dos transformadores em operação, ocorre a decomposição do papel e do óleo, gerando gases que se dissolvem no óleo e reduzem sua resistência dielétrica. Assim, concentrações de vários gases se dissolveram no óleo do transformador devido à decomposição do isolamento de óleo e papel (ABU BAKAR et al., 2014).

Dependendo da composição e da concentração de gás, o transformador de potência pode falhar. Normalmente, os gases presentes dissolvidos no óleo isolante nos transformadores são hidrogênio ( $H_2$ ), metano ( $CH_4$ ), etano ( $C_2H_6$ ), etileno ( $C_2H_4$ ), acetileno ( $C_2H_2$ ), dióxido de carbono ( $CO_2$ ), monóxido de carbono (CO), oxigênio ( $O_2$ ) e nitrogênio ( $N_2$ ) (CHAKRAVORTI et al., 2013).

Existem alguns métodos interpretativos baseados em DGA para detectar a falha do transformador de potência, que usam a relação entre dois gases para determinar possível problema. Os métodos são chave de gás; índices IEC; a representação gráfica, IEC 60599, Duval; Doernenburg, Rogers, entre outros (FARIA JUNIOR et al., 2015; BACHA et al., 2012;

CHEN et al., 2019). Como exemplo, a seguir, serão mostrados três desses métodos e suas respectivas concentrações de gás.

A Tabela 34 mostra a descrição da falha para concentração de gás, descrição do problema, valores normais e anormais para cada concentração de gás. Na Tabela 35, temos o diagnóstico de falhas pelo método da razão de Dornenburg (R1, R2, R3, R4) e tipo de falha, enquanto a Tabela 36 mostra a classificação de falhas usando códigos de razão IEC ( $C_2H_2/C_2H_4$ ,  $CH_4/H_2$ ,  $C_2H_4/C_2H_6$ ) e falha tipo (CHAKRAVORTI et al., 2013; BUSTAMANTE et al., 2019).

**Tabela 34 – Descrição da falha para concentração de gás**

<b>Fórmula Química</b>	<b>Normal</b>	<b>Alterado</b>	<b>Descrição do Problema</b>
H <sub>2</sub> (hidrogenio)	<150 ppm	>1000 ppm	Descarga elétrica (efeito corona, baixa descarga parcial)
CH <sub>4</sub> (metano)	<25 ppm	>80 ppm	Indicador secundário de um arco ou superaquecimento grave
N <sub>2</sub> (nitrogenio)	1-10 %	NA	-
O <sub>2</sub> (oxigenio)	0.03 %	>0.5 %	Falha na vedação do transformador
CO (monoxido de carbono)	<500 ppm	>1000 ppm	Decomposição de celulose
CO <sub>2</sub> (dioxido de carbono)	<10000 ppm	>15 000ppm	Decomposição de celulose
C <sub>2</sub> H <sub>6</sub> (etano)	<10 ppm	>35 ppm	Indicador secundário de falha térmica
C <sub>2</sub> H <sub>4</sub> (etileno)	<20 ppm	>100 ppm	Falha térmica (superaquecimento local)
C <sub>2</sub> H <sub>2</sub> (acetileno)	<15 ppm	>70 ppm	Falha elétrica (arco, faísca)

**Tabela 35 – Diagnóstico de falhas pelo método da razão de Dornenburg**

<b>Razão R1 (CH<sub>4</sub>/H<sub>2</sub>)</b>	<b>Razão R2 (C<sub>2</sub>H<sub>2</sub>/C<sub>2</sub>H<sub>4</sub>)</b>	<b>Razão R3 (C<sub>2</sub>H<sub>2</sub>/CH<sub>4</sub>)</b>	<b>Razão R4 (C<sub>2</sub>H<sub>6</sub>/C<sub>2</sub>H<sub>2</sub>)</b>	<b>Tipo de Falha</b>
> 1	< 0.75	< 0.3	> 0.4	Decomposição termal
< 0.1	Insignificante	< 0.3	> 0.4	Corona
> 0.1 & < 1	> 0.75	> 0.3	< 0.4	Arco

Tabela 36 – Classificação de falhas usando códigos de razão IEC

$C_2H_2/C_2H_4$	$CH_4/H_2$	$C_2H_6/C_2H_2$	Tipo de Falha
0	0	0	Envelhecimento normal, sem falhas
Insignificante	1	0	Descarga parcial de baixa densidade energética
1	1	0	descarga parcial de alta densidade energética
1	0	1	Descargas de baixa energia
1	0	2	Descargas de alta energia
0	0	1	Falha térmica $<150\text{ }^\circ\text{C}$
0	2	0	Falha térmica de $\geq 150\text{ }^\circ\text{C}$ e $\leq 300\text{ }^\circ\text{C}$
0	2	1	Falha térmica de $> 300\text{ }^\circ\text{C}$ e $\leq 700\text{ }^\circ\text{C}$
0	2	2	Falha térmica de $> 700\text{ }^\circ\text{C}$

## 2.2 Transformação discreta de *wavelet*

A transformação *wavelet* (WT) é a técnica de processamento de sinal generalizada. Nas últimas décadas, vários algoritmos de *wavelet* com suporte compacto foram criados por comunidades de análise matemática e processamento de sinais. De fato, vários trabalhos tentam motivar e explicar as ideias básicas dos *Wavelets*, o que as torna tão bem-sucedidas em muitas aplicações em diferentes áreas, quanto em algumas limitações (DAUBECHIES, 1988; DAUBECHIES e SWELDENS, 1998; JENSEN e LA COUR-HARBO, 2001; GOGOLEWSKI, 2020).

Principalmente, o WT é usado para análise de tendências, correlação e coerência entre duas séries temporais, análise espectral cruzada e *Wavelets* de verificação com base no espaço, enquanto há aplicações limitadas relacionadas à previsão (RHIF et al., 2019).

Transformação de *wavelet* discreta (DWT) é aquela para a qual as *wavelets* são discretamente amostradas, o que significa que este é um conjunto discreto de escalas e traduções de *wavelets* (RHIF et al., 2019). A ideia básica dessa técnica é explorar a correlação presente na maioria dos sinais da vida real para criar aproximação esparsa (PEREIRA et al., 2018). Portanto, o DWT possui muitas propriedades favoráveis que são úteis para pesquisadores no campo de mineração de dados de séries temporais (TANG e WU, 2013).

A dependência de transformar uma *wavelet* unidimensional discreta é apresentada na Equações 50, 51 e 52.

$$f(t) = \sum_k A_{m,k} \varphi_{m,k}(t) + \sum_{j=1}^m \sum_k D_{j,k} \Psi_{j,k}(t) \quad (50)$$

$$H(\omega) = \sum_{n \in \mathbb{Z}} h_n e^{-in\omega} \quad (51)$$

$$G(\omega) = \sum_{n \in \mathbb{Z}} g_n e^{-in\omega} \quad (52)$$

Em que  $H(\omega)$  é a função de transferência do filtro passa-alta,  $G(\omega)$  é a função de transferência do filtro passa-baixa e eles filtram os sinais de baixa e alta frequência, mantendo informações importantes do sinal original,  $h_n$ ,  $g_n$  são os coeficientes dependendo da *wavelet* mãe,  $A_{(m, k)}$  é um perfil aproximado,  $D_{(j, k)}$  é o perfil detalhado, “j”, “m” é o nível de decomposição da transformada *wavelet*,  $\Psi_{(j, k)}(t)$  é mãe *wavelet* e  $\varphi_{(m, k)}(t)$  é a função de escala.

As *wavelets Daubechies* (DW), *dbN* oferecem uma família de transformações ortogonais, em que N se refere ao número de momentos de fuga que geralmente variam de *db2* a *db22* e têm a capacidade de aproximar com precisão funções constantes e lineares, além de forma relativamente simples (TANG e WU, 2013; GOGOLEWSKI, 2020).

### 2.3 MODELO EXÓGENO AUTOREGRESSIVO NÃO LINEAR

Uma série temporal é qualquer conjunto de observações organizadas no tempo; geralmente, as medições são feitas em tempos espaçados de maneira uniforme, “p”. valores diários de poluição, valores mensais de temperatura, valores diários de consumo de eletricidade, índices diários de bolsa de valores, entre outros (BOX e JENKINS, 2015).

Um modelo autorregressivo não linear exógeno (NARX) na modelagem de séries temporais é aquele que possui entradas exógenas, nas quais o modelo relaciona valores passados da mesma série e valores atuais e passados da série acionadora (exógena), ou seja, séries determinadas externamente que influenciam as séries de interesse, conforme definido na Equação 53 para  $l$  variáveis exógenas.

$$y_t = F(y_{t-1}, y_{t-2}, \dots, y_{t-d_y}, \dots, u_{1,t}, u_{1,t-1}, \dots, u_{1,t-d_{u1}}, \dots, u_{l,t}, u_{l,t-1}, \dots, u_{l,t-d_{ul}}) + \varepsilon \quad (53)$$

Na qual  $y$  é a variável de saída,  $u_i$ ,  $i = 1, \dots, l$  são variáveis determinadas externamente,  $d_{ui}$  é a ordem da variável  $u_i$ ,  $\varepsilon$  é o termo do erro e  $F$  é uma função não linear, como a polinomial por exemplo.

Há tendência para a adoção de técnicas computacionais e muitas tentativas eficazes foram desenvolvidas, tais como: Pereira et al. (2018), que usaram o NARX para prever a concentração de gás no óleo do transformador de potência. Enquanto, Rahman et al. (2018) e NOORI et al. (2017) criaram um modelo autorregressivo usando modelos ANFIS para detectar, isolar e estimar a expectativa de vida útil de papel de transformador.

Como Li et al. (2016), que utilizaram o SVM para verificar o diagnóstico de falha do transformador de potência, na qual utilizaram SVM e algoritmo genético e Liu et al. (2019) usaram *wavelet least squares* SVM e algoritmo de competição imperialista para verificar falhas do transformador. Finalmente, Li et al. (2016) utilizaram o LSTM para fazer previsões para o estado de execução do transformador de potência.

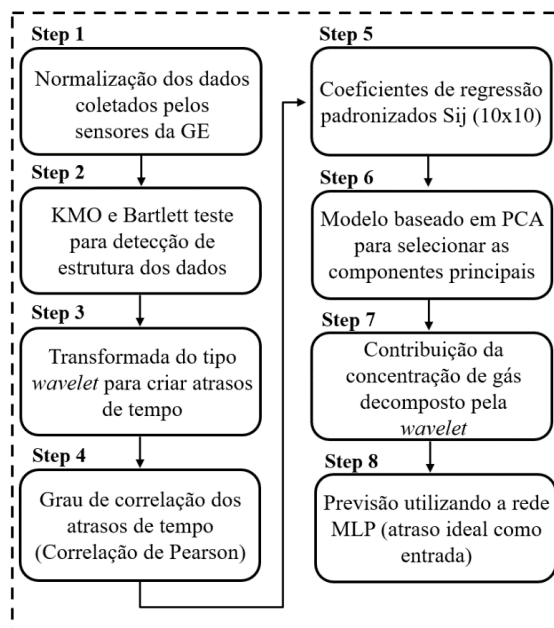
Como pode ser visto, existem alguns modelos autorregressivos que foram utilizados para auxiliar na previsão de DGA e o estado de saúde do transformador de potência (CIGRE, 2011). No entanto, em todos os trabalhos pesquisados, a ordem dos fatores é sempre a mesma, ou seja,  $d_y = d_{u1} = \dots = d_{ul}$  na Equação 53.

### 3. MATERIAIS E MÉTODOS

A abordagem proposta se baseia na transformação do tipo *wavelet* para otimizar a ordem do fator (concentrações de gases) em modelo autorregressivo não linear com variáveis exógenas. Significa definir a ordem ideal  $d_y, d_{u1}, \dots, d_{ul}$  para cada concentração de gás.

Assim, a abordagem proposta possui as seguintes etapas: etapa 1, aquisição da concentração de gás e normalização dos dados; passo 2, teste KMO e Bartlett; etapa 3, decomposição do tipo *wavelet* da concentração de gás; etapa 4, correlação de Pearson; etapa 5, coeficientes de regressão padronizados; etapa 6, um modelo usando PCA para selecionar o componente principal; etapa 7, cálculo da taxa de contribuição para cada nível de decomposição da *wavelet* e, finalmente; etapa 8, previsão usando o melhor atraso de tempo como entrada em uma rede MLP. Todas essas etapas estão ilustradas na Figura 46 e são descritas em detalhes da seguinte maneira.

**Figura 46 – Modelo para seleção e taxa de contribuição da concentração e previsão de gases**



Etapa 1: geralmente, técnicas de interpretação como o triângulo Duval são aplicadas às informações sobre a concentração de gases no óleo do transformador, que são coletadas usando um equipamento como *Morgan Calisto*, *Luman Sense Smart DGA*, *GE Transfix*, *Qualitrol DGA 150* ou outros. (EL-HOSHY et al., 2019).

Inicialmente, este trabalho coletou um conjunto de 190 dados históricos de gás dissolvido em óleo de um transformador equipado com um *GE Kelman-Transfix* (*GE – General Electric*, São Paulo, Brasil) e *GE Intellix BMT 330* (*GE – General Electric*, São Paulo, Brasil). Nessa etapa, as variáveis apontadas por Bacha et al. (2012); Lin et al. (2011); Liu et al. (2019); Noori et al. (2017) são  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $O_2$ ,  $H_2$ . No entanto, as concentrações de  $H_2O$  e gás combinado foram adicionadas como entrada, resultando em dez variáveis. Antes da próxima etapa, todos os dados foram normalizados entre 0 e 1.

Etapa 2: o teste KMO é aplicado para verificar a amostragem de adequação da medida para cada variável no modelo Cerny e Kaiser (1977) e o teste de Bartlett para testar a hipótese de que a matriz de correlação seja matriz de identidade, o que indicaria que as variáveis não são relacionadas e, portanto, inadequadas para a estrutura detecção (BARTLETT, 1937).

O KMO (1977) é critério para identificar se o modelo de análise fatorial em uso está adequadamente ajustado aos dados, testando a consistência geral dos dados (CERNY e KAISER, 1977). Enquanto o teste de esfericidade de Bartlett, técnica criada por Maurice Stevenson Bartlett em 1937, indica a força do relacionamento entre variáveis.

Etapa 3: nessa fase, o DWT é usado de duas formas. No primeiro, cada concentração de gás é decomposta mantendo o nível de decomposição em 1 e alterando a *wavelet* de  $db_2, db_4, \dots$  para  $db_{20}$ , a fim de criar aproximações suaves da concentração original de gás usando os filtros de baixa frequência. Além disso, a transformação do tipo *wavelet* é aplicada nas concentrações de gás em ordem cronológica reversa, de modo que cada amostra da aproximação seja criada com valores passados do sinal original.

Considerando “m” amostras de uma série temporal em ordem cronológica reversa, que são as amostras mais recentes no início,  $S = (s_t, s_{(t-1)}, \dots, s_{(t-k)}, \dots, s_{(t-m+k)}, \dots, s_{(t-m+2)}, s_{(t-m+1)})$  e um filtro *Wavelet* passagem baixa “H” de comprimento “k”,  $H = (h_0, h_1, \dots, h_{(k-1)})$ ,  $k \ll m$ , na qual a Equação 54 define a aplicação da transformação ao sinal S para criar uma aproximação  $S'_{dbk} = (s'_t, s'_{t-1}, \dots, s'_{t-(m/2)+1})$  com atraso de tempo  $k - 1$ , conforme proposto neste trabalho.

$$s'_{t-j} = \sum_{i=0}^{k-1} h_i s_{t-i-jk}, j=0, \dots, \frac{m}{2} - 1 \quad (54)$$

As aproximações  $S'_{dbk}$ ,  $k=2, 4, \dots, 20$ , com metade do comprimento do sinal original, “m/2”, para cada *Wavelets* de *Daubechies* de  $db_2$  a  $db_{20}$ , são criados resultando em 10 aproximações para cada série temporal “S”. Aqui, temos 190 amostras de cada concentração de gás.

Etapa 4: ao contrário de Wang et al. (2015), que utilizaram o coeficiente de correlação de Pearson entre o parâmetro de característica constante e o candidato dos parâmetros de característica variável para verificar a concentração de gás que apresenta a melhor correlação com falhas elétricas, este trabalho usa a correlação de Pearson para calcular a relação entre as várias aproximações criadas para concentrações de gases com diferentes atrasos no tempo (*wavelets* diferentes comprimentos). Assim, essa etapa resulta em uma matriz X com 110 colunas e 190 linhas, de modo que as 110 colunas representem o tempo “t”,  $t-2$  a  $t-20$  de cada concentração de gás, o que gera 110 variáveis de entrada.

Etapas 5 e 6: nessas etapas, aplicamos o PCA na matriz “A” criada a partir da relação entre as entradas  $X_j$  (concentrações de gás atrasadas no tempo  $t-2$  a  $t-20$  de acordo com a transformação tipo *wavelet* e a saída)  $Y_i$  (uma concentração de gás em tempo instante t). Portanto, os valores de “A” são calculados como coeficientes de regressão padronizados  $a_{ij}$  (Equação 55) para cada entrada e saída, descrevendo a relação entre a concentração de um determinado gás e as aproximações criadas para todos os outros gases em diferentes atrasos gerados pela transformação do tipo *wavelet*. Portanto, é criada uma matriz quadrada para cada

concentração de gás, na qual o PCA é aplicado para selecionar os principais componentes que representam pelo menos 99% da variação dos dados originais, gerando um PCA supervisionado (SPCA) (LIU et al., 2019; SCHIMIT e PEREIRA, 2018).

$$a_{ij} = \frac{X_j^T Y_i}{\sqrt{X_j^T Y_i}} \quad (55)$$

Etapa 7: a contribuição de cada atraso de tempo é calculada da seguinte forma:  $C_i(j) = \sum_{i=1}^p (A_j^T A_j') \lambda_i$ , em que “A” representa os dados de entrada,  $\lambda$  valores próprios correspondentes, A' é a representação de “A” no espaço do componente principal e “p” é o componente principal mais importante (SCHIMIT e PEREIRA, 2018).

Etapa 8: uma rede neural MLP é treinada com o algoritmo *backpropagation* com 100 épocas, uma camada de entrada, uma camada oculta e uma camada de saída. Os neurônios na camada oculta foram utilizados seguindo duas abordagens, a primeira desenvolvida por Hunter et al. (2012), que propõem um método usando  $N_h = 2^n - 1$ , e a segunda, proposta no trabalho de Sheela e Deepa (2013), que usam  $N_h = (4n^2 + 3) / (n^2 - 8)$ , em que  $N_h$  corresponde aos melhores números de neurônios e, números de parâmetros de entrada.

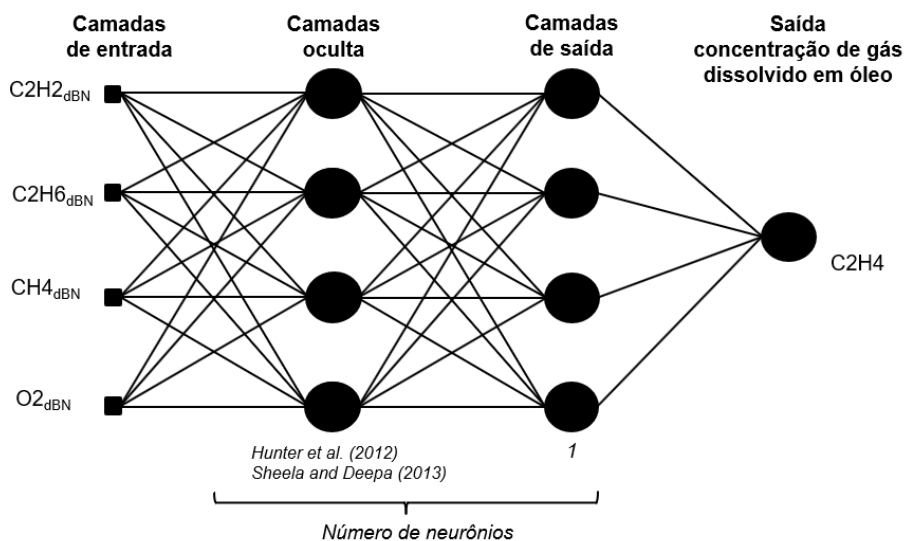
Ao contrário de Dong et al. (2019), que normalizaram os dados de entrada entre -1 e 1 para aplicar um algoritmo meta-heurístico de base populacional para otimizar a estrutura da rede neural MLP com algoritmo de propagação reversa. Foram empregados atrasos ideais feitos com a *wavelet* como entrada em um MLP com algoritmo *backpropagation*.

Para testar a relação temporal entre os gases subjacentes ao processo de degradação do óleo isolante e do material isolante sólido, foram escolhidas cinco concentrações de gases, como os principais métodos de relações IEC, relações de *Rogers e Dornenburg*, basicamente, usando  $C_2H_2$ ,  $C_2H_4$ ,  $C_2H_6$ ,  $CH_4$ ,  $H_2$  a identificar possíveis falhas do transformador de potência.

A Figura 47 mostra um exemplo da arquitetura de rede neural para prever a concentração de gás  $C_2H_4$ , em que o tempo ideal atrasa  $C_2H_2$ ,  $C_2H_6$ ,  $CH_4$  e  $O_2$ , selecionado de acordo com a correlação de Pearson e o PCA.

Em relação à saída, foram selecionadas 95 amostras relacionadas a dias ímpares para criar a matriz “A”, além de treinar e testar o modelo de previsão, pois é necessário colocar os dados de entrada e saída com o mesmo comprimento.

Figura 47 – Concentração de gás C<sub>2</sub>H<sub>4</sub> decomposta pelo *wavelet db2 a db20*



Experimentos semelhantes foram realizados para os outros gases: quando a saída é C<sub>2</sub>H<sub>2</sub>, a entrada é a aproximação de C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, CH<sub>4</sub> e O<sub>2</sub> com seus respectivos atrasos ótimos definidos pela abordagem proposta, quando a saída é C<sub>2</sub>H<sub>6</sub>, a entrada é a aproximação de *wavelet* dos gases C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, CH<sub>4</sub> e O<sub>2</sub> e assim por diante.

## 6. RESULTADOS

Primeiramente, avaliamos os resultados do teste KMO e Bartlett. Na Tabela 37, o teste KMO indicou 0,743, ao mesmo tempo que o teste de Bartlett indicou 0 e, por esse motivo, esses dados são adequados para a detecção da estrutura de dados, enquanto o teste de Bartlett indica que a análise fatorial pode ser útil para seus dados.

Tabela 37 – Teste de esfericidade KMO e Bartlett

Teste KMO e Bartlett		
Medida de adequação e amostragem KMO		0,743
	Aprox. Square-Qui	418,644
Teste de esfericidade de Bartlett	gl	45
	Sig.	0

A próxima etapa mostra o resultado da taxa de seleção e contribuição da decomposição de cada variável. A Tabela 38 mostra a taxa de contribuição por concentração de gases (normalizada).

Cada variável apresenta uma taxa de importância diferente e, na primeira ordem, C<sub>2</sub>H<sub>6</sub> possui *wavelet db20*, significa que toda a concentração de gás deve atrasar no instante  $t-20$ . Enquanto o *db8* tem mais impacto no CH<sub>4</sub>, *db18* no O<sub>2</sub> e assim por diante (Tabela 38).

Tabela 38 – Ordem de importância, taxa de cada ordem de *wavelet* para concentração de gás

Ordem Importância	ordem da Wavelet	Gás	Taxa de Importância	Ordem Importância	ordem de Wavelet	Gás	Taxa de Importância
1	db20	C2H6	1,000	11	db4	GC	0,636
2	db8	CH4	0,858	12	db20	H2	0,581
3	db18	O2	0,848	13	db8	C2H6	0,575
4	db20	CH4	0,803	14	db8	H2	0,572
5	db16	O2	0,793	15	db20	C2H4	0,568
6	db12	CO2	0,791	16	db20	H2O	0,539
7	db10	CO2	0,779	17	db16	GC	0,53
8	db6	CH4	0,776	18	db18	GC	0,529
9	db14	O2	0,692	19	db14	H2	0,507
10	db2	GC	0,644	20	db20	CO	0,495

A correlação dos atrasos de tempo para cada concentração de gás pode ser observada na Tabela 39, em que os valores marcados em negrito e itálico são os atrasos de tempo que têm maior correlação com C<sub>2</sub>H<sub>2</sub>, C<sub>2</sub>H<sub>6</sub>, C<sub>2</sub>H<sub>4</sub>, H<sub>2</sub> e CH<sub>4</sub>.

Tabela 39 – Atraso no nível de correlação

Concentração de Gás (atraso)	C2H2	C2H6	C2H4	H2	CH4
C <sub>2</sub> H <sub>2</sub> ( <i>db2</i> )	0,01254	0,00116	<b><i>0,01904</i></b>	0,00029	0,00608
C <sub>2</sub> H <sub>2</sub> ( <i>db4</i> )	0,00449	0,00116	0,01796	0,00109	<b><i>0,00689</i></b>
C <sub>2</sub> H <sub>2</sub> ( <i>db6</i> )	0,00000	0,00032	0,01232	<b><i>0,00548</i></b>	0,00490
C <sub>2</sub> H <sub>2</sub> ( <i>db20</i> )	<b><i>0,01769</i></b>	<b><i>0,00563</i></b>	0,00000	0,00137	0,00029
C <sub>2</sub> H <sub>6</sub> ( <i>db4</i> )	0,01103	0,01440	0,03349	0,01061	<b><i>0,00314</i></b>
C <sub>2</sub> H <sub>6</sub> ( <i>db6</i> )	<b><i>0,02190</i></b>	<b><i>0,02045</i></b>	<b><i>0,04203</i></b>	<b><i>0,01232</i></b>	0,00281
C <sub>2</sub> H <sub>4</sub> ( <i>db8</i> )	0,00922	0,00017	0,00073	<b><i>0,00292</i></b>	0,00026
C <sub>2</sub> H <sub>4</sub> ( <i>db10</i> )	0,00865	0,00003	0,00130	0,00044	0,00044
C <sub>2</sub> H <sub>4</sub> ( <i>db12</i> )	<b><i>0,01082</i></b>	0,00010	<b><i>0,00240</i></b>	0,00000	<b><i>0,00185</i></b>
C <sub>2</sub> H <sub>4</sub> ( <i>db20</i> )	0,00706	<b><i>0,00144</i></b>	0,00130	0,00036	0,00010
H <sub>2</sub> ( <i>db2</i> )	0,00410	0,00130	0,00410	<b><i>0,00504</i></b>	0,00102
H <sub>2</sub> ( <i>db6</i> )	0,00068	<b><i>0,00348</i></b>	0,00384	0,00026	0,00109
H <sub>2</sub> ( <i>db8</i> )	0,00130	0,00250	0,00240	0,00048	0,00017
H <sub>2</sub> ( <i>db20</i> )	<b><i>0,01300</i></b>	0,00130	<b><i>0,00608</i></b>	0,00176	<b><i>0,00270</i></b>
CH <sub>4</sub> ( <i>db4</i> )	0,00005	0,00281	0,00036	<b><i>0,00336</i></b>	0,00044
CH <sub>4</sub> ( <i>db10</i> )	<b><i>0,00922</i></b>	0,00058	0,00032	0,00023	0,00144
CH <sub>4</sub> ( <i>db14</i> )	0,00044	0,00000	0,00006	0,00012	<b><i>0,00281</i></b>
CH <sub>4</sub> ( <i>db20</i> )	0,00336	<b><i>0,00336</i></b>	<b><i>0,00212</i></b>	0,00160	0,00020

A seguir, são apresentados os resultados da previsão da rede neural para concentrações de dois gases usando oito e 15 neurônios na camada oculta, uma vez que os métodos razões

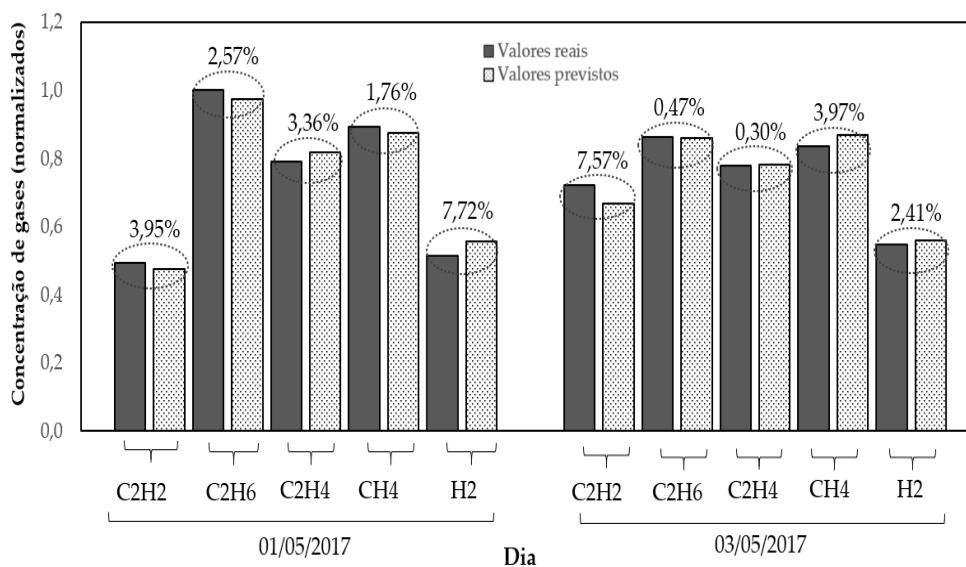
DGA IEC, razões Rogers e Dornenburg, usam basicamente os seguintes gases  $\text{CH}_4$ ,  $\text{H}_2$ ,  $\text{C}_2\text{H}_2$ ,  $\text{C}_2\text{H}_4$  e  $\text{C}_2\text{H}_6$  para analisar os problemas em potencial nos transformadores de potência.

Os resultados apresentados na Tabela 28 mostram um MAPE médio para dois dias de 1,525% para  $\text{C}_2\text{H}_6$  e 1,831% para  $\text{C}_2\text{H}_4$ . A Figura 48 compara os valores previstos com os valores reais para as cinco concentrações de gás. Como pode ser visto, a seleção do atraso de tempo ideal na concentração de gás pode melhorar a precisão da previsão, ao comparar as previsões com as variáveis de entrada ao mesmo tempo  $t-2$  e  $t-4$  (Tabela 40).

**Tabela 40 – Valores previstos com e sem a seleção do melhor tempo de atraso**

Número de neurônios	Entrada/Data	Concentração do gás $\text{C}_2\text{H}_6$			Concentração do gás $\text{C}_2\text{H}_4$		
		05/01/2017	05/03/2017	MAPE% médio	05/01/2017	05/03/2017	MAPE% médio
8 neurônios	Real	1	0,864	-	0,791	0,779	-
	Seleção do atraso de tempo	0,972	0,756	7,645	0,789	0,657	7,891
	$t-2$	0,842	0,849	8,811	0,749	0,753	4,308
	$t-4$	0,769	0,889	13	0,786	0,847	4,64
15 neurônios	Seleção do atraso de tempo	0,974	0,86	<b>1,525</b>	0,817	0,781	<b>1,831</b>
	$t-2$	0,818	0,919	12,294	0,794	0,805	1,909
	$t-4$	0,83	0,864	8,492	0,65	0,995	22,777

**Figura 48 – Valores reais e previstos da concentração de gás por dois dias**



## 7. DISCUSSÃO

Este estudo teve como objetivo estudar o atraso ótimo de cada concentração de gás que afeta os gases  $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  e  $C_2H_6$  (Tabela 38 e Tabela 39), nos quais posteriormente pode ser utilizada uma técnica DGA para detectar o defeito na potência transformador.

A abordagem usando transformação do tipo *wavelet* e SPCA mostra a taxa de contribuição com diferentes atrasos de tempo de cada concentração de gás, o que difere da proposta de trabalhos recentes (LIU et al., 2019; PEREIRA et al., 2018). Em Pereira et al. (2018), por exemplo, apesar de testar diferentes funções da *wavelet* e diferentes atrasos, todos os modelos adotaram o mesmo atraso de tempo para variáveis externas. Essa abordagem mostra a taxa e a ordem de importância e a ordem de *Wavelet* para dez concentrações de gases (Tabela 38), indicando que o db20 ( $t-20$ ), db8 ( $t-8$ ) e db18 ( $t-18$ ) são os três mais atrasos importantes para as concentrações de gás  $C_2H_6$ ,  $CH_4$  e  $O_2$ , respectivamente. Esse resultado mostra que o efeito que determinado gás sofre com outros gases varia ao longo do tempo de maneira diferente para cada gás.

A correlação de Pearson foi utilizada para considerar o impacto de cada atraso de tempo, usando diferentes atrasos de tempo  $t-2$  a  $t-20$  em cada concentração de gás, mostrando que, por exemplo, para prever as concentrações de  $C_2H_2$ , os melhores atrasos das concentrações dos gases são:  $t-12$  para  $C_2H_4$ ,  $t-6$  para  $C_2H_6$ ,  $t-10$  para  $CH_4$  e  $t-8$  para  $H_2$ .

É importante destacar que o modelo autorregressivo tradicional que adota o mesmo atraso para todas as variáveis não teria identificado essa relação. Além disso, este resultado é muito importante para a calibração dos sistemas de monitoramento, pois indica que qualquer variação no  $C_2H_4$ , por exemplo, levará cerca de 12 unidades de tempo para refletir sobre a concentração de  $C_2H_2$ . Análise semelhante se aplica a outros gases.

Um tipo semelhante de relação de diferentes gases foi estudado em Lu et al. (2019) e Wang et al. (2015). Em Lu et al. (2019), os autores estudaram a correlação entre as cinco concentrações de gás, aplicando o valor do grau relacional cinza para revelar as relações entre as características do gás. Esses autores mostraram que a análise da relação cinza é eficiente na seleção e remoção de recursos redundantes do conjunto de variáveis de entrada. No entanto, os autores não consideram nenhum atraso na amostragem das séries de entrada de concentrações de gases. Por outro lado, Wang et al. (2015) utilizaram coeficientes de correlação da concentração de gás CO como parâmetro característico constante para a correlação das análises de séries temporais e  $H_2$ ,  $CH_4$ ,  $C_2H_2$ ,  $C_2H_4$  e  $C_2H_6$  como parâmetros característicos variáveis a serem utilizados para distinguir falhas elétricas de falhas térmicas.

Entretanto, abordagens baseadas em modelos autorregressivos aplicam a mesma ordem para todas as variáveis de entrada e não levam em consideração a relação de atraso de tempo entre as concentrações de gás. Não obstante, vimos que a seleção ideal do atraso de tempo para cada concentração de gás afeta a saída.

Em relação à acurácia da previsão, essa abordagem mostra algumas previsões melhores do que Liu et al. (2018), Lu et al. (2019) e Wang et al. (2015) (Tabela 41).

**Tabela 41 – Comparação das concentrações previstas de gás**

MAPE médio (%)						
Autores	Tipo de técnicas	C <sub>2</sub> H <sub>2</sub>	C <sub>2</sub> H <sub>4</sub>	C <sub>2</sub> H <sub>6</sub>	CH <sub>4</sub>	H <sub>2</sub>
Wang et al., 2015	Correlação de séries temporais	38,900	42,100	22,200	42,100	11,100
Lin et al., 2018	LSTM_DBN Network	2,450	1,450	2,100	0,260	1,890
Lu et al., 2018	<i>ANN, SVM, LSSVM and Gaussian process regression</i>	6,433	7,375	5,913	5,500	6,313
Zhang et al., 2018	RBFNN	4,310	5,670	5,520	3,940	4,640
	LSSVM(RBF)	3,960	5,420	2,330	1,690	3,130
Pereira et al., 2018	NAR-DWT	0,330	0,320	0,290	0,100	0,460
Liu et al., 2019	Wavelet Least SVM and Imperialist Competition Algorithm	4,168	0.1684	1,993	0.9675	1,854
Esse trabalho	Transformação tipo <i>wavelet MLP neural network</i>	5,763	1,831	1,525	2,869	5,069

É importante destacar o baixo custo computacional do modelo proposto, pois o modelo usado neste artigo leva alguns segundos para ser executado. No exemplo acima, com relação à previsão da concentração de C<sub>2</sub>H<sub>2</sub>, as variáveis de entrada C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, CH<sub>4</sub> e H<sub>2</sub> têm 12, 6, 10 e 8 variáveis de entrada, já que o atraso ideal para prever a concentração de gás C<sub>2</sub>H<sub>2</sub> é  $t-12$ ,  $t-6$ ,  $t-10$  e  $t-8$ , respectivamente, conforme definido na Equação 52, foi utilizado simplesmente nas quatro aproximações correspondentes criadas pela transformação tipo *wavelet* para cada gás exógeno.

## 8. CONCLUSÕES

Este trabalho apresentou uma abordagem baseada na transformação do tipo *wavelet* para determinar o atraso de tempo ideal para cada concentração de gás e usá-lo como modelo de regressão em rede MLP.

A aplicação da transformação do tipo *wavelet* cria aproximações esparsas das concentrações de gases com diferentes atrasos de tempo e permite definir um modelo de previsão preciso e computacionalmente eficiente. O modelo de previsão tem alta precisão, pois se concentra na tendência dos sinais de entrada das aproximações sem ruído calculadas pela transformação *wavelet*. Além disso, reduz o número de variáveis de entrada no modelo de regressão automática ao usar as aproximações resultantes de transformações do tipo *wavelet* com diferentes comprimentos, já que consideram o atraso de tempo determinado para cada variável.

Além de simplesmente avaliar a relação de diferentes gases, a proposta possibilita identificar a relação entre um gás específico e as aproximações atrasadas de todos os outros gases. Esse resultado é fundamental para os sistemas de monitoramento, pois indica que as variações em diferentes gases levam tempos diferentes para serem refletidas nas concentrações de outros gases sob observação. Como exemplo, observou-se que as variações no etano afetam as concentrações de acetileno mais rapidamente do que as variações no etileno. Assim, a abordagem amplia a discussão da relação temporal entre as concentrações de gases, proporcionando expansão do conhecimento da relação temporal entre os gases subjacentes ao processo de degradação do óleo isolante e do material isolante sólido.

**Contribuições dos autores:** F.E.B. e F.H.P., conceitualização, metodologia, validação, análise formal, investigação, criação e escrita de dados - preparação do rascunho original; F.H.P., F.A.Z.G., S.I.N. e G.F.M.S., supervisão e redação - revisão e edição, F.H.P e S.I.N. Administração de projetos; validação, redação de I.E.C, J.C.S e S.N.J. revisão e edição. Todos os autores leram e concordaram com a versão publicada do manuscrito.

**Conflitos de interesse:** Os autores declaram não haver conflito de interesse. Os patrocinadores não tiveram nenhum papel no desenho, execução, interpretação ou redação do estudo

**Nomenclatura:**

ANFIS - Inferência Neuro Difusa Adaptável

BPNN - Rede Neural com algoritmo *backpropagation*

C<sub>2</sub>H<sub>2</sub> - Gás acetileno

C<sub>2</sub>H<sub>4</sub> - Gás etileno

C<sub>2</sub>H<sub>6</sub> - Gás etano

CH<sub>4</sub> - Gás metano

CO - Gás monóxido de carbono

CO<sub>2</sub> - Gás dióxido de carbono  
 Db - *Daubechies wavelets*  
 DGA - Análise de Gás Dissolvido  
 DGC - Concentração de gás dissolvido  
 DWT - Transformação de *wavelet* discreta  
 FIS - Sistema de Inferência *Fuzzy*  
 GE - *General Electric*  
 H<sub>2</sub> - Gás hidrogênio  
 IEC - Comissão Eletrotécnica Internacional  
 KMO – *Kaiser-Meyer-Olkin*  
 LS-SVM - Suporte para mínimos quadrados da máquina vetorial  
 LSTM - Memória de curto prazo  
 MAPE - Erro médio percentual absoluto  
 MLP - *Perceptron* de múltiplas camadas  
 NARX - Não-linear autorregressivo exógeno  
 PCA - Análise de Componentes Principais  
 PTE - Transformadores de potência  
 SPCA - Análise de componentes principais esparsos da  
 SVM - Máquina de vetores de suporte  
 TDCG - Gás Combustível Total Dissolvido  
 W-LSSVR - Os mínimos quadrados da máquina vetorial suportam regressão *Wavelet*  
 WT – Transformação do tipo *wavelet*  
 $\psi_{j,k}(t)$  - *Wavelet* mãe  
 $\varphi_{m,k}(t)$  - Função de escala  
 H( $\omega$ ) - Função de transferência do filtro passa-alta  
 G( $\omega$ ) - Função de transferência do filtro passa-baixa  
 $h_n$  - Coeficientes passa-alta  
 $g_n$  - Coeficientes passa-baixa  
 $A_{m,k}$  - Perfil aproximado  
 $D_{j,k}$  - Detalhe do perfil  
 $j, m$  - Nível de decomposição da transformada *wavelet*.

## REFERÊNCIAS

ABU BAKAR, N.; ABU-SIADA, A.; ISLAM, S. A review of dissolved gas analysis measurement and interpretation techniques. **IEEE Electrical Insulation Magazine** 2014, 30 (3), pp. 39-49.

BACHA, K.; SOUAHLIA, S.; GOSSA, M. Power transformer fault diagnosis based on dissolved gas analysis by support vector machine. **Electric Power Systems Research** 2012, 83, 73–79. 10.1016/j.epsr.2011.09.012.

BARTLETT, M. S. Properties of sufficiency and statistical tests. **Proceedings of the Royal Statistical Society** 1937 - Serie A, 60:268-282, 1937.

BOX, G.E.P., JENKINS, G.M., **Time Series Analysis: Forecasting and Control**, New Jersey: HoldenDay, 2015. 756p.

CERNY, C.A.; KAISER, H.F. (1977), **A study of a measure of sampling adequacy for factor-analytic correlation matrices**. *Multivariate Behavioral Research* 1977, 12(1), 43-47.

TANG, W. H., WU, Q. H. (2013). **Condition Monitoring and Assessment of Power Transformers Using Computational Intelligence**. Springer Publishing Company, Incorporated.

CHEN, Z.; ZHANG, X.; XIONG, H.; CHEN, D.; CHENG, H.; TANG, J.X.; TIAN, Y.; XIAO, S. Dissolved Gas Analysis in Transformer Oil Using Pt-Doped WSe<sub>2</sub> Monolayer Based on First Principles Method. **IEEE Access** 2019, 7, 72012-72019.

CHENG, L.; YU, T. **Dissolved Gas Analysis Principle Based Intelligent Approaches to Fault Diagnosis and Decision Making of Large Oil-Immersed Power Transformers: A Survey**. Preprints 2018, 2018040109 (doi:10.20944/preprints201804.0109.v1).

CIGRE. **Guide for Transformer Maintenance**; CIGRE: Paris, France, 2011; Volume 445.

HUNTER, D.; YU, H.; PUKISH III, M. S.; KOLBUSZ, J.; WILAMOWSKI, B. M. “Selection of proper neural network sizes and architectures: a comparative study,” **IEEE Transactions on Industrial Informatics** 2012, vol. 8, no. 2, pp. 228–240, 2012.

DAUBECHIES I. Orthonormal bases of compactly supported *wavelets*. **Commun. Pure Appl. Math.**, 41, pp. 909-996, 1988.

DAUBECHIES, I.; SWELDENS, W. Factoring *wavelet* transforms into lifting steps. **The Journal of Fourier Analysis and Applications**, 4, 247–269, 1998.

DE OLIVEIRA JR, J. I.; DA ROCHA, J. C. F.; GUIMARAES, A. M.; DA FONSECA, A. F. (2015). A PCA and SPCA based procedure to variable selection in agriculture. **Revista Brasileira de Computação Aplicada** 2015, 7 (1), 30-41.

DONG, H.; YANG, X.; LI, A.; XIE, Z.; ZUO, Y. Bio-Inspired PHM Model for Diagnostics of Faults in Power Transformers Using Dissolved Gas-in-Oil Data. **Sensors** 2019. 19. 845. 10.3390/s19040845.

EL-HOSHY, S.; GOUDA, O.; EL-TAMALY, H. (2019). Condition Assessment of Power Transformers Based on Dissolved Gas Analysis. **IET Generation, Transmission & Distribution** 2019. 13. 10.1049/iet-gtd.2018.6168.

FARIA JUNIOR, H.; COSTA, J.; OLIVAS, J. (2015). A review of monitoring methods for predictive maintenance of electric power transformers based on dissolved gas analysis. **Renewable and Sustainable Energy Reviews** 2015. 46. 10.1016/j.rser.2015.02.052.

GOGOLEWSKI, D. Influence of the edge effect on the *wavelet* analysis process. **Measurement** 2020, 152, 107314.

ZHENG, H.; ZHANG, Y.; LIU, J.; WEI, H.; ZHAO, J.; LIAO, R. A novel model based on *wavelet* LS-SVM integrated improved PSO algorithm for forecasting of dissolved gas contents in power transformers. **Electric Power Systems Research** 2018. 155. 10.1016/j.epsr.2017.10.010.

JENSEN, A.; LA COUR-HARBO, A. The Discrete *Wavelet* Transform, **Ripples in Mathematics**, Springer, Berlin 2001.pp-257.

LIN, J.; LEI, S.; YINGJIE, Y.; GEHAO, S.; DA, X.; XIUCHEN, J. Prediction Method for Power Transformer Running State Based on LSTM\_DBN Network. **Energies**, MDPI, Open Access Journal 2018, vol. 11(7), pages 1-14, July.

LI, J.; QIAOGEN, Z.; WANG, K.; WANG, J.; ZHOU, T.; ZHANG, Y. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. **IEEE Transactions on Dielectrics and Electrical Insulation** 2016. 23. 1198-1206. 10.1109/TDEI.2015.005277.

- LIN, C.H.; CHEN, J.L.; HUANG, P.Z. (2011). Dissolved gases forecast to enhance oil-immersed transformer fault diagnosis with grey prediction-clustering analysis. **Expert Systems** 2011. 28. 123-137. 10.1111/j.1468-0394.2010.00542.x.
- LIU, J.; ZHENG, H.; ZHANG, Y.; LI, X.; FANG, J.; LIU, Y.; LIAO, C.; LI, Y.; ZHAO, J. Dissolved Gases Forecasting Based on *Wavelet* Least Squares Support Vector Regression and Imperialist Competition Algorithm for Assessing Incipient Faults of Transformer Polymer Insulation. **Polymers** (Basel) 2019,11(1):85. Published 2019 Jan 8. doi:10.3390/polym11010085.
- LIU, C.; ZHANG, H.; XIE, Z.; WANG, Z.; CHEN, L.; LIN, X.; DING, H.; LI, Z.; XIONG, W. (2018). Combined Forecasting Method of Dissolved Gases Concentration and Its Application in Condition-Based Maintenance. **IEEE Transactions on Power Delivery** 2018. PP. 1-1. 10.1109/TPWRD.2018.2881747.
- LU, S. X.; LIN, G.; QUE, H.; LI, M. J. J.; WEI, C. H.; WANG, J. K. Grey relational analysis using Gaussian process regression method for dissolved gas concentration prediction. **Int. J. Mach. Learn. & Cyber** 2019. 10, 1313–1322. <https://doi.org/10.1007/s13042-018-0812-y>.
- NOORI, M.; EFFATNEJAD, R.; HAJIHOSSEINI, P. (2017). Using Dissolved Gas Analysis Results to Detect and Isolate the Internal Faults of Power Transformers by Applying a *Fuzzy* Logic Method. **IET Generation, Transmission & Distribution** 2017. 11. 10.1049/iet-gtd.2017.0028.
- RAHMAN, A. P.; KARUNIKA, D.; SUWARNO.; HARRY, G. 2017. "Transformer Paper Expected Life Estimation Using ANFIS Based on Oil Characteristics and Dissolved Gases (Case Study: Indonesian Transformers)," **Energies**, 2018, MDPI, Open Access Journal, vol. 10(8), pages 1-18, August.
- RHIF, M.; ABBES, A.; FARAH, I.; MARTINEZ, B.; SANG, Y. F. (2019). *Wavelet* Transform Application for/in Non-Stationary Time-Series Analysis: A Review. **Applied Sciences** 2019. 9. 1347. 10.3390/app9071345.
- SCHIMIT, P. H.; PEREIRA, F. H. Disease spreading in complex networks: A numerical study with Principal Component Analysis. **Expert Systems with Applications** 2018, 97, 41-50.
- SHEELA, K.G.; DEEPA, S.N.; Review on methods to fix number of hidden neurons in neural networks. **Math Prob Eng** 2013 (Article ID 425740, 11 pages).

CHAKRAVORTI, Sivaji.; DEY, Debangshu. Chatterjee, Biswendu. (2013). **Recent Trends in the Condition Monitoring of Transformers: Theory, Implementation and Analysis**. 10.1007/978-1-4471-5550-8.

WANG, X.; YANG, Q.; LI, R.; ZHANG, C.; LI, Y. Zhang, "Diagnosis of solid insulation deterioration for power transformers with dissolved gas analysis-based time series correlation", **IET Sci. Meas. Technol.**, vol. 9, no. 4, pp. 393-399, 2015.

PEREIRA, F.H.; BEZERRA, F. E.; JUNIOR, S.; SANTOS, J.; CHABU, I.; SOUZA, G.F.M.; MICERINO, F.; NABETA, S.I. Nonlinear Autoregressive Neural Network Models for Prediction of Transformer Oil-Dissolved Gas Concentrations. **Energies** 2018, 11, 1691.

BUSTAMANTE, S.; MANANA, M.; ARROYO, A.A.; Castro, P.; Laso, A.; and Martínez, R. "Dissolved Gas Analysis Equipment for Online Monitoring of Transformer Oil: A Review." **Sensors** (2019).

DIACONESCU, E. "The use of NARX neural networks to predict chaotic timeseries", *WSEAS Transactions on Computer Research* 2008, 3 (3), 182–191.

LIN, T., HORNE, B.G., TINO, P., GILES, C.L. "Learning long-term dependencies in NARX recurrent neural networks", **IEEE Transactions on Neural Networks**, 1996, 7(6), 1329-1351

GAO, Y., MENG, J. E. "NARMAX time series model prediction: feedforward and recurrent *fuzzy* neural network approaches", **Fuzzy Sets and Systems**, 2005, 150(2), 331-350.

LIN, T., GILES, C.L., HORNE, B.G., KUNG, S.Y. "A Delay Damage Model Selection Algorithm for NARX Neural Networks", *IEEE Transactions on Signal Processing*, "Special Issue on Neural Networks", 1997, 45(11), 2719-2730.

SIEGELMANN, H. T., HORNE, B. G. and GILES, C. L. "Computational capabilities of recurrent NARX neural networks", **IEEE Transactions on Systems, Man and Cybernetics, Part B**, 1997, 27(2), 208-215.