UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

FRANCISCO CESARIO NETO

DESENVOLVIMENTO DE ESTRATÉGIAS PARA TRATAMENTO E PREVISÃO DE DADOS, USANDO SÉRIES TEMPORAIS, BASEADAS EM TÉCNICAS ESTATÍSTICAS E DE SISTEMAS INTELIGENTES, PARA ESTAÇÕES DE TRATAMENTO DE EFLUENTES LÍQUIDOS

Recife
2018

FRANCISCO CESARIO NETO

DESENVOLVIMENTO DE ESTRATÉGIAS PARA TRATAMENTO E PREVISÃO DE DADOS, USANDO SÉRIES TEMPORAIS, BASEADAS EM TÉCNICAS ESTATÍSTICAS E DE SISTEMAS INTELIGENTES, PARA ESTAÇÕES DE TRATAMENTO DE EFLUENTES LÍQUIDOS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Engenharia em Engenharia Química.

Área de concentração: Engenharia Ambiental.

Orientador: Profº. Mohand Benachour

Coorientadores: Profª. Dra. Silvana Carvalho de Souza Calado

Profº. Dr. Frede de Oliveira Carvalho

Recife
2018

Catalogação na fonte
Bibliotecária Maria Luiza de Moura Ferreira, CRB-4 / 1469

C421d Cesario Neto, Francisco.
Desenvolvimento de estratégias para tratamento e previsão de dados, usando séries temporais, baseadas em técnicas estatísticas e de sistemas inteligentes, para estações de tratamento de efluentes líquidos / Francisco Cesario Neto. - 2018.
212 folhas, il., tabs., abr., sigl. e símb.

Orientador: Prof. Dr. Mohand Benachour.
Coorientadores: Profª. Dra. Silvana Carvalho de Souza Calado.
Dr. Frede de Oliveira Carvalho.

Tese (Doutorado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia Química, 2018.
Inclui Referências e Apêndices.

1. Engenharia Química. 2. Análise de componentes principais. 3. Redes neurais artificiais. 4. Rede neural convolutiva. 5. Tratamento biológico de efluentes. 6. Wavelet. I. Benachour, Mohand (Orientador). II. Calado, Silvana Carvalho de Souza (Coorientadora). III. Carvalho, Frede de Oliveira (Coorientador). IV. Título.

UFPE

660.2 CDD (22. ed.) BCTG/2018-399

FRANCISCO CESARIO NETO

DESENVOLVIMENTO DE ESTRATÉGIAS PARA TRATAMENTO E PREVISÃO DE DADOS, USANDO SÉRIES TEMPORAIS, BASEADAS EM TÉCNICAS ESTATÍSTICAS E DE SISTEMAS INTELIGENTES, PARA ESTAÇÕES DE TRATAMENTO DE EFLUENTES LÍQUIDOS

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Engenharia Química.

Aprovada em: 30/07/2018.[S1]

BANCA EXAMINADORA

_____

Profº. Dr. Mohand Benachour (Orientador)
Universidade Federal de Pernambuco -DEQ

_____

Profª. Drª.Silvana Carvalho de Souza Calado (Co-orientadora)
Universidade Federal de Pernambuco -DEQ

_____

Profº. Dr. Frede de Oliveira Carvalho (Co-orientador)
Universidade Federal de Alagoas -DEQ

_____

Profº. Dr. Gilson Lima da Silva (Examinador Externo)
Universidade Federal de Pernambuco -DEP

_____

Profº. Dr. José Marcos Francisco da Silva (Examinador Interno)
Universidade Federal de Pernambuco -DEQ

_____

Profº. Dr. Nelson Medeiros de Lima Filho (Examinador Interno)
Universidade Federal de Pernambuco -DEQ

_____

Profª. Dra. Otidene Rossinter Sá Rocha (Examinadora Interna)
Universidade Federal de Pernambuco -DEQ

_____

Profª. Dra. Yana Batista Brandão (Examinadora Externa)
Universidade Federal Rural de Pernambuco -UACSA

Dedico este trabalho, a Deus, a minha família, meus alunos, meus professores e amigos.

AGRADECIMENTOS

À Deus por tudo de maravilhoso que ele criou em mim e na natureza.

Aos meus pais, Costantino Cesario (em memória) e Ivanilda Martins Cesario (minha mãe, fã e também meu ídolo) e minhas filhas e netos.

A todos meus professores desde o Jardim da infância até a pós-graduação e todos o que me ajudaram na execução deste trabalho, em especial ao professor Mohand Benachour ao qual sem seu incentivo jamais teria chegado a este fim. Ficando sempre associada a ele a frase, quando estava prestes a desistir de tentar o doutorado “Não desista Chico vai em frente até o fim” ... e até o fim segui, graças ao seu incentivo. Também agradeço muito a professora Silvana Calado que se dispôs a me ajudar, facilitando meu ingresso no grupo de pesquisa

A todos os meus amigos por todos os momentos de aprendizado e troca de ideias, em especial Juan Felipe Alba e Josivan Pedro da Silva

A Simone Guitel Fleischman

Ao Prof. Frede Carvalho por todas as orientações em redes neurais.

Ao Prof. José Amigo pelo ensino da programação em MATLAB, fundamental para adaptar meus dados aos programas exemplos e obter resultados.

Aos colegas funcionários e ex-funcionários do grupo internacional de Meio-ambiente White Martins Brasil e da Neotex Soluções Ambientais, que na troca de ideias e experiências operações de plantas e cursos ministrados, que tanto me ensinaram.

A Jacques André Conchon (em memória)

Aos colegas professores do Instituto Federal de Ensino de Pernambuco no incentivo e cobertura da realização deste trabalho.

A SOLAR engarrafadora de bebidas pelos dados fornecidos e de grande importância para o desenvolvimento deste trabalho

Por fim, a todos que fizeram parte da minha caminhada, vocês contribuíram de forma ímpar para o meu desenvolvimento e a concretização deste sonho.

RESUMO

A operação de uma planta de tratamento de efluentes tem características próprias, especialmente em se tratando de processos biológicos. Os processos industriais semelhantes, com os mesmos projetos de estação, têm as suas características afetadas por alguma variável intrínseca, específica daqueles processos. Assim, as variáveis e ajustes de operação, para melhor eficiência de uma planta de tratamento de efluente biológico são realizadas dentro de faixas específicas e de acordo com as características desta planta, as quais envolvem uma série de análises físico-químicas, ajustes de equipamentos e de processos. Normalmente, quando uma ou mais variáveis saem do controle, podem ou não afetar outras variáveis. Por exemplo, se uma planta apresenta perda de sólidos em suspensão na saída de um decantador secundário, em um processo de tratamento biológico com lodos ativados, podem acontecer várias causas e ajustes para esta variável fora de controle, tais como sobrecarga hidráulica no decantador, sobrecarga de sólidos, desnitrificação, oxigenação inadequada no reator biológico, presença de microrganismos filamentosos, quebra de raspador mecânico, idade de lodo muito elevada, excesso de densidade de potência no reator biológico e fator de carga inadequado. Muitas vezes, o operador realiza o ajuste de uma variável por vez e espera os resultados, para que então se possa fazer o ajuste em outra variável, gerando demora no ajuste da planta e riscos ambientais. A interferência de uma variável sobre a outra pode gerar ou não consequências de perda de eficiência no tratamento biológico. Assim, fazer a previsão ou uma simulação do que ocorreria com a mudança destas variáveis, considerando as características intrínsecas de uma estação, seria uma ferramenta de grande valia no controle operacional. Neste trabalho, foram utilizadas algumas ferramentas estatísticas convencionais (média, média móvel, média móvel dupla, amortecimento exponencial simples, amortecimento exponencial duplo, modelo autoregressivo, integrativo com média móvel), para previsão em séries temporais e também técnicas que envolvem Inteligência artificial como redes neurais artificiais (RNA), vetor suporte de máquina (VSM), suavização por wavelet associada a RNA, redes neurais convolutivas (RNC) de aprendizado profundo, se buscando verificar os resultados, que mais se aproximem de uma

previsão futura e também uma classificação dos resultados da demanda química de oxigênio (DQO) de saída de uma estação de tratamento de efluentes (ETE) e uso de técnicas como avaliação dos componentes principais (ACP), classificando e correlacionando as variáveis com os resultados, conseguindo valores de previsão, com erros na faixa entre 24 a 30% de erro médio relativo, valores bem abaixo do desvio padrão (56,7%), apresentado para esta variável no período e classificação de previsão futura acima de 83% de acurácia. Com as técnicas estatísticas e de inteligência artificial é possível ter noção tanto quantitativa, quanto qualitativa, do que ocorrerá um passo à frente ou no dia seguinte com a DQO de saída de um reator anaeróbio, e também simular o que ocorrerá ao se alterar uma das diversas variáveis de controle de uma estação de tratamento de efluentes.

Palavras-Chave**:** Análise de componentes principais. Redes neurais artificiais. Rede neural convolutiva. Séries temporais. Tratamento biológico de efluentes. Wavelet.

ABSTRACT

The operation of an effluent treatment plant has its own characteristics, especially when dealing with biological processes. Similar industrial processes, with the same station designs, have their characteristics affected by some intrinsic variable, specific to those processes. Thus, the operating variables and adjustments for better efficiency of a biological effluent treatment plant are carried out within specific ranges and according to the characteristics of this plant, which involve a séries of physical-chemical analyzes, equipment adjustments and of processes. Normally, when one or more variables go out of control, they may or may not affect other variables. For example, if a plant exhibits loss of suspended solids at the outlet of a secondary decanter, in a biological treatment process with activated sludges, various causes and adjustments for this out-of-control variable can occur, such as hydraulic overload in the decanter, overload of solids, denitrification, inadequate oxygenation in the biological reactor, presence of filamentous microorganisms, mechanical scraper cracking, very high sludge age, excess of power density in the biological reactor and inadequate relationship between food and microorganism. Often, the operator performs the adjustment of one variable at a time and waits for the results, so that the adjustment can be made in another variable, generating delay in the adjustment of the plant and environmental risks. The interference of one variable over the other may or may not lead to loss of efficiency in biological treatment. Thus, forecasting or a simulation of what would happen with the change of these variables, considering the intrinsic characteristics of a station, would be a valuable tool in operational control. In this work, we used some conventional statistical tools (mean, moving average, double moving average, simple exponential damping, double exponential damping, autoregressive model, integrative with moving average), for time séries forecasting and also techniques that involve artificial intelligence as neural networks (NN), support vector machine (SVM), wavelet smoothing associated with RNA, deep learning convolutional neural networks (CNN), seeking to verify the results, which are closer to a future forecast and also a classification of results of the chemical oxygen demand (COD) from an effluent treatment plant (ETP) and use of techniques such as principal component analysis (PCA), classifying and

correlating the variables with the results, achieving predicted values with errors in the range of 24 to 30% relative mean error, values well below the deviation (56.7%), presented for this variable in the period and future forecast classification above 83% of accuracy. With statistical and artificial intelligence techniques it is possible to have both quantitative and qualitative notions of what will happen one step ahead or the next day with the exit COD of an anaerobic reactor, and also to simulate what will happen when one of the control variables of an effluent treatment plant.

Keywords: Principal component analysis. Artificial neural networks. Biological treatment of effluents. Convolutive neural network. Time séries. Wavelet.

LISTA DE FIGURAS

Figura 1 - Desenho esquemático de um processo com lodos ativados..... 37

Figura 2 - Os caminhos de reação bioquímica descritos em Modelo de Degradação Biológica 1 (ADM1) do International Water Association (IWA), com os seguintes grupos microbianos:(1) degradadores de açúcar, (2) degradadores de aminoácidos, (3) degradadores de ácidos graxos de cadeia longa (4) degradadores de ácido propiònico, (5) degradadores de ácido butírico e valérico, (6) metanógenos acetoclásticos e (7) metanógenos hidrogenotróficos..... 39

Figura 3 - Série temporal hipotética e sua decomposição para previsão..... 47

Figura 4 - Ilustração da possível divergência de uma previsão extrapolada..... 58

Figura 5 - Transformação de série não estacionária em série estacionária usando a primeira diferença..... 61

Figura 6 - Neurônio Biológico..... 63

Figura 7 - Conexões sinápticas entre neurônios..... 64

Figura 8 - Representação esquemática comparativa entre o neurônio real e neurônio artificial..... 65

Figura 9 - Esquema básico de uma rede neural, mostrando os bias e os pesos sinápticos..... 66

Figura 10 - Tipologia de uma rede neural tipo Multilayer Perceptron (MLP)..... 68

Figura 11 - Representação gráfica da função Erro x Pesos sinápticos..... 69

Figura 12 - Função de ativação degrau..... 70

Figura 13 - Função de ativação degrau bipolar..... 70

Figura 14 - Função de ativação rampa simétrica..... 71
Figura 15 - Função de ativação Logística ou sigmóide..... 71
Figura 16 - Função de ativação Hiperbólica..... 71
Figura 17 - Função de ativação Gaussiana..... 72
Figura 18 - Função de ativação Linear..... 72
Figura 19 - Interpretação geométrica da regra delta..... 73
Figura 20 - Regra delta considerando dois pesos..... 73
Figura 21 - Margem de separação entre hiperplanos nas fronteiras das classes positiva (quadrado preenchido) e negativa (círculo preenchido). No SVM queremos que a margem M seja maximizada. Os eixos ordenados x1 e x2 representam as dimensões das amostras no espaço 2..... 76
Figura 22 - As variáveis de folga e o hiperplano separador ótimo (OSH)..... 77
Figura 23 - Transformação: problema não linearmente..... 78
Figura 24 - Feature (características) ou classificador (Kernel) treináveis..... 82
Figura 25 - Dados esperados (linha verde) versus Dados com ruído (linha vermelha) 86
Figura 26 - Curva suavizada em azul obtida pelos dados com ruído da Figura 25, versus curva desejada em verde feito por Interpolação polinomial de primeiro grau de 11 pontos sequenciados por vez..... 87
Figura 27 - Curva suavizada em azul obtida pelos dados com ruído da Figura 25 versus curva desejada em verde, feito por Interpolação polinomial de segundo grau de 11 pontos sequenciados por vez..... 87
Figura 28 - Curva suavizada em azul versus curva desejada em verde feito por Interpolação polinomial de segundo grau de 21 pontos sequenciados por vez..... 88
Figura 29 - Projeção ortogonal do vetor υ no vetor υ 0..... 90
Figura 30 - Esquema de transformação de tempo em frequência por transformada de Fourier..... 92

Figura 31 - Esquema das janelas de divisão tempo frequência da transformada de Fourier Janelada..... 94
Figura 32 - Esquema das janelas de divisão tempo frequência da transformada Wavelet..... 94
Figura 33 - Famílias wavelets..... 96
Figura 34 - Processo de decomposição de um sinal em filtros passa alta e filtros passa baixa..... 97
Figura 35 - Representação geométrica das componentes principais..... 99
Figura 36 - Série temporal-valores observados e previstos usando KSOM-ANFIS..... 104
Figura 37 - a) Correlação para previsão de apenas um dia a frente da DQO, usando rede neural. b) Gráfico com previsão para um dia à frente..... 106
Figura 38 - Arquitetura de EBPN para previsão..... 107
Figura 39 - Variação do erro médio quadrático com a previsão futura..... 109
Figura 40 - Previsão da vazão de entrada sem defasagem de tempo..... 109
Figura 41 - Previsão da vazão de entrada com defasagem de 30 min..... 110
Figura 42 - Previsão com defasagem de 180 minutos..... 110
Figura 43 - Esquema usado para previsão n passos à frente, usando os últimos cinco dias..... 112
Figura 44 - Fluxograma com as etapas executadas para o desenvolvimento da metodologia..... 114
Figura 45 - Esquema com localização dos pontos de amostragem e as análises..... 116
Figura 46 - Representações gráficas de matrizes transformadas em Figuras, em três tons de cinza, com 10 linhas, representando dez dados das variáveis de controle de uma ETE e cinco linhas, representando cinco dias anteriores..... 120
Figura 47 - Tela do software mostrando as variações, em distribuição normal, da DQO do tanque de equalização..... 124

Figura 48 - Tela do software, mostrando as variações da distribuição normal, da vazão de entrada do Tanque de Metanização (TM)..... 124
Figura 49 - Tela do software mostrando as variações da carga orgânica aplicada no Tanque de Metanização (TM)..... 125
Figura 50 - Tela do software mostrando as variações do pH de entrada do Tanque de Metanização (TM)..... 125
Figura 51 - Tela do software mostrando as variações na distribuição normal, da Alcalinidade no ponto 2 do Tanque de Metanização (TM)..... 126
Figura 52 - Tela do software mostrando as variações na distribuição normal, dos ácidos graxos voláteis no Tanque de Metanização (TM)..... 126
Figura 53 - Tela do software mostrando as variações na distribuição normal da relação ácidos graxos voláteis e alcalinidade do Tanque de Metanização (TM)..... 127
Figura 54 - Tela do software mostrando as variações do pH do ponto 2 no Tanque de Metanização (TM)..... 127
Figura 55 - Tela do software mostrando as variações da temperatura no Tanque de Metanização (TM)..... 128
Figura 56 - Tela do software, mostrando as variações da distribuição normal, da DQO de saída do Tanque de Metanização (TM)..... 129
Figura 57 - Preenchimento de dados em falta, DQO de entrada do TM..... 131
Figura 58 - Preenchimento de dados em falta, vazão de entrada do TM..... 132
Figura 59 - Preenchimento de dados em falta, carga orgânica aplicada na entrada do TM..... 132
Figura 60 - Preenchimento de dados em falta, pH ajustado na entrada do TM..... 133
Figura 61 - Preenchimento de dados em falta, ácido graxos voláteis no ponto 2 do TM..... 133
Figura 62 - Preenchimento de dados em falta, alcalinidade no ponto 2 do TM..... 134
Figura 63 - Preenchimento de dados em falta, temperatura do TM..... 134
Figura 64 - Preenchimento de dados em falta, pH no ponto 2 do TM..... 135

Figura 65 - Preenchimento de dados em falta, relação entre AGV x alcalinidade do TM..... 135
Figura 66 - Preenchimento de dados em falta, DQO de saída do TM..... 136
Figura 67 - Variável DQO de saída do Tanque de Metanização (TM) unidimensionalizadas pelo método z-score (desvio padrão)..... 137
Figura 68 - Variância dos sete primeiros componentes principais..... 138
Figura 69 - Análise de componentes principais, usando todas as entradas, componente1 x componente2..... 139
Figura 70 - Análise de componentes principais, usando todas as entradas, componente1 x componente3..... 140
Figura 71 - Análise de componentes principais, usando todas as entradas, Componente2 x componente3..... 140
Figura 72 - Previsão univariada, um passo à frente, usando média acumulada e apenas os 30 dias de teste (eixo x)..... 143
Figura 73 - Previsão univariada, um passo à frente, usando média acumulada e considerando todo o período da série (no eixo x estão os números de dias amostrados)..... 144
Figura 74 - Previsão univariada, um passo à frente, utilizando média móvel simples com N=2..... 145
Figura 75 - Previsão univariada, um passo à frente, usando média móvel simples com N=2, considerando todo o período da série (eixo x-em dias)..... 145
Figura 76 - Previsão univariada, um passo à frente, utilizando média móvel simples com N=3..... 146
Figura 77 - Previsão univariada, para todo o período, um passo à frente, usando média móvel para N=3..... 147
Figura 78 - Previsão univariada, um passo à frente, utilizando média móvel simples com N=4..... 147
Figura 79 - Previsão univariada para todo o período, um passo à frente, usando média móvel simples com N=4..... 148

Figura 80 - Previsão univariada, um passo à frente, usando média móvel simples com N=10..... 149
Figura 81 - Previsão univariada, um passo à frente, para todo período, usando média móvel simples com N=10..... 150
Figura 82 - Previsão univariada, um passo à frente, usando média móvel simples com N=15..... 151
Figura 83 - Previsão univariada, para todo o período, um passo à frente, usando média móvel simples com N=15..... 151
Figura 84 - Previsão univariada, um passo à frente, usando média móvel dupla... 152
Figura 85 - Previsão univariada, para todo o período, um passo à frente, usando média móvel dupla..... 153
Figura 86 - Previsão univariada, um passo à frente, usando amortecimento exponencial simples..... 154
Figura 87 - Previsão univariada, um passo à frente de todo o período, usando amortecimento exponencial simples..... 155
Figura 88 - Previsão univariada, um passo à frente, do período de testes, usando Amortecimento Exponencial Duplo (AED) de Holt..... 156
Figura 89 - Previsão univariada, um passo à frente de todo o período, usando Amortecimento Exponencial Duplo (AED) de Holt..... 156
Figura 90 - Previsão univariada, um passo à frente, no período de teste, usando o amortecimento exponencial duplo (AED) de Brown..... 157
Figura 91 - Previsão univariada, um passo à frente, para todo o período, usando Amortecimento Exponencial Duplo (AED) de Brow..... 158
Figura 92 - Teste ARIMA (1,1,1)..... 159
Figura 93 - Todo o período ARIMA (1,1,1)..... 159
Figura 94 - Teste ARIMA (2,1,1)..... 160
Figura 95 - Todo o período ARIMA (2,1,1)..... 160
Figura 96 - Teste ARIMA (2,1,2)..... 161
Figura 97 - Todo o período ARIMA (2,1,2)..... 162

Figura 98 - Teste ARIMA (2,2,2)..... 162
Figura 99 - Todo o período ARIMA (2,2,2)..... 163
Figura 100 - Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO um dia antes..... 164
Figura 101 - Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO dois dias antes..... 164
Figura 102 - Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO três dias antes..... 165
Figura 103 - Tela do Matlab, resultado dos testes usando rede Feedforward, um passo à frente e um dia anterior como entrada (NNCO=3)..... 166
Figura 104 - Tela do Matlab, resultado dos testes usando rede Feedfrward, um passo à frente e dois dias anteriores como entrada.(NNCO=5)..... 167
Figura 105 - Tela do Matlab, resultado dos testes usando rede Feedforward, um passo à frente e três dias anteriores como entrada (NNCO=7)..... 168
Figura 106 - Decomposição e reconstrução do sinal original usando a transformada wavelet aplicada a DQO de saída do TM, usando 2 filtros de alta frequência e 2 filtros de baixa frequência..... 169
Figura 107 - Correlação e histograma de rede Feedforward, com as variáveis transformadas wavelet com dois filtros de baixa e dois filtros de alta frequência..... 170
Figura 108 - Resultado dos testes um passo à frente, usando transformada wavelet, em método univariado, usando 2 filtros de baixa e 2 filtros de alta frequência..... 170
Figura 109 - Decomposição e reconstrução do sinal original usando a transformada wavelet aplicada a DQO de saída do TM, usando 4 filtros de alta frequência e 4 filtros de baixa frequência..... 171
Figura 110 - Resultado dos testes um passo à frente, usando transformada wavelet, em método univariado, usando 4 filtros de baixa e 4 filtros de alta frequência..... 172

Figura 111 - Curvas de correlação e histograma de erro da rede neural..... 174
Figura 112 - Resultado dos testes usando apenas feedforward e um dia anterior na previsão..... 175
Figura 113 - Resultado do período de teste da rede neural suavizada com transformada wavelet, com 2 filtros de baixa e 2 filtros de alta, e 81 neurônios na camada oculta (teste1)..... 176
Figura 114 - Resultado do período de testes da rede neural suavizada com transformada wavelet com 2 filtros de baixa e 2 filtros de alta e 81 neurônios na camada oculta (teste2)..... 176
Figura 115 - Resultado do período de teste da rede neural, suavizada com transformada wavelet com 4 filtros de baixa e 4 filtros de alta frequência e 161 neurônios na camada oculta (teste 1)..... 177
Figura 116 - Resultado da regressão e período de teste da rede neural, suavizada com transformada wavelet com 4 filtros de baixa e 4 filtros de alta frequência e 161 neurônios na camada oculta (teste 2)..... 177
Figura 117 - Previsão, um passo à frente, usando todas as dez variáveis na entrada, e usando o vetor suporte de máquina..... 178
Figura 118 - Matriz de confusão e configuração da rede neural para classificação dos dados..... 185

LISTA DE TABELAS

Tabela 1- Resultado dos erros relativos entre as médias dos valores práticos e teóricos da DQO em mg.L-1..... 41

Tabela 2 - Valores de média e desvio padrão da ETE (lodos ativados)..... 42

Tabela 3 - Tabela de Dickey-Fuller..... 51

Tabela 4 - Propriedade temporal do sinal e sua respectiva representação em análise de Fourier..... 93

Tabela 5 - Resultado das séries..... 102

Tabela 6 - Desempenho do sistema ANFIS..... 103

Tabela 7 - Variáveis usadas para modelagem concentração COD do efluente..... 105

Tabela 8 - O coeficiente de correlação (r), e erro de raiz quadrada (RMSE) entre os valores reais de DQO de uma planta de tratamento de lodos ativados e as previsões, análise de rede neural..... 106

Tabela 9 - Resultados da previsão..... 108

Tabela 10 - Resultados da previsão de SST, um dia a frente e os erros médios absolutos e relativos..... 111

Tabela 11 - Resultados para rede neural..... 112

Tabela 12 - Resultado da previsão para passos à frente de TSS, usando uma rede neural MLP com treinamento de aprendizagem dinâmica..... 113

Tabela 13 - Dados estatísticos das variáveis de controle do Tanque de Metanização (TM)..... 114

Tabela 14 - Percentual do desvio padrão em relação à média..... 115

Tabela 15 - Correlação de Pearson (R), da DQO de saída um passo à frente (um dia à frente) e as outras variáveis de controle..... 130

Tabela 16 - Tela do software, correlação de Pearson da DQO de saída um passo à frente e a mesma variável nos 5 dias anteriores..... 130

Tabela 17 - Métodos univariados clássicos que não usam rede neural..... 179

Tabela 18 - Métodos univariados com redes neurais feedforward..... 180

Tabela 19 - Métodos univariados com transformada wavelet e feedforward..... 180
Tabela 20 - Métodos multivariados..... 181
Tabela 21 - Resultado dos testes com as redes CNN..... 182
Tabela 22 - Tabela com resumo dos resultados gerados no MATLAB, para previsão de 15 dias..... 183
Tabela 23 - Resultado dos testes de classificação usando SVM, com um dia de atraso..... 184

LISTA DE ABREVIATURAS E SIGLAS

ADM1	*Anaerobic Digestion Model Nº 1* (Modelo de digestão anaeróbica Nº 1);
AED	Amortecimento Exponencial Duplo;
AES	Amortecimento Exponencial Simples;
ANFIS	*Adaptive Network-based Fuzzy Inference System* (modelagem que envolve lógica difusa e redes neurais);
ARIMA	Modelo autoregressivo, integrativo com média móvel;
ARMA	Modelo autoregressivo, com média móvel;
BOD	Biochemical Oxygen Demand (demanda bioquímica de oxigênio);
CNN	Rede Neural Convolutiva;
COD	*Chemical Oxygen Demand* (demanda química de oxigênio);
DBO	Demanda Bioquímica de Oxigênio;
DNA	Ácido desoxirribonucleico;
DO	*Dissolved oxygen* (oxigênio dissolvido);
DQO	Demanda Química de Oxigênio;
DSVI	*Dilute Sludge Volume Index* (índice volumétrico de lodo diluído);
EBPN	*Error Back Propagation Network* (rede neural que corrige o erro calculado, ajustando a rede de traz para frente);
ETAR	Estação de Tratamento de Águas Residuárias;
ETE	Estação de Tratamento de Efluentes;
F/M	Relação entre o substrato biodegradável (*food*) e a quantidade de microrganismos;
IVL	Índice Volumétrico de lodo;
IWA	*International Water Association* (associação internacional de águas);
K-NN	*k-Nearest Neighbors* (técnica de previsão usando o vizinho mais próximo);
KSOM	*Kohonen Self Organising Map* (rede com mapas organizáveis de Kohonem);
MAE	*Mean Absolute Error* (erro médio absoluto);

MAPE	*Mean Absolute Percentage Error* (erro percentual médio absoluto);
MARS	*Multi-variate Adaptive Regression Spline* (técnica de previsão usando multivariáveis);
mg.L	miligrama por litro;
MLP	*Multilayer Perceptron* (rede neural *perceptron* de múltiplas camadas);
MLSS	*Mixed Liquor Suspended Solids* (sólidos suspensos no reator Biologico);
MMD	Média Móvel Dupla;
MMP	Média Móvel Ponderada;
MMQ	Método dos Mínimos Quadrados;
MMS	Média Móvel Simples;
MSE	*Mean Squared Error* (erro médio quadrático);
NN	*Neural Networks* (rede neural);
NNCO	Número de neurônios na camada oculta de uma rede neural;
OD	Oxigênio Dissolvido;
pH	Potencial hidrogênio iônico;
PLS	Mínimos quadrados parciais;
PT	Fósforo Total;
PTCO	Unidades platina cobalto para medir cor;
RAS	*Return Activated Sludge* (refere-se ao reciclo de lodo, em processos de lodos ativados);
RBFN	*Radial Basis Function Network* (rede neural de base radial);
RF	*Random Forest* (técnica de previsão usando chamadas florestas aleatórias, que fazem ensacamento de características, para previsão e classificação);
RMSE	*Root Mean Square Error* (raiz quadrada do erro médio);
RNA	Rede neural artificial;
SF	Sólidos Fixos;
SSD	Sólidos em Suspensão Decantáveis;
SSTR	Sólidos Suspensos Totais no Reciclo de lodo;
SST	Sólidos Suspensos Totais;

SSV	Solidos Suspensos Voláteis;
SSVI	*Stirred Specific Volume Index* (volume de sólidos sedimentáveis em cone de Inhoff);
SSVR	Solidos Suspensos Voláteis no Reciclo de lodo;
ST	Sólidos Totais Dissolvidos;
SV	Sólidos Voláteis;
SVI	*Sludge Volumetric Index (*Indice volumétrico de lodo);
SVM	*Support Vector Machine* (vetor suporte de máquina);
TA	Temperatura Ambiente;
TE	Temperatura do Efluente;
TSS	Sólidos suspensos totais;
UASB	*Upflow Anaerobic Sludge Blanket* (reator de fluxo ascendente de manta de lodo);
UHASB	*Upward Flow Hybrid Anaerobic Sludge Blanket* (reator anaeróbio de fluxo ascendente hibrido).

LISTA DE SÍMBOLOS

LATINOS:

b	Valor das bias que inicialmente é atribuído aleatoriamente em uma rede neural;
b_t	é um fator de ajuste adicional usado no método AED;
C_t	Componente cíclica ou sazonal que se repete a cada intervalo fixo de tempo em série temporal;
d_j	Valor desejado na saída de uma rede neural;
d_t	Resposta desejada para o exemplo *t,* na aferição de erro;
$\bar{E}$	Erro global médio para um conjunto de treinamento de uma rede neural;
E(n)	Erro global instantâneo na interação n, de uma rede neural;
e_j	Erro obtido na rede neural;
i	Se refere à camada, em rede neural;
I_t	Valor do componente irregular em t;
j	Se refere a posição do neurônio na camada, em rede neural;
N	Número de exemplos do conjunto de dados, para aferição de erro;
n	Número de neurônios da camada anterior, em rede neural;
N_d	Número de dados que vão ser amostrados, na definição de confiabilidade da amostra;
n_i	Número de indivíduos na amostra, na definição de confiabilidade da amostra;

p	Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar, na definição de confiabilidade da amostra;
P_{t+1}	Valor da previsão usado AES;
Rt	Uma componente aleatória ou ruído (a parte não explicada), usada em série temporal;
St	Valor do componente sazonal em t, em série temporal;
Tt	Uma componente de tendência em série temporal;
V_{ij}	Valor de um sinal propagado em uma rede neural;
W	Peso atribuído a cada sinal que sai do neurônio da camada anterior. Em rede neural;
X_t	Variável vizinha que se quer determinar usando série temporal;
Y_j (n)	Valor de saída de uma rede neural na interação n;
Yt	Resposta calculada pelo método para o exemplo *t,* na aferição de erro;
Z	Valor do desvio, que corresponde ao grau de confiança desejado, na definição de confiabilidade da amostra;
Z_t,Z_{t-1},Z_{t-2},Z_{t-3}	São valores reais observados ou calculados da própria variável que se deseja estimar, a cada intervalo (0,1,2,3...) de tempo t anterior em uma série temporal;
Z_{t+h}	São valores da previsão futura após intervalo de tempo h, usado em série temporal.

GREGOS:

α	é o coeficiente de amortecimento (0£ a £ 1), usado na previsão por amortecimento exponencial simples (AES) e amortecimento exponencial duplo (AED);
$\delta_j(n)$	Função do gradiente local do neurônio j, na interação n de uma rede neural;
σ_j	Desvio-padrão da variável X;
ε	Margem de erro ou erro máximo da estimativa. Identifica a diferença máxima entre a proporção amostral e a verdadeira proporção populacional (p), na definição de confiabilidade da amostra;
ε_i	Folga usada para cálculo do hiperplano separador otimo usada em SVM;
$\xi(B)$	Operador auto regressivo não estacionário de ordem p+d, usado nos modelos ARIMA e ARMA;
ε_t	Erro ou ruído do processo em t, em série temporal;
$\varphi(a_0, a_1, a_{2,\ldots}, a_m)$	Função do resíduo da distância entre pontos e um polinômio interpolador;
ω	Sinal de saída do neurônio, que envia o sinal ao próximo neurônio ou a saída da rede, em rede neural;
$\emptyset(x_i)$	Função Kernel pré definida, usada em SVM e CNN;
ϑ	Coeficiente de inclinação da série temporal no teste de estacionariedade de Dick-Fuller
μ_j	Média da variável *Xj*;
τ	Número tau usado no teste de estacionariedade de Dickey-Fuller.

SUMÁRIO

1	**INTRODUÇÃO**.....	32
2	**REVISÂO DA LITERATURA**.....	37
2.1	O PROCESSO DE LODOS ATIVADOS.....	37
2.2	REATOR ANAERÓBIO DE FLUXO ASCENDENTE.....	38
2.3	PRINCIPAIS VARIÁVEIS DE CONTROLE E MONITORAMENTO BIOLÓGICO EM ESTAÇÃO DE TRATAMENTO DE EFLUENTES (ETE).....	40
2.4	FUNDAMENTAÇÃO TEÓRICA.....	45
2.4.1	**Séries temporais**.....	45
2.4.2	**Método de aferição dos resultados**.....	46
2.4.3	**Séries temporais estacionárias e não estacionárias**.....	46
2.4.3.1	*Teste de estacionariedade da série temporal*.....	49
2.4.4	**Séries temporais univariada e multivariadas**.....	52
2.4.5	**Série Temporal Discreta e Contínua**.....	53
2.4.6	**Modelos de previsão, preenchimento de dados faltantes e suavização em séries temporais**.....	53
2.4.6.1	*Média*.....	54
2.4.6.2	*Média Móvel Simples (MMS)*.....	54
2.4.6.3	*Média móvel ponderada (MMP)*.....	55
2.4.6.4	*Média móvel dupla (MMD)*.....	55
2.4.6.5	*Interpolação polinomial (mínimos quadrados)*.....	56
2.4.6.6	*Extrapolação de dados*.....	58
2.4.6.7	*Amortecimento exponencial simples (AES)*.....	59
2.4.6.8	*Amortecimento Exponencial Duplo (Método de Brown) - AED*.....	59

2.4.6.9	*ARIMA (Auto Regressivo integrado de média móvel)*.....	60
2.4.7	**Redes Neurais Artificiais**.....	62
2.4.8	**Vetor suporte de Maquina (SVM)**.....	75
2.4.9	**Rede Neural Convolutiva (CNN)**.....	81
2.4.10	**Necessidade de pré-tratamento dos dados**.....	83
2.4.10.1	*Normalização dos dados*.....	83
2.4.10.2	*Preenchimento de dados faltantes*.....	84
2.4.10.3	*Suavização dos dados para redução do ruído*.....	85
2.4.10.4	*Transformada de Fourier clássica*.....	88
2.4.10.5	*A forma complexa da série de Fourier*.....	89
2.4.10.6	*A transformada de Fourier continua e discreta*.....	91
2.4.10.7	*A transformada de Fourier janelada*.....	93
2.4.10.8	*A transformada wavelet continua e sua inversa*.....	94
2.4.10.9	*A transformada wavelet discreta*.....	97
2.4.10.11	*A Análise dos componentes principais (ACP)*.....	98
2.5	TRABALHOS DESENVOLVIDOS PARA PREVISÃO DE DADOS EM ESTAÇÕES DE TRATAMENTO DE EFLUENTES.....	100
3	**METODOLOGIA**.....	114
3.1	COLETA DE DADOS.....	115
3.2	DESENVOLVIMENTO DE SOFTWARE PARA TRATAR OS DADOS E EXECUÇÃO DOS MÉTODOS.....	116
3.3	VERIFICAÇÃO DE DADOS ESTATÍSTICOS DE DISTRIBUIÇÃO NORMAL DAS VARIÁVEIS ENVOLVIDAS ECORRELAÇÃO DE PEARSON.....	117
3.4	PRÉ-TRATAMENTO DE DADOS.....	117
3.5	CRITERIOS DE DESEMPENHO E AVALIAÇÕES DAS SÉRIES TEMPORAIS PARA PREVISÃO UM PASSO A FRENTE.....	118

3.6	CRITÉRIOS DE DESEMPENHO E AVALIAÇÕES PARA CLASSIFICAÇÃO DOS DADOS UM PASSO A FRENTE.	119
3.7	CRITERIOS DE DESEMPENHO E AVALIAÇÕES DAS SÉREIS TEMPORAIS PARA CLASSIFICAÇÃO E AVALIAÇÃO DOS COMPONENTES PRINCIPAIS.....	121
4	**RESULTADOS E DISCUSSÃO**.....	122
4.1	VERIFICAÇÃO DE DADOS ESTATÍSTICOS DE DISTRIBUIÇÃO NORMAL DAS VARIÁVEIS ENVOLVIDAS.....	122
4.2	VERIFICAÇÃO DA CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS UM PASSO À FRENTE E OS DIAS ANTERIORES.....	129
4.3	PRÉ-TRATAMENTO DE DADOS.....	131
4.3.1	**Preenchimento dos dados faltantes nas séries temporais**.....	131
4.3.2	**Normalização dos dados, para trabalho com multivariáveis**.	136
4.3.3	**Teste de estacionariedade de Dickey-Fuller**.....	137
4.4	CLASSIFICAÇÃO E AVALIAÇÃO DOS COMPONENTES PRINCIPAIS DOS DADOS OBTIDOS.....	138
4.5	PREVISÃO UNIVARIADA UM PASSO À FRENTE DA DQO DE SAIDA.....	141
4.5.1	**Previsão univariada usando média acumulada**.....	142
4.5.2	**Previsão univariada usando a média móvel simples**.....	144
4.5.3	**Previsão univariada usando a média móvel dupla**.....	152
4.5.4	**Previsão univariada usando o amortecimento exponencial simples (AES)**.....	154
4.5.5	**Previsão univariada usando o amortecimento exponencial duplo de Holt (AED de Holt)**.....	155
4.5.6	**Previsão univariada usando amortecimento exponencial duplo de Brown (AED de Brown)**.....	157

4.5.8	**Previsão univariada usando ARIMA**.....	158
4.5.9	**Previsão univariada usando vetor suporte de máquina**.....	163
4.5.7	**Previsão univariada usando rede neural feedforward**.....	165
4.5.10	**Previsão univariada usando a transformada wavelet e rede neural feedforward**.....	168
4.6	PREVISÃO MULTIVARIADA, UM PASSO À FRENTE, DA DQO DE SAÍDA.....	172
4.6.1	**Previsão Multivariada usando apenas feedforward e um dia anterior**.....	173
4.6.2	**Previsão multivariada usando transformada wavelet**.....	175
4.6.3	**Previsão multivariada um passo à frente, usando SVM**.....	178
4.7	RESUMO DOS TESTES UM PASSO À FRENTE UNIDIMENSIONAL E MULTIDIMENSIONAL.....	179
4.8	CLASSIFICAÇÃO DE DADOS UM PASSO A FRENTE, USANDO REDES NEURAIS CONVOLUTIVAS.....	181
4.9	CLASSIFICAÇÃO DE DADOS, UM PASSO A FRENTE, USANDO VETOR SUPORTE DE MÁQUINA.....	183
4.10	CLASSIFICAÇÃO DE DADOS, UM PASSO A FRENTE, USANDO REDE NEURAL FEEDFORWARD.....	184
4.11	COMPARAÇÃO DOS RESULTADOS COM OS DADOS ENCONTRADOS EM LITERATURA.....	186
5	**CONCLUSÕES E PERSPECTIVAS**.....	189
5.1	CONCLUSÕES.....	189
5.2	SUGESTÃO PARA TRABALHOS FUTUROS.....	190
	REFERÊNCIAS.....	191
	APENDICE A - PROGRAMA QUE PREENCHE OS DADOS FALTANTES E NORMALIZA OS DADOS USANDO Z-SCORE.....	199

APÊNDICE B - PROGRAMA QUE TRANSFORMA OS DADOS EM FIGURA E USA A REDE CONVOLUTIVAS..... 201
APÊNDICE C- PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO *FEEDFWARD* UNIVARIADO..... 205
APÊNDICE D - MÉTODO UNIVARIADO USANDO WAVELET COM QUATRO FILTROS DE BAIXA E QUATRO FILTROS DE ALTA FREQUÊNCIA..... 206
APÊNDICE E - PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO MODELO *WAVELET/FEEDFORWARD* DADOS NORMALIZADOS, MULTIVARIADO..... 208
APÊNDICE F - PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO MODELO SUPPORT VECTOR MACHINE REGRESSION, DADOS NORMALIZADOS..... 210
APÊNDICE G - TELA DO SOFTWARE GRETL COM TESTE DE ESTACIONARIEDADE DE DICKEY-FULLER... 212

1 INTRODUÇÃO

Despoluir a água é um grande desafio em função da prevista escassez de água doce no planeta e também pelas doenças a esta associada.

No Brasil, doenças relacionadas ao saneamento básico inadequado foram responsáveis, em média, por 13.449 óbitos por ano ao longo do período de 2001 a 2009, 1,31% dos óbitos ocorridos no período. As diarréias e a doença de Chagas representaram mais de 83% do total de óbitos por causas definidas relacionadas ao saneamento básico inadequado, ou seja, mais de nove mil mortes por ano (TEIXEIRA *et al.,* 2014).

Assim, o tratamento da água poluída é hoje de grande importância em função dos desafios de sustentabilidade e também melhoria na saúde e qualidade de vida das pessoas.

Conforme Leme (2014), diversos agentes patogênicos transmissores de doenças para o homem estão presentes nos esgotos, entre eles: Ascaris spp e Enterobius spp, causadores das 17 verminoses; Ancylostomaspp, causadores do amarelão; Bacillusanthracis, causador do antraz; Brucellassp, causadores da brucelose, Entamoebahistolyticae Giardiaspp, causadores da diarreia; Leptospiraspp, causadores da leptospirose; Mycobacterium tuberculosis, causador da tuberculose, Salmonellaspp, causadoras de febre tifoide e paratifoide, Schistosomaspp, causadores de esquistossomose, Vibriocholerae, causador da cólera; evírus, causadores da poliomielite, hepatite infecciosa e outros.

Diversas técnicas envolvendo processos químicos, físicos e biológicos, são utilizadas para despoluir a água doce, dependendo da natureza do poluente e as características finais e uso da água a ser tratada.

Os tratamentos de efluentes com poluentes orgânicos biodegradáveis, usando processos biológicos vêm sendo desenvolvidos especialmente durante estes últimos cem anos, utilizando técnicas diversas que usam microrganismos aeróbios, anaeróbios e facultativos com base no que faz a natureza, para separar transformando em gases e biomassa a matéria poluente.

Como em todo processo que utiliza microrganismo vivo, o controle de uma estação de tratamento de efluentes necessita de ajustes diversos e monitoramento de várias variáveis, para que a estação de tratamento funcione.

Normalmente, no processo biológico, quando uma variável não monitorada e indesejada ocorre, ou mesmo um acidente de difícil detecção por parte do operador, como por exemplo:

-Incremento de um novo produto químico na produção, cujos resíduos vão para a Estação de tratamento de efluentes (ETE) e que afetam o metabolismo dos microrganismos;

-Quebra de uma parte mecânica submersa no processo de tratamento de efluentes, por exemplo, um dos raspadores mecânicos de um decantador secundário.

Estes acidentes, provocam mudanças nas variáveis controladas, causando um efeito de descontrole, que podem ser detectados imediatamente, se a velocidade na mudança for rápida, ou lentamente, se o efeito causar mudanças, que podem ser com o tempo, adaptadas pelos microrganismos, mas sentidas no processo por uma variação na média desta variável com o tempo.

Quando uma ou mais variáveis monitoradas ou não, saem de controle, estas podem ou não afetar outras variáveis, por exemplo, se uma planta apresenta perda de sólidos em suspensão na saída de um decantador secundário, em um processo de tratamento biológico com lodos ativados, podem ocorrer várias possibilidades e ajustes para esta causa, como:

- Sobrecarga hidráulica no decantador;
- Sobrecarga de sólidos;
-Desnitrificação;
-Oxigenação inadequada no reator biológico;
-Presença de microrganismos filamentosos;
-Quebra de raspador mecânico;
-Idade de lodo muito elevada;
- Excesso de densidade de potência no reator biológico;
- Fator de carga inadequado.

Muitas vezes para o operador se aconselha o ajuste de uma variável por vez e esperar os resultados, para que então se possa fazer o ajuste em outra variável, gerando demora no ajuste da planta e riscos ambientais.

A interferência de uma variável sobre a outra pode gerar ou não consequências de perda de eficiência na estação. Assim fazer a previsão, com respostas dentro de uma faixa menor que os desvios padrões das variáveis de um processo estável, ou mesmo uma simulação do que ocorreria com a mudança destas variáveis, considerando as características intrínsecas de uma estação, seria uma ferramenta de grande valia no controle operacional, permitindo uma melhor tomada de decisão, melhores ajustes do processo, antecipando riscos e simulando processos ou

mesmo desenvolvendo métodos de controle automático de uma estação de tratamento de efluentes.

Normalmente, os processos de tratamento de efluentes com cargas orgânicas biodegradáveis, quando possível, seguem a sequência do tratamento de menor custo e envolvem inicialmente separações simples de componentes que estão misturados à agua de forma heterogênea, usando para isto, por exemplo, processos físicos simples como peneiramento, decantação, flotação e separação física por diferença de densidade do material que não se solubiliza com a água.

Após a separação do material heterogêneo, temos uma mistura homogênea ao qual se utiliza microrganismos, que em seu metabolismo convertem em biomassa e gás o material poluente biodegradável misturado a água, de forma que a mistura antes homogênea se torme heterogênea e passível de ser separada da água por processos fisicos. Entre estas técnicas, principalmente, temos: as lagoas de tratamento de efluentes, processos anaeróbios de tratamento e sistemas de lodos ativados.

Os processos biológicos precisam de acompanhamento analítico e operacional de diversas variáveis para manter a estação com resultados operacionais previstos em projetos.

Este trabalho, utiliza métodos de previsão da variável de saída de uma estação operada (variável alvo), usando sistemas anaeróbios de tratamento de efluentes.

O método poderá ser usado também em outras áreas ou processos, tais como: lodos ativados, lagoas de tratamento biológico, produção, controle de qualidade e econometria.

Também se utilizam algumas ferramentas estatísticas usadas em séries temporais e associam-se estas técnicas as redes neurais, buscando verificar os resultados, que mais se aproximem de uma previsão futura dos resultados da DQO de saída, de uma estação de tratamento de efluentes. Evidente, que não se pode esgotar todos os métodos de previsão existentes, mas se pretende chegar a resultados que possibilitem uma margem de confiança maior, empregando os métodos simplesmente e combinações destes métodos, usando rede neural com algoritimo de retropropagação (*feedforward*), com uso de transformadas onduletas (wavelet) e outras técnicas como por exemplo a do vetor suporte de máquina (VSM).

Com o objetivo de responder se existe ganhos na previsão, utilizou-se alguns métodos convencionais de forma univariada ou multivariada ou a combinação dos métodos, junto com redes neurais artificiais feedforward, associado a técnica wavelet.

Foi utilizado também, redes neurais convolutivas (aprendizado profundo), na classificação do que vai acontecer com a DQO de saída de uma ETE.

É importante ressaltar que, uma previsão de uma variável sujeita a erros humanos, de material e de métodos, como é o caso das variáveis de controle de uma estação de tratamento de efluentes, tem resultados sempre dentro de um campo probabilístico.

Porém, se procurou um método matemático, que tenha condições de responder ou simular o processo dentro de uma faixa normal de operação, e que associe as variáveis existentes e suas mudanças a resultados quantitativos e qualitativos do resultado esperado.

Em uma estação de tratamento de efluentes, por exemplo, responder a questionamentos em casos de funcionamento normal da planta como o que aconteceria no caso de mudanças de variáveis de controle do processo, a exemplo a temperatura no reator anaeróbio de fluxo ascendente, se subisse de 27 para 29ºC, o que se deve esperar da demanda química de oxigênio (DQO) de saída, estando as outras variáveis dentro dos seus valores normais de operação ?

Segundo Masters (1994) apud Peixoto (2013), existem dezenas de redes neurais descritos na literatura, ao qual o processo de seleção pode parecer uma tarefa difícil, diante do número de possibilidades, porém é pouco provável que poucos modelos possam fornecer uma solução excelente.

Usou-se um método, que apresenta uma resposta dentro de um espaço probabilístico de confiança, na estimativa e classificação dos dados previstos, permitindo, ao usar um ou vários métodos, se ter uma noção do que se espera em regime normal de operação em relação a resultados, considerando a variação dos ruídos presentes. Se as respostas forem fora desta faixa se deve ficar em alerta a mudanças em materiais, métodos e processo.

Não é realizada uma comparação de qual método é o melhor, mas se ter noção de até quanto se pode chegar, dentro das variações normais do processo, ou resultados que possam representar dentro de faixas estatísticas, valores de previsão, dentro dos desvios padrões dos valores amostrados em dois anos, e que possam ser reproduzidos mantendo variações próximas da mesma faixa.

Os métodos usados, são válidos para faixas de variação pertencentes aos históricos registrados das análises, em uma planta bem operada, pois a rede é treinada com os valores registrados das análises de operação, um valor da previsão com resultado muito diferente pode ser indicativo de que uma variável não monitorada está afetando o resultado. Então, neste caso,

o operador deve investigar o motivo desta variação, ou ainda mudança brusca no processo produtivo, como a entrada de um novo produto ou ainda quebra de equipamentos no tratamento de efluentes.

O objetivo principal deste trabalho foi de se desenvolver uma metodologia de controle e previsão do comportamento de eficiência de um sistema biológico de tratamento de efluentes, usando técnicas estatísticas e alguns sistemas de inteligência artificial, permitindo maior controle sobre a ETE (no caso deste trabalho, se usou um tratamento anaerobio de fluxo ascendente, podendo o método ser usado em outros processos biológicos) e simulações do processo dentro de padrões e características de uma planta em condição normal de operação.

Como objetivos específicos foram necessários coletar dados de dois anos de uma estação de tratamento de efluentes, o desenvolvimento de software para lidar com o volume de dados e realizar a avaliação estatística dos dados, e executar diversos métodos de previsão e classificação em séries temporais.

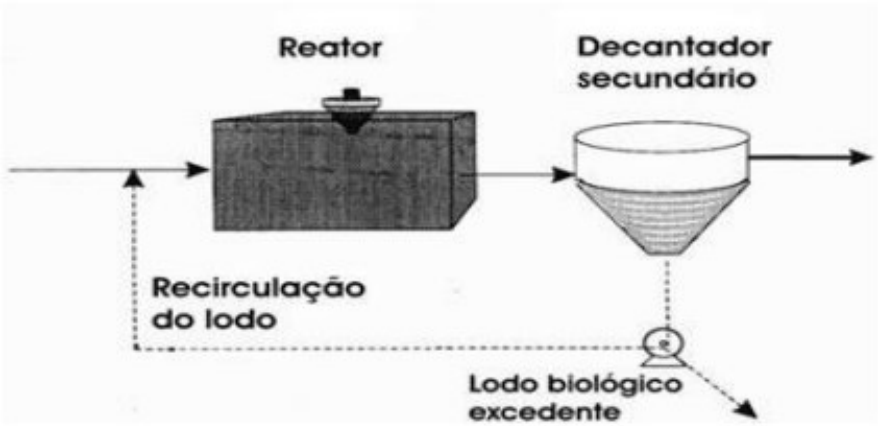
2 REVISÂO DA LITERATURA

Existem diversos tipos de modalidades de estações de tratamento de efluentes biológicos, citaremos resumidamente as principais.

2.1 O PROCESSO DE LODOS ATIVADOS.

Segundo Sperling (2012), o sistema de lodos ativados (Figura 1) é amplamente divulgado, a nível mundial, para o tratamento de despejos domésticos e industriais, e é usado em situações em que são necessários uma qualidade elevada do efluente tratado e requisitos mínimos de área, porém exigem uma operação mais sofisticada e um índice maior de mecanização.

Figura 1-Desenho esquemático de um processo com lodos ativados.



Fonte: Sperling (2012)

Após passar pelas etapas de separação da mistura heterogênea, o efluente homogêneo, contendo água e matéria orgânica biodegradável, é lançado em um reator biológico contendo microrganismos aeróbios, que em condições favoráveis ao seu desenvolvimento (principalmente: oxigênio dissolvido (OD), temperatura, nitrogênio, fósforo, controle de alimentação e ausência ou controle de materiais recalcitrantes). Após esta etapa, ocorre a metabolização da matéria orgânica solúvel na água, usando esta, como fonte de produção de energia, gerando gases no processo, como subproduto do metabolismo e convertendo a massa biodegradável homogênea em uma massa heterogênea. Desta forma, o material biodegradável

solúvel é convertido em uma mistura de gases e microrganismos que se separam da água. A separação dos gases é feita no próprio reator biológico e as dos microrganismos, denominados lodos no processo, é feita por decantação, pois, em certas condições operacionais, o lodo biológico tem a capacidade de flocular e se depositar no fundo, aonde ocorre a remoção mecânica de uma parte calculada do lodo (denominada descarte). Uma outra parte dos microrganismos retornam ao reator biológico ou tanque de aeração, permitindo aumento da concentração da biomassa (microrganismos) no reator biológico.

Segundo Sperling (2012), o processo de recirculação do lodo permite um aumento de de até dez vezes à concentração de microrganismos de uma lagoa aerada, permitindo ao processo de lodos ativados, tratar em volumes bem menores a mesma carga orgânica. Por ser um processo que utiliza microrganismos, uma série de parâmetros devem ser monitorados e ajustados, para dar condições de um ambiente favorável ao metabolismo e multiplicação celular, além de condições de boa decantabilidade da mistura para separação da água tratada no processo de lodos ativados, sendo o seu controle fundamental nos resultados de uma Estação de Tratamento de Efluentes (ETE).

2.2 REATOR ANAERÓBIO DE FLUXO ASCENDENTE.

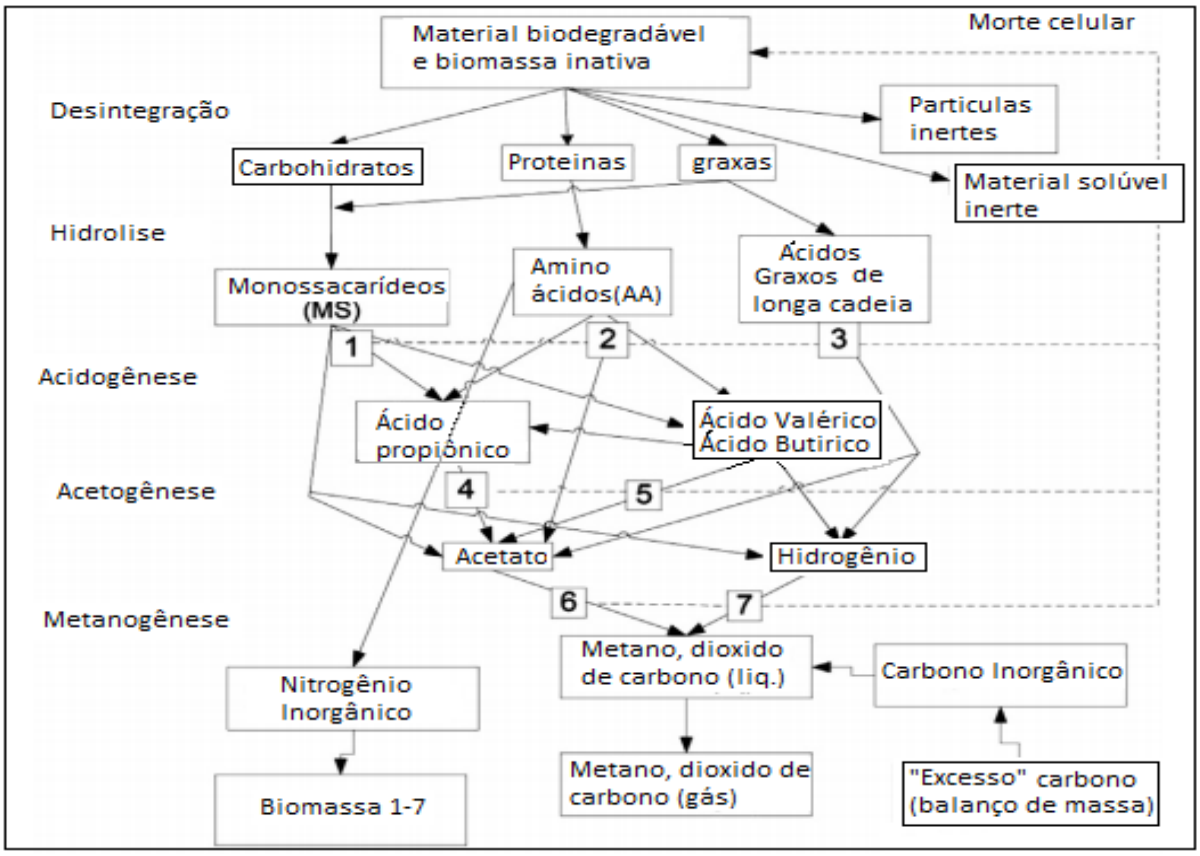
Os microrganismos anaeróbios também utilizam a matéria orgânica biodegradável em seu metabolismo, gerando como subproduto gases e células novas (biomassa biológica ou lodo).

O reator de fluxo ascendente de manta de lodo UASB (*Upflow Anaerobic Sludge Blanket*, em inglês), consiste de um tanque de fluxo ascendente, no qual os microrganismos presentes no manto de lodo do reator, na forma de flocos ou grânulos retêm os sólidos e convertem a matéria orgâniga biodegradável (expressa em Demanda Bioquimica de Oxigênio ou DBO), em biogás, microrganismos e água. Na sua parte superior há um separador trifásico (sólido-líquido-gás), ao qual ocorre a remoção do gás produzido, assim como, a sedimentação e retorno automático do lodo à câmara de digestão (FAEDO, 2010).

Na Figura 2, está representado um esquema das etapas e reações bioquímicas, que ocorrem em um reator biológico anaeróbio. Observa-se que existe no processo, diferentes microrganismos e diferentes etapas de degradação orgânica, até se chegar a redução a gás da matéria orgânica biodegradável, exigindo um controle forte sobre a alimentação, temperatura e pH do sistema.

Assim, dentro de uma faixa de projeto do reator, se tenta manter a maior eficiência na remoção da demanda química de oxigênio (DQO) de entrada, também podem ocorrer variações nos parâmetros de controle que podem afetar, os diversos equilíbrios e população microbiana, que são responsáveis pela conversão e separação da matéria biodegradável a gás, afetando o resultado da DQO de saída do sistema.

Figura 2 -Os caminhos de reação bioquímica descritos em Modelo de Degradação Biológica 1 (ADM1) do International Water Association (IWA), com os seguintes grupos microbianos:(1) degradadores de açúcar, (2) degradadores de aminoácidos, (3) degradadores de ácidos graxos de cadeia longa (4) degradadores de ácido propiònico, (5) degradadores de ácido butírico e valérico, (6) metanógenos acetoclásticos e (7) metanógenos hidrogenotróficos.



Fonte: Adaptado de Lauwers et al. (2013).

A digestão anaeróbica permite o tratamento de resíduos (água) e a produção de energia sob a forma de biogás. A implementação bem sucedida deste processo levou a um crescente interesse em todo o mundo. No entanto, a digestão anaeróbica é um processo biológico complexo, ao qual centenas de populações microbianas estão envolvidas e cujo início e operação são questões que exigem grande controle de operação. Para entender melhor a dinâmica do processo e otimizar as condições operacionais, a disponibilidade de modelos dinâmicos é de

suma importância. Tais modelos devem ser inferidos a partir do conhecimento prévio e dados experimentais reais coletados destas estações (DONOSO-BRAVO *et al.,* 2011).

A digestão anaeróbica é uma excelente técnica para a valorização energética de vários tipos de biomassa, incluindo formas de resíduos. Devido à sua natureza complexa, a otimização e o desenvolvimento de processos adicionais desta tecnologia acompanham a disponibilidade de modelos matemáticos, tanto para fins de simulação, quanto de controle. Ao longo dos anos, a variedade de modelos matemáticos desenvolvidos aumentou, assim como, a sua complexidade (LAUWERS *et al.,* 2013).

Segundo Brandão (2014), a digestão anaeróbia tem como objetivo a decomposição da matéria orgânica e ou inorgânica na ausência de oxigénio molecular. Os micro-organismos responsáveis pela decomposição da matéria orgânica dividem-se frequentemente em dois grupos:

– Acidogénicos: este primeiro grupo transforma os ácidos simples, álcoois e aldeídos.

– Metanogénicos: este segundo grupo transforma os ácidos orgânicos simples formados pelo primeiro grupo em gás neutro, anídrico carbónico e água como produtos finais, conseguindo-se dessa forma uma matéria sólida bem estabilizada, apta para a sua descarga para uma zona de secagem e desidratação ou para o seu aproveitamento agrícola.

O correto dimensionamento de um processo de digestão anaeróbia tenderá a manter num equilíbrio dinâmico os dois grupos de micro-organismos, para que, em função da matéria orgânica a digerir, da taxa de crescimento dos micro-organismos e da temperatura de funcionamento do sistema, se obtenha um aproveitamento ótimo do seu depurador, conseguindo-se uma elevada taxa de gaseificação e uma correta estabilização de resíduo, com o melhor retorno económico.

2.3 PRINCIPAIS VARIÁVEIS DE CONTROLE E MONITORAMENTO BIOLÓGICO EM ESTAÇÃO DE TRATAMENTO DE EFLUENTES (ETE)

As principais variáveis de controle e monitoramento mais usadas no processo de tratamento anaeróbio e aeróbio de efluentes, sofrem influências da metodologia usada, qualidade dos reagentes químicos usados, tomadas de amostra e erros amostrais envolvidos. Estas influências geram variações, que nas cartas de controle apresentam os chamados ruídos.

Um exemplo de erro, está no trabalho de Bado *et al.* (2012), que compara os erros entre os valores medidos por métodos colorimétricos e titulométricos, na análise da demanda química de oxigênio (DQO), conforme os resultados da Tabela 1.

O erro relativo na análise aumenta a medida que a DQO diminui. Dados práticos mostram que, mesmo usando a mesma metodologia e a mesma amostra haverá diferença nos resultados.

Uma série temporal, dentro destas possibilidades de erro, não faz previsão exata dos valores futuros, mas se tenta chegar a modelos que mais se aproximem dos valores reais obtidos. O mesmo modelo pode prever tendências de resultados e correlação entre as variáveis de controle, que podem ser de grande utilidade no controle e automação da ETE.

Tabela 1 -Resultado dos erros relativos entre as médias dos valores práticos e teóricos da DQO em $mg.L^{-1}$

DQO Testada	DQO Teórica	DQO Colorimétrico	Erro relativo entre colorimétrica e DQO teórica%	DQO Titulométrico	Erro relativo entre Titulométrica e DQO teórica%
193,00	195,00	173,23	11,20	163,73	16,00
200,00	202,00	190,91	5,76	213,68	5,47
500,00	501,41	513,89	2,49	509,86	1,69

Fonte: Bado, Pérsio e Lindino (2012)

Na Tabela 2, se encontra valores levantados por Bousfield, Silva e Chapuis (2011) em uma estação de tratamento de efluentes com lodos ativados, observando os valores de desvio e seu percentual de variação em relação a média destes valores. Se observa variações diversas por variável onde os valores variam de 2,41% até 88,71%. Assim, a escolha da variável a se determinar também pode resultar em valores de maior ou menor erro de previsão de acordo com a variável escolhida.

Tabela 2-Valores de média e desvio padrão da ETE (lodos ativados)

Variável	Entrada	Desvio	%Desvio
Demanda Bioquimica de Oxigênio $_5$ (mg.L^{-1})	400,13	92,26	23,06%
Demanda Química de Oxigênio (mg.L^{-1})	1423	258,5	18,17%
Fósforo Total (mg.L^{-1})	3,95	1,7	43,04%
Nitrogênio Amoniacal (mg.L^{-1})	11,82	5,22	44,16%
Oxigênio Dissolvido (mg.L^{-1})	3,31	0,64	19,34%
pH	9,71	0,42	4,33%
Sólidos Sedimentaveis (mg.L^{-1})	0,62	0,55	88,71%
Sólidos em Suspensão (mg.L^{-1})	775,73	255,92	32,99%
Solidos Totais (mg.L^{-1})	1318,76	245,59	18,62%
Temperatura (ºC)	39,93	2,59	6,49%
Turbidez (NTU)	501,49	30,12	6,01%
Cor (PTCO)	2690,08	164,6	6,12%
Variável	Saída	Desvio	%Desvio
Demanda Bioquímica de Oxigênio $_5$ (mg.L^{-1})	23,42	5,83	24,89%
Demanda Química de Oxigênio (mg.L^{-1})	98,31	15,26	15,52%
Fósforo TotalT (mg.L^{-1})	0,55	0,13	23,64%
Nitrogênio Amoniacal (mg.L^{-1})	7,44	4,41	59,27%
Oxigênio Dissolvido (mg.L^{-1})	5,49	0,83	15,12%
pH	6,65	0,16	2,41%
Sólidos Sedimentáveis (mg.L^{-1})	0,13	0,10	76,92%
Solidos em Suspensão (mg.L^{-1})	174,24	37,41	21,47%
Solidos Totais (mg.L^{-1})	1270,07	113	8,90%
Temperatura (ºC)	31,61	2,34	7,40%
Turbidez (NTU)	21,95	7,71	35,13%
Cor (PTCO)	118,17	39,27	33,23%

Fonte: Adaptado de Bousfield, Silva e Chapuis (2011)

A vazão dos efluentes é uma variável na qual, o processo produtivo pode ter grande influência, pode ser trabalhada na fonte de geração por redução de consumo de água nas lavagens de equipamentos ou melhoria da eficiência nestas lavagens. A vazão tem variação por erros de leitura nos equipamentos de medição, de acordo com a precisão de leitura destes equipamentos e também por ingresso de outras fontes de água ou efluentes não previstas

O aumento de vazão pode provocar sobrecarga hidráulica nas estações de tratamento de efluentes, arrastar os sólidos em suspensão e provocar perda de eficiência tanto em processos anaeróbios, como em processos aeróbios e facultativos.

Uma vazão reduzida pode deixar o tempo de residência hidráulica do sistema projetado muito alto e em caso de decantadores secundários, isto pode provocar problemas de decantação de lodo por redução do nível de oxigênio dissolvido no sistema.

A Demanda Bioquímica de Oxigênio é uma variável que sofre influência da carga e do tipo de material biodegradável, que é arrastado pela água de lavagem de equipamentos ou água de lixiviação do material biodegradável. A DBO pode ser trabalhada por melhoria na eficiência do processo produtivo, redução do contato com a água de resíduos lixiviáveis e outras ações de substituição ou redução da carga do resíduo em contato com a água de lavagem. Como toda análise química, a DBO apresenta variações de acordo com a tomada de amostra e também com a qualidade dos reagentes utilizados.

A Demanda Química de Oxigênio (DQO), é uma variável que apresenta influência da carga e do tipo de material orgânico ou inorgânico que consome oxigênio e que é arrastado pela água de lavagem de equipamentos ou água de lixiviação deste material.

Segundo Ferreira e Coraiola (2008), o interferente mais comum para o teste da DQO é a presença de cloreto, que pode precipitar a prata que é utilizada como catalisador na determinação da DQO, assim, a presença desses elementos pode alterar o resultado das análises.

O nitrogênio é um elemento indispensável para o crescimento de algas e quando em elevadas concentrações em lagos e represas pode conduzir a um crescimento exagerado desses organismos (processo denominado eutrofização). Existe nas estações de tratamentos de efluentes a necessidade da adição de nitrogênio, pois, este elemento, faz parte das células dos microrganismos.

O nitrogênio presente nas águas residuárias, fica sujeito a processos bioquímicos e se convertem da amônia a nitrito e deste a nitrato, o que possibilita um consumo de oxigênio dissolvido do meio, podendo afetar tanto a vida aquática, quanto os níveis de oxigênio dissolvido numa estação de tratamento de efluentes com meios aeróbios, afetando a alcalinidade e prejudicando a sedimentação do lodo (SPERLING, 2012).

Quando ocorre desnitrificação no decantador secundário, as bolhas de nitrogênio podem arrastar sólidos em suspensão para a lâmina d'água superior, gerando perda de sólidos no sistema.

Formas de Nitrogênio (SPERLING, 2012):

Nitrogênio total = N (orgânico)+ $N\text{-}NO_2^-$ + $N\text{-}NO_3^-$, sendo:

Equilíbrio da amônia:

$N\text{-}NH3 + H^+ \longrightarrow N\text{-}NH_4^+$

Oxidação da amônia a nitrito:

$2\ N\text{-}NH_4^+ + 3\ O_2 \longrightarrow 2\ N\text{-}NO_2^- + 4\ H^+ + 2\ H_2O$ + Energia

Oxidação do nitrito a nitrato:

$2\ N\text{-}NO_2^- + O_2 \longrightarrow 2\ N\text{-}NO_3^-$ + Energia

A temperatura é uma medida da intensidade de calor. A importância da temperatura está relacionada as taxas de reações físicas, químicas e biológicas. A elevação de temperatura, diminui a solubilidade dos gases e aumenta a taxa de transferência de gases, o que pode gerar mau cheiro no caso de liberação de gases com odores desagradáveis, assim como temperaturas baixas tornam o ambiente com dificuldade de degradação (SPERLING, 2012).

Uma temperatura acima de 30 ºC, pode provocar inibição enzimática no meio, reduzindo a população de microrganismos ativos na estação de tratamento de efluentes. Amostras retiradas de lodo de estação de tratamento de efluentes mostram diferenças visíveis na microbiota, por alteração de apenas um grau na temperatura.

A temperatura tem grande influência na composição da comunidade de bactérias nitrificantes (SIRIPONG; RITTMANN, 2007; CHO *et al.,* 2014).

Temperaturas mais altas (> 15°C) favorecem o processo de nitrificação, especialmente quando a idade do lodo é baixa, enquanto que para elevada idade de lodo, quase não exercem influência significativa. Para temperaturas menores que 15°C e idade do lodo superior a 20 dias, o processo é instável (KOMOROWSKA-KAUFMAN; MAJCHEREK; KLACZYN, 2006).

O processo de desnitrificação, se ocorrer no decantador secundário de uma ETE com lodos ativados, pode provocar perda de sólidos suspensos totais em função da geração de nitrogênio, que arrastam os sólidos para superfície, e esta perda aumenta os resultados da DBO e DQO em função da presença de material particulado (HREIZ; LATIFI; ROCHE, 2015).

A idade de lodo é teoricamente, a quantidade de dias que os sólidos em suspensão voláteis ficam retidos no sistema, expresso pela relação entre a quantidade de sólidos em suspensão voláteis, em uma estação de lodos ativados e a quantidade de sólidos em suspensão voláteis descartada por dia. A carga de DQO na entrada de uma estação de tratamentos de efluentes e a idade do lodo influenciam nos valores da relação SSV/SST no tanque de aeração, sendo que menores cargas de matéria orgânica e maiores idades do lodo correspondem a menores razões de SSV/SST (FAN *et al.,* 2015).

2.4 FUNDAMENTAÇÃO TEÓRICA

Diversos temas envolvendo séries temporais e métodos de previsão diversos, tem sido usado em alguns trabalhos de pesquisa. Este trabalho, teve como base os métodos ligados a previsões ambientais e em estações de tratamento de efluentes, entre estes series temporais e uso de redes neurais.

2.4.1 -**Séries temporais**

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo, cujas características são que as observações vizinhas, tem uma dependência as quais queremos modelar (EHLERS, 2007).

Na decomposição clássica de uma série temporal são usadas três variáveis, como método para se determinar as variáveis vizinhas (X_t), conforme a equação (2.1):

$$X_t = T_t + C_t + R_t \tag{2.1}$$

Sendo:

T_t é uma componente de tendência;

C_t é uma componente cíclica ou sazonal que se repete a cada intervalo fixo de tempo;

R_t é uma componente aleatória ou ruído (a parte não explicada, que se espera, ser puramente aleatória).

2.4.2 Método de aferição dos resultados

Para se aferir os resultados obtidos de uma rede neural, verificando-se a diferença entre o resultado do método e os valores reais observados, são utilizados os seguintes métodos: MAPE, Erro Percentual Médio Relativo – cujo cálculo é feito conforme a Equação 2.2

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{d_t - Y_t}{d_t}\right| .100 \quad (2.2)$$

O RMSE - Raiz do Erro Médio Quadrado calculado através da Equação 2.3.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(d_t - y_t)^2}. \quad (2.3)$$

Sendo:

N o número de exemplos do conjunto de dados,

t o exemplo atual, apresentado à rede;

d_t a resposta desejada para o exemplo *t;* e

Y_t a resposta calculada pelo método, para o exemplo *t.*

2.4.3 Séries temporais estacionárias e não estacionárias.

Segundo Morettin e Tolloi (2006), quando uma série temporal se desenvolve no tempo aleatoriamente ao redor de uma média constante é dita estacionária. Na prática, a maioria das séries que encontramos apresenta algum tipo de não estacionariedade, por exemplo, tendência.

Segundo Stock e Watson (2004), para uma série de dados ser estacionária, suas variáveis não podem apresentar tendências e devem ser estáveis ao longo do tempo. Para as séries temporais, essa característica é fundamental para previsão do futuro, é importante que as variáveis sejam estacionárias ou passíveis de sua estacionariedade uma vez que as previsões futuras são baseadas nas previsões passadas.

Segundo Oliveira (2012), a maioria dos procedimentos de análise estatística de séries temporais supõe que estas sejam estacionárias, portanto se usam métodos para transformar os dados originais, se acaso estes não formarem uma série estacionária.

A transformação mais comum consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária. A primeira diferença Δ Z(t) é definida por;

$$\Delta Z_t = Z_t - Z_{t-1} \quad (2.4)$$

então, a segunda diferença é

$$\Delta^2 Z_t = \Delta(\Delta Z_t) = \Delta(Z_t - Z_{t-1}) \quad (2.5)$$

$$\Delta^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \quad (2.6)$$

De modo geral, a enésima diferença de Z(t) é

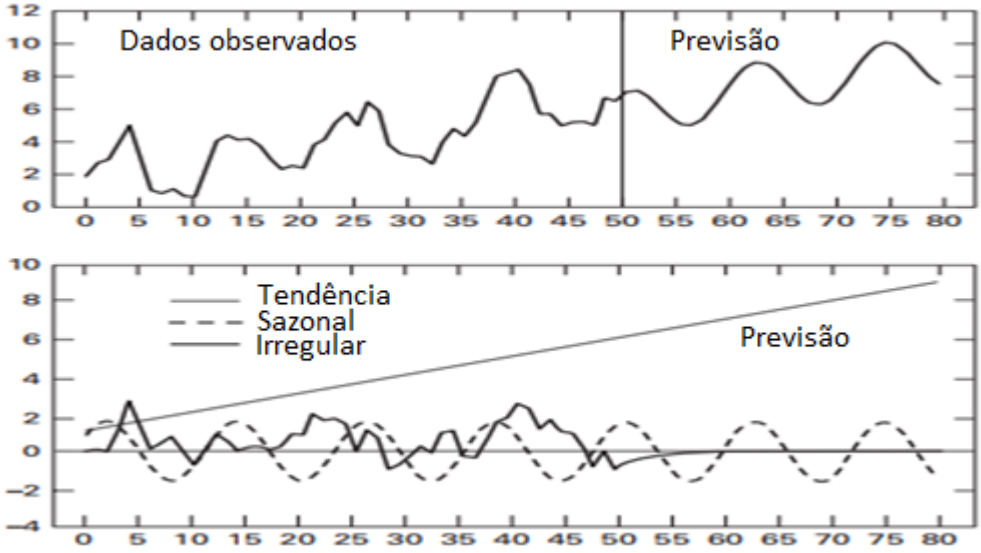
$$\Delta^2 Z_t = \Delta(\Delta^{n-1} Z_t) \quad (2.7)$$

Normalmente, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária.

Segundo Enders (2014), a visão moderna sustenta que uma série que contém elementos estocásticos terá componentes de tendência, sazonal e irregular. E o uso de técnicas matematicas é aconselhável para evitar estas complicações e para se fazer a projeção da tendência e dos componentes sazonais.

Na Figura 3, observamos que a tendência altera a média da série e o componente sazonal confere um regular padrão cíclico, com picos ocorrendo a cada doze unidades de tempo. Na prática, a tendência e os componentes sazonais não serão as funções deterministas simplistas mostradas.

Figura 3-Série temporal hipotética e sua decomposição para previsão.



Fonte: adaptado de Enders (2014)

A metodologia geral utilizada para fazer essas previsões implica encontrar a Equação de movimento conduzindo um processo estocástico e usando essa Equação, para prever subsequentes resultados. A exemplo da Figura 3, poderiamos dividir em três componentes a série temporal e definir suas equações como segue:

Tendência: $T_t = 1 + 0.1\,t$

Sazonalidade: $S_t = 1.6\ \mathrm{sen}(t\pi/6)$

Componente Irregular: $I_t = 0.7I_{t-1} + \varepsilon_t$

Sendo: T_t valor do componente de tendência no período t

S_t valor do componente sazonal em t

I_t valor do componente irregular em t

ε_t erro ou ruído do processo em t

Pela característica de otimização de eficiência, os resultados de remoção de carga orgânica biodegradável de estações de tratamento de efluentes biológicos tendem a uma série estacionária, podendo ter períodos de tendência no início do processo, mas com estabilidade de suas variáveis, em torno de valores médios de ajustes previstos no projeto.

Segundo Freitas (2007), as séries temporais com componentes sazonais, podem também se tentar torna-las estacionárias, por meio de uma transformação nos dados, definindo-se uma nova variável:

$$Z = (Y_t - \bar{Y})/S \qquad (2.8)$$

Sendo, $\bar{Y}$ e S respectivamente, a média e o desvio padrão amostrados. A estacionariedade é induzida pela Equação anterior. Esta transformação é aplicada para que a classe de modelos lineares e estacionários, possam ser empregados, uma vez que já existem testes estatísticos bem fundamentados nesta área.

Este artificio foi usado neste trabalho para se unidimensionalizar os dados.

2.4.3.1 *Teste de estacionariedade da série temporal***.**

Para se verificar se a série temporal pode ser considerada estacionária um dos testes que é utilizado é o de Dickey-Fuller

Consideramos o processo estocástico de forma

$$Y_{i\,t} = \vartheta Y_{i\,t-1} + \epsilon_i \tag{2.9}$$

Sendo, $|\vartheta| \leq 1$ e ε_i ruído branco. Se $|\vartheta| = 1$, temos o que é chamado de raiz unitária. Em particular, se $\vartheta = 1$, temos uma caminhada aleatória (sem deriva), que não é estacionária. Na verdade, se $|\vartheta| = 1$, o processo não é estacionário, enquanto se $|\vartheta| < 1$, o processo é estacionário.

O teste Dickey-Fuller é uma maneira de determinar se o processo acima possui uma unidade de raiz. A abordagem utilizada é bastante direta. Primeiro, calcule a primeira diferença, ou seja,

$$Y_{i\,t} - Y_{i\,t-1} = \vartheta Y_{i\,t-1} + \epsilon_i - Y_{i\,t-1} \tag{2.10}$$

Ou

$$Y_{i\,t} - Y_{i\,t-1} = (\vartheta - 1) Y_{i\,t-1} + \epsilon_i \tag{2.11}$$

Se usarmos o operador delta, definido por $\Delta Yi_{\,t} = Y\,i_{\,t} - Yi_{\,t-1}$ e definir $\beta = \vartheta - 1$, então a Equação 2.11, torna-se a Equação de regressão linear, ao qual β pode ser calculado usando o método dos mínimos quadrados.

$$\Delta Y_{i\,t} = \beta Y_{i\,t-1} + \epsilon_i \tag{2.12}$$

Ao qual $\beta \leq 0$, e portanto, o teste para ϑ é transformado em um teste para o parâmetro de inclinação

$\beta = 0$. Assim, temos um teste de uma união (uma vez que β não pode ser positivo), ao qual:

H_0: $\beta = 0$ (equivalente a $\vartheta = 1$)

H_1: $\beta < 0$ (equivalente a $\vartheta < 1$)

As verificações das hipóteses é de acordo com o numero Tau (τ), dado pela relação entre a inclinação β (obtida pelo método dos mínimos quadrados) e o desvio padrão da série S (Equação 2.13)

$$\tau = \frac{\beta}{S} \tag{2.13}$$

Séries estacionárias apresentarão menor desvio (S), pois tendem a oscilar em torno de uma média, com pontos acima e abaixo da mesma, se aproximando mais de uma destruição normal, ao contrário das séries não estacionárias.

Observa-se pela Equação 2.13, que quanto maior o desvio, menor o valor absoluto de τ, e maior a tendência de uma raiz unitária, e a série não ser estacionária e quanto menor negativamente o valor de τ, mais tendência a estacionariedade.

A intuição por trás do teste é a seguinte. Se a série é estacionária (ou tendência estacionária), então ele tem uma tendência a retornar a uma média constante (ou deterministicamente constante). Portanto, valores grandes tenderão a ser seguidos por valores pequenos (mudanças negativas) e valores pequenos por valores grandes (mudanças positivas). Assim, o nível da série será um preditor significativo da mudança do próximo período e terá um coeficiente negativo. Se, por outro lado, a série é integrada, então mudanças positivas e mudanças negativas ocorrerão com probabilidades que não dependem do nível atual da série; em um passeio aleatório .

Podemos usar a abordagem de regressão linear usual, exceto que quando a hipótese nula mantém o coeficiente t não segue uma distribuição normal e, portanto, não podemos usar o teste t típico. Em vez disso, esse coeficiente segue uma distribuição de Tau (τ) e, portanto, nosso teste consiste em determinar se a estatística τ (que é equivalente a t estatística t usual) é menor do que τ_{crit}, com base em uma Tabela de valores críticos das estatísticas de τ , mostradas na Tabela de Dickey-Fuller (Tabela 3).

Se o valor τ calculado for inferior ao valor crítico na Tabela de valores críticos, então temos um resultado significativo; Caso contrário, aceitamos a hipótese nula de que existe uma unidade de raiz e as séries temporais não são estacionárias.

Existem as seguintes versões do teste Dickey-Fuller (REAL-STATISTICS, 2018):

Tipo 0	Sem constante, sem tendência	$\Delta Y\, i = \beta_1\ Y\, i_{-1} + \varepsilon_i$
Tipo 1	Constante, nenhuma tendência	$\Delta Y\, i = \beta_0 + \beta_1\ Y\, i_{-1} + \varepsilon_i$
Tipo 2	Constante e tendência	$\Delta Y\, i = \beta_0 + \beta_1\ Y\, i_{-1} + \beta_2\, i + \varepsilon_i$

Cada versão do teste usa um conjunto diferente de valores críticos, como mostrado na Tabela Dickey-Fuller. É importante selecionar a versão correta do teste para a série temporal em análise. Observamos que o teste de tipo 2 pressupõe que existe um termo constante (que pode ser significativamente igual a zero).

Tabela 3 -Tabela de Dickey-Fuller

Distribuição Empírica Cumulativa de Tau				
		Nível de Significância		
N	1%	2.5%	5%	10%
Estatística Tau: Sem intercepto e sem tendência				
25,00	-2,66	-2,26	-1,95	-1,60
50,00	-2,62	-2,25	-1,95	-1,61
100,00	-2,60	-2,24	-1,95	-1,61
250,00	-2,58	-2,23	-1,95	-1,62
300,00	-2,58	-2,23	-1,95	-1,62
infinito	-2,58	-2,23	-1,95	-1,62
Estatística Tau: Com intercepto e sem tendência				
25,00	-3,75	-3,33	-3,00	-2,62
50,00	-3,58	-3,22	-2,93	-2,60
100,00	-3,51	-3,17	-2,89	-2,58
250,00	-3,46	-3,14	-2,88	-2,57
300,00	-3,44	-3,13	-2,87	-2,57
infinito	-3,43	-3,12	-2,86	-2,57
Estatística Tau: Com intercepto e com tendência				
25,00	-4,38	-3,95	-3,60	-3,24
50,00	-4,15	-3,80	-3,50	-3,18
100,00	-4,04	-3,73	-3,45	-3,15
250,00	-3,99	-3,69	-3,43	-3,13
300,00	-3,98	-3,68	-3,42	-3,13
infinito	-3,96	-3,66	-3,41	-3,12

Fonte: Adaptado de Real-estatísticas (2018)

2.4.4 Séries temporais univariada e multivariadas

Quando utilizamos uma só variável uma função de previsão é representada da seguinte forma:

$$Z_{t+h} = f(Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}, \dots) \quad (2.14)$$

Sendo: Z_{t+h} valores de previsão futura após intervalo de tempo h.

$Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}$, valores reais observados ou calculados da própria variável que se deseja estimar, a cada intervalo (0,1,2,3...) de tempo h anterior.

Em problemas reais, não é incomum que o valor de uma série seja afetado também por variáveis extrínsecas a ela, como, por exemplo, o resultado do pH de entrada de um reator anaeróbio afetando a eficiência de remoção de DQO.

Em tal situação, temos uma série temporal múltipla ou multivariada, Ao qual pode-se reescrever a Equação 2.14 como:

$$Z1_{t+h} = f(Z1_t, Z2_t, Z1_{t-1}, Z2_{t-1}, Z1_{t-2}, Z2_{t-2}, Z1_{t-3}, Z2_{t-3}, \dots) \quad (2.15)$$

Sendo: Z1 e Z2 duas variáveis distintas, as quais, o valor futuro $Z1_{t+h}$, será estimado em função de seus próprios valores observados ou calculados e ainda pelos valores observados ou calculados de outras variáveis que tem influência sobre o seu resultado ou que seja correlacionada a variável que se deseja estimar.

Os modelos de previsão multivariados procuram, de maneira análoga ao caso univariado, aproximar as S funções fs(.) de previsão das S séries temporais. Em contrapartida ao uso de múltiplos modelos univariados, esta abordagem permite capturar relações de interdependência entre as séries, resultando, potencialmente, em uma maior acurácia.

Vale salientar que, em alguns casos, a abordagem multivariada pode não apresentar ganhos de desempenho, ou até ser inferior ao de modelos univariados. Ocorrendo inclusão de variáveis que não se correlacionam, poderia haver uma diminuição na eficiência de precisão dos dados, então métodos de verificação de correlação devem ser usados antes de se usar modelos multivariados. Outra verificação se faz quanto ao período de previsões mais longas podem dar resultados fora do padrão, períodos de previsões curtos são mais próximos do real.

Segundo Freitas e Oliveira (2016), séries temporais hierárquicas (ou agrupadas, quando possuem mais de uma variável de agregação), são as que podem ser organizadas em níveis, por exemplo: o tipo de produto, região geográfica, classe de consumo, dentre outros. As informações

referentes à previsão destas séries, são fundamentais para a tomada de decisão seja no nível gerencial ou operacional de todo tipo de negócio. Para atender a essas informações, são utilizadas técnicas de previsão hierárquica, que têm como foco reduzir os custos e melhorar a acurácia da previsão.

2.4.5 Série Temporal Discreta e Contínua

Uma série temporal é dita discreta quando as observações são feitas em instantes de tempos específicos, usualmente igualmente espaçados (Exemplo: temperatura média mensal) (EHLERS, 2009).

Uma série temporal é dita ser contínua quando as observações são feitas continuamente no tempo. Exemplo: medidas de temperatura a partir de um termógrafo (EHLERS, 2009). Séries contínuas podem ser discretizadas, ou seja, seus valores são registrados a certos intervalos de tempo

As séries temporais discretas podem surgir de várias formas, por exemplo, médias de valores diários de dados de temperatura do ar, umidade relativa do ar.

Os dados de uma estação de tratamento de efluentes podem ser contínuos ou discretos, as análises feitas em laboratório, são discretas.

2.4.6 Modelos de previsão, preenchimento de dados faltantes e suavização em séries temporais.

Existem vários modelos de séries temporais, envolvendo sistemas simples de previsão, preenchimento de dados faltantes e suavização das curvas para redução de ruídos, dentro destes modelos podemos citar:

Média (método usado em séries estacionárias);

Média Móvel Simples (método usado em séries estacionárias);

Média Móvel Ponderada (método usado em séries estacionárias);

Média Móvel Dupla (método usado em séries estacionárias);

Previsão por interpolação polinomial (mínimos quadrados)

Amortecimento exponencial simples (método usado em séries estacionárias);

Amortecimento exponencial duplo (Brow);

Amortecimento exponencial duplo de Holt;

ARIMA (Modelo autoregressivo, integrativo com média móvel);

SVM (*Support Vector Machine,* vetor suporte de máquina);

Redes Neurais (NN);

Feedforward ;

CNN (Rede Neural Convolutiva);

Wavelet Feedforward;

Wavelet SVM;

Acumulada SVM;

Acumulada *Feedforward.*

2.4.6.1 Média

A previsão pela média consiste em usar como previsão do valor próximo futuro o valor observado das últimas médias do período

Trata-se de um método univariado ao qual o valor futuro é dado pela expressão:

$$Z_{t+h} = \frac{\sum_0^n (Y_{t-nh})}{n} \quad (2.16)$$

Sendo: n é o número de intervalos de tempos anterior observados.

Z_{t+h} são valores previsão futura após intervalo de tempo h.

$Z_{t-nh,}$ são valores reais observados da própria variável que se deseja estimar, a cada intervalo de tempo h anterior.

2.4.6.2 Média Móvel Simples (MMS)

Segundo Zan e Sellito (2007), na previsão pelo método da média móvel simples, calcula-se a média de uma série temporal para n períodos de tempos reais mais recentes conforme a Equação 2.17.

$$Z_{t+h} = \frac{Z_t + Z_{t-1} + Z_{t-2} \ldots Z_{t-(n+1)}}{n} \quad (2.17)$$

Na qual:

n é número total de períodos na média;

2.4.6.3 Média móvel ponderada (MMP)

Segundo Zan e Sellitto (2007), na previsão por média móvel ponderada se usa pesos dentro de um período anterior n, usados na previsão ao qual se atribui pesos maiores aos dados mais recentes tendo como regra que o somatório dos pesos deve ser igual a 1.

$$Z_{t+h} = P1Z_t + P2Z_{t-1} + P3Z_{t-2} \ldots PnZ_{t-(n+1)} \quad (2.18)$$

na qual:

n Número total de períodos na média;

Z_{t+h} Previsão para o período t+1.

P1+P2+....+Pn = 1

2.4.6.4 Média móvel dupla (MMD)

Utilizando-se a média móvel dupla se faz necessário se calcular a média móvel simples (M_t)

$$M_t = \frac{Z_t + Z_{t-1} + Z_{t-2} \ldots Z_{t-(n+1)}}{n} \quad (2.19)$$

Em seguida se faz a segunda média com os valores das médias anteriores, ou seja, a média móvel das médias moveis (M'_t).

$$M'_t = \frac{M_t + M_{t-1} + M_{t-2} \ldots M_{t-(n+1)}}{n} \quad (2.20)$$

Esta série é usada quando existe uma tendência dos dados, assim se calcula a_t e b_t, ao qual:

$$a_t = M_t + (M_t - M'_t) = 2\,M_t - M'_t \quad (2.21)$$

$$b_t = \frac{2}{n-1}(M_t - M'_t) \quad (2.22)$$

Sendo a previsão dada por:

$$Z_{t+h} = a_t + h\,b_t \quad (2.23)$$

2.4.6.5 Interpolação polinomial (mínimos quadrados)

A previsão polinomial pelos métodos dos mínimos quadrados, seja ela de forma linear ou não, é muito usada para prever tendências em séries temporais.

O método consiste em tornar mínimo o somatório dos quadrados das distâncias entre um polinômio $P(x) = A_0 + A_1 x_j^{i_1} + A_2 x_j^{i_1} + \cdots + A_m x_j^{i_m}$ e os pontos (x_j, y_j), interpolando a função, que é determinada por cálculo dos coeficientes $A_0, A_1 \ldots A_m$, conforme segue:

$$P(x_j) = A_0 + A_1 x_j^{i_1} + A_2 x_j^{i_1} + \cdots + A_m x_j^{i_m} \quad (2.24)$$

Ou ainda $$P(x) = \sum_{m=0}^{m} A_m x_j^{i_m} \quad (2.25)$$

Sendo:

y_j valor real observado da variável que se quer interpolar

$A_0, A_1, A_2 \ldots\ldots A_m$, constantes do polinômio interpolador, a serem determinadas;

$i_1, i_2, i_3 \ldots\ldots i_m$, coeficientes do polinômio de grau m, que são números reais (inteiros ou não, positivos, negativos ou nulos), usados na escolha do polinômio interpolador;

m número de termos do polinômio interpolador (variando de 0 a m);

J número de pontos observados J=1, 2, 3, ...,,n.

Temos que a distância entre o ponto de coordenadas (x_j, y_j) e P(x_j) é dado por R_j, conforme Equação 2.26:

$$Rj = p(x_j) - f(x_j) \quad (2.26)$$

Ou

$$Rj = A_0 + A_1 x_j^{i_1} + A_2 x_j^{i_1} + \cdots + A_m x_j^{i_m} - f(x_j) \quad (2.27)$$

Ao qual: J= 0, 1, 2, 3,.....,n;

n = número de pontos.

A função do resíduo da distância, entre o polinômio e os pontos será:

$$\varphi(a_0, a_1, a_{2,\ldots}, a_m) = \sum_{j=0}^{n} R_j^2 = \varphi \quad (2.28)$$

Podemos minimizar o resíduo entre o polinômio e os pontos, para isto a derivada da função 2.28, deve se igualar a zero.

$$\frac{\partial \varphi}{\partial a_0} = 0, \ \frac{\partial \varphi}{\partial a_1} = 0, \ \ldots, \frac{\partial \varphi}{\partial a_m} = 0 \quad (2.29)$$

Assim também como as derivadas parciais de φ, pela regra da cadeia, obtemos o sistema 2.30:

$$\frac{\partial \varphi}{\partial a_0} = \sum_{j=0}^{n} \frac{\partial \varphi}{\partial R_j}\frac{\partial R_j}{\partial a_0} = 0, \frac{\partial \varphi}{\partial a_1} = \sum_{j=0}^{n} \frac{\partial \varphi}{\partial R_j}\frac{\partial R_j}{\partial a_1} = 0, \bullet\bullet\bullet , \frac{\partial \varphi}{\partial a_m} = \sum_{j=0}^{n} \frac{\partial \varphi}{\partial R_j}\frac{\partial R_j}{\partial a_m} = 0$$

(2.30)

Resolvendo as derivadas parciais, de 2.29, obtemos:

$$\frac{\partial \varphi}{\partial R_j} = 2\sum_{j=0}^{n} R_j \quad (2.31)$$

E de 2.27, temos:

$$\frac{\partial R_j}{\partial a_m} = x_j^{i_m} \quad \text{ao qual m= 0, 1, 2, ..., m} \quad (2.32)$$

Substituindo em 2.4.6.15 usando as equações 2.4.31 e 2.4.32, temos o sistema: $\sum_{j=0}^{n} R_j\ x_j^0 = R_0\ x_j^0 + R_1\ x_j^0 + R_2\, x_j^0 + \ldots + R_m\ x_j^0 = 0$ $\sum_{j=0}^{n} R_j\ x_j^1 = R_0\ x_j^1 + R_1\ x_j^1 + R_2\, x_j^1 + \ldots + R_m\ x_j^1 = 0$

..... (2.33)

.....

$$\sum_{j=0}^{n} R_j\ x_j^{i_m} = R_0\ x_j^{i_m} + R_1\ x_j^{i_m} + R_2\, x_j^{i_m} + \ldots + R_m\ x_j^{i_m} = 0$$

Substituindo o valor de Rj no sistema de equações 2.33, temos o sistema 2.34:

$$\sum_{j=0}^{n} a_0 x_j^0 + \sum_{j=0}^{n} x_j^1 a_1 + \sum_{j=0}^{n} x_j^2 a_2 + \cdots + \sum_{j=0}^{n} x_j^m a_m - \sum_{j=0}^{n} f(x_j) x_j^0$$

$$\sum_{j=0}^{n} a_0 x_j^1 + \sum_{j=0}^{n} x_j^2 a_1 + \sum_{j=0}^{n} x_j^3 a_2 + \cdots + \sum_{j=0}^{n} x_j^{m+1} a_m - \sum_{j=0}^{1} f(x_j) x_j^1$$

..... (2.34)

.....

$$\sum_{j=0}^{n} a_0 x_j^m + \sum_{j=0}^{n} x_j^{m+1} a_1 + \sum_{j=0}^{n} x_j^{m+2} a_2 + \cdots + \sum_{j=0}^{n} x_j^{m+m} a_m - \sum_{j=0}^{n} f(x_j) x_j^m$$

Definindo:

$S_t = \sum_{j=0}^{n} x_j^t$ com t=0, 1,2,3,....,2m

$K_l = \sum_{j=0}^{n} f(x_j) x_j^l$ com l=0,1,2,3,...,m

Obtemos um sistema de equações lineares, que pode ser escrita da forma:

$$S_0\ a_0 + S_1\ a_1 + \cdots S_m\, a_m = K_0$$

$$S_1\ a_0 + S_2\ a_1 + \cdots S_{m+1}\, a_m = K_0$$

..... (2.35)

.....

$$S_m\, a_0 + S_{m+1}\ a_1 + \cdots S_{m+m}\, a_m = K_m$$

Resolvendo o sistema linear 2.35, temos os valores de $a_0, a_1, \ldots, a_m$, que são as constantes do polinômio de interpolação P(x).

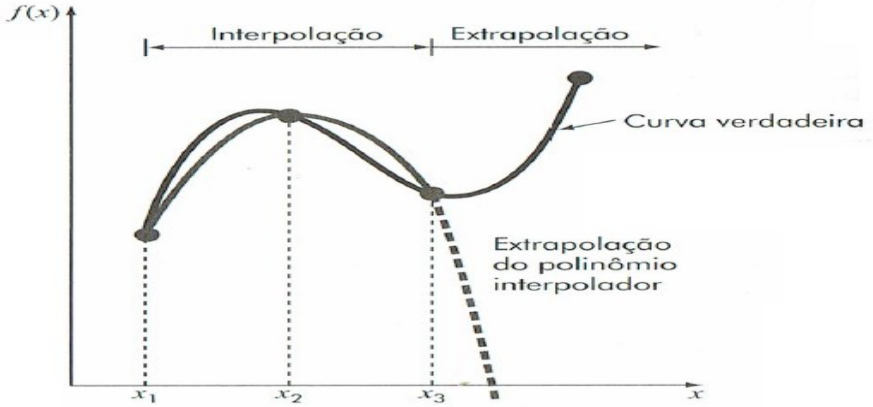
2.4.6.6 Extrapolação de dados

Durante as pesquisas foram encontrados trabalhos citando ajustes de modelos matemáticos, inclusive rede neural, mostrando como predição os erros e correlações de treinamento da rede como erro de previsão do método. Na verdade, um modelo matemático pode se ajustar perfeitamente aos pontos de treinamento, porém os resultados para pontos fora do treinamento (caso de previsão), nem sempre têm valores ajustados ao modelo, em função de extrapolação de dados. Isto também ocorre em redes neurais, especialmente quando se eleva muito a quantidade de neurônios ou graus de liberdade de uma rede.

O software MATLAB, separa alguns dados para validação e teste da rede neural, que são independentes dos dados usados no treinamento, para mostrar se o modelo realmente reproduz as correlações existentes, em dados que não pertencem aos de ajuste do modelo, assim também como parar o treinamento, no caso dos erros médios quadráticos de validação começarem a subir ou não apresentar mudanças durante certo número de interações da rede.

Segundo Chapra (2013), a extrapolação é o processo de fazer uma estimativa de um valor f(x)que está fora dos pontos conhecidos, x_1, x_2, x_3....x_n. Conforme a Figura 4, a extrapolação estende a curva a uma região desconhecida, divergindo da previsão. Polinômios de graus elevados usados para ajustar uma curva, podem também levar a extrapolação ou oscilações dos dados entre dois pontos (equações que apesar de dar resultados dentro dos pontos usados para cálculo do polinômio, podem dar valores muito fora da curva desejada).

Figura 4-Ilustração da possível divergência de uma previsão extrapolada



Fonte: Chapra (2013)

2.4.6.7 Amortecimento exponencial simples (AES)

O amortecimento exponencial simples é utilizado em séries temporais que não têm tendência ao qual os dados oscilam em torno de uma média. A previsão (P_{t+1}) é dada por:

$$P_{t+1} = \alpha Z_t + \alpha (1-\alpha) Z_{t-1} + \alpha (1-\alpha)^2 Z_{t-2} + \cdots \quad (2.36)$$

Ou ainda:

$$P_{t+1} = \alpha Z_t + (1-\alpha) P_t \quad (2.37)$$

Sendo:

α o coeficiente de amortecimento (0≤ α ≤ 1);

n Número total de períodos na média;

$Z_{t-nh,}$ valores reais observados da própria variável que se deseja estimar, a cada intervalo de tempo h anterior.

Quando α é próximo de 1 a previsão é mais sensível ao último valor observado.

Geralmente adota-se $P_0 = Z_0$ ou $P_0 = (\Sigma Z_t)/n$

2.4.6.8 Amortecimento Exponencial Duplo (Método de Brown) -AED

Este método é utilizado em séries temporais com tendência, ao qual existem atribuições diferentes aos dados históricos da série. Se calculam dois amortecimentos, ao qual o primeiro amortecimento (A_t) dado por:

$$A_t = \alpha Z_t + (1-\alpha) A_{t-1} \quad (2.38)$$

O segundo amortecimento (A'_t) é dado por:

$$A'_t = \alpha A_t + (1-\alpha) A'_{t-1} \quad (2.39)$$

Calculado os amortecimentos, se determina suas diferenças (a_t), dada por:

$$a_t = 2 A_t - A'_t \quad (2.40)$$

E um fator de ajuste adicional (b_t), dado por:

$$b_t = \frac{\alpha}{1-\alpha}(A_t - A'_t) \quad (2.41)$$

Sendo a previsão dada por:

$$Z_{t+h} = a_t + h\, b_t \quad (2.42)$$

Sendo:

h número de períodos futuros a serem previstos

2.4.6.9 ARIMA (Auto Regressivo integrado de média móvel)

Segundo Oliveira (2007), o modelo Box e Jenkins ARIMA (Auto Regressivo integrado de média móvel), são modelos utilizados em séries estacionárias ou não, existindo vários programas de computador desenvolvidos usando este método.

Segundo Franco (2011), um modelo Auto Regressivo (AR), se deve ao fato de que Y_t no instante t é função dos Y's nos instantes anteriores a t.

$$\mathrm{AR(p)}: Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} \quad (2.43)$$

O nome Médias Móveis (MA), vem do fato que Y_t é uma função soma algébrica ponderada dos μ_t que se movem no tempo.

$$\mathrm{MA(q)}: Y_t = \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \quad (2.44)$$

O nome ARMA, designa o modelo que tem tanto uma parte AR ($\varphi(\cdot) \neq 1$) como uma parte MA ($\theta(\cdot) \neq 1$).

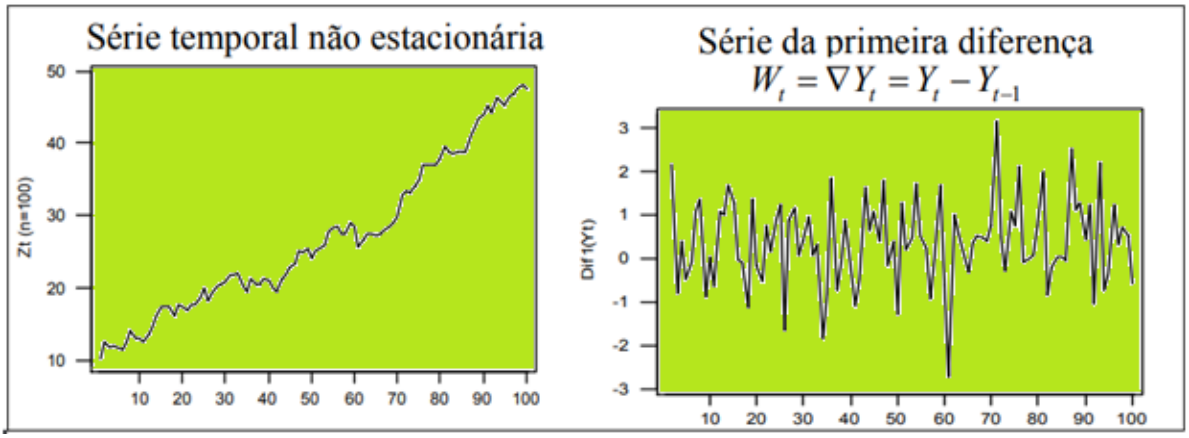
$$ARMA(p, q):$$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \mu_t - \theta_1 \mu_{t-1} - \theta_2 \mu_{t-2} - \cdots - \theta_q \mu_{t-q} \quad (2.45)$$

O modelo ARIMA supõe que a d-ésima diferença da série Yt pode ser representada por um modelo ARMA, estacionário e inversível. Na maioria dos casos usuais, d = 1 ou 2.

Na Figura 5, se encontra a transformação, considerando uma diferença, em uma série com tendência.

Figura 5-Transformação de série não estacionária em série estacionária usando a primeira diferença.



Fonte:Adaptado de Oliveira (2007)

Os passos básicos da Metodologia Box e Jenkins, seguem uma metodologia que envolvem etapas básicas a seguir (OLIVEIRA, 2007):

- Diferenciar a série para se obter estacionariedade, para isto se plota a curva do correlograma da série a partir de diferenças Δy, $\Delta^2 y$,... $\Delta^n y$. Este processo se faz até o correlograma cair dentro de faixas que demonstrem estacionariedade;

- Escolher um modelo para decidir a ordem apropriada AR e MA, baseado no correlograma feito com as diferenças. O correlograma de um modelo MA é zero depois de um certo ponto, enquanto o de um processo AR declina geometricamente. Os correlogramas do processo ARMA apesar de demostrar diferentes padrões, eles caem depois de certo ponto;

- Estimativa do modelo. O próximo passo é a estimativa do modelo identificado no passo anterior. A estimativa de um modelo AR é direta, ele é utilizado o método dos mínimos quadrados ordinários (MQO) para minimizar a soma dos erros quadrado, $\sum \epsilon^2$. No caso dos modelos MA, não se pode escrever a soma dos erros quadrado simplesmente como uma função dos dados observados e dos parâmetros como no modelo. O que se pode fazer é escrever a matriz de covariância do erro da média-móvel e, assumindo normalidade, usar o método de máxima verossimilhança para estimativa (MVE). Para modelos ARMA, novamente aparece o problema com o componente MA, porém pode-se também utilizar o (MVE);

-Diagnóstico. Quando um modelo AR, MA ou ARMA é ajustado para uma dada série temporal, existem dois critérios frequentemente usados que refletem a qualidade do ajuste e o número de parâmetros estimado. Um deles é o Critério de Informação de Akaike (AIC) e o outro

é o Critério de Informação Bayesiano de Schwartz (BIC). Se k é o número total de parâmetros estimado, tem-se que:

$$AIC(k) = n \log \hat{\sigma}^2 + 2k \quad (2.46)$$

$$BIC(k) = n \log \hat{\sigma}^2 + k \log n \quad (2.47)$$

Sendo: n o tamanho da amostra;

$$RSS = \sum \hat{\varepsilon}_t^2 \quad (2.48)$$

$$\hat{\sigma}^2 = RSS/(n-p) \quad (2.49)$$

p número escolhido de dias anteriores usado no modelo auto regressivo (AR)

Se vários modelos ARMA estão sendo considerados, escolhe-se aquele com os menores AIC e BIC. Os dois critérios podem levar a diferentes conclusões;

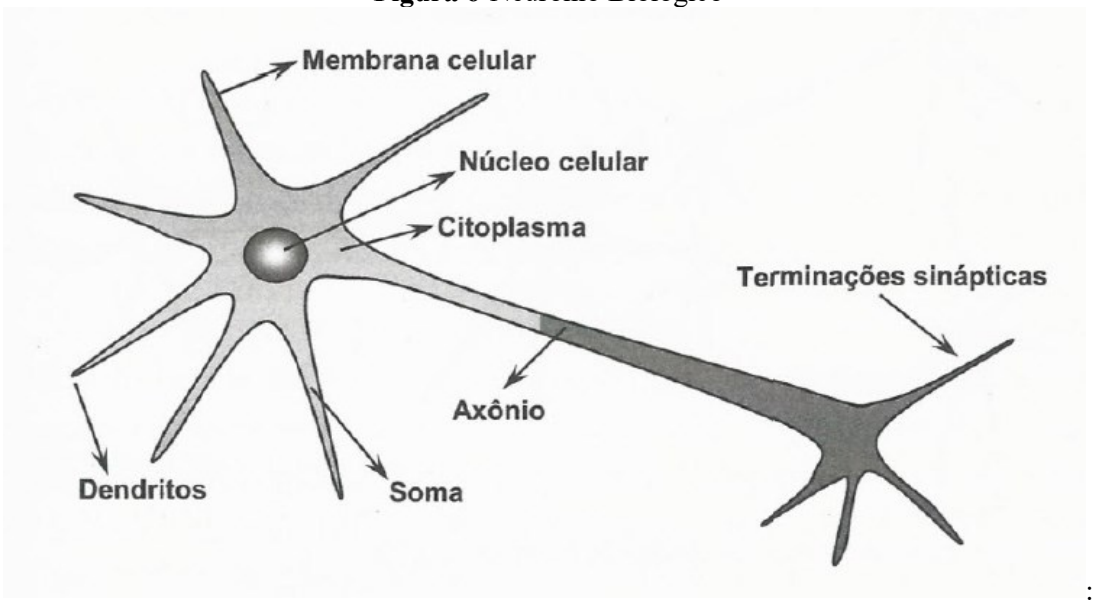
-Previsão. Supondo que um modelo tenha sido estimado com as observações, quer se prever Y_{n+k} , chamada de previsão de k-períodos à frente. Primeiro é necessário escrever a expressão para Y_{n+k} então substituir todos os valores futuros Y_{n+j} $(0 < j < k)$ pelas suas previsões e ϵ_{n+j} (j > 0) por zero (uma vez que seu valor esperado é zero). Também são substituídos todos os ϵ_{n-j} (j ≥ 0), pelos resíduos preditos.

2.4.7 Redes Neurais Artificiais

As redes neurais artificiais são modelos matemáticos inspirados no sistema nervoso dos seres vivos. Possuem capacidade de manutenção e aquisição de conhecimento, baseado em informações, que podem ser processadas em um conjunto de unidades por um computador.

O processamento de informações no cérebro humano é regido por elementos processadores biológicos que operam em paralelo e os modelos artificiais são sequenciais. O neurônio biológico é constituído por dendritos, pela membrana celular, núcleo celular e citoplasma e ligado a este está o axônio com suas terminações sinápticas conforme esquematizado na Figura 6.

Figura 6-Neurônio Biológico



:

Fonte: Silva, Spatti e Flauzino (2010)

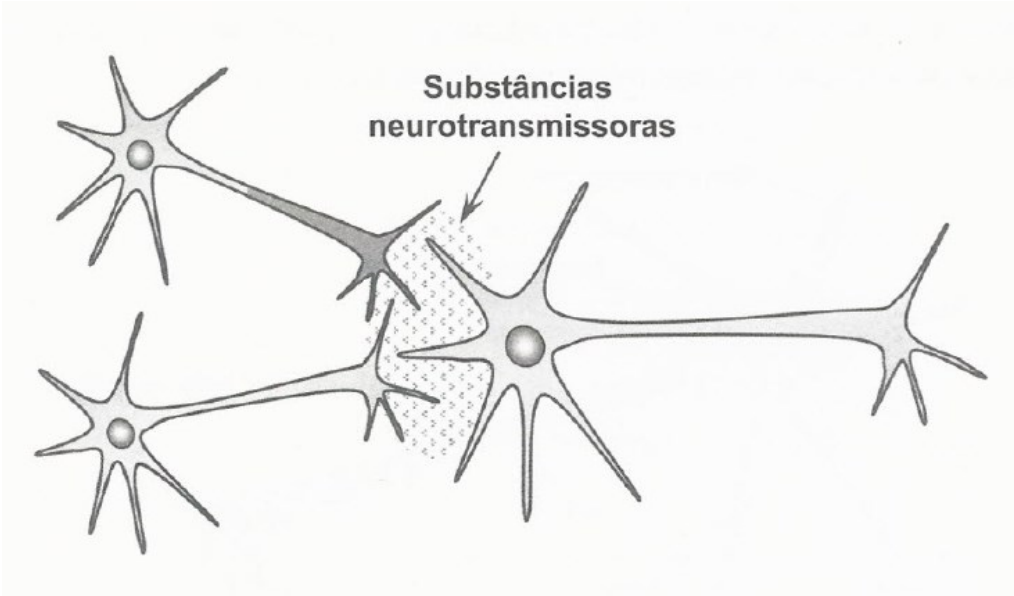
Os dendritos são constituídos por vários finos prolongamentos que formam a árvore dendrital, e a principal função desses dendritos, consiste em captar de forma contínua os estímulos (impulsos elétricos) vindos de diversos outros neurônios ou do próprio meio externo, ao qual os mesmos estão em contato (neurônios sensitivos).

O corpo celular incumbência de processar as informações (sinais elétricos) advindas dos dendritos a fim de produzir o potencial de ativação, que indicará se um neurônio poderá disparar um impulso elétrico ao longo do axônio.

O axônio é constituído por um único prolongamento cuja missão é conduzir os impulsos elétricos para outros neurônios conectores ou para aqueles que se conectam diretamente com tecido muscular chamados de neurônios efetuadores (Figura 7).

As terminações do axônio são constituídas de ramificações denominadas de sinápticas e estas transferem os impulsos elétricos recebidos para outros detritos e daí o sinal é transmitido para sinapse de outros neurônios (SILVA; SPATI; FLAUZINO, 2010).

Figura 7-Conexões sinápticas entre neurônios



Fonte: Silva, Spatti e Flauzino (2010)

Similar ao cérebro humano em uma rede neural artificial os dados de entrada de um modelo matemático com seus respectivos pesos fazem o papel dos sinais elétricos recebidos pelos dendritos. Pois estes dados de acordo com o somatório dos pesos atribuídos a cada entrada, são comparados com uma função matemática denominada função de ativação (função esta, similar a capacidade do neurônio biológico de medir os impulsos elétricos recebidos e conforme o diferencial de potencial elétrico, ele enviar ou ativar o sinal para os axônios).

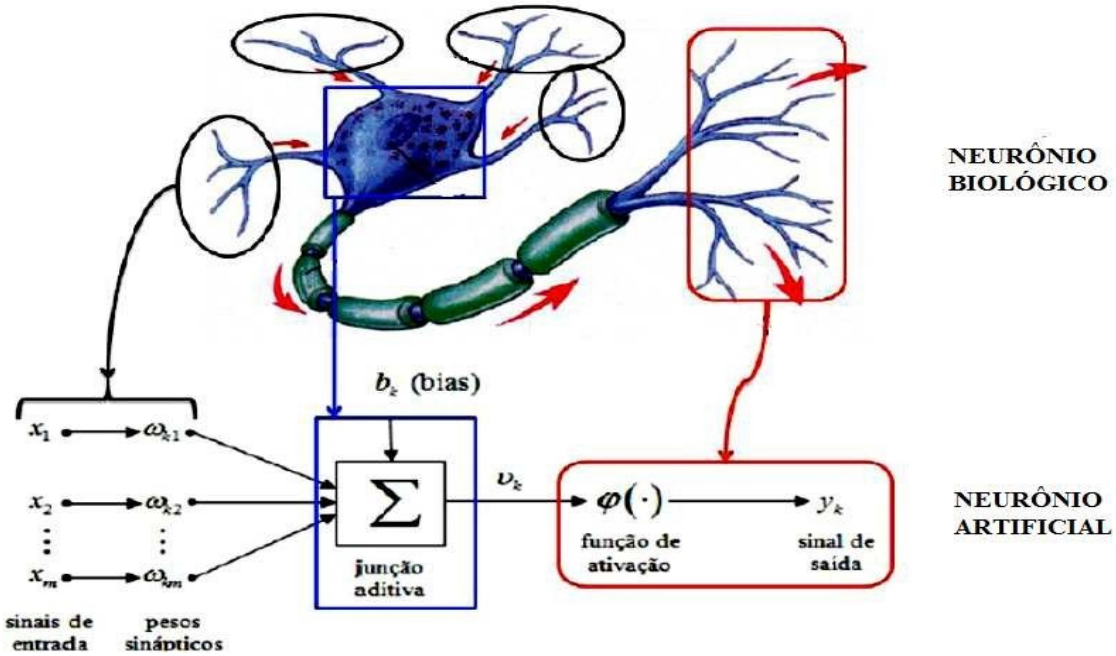
No neurônio artificial uma vez ativada a função matemática, se associa a saída da função a um novo peso e o resultado desta nova função servirá para alimentar novos neurônios em outras camadas da rede neural artificial, o processo continua até se atingir a camada de saída ao qual o resultado do sinal é comparado com padrões definidos em períodos de treinamento da rede, não se atingindo um valor de erro menor ao desejável na saída os pesos são reajustados novamente na rede até os valores convergirem ou se aproximarem ao máximo do valor desejado na saída.

A diferença entre o sinal de saída e o desejado dentro do período de treinamento da rede neural é tratada matematicamente, de forma que a diferença seja minimizada na camada de saída, para isto se faz os reajustes dos pesos em cada entrada dos neurônios artificiais.

A Figura 8, apresenta uma representação esquemática comparativa entre um neurônio real e um neurônio artificial.

Segundo Braga, Carvalho e Ludermir (2007) as principais utilizações de redes neurais são para tarefas de classificação, tais como, reconhecimento de caracteres, reconhecimento de imagens, diagnósticos médico, diagnósticos de equipamentos, análise de risco de crédito, detecção de fraudes e detecção de falhas em sistemas industriais, categorização, agrupamento de sequência de DNA, mineração ou classificação de dados, análise de expressão gênica e agrupamento de clientes, previsão de tempo, previsão financeira, modelagem de sistemas dinâmicos e previsão de sequência de DNA.

Figura 8-Representação esquemática comparativa entre o neurônio real e neurônio artificial.



Fonte: Souza (2011)

Uma das redes neurais que é muito utilizada para previsão em séries temporais com variáveis múltiplas com tendências e sazonalidades é a que usa o algoritmo de retropropagação do erro.

Segundo Ludwig Jr e Montgomery (2007), um algoritmo de retropropagação pode ser descrito em cinco etapas a seguir:

A primeira etapa, consiste em se arbitrar aleatoriamente pesos sinápticos e níveis de bias, em distribuição uniforme de forma que a média destes pesos seja zero.

$$\frac{\sum_1^n W_{ijk}}{n} = 0 \quad (2.50)$$

$$\frac{\sum_1^n b_{ij}}{n} = 0 \quad (2.51)$$

Sendo:

i se refere à camada;

J se refere a posição do neurônio na camada;

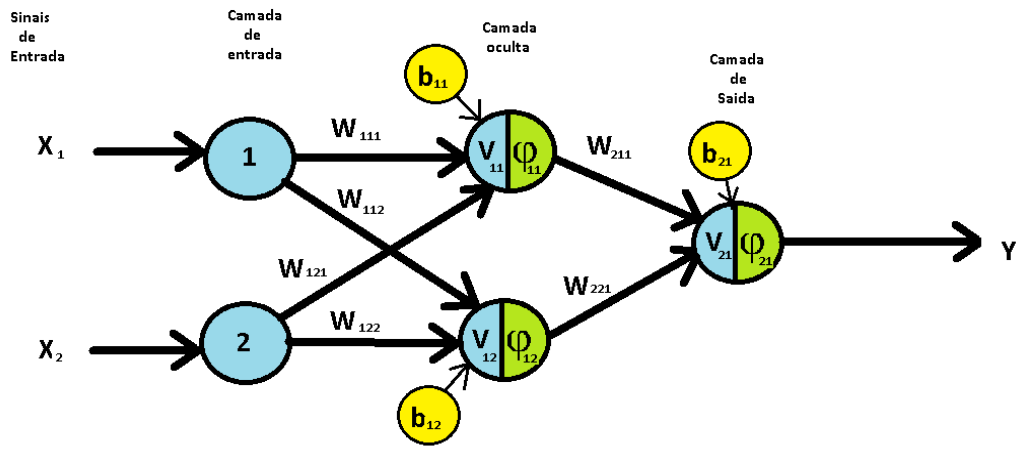
K se refere a linha de destino da propagação do sinal;

W se refere ao peso sináptico que inicialmente é atribuído aleatoriamente;

b se refere ao valor das bias que inicialmente é atribuído aleatoriamente.

Um esquema básico de uma rede neural, com uma camada de entrada, uma camada oculta e uma camada de saída, mostrando os pesos sinápticos e as bias pode ser visto na Figura 9.

Figura 9-Esquema básico de uma rede neural, mostrando os bias e os pesos sinápticos que devem ser arbitrados inicialmente



Fonte: o autor

A segunda etapa consiste em utilizar os exemplos de treinamento, ao qual para cada um, realiza-se a propagação dos sinais e a retropropagação dos erros com correção dos pesos e níveis de bias.

A propagação dos sinais segue conforme Equação 2.52 e 2.53 a seguir:

Propagação para a primeira camada ou camada de entrada:

$$V_{ij} = \sum_{j=1}^{n} W_{ijk} * \omega_{(i-1)j} + b_{ij} \quad (2.52)$$

Sendo:

n o número de neurônios da camada anterior à camada i;

i refere à camada;

j refere a posição do neurônio na camada;

W peso atribuído a cada sinal que sai do neurônio da camada anterior;

ω refere ao sinal de saída do neurônio que envia o sinal ao próximo neurônio ou a saída da rede (neste caso está se considerando uma saída sigmoide).

Nota: Para a primeira camada ou camada de entrada os valores de δ são iguais aos do sinal de entrada

$$\omega_{ij} = \frac{1}{1+\epsilon^{-V_{ij}}} \quad (2.53)$$

Na Figura 10, temos a tipologia de uma rede neural *Multilayer Perceptron* (MLP), o sinal se propaga até se obter o sinal de saída aplicado a última camada s, que é comparada por um valor desejado (d) ou valor obtido durante o período de experiência ou treinamento da rede.

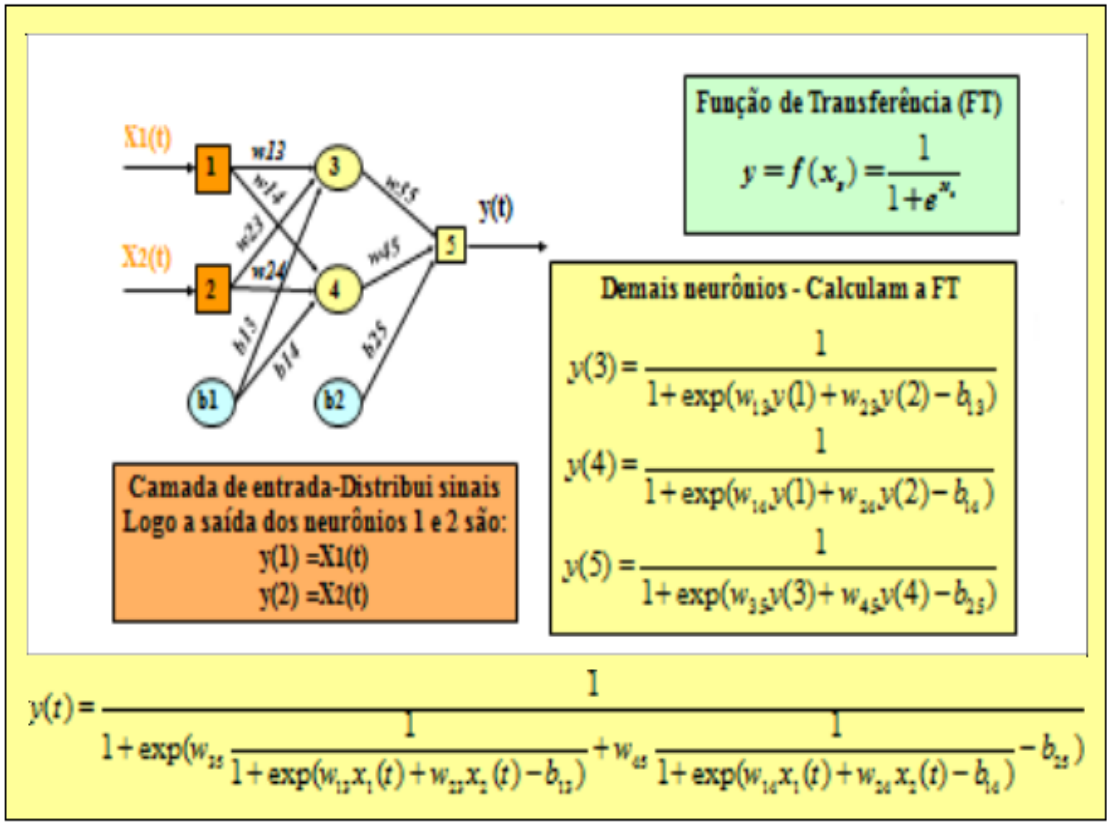
A diferença entre o sinal de saída (Y_j (n)) e o sinal desejado é o erro (e) obtido na iteração (n) da rede neural.

$$e_j = d_j - Y_j(n) \quad (2.54)$$

O sinal de erro global e instantâneo E(n) da rede, para os j neurônios é dado pela Equação 2.55, ao qual o erro de cada neurônio é elevado ao quadrado, para se evitar que um erro negativo de um neurônio em relação a resposta desejada compense ou reduza o valor positivo encontrado em outra saída ou neurônio de saída.

$$E(n) = \frac{1}{2}\sum_{1}^{j} e_j(n)^2 \quad (2.55)$$

Figura 10-Tipologia de uma rede neural tipo *Multilayer Perceptron* (MLP)



Fonte:Adaptdo de Carvalho et al. (2007).

O erro global médio para todo um conjunto de treinamento ($\bar{E}$) ou seja para um número N de interações é dado pela média aritmética dos erros globais instantâneos, conforme Equação 2.56;

$$\bar{E} = \frac{1}{2}\sum_{n=1}^{N} E(n) \tag{2.56}$$

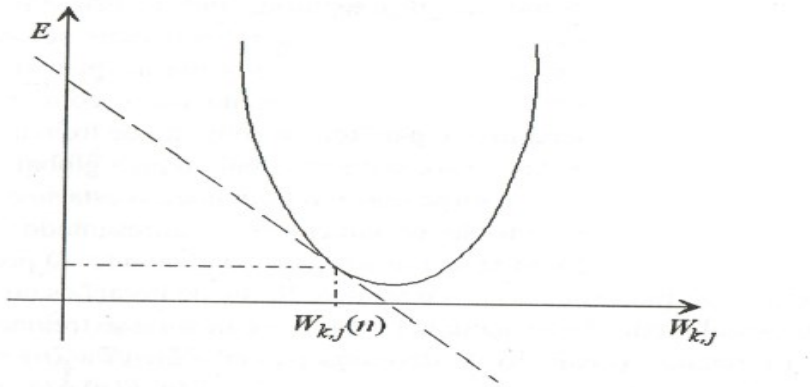
Segundo Ludwig Jr e Montgomery (2007) o erro global instantâneo médio e o erro global médio é função dos pesos sinápticos e das bias. Assim se procura ajustar estes parâmetros de forma a se ter o mínimo erro desejado.

A Equação do erro global instantâneo pode ser escrita em função dos pesos:

$$E = \frac{1}{2}\sum_{n}(d_n - \omega(\sum_k W_{ok}\ \omega(...\omega(\sum_{kj} W_{ok}\ \omega(...\omega(\sum_l W_{li}\ x_i))))))^2 \tag{2.57}$$

A derivada da função erro instantâneo E(n) na enésima interação (n) em relação ao peso sináptico (W_{kj}) é representado pela tangente na Figura 11.

Figura 11 -Representação gráfica da função Erro x Pesos sinápticos



Fonte: Ludwig Jr e Montgomery (2007)

Derivadas parciais são usadas então para se ajustar os pesos sinápticos de forma a se obter um mínimo de erro ao qual as variações de pesos sofrem correções a cada interação(n), segundo uma taxa de aprendizado (n) conforme Equação 2.58, conhecida como a regra delta:

$$\Delta W_{ij}(n) = -n\frac{\partial E(n)}{\partial W_{ij}(n)} \tag{2.58}$$

Aplicando-se a regra da cadeia obtemos:

$$\frac{\partial E(n)}{\partial W_{ij}(n)} = \frac{\partial E(n)}{\partial e_j(n)}\frac{\partial e_j(n)}{\partial y_j(n)}\frac{\partial y_j(n)}{\partial V_j(n)}\frac{\partial V_j(n)}{\partial W_{ij}(n)} \tag{2.59}$$

O gradiente local do neurônio $\partial E(n)/\partial V_j(n)$, pode ser expresso como:

$$\frac{\partial E(n)}{\partial V_j(n)} = \frac{\partial E(n)}{\partial e_j(n)}\frac{\partial e_j(n)}{\partial y_j(n)}\frac{\partial y_j(n)}{\partial V_j(n)} \tag{2.60}$$

Usando a Equação 2.54, verificamos que $\partial E(n)/\partial Y_j(n)$, pode ser expresso como:

$$\frac{\partial E(n)}{\partial Y_j(n)} = -1 \tag{2.61}$$

O sinal de saída do j neurônio na interação n ($Y_j(n)$), pode ser dado por:

$$Y_j(n) = \omega_j(V_j(n)) \tag{2.62}$$

Usando a Equação 2.62, verificamos que:

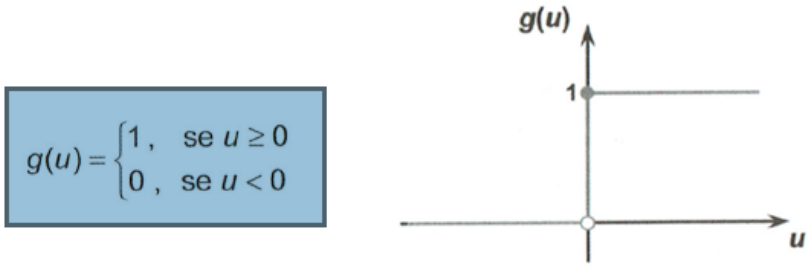
$$\frac{\partial y_j(n)}{\partial V_j(n)} = \omega_j{}'(V_j(n)) \tag{2.63}$$

Mostra-se assim que o valor da derivada parcial, $\omega_j{}'(V_j(n))$ depende do valor da função de transferência ou função de ativação escolhida para cada neurônio.

Segundo Silva, Spatti e Flauzino (2010), o objetivo da função de ativação é limitar a saída de um neurônio, dentro de valores razoáveis a serem assumidos pela própria imagem funcional

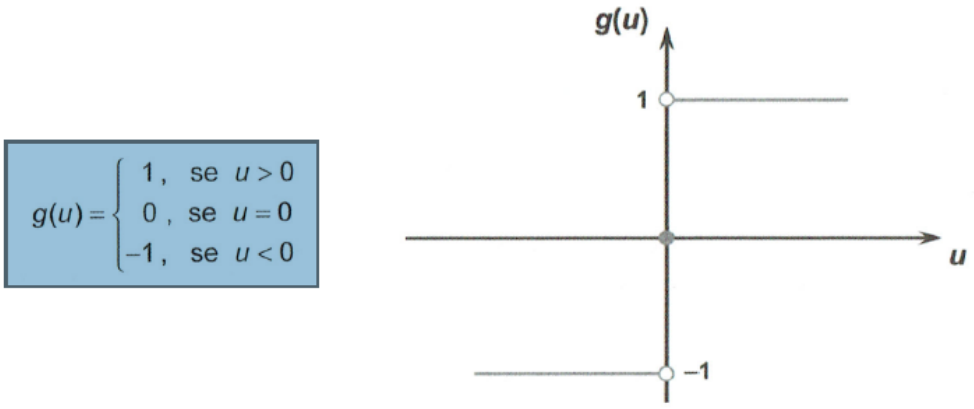
As funções de ativação podem ser parcialmente diferenciáveis ou totalmente diferenciáveis, entra as parcialmente diferenciáveis temos a função degrau (Figura 12), função de ativação degrau bipolar (Figura 13) ou função sinal, função rampa simétrica (Figura 14). As funções totalmente diferenciáveis mais usadas são as funções logísticas (Figura 15), função tangente hiperbólica (Figura 16), função gaussiana (Figura17) e a função de ativação linear (Figura 18).

Figura 12-Função de ativação degrau



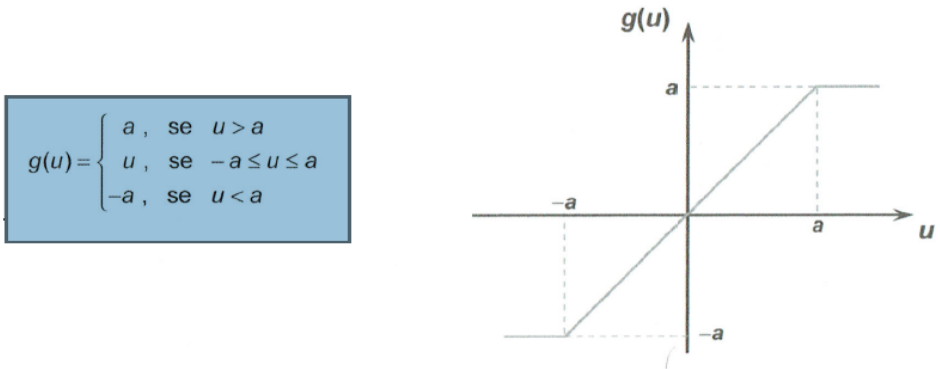
Fonte: Silva, Spatti e Flauzino (2010)

Figura 13-Função de ativação degrau bipolar



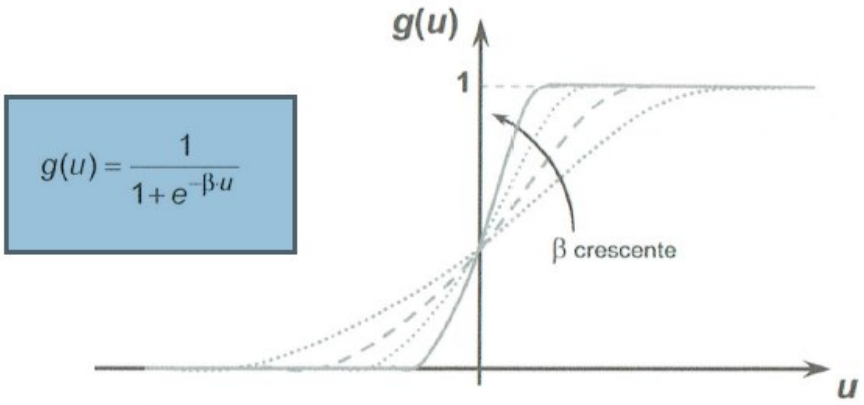
Fonte: Silva, Spatti e Flauzino (2010)

Figura 14-Função de ativação rampa simétrica



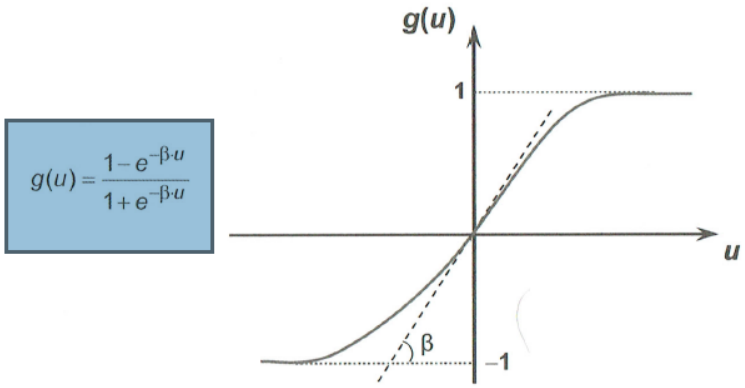
Fonte: Silva, Spatti e Flauzino (2010)

Figura 15-Função de ativação Logística ou sigmóide



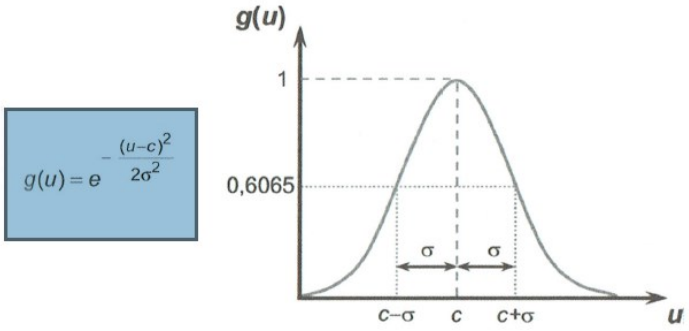
Fonte: Silva, Spatti e Flauzino (2010)

Figura 16-Função de ativação Hiperbólica



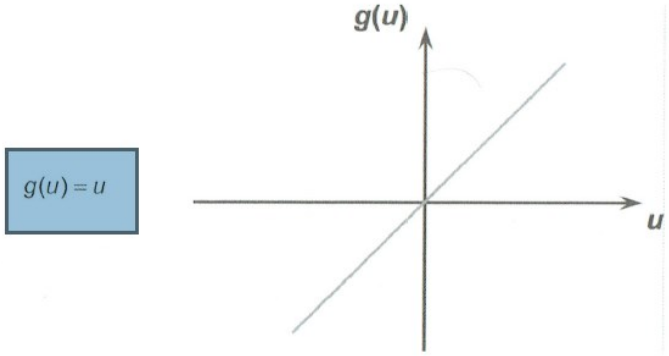
Fonte: Silva, Spatti e Flauzino (2010)

Figura 17-Função de ativação Gaussiana



Fonte: Silva, Spatti e Flauzino (2010)

Figura 18-Função de ativação Linear



Fonte: Silva, Spatti e Flauzino (2010)

Se por exemplo a função for sigmoide ao qual a função de transferência correspondente é dada pela Equação 2.63:

Sua derivada parcial seria conforme Equação 2.64:

$$\omega_j{}'\left(V_j(n)\right) = \frac{e^{-V_j(n)}}{(1+e^{-V_j(n)})^2} \tag{2.64}$$

A função do gradiente local do neurônio j, δ_j (n) pode ser reescrita como:

$$\delta_j(n) = \frac{\partial E(n)}{\partial V_j(n)} = -e_j(n)\omega_j{}'(V_j(n)) \tag{2.65}$$

Desenvolvendo-se $\partial V_j(n)/\,\partial W_{ij}(n)$, em que:

$$V_j(n) = \sum_i W_{ij} * \mathrm{Y}_i(n) \tag{2.66}$$

ou: $\frac{\partial V_j(n)}{\partial W_{ij}(n)} = \mathrm{Y}_i(n)$ (2.67)

Assim, a Equação 2.59, fica reescrita da seguinte forma:

$$\frac{\partial E(n)}{\partial W_{ij}(n)} = -e_j(n)\omega_j{}'(V_j(n))\mathrm{Y}_i(n) \quad (2.68)$$

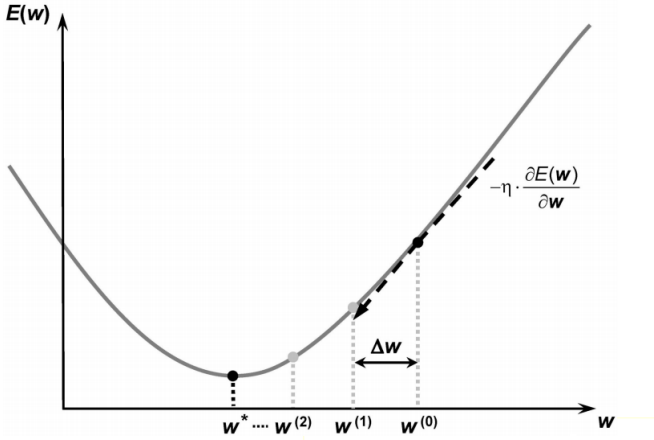
Ou ainda pela Equação 2.65, temos:

$$\frac{\partial E(n)}{\partial W_{ij}(n)} = \delta_j(n)\mathrm{Y}_i(n) \quad (2.69)$$

Segundo Ludwig Jr e Montgomery (2007), com isto se faz as correções de todos os pesos sinápticos e níveis de bias dos neurônios das últimas camadas.

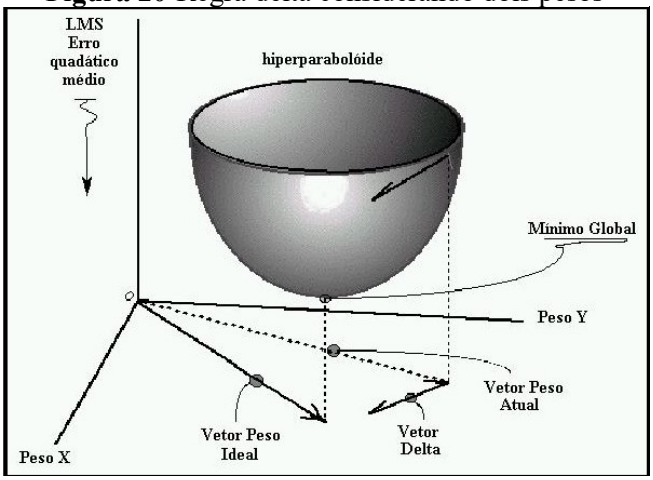
As correções se fazem, através de um vetor delta, tangente a curva do erro quadrático médio e os respectivos pesos e no sentido do erro mínimo quadrático, ou contrário ao gradiente, conforme ilustrado na Figura 19 e 20.

Figura 19- Interpretação geométrica da regra delta



Fonte: Silva, Spatti e Flauzino (2010)

Figura 20-Regra delta considerando dois pesos



Fonte: ICMC (2018)

Aplicand**o**-se a Equação 2.58 e 2.68, temos:

$$\Delta W_{ij}(n) = n\, e_j(n)\omega_j{}'(V_j(n))\mathrm{Y}_i(n) \tag{2.70}$$

Ou ainda pela Equação 2.69:

$$\Delta W_{ij}(n) = -n\, \delta_j(n)\mathrm{Y}_i(n) \tag{2.71}$$

Este processo de correção dos pesos se inicia na última camada e se propaga para as camadas imediatamente anterior, de forma que todos os pesos são reajustados no sentido de minimização do erro.

Valores infinitesimais dos sinais obtidos após correção dos pesos são calculados.

$$\partial V_i(n) = \frac{\partial V_i(n)}{\partial W_{ki}(n)} \partial W_{ki}(n) \tag{2.72}$$

E consequentemente o valor de $\mathrm{Y}_i(n)$, seria também corrigido:

$$\partial Y_i(n) = \frac{\partial Y_i(n)}{\partial V_i(n)}\, \partial V_i(n) \tag{2.73}$$

Considerando a Equação 2.72 e 2.73, temos:

$$\frac{\partial Y_i(n)}{\partial W_{ki}(n)} = \frac{\partial Y_i(n)}{\partial V_i(n)} \frac{\partial V_i(n)}{\partial W_{ki}(n)} \tag{2.74}$$

ou:

$$\frac{\partial Y_i(n)}{\partial W_{ki}(n)} = \omega_j{}'(V_j(n))\mathrm{Y}_k(n) \tag{2.75}$$

Assim, temos que:

$$\partial Y_i(n) = \omega_j{}'(V_j(n))\mathrm{Y}_k(n)\, \partial W_{ki}(n) \tag{2.76}$$

O valor de $\partial Y_i(n)$ é o acréscimo sofrido pelo neurônio i, após receber o acréscimo infinitesimal $\partial W_{ki}(n)$ ao seu peso $W_{ki}(n)$, distribuindo este sinal para todos os neurônios a ele ligado das camadas anteriores. Assim cada neurônio da camada J,

$$\frac{\partial V_j(n)}{\partial Y_i(n)} = W_{ij}(n) \tag{2.77}$$

Assim usando a Equação 2.72 e 2.75, temos que:

$$\partial V_j(n) = \omega_j{}'(V_j(n))\mathrm{Y}_k(n)\, \partial W_{ki}(n) W_{ij}(n) \tag{2.78}$$

Pode-se então corrigir os sinais de saída, que vai gerar um novo erro E(n):

$$\partial E_j(n) = \delta_j(n)\partial V_j(n) \tag{2.79}$$

Mas é necessário somar todas as parcelas oriundas dos neurônios da camada j, assim:

$$\partial E(n) = \sum_{j=1}^{N} E_j(n) \tag{2.80}$$

Ou ainda, usando a Equação 2.79, temos:

$$\partial E(n) = \sum_{j=1}^{N} \delta_j(n) \partial V_j(n) \quad (2.81)$$

Da Equação 2.78, temos:

$$\partial E(n) = \sum_{j=1}^{N} \delta_j(n) \omega_j'(V_j(n)) \mathrm{Y}_k(n)\, \partial W_{ki}(n) W_{ij}(n) \quad (2.82)$$

Ou:

$$\frac{\partial E(n)}{\partial W_{ki}(n)} = \omega_j' \left(V_j(n)\right) \mathrm{Y}_k(n) \sum_{j=1}^{N} \delta_j(n) W_{ij}(n) \quad (2.83)$$

Ou ainda:

$$E(n) = \omega_j'(V_j(n)) \mathrm{Y}_k(n)\, \partial W_{ki}(n) \sum_{j=1}^{N} \delta_j(n) W_{ij}(n) \quad (2.84)$$

Usando a regra delta, ou Equação 2.58 e 2.83, temos a Equação que faz as correções de peso da rede neural:

$$\Delta W_{ij}(n) = -n\omega_j' \left(V_j(n)\right) \mathrm{Y}_k(n) \sum_{j=1}^{N} \delta_j(n) W_{ij}(n) \quad (2.85)$$

2.4.8 Vetor suporte de Maquina (SVM)

Máquinas de Vetores de Suporte (SVMs), que foram introduzidas por Vapnik (1998). no início dos anos 90, provaram ser uma técnica eficaz e promissora para a separação de dados.

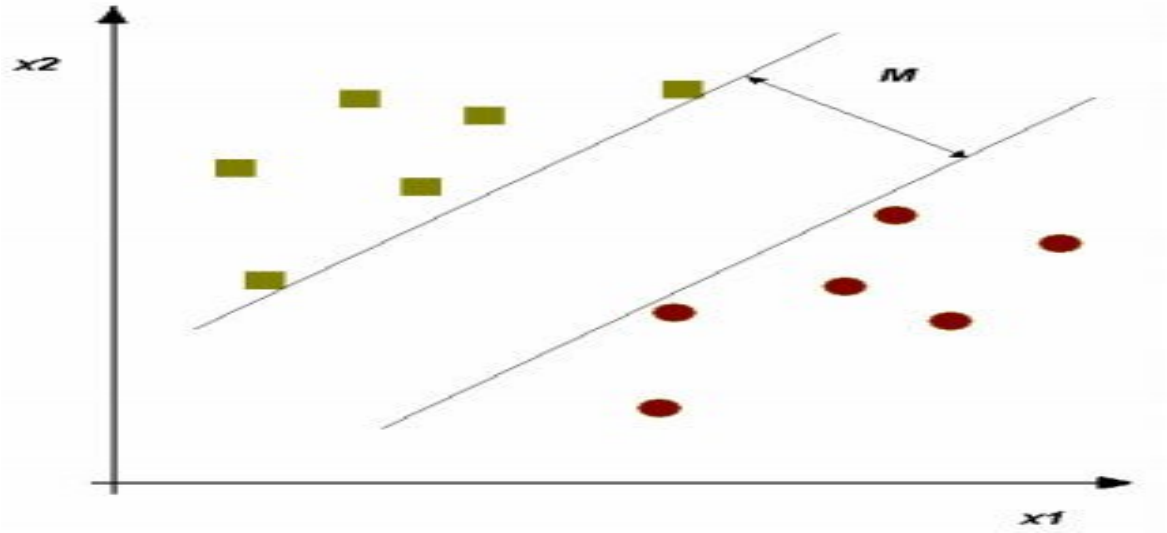
SVMs têm a capacidade de resolver problemas de classificação e regressão, adquirindo com o aprendizado na etapa de treinamento, a capacidade de generalização. Considerando um problema binário, o objetivo da SVM será de separar as instâncias das duas classes através de uma função que será obtida a partir dos exemplos conhecidos na fase de treinamento. O objetivo será produzir um classificador que funcione de forma adequada com exemplos não conhecidos, adquirindo assim a capacidade de predizer as saídas de futuras novas entradas (OLIVEIRA-JUNIOR (2010)),

Os SVMs recentemente fizeram muitos avanços em suas teorias estudos e implementações de algoritmos. Eles foram aplicados com sucesso em muitos campos, tais como categorização de texto, reconhecimento de fala, controle remoto, detecção de análise de imagem, previsão de séries temporais, segurança da informação e outros (DENG, TIAN e ZHANG, 2013)

Segundo Kinto (2011), a formulação do SVM resulta em um problema de otimização quadrática e envolve o princípio de minimização do risco estrutural (*structural risk minimization*), em vez do princípio de minimização do risco empírico (ERM, *empirical risk minimization*), muito empregado pelas redes neurais artificiais.

A maximização da margem M de separação entre as superfícies positivas e negativas está ilustrada na Figura 21.

Figura 21- Margem de separação entre hiperplanos nas fronteiras das classes positiva (quadrado preenchido) e negativa (círculo preenchido). No SVM queremos que a margem M seja maximizada. Os eixos ordenados x1 e x2 representam as dimensões das amostras no espaço 2



Fonte: Kinto, 2011

Dado que (no espaço 2D):

- O semi-plano positivo é separado pela reta: $\boldsymbol{xp}: \boldsymbol{w^t}.\boldsymbol{xp} + b = +1$ (2.86)
- O semi-plano negativo é separado pela reta: $\boldsymbol{xn}: \boldsymbol{w^t}.\boldsymbol{xn} + b = +1$ (2.87)

$$\boldsymbol{xp} = \boldsymbol{xn} + \boldsymbol{\lambda w} \quad (2.88)$$

Os parâmetros **w** e **b** são, respectivamente, o peso e o bias. Semelhante as redes neurais, eles serão ajustados durante o processo de treinamento do SVM, sendo que xp representa as amostras da classe positiva e **xn** representa as amostras da classe negativa.

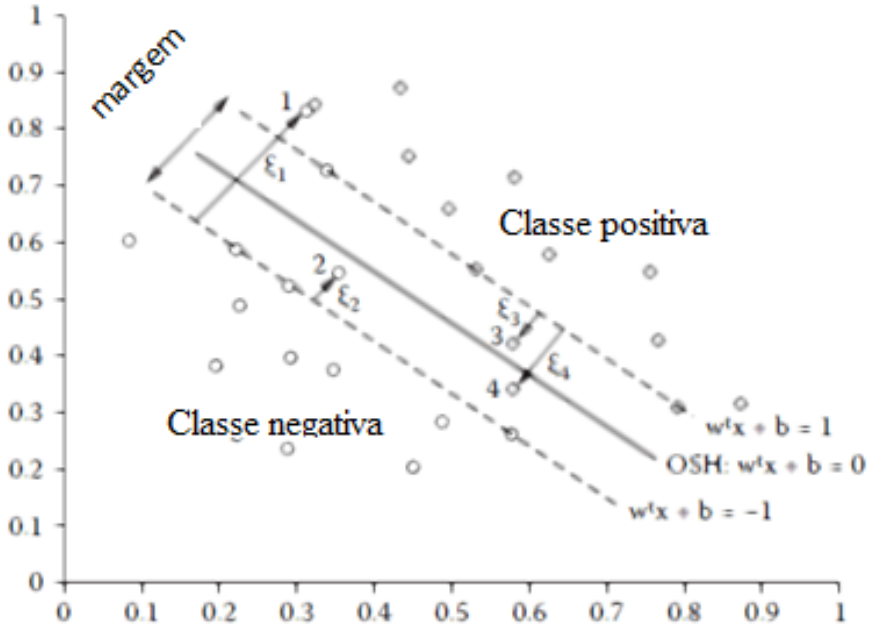
Pode-se calcular a margem M em função de w e b, como a seguir:

$$M = \boldsymbol{xp} - \boldsymbol{xn} \quad (2.89)$$
$$M = |\boldsymbol{\lambda w}| \quad (2.90)$$
$$M = \boldsymbol{\lambda}\sqrt{\boldsymbol{w^t}.\boldsymbol{w}} \quad (2.91)$$

Segundo Liang *et al. (*2011), na maioria dos casos, os dados são linearmente inseparáveis devido à fundo desconhecido (ruidos), ou não-linearidade intrínseca. Conforme mostrado na Figura 22, não é possível construir um hiperplano, que possa separar-se linearmente as duas classes sem erro.

O objetivo é encontrar uma função $f(x)$ que desvia de y_n por um valor não superior a ε para cada ponto de treinamento x e, ao mesmo tempo, é o mais plano possível.

Figura 22 -As variáveis de folga e o hiperplano separador ótimo (OSH) para o caso linearmente inseparável.



Fonte: Adaptado de Liang *et al. (*2011)

A variável de folga (ε_i) foi introduzida por Cortés e Vapnik para construir o hiperpalano separador ótimo (OSH), considerando os inevitáveis erros ou ruidos medidos que comportam os dados. A restrição a desigualdade da OSH pode ser expressa na forma:

$(\boldsymbol{w^t x_i} + b).\, y_i \geq 1 - \varepsilon_i$ sujeito as condições: $\varepsilon_i > 0$ e i=1,2,3,....n (2.92)

Sendo ε_i é a variável de folga, que é uma medida de quão longe a amostra se afasta do hiperplano limite (Figura 22). Em seguida, a construção da OSH pode ser expressa como o problema de otimização (LIANG *et al.,, 2011)*:

$$\text{minimizar} \qquad \frac{1}{2}\| w\| + C \sum_1^N \varepsilon_i \tag{2.93}$$

sujeito a: $(\boldsymbol{w}^t \boldsymbol{x_i} + b).y_i \geq 1 - \varepsilon_i$ e $\varepsilon_i \geq 0$

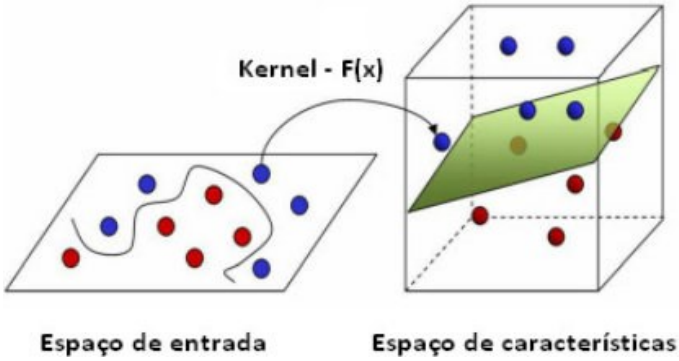
Sendo C um fator penalizador que controla a troca entre o erro de treinamento e a margem.

Existirão situações nas quais um classificador linear como apresentado não encontrará alguma maneira de separar as amostras. Nesse caso é necessário usar um classificador com função de separação não linear (LOPES, 2016)

À medida que se aumenta o espaço da dimensão do problema, aumenta também a probabilidade desse problema se tornar linearmente separável em relação a um espaço de baixa dimensão. As funções de Kernel têm a finalidade de projetar os vetores de características de entrada em um espaço de características de alta dimensão para classificação de problemas que se encontram em espaços não linearmente separáveis. Entretanto, para obter uma boa distribuição para esse tipo de problema, é necessário um conjunto de treinamento com um elevado número de instâncias (GONÇALVES, 2010, *apud* LOPES 2016)

A Figura 23, apresenta o processo de transformação de um domínio não linearmente separável, em um problema linearmente separável através do aumento da dimensão, ao qual é feito um mapeamento por uma função de Kernel F(x).

Figura 23 - Transformação: problema não linearmente separável em um problema linearmente separável



Fonte: Adaptado de Rebelo, 2008.

Segundo Gonçalves (2018), uma função Kernel recebe dois dados de entrada xi e xj e calcula o produto interno destes dados no espaço de características

$$K\left(x_i x_j\right) = \langle \emptyset(x_i).\emptyset(x_j) \rangle \tag{2.94}$$

No qual $\emptyset$ (.) deve pertencer a um domínio ao qual seja possível o cálculo do produto interno. Essas funções também satisfazem as condições do Teorema de Mercer.

Teorema de Mercer: uma determinada função é definida como sendo de Kernel, se a matriz K é positivamente definida (autovalores maiores que zero), sendo K é obtida por:

$$K = K_{ij} = K(x_i, x_j) \tag{2.95}$$

Segundo Kinto (2015) existem três funções Kernels mais usadas, a função linear dada pela Equação 2.96, usadas para classes linearmente separáveis, e as outras duas, a Kernel polinomial e a Kernel de base gaussiana radial (RBF), expressas pelas equações 2.97 e 2.98, respectivamente.

$$K(\boldsymbol{x}.\boldsymbol{x}') = (\boldsymbol{x}^{\boldsymbol{t}}.\boldsymbol{x}) \tag{2.96}$$

$$K(\boldsymbol{x}.\boldsymbol{x}') = \left(1 + (\boldsymbol{x}^{\boldsymbol{t}}.\boldsymbol{x})\right)^{p} \tag{2.97}$$

Sendo: p grau da função polinomial

$$K(\boldsymbol{x}.\boldsymbol{x}') = \exp(\frac{-1}{2\gamma}\|\boldsymbol{x} - \boldsymbol{x}'\|^2) \tag{2.98}$$

Sendo: γ (gama) a largura da função base radial gaussiana.

Segundo Lopes (2016), para construir um hiperplano ótimo, o SVM emprega um algoritmo iterativo de formação, que é utilizado para minimizar uma função de erro. De acordo com a forma da função de erro, os modelos SVM podem ser classificados em 4 grupos distintos, sendo dois de classificação e dois de regressão:

- Classificação SVM Tipo 1

Também conhecida como Classificação C-SVM. Para este tipo de SVM, o treinamento envolve a minimização da função de erro:

$$\frac{1}{2}w^t.w + C\sum_1^N \varepsilon_i \tag{2.99}$$

sujeita às restrições:

$(\boldsymbol{w}^t\emptyset(\boldsymbol{x_i}) + b).y_i \geq 1 - \varepsilon_i \quad e \;\; \varepsilon_i \geq 0 \quad$ i=1,2,...,N

Sendo C a constante de capacidade, w o vetor de coeficientes, b uma constante e representa parâmetros para tratamento de dados não-separáveis (dados de entrada). O índice i classifica os N casos de treinamento. Nota-se que $y_i \in \mp 1$, representa os rótulos de classe e x_i representa as variáveis dependentes. A função de Kernel $\emptyset(\boldsymbol{x_i})$ é usada para transformar os dados de entrada

(independentes) para o espaço de características. Deve se notar que quanto maior for o *C*, mais a função erro é penalizada. Assim, *C* deve ser escolhido com cuidado para evitar uma extrapolação de dados.

Classificação SVM Tipo 2

Também conhecida como Classificação nu-SVM. A modelo de Classificação SVM Tipo 2, minimiza a função de erro

$$\frac{1}{2}w^t.w - \upsilon\rho + \frac{1}{N}\sum_1^N \varepsilon_i \qquad (2.100)$$

sujeita às restrições:

$$(\boldsymbol{w}^t\varphi(\boldsymbol{x_i}) + b).y_i \geq \rho - \varepsilon_i \quad e \;\; \varepsilon_i \geq 0 \;\; i = 1,2,\dots,N \quad e \;\; \rho \geq 0$$

Regressão SVM Tipo 1

Na Regressão SVM, tem-se que estimar a dependência funcional da variável dependente *y* relacionada a um conjunto de independentes variáveis *x*. Logo, assume-se que a relação entre as variáveis dependentes e independentes é dada por uma função determinística função *f adicionada* de algum ruído:

$$y = f(x) + ruído \qquad (2.101)$$

A tarefa é então procurar uma forma funcional para *f* que pode corretamente predizer novos casos nos quais o SVM não tenha sido apresentado antes. Isso pode ser alcançado treinando o modelo SVM em um conjunto de amostras, isto é, treinando o conjunto, um processo que envolve, como na classificação, a otimização sequencial de uma função de erro. Dependendo da definição dessa função de erro, tem-se a Regressão SVM Tipo 1, também conhecida como Épsilon-SVM Regressão, que tem sua função de erro dada por:

$$\frac{1}{2}w^t.w + C\sum_1^N \varepsilon_i + \sum_1^N \varepsilon_i^* \qquad (2.102)$$

A função erro é minimizada, sujeita às seguintes restrições:

$$\boldsymbol{w}^t\emptyset(\boldsymbol{x_i}) + b - y_i \leq \; \varepsilon + \; \varepsilon_i^*$$

$$y_i - \boldsymbol{w}^t\emptyset(\boldsymbol{x_i}) - b_i \leq \; \varepsilon + \; \varepsilon_i$$

$$\varepsilon_i, \varepsilon_i^* \geq 0\,, \qquad i = 1,2,\dots,N$$

ε_i é uma variável de folga ou ruido

Regressão SVM Tipo 2

Também conhecida como Regressão nu-SVM. Para esse modelo de SVM, a função erro é dada por:

$$\frac{1}{2}\boldsymbol{w}^t.\boldsymbol{w} - C(\upsilon\varepsilon + \frac{1}{N}\sum_1^N(\varepsilon_i + \varepsilon_i^*)) \quad (2.103)$$

A função erro é minimizada, sujeita às seguintes restrições:

$$(\boldsymbol{w}^t\emptyset(\boldsymbol{x_i}) + b) - y_i \leq \varepsilon + \varepsilon_i$$

$$y_i - (\boldsymbol{w}^t\emptyset(\boldsymbol{x_i}) + b_i) \leq \varepsilon + \varepsilon_i^*$$

$$\varepsilon_i, \varepsilon_i^* \geq 0\,, \qquad i = 1,2,\dots,N\,, \quad \varepsilon \geq 0$$

2.4.9 Rede Neural Convolutiva (CNN)

Segundo Bezerra (2018), aprendizagem Profunda é uma subárea de Aprendizagem de Máquina que investiga técnicas para simular o comportamento do cérebro humano em tarefas como reconhecimento visual, reconhecimento de fala e processamento de linguagem natural. Algoritmos de aprendizagem profunda, objetivam produzir representações hierárquicas de alto nível dos dados de entrada, por meio de camadas de processamento sequencial, em uma rede neural artificial.

Segundo Rocha e Melidiú (2015), diversos fatores contribuem para tornar esta tarefa complexa, como a oclusão parcial do objeto, diferenças de iluminação, variações em cenas de fundo, além das diferentes configurações que objetos do mesmo tipo podem apresentar. Para associar a relação entre os valores dos pixels e o conteúdo da imagem geralmente há uma primeira etapa de pré-processamento, que inclui obter uma descrição do conteúdo da imagem mais apropriada que os pixels. Para isto, aplicam diferentes extratores de características, a depender da tarefa em questão.

É possível transformar números em imagens, imagens estas, que o nosso cérebro não é capaz de identificar, porém podemos associar a imagem a uma classificação feita e reconhecida no computador.

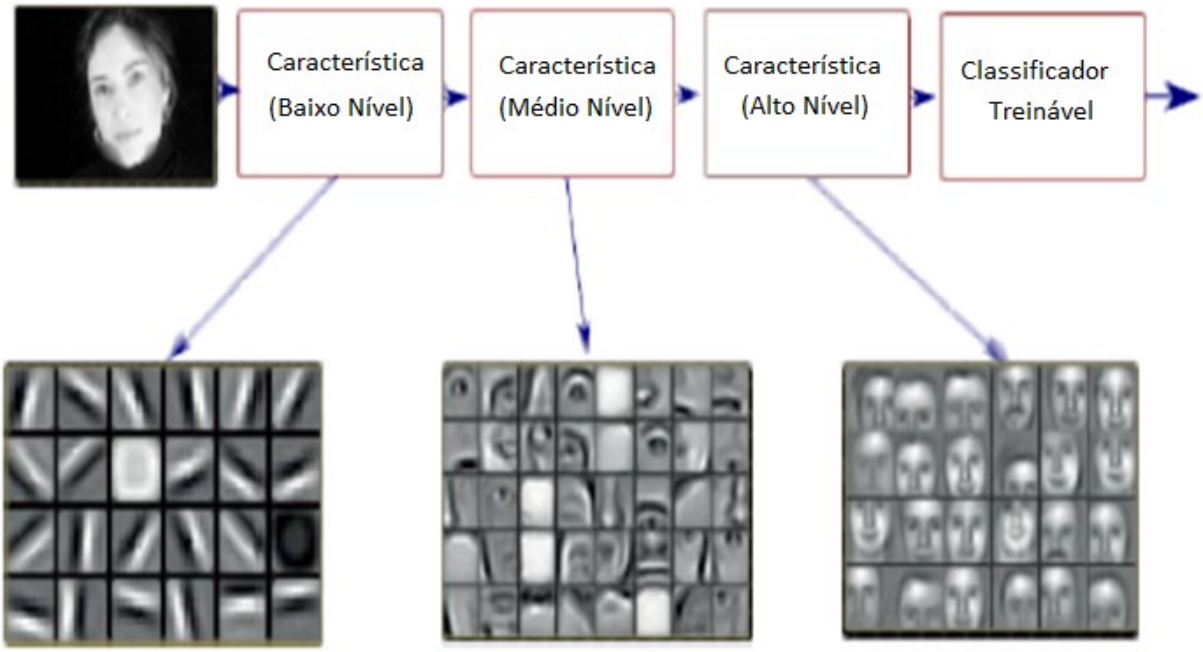
As redes neurais convolutivas são redes do tipo *Feedforward* na qual cada camada aprende representações hierárquicas em vários níveis de abstração (VILLALOBOS 2016)

. Cada neurônio da saída de uma camada de convolução Z_i^K está associado a uma região da imagem de entrada Xi e a um conjunto de parâmetros, chamados Kernels e cada camada de convolução possui um Kernel.

O valor no neurônio de saída é uma função do produto interno entre um dos Kernels W^k e essa região da entrada Xi , ou seja $Z_i^K = f(W^k \times Xi)$. Chamamos o conjunto de todas as saídas Z^k associadas a um Kernel W^k de *Feature Map* k. Para uma entrada X de dimensão w × h e um Kernel W de dimensão s × s, o *Feature Map* do kernel computa w − s + 1 × h − s + 1 valores de saída.

A função de disparo é computada com o mesmo parâmetro W em diferentes regiões da imagem. A camada de convolução funciona como uma extratora de características, especializada em imagens (ROCHA,2015)

Figura 24 -*Feature* (características) ou classificador (Kernel) treináveis.



Fonte: Villalobos (2016) adaptada

Na Figura 24, se mostra uma estrutura hierárquica de 3 níveis de representação. A entrada, o primeiro nível de representação são os pixels da imagem (a foto mesma), no primeiro

feature transform (Baixo Nível), são detectados tipicamente a presença ou ausência de bordas com também orientação especifica e locação da imagem.

O segundo *feature transform* (Médio Nível) extrai as partes de um objeto e o terceiro *feature transform* (Alto Nível) detecta objeto ou combinação das partes.

Os *feature maps* de Alto Nível são mais globais e invariantes. Os *Feature maps* de Baixo Nível são compartilhados entre os diferentes *Feature maps* das camadas posteriores (VILLALOBOS, 2016).

2.4.10 Necessidade de pré-tratamento dos dados.

2.4.10.1 *Normalização dos dados*

Em uma estação de tratamento de efluentes são medidos dados importantíssimos para o controle da Estação, como pH, temperatura, concentrações diversas de Demanda Bioquímica e Química de oxigênio e vazão.

Cada uma dessas variáveis tem suas unidades de medida diferentes e para se observar a influência de uma variável sobre as demais, especialmente se tratando de sistemas multivariável, técnicas de normalização dos dados são utilizados.

Campana (2010) propôs um modelo de transformação de dados para uso em Análise de Componentes Principais, que consiste em se transformar a variável em uma multiplicação do desvio reduzido pelo seu coeficiente de variação:

$$Z_{ij}^{*} = Z_{ij} * CV_j \quad (2.104)$$

para i=1, 2,..., n observações e j=1, 2,..., p variáveis, sendo:

Z_{ij}^{*} : o valor da *i*-ésima observação da variável *Xj* transformada;

$Z_{ij} = \frac{X_{ij}-\mu_j}{\sigma_j}$ o valor da $i-$ ésima observação da variável Xj padronizada; (2.105)

X_{ij} : o valor da i-ésima observação da variável Xj;

μ_j: a média da variável *Xj*; σ_j: é o desvio-padrão da variável *X*

$CV_j = \frac{\sigma_j}{\mu_j}$; é o coeficiente de variação da variável *Xj*; (2.106)

Substituindo (2.40.1) e (2.4.10.3) em (2.4.10.2), temos:

$$Z_{ij}^{*} = \frac{X_{ij}-\mu_j}{\mu_j} \quad (2.107)$$

Assim, as variáveis transformadas passam a estar numa mesma escala de medida, com médias iguais a zero, porém com variâncias diferentes de um e proporcionais aos seus respectivos CVs.

Este modelo pode ser usado também para unidimensionalizar dados de séries temporais.

Outro modelo é o uso da diferença entre o máximo e mínimo da série

$$Z_{ij} = \frac{X_{ij} - minimo}{máximo - minimo} \qquad (2.108)$$

Segundo Rocha (2015), a normalização dos dados se faz necessária para trabalhos com redes neurais convolutivas, porque o algoritmo que ajusta os parâmetros do modelo usa os valores dos dados de entrada no cálculo do ajuste, se os dados de entrada possuem valores muito altos ou muito baixos o modelo não consegue generalizar de modo satisfatório.

2.4.10.2 *Preenchimento de dados faltantes.*

Em função de quebra de equipamento de análises, feriados e problemas analíticos diversos, é normal acontecer de dias sem ser computados alguns resultados, isto cria uma situação de descontinuidade em algumas séries temporais.

Segundo Santos (2012), uma falha na série temporal pode comprometer a análise e a interpretação de dados.

Modelos de séries temporais unidimensionais são usados no preenchimento de dados faltantes em variáveis continuas (caso de séries temporais).

FERNANDEZ (2007), criou falhas em séries temporais que estavam completas e aplicou métodos estatísticos de análise de regressão múltipla, média simples, média de três estações, e outros métodos para prever dados faltantes na série e verificar qual método teve melhor resultado.

2.4.10.3 *Suavização dos dados para redução do ruído*

É conveniente se obter métodos que permitam um bom tratamento de remoção (ou atenuação) máxima do ruído indesejável, sem causar descaracterização da informação de interesse.

O objetivo da suavização é produzir mudanças lentas de valor em nossos dados, de modo que seja mais fácil ver tendências, sem muita interferência de ruídos da série original, diversas técnicas de suavização são utilizadas para se obter uma curva suavizada, entre elas métodos de média móvel, interpolação polinomial e transformadas *wavelets.*

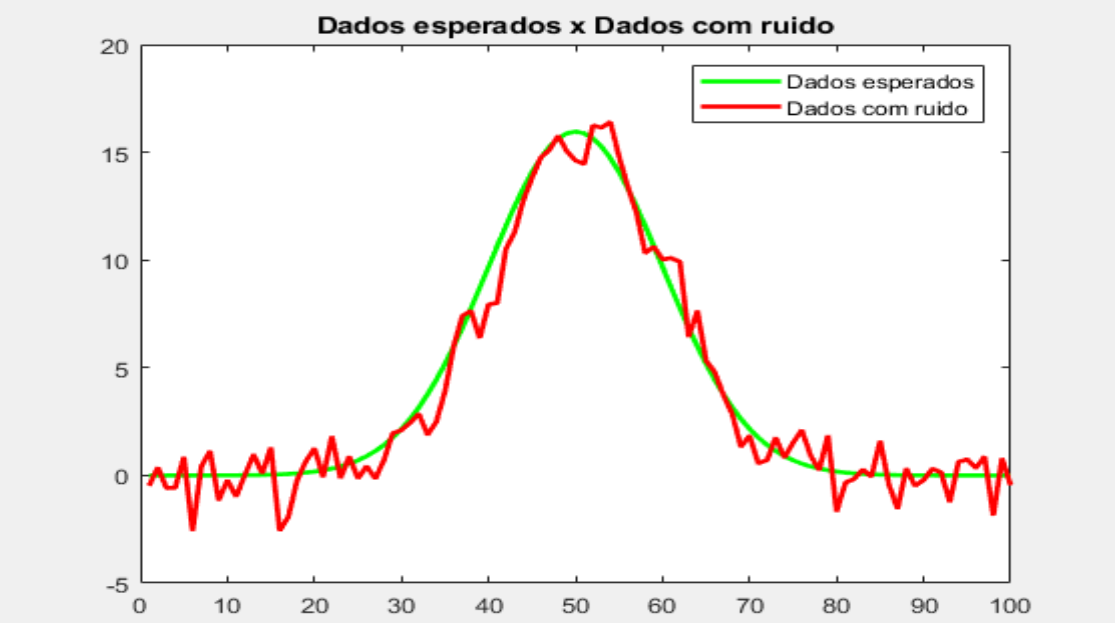
Segundo Ghosh e Raychaudhuri (2017), seja qual for a fonte de dados de séries temporais, é esperado que alguma quantidade de ruído sejam incorporadas. A análise de tais dados em presença de ruído geralmente pode deixar de fornecer informações precisas. O método de filtragem de dados de séries temporais é uma ferramenta para limpar esses erros, pois podemos com isto tornar os dados compatíveis para análises futuras.

É necessário o amplo desenvolvimento de metodologias e ferramentas que possibilitem processar sinais de maneiras particulares ao invés de somente analisar os sistemas existentes, como acrescenta Oppenheim e Willsky (2010), “um contexto muito comum em que esses problemas surgem é no projeto de sistemas para melhorar ou restaurar sinais que foram degradados de alguma maneira”.

. Para ilustrar o objetivo de uma suavização vamos supor que se espere de um dado qualquer um comportamento conforme a linha verde na Figura 25, porém em função de variações nas análises e erros diversos (ruidos), os dados se mostrem conforme o da linha vermelha na mesma Figura. Para se ter uma aproximação da curva sem ruido (suavizada), se usou umas das técnicas de suavização que se consiste em interpolar polinômios do segundo grau a cada 11 pontos em sequência aos dados obtidos com ruido, ficando a nova curva cosntituida do 1 ao 2 ponto com a interpolação dos 11 primeiros obtidos e a ligação do 2 ao terceiro ponto com a interpolação dos pontos de numero 2 ao numero 12, e assim sucessivamente cobrindo todos os intervalos e pontos análizados. Na Figura 26, temos a interpolação de 11 pontos em polinomio de 1 grau, na Figura 27, temos a interpolação também de 11 pontos na sequencia com um polinômio do segundo grau (observa-se que houve uma aproximação melhor dos dados obtidos em relação ao polinômio do primeiro grau. Na Figura 28, temos um polinomio do segundo grau

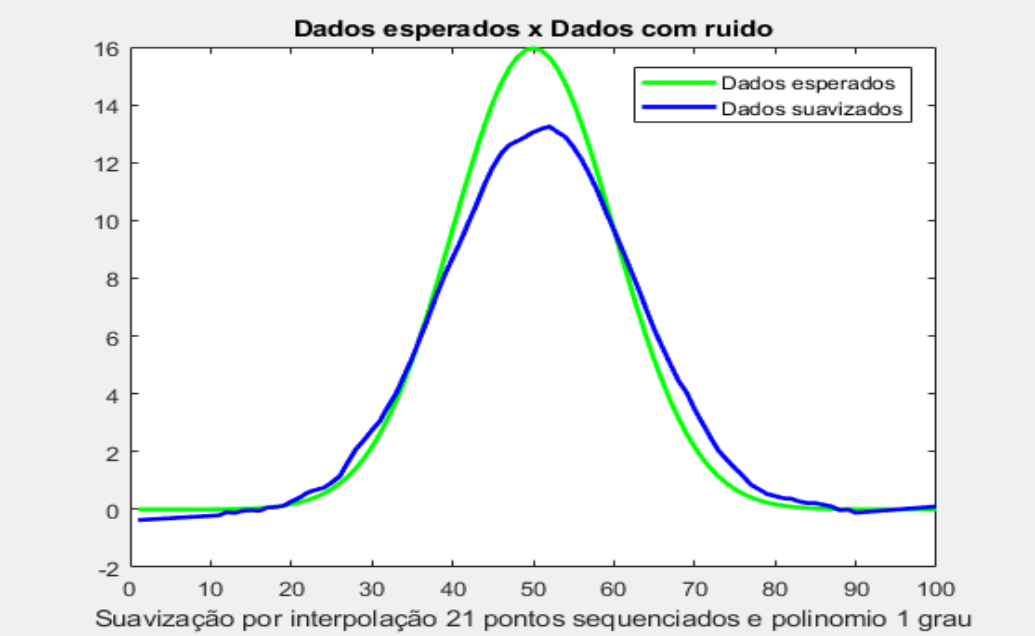
so que com 21 dias de sequencia. A escolha destas combinações podem gerar aumento ou diminuição dos erros em relação ao resultado esperado.

Figura 25- Dados esperados (linha verde) *versus* Dados com ruído (linha vermelha)



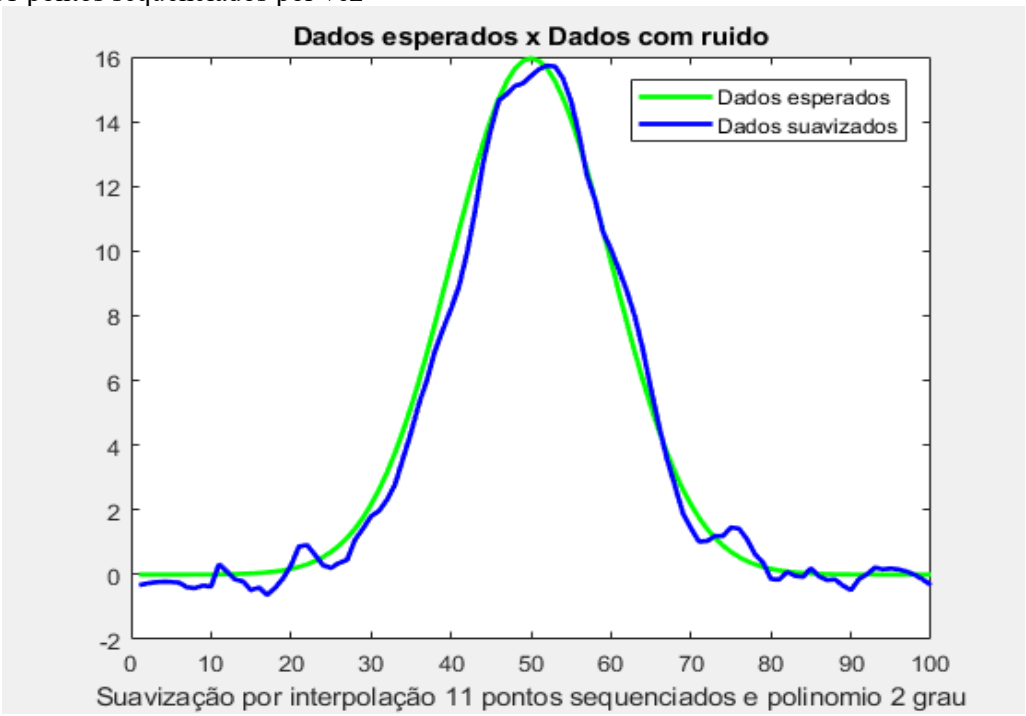
Fonte: O autor

Figura 26-Curva suavizada em azul obtida pelos dados com ruído da Figura 25 versus curva desejada em verde feito por Interpolação polinomial de primeiro grau de 11 pontos sequenciados por vez



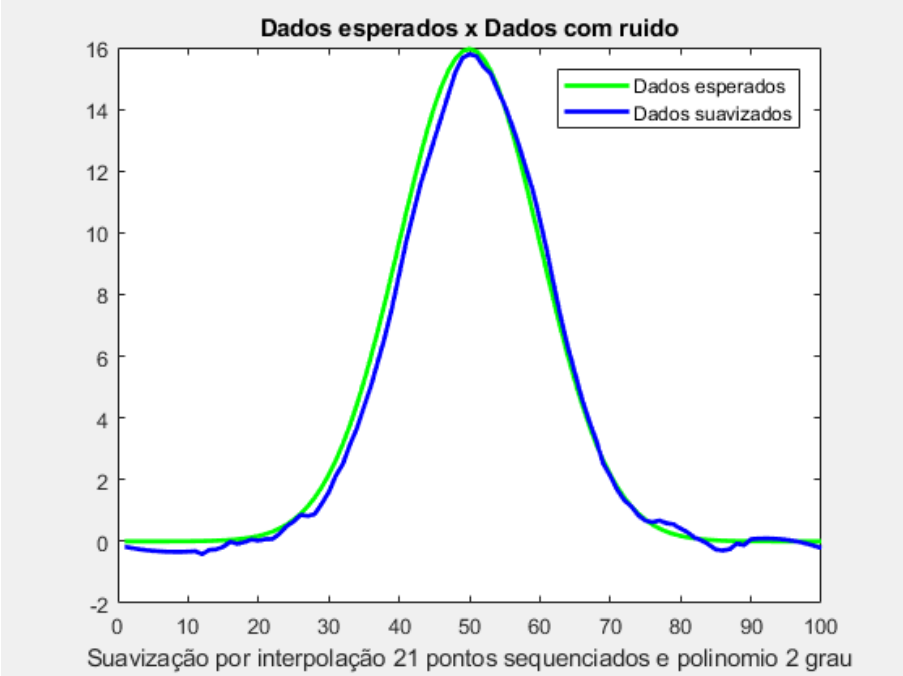
Fonte: O autor

Figura 27 - Curva suavizada em azul obtida pelos dados com ruído da Figura 25 *versus* curva desejada em verde, feito por Interpolação polinomia de segundo grau de 11 pontos sequenciados por vez



Fonte: O autor

Figura 28 - Curva suavizada em azul *versus* curva desejada em verde feito por Interpolação polinomial de segundo grau de 21 pontos sequenciados por vez



Fonte: O autor

2.4.10.4 Transformada de Fourier clássica.

Segundo Morettin e Toloi (2006), a análise de Fourier é uma das formas mais tradicionais para tratamento de sinais e séries temporais. Esta técnica foi criada por Jean Baptiste Joseph Fourier e publicada em 1822 no seu trabalho intitulado *Theorie Analitique de la Chaleur*. Fourier dedicou-se na resolução das equações diferenciais que regem a transferência de calor utilizando uma técnica de séries de senos e cossenos (séries de Fourier), para resolver seus problemas.

$$f(x) = \frac{A_0}{2} + \sum_{n=1}^{\infty} a_n \cos\ (n.x) + \sum_{n=1}^{\infty} b_n \operatorname{sen}\ (n.x\ ) \qquad (2.109)$$

Os coeficientes A_0 , a_n e b_n são relacionadas com funções de f (x) definidas por integrais.

$$a_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \cos(nt\ ) dt \qquad (2.110)$$

$$b_n = \frac{1}{\pi} \int_0^{2\pi} f(t) \operatorname{sen}(nt)\ dt \qquad (2.111)$$

Assim usando a análise de Fourier Canônica, uma série temporal pode ser decomposta ortogonalmente em termos de senos e cossenos definidos em diversas bandas de frequências, Morettin e Toloi (2006)

2.4.10.5 A forma complexa da série de Fourier

Segundo Boggess e Narcowich (2009), muitas vezes, é mais conveniente expressar a série de Fourier na sua forma complexa usando as exponenciais complexas, e^{inx}, n ϵ Z, ao qual Z denota os inteiros {. . . - 2, - 1, 0, 1, 2,. . .}

A exponencial complexa tem a seguinte definição:

Para qualquer número real t, o exponencial complexo é

$$\mathrm{e}^{it} = \cos(\mathrm{t}) + \mathrm{i}\sin(\mathrm{t}) \tag{2.112}$$

Sendol: $i = \sqrt{-1}$

Usando a série de Taylor e expandindo os termos cos(t) e sen(t), na Equação 2.112, temos:

$$\mathrm{e}^{it} = (1 - \frac{t^2}{2!} + \frac{t^4}{4!} + \cdots) + \mathrm{i}\,(\mathrm{t} - \frac{t^3}{3!} + \frac{t^5}{5!} + \cdots) \tag{2.113}$$

Para cada t e s ϵ $\mathbb{R}$, temos:

$$\mathrm{e}^{i(t+2\Pi)} = \mathrm{e}^{it} \tag{2.114}$$

$$|\mathrm{e}^{it}| = 1 \tag{2.115}$$

$$\overline{\mathrm{e}^{it}} = \mathrm{e}^{-it} \tag{2.116}$$

$$\mathrm{e}^{it}.\mathrm{e}^{is} = \mathrm{e}^{i(t+s)} \tag{2.117}$$

$$\frac{\mathrm{e}^{it}}{\mathrm{e}^{is}} = \mathrm{e}^{i(t-s)} \tag{2.118}$$

$$\frac{d}{d_t}(\mathrm{e}^{it}) = i\,\mathrm{e}^{it} \tag{2.119}$$

Para dois vetores X e Y, serem considerados ortogonais no espaço tridimensional ($\mathbb{R}^3$), seu produto interno, expresso pela Equação 2.120, deve ser igual a zero (BOGGES e NARCOWICH, 2009).

$$\langle X, Y\rangle = \|X\|\|Y\|.\cos(\theta) \tag{2.120}$$

Sendo θ é o ângulo entre X e Y

Considerando o produto interno no espaço bidimensional $\mathrm{L}^2[-\pi, \pi]$ temos a Equação 2.121:

$$\frac{1}{2\pi}\langle \mathrm{e}^{int}, \mathrm{e}^{imt}\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{int}\overline{\mathrm{e}^{\imath mt}}\,dt \tag{2.121}$$

Da Equação 2.121, temos:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{int}\overline{\mathrm{e}^{imt}}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{int}\mathrm{e}^{-imt}\,dt \tag{2.122}$$

$$\int_{-\pi}^{\pi} e^{int} e^{-imt} dt = \int_{-\pi}^{\pi} e^{(n-m)it} dt \quad (2.123)$$

$$\int_{-\pi}^{\pi} e^{(n-m)it}\ dt = \frac{e^{(n-m)it}}{(n-m)i}\Big|_{-\pi}^{\pi} \quad (2.124)$$

Temos para n≠m, da Equação 2.124:

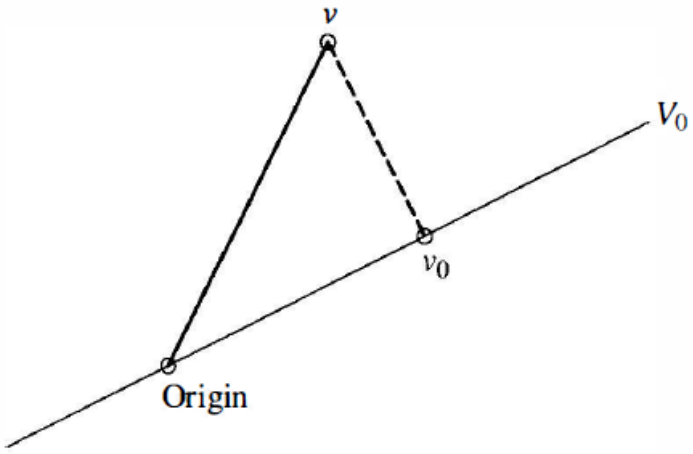
$$\frac{e^{(n-m)it}}{(n-m)i}\Big|_{-\pi}^{\pi} = 0 \quad (2.125)$$

Para n=m da Equação 2.123, temos:

$$\int_{-\pi}^{\pi} e^{(n-m)it} dt = \int_{-\pi}^{\pi} 1 dt = 2\pi \quad (2.126)$$

Supondo que υ seja um vetor projetado em um espaço V_0, N-dimensional com base ortogonal (e_1,e_2,e_3,....e_n), pertencente ao vetor V_0, como ilustrado na Figura 29.

Figura 29-Projeção ortogonal do vetor υ no vetor υ_0



Fonte: Boggess e Narcowich 2009

A projeção ortogonal do vetor $\upsilon \in V$ no vetor V_0, pode ser dada por:

$$V_o = \sum_{j=1}^{N} \alpha_j\, e_j \quad (2.127)$$

Sendo: $\alpha_j = \langle \upsilon, e_j \rangle$ ou seja o produto interno do vetor υ com o vetor e_j .

Se $f(t) = \sum_{n=-\infty}^{\infty} \alpha_n\, e^{int}$ no intervalo $-\pi <= t <= \pi$

$$\alpha_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\, \mathrm{e}^{-int} dt \qquad (2.128)$$

A Equação 2.128 expressa uma versão complexa da série de Fourier (BOGGESS e NARCOWICH, 2009).

2.4.10.6 A transformada de Fourier continua e discreta

A transformada Fourier é usada na transformação de série temporal em frequência (GOPI, 2018), tornando algumas características mais evidentes no domínio transformado como por exemplo o ruído apresentado na série. A Figura 30, mostra um esquema de transformação da variável temporal em frequência.

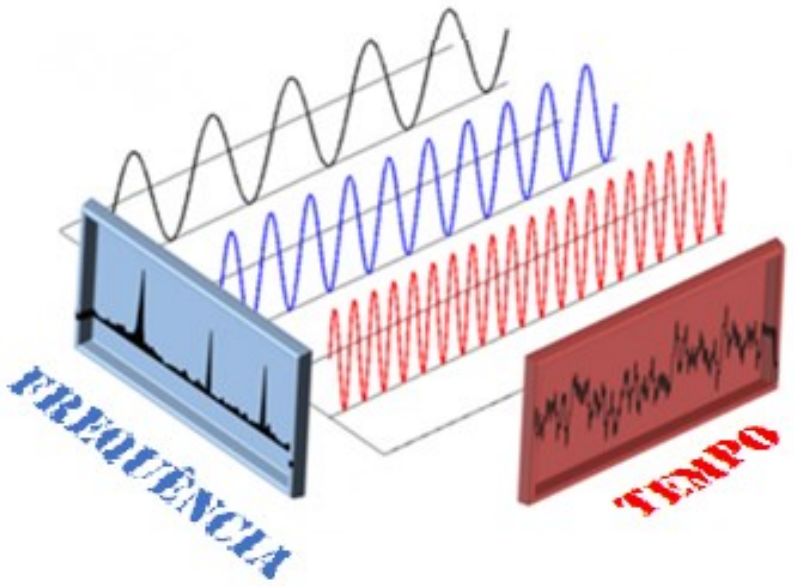
A transformada de Fourier continua é expressa pela Equação 2.129, esta Equação permite transformar o conjunto de pontos x no domínio do tempo, em um conjunto de pontos X, correspondentes no domínio da frequência (SOARES, 2008)

$$\bar{X}(f) = \int_{-\infty}^{\infty} \bar{x}(t)\, \mathrm{e}^{-i2\pi ft} dt \qquad (2.129)$$

Sendo: $i = \sqrt{-1}$

$f =$ frequência do sinal no domínio do tempo t.

Figura 30-Esquema de transformação de tempo em frequência por transformada de Fourier.



Fonte: ENSUS (2017) adaptado

A Transformada inversa de Fourier, que recupera a função original é representada na Equação 2.130.

$$\bar{x}(t) = \int_{-\infty}^{\infty} \bar{X}(f)\, e^{i2\pi f t} df \qquad (2.130)$$

A partir do conjunto de pontos discretos (maioria das análises feitas nas estações de tratamento de efluentes) é possível aplicar o método de aproximação dos coeficientes, conhecida como a transformada discreta de Fourier (DFT), expressa pela Equação 2.131.

$$\bar{X}(f) = \frac{1}{N}\sum_{n=0}^{N-1} \bar{x}(t)\, e^{\frac{-i2n\pi f}{N}} \qquad (2.131)$$

A transformada discreta inversa de Fourier (IDFT), recupera o sinal de domínio da frequência e o transforma novamente em um sinal de domínio do tempo, e é expressa pela Equação 2.132.

$$\bar{x}(t) = \frac{1}{N}\sum_{n=0}^{N-1} \bar{X}(f)\, e^{\frac{i2n\pi f}{N}} \qquad (2.132)$$

Segundo Gurgel (2013), quatro representações distintas de Fourier são bem conhecidas na literatura, cada uma delas é apropriada a uma classe diferente de sinais e são definidas pelas suas propriedades temporais conforme a Tabela 4. Por intermédio da série de Fourier, qualquer função

periódica, por mais complexa que seja, tem sua representação equivalente como uma soma de componentes senoidais de frequências distintas.

Tabela 4 - Propriedade temporal do sinal e sua respectiva representação em análise de Fourier

Propriedade de tempo	Periódica	Não periódica
Contínuo	Série de Fourier (SF)	Transformada de Fourier (TF)
Discreto	Série de Fourier de tempo discreto (SFTD)	Transformada de Fourier de tempo discreto (TFTD)

Fonte: adaptado de Gurgel (2013).

2.4.10.7 A transformada de Fourier janelada.

Segundo Soares (2008), para se permitir a localização temporal da frequência e buscar uma forma de melhorar o desempenho da transformada de Fourier, se usa da divisão da série em segmentos, ao qual a localização temporal das frequências pode ser obtida por janelamentos da série x em instantes t, definidos na Equação 2.133, por uma função $\bar{\omega}$. Este procedimento é definido como a Transformada de Fourier Janelada (STFT- Short Time Fourier Transform), cujo esquema de janelamento é ilustrado na Figura 31.

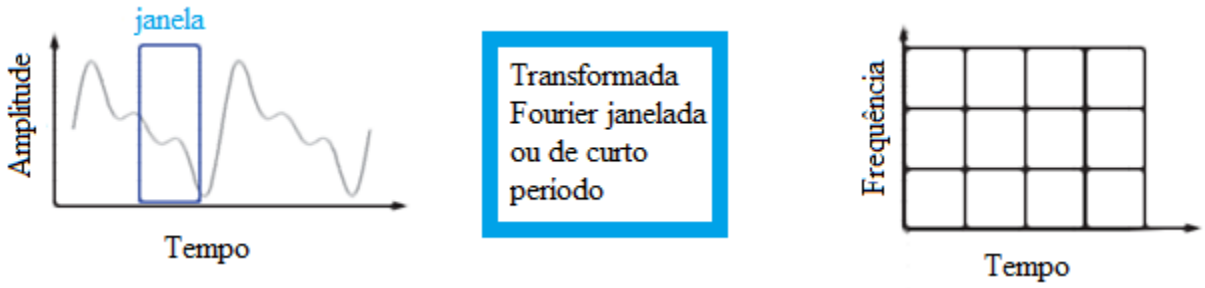
$$F(p,f) = \sum_{t=0}^{N-1} x_t \, \bar{\omega}_{(t-p)} e^{-ift} \tag{2.133}$$

Sendo:

$\bar{\omega}$ uma função de janelamento responsável pela delimitação do trecho considerado do sinal

p= posição da janela dentro do sinal

Figura 31-Esquema das janelas de divisão tempo frequência da transformada de Fourier Janelada



Fonte : Misiti *et al.* (2017) adaptado

A dificuldade de se definir o tamanho da janela a ser usada e o tamanho constante ao longo do sinal faz com que a transformada Wavelet tenha maior vantagem na aplicação que a transformada de Fourier Janelada.

2.4.10.8 A transformada wavelet continua e sua inversa

Uma wavelet é uma forma de onda de duração efetivamente limitada que tem uma média valor de zero. Enquanto as funções seno e coseno da transformada de Fourier são suaves e previsíveis, as wavelets tendem a ser irregulares e assimétricas (MISITI *et al.,* 2017).

A análise de Fourier consiste em dividir um sinal em ondas senodais de várias frequências. Da mesma forma, a análise wavelet é a quebra de um sinal para versões deslocadas e redimensionadas da wavelet original (ou mãe). Na Figura 32, se encontra o esquema da transformada Wavelet.

Figura 32-Esquema das janelas de divisão tempo frequência da transformada *Wavelet*

Fonte : Misiti *et al*. (2017)

A análise *wavelet* representa o próximo passo lógico: uma técnica de janelas com regiões de tamanho variável. A análise *wavelet* permite o uso de intervalos de tempo longos ao

qual queremos informações mais precisas de baixa freqüência e regiões mais curtas, ao qual queremos informações de alta freqüência (MISITI *et al.,* 2017).

A análise de wavelets geralmente pode comprimir ou desativar um sinal, sem degradação apreciável, pode ser usada na remoção de ruidos em Figuras e em dados temporais, como é o caso de estações de tratametno de efluentes.

Segundo Malat (2009), transformada wavelet é construida a partir de uma função wavelet mãe ψ(t), que tem uma média valor de zero.

$$\int_{-\infty}^{\infty} \psi(t)\, dt = 0 \tag{2.134}$$

Que é dilatado com um parâmetro de escala s e posicionado por u:

$$D = \{\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-u}{s})\}_{u\in\mathbb{R},s>0} \tag{2.135}$$

A transformada continua Wavelet da função f(t) em uma escala s na posição u é a projeção de f(t) em sua correspondente Janela:

$$Wf(u,s) =< f, \psi_{u,s} >= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\psi^*(\frac{t-u}{s})dt \tag{2.136}$$

Sendo: $f(t)$ sinal a ser decomposto

ψ^* função base wavelet com duração limitada no tempo.

s parâmetro de escala (dilatação ou compressão)

u parâmetro de deslocamento ou de posicionamento.

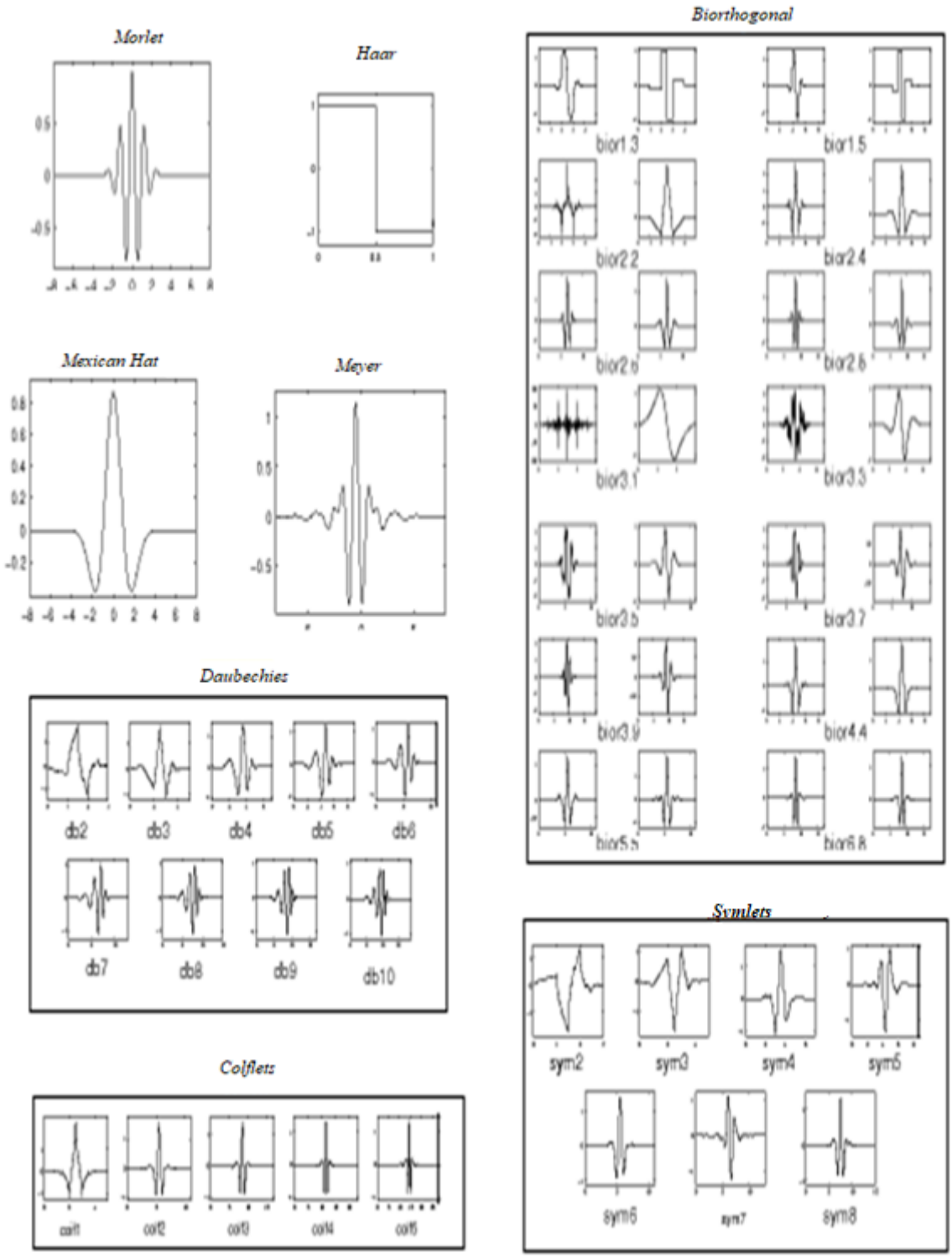
A transformada inversa x(t) da função *wavelet* continua é expressa por:

$$x(t) = \frac{C_\psi^{-1}}{s^2}\iint_{-\infty}^{\infty} Wf(u,s)\,\psi^*(\frac{t-u}{s})duds \tag{2.137}$$

Sendo: C_ψ^{-1} Constante que depende da função *wavelet* (ψ) escolhida.

Na Figura 33 se ilustra algumas famílias *wavelet*

Figura 33-Familias *wavelets*



Fonte: Mathworks (2017)

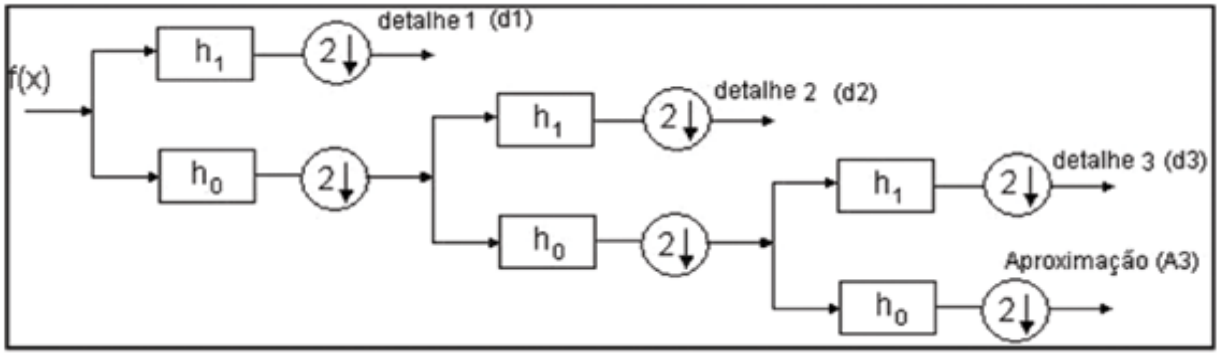
2.4.10.9 A transformada wavelet discreta

A transformada *wavelet* é uma ferramenta matemática de análise de sinais, que por meio de operações de escalonamentos e translações em uma função base (wavelet mãe), decompõe um sinal em diferentes faixas de frequências. Na maioria das aplicações práticas, predomina a utilização da TWD (Transformada wavelet discreta), comumente implementada na forma de um banco de filtros escalonados em oitavas (Mallat, 2009).

Segundo Silva, Oleskovicz e Coury (2005). uma das formas para implementação da TWD consiste no sucessivo agrupamento de pares de filtros passa baixa e passa alta em cada estágio de escala da TW. Isto pode ser visto como sucessivas aproximações da mesma função, ao qual, cada aproximação ressalta importantes informações relativas a uma dada escala (faixa de frequência).

A Figura 34, ilustra um processo de análise multiresolução em três níveis de resolução. Os detalhes e aproximações do sinal original $f(x)$ são obtidos por meio de bancos de filtros, os quais são formados por filtros passa-baixa (h_0) e passa-alta (h_1). Um filtro passa-baixa remove os componentes de altas frequências, enquanto o filtro passa alta separa o conteúdo de alta frequência no sinal sendo analisado (SILVA, OLESKOVICZ E COURY, 2005).

Figura 34-Processo de decomposição de um sinal em filtros passa alta e filtros passa baixa



Fonte: Silva, Oleskovicz e Coury (2005).

Aplicando-se a TWD a um sinal discreto c_o, com k_f amostras e fs frequência de amostragem, tem-se:

$$c_j(k) = \sum_n h(n - 2K)\, c_{j-1}(n) \tag{2.138}$$

$$d_j(k) = \sum_n g(n - 2K)\, c_{j-1}(n) \tag{2.139}$$

sendo c_je d_j os coeficientes de aproximação e *wavelet* da escala j, respectivamente; h(k) e g(k) os filtros passa-baixa e passa-alta, respectivamente (COSTA *et al*., 2010).

O método proposto baseia-se no uso da energia janelada dos coeficientes *wavelet* da primeira escala. Em uma escala j, a energia janelada é calculada através de uma janela deslizante nos coeficientes *wavelet*, deslocando-se um coeficiente a cada passo de tempo (COSTA et al., 2007; COSTA et al., 2008), como segue:

$$e_j(k) = \sum_{n=1}^{\frac{w}{2^j}} d_j^2\,(n), \quad se \;\; 1 \leq k \leq \frac{w}{2^j} \tag{2.140}$$

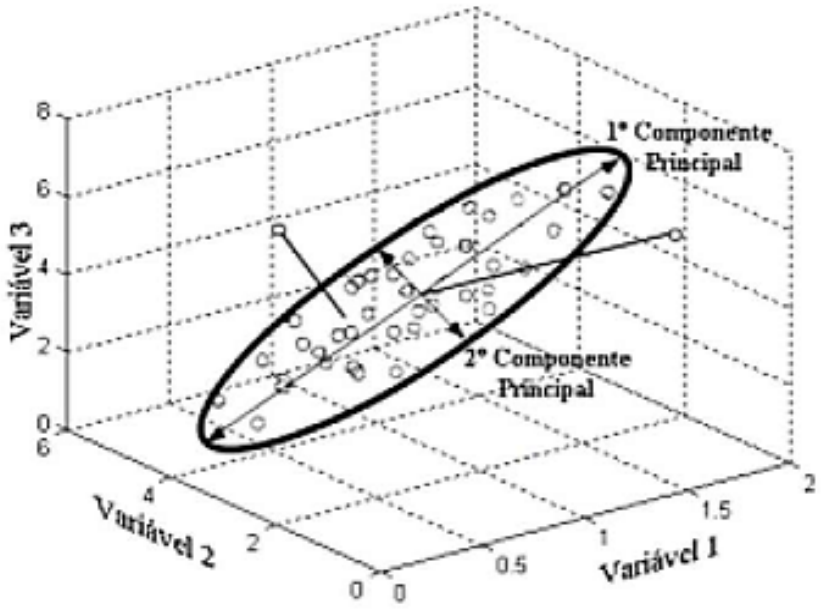
$$e_j(k) = \sum_{n=k-\frac{w}{2^j}+1}^{K} d_j^2\,(n), \quad se \;\; \frac{w}{2^j} \leq k \leq \frac{K_f}{2^j} \tag{2.141}$$

Sendo w o tamanho da janela, k= {1, 2, . . ., $k_f/2^j$} e k_f o número de amostras do sinal original. O número de amostras do sinal original deve ser múltiplo de 2^j de modo que k seja um número inteiro.

2.4.10.11 *A Análise dos componentes principais (ACP)*

A Análise de componentes principais (ACP) é uma técnica da estatística multivariada que consiste em transformar um conjunto de variáveis originais, por projeção em planos ortogonais, em outro conjunto de variáveis de mesma dimensão denominadas de componentes principais. Os componentes principais apresentam propriedades importantes: cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e estimados com o propósito de reter, em ordem de estimativa, o máximo de informação, em termos da variação total contida nos dados, podendo facilitar a interpretação da independência entre elas (HONGYU, 2015). A representação geométrica dos componentes principais é ilustrada na Figura 35.

Figura 35-Representação geométrica das componentes principais



Fonte:Oliveira (2010), *apud* Marques (2011).

A ACP pode ser caracterizada também, através das distancias dos dados a um determinado subespaço ortogonal. Assim, a primeira componente principal corresponde a linha que minimiza a soma das distâncias quadradas de pontos de dados. A segunda componente principal é definida como a linha que, combinada com a primeira, minimiza a soma do quadrado da distância dos pontos para esta combinação (ALFARO *et al.,* 2014).

Segundo Marques (2011), as primeiras componentes principais explicam a maioria da variância total contida no conjunto de dados, e podem ser usadas para representa-los. Assim, o valor da i-ésima componente principal de um conjunto de m variáveis pode ser determinado segundo a Equação 2.142, sendo Z_i a componente principal, a_{ij} são os pesos e X_j são as variáveis originais.

$$Z_i = \ a_{i1}.X_1 + a_{i2}.X_2 + \cdots + a_{im}.X_m \tag{2.142}$$

É possível representar as variáveis nos planos ortogonais, que revelam as covariâncias entre as variáveis originais, em cada uma das componentes principais. mapeando a estrutura de

covariância dos dados. Apenas os dados são extraídos, os ruídos são demonstrados nos gráficos dos resíduos. (RAJALAHTI;KVALHEIM, 2011; SOUZA *et al.,* 2013)

2.5 TRABALHOS DESENVOLVIDOS PARA PREVISÃO DE DADOS EM ESTAÇÕES DE TRATAMENTO DE EFLUENTES.

Nos últimos anos, estudos de modelos inteligentes relacionados a Estação de Tratamento de Águas Residuárias (ETAR) tornou-se popular por causa da crescente preocupação sobre o meio ambiente. Os desenvolvimentos em métodos inteligentes tornam possível a sua utilização em modelagem de sistemas complexos. A modelagem inteligente foi usada em primeiro lugar para aumentar a robustez dos modelos existentes, mas agora é utilizada para se obter novos modelos. Os modelos matemáticos complexos incluem muitos dados bioquímicos dos processos. No entanto, devido à escassez de medidas, é quase impossível obter estimativas confiáveis de parâmetros dinâmicos desconhecidos, portanto são necessários modelos gerenciáveis. Tão consideráveis esforços foram gastos para mudar ou modificar o tradicional controle de processos de tratamento de águas residuais (MULLER *et al*., 2002).

Operação inadequada de uma estação de tratamento de efluentes pode gerar graves problemas ambientais e públicos, problemas de saúde, como seu efluente para um corpo de água receptor. Pode causar ou espalhar várias doenças para os seres humanos. Um melhor controle de uma ETAR pode ser alcançado através do desenvolvimento de modelos robustos para prever o desempenho da planta com base na observação passada, de certos parâmetros fundamentais. Contudo, modelar uma ETAR é uma tarefa difícil devido à complexidade dos processos de tratamento. O complexo físico, biológico e processos químicos envolvidos em águas residuais exibem comportamentos não-lineares que são difíceis de descrever por modelos matemáticos lineares (HAMED, KHALAFALLAH E HASSANIEN, 2004)

Segundo Hanbay,Turkoglu e Demir (2008), muitos estudos foram realizados em águas residuais, baseado em métodos inteligentes. As pesquisas estão relacionadas à modelagem da planta de tratamento de efluentes. As invertigações são sobre previsões de parâmetros de Estações de Tratamento de Águas Residuárias (ETAR), estimando as características dos parâmetros de controle e de saída da ETAR.

A combinação dos métodos de pré-tratamento de dados para obtenção de melhores

estruturas usando as redes neurais tem sido muito utilizada para prognóstico de séries temporais, dando origem ao que é intitulado de sistemas híbridos. Os sistemas híbridos são a junção de duas ou mais técnicas em uma mesma metodologia, destacando ser essa uma tendência atual de investigação (KHASHEI; BIJARI, 2012).

Bousfield, Silva e Chapuis (2011), fez uso de uma rede neural artificial para previsão do volume de lodo gerado em estações de efluente têxtil, coletando dados de análises realizadas no laboratório no período de 2002 a 2007, com os seguintes parâmetros: temperatura (Temp), oxigênio dissolvido (OD), pH, turbidez (Turb) e cor, fósforo total (PT), nitrogênio amoniacal (NAm), sólidos suspensos (Ssus), demanda bioquímica de oxigênio (DBO5), demanda química de oxigênio (DQO), sólidos sedimentáveis (SSed) e sólidos totais (ST).

Martínez *et al*. (2017) utilizaram séries temporais usando a média móvel em 2ª ordem e terceira ordem e suavização exponencial, com objetivo de fazer a previsão da Demanda Bioquímica de Oxigênio (DBO) na saída de uma lagoa de estabilização no México, conseguindo os resultados registrados na Tabela 5, na qual os números que aparecem na Tabela nas ultimas linhas da previsão são a previsão futura de um período a frente dos dados e os dados da última linha dos erros, são a soma dos erros na Tabela.

É necessário esclarecer que foram usados 17 meses de valores conhecidos de DBO para fazer a previsão do decimo oitavo (um período a frente) mês, sendo utilizados os valores de 96.8, 86, 57.9, 88.6, 86.4, 77, 53, 82, 71, 61, 97, 71, 80; 95, 111, 76 e 81 $mg.L^{-1}$ de DBO na previsão .

Tabela 5 -Resultado das séries

Média móvel de Ordem 2		Média móvel de Ordem 3		Suavização Exponencial	
Previsão de	Típico erro	Previsão de	Típico erro	Previsão de	Típico erro
***N/D**	***N/D**	***N/D**	***N/D**	***N/D**	***N/D**
91	***N/D**	***N/D**	***N/D**	97	***N/D**
72	11	80	***N/D**	91	***N/D**
73	15	78	***N/D**	75	***N/D**
88	11	73	15	32	22
82	3	84	9	84	21
65	9	72	13	81	9
68	13	71	13	67	17
77	11	69	13	74	19
66	5	71	9	73	18
79	13	76	13	67	11
84	16	76	14	82	19
76	10	83	12	76	20
88	6	82	8	78	19
103	8	95	12	87	12
94	14	94	16	99	17
79	13	89	15	87	22
	157		**162**		**225**

Fonte: Martínez *et al.* (2017)
Nota: * N/D não definido.

Chegou-se à conclusão neste estudo que a média móvel de ordem 2 dava um melhor prognóstico para o valor do decimo oitavo mês (ou um período a frente).

Segundo Rustum e Adeloye (2006), a necessidade de modelos para controle de estações de tratamento de águas residuais e as limitações dos modelos determinísticos atuais exigem a aplicação de técnicas de Inteligência Artificial (AI) para o modelagem do processo, para isto usaram um sistema de logica difusa (FUZZY) e rede neural que usa o mapa auto organizável de Kohonem com software MATLAB, para previsão da Demanda Bioquímica de Oxigênio (DBO) de uma estação de tratamento de efluentes por lodos ativados em Edimburgo (Escócia). Utilizou-se 1066 dados da ETE, variando-se a entrada entre três e no máximo cinco parâmetros. Usou-se como dados de entrada os parâmetros: DBO de entrada; oxigênio dissolvido; sólidos suspensão solúveis no recilco de lodo; índice volumétrico de lodo e o fator de carga (F/M) obtendo-se os resultados descritos na Tabela 6.

Tabela 6- Desempenho do sistema ANFIS

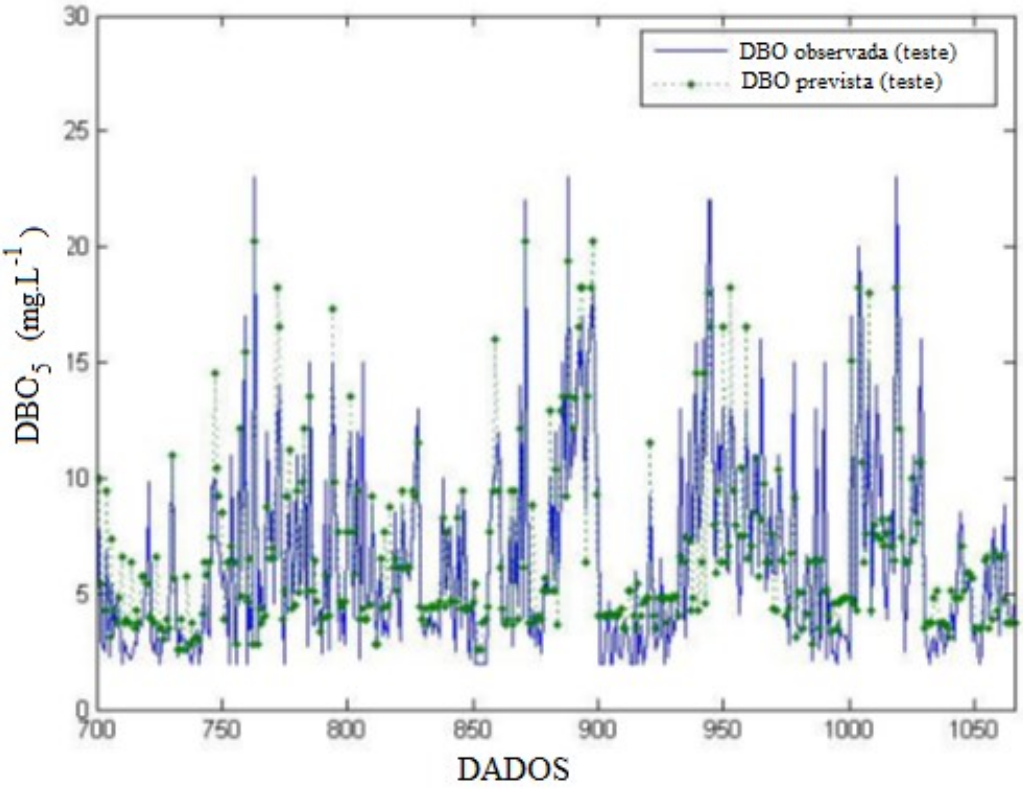
	Parâmetros de entrada		Correlação(%)			Mínimo		
	Funções de associação		Treinamento	Validação	Teste	Treinamento	Validação	Teste
		Variáveis observadas				2 0	2 0	2 0
M1	2	DBO de entrada, OD, SSR, IVL, F/M	81	78	76	0,6	-0,1	0,9
M2	3	DBO de entrada, SSR, IVL, F/M	73	77	72	-13,0	1,6	-0,1
M3	2	DBO de entrada, SSR, IVL, F/M	54	69	66	2,5	1,4	0,3
M4	5	DBO de entrada, IVL, F/M	64	65	62	1,3	3,0	1,9
M5	4	DBO de entrada, IVL, F/M	75	69	64	-4,5	3,2	2,0
M6	3	DBO de entrada, IVL, F/M	63	67	64	0,7	3,2	1,9
M7	2	DBO de entrada, IVL, F/M	48	61	63	4,1	-0,6	1,2
M8	2	DBO de entrada, SSR, SSD, IVL, F/M	94	89	90	3,2	3,5	1,4
M9	3	DBO de entrada, SSR, IVL, F/M	95	91	92	2,8	3,5	2,6
M10	2	DBO de entrada, SSR, IVL, F/M	90	86	86	2,4	2,3	2,1
M11	4	DBO de entrada, IVL, F/M	93	89	89	1,0	3,0	2,3
M12	3	DBO de entrada, IVL, F/M	90	87	83	1,8	2,7	0,1
M13	2	DBO de entrada, IVL, F/M	80	79	77	2,6	2,5	2,6

Fonte:Rustum e Adeloye (2006) adaptado.

Os modelos de dados brutos de ANFIS são denotados por M1-M7 na Tabela 6, ao qual se observa baixa correlação e valores negativos da DBO de saída.

Os modelos de KSOM-ANFIS híbrido são denotados por M8-M13 e apresentam melhor correlação, sendo escolhido o modelo M9 para realizar previsões no método, com os resultados da Figura 36.

Figura 36- Série temporal-valores observados e previstos usando KSOM-ANFIS



Fonte: Rustum e Adeloye (2006) adaptado.

Neste estudo se usou 566 dados ao todo, 200 dados para validação e 366 de treinamento e teste. Os dados de entrada da rede, são aqueles que correspondem a maior correlação entre os parâmetros de entrada e a DBO de saída que se desejava prever. A rede ao que se entende é sempre alimentada por dados reais para uma previsão da DBO de saída.

Heikkinem, Heikkinem e Hiltunem (2008) usaram Perceptron Múltiplas Camadas (MLP), para fazer a previsão da DBO de saída com até cinco dias de previsão a frente, de uma estação de tratamento de efluentes de uma fábrica de celulose, com sistema de tratamento por lodos ativados na vazão de mil litros por segundo. Para tal utilizou-se como dados entrada da rede 22 variáveis de entrada (Tabela 7) e 1378 dados em uma rede neural (MLP), com uma camada oculta.

Tabela 7 - Variáveis usadas para modelagem concentração COD do efluente

Variáveis	unidade
1 Vazão de reciclo de lodo	$m3.d^{-1}$
2 Vazão do efluente	$m3.d^{-1}$
3 Sólidos totais na entrada da aeração	$g.m^{-3}$
4 Sólidos totais no tanque de aeração	$g.m^{-3}$
5 Sólidos totais na saida	$g.m^{-3}$
6 Sólidostotais no reciclo de lodo	$g.m^{-3}$
7 DQO na entrada di tanque de aeração	$g.m^{-3}$
8 DQO de saida	$g.m^{-3}$
9 pH na entrada do tanque de aeração	
10 pH na saida	
11 Condutividade na entrada do tanque de aeração	$mS.m^{-1}$
12 Nitrogênio na entrada do tanque de aeração	$g.m^{-3}$
13 Nitrogênio na saida	$g.m^{-3}$
14 Fósforo na entrada do tanque de aeração	$g.m^{-3}$
15 Fósforo na saida	$g.m^{-3}$
15 Oxigênio dissolvido na aeração	$g.m^{-3}$
17 Temperatura na aeração	ºC
18 Sedimentabilidade do lodo	$ml.L^{-1}$
19 Índice Volumétrico de Solidos Decantáveis	$ml.g^{-1}$
20 Índice Volumetrico de lodo (IVL)	$ml.g^{-1}$
21 Idade de Lodo	d
22 Carga de DQO do lodo na entrada	$(Kg.DQO.Kg.SSV^{-1}).d^{-1}$

Fonte: Adaptado de Heikkinem, Heikkinem e Hiltunem (2008)

Para tratamento dos dados, os autores aplicaram uma interpolação linear para preencher dados em branco (na ausência de dados devido à quebra de equipamentos), usou-se 950 dados para treinamento da rede e o restante para validação dos dados. Os resultados da previsão para até cinco dias à frente estão na Tabela 8.

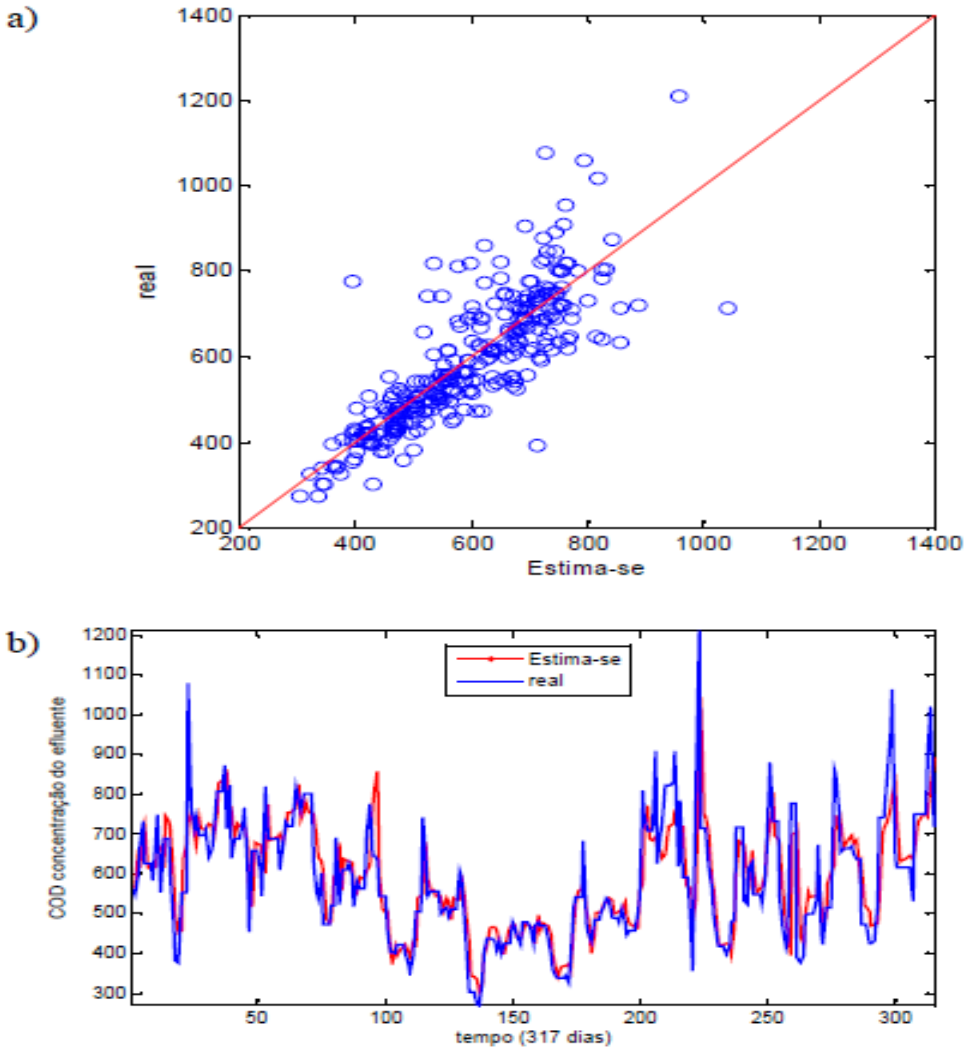
Observou-se correlação diminui conforme os dias de previsão futuro pretendidos variando de 0,43 para previsão de cinco dias a frente e 0,83 (melhor previsão para um dia a frente). Os dados de correlação e previsão para um dia a frente estão na Figura 37 a e b.

Tabela 8 -O coeficiente de correlação (r), e erro de raiz quadrada (RMSE) entre os valores reais de DQO de uma planta de tratamento de lodos ativados e as previsões, análise de rede neural

Previsão	r	RMSE
t-1	0,83	85,3
t-2	0,69	112,4
t-3	0,55	127,8
t-4	0,50	137,7
t-5	0,44	138,6

Fonte: Heikkinem, Heikkinem e Hiltunem (2008) (adaptado)

Figura 37- a) Correlação para previsão de apenas um dia a frente da DQO, usando rede neural. b) Gráfico com previsão para um dia a frente.

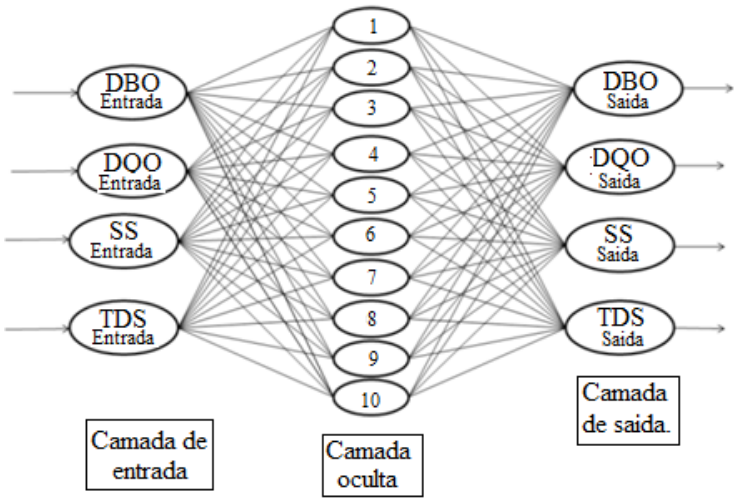


Fonte: Heikkinem, Heikkinem e Hiltunem (2008)

Observamos na Tabela 8, que a previsão um dia a frente tem correlação maior e um erro menor que os outros dias, aumentando o erro à medida que se aumenta o número de dias na previsão futura.

Sindhu *et al.,(*2012) utilizaram um modelo baseado em redes neurais EBPN (*Error Back Propagation Network*) e RBFN(*Radial Basis Function Network*) para previsão de resultados de uma planta piloto usando um reator anaeróbio de fluxo ascendente com um volume de operação de 56,52 L, uma altura de 0,9 m e diâmetro interno de 0,30 m, alimentado com pré-esgotos domésticos selecionados da área de Nehru Nagar, Bhilai, C.G., Índia misturado com resíduos vegetais pulverizado. Neste estudo que usa uma rede neural não em série temporal, mas dados de entrada, para prever na saída da planta, a demanda bioquímica de oxigênio (DBO), demanda química de oxigênio (DQO), sólidos (SS) em suspensão e sólidos totais dissolvidos (STD). Os resultados de simulação indicaram que o erro médio percentual absoluto (MAPE) de 11.86, 15.53, 26.67 e 22.26 para DBO, DQO, SS e TDS. A rede neural usou 10 neurônios na camada escondida, Foram coletados dados de janeiro de 2009 e dezembro de 2010, usou-se 489 dados dos quais 291 dados são considerados para o treinamento, enquanto 198 dados são considerados para testar o modelo de ANN (Figura 38).

Figura 38-Arquitetura de EBPN para previsão



Fonte: Sindhu *et al. (*2012)

Observamos, que a previsão obtida no trabalho de Sindhu. *et al*. (2012), não se trata de uma previsão em série temporal, mas de dados estáticos de entrada *versus* dados de saída previstos, pois usa-se as análises de entrada com objetivo de se prever as mesmas variáveis na saída.

Turner e Edebali (2015) fizeram uma modelagem para uma estação de tratamento de efluentes da cidade de Konya na Turquia, se utilizou regressão multilinear e rede neural artificial com diferentes arquiteturas no software SPSS e MATLAB. Todos os dados foram obtidos da estação de tratamento de águas residuais, com registros diários de quatro meses. A eficiência de tratamento da planta foi determinada levando em consideração os valores de entrada de pH, temperatura, DQO, SST e DBO com valores de saída de DQO. Para comparar a performance do modelo, o coeficiente de determinação (R^2) e *Mean Squared Error* (MSE) foram utilizados. No método de regressão multilinear, para compreender os efeitos dos parâmetros testados, a função de regressão foi desenvolvida. A maior eficiência de predição foi obtida em duas camadas ocultas nos modelos da Rede Neural Artificial. De acordo com o estudo de modelagem, os modelos da Rede Neural Artificial deram resultados mais satisfatórios do que o modelo de Regressão Multilinear.

Wei, Kusiak e Hosseini (2013) utilizaram dados de radares meteorológicos, precipitação e velocidade dos ventos, para verificar efeitos de variação de vazão de entrada em uma Estação de tratamento de efluentes sanitários da cidade de Des Moines, Iowa, operada para servir uma população de 317.930 habitantes. Para isto usou-se dados registrados por radares meterológicos em período de três anos e três meses realizando modelagem por métodos estatísticos e várias redes neurais (MLP) na previsão de vazão de até 180 minutos a frente, obtendo melhores resultados da rede neural, com 60 dados de entrada, uma camada oculta de 58 neurônios e uma variável de saída, a vazão de entrada da estação, foi construído os resultados da Tabela 9, ao qual são apresentados valores de correlação para os resultados no tempo t, t+15, t+30, t+60, t+90, t+120, t+150 e t+180 ao qual o tempo foi medido com avanço de minutos, e medidos o coeficiente de correlação, o erro absoluto médio e o erro médio quadrático.

Tabela 9-Resultados da previsão

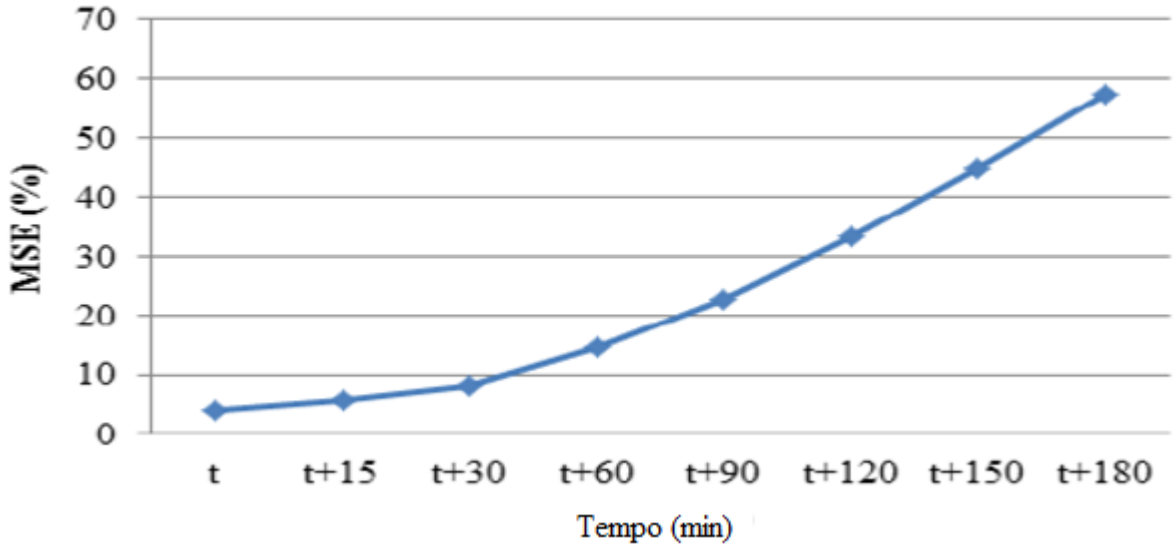
	T+0 (teste)	T+15	T+30	T+60	T+90	T+120	T+150	T+180
Correlação (R)	0,98	0,98	0,97	0,95	0,93	0,9	0,87	0,83
MAE	1,09	1,48	1,89	2,75	3,61	4,46	5,26	6,02
MSE	4,21	5,83	8,2	14,59	22,95	33,21	44,88	57,39

Fonte: Wei, Kusiak e Hosseini (2013) (adaptado)

Na Figura 39, esta apresentação da evolução erro médio quadrático para previsão, variando com o tempo futuro. Observamos que quanto mais distante a previsão futura maior é o erro, conforme também ilustrado na Figura 40, 41 e 42, nos quais observa-se a defasagem entre a previsão e os dados reais com o aumento do tempo.

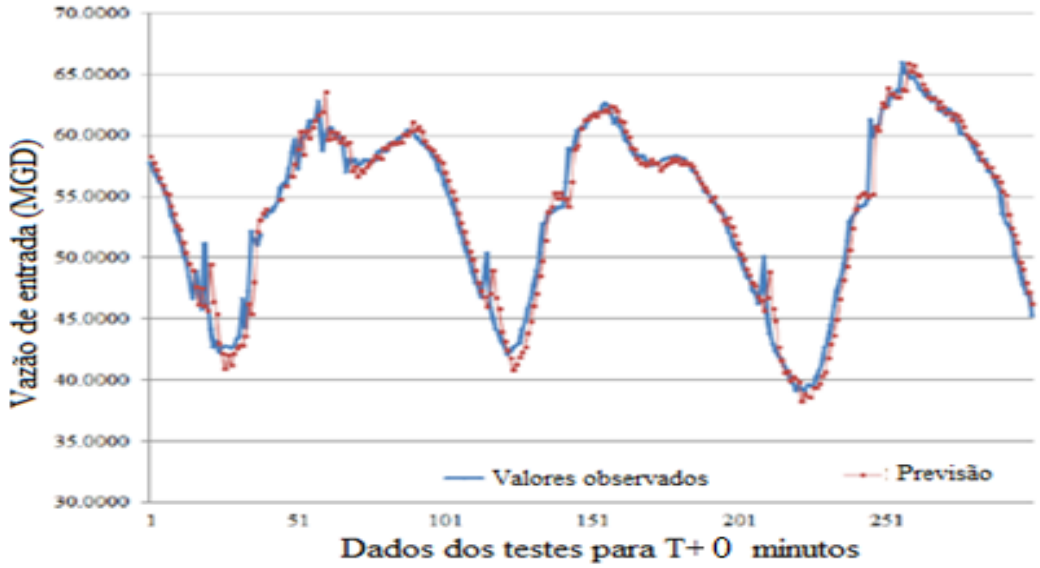
Os chamados passos à frente podem ser em minutos, dias, meses ou anos, dependem da unidade usada no estudo.

Figura 39-Variação do erro médio quadrático com a previsão futura



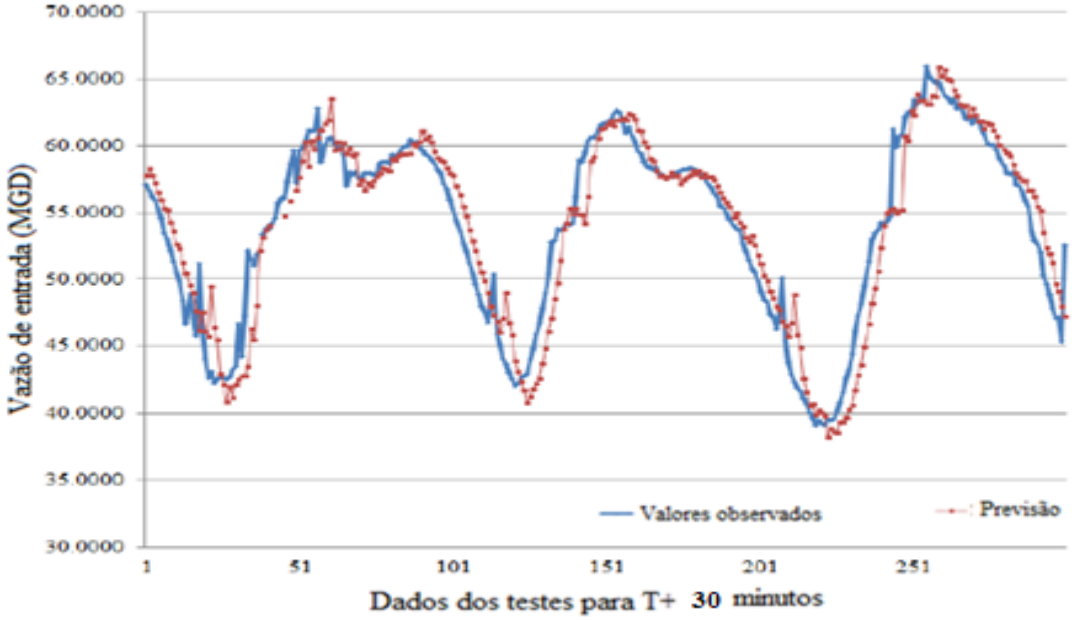
Fonte: adaptado de Wei, Kusiak e Hosseini (2013).

Figura 40-Previsão da vazão de entrada sem defasagem de tempo.



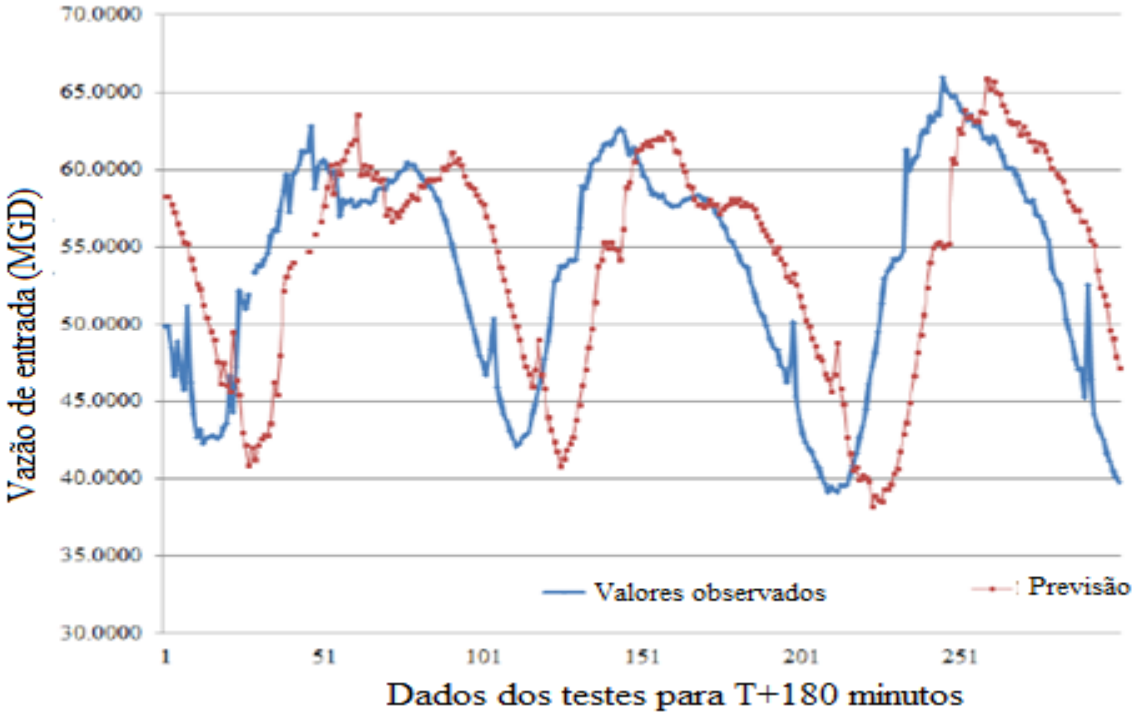
Fonte: Adaptado de Wei, Kusiak e Hosseini (2013)

Figura 41-Previsão da vazão de entrada com defasagem de 30 min.



Fonte: Adaptado de Wei, Kusiak e Hosseini (2013).

Figura 42-Previsão com defasagem de 180 minutos.



Fonte: Adaptado de Wei, Kusiak e Hosseini (2013).

Wei (2013) realizou um estudo de previsão de dados em estações de tratamento de efluentes, usando modelos matemáticos com redes neurais entre esses a previsão dos sólidos em suspensão voláteis de uma estação com lodos ativados. Utilizou-se cinco métodos para previsão, ou seja, k-*nearest neighbors* (k-NN); *multi-variate adaptive regression spline* (MARS); *neural network* (NN); *support vector machine* (SVM); e *random forest* (RF) algoritmo, cujo os resultados estão citados na Tabela 10.

Os dados da Tabela 10, são para apenas um dia de previsão futura ou um passo à frente, observa-se que das técnicas a que melhor se adaptou para previsão dos Sólidos em Suspensão Totais (SST) em uma ETE foi a rede MLP. Em seguida, foi feito um estudo para previsão de sete dias a frente usando o melhor método na pesquisa.

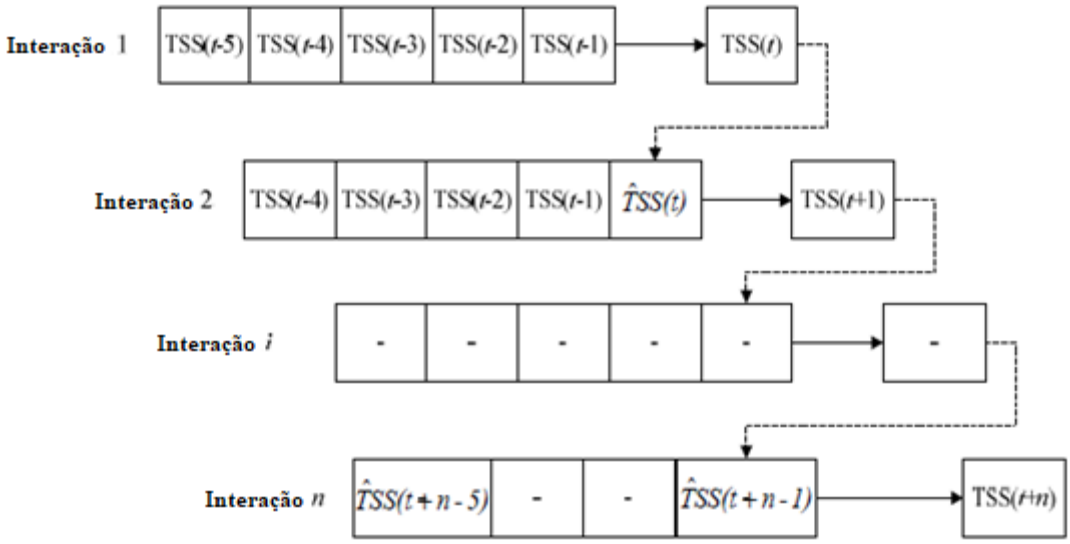
Tabela 10 -Resultado da previsão de SST, um dia a frente e os erros médios absolutos e relativos

Algoritimo	MAE	MRE%
K-NN (K=10)	62,15	26,46%
RF	52,19	21,66%
NN	38,88	16,15%
MARS	44,59	18,29%
SVM	61,36	26,10%

Fonte: Wei (2013)

Para se fazer a previsão futura na série temporal dos sólidos, Wei(2013) usou um rede neural com cinco dias anteriores de sólidos suspensos voláteis (modelo que envolve apenas a variável de estudo ou modelo univariável), visando a previsão de um passo à frente conforme Figura 43. O valor encontrado na primeira previsão alimentou a rede neural treinada para se obter o valor futuro na segunda predição ou dois passos à frente. Repetindo o procedimento sucessivamente até os dias a frente desejados. Evidenciou-se o aumento do erro na previsão com os resultados listados na Tabela 11, na medida em que se aumentava o número de dias a frente ou os passos a frete na previsão.

Figura 43-Esquema usado para previsão n passos à frente, usando os últimos cinco dias.



Fonte Wei(2013)

Tabela 11-Resultados para rede neural.

Passos à frente (dias)	MAE	MRE%	Estrutura MLP	Função de ativação da camada oculta	Função de ativação da camada de saida
1,00	44,04	18,54	MLP 5-12-1	Tangente hiperbolica	Exponencial
2,00	46,15	19,19	MLP 5-21-1	Identidade	Exponencial
3,00	46,80	19,60	MLP 5-3-1	Logistica	Identidade
4,00	47,05	23,14	MLP 5-25-1	Exponencial	Identidade
5,00	49,99	23,82	MLP 5-25-1	Exponencial	Exponencial
6,00	51,22	25,74	MLP 5-13-1	Tangente hiperbolica	Tangente hiperbolica
7,00	50,76	26,58	MLP 5-25-1	Identidade	Exponencial

Fonte Wei (2013) adaptado

Na Tabela 11, tem os resultados de previsão uma semana à frente, com erro relativo de 26,58%, para sete dias à frente e erro absoluto de 50,76 $mg.L^{-1}$.

A Tabela 12, ilustra a melhoria do erro ao longo do tempo para o aprendizado dinâmico ou seja, a rede recebe treinamento a cada dia que em que se é calculado um valor futuro, se adicionando os valores reais aos mais recentes calculados. Ao aplicar o esquema iterativo de aprendizagem NN, uma melhoria de 5% no erro absoluto médio uma melhora de 4% na media do erro absoluto relativo. Os resultados apresentados na Tabela 12, indicam que o esquema de aprendizagem iterativo pode ser útil para fazer previsões em longo prazo.

Tabela 12 - Resultado da previsão para passos à frente de TSS, usando uma rede neural MLP com treinamento de aprendizagem dinâmica.

Passos à frente (dias)	MAE	MRE (%)
t+1	40,95	21,85
t+2	44,32	24,04
t+3	45,95	24,32
t+4	47,70	25,88
t+5	49,38	27,01
t+6	49,66	27,20
t+7	52,30	27,55

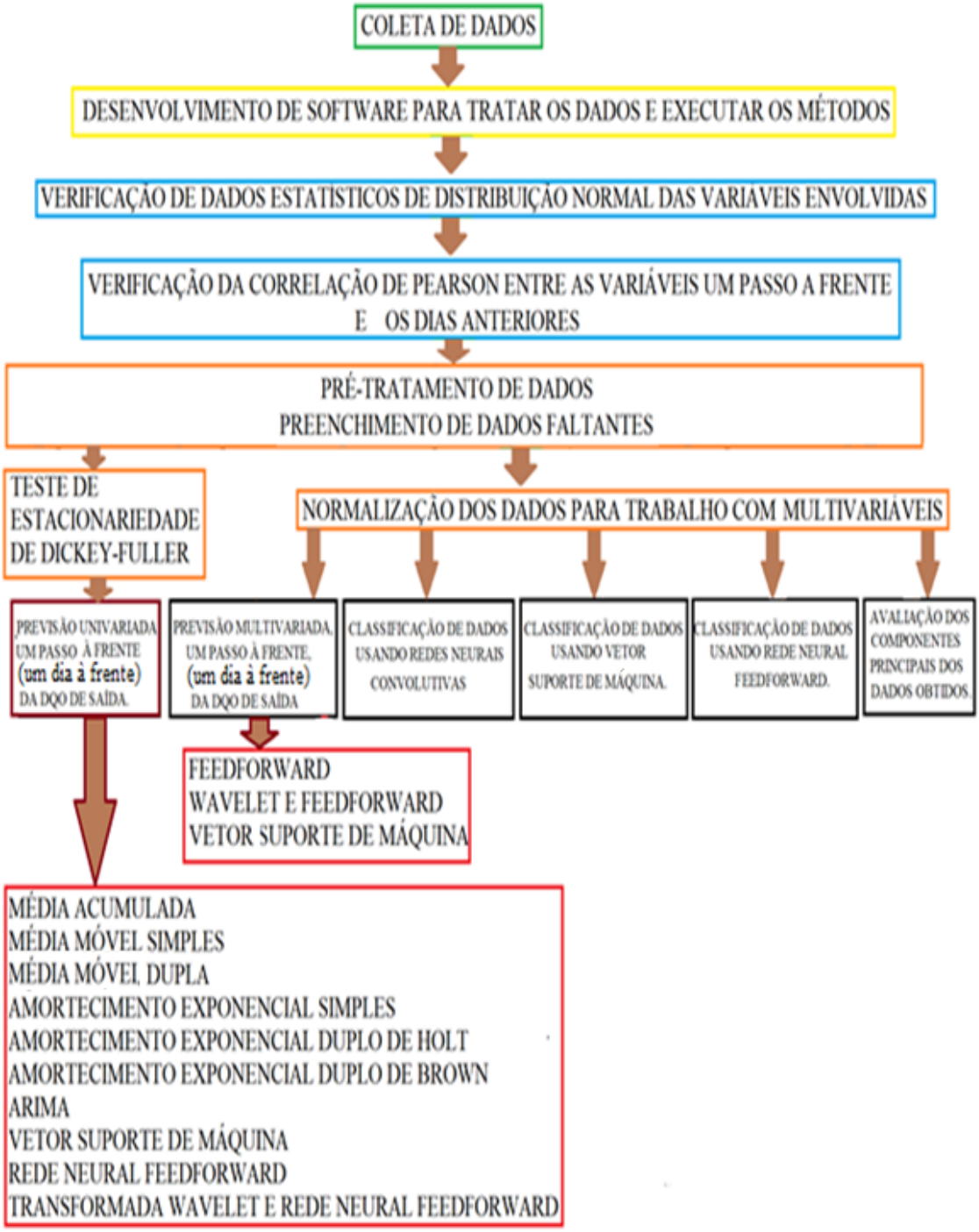
Fonte: Wei (2013) adaptado

Sridevi et al. (2014) investigou-se a biodegradação e a produção do biohidrogênio num reator híbrido de fluxo ascendente com manta de lodo anaeróbio usado para águas residuais de substrato de destilaria, com uma taxa máxima de produção de hidrogênio igual a 1300 ml.d-1 . Foi desenvolvido e validado uma rede neural back-propagation com topologia de rede de 4-20-1 por meio do algoritmo LevenbergMarquardt, utilizando um total de 231 pontos de dados para analisar o desempenho do reator em fase de aclimatação e operação.

3 METODOLOGIA

Na Figura 44, se encontra o fluxograma com as etapas executadas para o desenvolvimento da metodologia esperada, ao qual cada etapa será descrita.

Figura 44 -fluxograma com as etapas executadas para o desenvolvimento da metodologia



O objetivo principal deste trabalho é desenvolver uma metodologia de controle e previsão do comportamento de eficiência de um sistema biológico de tratamento de efluentes, usando técnicas estatísticas e alguns sistemas de inteligência artificial, permitindo maior controle sobre a ETE (no caso em estudo um reator anaerobio de fluxo ascendente) e simulações do processo dentro de padrões e características de uma planta em condição normal de operação

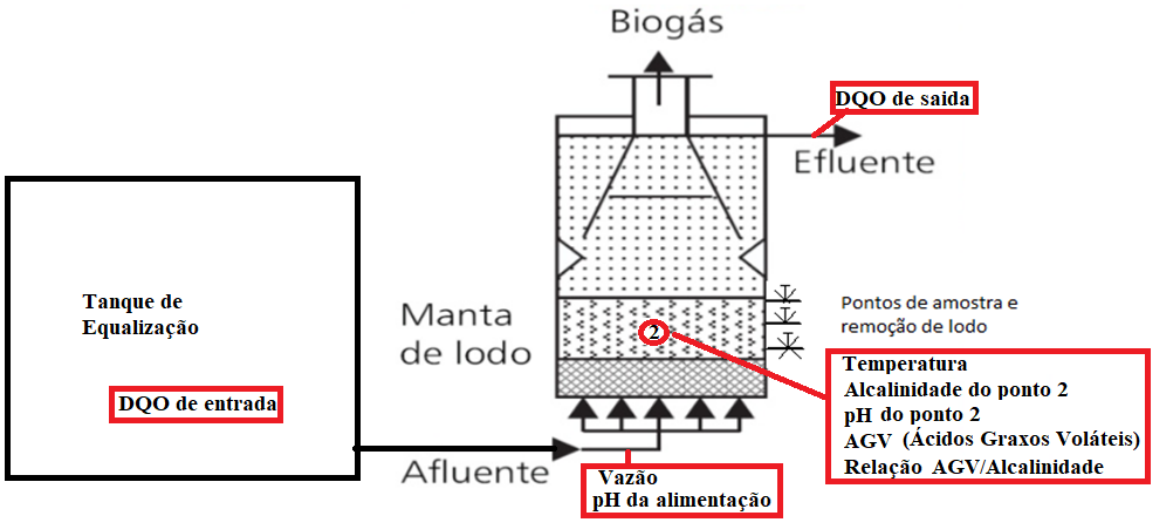
Além de uso de técnicas de previsão usando uma variável (modelo univariado) em série temporal, a previsão futura será baseada em outras variáveis que afetam o comportamento da variável em estudo (uso de métodos multivariados), visando a minimização dos erros nas previsões e objetivando previsões um passo à frente, que neste estudo foi o período de um dia a frente (usado tanto para previsão como classificação).

3.1 COLETA DE DADOS

Para elaboração da metodologia, foram coletados 725 dados de dez variáveis de controle de um reator anaeróbio de fluxo ascendente (7250 dados ao todo), durante o período de dois anos (2014 e 2015). Estes efluentes industriais foram de uma engarrafadora de refrigerantes, ao qual as variáveis usadas nos métodos foram: DQO do tanque de Equalização (ou DQO de entrada do reator anaeróbio medida em $mg.L^{-1}$); Vazão de alimentação do tanque anaeróbio (medida em $m^3.h^{-1}$); pH de alimentação do reator anaeróbio; pH no ponto 2 ou ponto intermediário da manta de lodo no reator; Ácidos Graxos Voláteis(AGV); alcalinidade; carga orgânica aplicada (Produto da vazão de entrada pela DQO de entrada do reator anaeróbio); relação AGV/alcalinidade; temperatura do reator e DQO de saída.

Na Figura 45, se encontra o esquema dos locais ao qual estes dados foram coletados.

Figura 45 -Esquema com localização dos pontos de amostragem e as análises



3.2 DESENVOLVIMENTO DE SOFTWARE PARA TRATAR OS DADOS E EXECUÇÃO DOS MÉTODOS

Para manipulação de um grande volume de dados e desenvolvimento do método foi necessário a elaboração de alguns softwares, que operam no sistema Windows.

Foi usado a linguagem de programação para VISUAL BASIC versão 5.0, na avaliação da distribuição normal das curvas, e correlação de Pearson entre as variáveis envolvidas no processo.

Alguns métodos estatísticos mais básicos como as séries temporais usando média, média móvel, amortecimento exponencial simples, amortecimento exponencial duplo foram desenvolvidos diretamente usando a planilha Excel (2016). Nos testes de estacionariedade das séries foram usados os *softwares Real-Statistics, Excel e GRETL*.

Para as redes neurais feedforward, vetor suporte de máquina, Wavelet e feedforward, bem como a classificação dos dados usando redes neurais convolutivas de aprendizado profundo (CNN), SVM, feedforward e PCA se usou o MATLAB versão 2017a em sua linguagem de programação.

3.3 VERIFICAÇÃO DE DADOS ESTATÍSTICOS DE DISTRIBUIÇÃO NORMAL DAS VARIÁVEIS ENVOLVIDAS ECORRELAÇÃO DE PEARSON.

Todas as variáveis envolvidas no processo foram analisadas quanto ao seu comportamento no período de coleta, para verificar os respectivos desvios padrões, média e faixa de amplitude, o período foi dividido em três épocas, com objetivo de se verificar mudanças nas variáveis de controle e comparar resultados dos métodos com o desvio normal do processo.

Foram verificadas as correlações de Pearson entre as variáveis um passo à frente e os dias anteriores, para se determinar a influência entre os dados passados e futuros e ainda a correlação entre as demais variáveis e a variável que se deseja a previsão (no caso a DQO de saída do tanque Tanque de Metanização (TM)).

3.4 PRÉ-TRATAMENTO DE DADOS

As análises de laboratório e de controle, realizadas em uma planta podem ter erros operacionais e também inclusão de dias atípicos, como, por exemplo, foi visto neste trabalho:

-O operador transcreveu por engano um pH de 7,56 para um valor de 75,6;

-Nos feriados, ou dias em que a indústria não produziu, a vazão e a Demanda Química de Oxigênio de entrada da estação de tratamento de efluentes, teve seus valores registrados como zero, porém para se manter vivos os microrganismos consumiram biomassa, que equivaleria a uma carga orgânica mínima aplicada;

-Por problemas nos equipamentos no laboratório, uma das análises de controle, não foi registrada.

Essas análises, para efeito de previsão ou de comportamento das séries foram interpoladas, para se mitigar o efeito da descontinuidade dos dados na série temporal. A escolha do método de interpolação foi o da média móvel em virtude de apresentar dados positivos evitando-se dados extrapolados, por outros métodos aplicados, como se descreve na seção (2.4.6.2), porém outros métodos univariados podem ser usados, inclusive a aplicação de redes neurais, para preenchimento dos dados não registrados.

Para se trabalhar com métodos que utilizem mais de uma variável (multidimensional), como as classificações, previsões de um passo à frente em séries temporais e Análise de Componentes Principais, as variáveis foram normalizadas utilizando a Equação 2.107.

Nas séries temporais em que se utilizou a transformada *wavelet* com redes neurais, tanto os dados de entrada, como os de saída da rede neural, foram transformados de dados temporais, para números envolvendo tempo e frequência, e após a obtenção dos resultados foram convertidos novamente aos valores originais.

3.5 CRITERIOS DE DESEMPENHO E AVALIAÇÕES DAS SÉRIES TEMPORAIS PARA PREVISÃO UM PASSO A FRENTE

Para se avaliar as séries temporais na previsão de um passo à frente (um dia a frente), se utilizou os dados da média diária das análises de laboratório, em período de dois anos de coleta, se separando para todos os métodos os últimos 30 dados das médias diárias das análises, para avaliação dos erros neste período. A previsão, no caso das redes neurais, é feita através de treinamento da rede usando os dados de um ano e onze meses e se verifica a previsão para o ultimo mês, calculando-se a previsão encontrada da média diária a cada dia, para se comparar com os dados reais. Assim que é feita a previsão de um dia a frente, os dados reais daquele dia, são substituídos na série para a próxima previsão e treinamento do método, até se completar um mês de previsão.

No caso das redes neurais em função da aleatoriedade das escolhas dos pesos iniciais e também dos dados escolhidos pela rede para testar a convergência do método, ocorre resultados diferentes para cada treinamento. Assim, no caso das redes neurais foram feitos dois testes de verificação, como comparação, para avaliação dos resultados.

Alguns métodos de previsão univariados são aplicados apenas para séries estacionárias, como é o caso da média móvel. Em uma estação de tratamento de efluentes se procura manter os resultados de saída sempre com valores máximos de remoção da carga orgânica, ao qual o projeto pode oferecer, assim é de se esperar que as variáveis de saída como a DQO tenham estacionariedade, oscilando em torno deste valor máximo de eficiência. Utilizou-se então o teste de Dickey-Fuller para se verificar a estacionariedade da série.

3.6 CRITÉRIOS DE DESEMPENHO E AVALIAÇÕES PARA CLASSIFICAÇÃO DOS DADOS UM PASSO A FRENTE.

Foi verificado neste item, observando os resultados das previsões nas séries temporais, que os resultados mostram tendência de se seguir os últimos valores reais usados na previsão, por exemplo: se o último valor da série subiu a previsão normalmente indica que o valor futuro previsto subirá também, ficando a previsão um passo à frente se comportando como uma curva defasada do dia anterior. Observou-se que nem sempre é este o comportamento real da previsão, muitas vezes ocorre a inversão dos resultados do dia anterior, uma vez que os valores devem oscilar em torno de uma média, no caso de estacionariedade da série. Para se obter mais informações sobre a previsão, foi feita uma previsão usando redes neurais convolutivas (CNN), *feedforward* e vetor suporte de máquina, na classificação e previsão futura, em valores acima ou abaixo da média.

No caso a escolha de simplificar a previsão em dois valores apenas (acima e abaixo da média), foi em função do número de dados restritos para realização do método.

Uma rede neural CNN usa imagens como dados de entrada computacional e se tem como base, no mínimo 1000 imagens para cada classificação, porém com 725 dados e ainda retirando 30 dias para teste poderíamos usar apenas 346 imagens considerando duas classificações.

Na rede CNN, se usou também artifícios para simplificar as Figuras, como a transformação de números divididos em três cores básicas de cinza sendo cada cor representada por faixas.

Uma faixa para valores um dia a frente, abaixo de 80% do desvio padrão da média, outra faixa entre valores 80% do desvio padrão abaixo da média e 80% do desvio padrão acima da média e o terceiro valor acima de 80% do desvio padrão da média.

Este procedimento permitiu a criação de Figuras com três cores distintas, que foram associadas a duas classificações (abaixo ou acima da média, para o valor um passo à frente). A redução do número de cores permitiu reduzir os valores numéricos da matriz que representa as figuras, tornado-as mais fáceis de serem codificadas e processadas.

Assim, as matrizes numéricas, que simbolizam as Figuras foram testadas também com dias de atraso, formando pixéis de tamanho diferente, como por exemplo, uma Figura com 10 linhas e 15 colunas, sendo cada ponto um pixel com a cor simbolizando acima da média e abaixo da

média (‘sobe’ ou ‘desce’), cada linha na Figura, simbolizando uma análise e cada coluna um dia de atraso, a Figura 10x15 significa 10 análises e 15 dias anteriores, associado a valores futuros que informam se este vai subir (acima da média) ou descer (abaixo da média).

Um exemplo de transformação dos dados em imagem está na Figura 46, que representa cada coluna um dia de atraso e cada linha uma variável, no caso deste exemplo, temos 10 variáveis transformadas (10 linhas) e 5 dias de atraso (cinco colunas) e as cores em cada pixel, representando um resultado da análise da variável correspondente. Cores branca informando se a variável ficou abaixo da média em menos de 80% do desvio padrão, cinza se esta ficou entre a média e mais ou menos 80% do desvio padrão e preta para valores acima, de 80% do desvio padrão acima da média. Estas Figuras foram associadas a duas classificações futuras, se o valor real estava acima ou abaixo da média

Figura 46- Representações gráficas de matrizes transformadas em Figuras, em três tons de cinza, com 10 linhas, representando dez dados das variáveis de controle de uma ETE e cinco linhas, representando cinco dias anteriores.

Para a classificação com o vetor suporte de máquina e rede neural *feedforward* não se usou nenhum procedimento de preparo dos dados, sendo aplicado diretamente o método classificando da mesma forma em duas regiões acima e abaixo da média, para prever os resultados.

3.7 CRITERIOS DE DESEMPENHO E AVALIAÇÕES DAS SÉREIS TEMPORAIS PARA CLASSIFICAÇÃO E AVALIAÇÃO DOS COMPONENTES PRINCIPAIS.

A utilização das técnicas de avaliação dos componentes principais, permitiu uma visão geral das influências das variáveis no comportamento do resultado de uma estação de tratamento de efluentes.

Na metodologia se programou o soft para colorir os pontos, identificando na variável DQO de saída, resultados maiores que um desvio padrão da média em cor vermelha, dados entre um desvio padrão acima da média e um desvio padrão abaixo da média na cor amarela e dados abaixo da média, mais que um desvio padrão, na cor verde. Com isto se permitiu visualizar dentro do gráfico das componentes principais, o que pode acontecer com os resultados da DQO de saída, com a variação de seus componentes.

4. RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentadosos resultados obtidos com a execução das diferentes etapas usadas na metodologia apresentada.

4.1 VERIFICAÇÃO DE DADOS ESTATÍSTICOS DE DISTRIBUIÇÃO NORMAL DAS VARIÁVEIS ENVOLVIDAS.

Foi desenvolvido um programa em linguagem BASIC e usando o Visual Basic 5.0 para se avaliar a verificação gráfica das variações das componentes de controle do processo, de modo a verificar o comportamento dos dados e suas variações durante o período de coleta.

Verificando dados estatísticos como o desvio padrão, sendo o período total, dividido e avaliado em termos de desvio geral e três períodos de igual tamanho ou representação para verificar as mudanças sofridas durante o período e as dimensões de variações de amplitudes e desvio padrão de cada variável.

A Tabela 13, mostra as variações a nível de desvio padrão das variáveis do processo e o seu nível de amplitude.

Se observa na Tabela 14, que a DQO de saída do Tanque de Metanização (TM) tem um desvio padrão que representa uma variação de até 57,74 % em relação à média, enquanto que uma variável que pode ser ajustada como o pH de alimentação no Tanque de Metanização (TM) possui desvio padrão com variação de apenas 2,56% em relação a sua média. Logo, variáveis com dependência maior de outras variáveis e de difícil controle, sofrerão mais variações que as menos dependentes e de controle ajustáveis.

Assim, é de se esperar que os resultados dos erros relativos dos métodos utilizados para previsão, também oscilem em função da variável alvo. No caso a variável alvo ´que se deseja obter a previsão é a DQO de saída, um dia a frente que possui um desvio padrão de 57,74%. Esta variável alvo tem grande oscilação e se espera resultados de previsão inferiores ao seu desvio padrão.

A DQO de saída sofre influência da variação do volume de produção e do tipo de refrigerante que esta sendo processado, além de outras variações resultante de métodos de controle operacional da estação e métodos analíticos, envolvendo o resultado desta variável.

Tabela 13- Dados estatísticos das variáveis de controle do Tanque de Metanização (TM).

Variáveis	Média Geral	Desvio Padrão	Valor Máximo	Valor Mínimo
DQO_EEQ(mg.L^{-1})	3769,04	929,53	7240,00	1056,00
Vazão_TM(m^3.h^{-1})	47,60	21,30	106,67	0,00
Carga Orgânica Aplicada_TM (Kg DQO.d^{-1})	4314,19	2199,28	12910,70	0,00
pH de alimentação do TM	7,35	0,19	8,36	6,83
Alcalinidade do TM no ponto2 (mg.L-1)	1089,68	258,36	1893,00	603,33
AGV no ponto 2 do TM (mg.L-1)	160,66	108,91	609,66	29,33
Relação AGV x alcalinidade TM	0,15	0,10	0,64	2,96
pH do ponto 2 do TM	6,77	0,18	7,23	6,11
Temperatura do Reator TM (ºC)	35,25	2,08	41,32	30,50
DQO de saida do TM (mgO_2.L^{-1})	474,77	274,13	1640,00	75,00

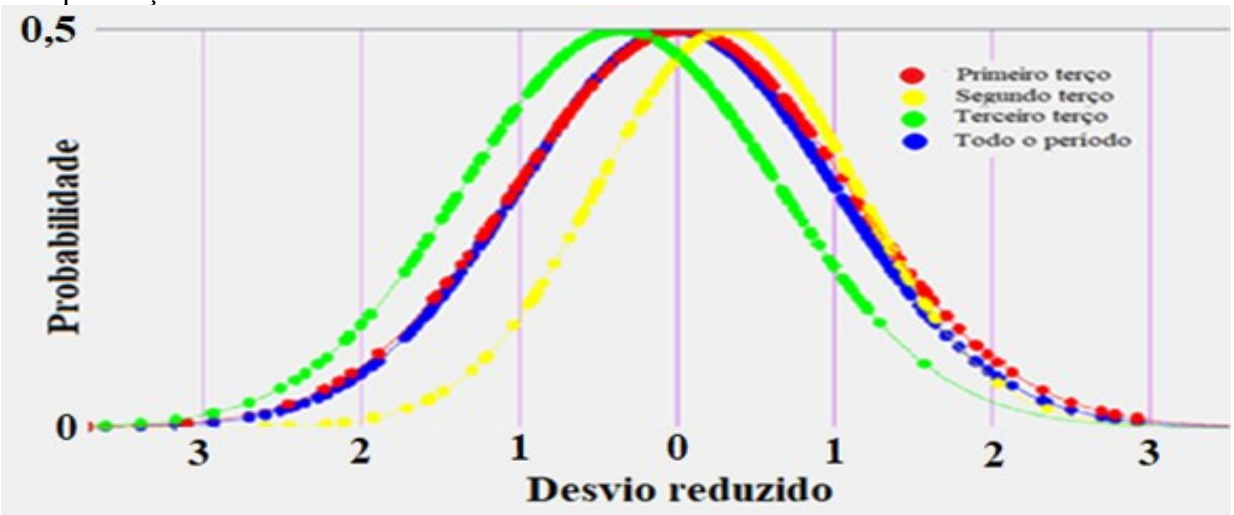
Tabela 14- Percentual do desvio padrão em relação à média

Variáveis	Média Geral	Desvio Padrão	% Desvio
DQO_EEQ(mg.L^{-1})	3769,04	929,53	24,66
Vazão_TM(m^3.h^{-1})	47,60	21,30	44,75
Carga Orgânica Aplicada_TM (Kg DQO.d^{-1})	4314,19	2199,28	50,98
pH de alimentação do TM	7,35	0,19	2,56
Alcalinidade do TM no ponto2 (mg.L-1)	1089,68	258,36	23,71
AGV no ponto 2 do TM (mg.L-1)	160,66	108,91	67,79
Relação AGV x alcalinidade TM	0,15	0,10	67,79
pH do ponto 2 do TM	6,77	0,18	2,70
Temperatura do Reator TM (ºC)	35,25	2,08	5,90
DQO de saida do TM (mgO_2.L^{-1})	474,77	274,13	57,74

Para se verificar a distribuição normal destes dados e mudanças durante o período de avaliação, usou-se de medir o desvio padrão reduzido das variáveis (Equação 2.107), dividiu-se os dados coletados de dois anos, em três períodos de oito meses, ou seja, primeiro terço, segundo terço e terceiro terço e ainda todo o período (linhas vermelha, amarela, verde e azul respectivamente, da Figura 47 a 56). Verificou-se assim, as mudanças no processo, a nível de distribuição normal dos dados e mudanças na média do período, cada divisão, em traços azuis nas Figuras, corresponde a um desvio padrão de todo o período.

Observa-se na Figura 47, que a DQO equalizada, se manteve dentro padrões de variação, se mantendo a média e as amplitudes, durante todo o período, apesar desta variável não ser controlada a equalização oferece uma boa mistura considerando as variações do processo.

Figura 47-Tela do software mostrando as variações, em distribuição normal, da DQO do tanque de equalização.

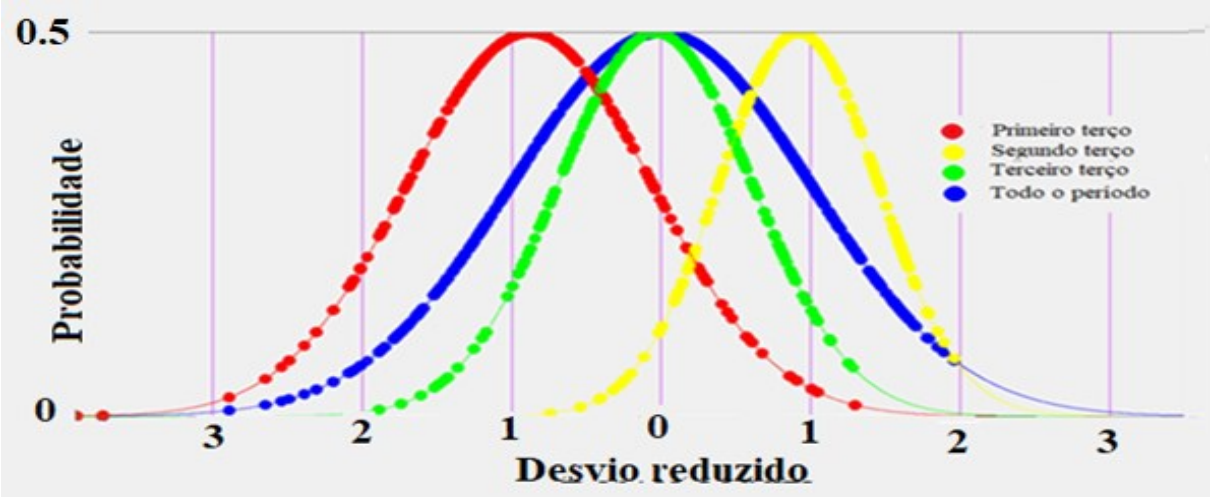


A Figura 48, mostra que a vazão teve variações na média durante o período, estes ajustes são efetuados, objetivando se maximizar a eficiência de remoção de carga biodegradável, na estação de tratamento de efluentes.

Figura 48-Tela do software, mostrando as variações da distribuição normal, da vazão de entrada do Tanque de Metanização (TM).

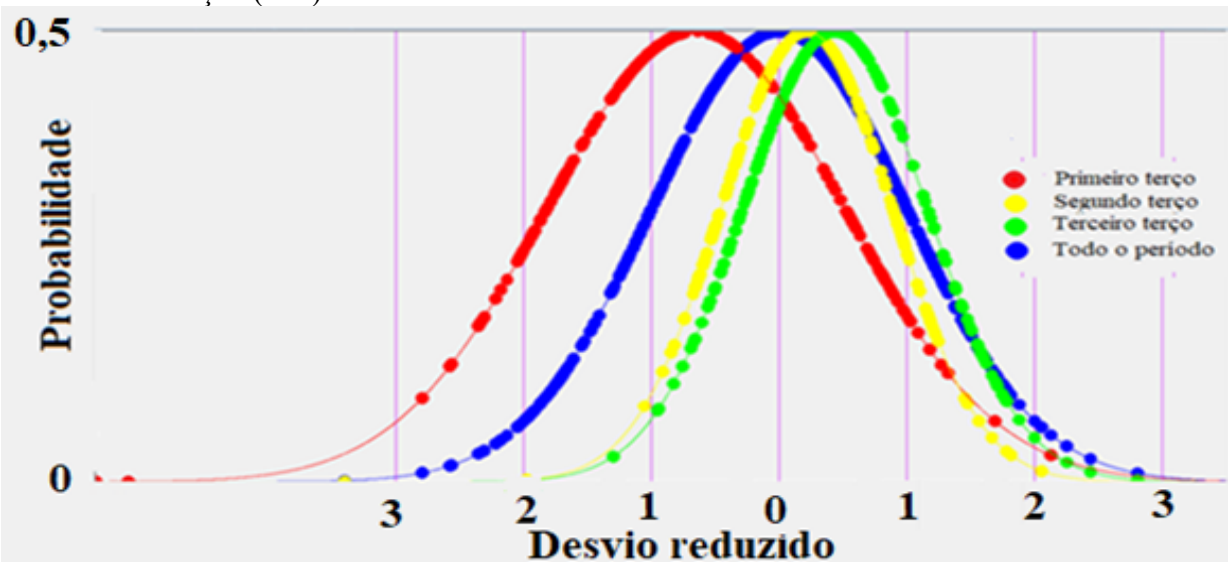
Observa-se na Figura 49, que no período que a carga orgânica aplicada acompanha as oscilações de vazão, de forma que os ajustes permitam o melhor desempenho da ETE, na remoção da carga orgânica.

Figura 49-Tela do software mostrando as variações da carga orgânica aplicada no Tanque de Metanização (TM).



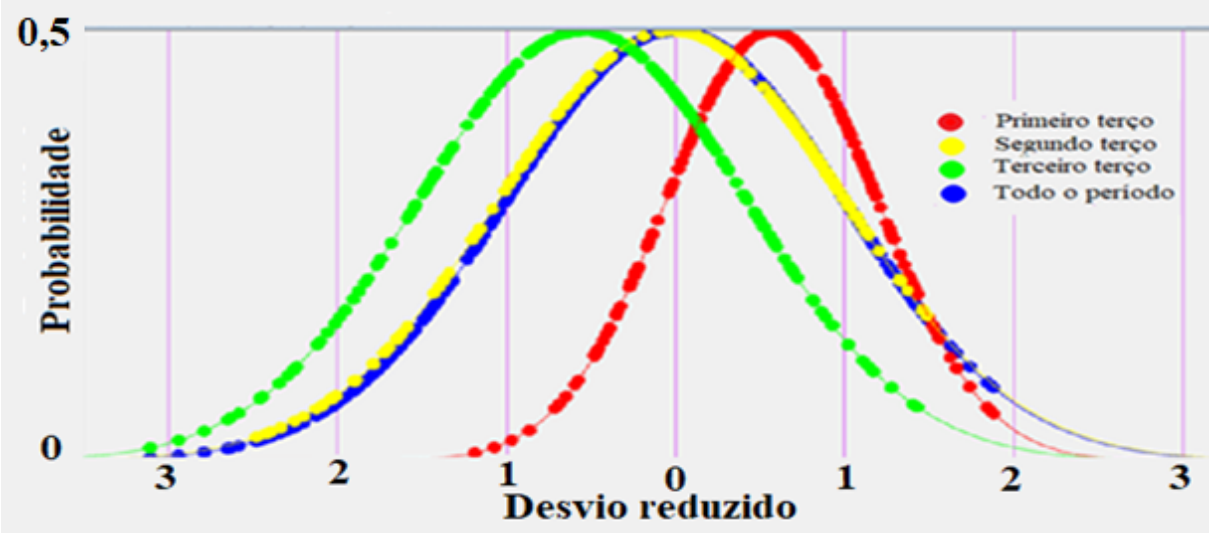
Observa-se na Figura 50, que variações do pH foram realizadas durante o período, permitindo um controle e ajustes de etapas fundamentais entre as fases de acidificação e metanização. O estreitamento da amplitude mostra que houve necessidades de ajustes no primeiro terço do período e maior estabilidade do pH de entrada à medida que o processo se desenvolveu.

Figura 50-Tela do software mostrando as variações do pH de entrada do Tanque de Metanização (TM).



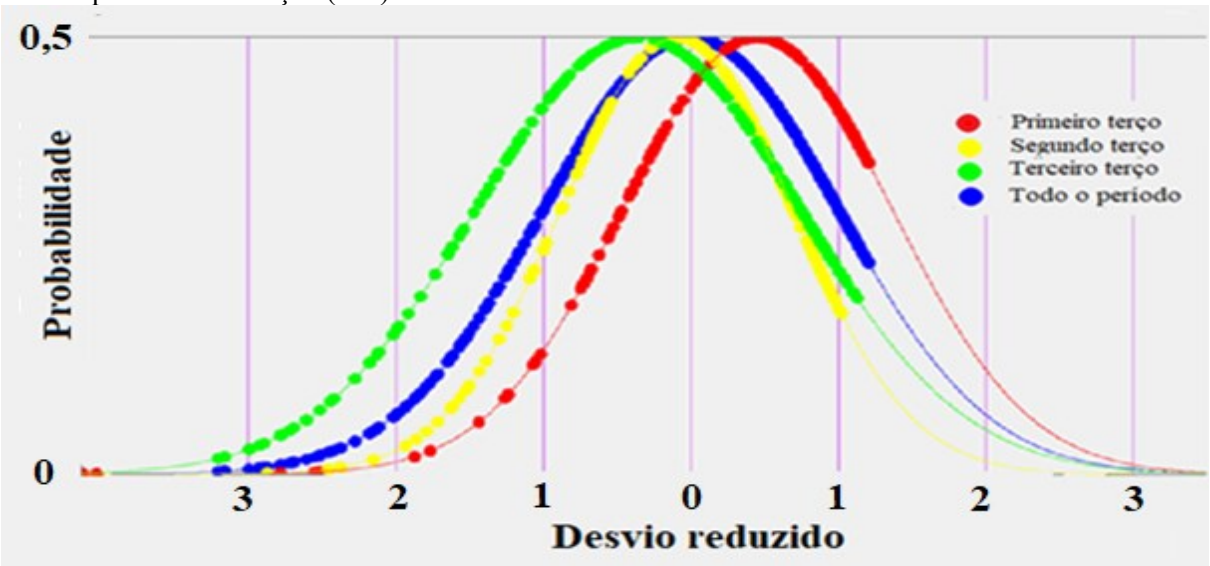
Observa-se na Figura 51, que houve variação na média menor do que um desvio padrão e aumento da amplitude para alcalinidade no ponto intermediário da manta de lodo (ponto de coleta 2)

Figura 51-Tela do software mostrando as variações na distribuição normal, da Alcalinidade no ponto 2 do Tanque de Metanização (TM)



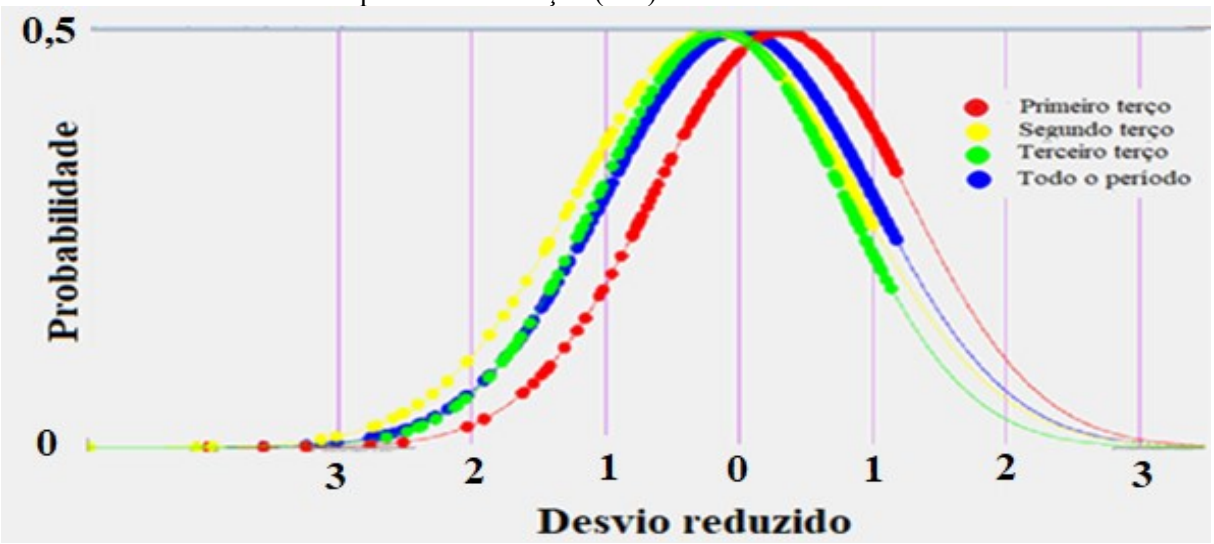
Observa-se na Figura 52, o teor de ácidos graxos voláteis sofreu alterações menor um desvio padrão na média, sofrendo pouca alteração na amplitude.

Figura 52-Tela do software mostrando as variações na distribuição normal, dos ácidos graxos voláteis no Tanque de Metanização (TM).



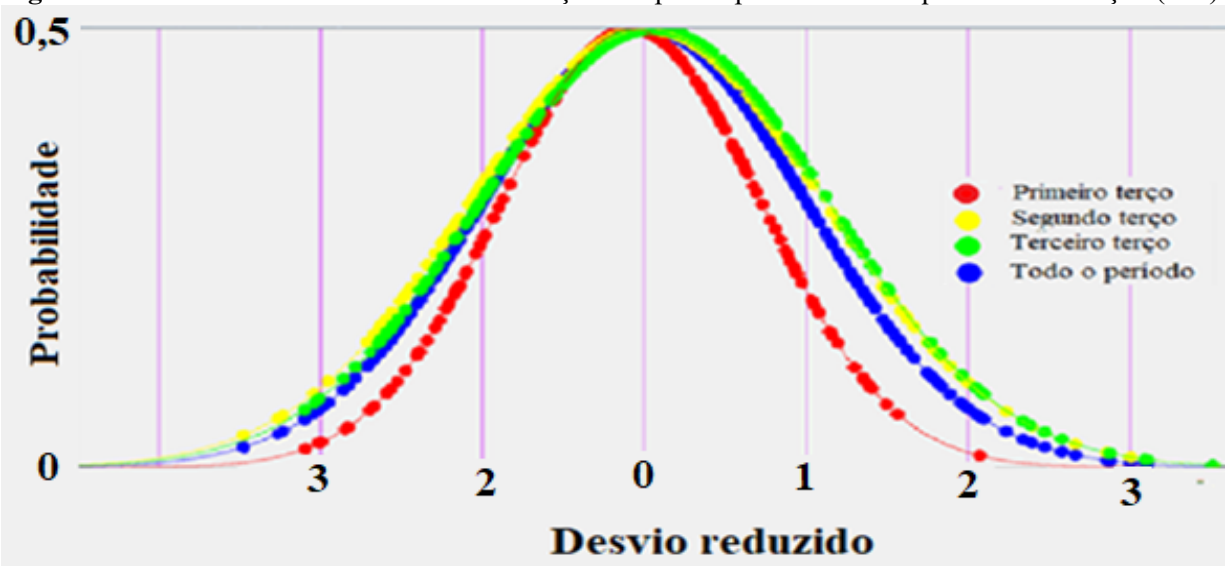
Observa-se na Figura 53, que apesar das variações da alcalinidade e dos ácidos graxos voláteis no processo, a relação entre estas duas variáveis foi mantida de forma a apresentar muito pouca variação durante o período.

Figura 53-Tela do software mostrando as variações na distribuição normal da relação ácidos graxos voláteis e alcalinidade do Tanque de Metanização (TM)



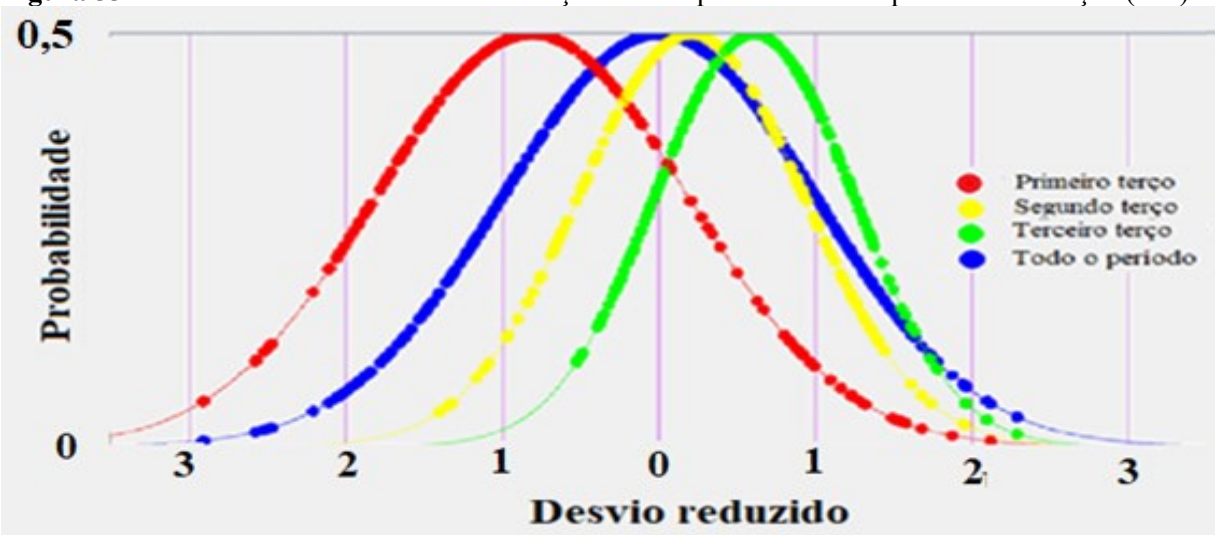
Observa-se na Figura 54, que apesar das variações de carga orgânica, vazão e pH ajustado na entrada o pH do ponto intermediário da camada de lodo (ponto 2) sofreu muito pouca variação na média, aumentando um pouco sua amplitude no final do período, demostrando um bom controle sobre esta variável.

Figura 54-Tela do software mostrando as variações do pH do ponto 2 no Tanque de Metanização (TM)



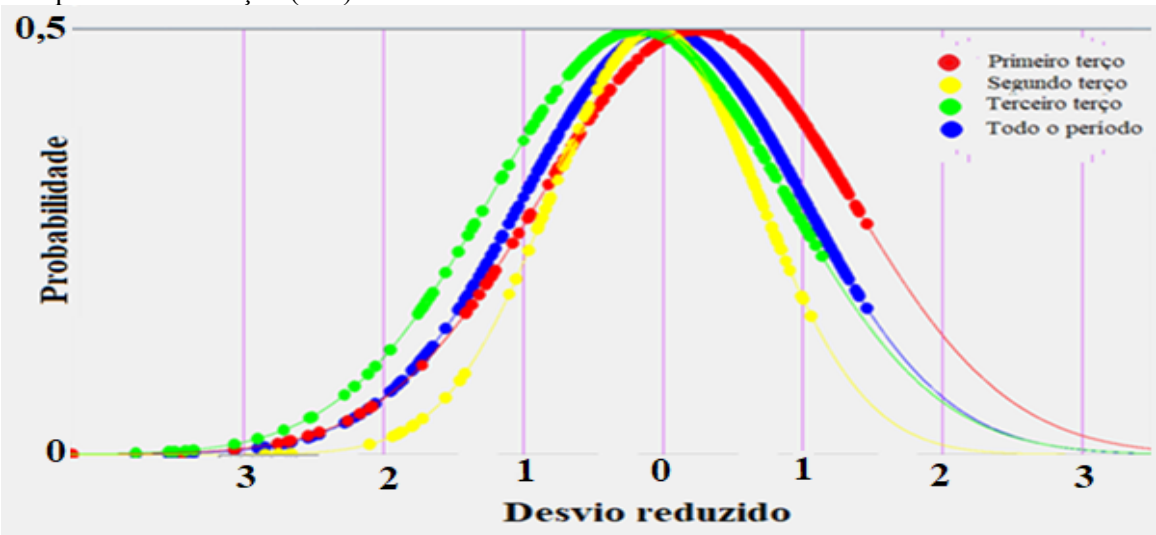
Observa-se na Figura 55, que a temperatura do Tanque de Metanização (TM) variou durante o período de ajuste do processo, as variações de carga, vazão e temperatura ambiente, afetam esta variável.

Figura 55-Tela do software mostrando as variações da temperatura no Tanque de Metanização (TM)



Observa-se na Figura 56, que a DQO de saída do Tanque de Metanização (TM), mesmo com todas as variações dos elementos que afetam seu desempenho na entrada, procura manter resultados de média e desvio padrão não muito defasados e próximos a uma média de eficiência de remoção, pois se trabalha ajustando as variáveis de controle, com objetivo de estar sempre buscando o melhor resultado possível de remoção de DQO. Estes ajustes fazem com que a série temporal da DQO de saída, oscile em torno de uma média de eficiência da ETE, tendendo a uma estacionariedade de resposta.

Figura 56-Tela do software, mostrando as variações da distribuição normal, da DQO de saída do Tanque de Metanização (TM)



4.2 VERIFICAÇÃO DA CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS UM PASSO À FRENTE E OS DIAS ANTERIORES

A Tabela 15, mostra os resultados do software desenvolvido em Visual Basic para se verificar as correlações existentes entre a variável medida um passo à frente e os valores medidos das outras variáveis. Nos dados da DQO de saída do Tanque de Metanização (TM) um passo à frente foi observado uma correlação de 81%, entre os Ácidos Graxos Voláteis (AGV) do Tanque de Metanização (TM) e a própria DQO de saída do Tanque de Metanização (TM) do dia anterior e de 82% entre os valores da DQO um dia antes e um dia depois.

Tabela 15 - Correlação de Pearson (R), da DQO de saída um passo à frente (um dia à frente) e as outras variáveis de controle

Variável (I+0)	R (DQO de saida do TM ($mgO_2.L^{-1}$) (I+1))
DQO_EEQ($mg.L^{-1}$)	0,27
Vazão_TM($m^3.h^{-1}$)	-0,12
Carga Orgânica Aplicada_TM (Kg DQO.d^{-1})	0,06
pH de alimentação do TM	0,08
Alcalinidade do TM no ponto2 (mg.L-1)	0,09
AGV no ponto 2 do TM (mg.L-1)	**0,81**
Relação AGV x alcalinidade TM	**0,80**
pH do ponto 2 do TM	-0,43
Temperatura do Reator TM	-0,19
DQO de saida do TM ($mgO_2.L^{-1}$)	**0,82**

Nota: O prefixo (I-0), se refere ao dia presente e o prefixo (I+1) ao dia futuro

A Tabela 16, mostra correlações existentes entre a variável medida um passo à frente (um dia à frente) e os valores medidos nos dias anteriores, neste caso avalia-se a DQO de saída do Tanque de Metanização (TM). Observa-se que as maiores correlações estão nos valores das variáveis até dois dias antes, nesta pesquisa foram verificados efeitos combinados até três dias antes.

Tabela 16 - Tela do software, correlação de Pearson da DQO de saída um passo à frente e a mesma variável nos 5 dias anterio

Váriáveis	R (DQO de saida do TM ($mgO_2.L^{-1}$) (I+1))
DQO de saida do TM ($mgO_2.L^{-1}$) (I-0)	**0,81**
DQO de saida do TM ($mgO_2.L^{-1}$) (I-1)	**0,68**
DQO de saida do TM ($mgO_2.L^{-1}$) (I-2)	0,58
DQO de saida do TM ($mgO_2.L^{-1}$) (I-3)	0,51
DQO de saida do TM ($mgO_2.L^{-1}$) (I-4)	0,49
DQO de saida do TM ($mgO_2.L^{-1}$) (I-5)	0,48

Nota: O prefixo (I-0), se refere ao dia presente e o prefixo (I+1) ao dia futuro

O uso de dois dias antes tem maior correlação com a variável futura, assim também, como os Ácidos Graxos Voláteis (AGV) no ponto 2, afetam com mais representatividade a DQO de saída.

4.3 PRÉ-TRATAMENTO DE DADOS

4.3.1 Preenchimento dos dados faltantes nas séries temporais

As Figura 57 a 66, mostram os resultados do preenchimento de dados faltantes, nas 10 variáveis de controle, usadas no Tanque de Metanização (TM).

No **Apêndice A** se encontra o programa usado para preenchimento de dados faltantes e normalização dos dados

Figura 57-Preenchimento de dados em falta, DQO de entrada do TM

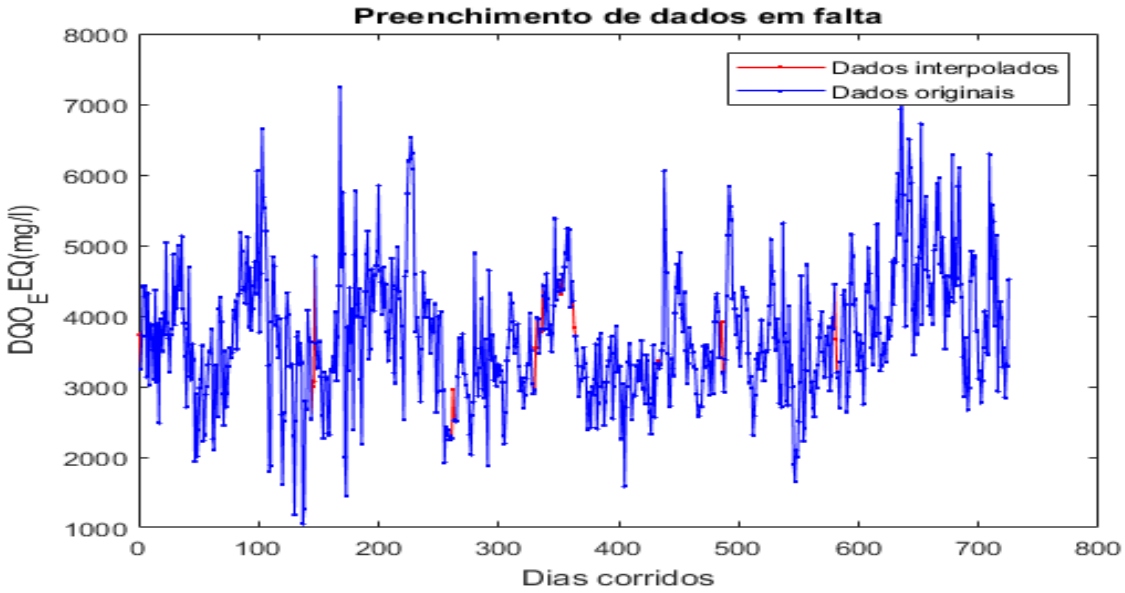


Figura 58-Preenchimento de dados em falta, vazão de entrada do TM

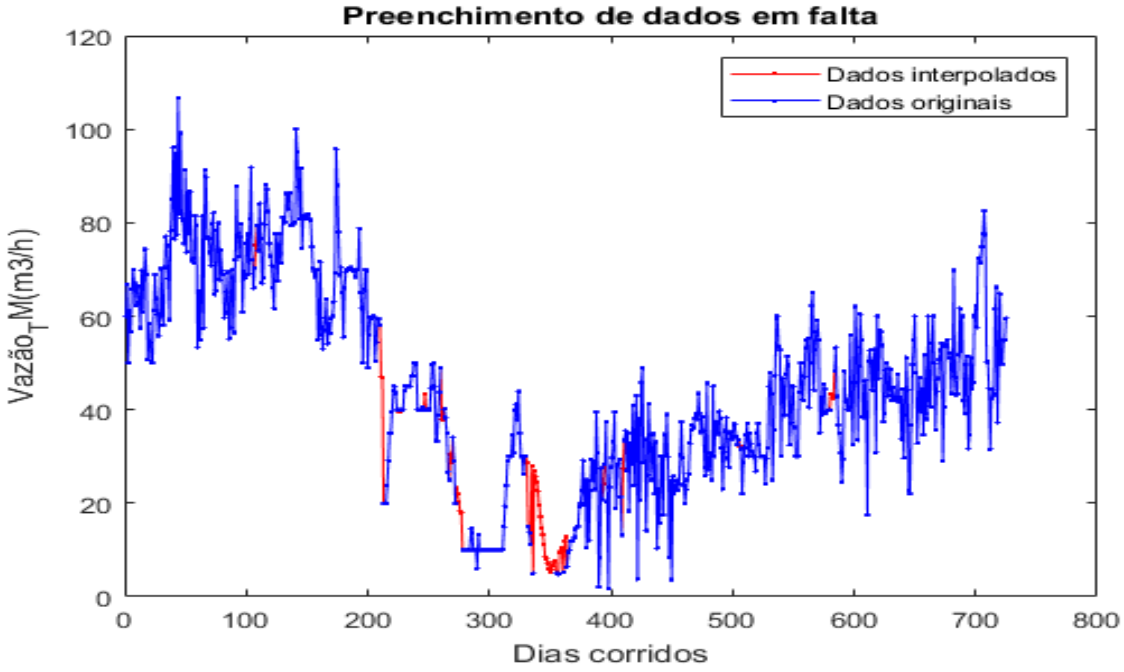


Figura 59-Preenchimento de dados em falta, carga orgânica aplicada na entrada do TM

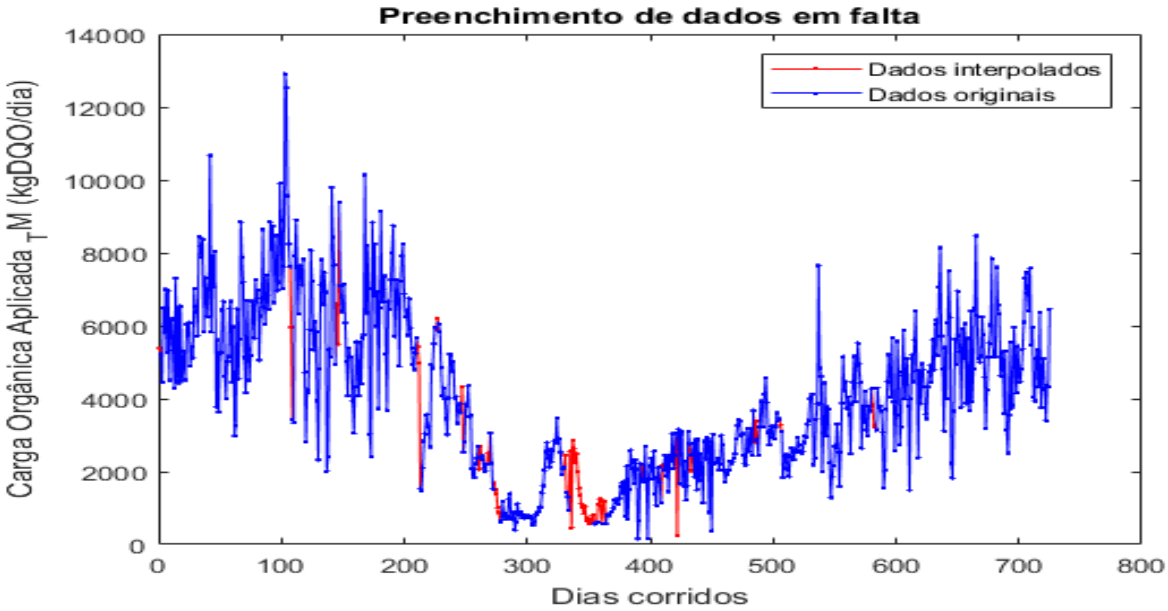


Figura 60-Preenchimento de dados em falta, pH ajustado na entrada do TM

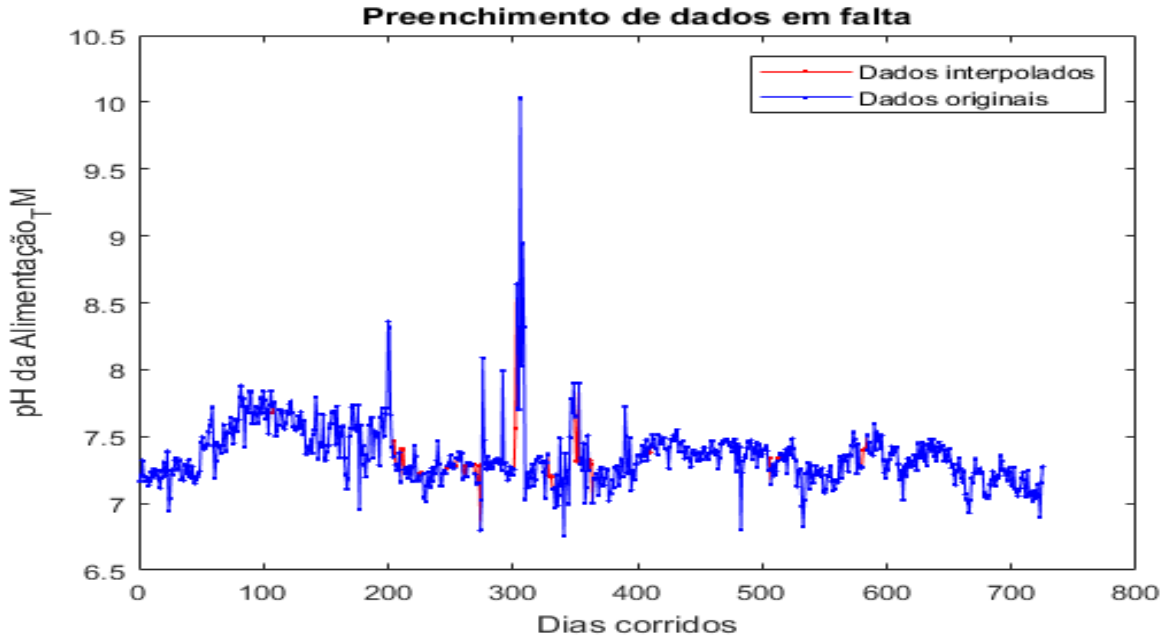


Figura 61-Preenchimento de dados em falta, ácido graxos voláteis no ponto 2 do TM

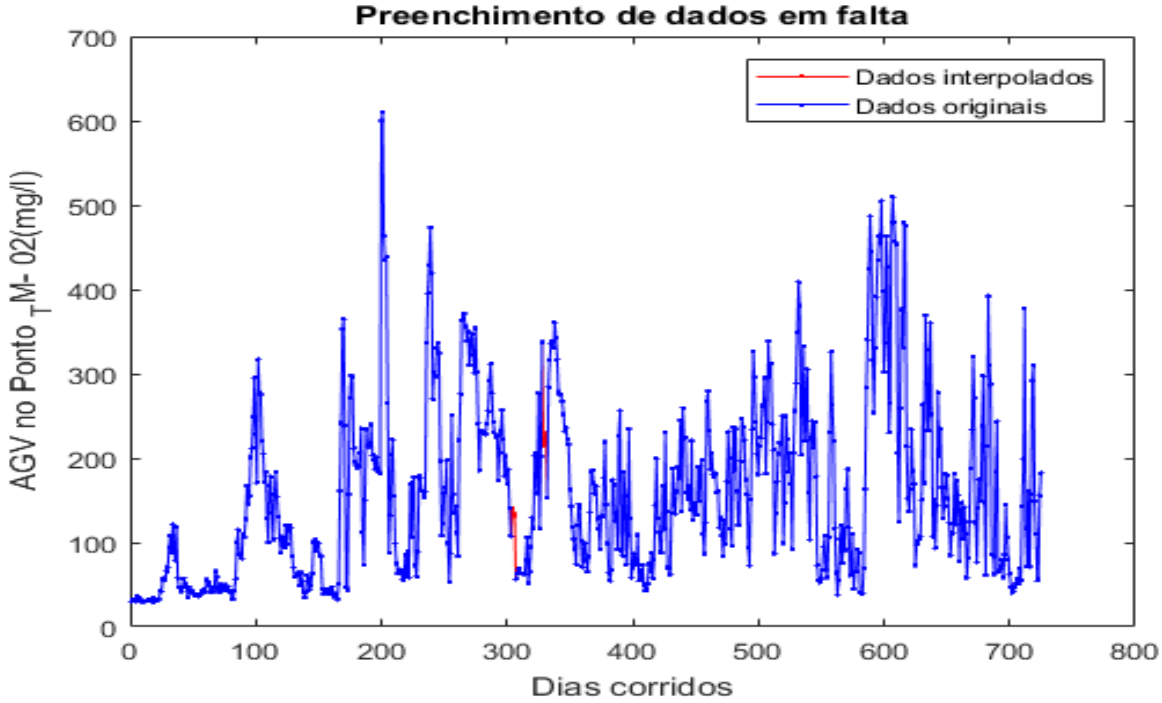


Figura 62 -Preenchimento de dados em falta, alcalinidade no ponto 2 do TM

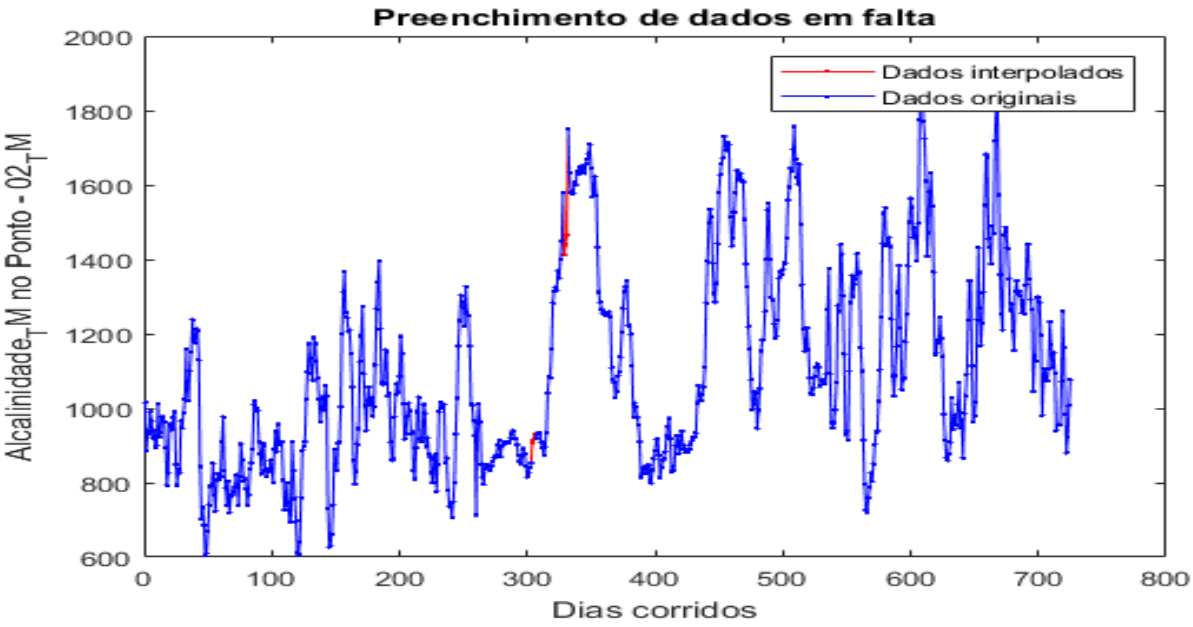


Figura 63 -Preenchimento de dados em falta, temperatura do TM

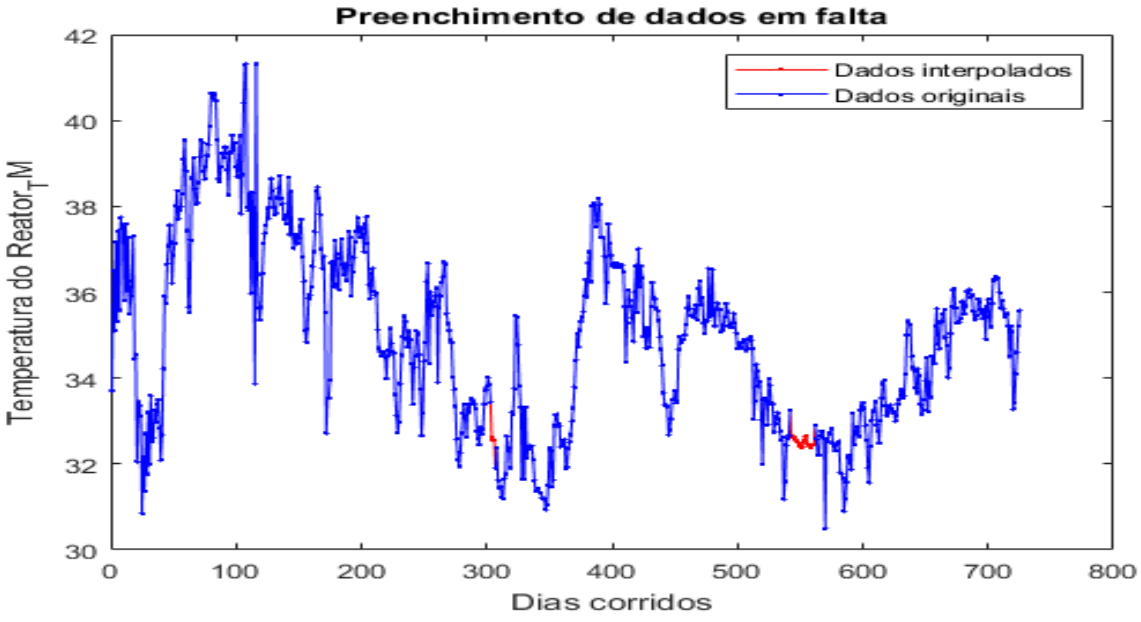


Figura 64 -Preenchimento de dados em falta, pH no ponto 2 do TM

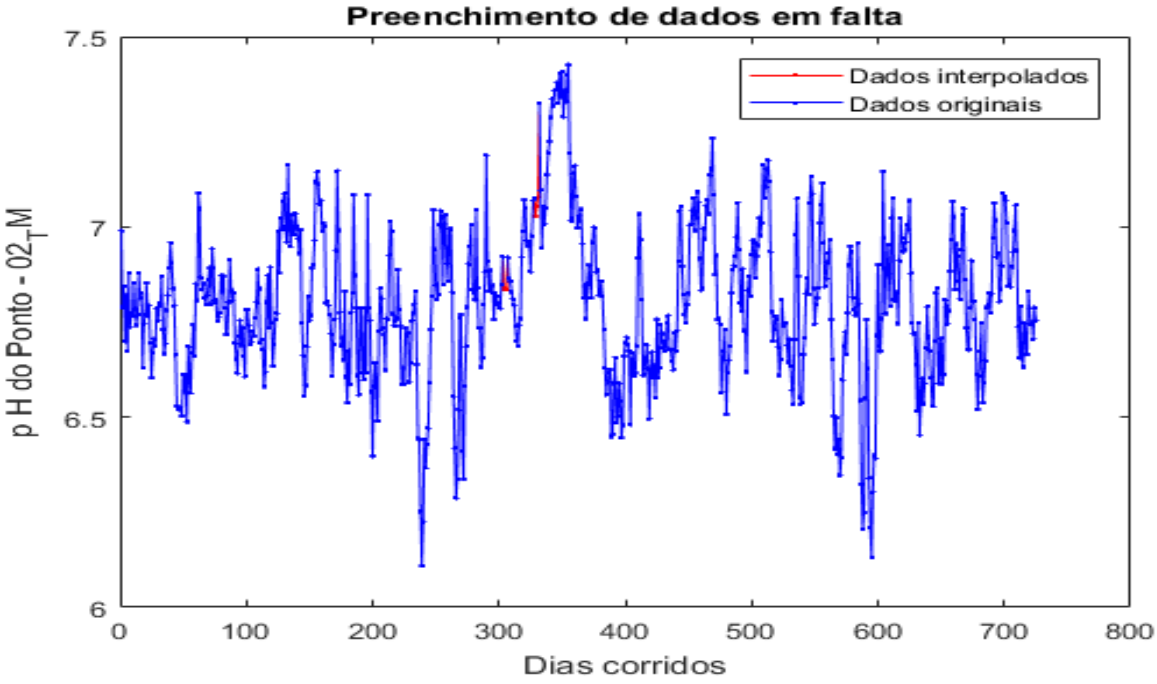


Figura 65 -Preenchimento de dados em falta, relação entre AGV x alcalinidade do TM

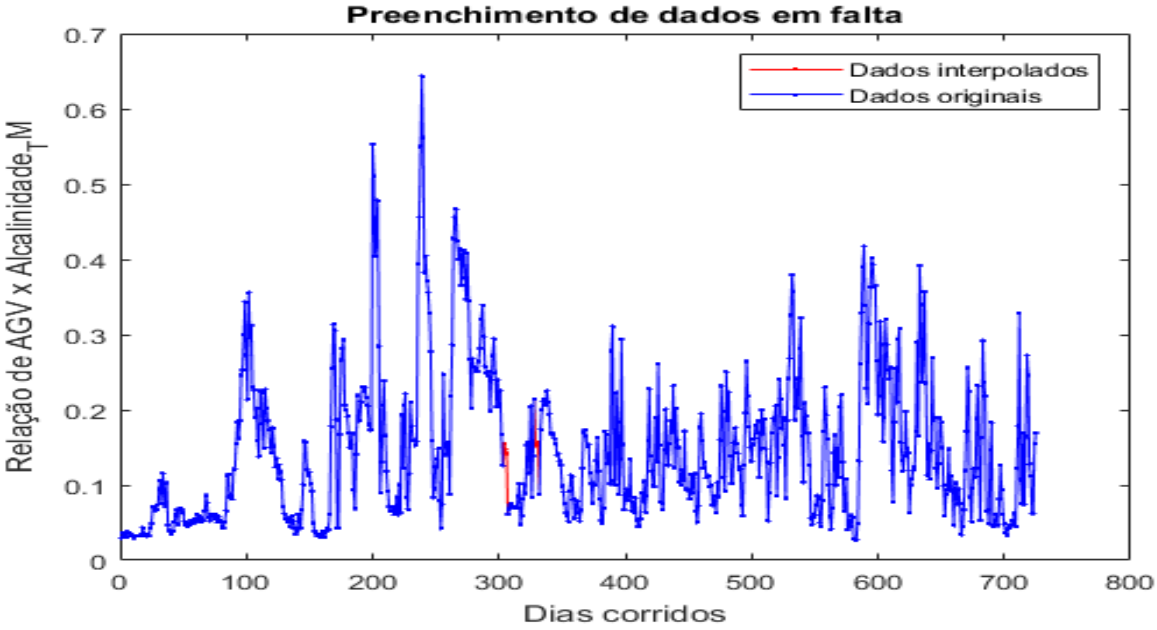
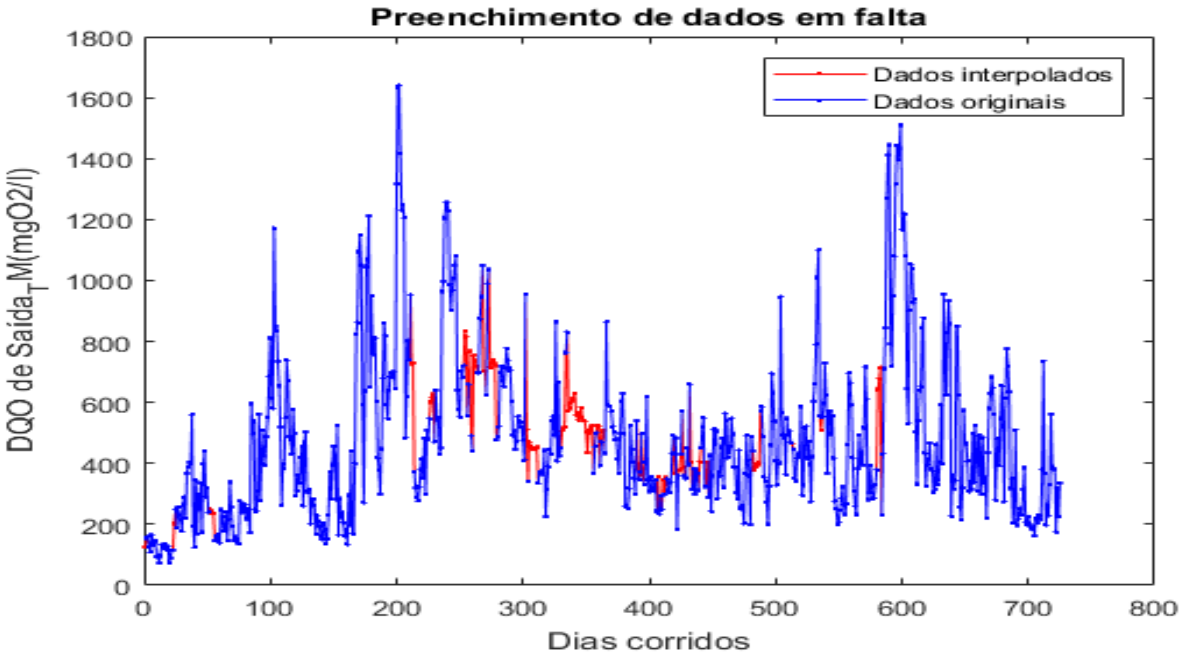


Figura 66 -Preenchimento de dados em falta, DQO de saída do TM

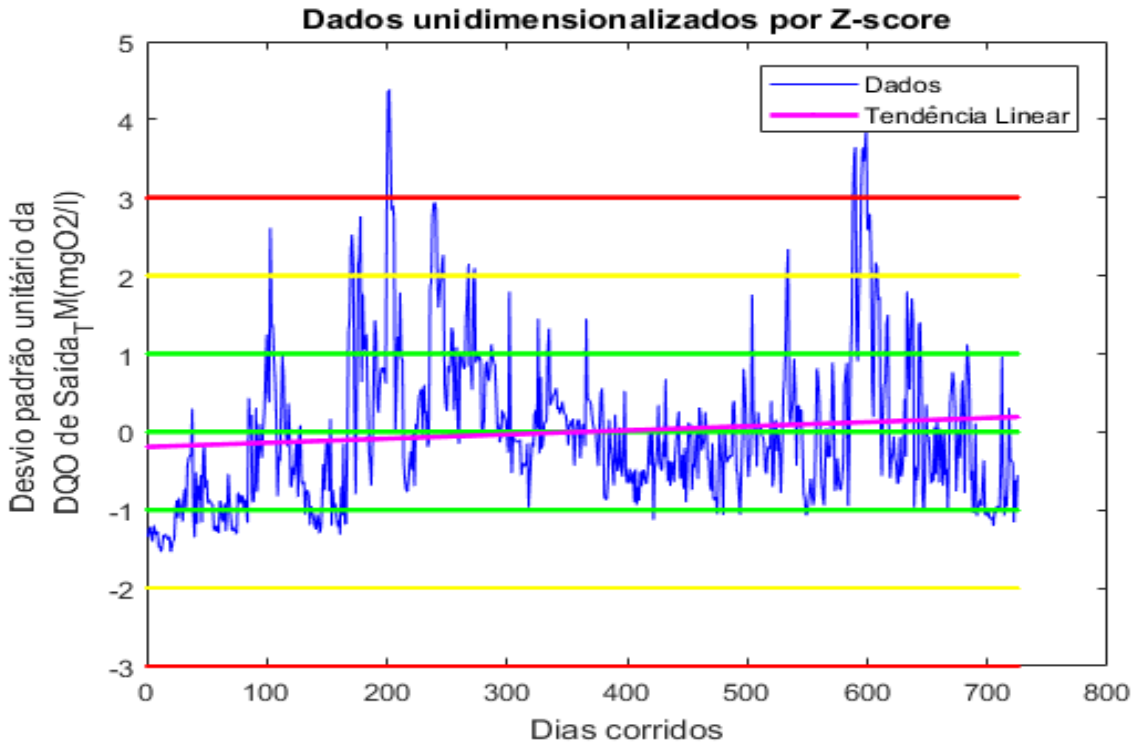


A quantidade de dados faltantes afeta a previsão, sendo mais um fator do aumento de erro na previsão se estes forem excessivos.

4.3.2 Normalização dos dados, para trabalho com multivariáveis

Para poder se comparar os efeitos de uma variável sobre as outras em métodos multivariados, todos os dados foram unidimensionalizados, assim a influência da variação de um pH que é em escala logarítmica, ou a temperatura ficaria sendo comparada com qualquer outra variável do processo. Para isto, os dados são transformados em números unitários de desvios padrões (método Z-score), conforme descrito na seção 2.7.1. Na Figura 67, temos a exemplo a normalização da DQO de saída do reator anaeróbio, mostrando que apesar da tendência da DQO de saída em aumentar, a média se mantem como zero, nas linhas verdes valores até um desvio padrão, na amarela dois desvios padrões e vermelha três desvios padrões. O programa usado no preenchimento dos dados faltantes e na normalização dos dados está no **Apêndice A** .

Figura 67 -Variável DQO de saída do Tanque de Metanização (TM) unidimensionalizadas pelo método z-score (desvio padrão)



4.3.3 Teste de estacionariedade de Dickey-Fuller

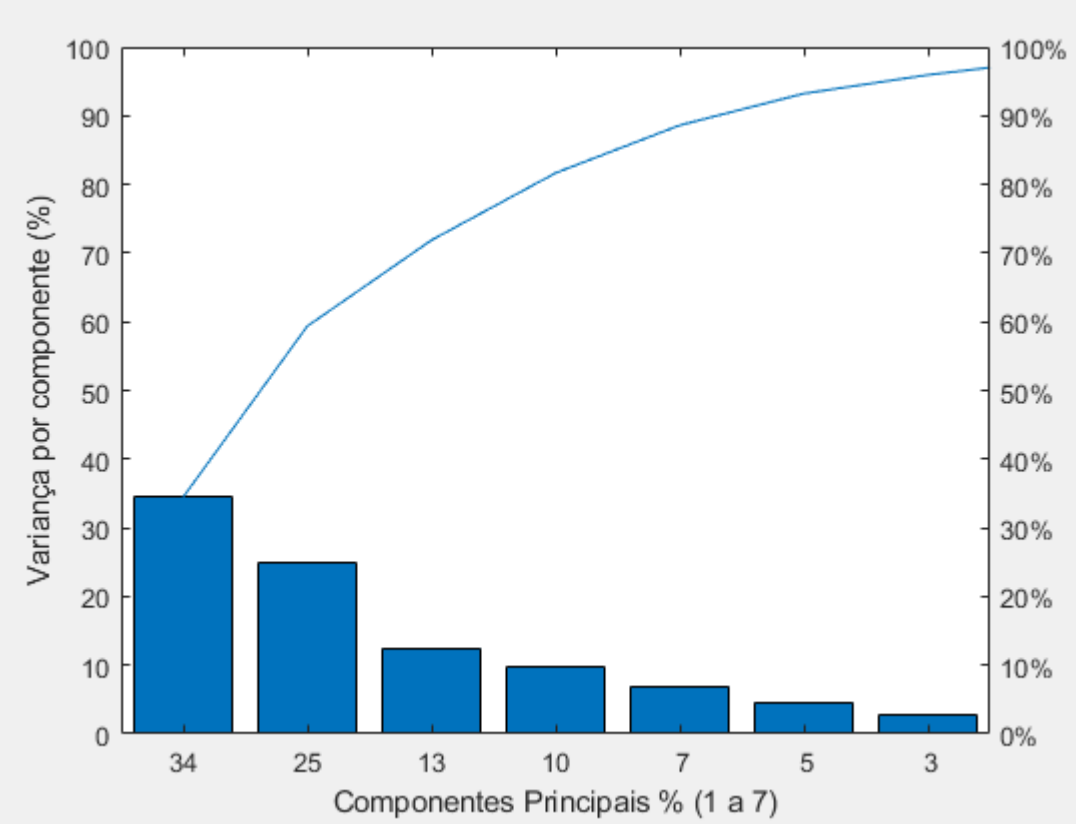
Para se verificar se a DQO de saída é uma série estacionária ou não, foi realizado o teste de Dickey-Fuller aumentado, conforme descrito na seção 2.4.2.1, usando o *software Gretl* aplicado a variável DQO de saída, obtendo os dados que se encontram no **Apêndice G**. Ao qual concluímos que mesmo para teste com constante e com constante e tendência, apresenta um valor para estatística de teste: τ (tau) de -4,555843 e -4,5327 (valores abaixo de -2,58 e -3,43 tabelados para significância alfa de 0,01% na Tabela 3), indicando a rejeição da hipótese de raiz unitária, definindo a série como estacionária.

O resultado da estacionariedade da série é esperado pois as variáveis são ajustadas para a máxima eficiência de remoção da DQO no Tanque de Metanização (TM).

4.4 CLASSIFICAÇÃO E AVALIAÇÃO DOS COMPONENTES PRINCIPAIS DOS DADOS OBTIDOS.

Para se verificar a influência de uma variável sobre a outra, e a interferência destas variáveis no resultado da estação de tratamento de efluente, foi feito uma Análise de Componentes Principais (ACP), como mostram as Figuras 69,70 e 71, com uma classificação nos dados da DQO de saída do Tanque de Metanização (TM), de maneira a identificar resultados maiores que um desvio padrão da média em cor vermelha, dados entre a média e um desvio padrão na cor amarela e dados abaixo da média mais que um desvio padrão na cor verde. A análise das variâncias dos componentes principais está na Figura 68.

Figura 68 -Variância dos sete primeiros componentes principais



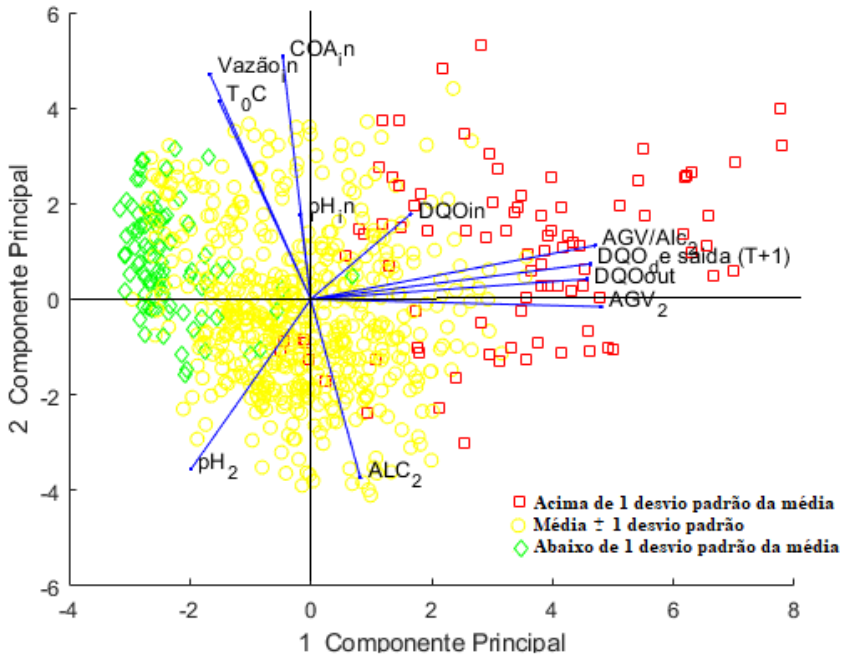
Na Figura 68, se observa que as três primeiras componentes respondem por 72% da variança do processo. A análise dos pares da primeira componente e da segunda componente (Figura 69), mostra uma separação clara das classes, assim como também a análise da primeira componente e da terceira componente (Figura 70), porém a análise da segunda componente e da terceira não oferece boa visualização da separação de classes (Figura 71).

Observa-se nas Figuras 68 e 69, que existem combinações claras, de faixas de resultados para DQO de saída conforme valores de suas variáveis, especialmente a carga orgânica de entrada (COA), a relação de Ácidos Graxos Voláteis com a alcalinidade (AGV/ALc), e os Ácidos Graxos Voláteis (AGV),elevam o resultado da DQO de saída (DQOout), conforme sobem, e o pH do ponto 1 e 2, quanto mais elevado tendem a um melhor resultado da estação.

A alcalinidade acompanha os valores de estabilidade da estação dentro da faixa de variação da planta, assim como a vazão de entrada que é controlada para manter uma carga orgânica aplicada mais uniforme.

O pH da entrada (pHin) e a carga orgânica aplicada (COAin), são variáveis controladas na ETE.

Figura 69 -Análise de componentes principais, usando todas as entradas, componente1 x componente2



Assim, o controle das fases de acidificação e metanização, se tornam essencial, para o bom resultado da estação. Esta ferramenta é importantíssima na resposta a perguntas de como se comportaria o resultado da ETE se aumentasse ou diminuísse alguma variável do sistema, assim também, como verificar se a estação está sendo bem operada.

Figura 70 -Análise de componentes principais, usando todas as entradas, componente1 x componente3

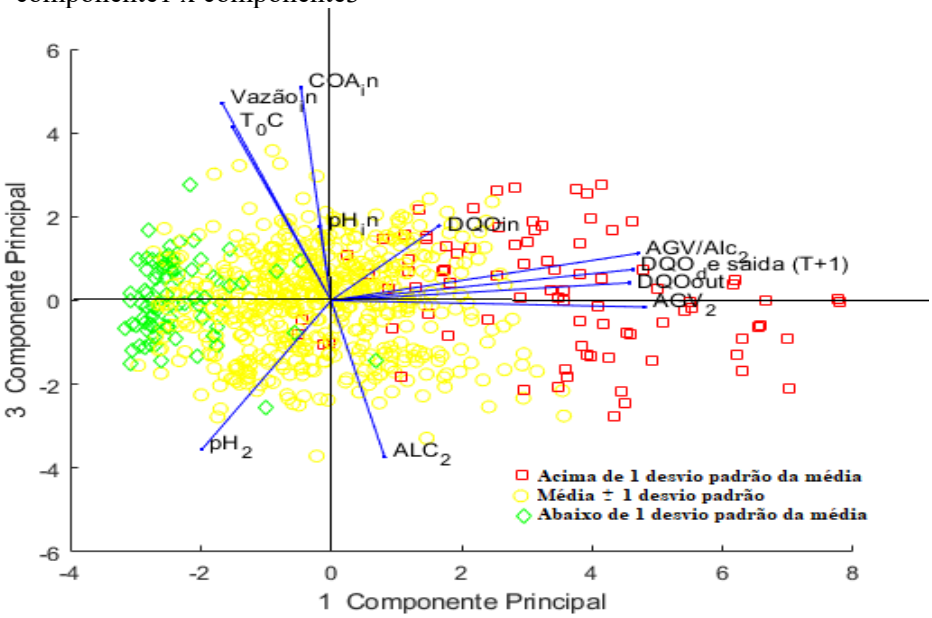
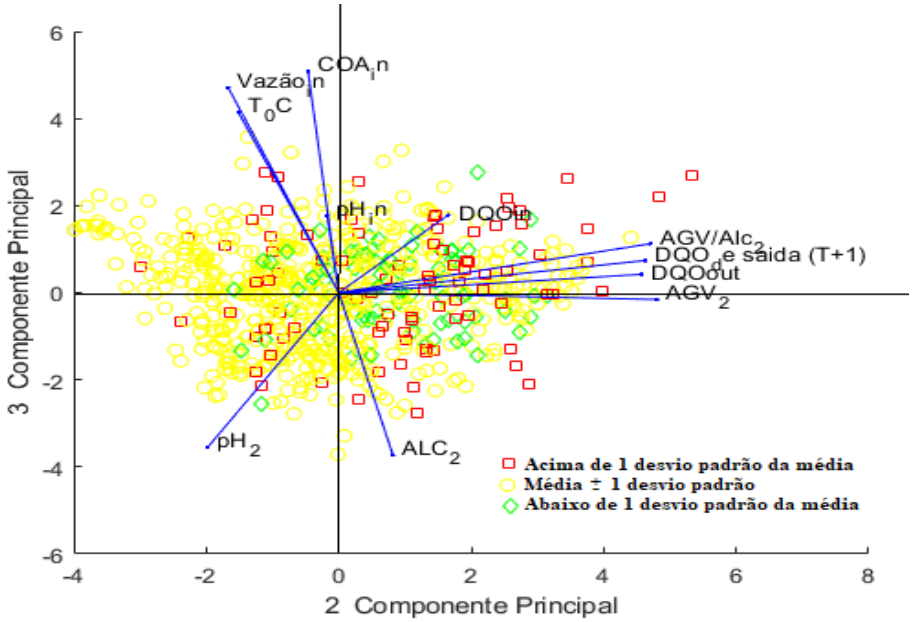


Figura 71 -Análise de componentes principais, usando todas as entradas, Componente2 x componente3



Esta ferramenta possibilita com a classificação realizada, se ter noção do comportamento da DQO de saída com as mudanças das demais variáveis do processo, permitindo ao operador por exemplo, visualizar o efeito do aumento da DQO na entrada como afeta os resultados levando a variável para regiões acima da média ou para zona vermelha do gráfico

4.5 PREVISÃO UNIVARIADA UM PASSO À FRENTE DA DQO DE SAIDA.

A decisão pelo método teve como base nas correlações entre os valores futuros Z_{t+1} na série e os valores passados $Z_t, Z_{t-1}, Z_{t-2}, Z_{t-3}$,...., para efeito de comparação, foi usado em todos os métodos a DQO de saída, cuja variação normal no período teve um desvio padrão que representa 57,7% de variação relativa de todo o período.

Foram utilizados na primeira previsão 695 dos 725 dados diários obtidos, ficando separados 30 dados do final da série independente do treinamento e dos dados usados pela rede neural, para comparação dos resultados, assim todos os dados estarão sujeitos a mesma condição do teste. Segundo Luchesa e Chaves Neto (2018) com esta amostragem se calcula um nível de confiança próximo de 76%, usando a Equação 4.5.1, com um erro máximo da estimativa de 30% (variação observada nos testes), para verificação do método, neste período (ultimo terço da série) a DQO de saída do Tanque de Metanização (TM) estava com um desvio padrão de 291,55 $mg.L^{-1}$, para uma média de DQO de 522,49 $mg.L^{-1}$, representando uma variança relativa de 55,8% do terceiro terço do período (isto em função da DQO de saída sofrer varições com mudança de produção, variações de processo e de métodos analíticos) . Uma amostragem maior, reduziria a quantidade de dados necessárias para treinamento da rede. A análise estatística prévia é importante para se saber, se os resultados dos métodos estão atingindo faixas dentro de um desvio padrão, quanto menor esta faixa, mais preciso será o método.

$$n_i = \frac{N_d.Z^2.p.(1-p)}{(N_d-1).\varepsilon^2+Z^2.p.(1-p)} \qquad (4.5.1)$$

Sendo:

n_i Número de indivíduos na amostra

Z Valor do desvio, que corresponde ao grau de confiança desejado.

p Proporção populacional de indivíduos que pertence a categoria que estamos interessados em estudar.

ε Margem de erro ou erro máximo da estimativa. Identifica a diferença máxima entre a proporção amostral e a verdadeira proporção populacional (p).

N_d número de dados que vão ser amostrados.

Nota-se que uma variável tem desvio padrão diferente da outra, assim não se pode comparar os resultados dos métodos de variáveis diferentes, isto muda até para processos ou plantas diferentes, para a mesma variável alvo.

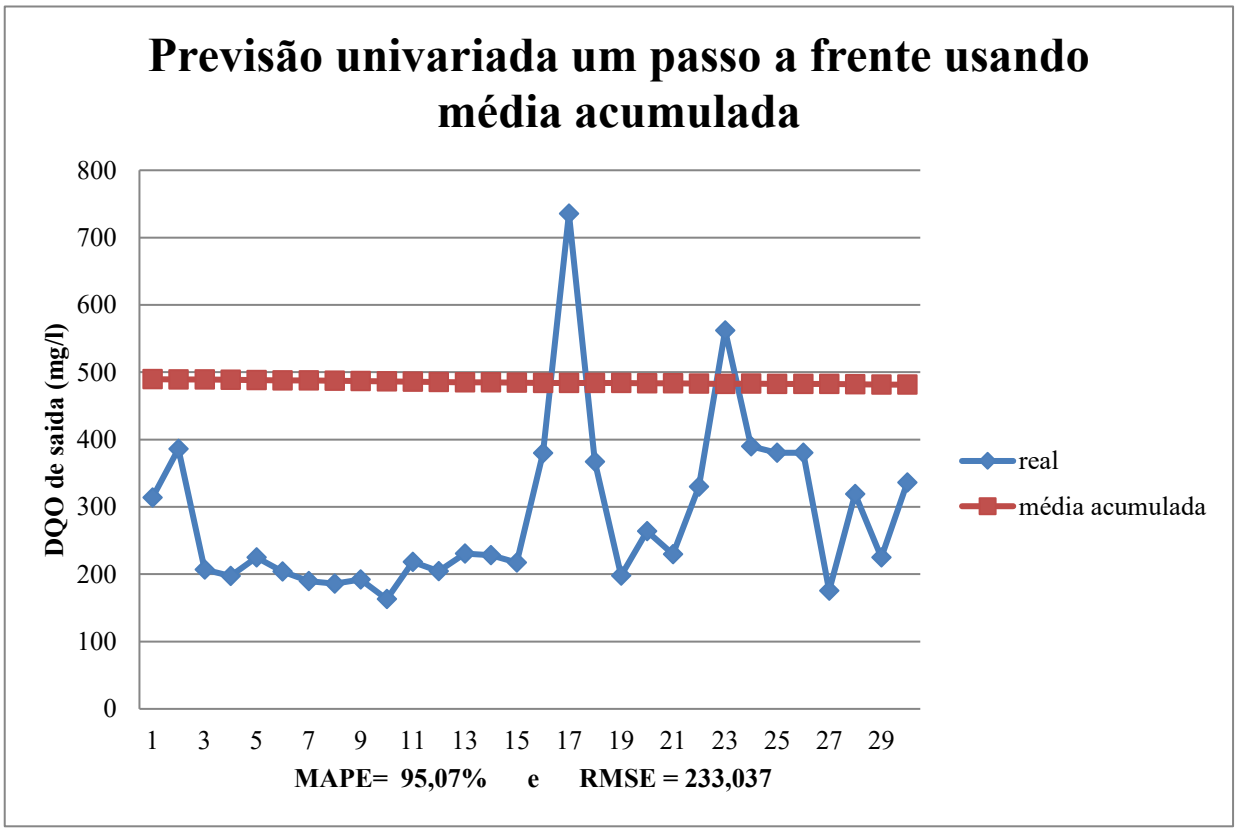
Assim que um dia futuro é calculado na série, os números usados na previsão foram incrementados por valores reais, em mais um dia, destes 30 dias remanescentes, para previsão do dia seguinte, ficando o intervalo de um dia como um passo da série a ser trabalhado. Para manipulação dos dados se usou uma planilha Excel em algumas séries e redes neurais em outras usando MATLAB.

Nas redes neurais utilizadas foi feito o treinamento da rede a cada dia durante o período de testes (retreinamento), visando uma simulação do que seria em caso prático a atualização dos dados.

4.5.1 Previsão univariada usando média acumulada

Na Figura 72, temos a previsão univariada, um passo à frente (neste trabalho, um dia à frente), usando média acumulada considerando apenas os 30 dias de testes.

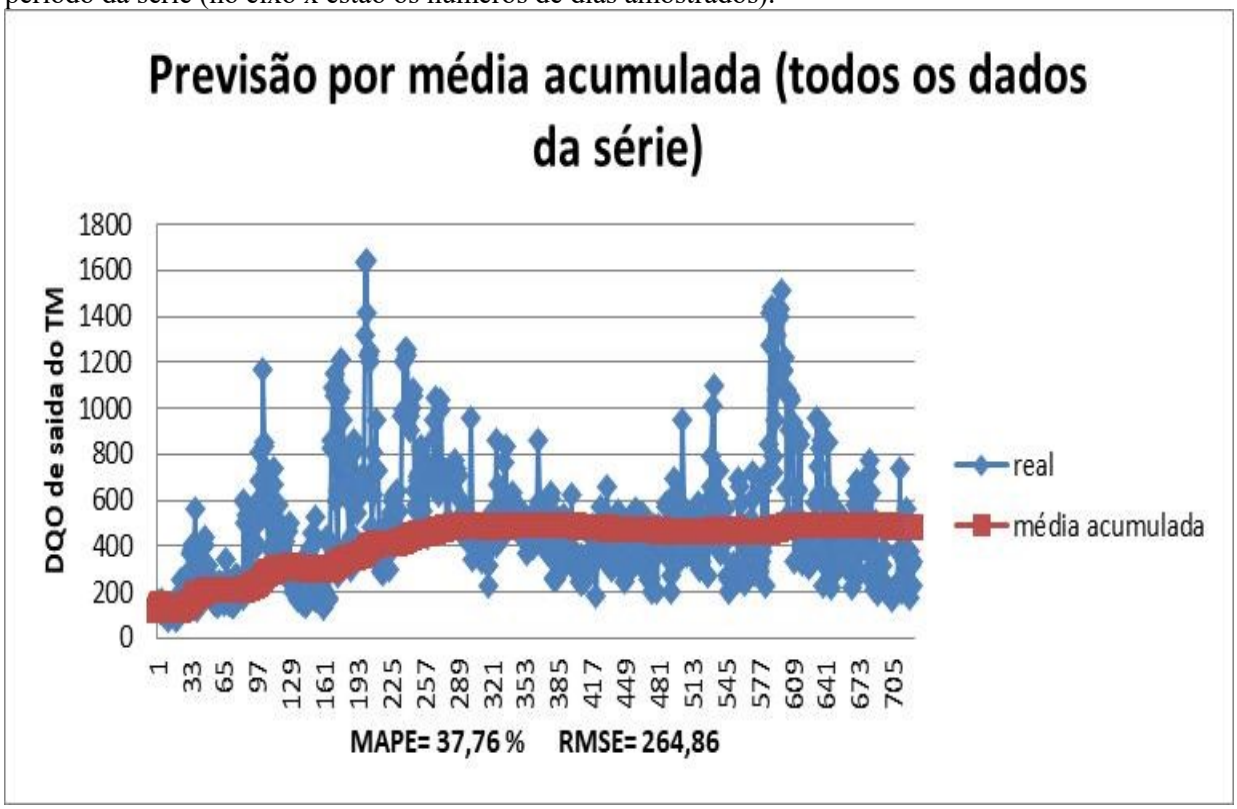
Figura 72 -Previsão univariada, um passo à frente, usando média acumulada e apenas os 30 dias de teste (eixo x)



Observa-se na Figura 72, que o erro apresentado neste método (MAPE de 95,07% e RMSE de 233,03) é superior ao desvio padrão computado no período de 57,7%. Este método exige muita estacionariedade da série para uma previsão pois este desvio mostra que houve em períodos anteriores mudanças em relação a média de todo o período.

Na Figura 73, temos a previsão univariada, um passo à frente, usando média acumulada considerando todo o período de dados.

Figura 73 -Previsão univariada, um passo à frente, usando média acumulada e considerando todo o período da série (no eixo x estão os números de dias amostrados).

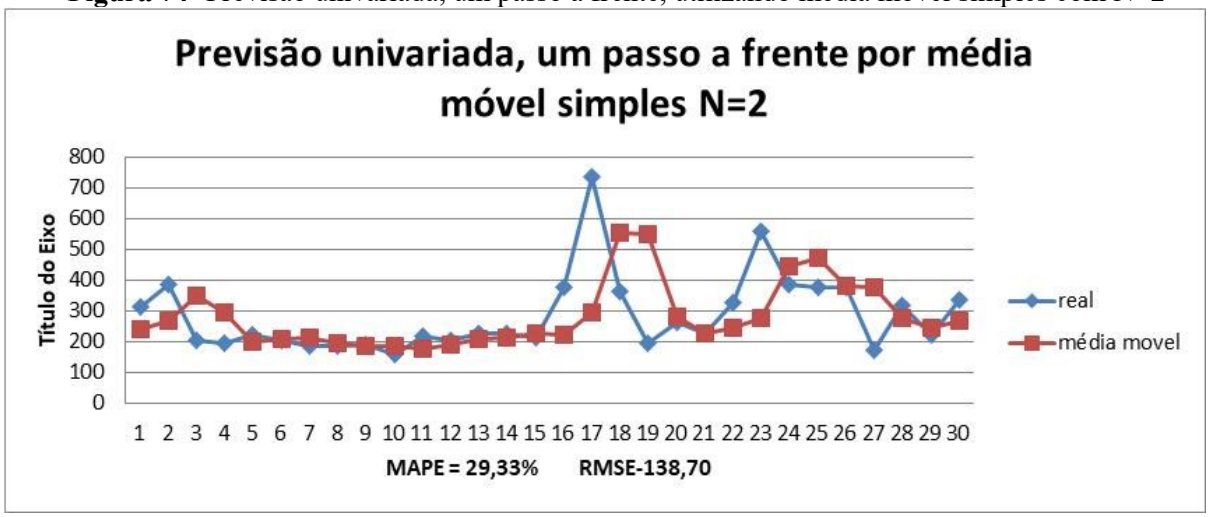


Observa-se na Figura 73, que apesar do erro ser menor considerando todo o período, ocorre que as previsões futuras (Figura 72), ficam sujeitas a erros maiores que um desvio padrão da série. Assim, se aconselha nas tomadas de decisão, observar que a análise de todo um período pode dar resultados de erros bem menores do que o da previsão, usando este método, que exige uma estacionariedade grande da série, para uma boa previsão.

4.5.2 Previsão univariada usando a média móvel simples

Na Figura 74, temos a previsão univariada para o período de teste, um passo à frente, usando média móvel para N=2, com o MAPE e o RMSE

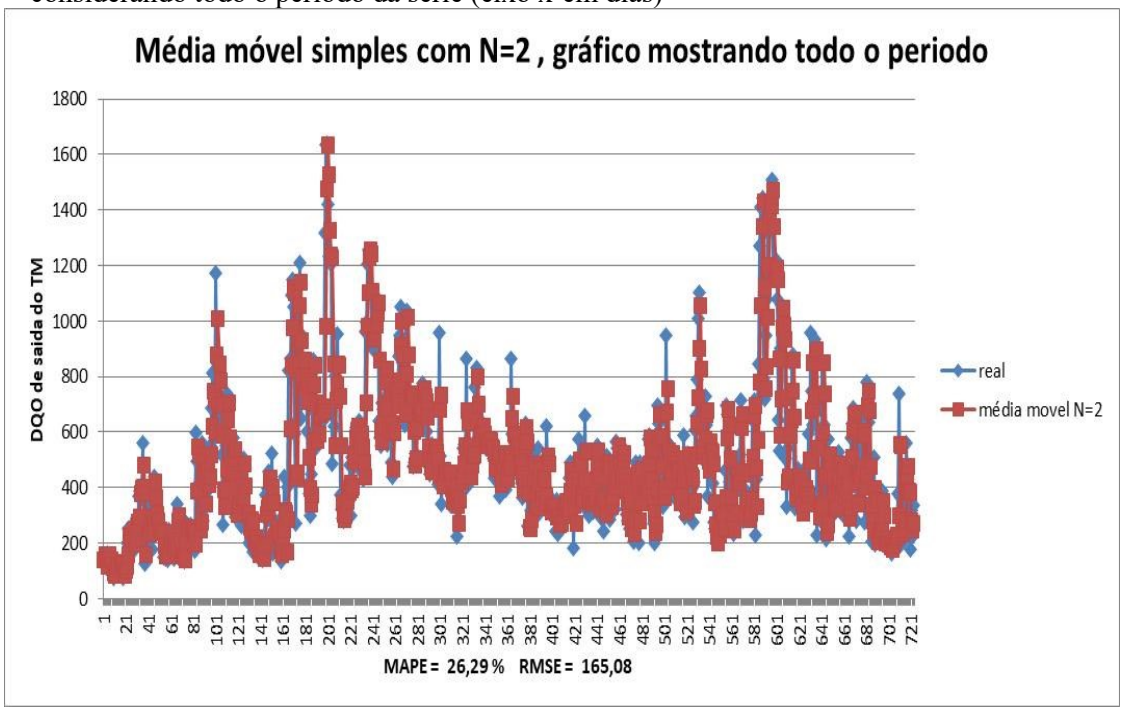
Figura 74 -Previsão univariada, um passo à frente, utilizando média móvel simples com N=2



A Figura 74 mostra que a média móvel (N=2), apresentou erros relativos de 29,33% e RMSE de 138,70, para o período de teste de 30 dias, estes erros são abaixo de um desvio padrão

Na Figura 75, temos a previsão univariada, um passo à frente, usando média móvel para N=2 de todo o período, com o MAPE e o RMSE.

Figura 75 -Previsão univariada, um passo à frente, usando média móvel simples com N=2, considerando todo o período da série (eixo x-em dias)



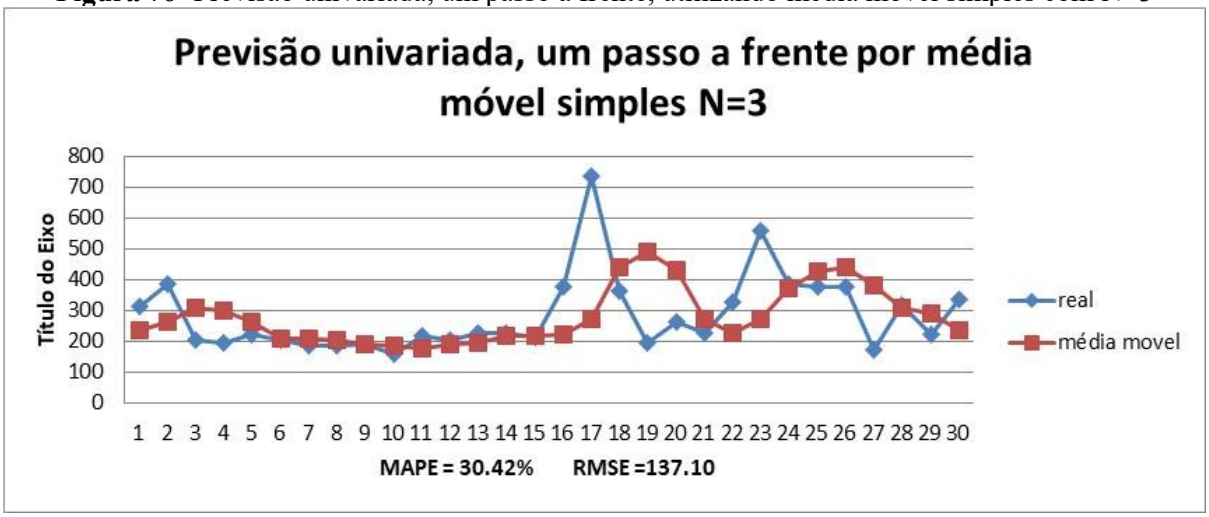
Observa-se na Figura 75, pequena variação para um erro menor entre os dados de todo o período e os dados para 30 dias de previsão, tanto no MAPE (26,29%) como o RMSE (165,08).

Na Figura 76, temos a previsão univariada, para o período de teste, um passo à frente, usando média móvel para n=3, com o MAPE e o RMSE.

Esta previsão serve para comparar o efeito da escolha de média móvel com níveis maiores de N.

Observa-se na Figura 76, pequena variação de erro considerando um aumento de N=2, para N=3, dos números usados na média móvel, tanto no MAPE (30,42%) como o RMSE (137,10). Este resultado ressalta o que foi visto nas correlações de Pearson que mostra que a correlação mais alta do valor da série apresentada está no penúltimo dia ou um dia antes (N=1).

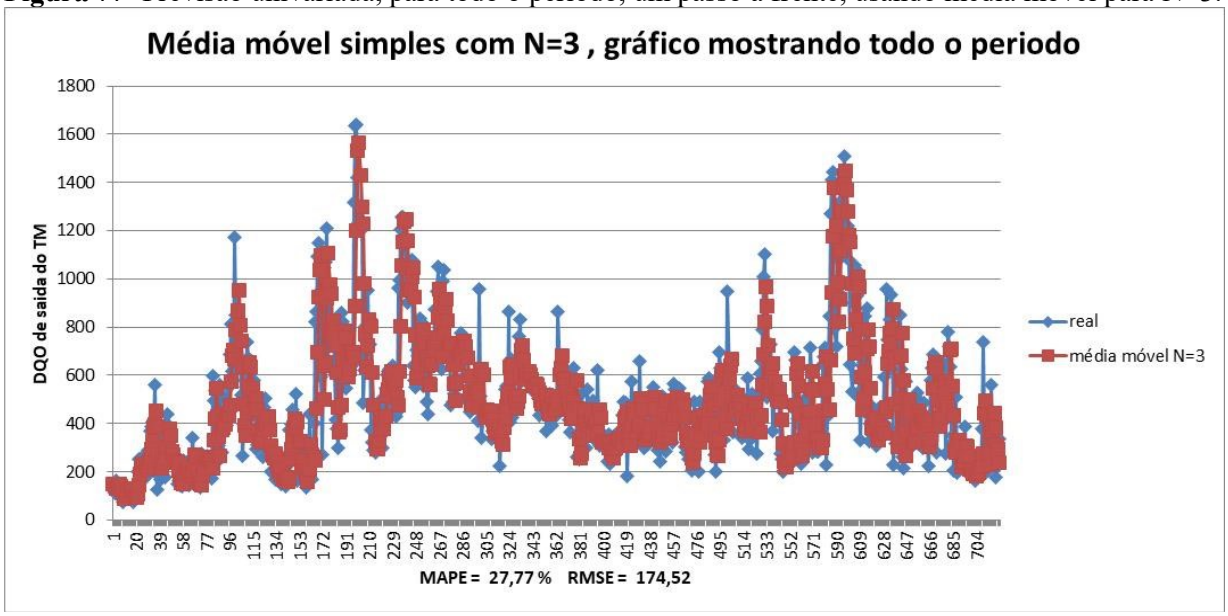
Figura 76 -Previsão univariada, um passo à frente, utilizando média móvel simples com N=3



Na Figura 77, temos a previsão univariada para todo o período, um passo à frente, usando média móvel para N=3, com o MAPE e o RMSE.

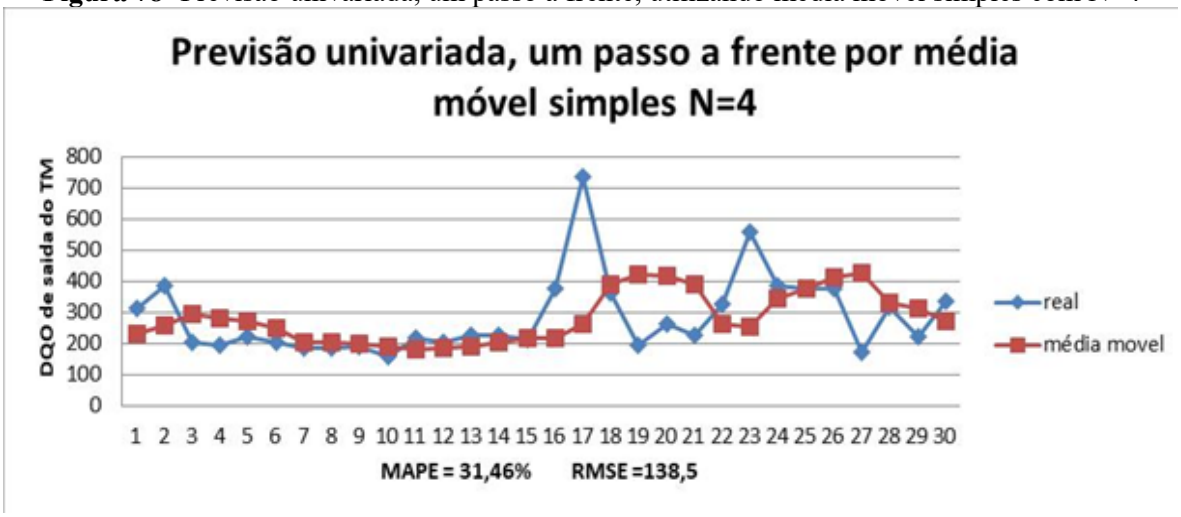
Novamente conforme mostrado na Figura 77, verifica-se que para todo o período o erro se apresenta menor nesta série que o do período de teste (MAPE de 27,77 e RMSE de 174,52).

Figura 77 -Previsão univariada, para todo o período, um passo à frente, usando média móvel para N=3.



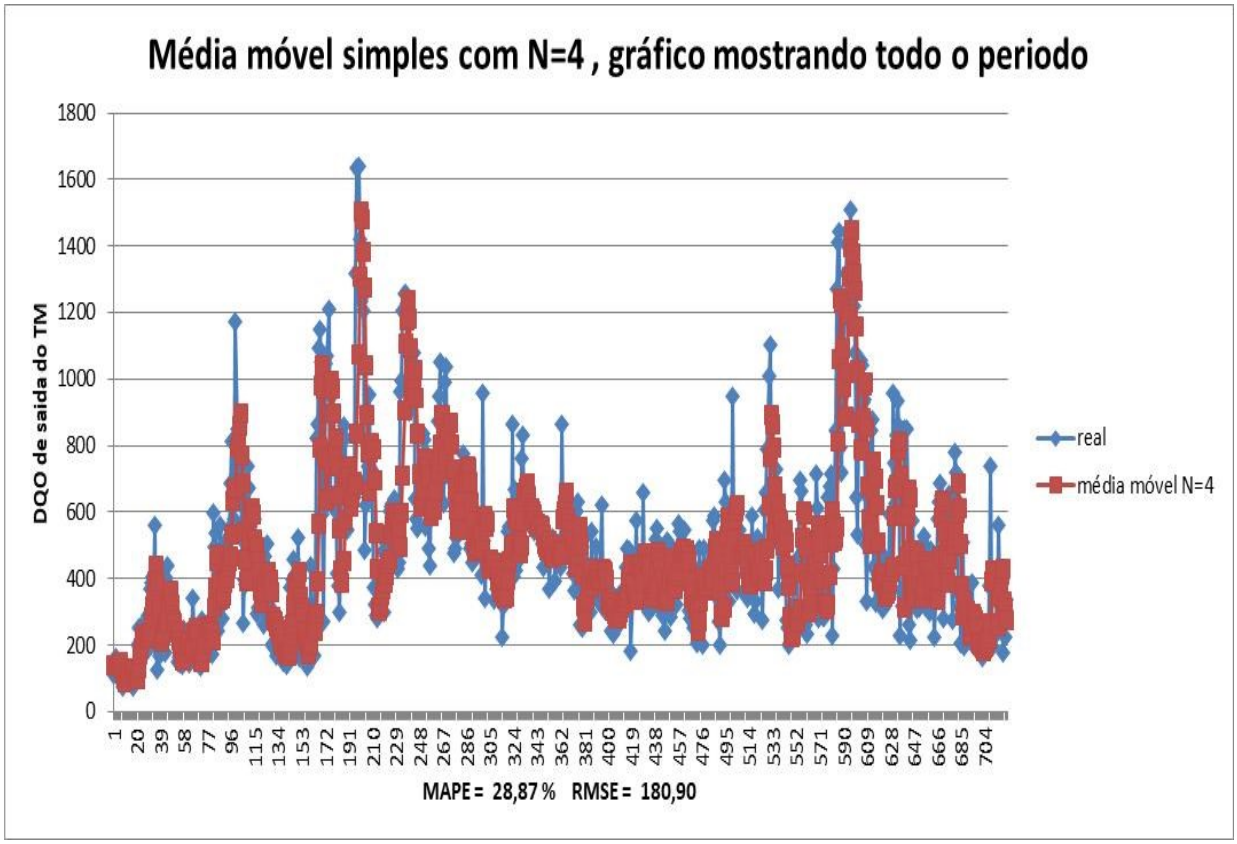
Na Figura 78, temos a previsão univariada, para o período de teste, um passo à frente, usando média móvel para N=4, com o MAPE e o RMSE

Figura 78 -Previsão univariada, um passo à frente, utilizando média móvel simples com N=4



Na Figura 78, verifica-se que para todo o período, o erro se apresenta menor também neste método, que o do período de teste (MAPE de 31,46 e RMSE de 138,5), mostrado na Figura 79.

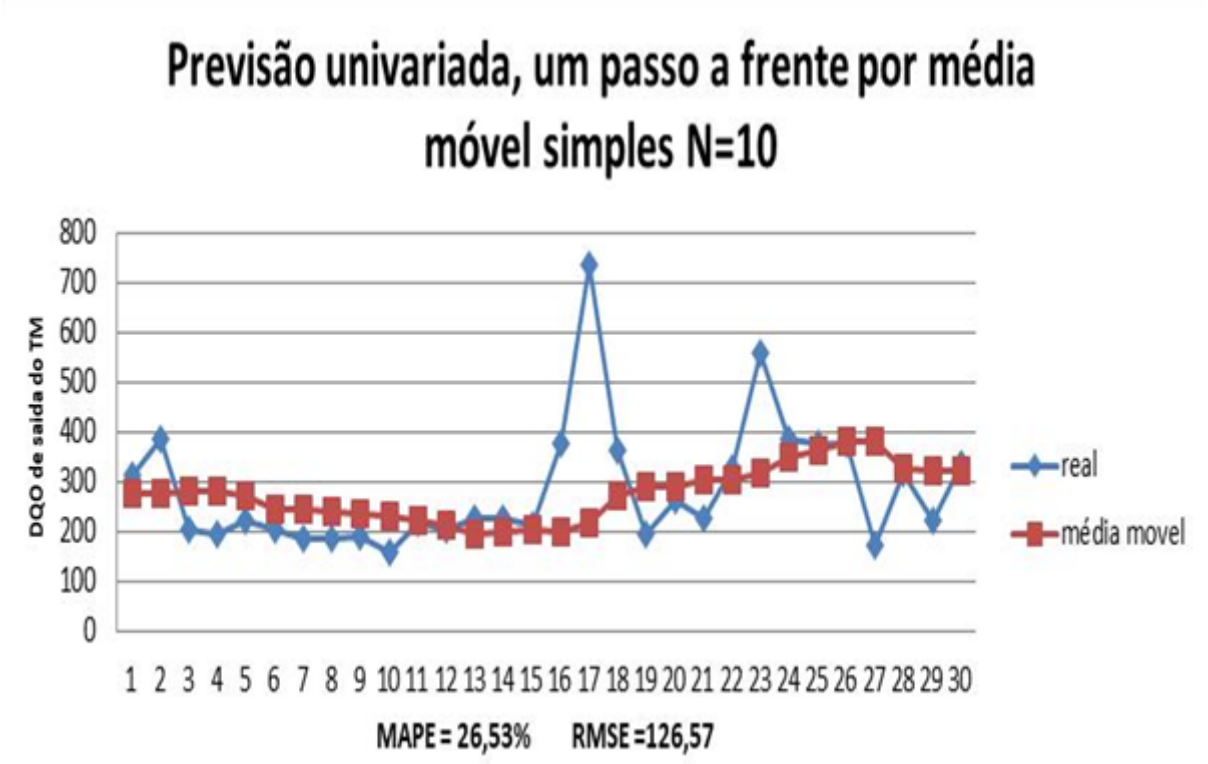
Figura 79 -Previsão univariada para todo o período, um passo à frente, usando média móvel simples com N=4



Novamente conforme mostrado na Figura 79, verifica-se que para todo o período o erro se apresenta menor nesta série que o do período de teste (MAPE de 28,77 e RMSE de 180,90).

Na Figura 80, temos a previsão univariada, para o período de teste, um passo à frente, usando média móvel para N=10, com o MAPE e o RMSE

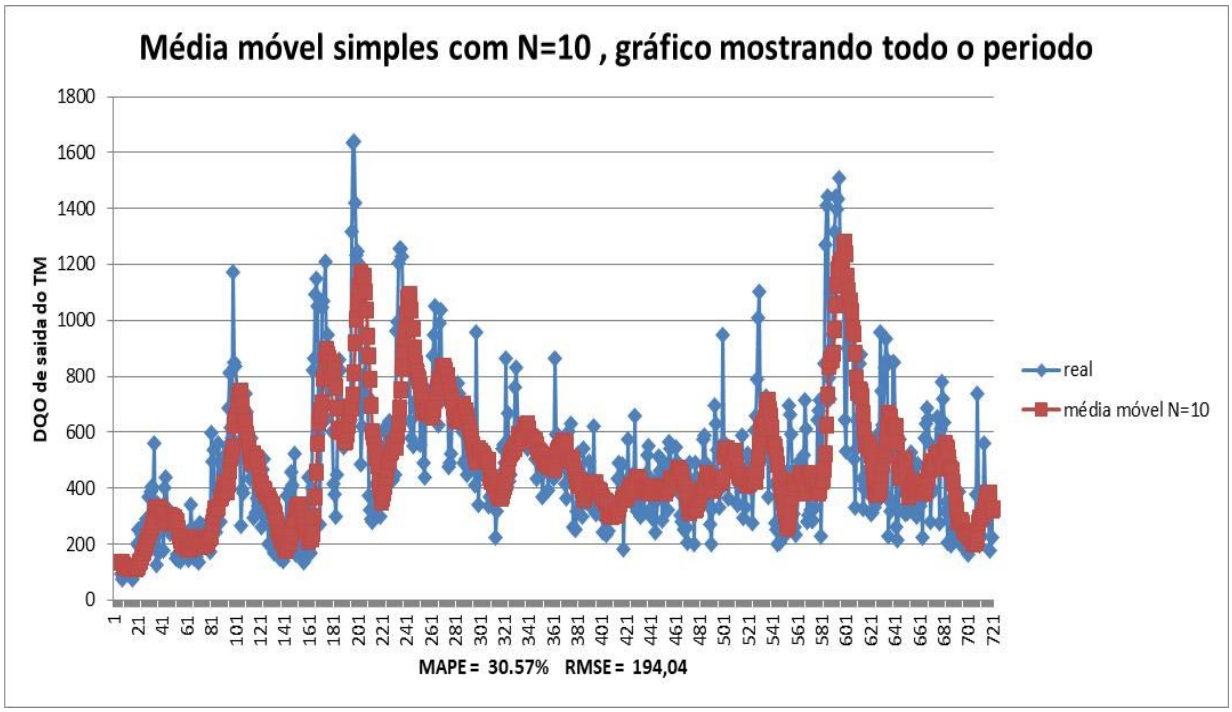
Figura 80 -Previsão univariada, um passo à frente, usando média móvel simples com N=10



Observa-se na Figura 80, que para N=10 o período de teste mostrou uma média móvel mais representativa dos 30 dias com um MAPE de 26,53% e RMSE de 126,57, porém as linhas de previsão são mais atenuadas e neste período específico coincidentemente o erro foi menor, mas não significa que em outros períodos estes valores se repitam. A verificação disto está no resultado em que a série para todo o período teve um erro maior do que o medido em N=2.

Na Figura 81, temos a previsão univariada, para todo o período, um passo à frente, usando média móvel para N=10, com o MAPE e o RMSE

Figura 81 -Previsão univariada, um passo à frente, para todo período, usando média móvel simples com N=10.

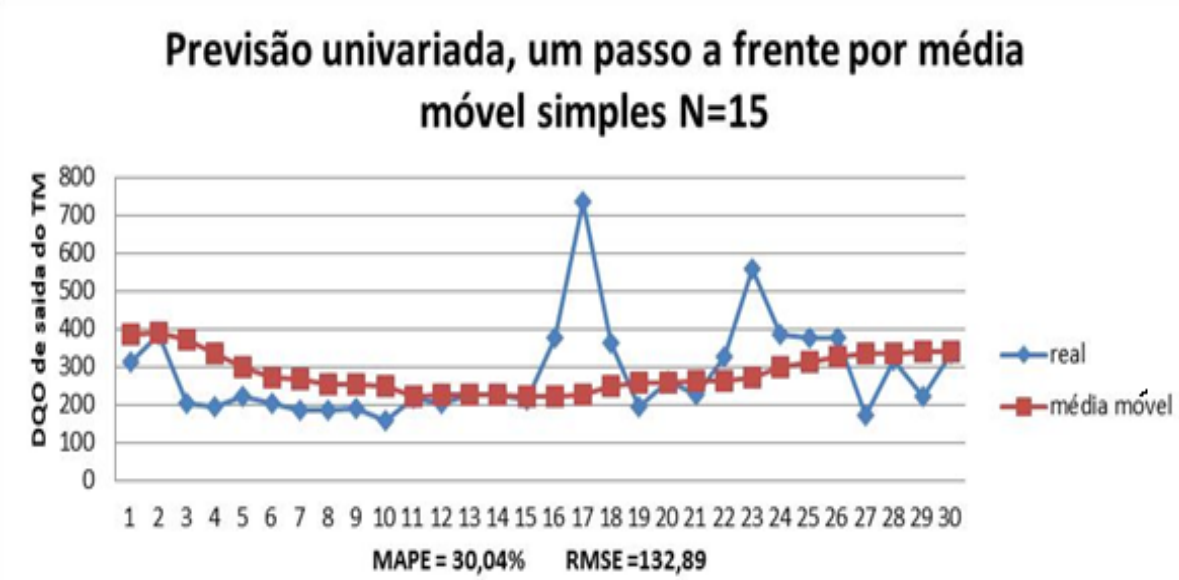


A Figura 81, mostra que o erro para todo período, neste caso é maior que o de período de teste, mostrando que o período escolhido, por coincidência, favoreceu este resultado, por ter tido pontos mais próximos da média dos últimos 10 dias.

Na Figura 81, temos a previsão univariada, para o período de teste, um passo à frente, usando média móvel para N=15, com o MAPE e o RMSE

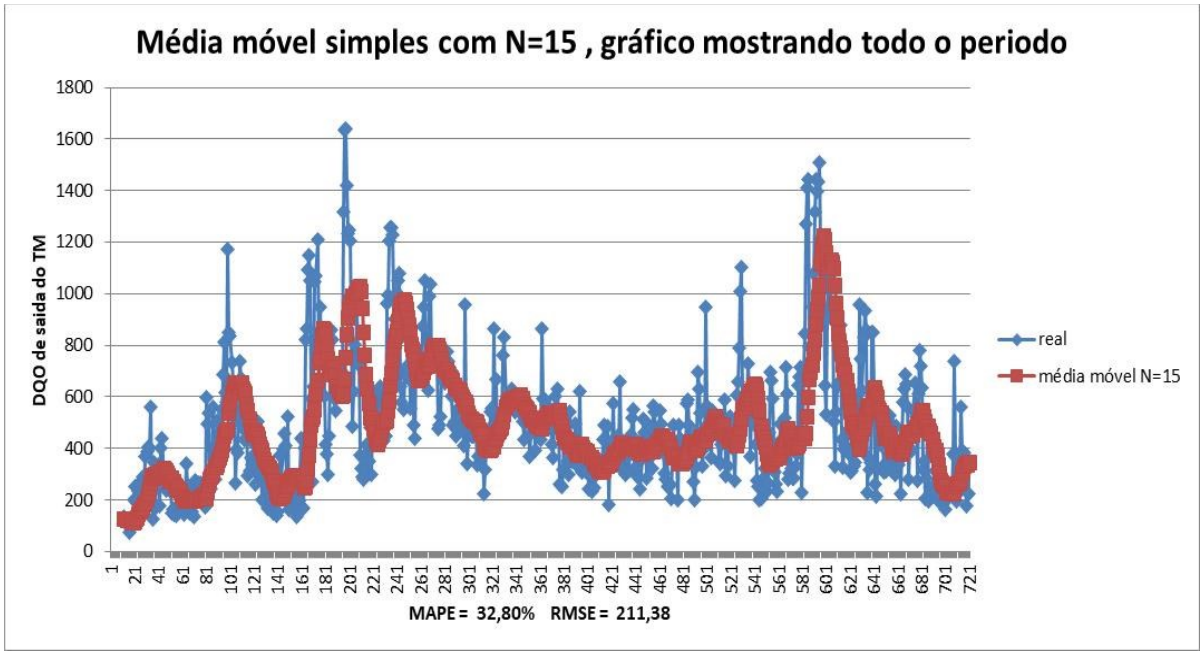
A Figura 82 mostra que aumentando a média móvel com N=15, o resultado do período já aumenta o resultado do MAPE (30,04%) e do RMSE (132,89), do que os valores para N=10, significando uma média móvel de 15 dias resulta em números mais distantes da média geral do período, levando ao aumento do erro.

Figura 82 - Previsão univariada, um passo à frente, usando média móvel simples com N=15



Na Figura 83, temos a previsão univariada, para todo o período, um passo à frente, usando média móvel para N=15, com o MAPE (32,80%) e o RMSE (211,38), mostrando que o aumento de N apresentou um erro relativo maior durante o período de teste, indicando que para médias de períodos longos, pode acontecer mudanças no processo, que o afastam de médias anteriores gerando erros maiores na previsão.

Figura 83 -Previsão univariada, para todo o período, um passo à frente, usando média móvel simples com **N**=15

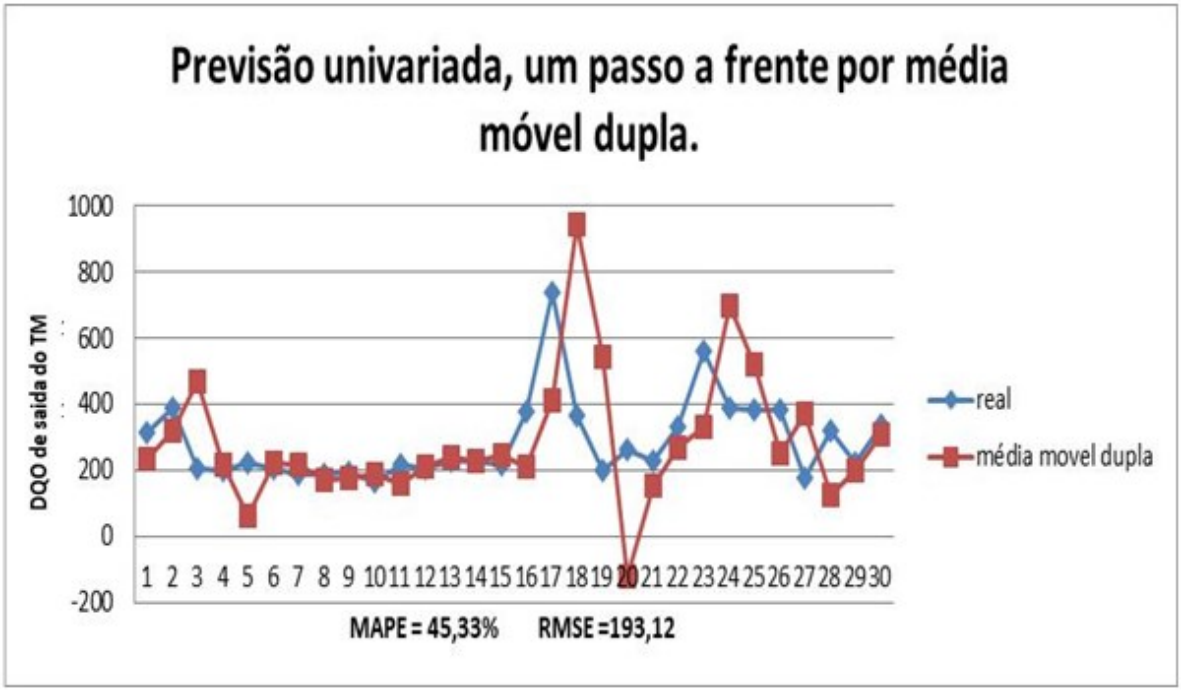


Observa-se no caso que a escolha do N ou número de dias usados na média móvel simples oferece resultados de erros variáveis com previsões de erro relativo abaixo de um desvio padrão, no caso da DQO de saída. O valor de N que der o menor MAPE e RMSE para todo o período e também para o período de teste, seria a escolha para representar melhor a série, como os resultados para N=2 (neste caso).

4.5.3 Previsão univariada usando a média móvel dupla

Na Figura 84, temos a previsão univariada, para o período de testes, um passo à frente, usando média móvel dupla, apresentando um MAPE de 45,33 e o RMSE (193,12), observa-se previsões negativas usando o modelo, mostrando extrapolação de dados.

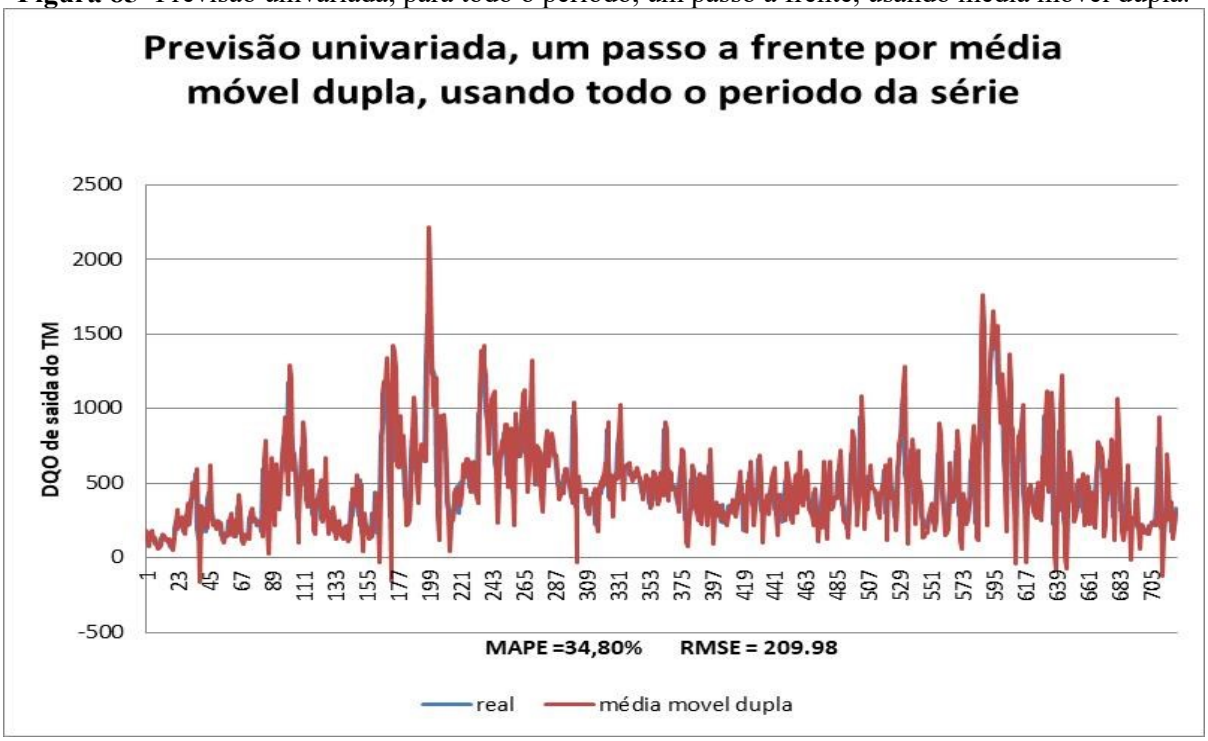
Figura 84 -Previsão univariada, um passo à frente, usando média móvel dupla.



Observa-se na Figura 84, que a média móvel dupla apresentou, neste caso, erros maiores de previsão que a média móvel simples com MAPE de 45,83% e RMSE de 193,12

Na Figura 85, temos a previsão univariada para todo o período, um passo à frente, usando média móvel dupla, com o MAPE e o RMSE, observa-se previsões negativas usando o modelo (pois existem pontos de previsão, que em função de ser calculado pelo dia anterior e estando eles com valores baixos ou próximos de zero, levam a uma previsão negativa do dia à frente), também extrapolação dos dados em função das componentes de tendência do método (a_t e b_t), conforme Equação 2.23.

Figura 85 -Previsão univariada, para todo o período, um passo à frente, usando média móvel dupla.

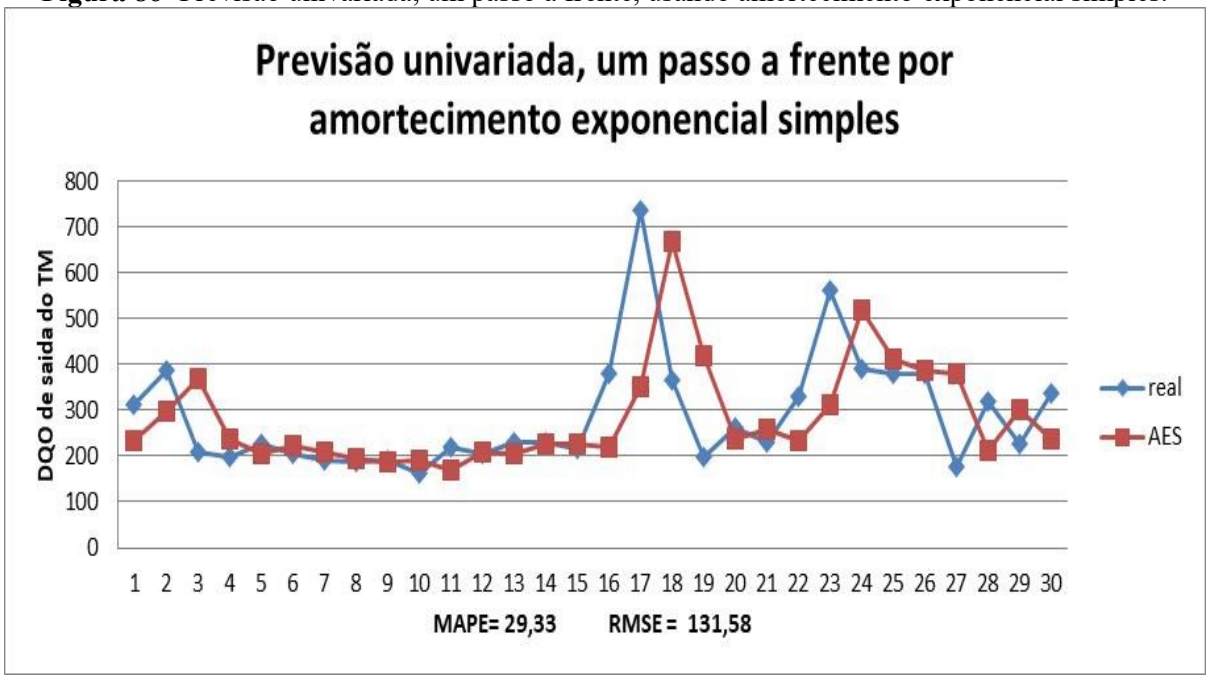


Na Figura 85, observamos também um MAPE de 34,80% para todo o período e RMSE de 209,88, e pontos de extrapolação de dados com valores negativos e a curva de previsão em vermelho, sobrepondo a curva com os dados reais em azul.

4.5.4 Previsão univariada usando o amortecimento exponencial simples (AES)

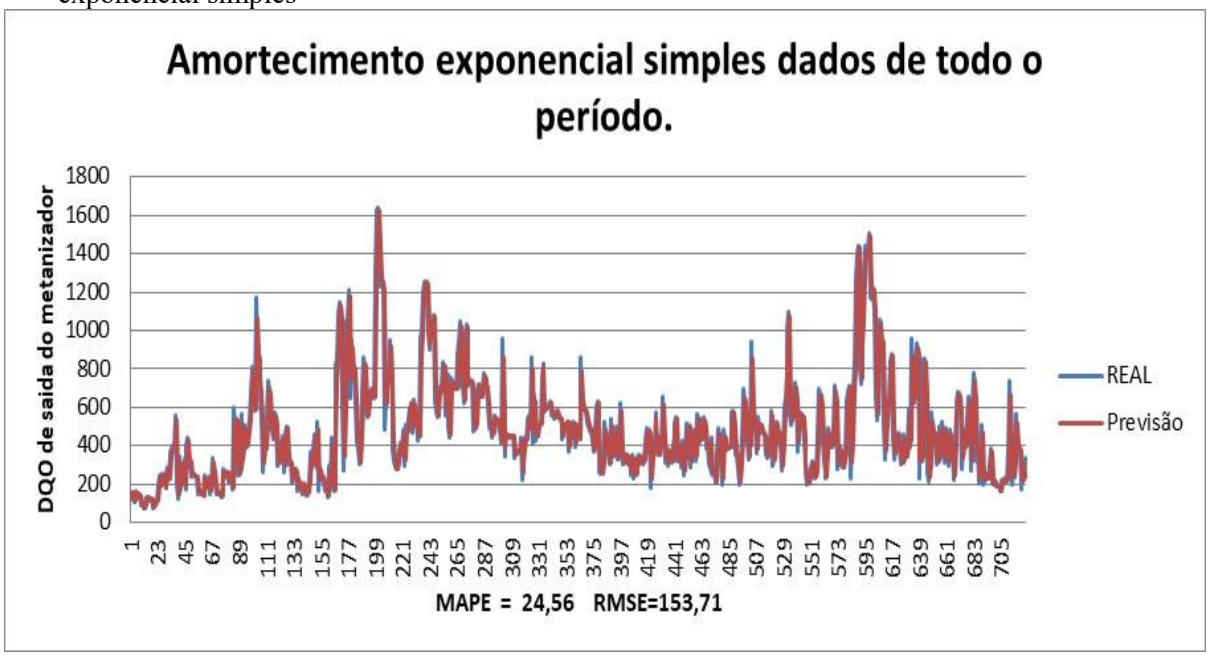
Na Figura 86, temos a previsão univariada, para o período de testes, um passo à frente, usando o amortecimento exponencial simples (AES), com o MAPE de 29,33% e o RMSE de 131,58, ao qual, os valores de α (ver Equação 2.37) ou coeficiente de amortecimento ($0 \leq \alpha \leq 1$), são otimizados pelo software solver do Excel (neste caso $\alpha = 0{,}8224$), para minimizar o erro.

Figura 86 -Previsão univariada, um passo à frente, usando amortecimento exponencial simples.



Na Figura 87, temos a previsão univariada para todo o período, um passo à frente, usando o amortecimento exponencial simples (AES), com o MAPE de 24,56 e o RMSE de 153,71.

Figura 87 -Previsão univariada, um passo à frente de todo o período, usando amortecimento exponencial simples

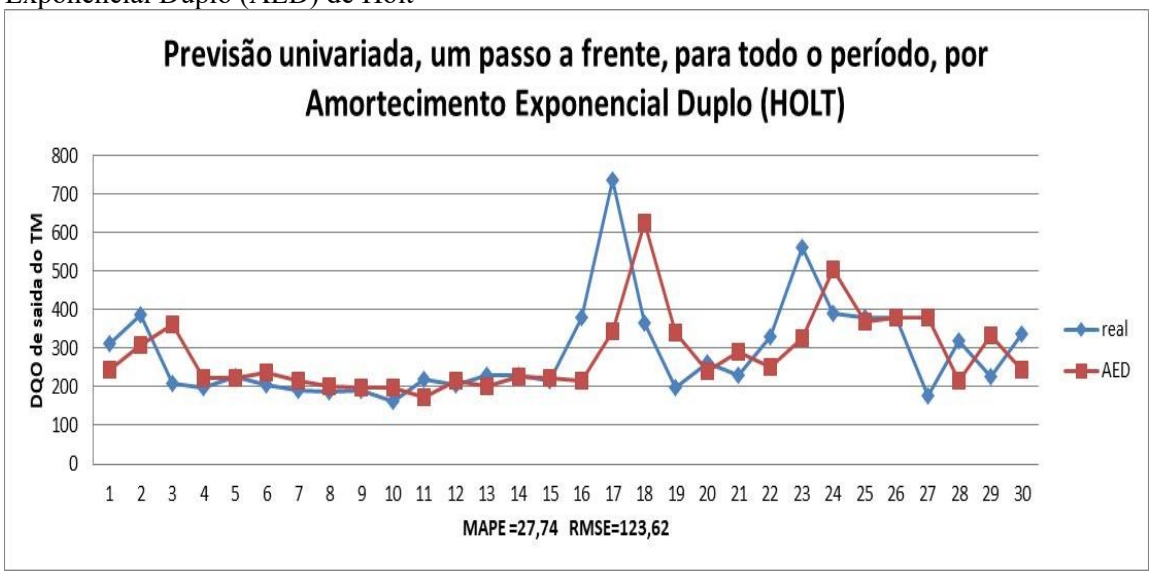


Para todo o período este método apresentou bons resultados em relação ao período total avaliado, mesmo para média simples para N=2.

4.5.5 Previsão univariada usando o amortecimento exponencial duplo de Holt (AED de Holt)

Na Figura 88, temos a previsão univariada, um passo à frente, no período de teste, usando o amortecimento exponencial duplo (AED) de Holt, com o MAPE (27,74%) e o RMSE (123,62). Foi usado o Soft Excel, com a macro do solver, para otimizar os parâmetros de tendência $\propto$ e β de modo a se obter o mínimo erro entre o real e a previsão ficando $\propto = 0{,}4979$ e $\beta = 0{,}5936$ Observa-se que, neste método se segue a tendência anterior da curva real sem desviar muito da projeção.

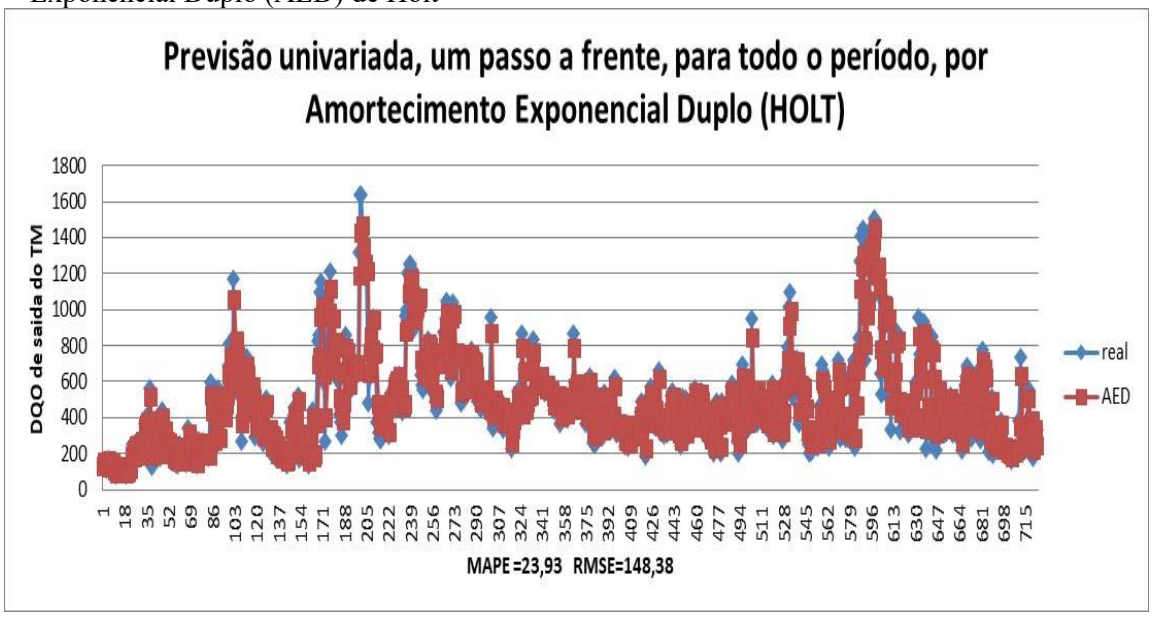
Figura 88 -Previsão univariada, um passo à frente, do período de testes, usando Amortecimento Exponencial Duplo (AED) de Holt



Para o período de testes, observa-se valores bem próximos dos observados nos últimos métodos, porém o período de teste pode favorecer temporariamente os resultados de um método em relação ao outro. A avaliação do período todo pode dar mais informações em relação aos resultados de uma forma geral.

Na Figura 89, temos a previsão univariada, um passo à frente, usando o amortecimento exponencial duplo (AED) de Holt, para todo o período, com o MAPE e o RMSE

Figura 89 -Previsão univariada, um passo à frente de todo o período, usando Amortecimento Exponencial Duplo (AED) de Holt

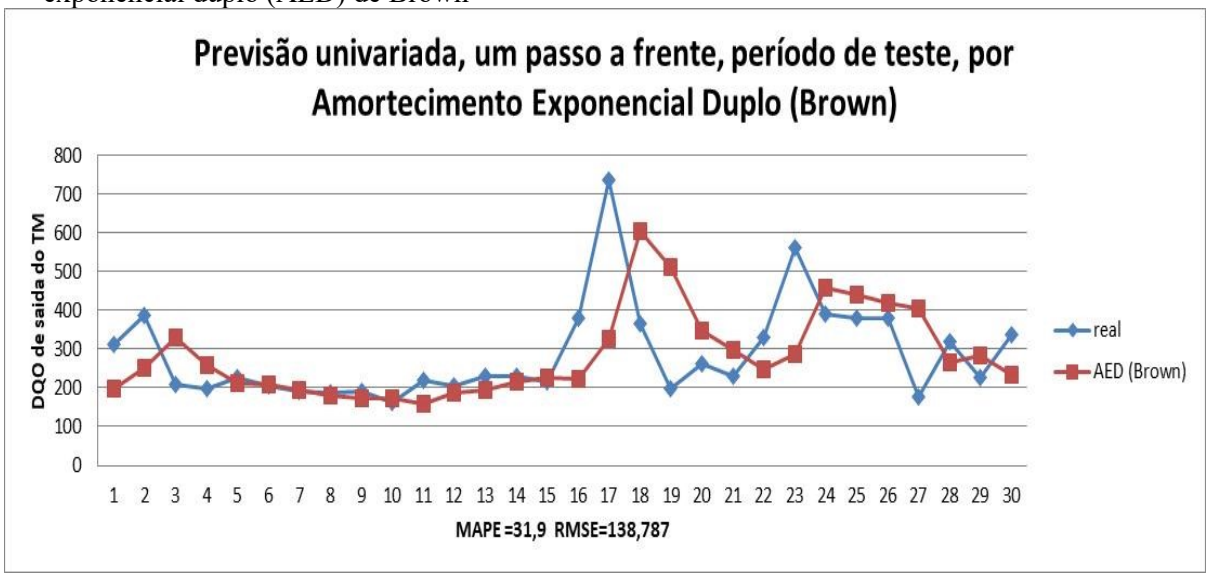


Observa-se neste caso um resultado mais baixo do MAPE (23,93) e RMSE de 148,38, considerando todo o período de teste

4.5.6 Previsão univariada usando amortecimento exponencial duplo de Brown (AED de Brown)

Na Figura 90, temos a previsão univariada, um passo à frente, no período de teste, usando o amortecimento exponencial duplo (AED) de Brown, com o MAPE e o RMSE, foi usado o Soft Excel com a macro do solver, para otimizar os parâmetros de tendência $\propto$ (ver Equação 2.41, de modo a se obter o mínimo erro entre o real e a previsão. Observa-se que neste método se segue a tendência anterior da curva real sem desviar muito da projeção.

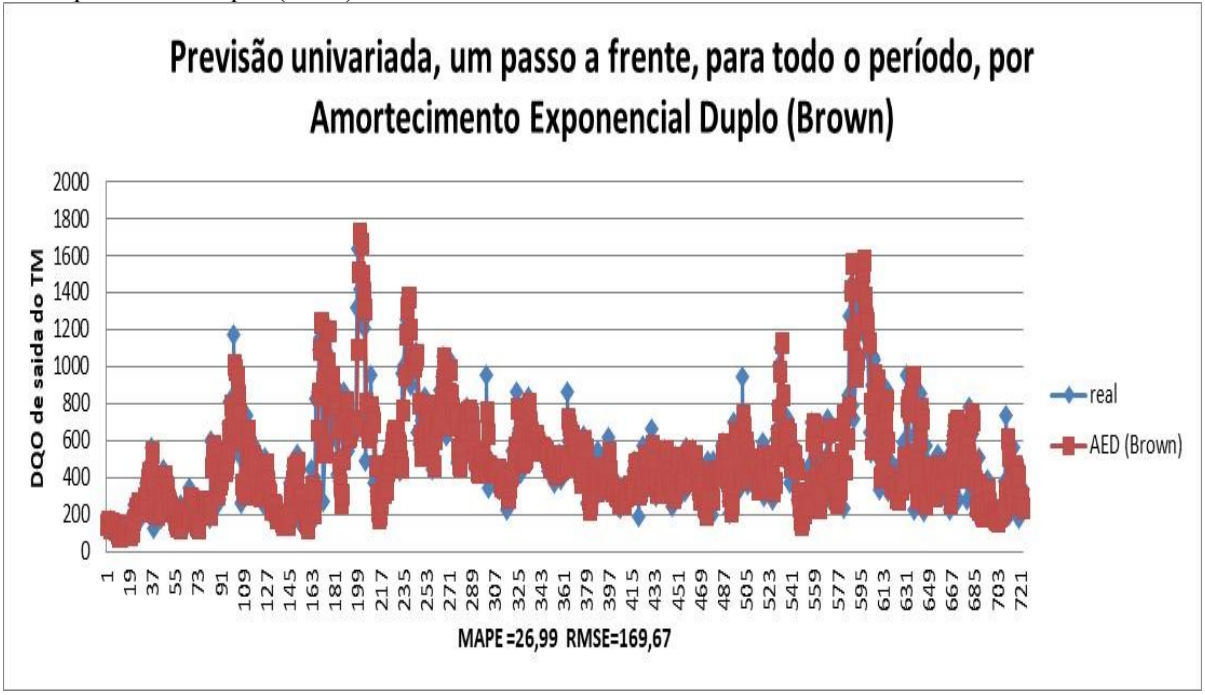
Figura 90 -Previsão univariada, um passo à frente, no período de teste, usando o amortecimento exponencial duplo (AED) de Brown



No período de teste, conforme a Figura 90, a previsão apresentou valores de MAPE de 31,9% e RMSE de 138,78, mostrando resultados próximos de outros métodos já testados.

Na Figura 91, temos a previsão univariada, um passo à frente, usando o amortecimento exponencial duplo (AED) de Brown, para todo o período, com o MAPE e o RMSE. Utilizou-se o valor de $\propto = 0{,}3201$, após uso da macro solver do excel para otimizar o parâmetro reduzindo-se os erros.

Figura 91 -Previsão univariada, um passo à frente, para todo o período, usando Amortecimento Exponencial Duplo (AED) de Brow



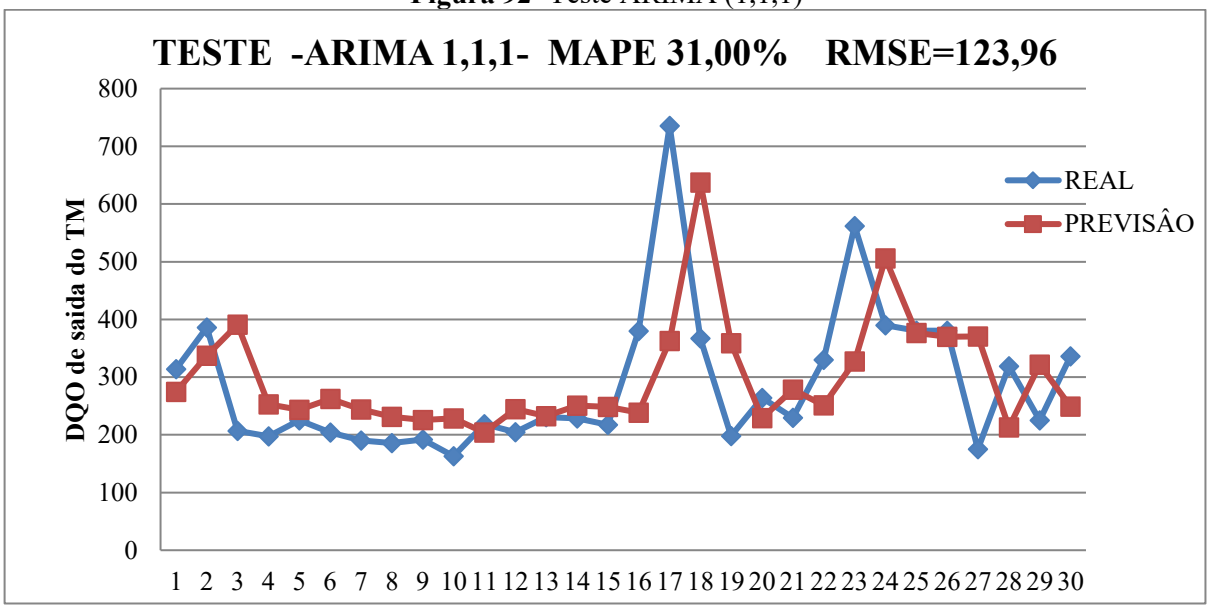
Para todo o período conforme Figura 91, o amortecimento duplo de Brow, apresentou um valor um pouco maior de MAPE (26,99%) e RMSE (169,67) do que a encontrada para o amortecimento exponencial duplo de Holt.

4.5.8 Previsão univariada usando ARIMA

Se utilizou o *soft Real Statistics Using Excel*, no modo ARIMA, para se verificar a previsão um passo à frente da DQO univariada, foram feitos quatro testes para comparação dos resultados no período de teste e em todo o período usando ARIMA (1,1,1) (Figura 92 e 93), ARIMA (2,1,1) (Figura 94 e 95), ARIMA (2,1,2) (Figura 96 e 97), ARIMA (2,2,2) (Figura 98 e 99).

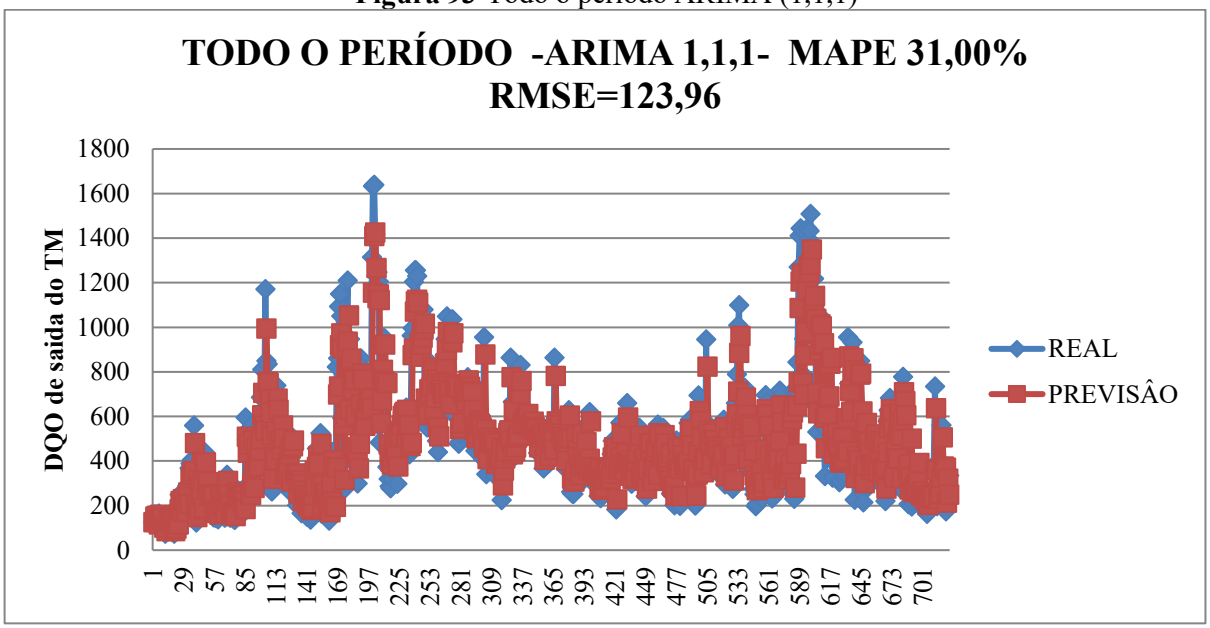
Os resultados do ARIMA (1,1,1), para o período de teste (figura 92) mostram dados de previsão com MAPE de 31,00% e RMSE de 123,96, estes dados ficam na faixa de resultados de outros métodos univariados já vistos

Figura 92- Teste ARIMA (1,1,1)



Os resultados do ARIMA (1,1,1), para o todo o período (Figura 93), mostram dados de previsão com MAPE de 31,00% e RMSE de 123,96, estes dados ficam na faixa de resultados de outros métodos univariados já vistos não apresentando um ganho diferenciável.

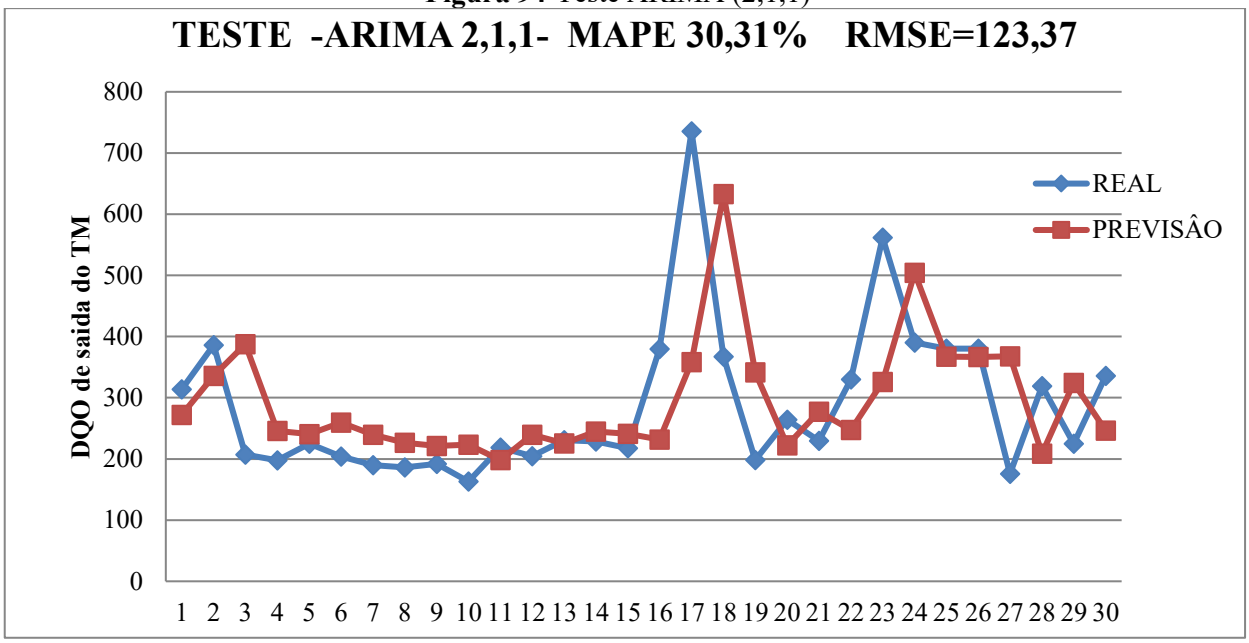
Figura 93-Todo o período ARIMA (1,1,1)



Os resultados do ARIMA (2,1,1), para o período de teste (figura 94) mostram dados de previsão com MAPE de 30,31% e RMSE de 123,37, estes dados ficam na faixa de resultados de

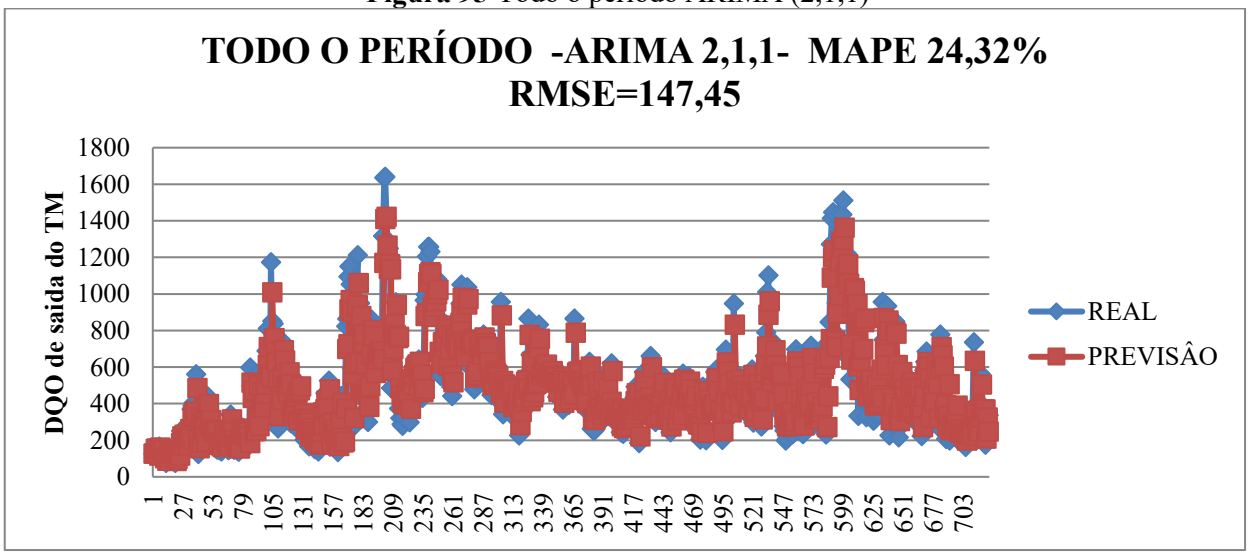
outros métodos univariados já vistos e apresenta pequeno ganho em relação ao modelo ARIMA (1,1,1)

Figura 94-Teste ARIMA (2,1,1)



Os resultados do ARIMA (2,1,1), na Figura 95, para o todo o período, mostram dados de previsão com MAPE de 24,32% e RMSE de 147,45, estes dados indicam uma melhoria de previsão, apresentando um ganho diferenciável, considerando em particular esta série temporal.

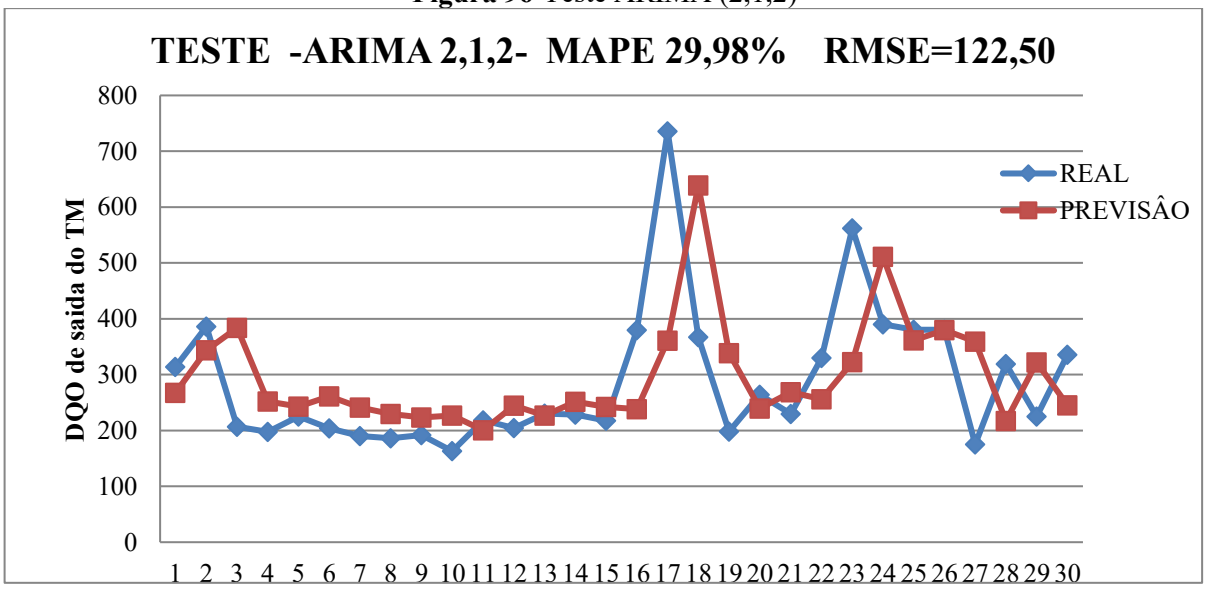
Figura 95-Todo o período ARIMA (2,1,1)



Os resultados do ARIMA (2,1,2), na Figura 96, para o período de teste, mostram dados de previsão com MAPE de 29,98% e RMSE de 122,50, estes dados indicam um ganho pequeno

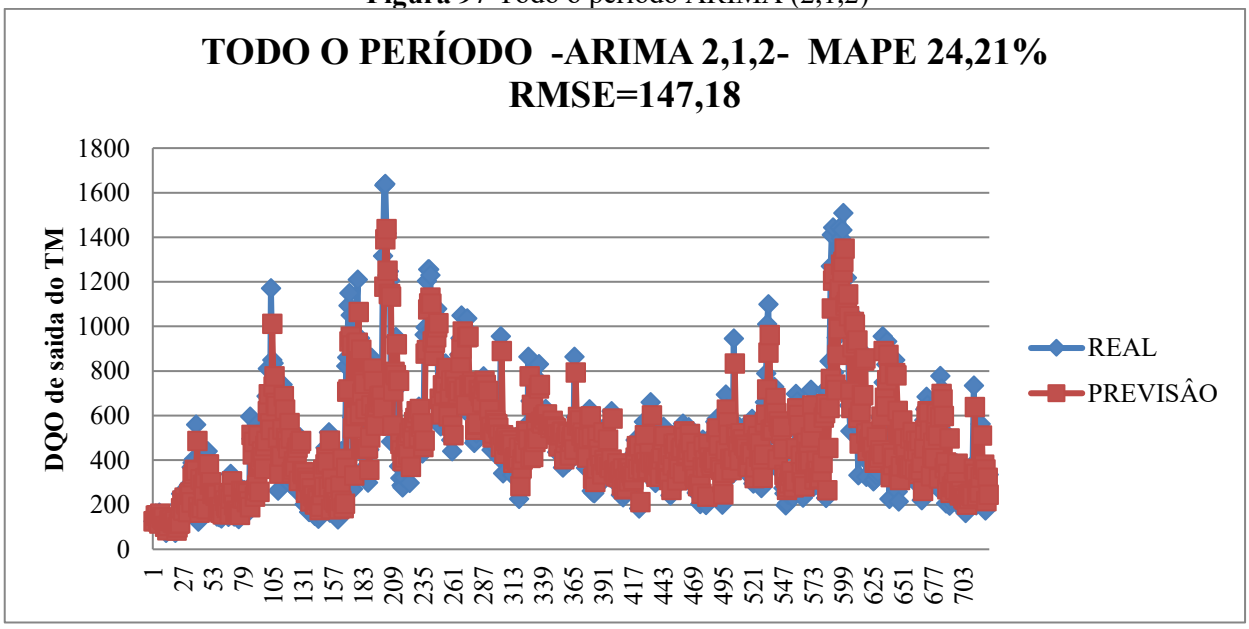
na previsão, em comparação com a ARIMA (2,1,1), considerando em particular esta série temporal.

Figura 96-Teste ARIMA (2,1,2)



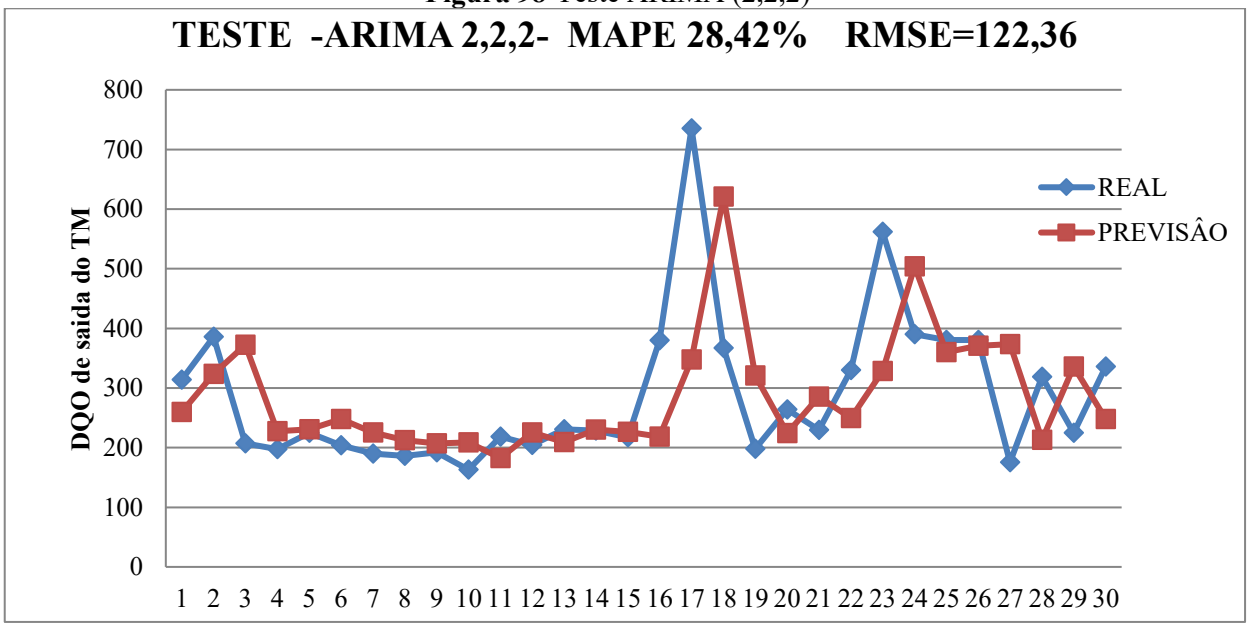
Os resultados do ARIMA (2,1,2), na Figura 97, para o todo o período, mostram dados de previsão com MAPE de 24,21% e RMSE de 147,28, estes dados indicam um ganho pequeno em relação a ARIMA (2,1,1)

Figura 97-Todo o período ARIMA (2,1,2)



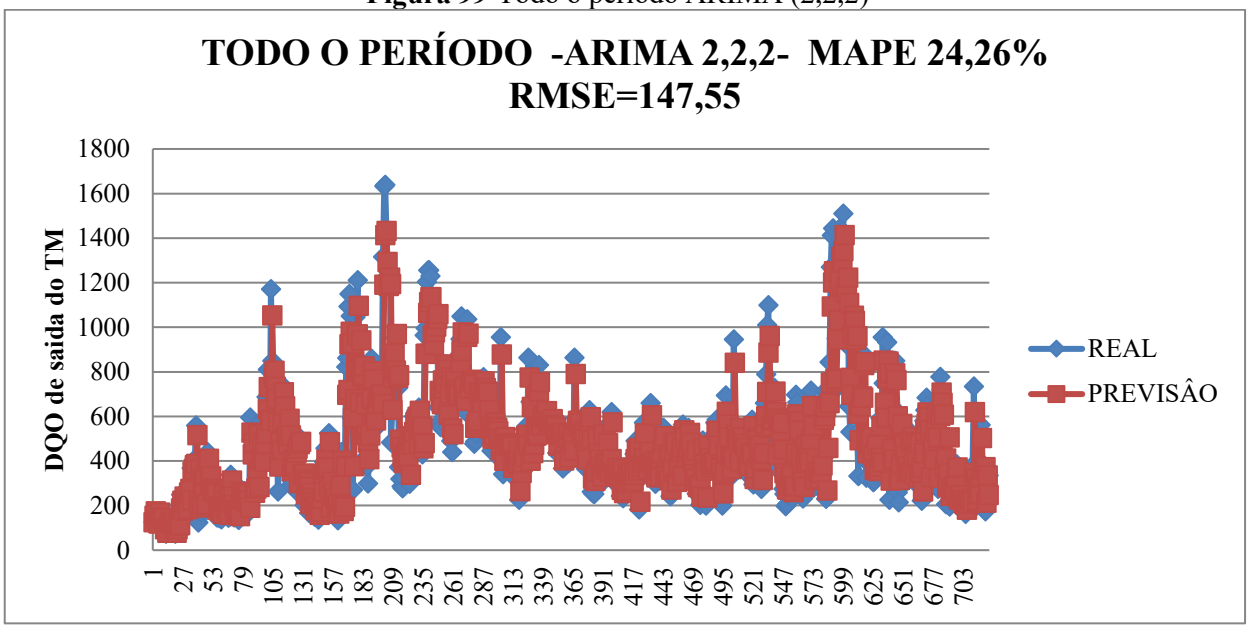
Os resultados do ARIMA (2,2,2), na Figura 98, para o período de teste, mostram dados de previsão com MAPE de 28,42% e RMSE de 122,36, estes dados indicam um ganho pequeno na previsão, em comparação com a ARIMA (2,1,2), considerando em particular esta série temporal.

Figura 98-Teste ARIMA (2,2,2)



Os resultados do ARIMA (2,2,2), na Figura 99, para o todo o período, mostram dados de previsão com MAPE de 24,26% e RMSE de 147,55, estes dados indicam uma pequena diferença a mais em relação a ARIMA (2,1,2)

Figura 99-Todo o período ARIMA (2,2,2)



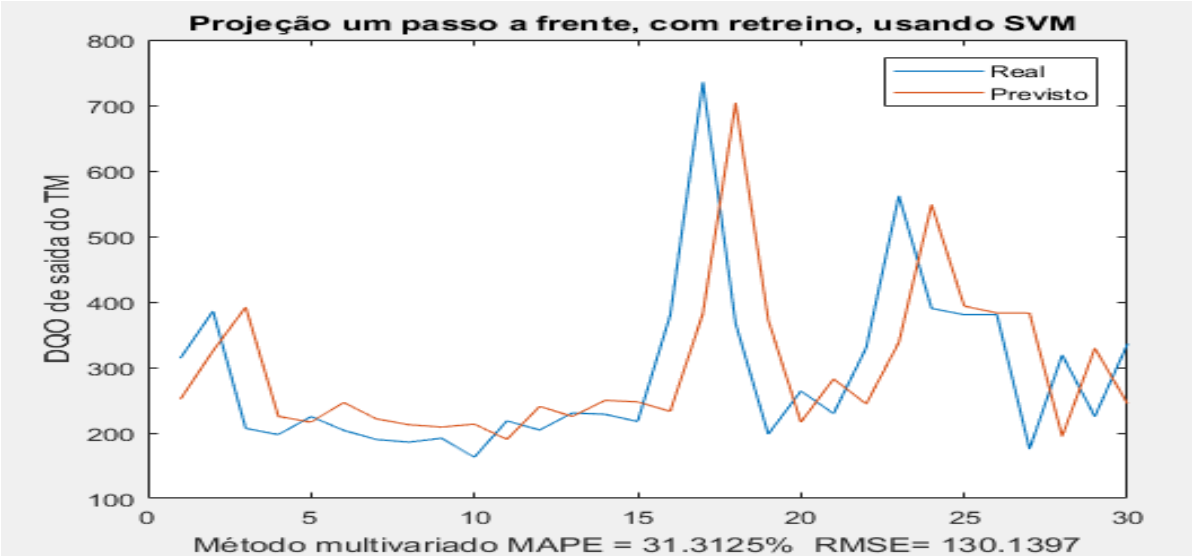
Observa-se nos métodos univariados, que se pode chegar a valores de MAPE entre 23 a 24% usando métodos de previsão convencionais

4.5.9 Previsão univariada usando vetor suporte de máquina

Foi realizado uma previsão univariada, para verificar a previsão da série um passo à frente (um dia a frente, considerando dados usando média diária), usando o MATLAB, cuja programação se encontra no **Apêndice G,** e a técnica do vetor suporte de maquina (VSM), usando um Kernel linear, cujo resultado está apresentado na Figura 100.

Avaliou-se apenas o período de testes, pois a metodologia tende a minimizar os erros usados no período de teste de modo a quase zerar os erros entre os dados usados no treinamento, ficando apenas as previsões sujeitas a erros. A mesma metodologia foi aplicada também a métodos que usam redes neurais pelo mesmo motivo.

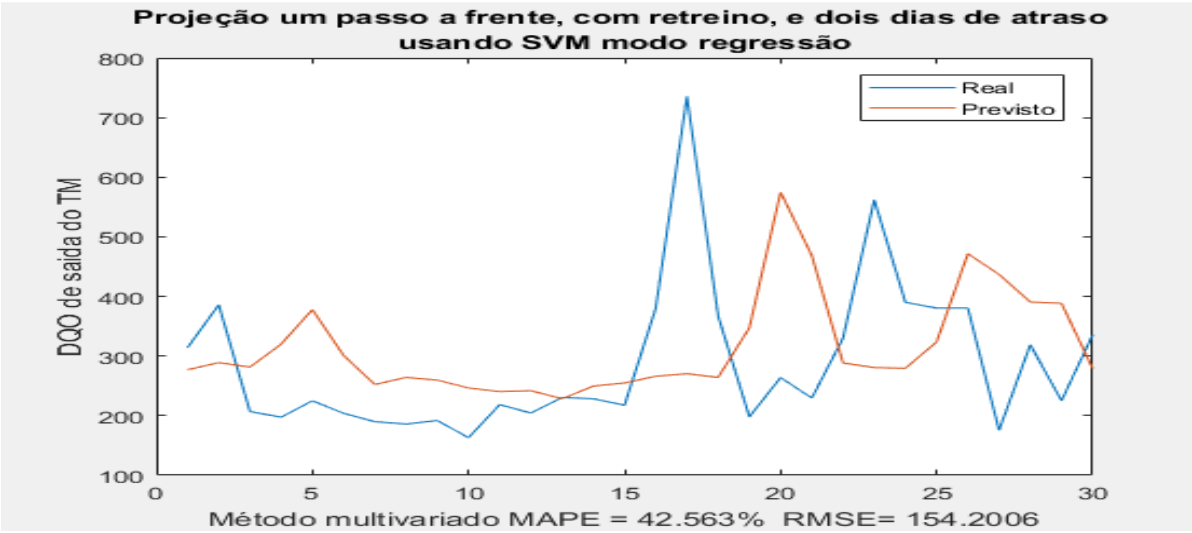
Figura 100 -Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO um dia antes.



A Figura 100, mostra o gráfico do período de teste usando SVM, que apresenta resultados de MAPE (31,31%) e RMSE (130,13), na faixa de resultados dos métodos convencionais, considerando o período de testes de 30 dias escolhido.

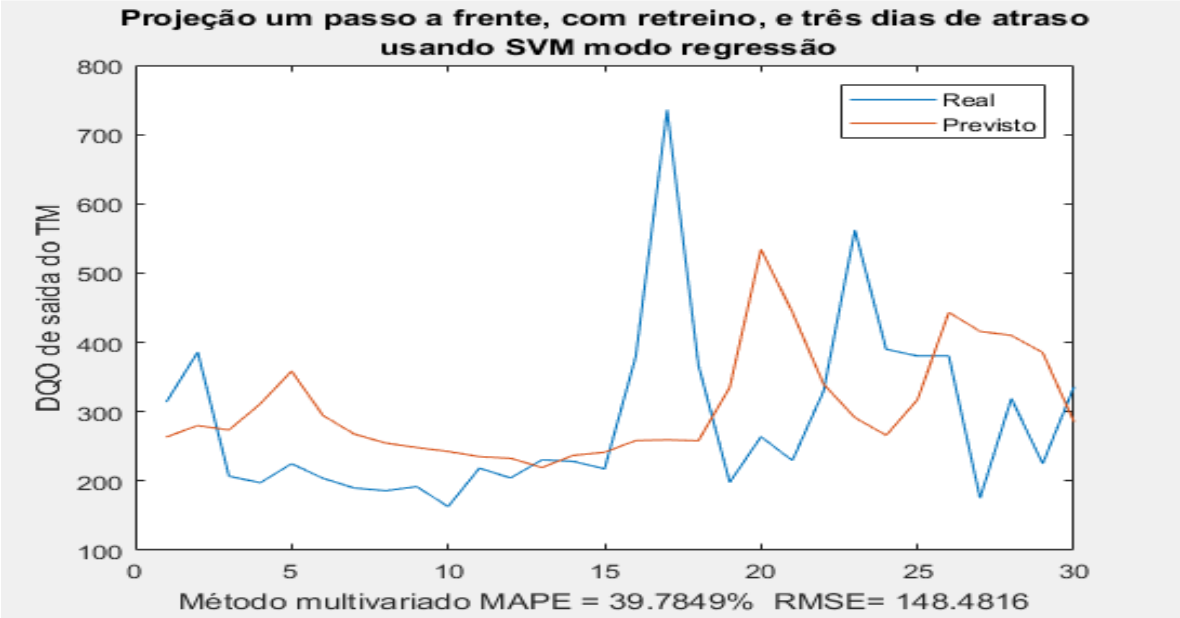
Nas Figuras 101 e 102 se tem a previsão usando dois e três dias de atraso respectivamente, para o período de testes.

Figura 101 -Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO dois dias antes.



Observa-se na Figura 101, que o uso de dois dias antes em relação a um dia antes, na entrada do método, não melhora os resultados de previsão, aumentando o MAPE (42,56%) e RMSE (154,20).

Figura 102 -Previsão univariada da DQO de saída do TM, usando SVM e dados de entrada da DQO três dias antes.



Observa-se na Figura 102, que o uso de três dias antes em relação a um dia antes, na entrada do método, não melhora os resultados de previsão, aumentando o MAPE (39,78%) e RMSE (148,48).

4.5.7 Previsão univariada usando rede neural feedforward

Foi usado um passo à frente (um dia à frente) em uma série univariada, usando redes neurais *feedforward*, com uma camada oculta, uma camada de entrada e uma de saída. Em relação a quantidade de neurônios usados na camada oculta, foi usado, para todas as redes neurais, a regra sugerida Hecth-Nielsen (1990), que sugere a quantidade de (2i+1) neurônios na camada oculta, sendo i o número de neurônios (variáveis) na camada de entrada. Esta rede foi retreinada com os dados do período à medida que os dias em testes estavam sendo incrementados visando se obter a projeção de um dia a frente.

As variáveis são unidimensionalizadas, assim as funções tem valores reais entre zero e um, em função disto se usou combinações de função de ativação logística totalmente

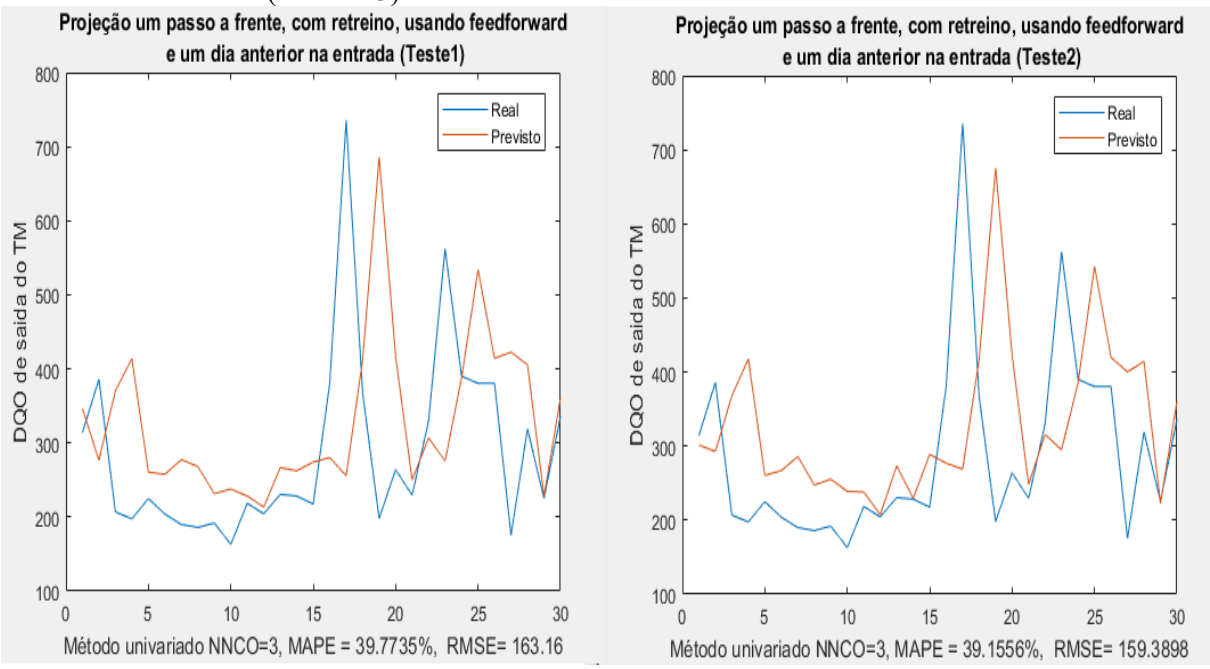
diferenciável, dentro da faixa entre zero e um para os neurônios da camada oculta e uma função linear para os neurônios de saída.

Observou-se durante a pesquisa, no caso em função do processo de interação das redes neurais e dos ruídos da variável alvo ou objeto (no caso a DQO de saída do tanque de metanização), que os resultados de rede podem variar mesmo se mantendo a mesma estrutura usada e o número de neurônios, ao se rodar nova interação, em função de cada treinamento existir uma interação diferente, com valores de pesos diferentes.

Assim, foi utilizada uma contraprova nos testes, ao qual se usa a mesma rede neural, para se verificar se existe muita variação entre os resultados obtidos.

A Figura 103, mostra duas previsões, para o período de teste, feitas em duplicatas devido a não reprodutibilidade dos testes neste método, usando na rede um dia anterior na entrada da rede, e numero de neurônios na camada oculta (NNCO) de 3.

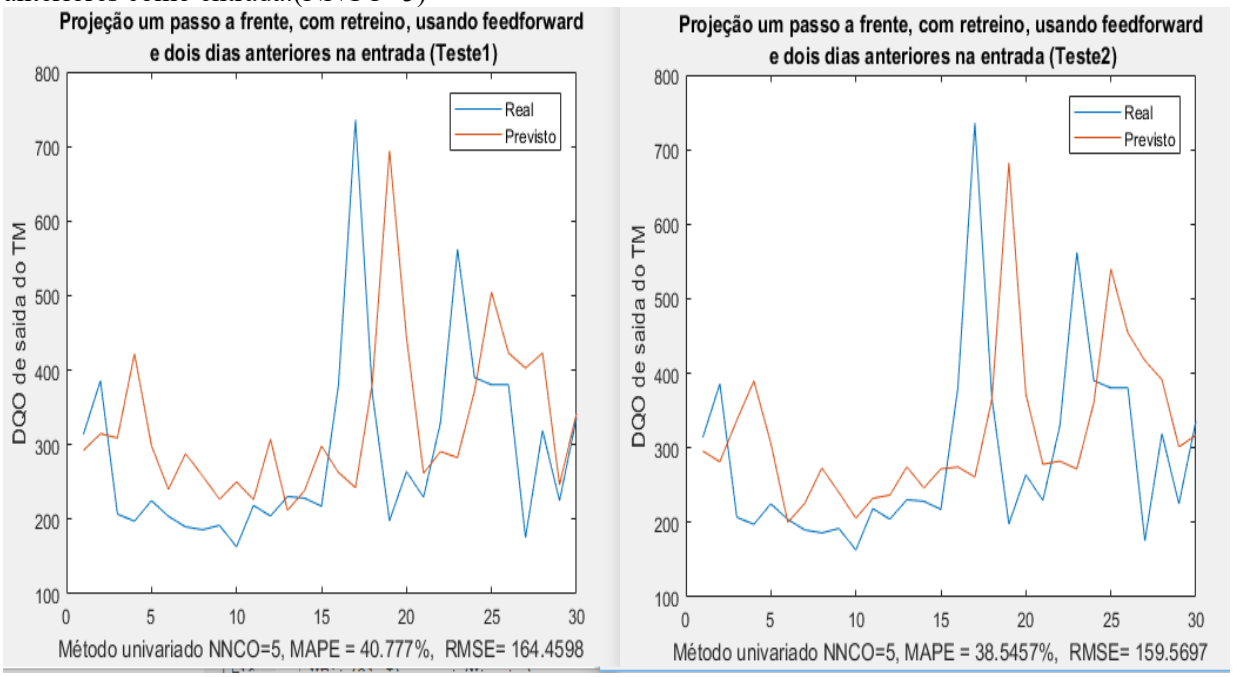
Figura 103-Tela do Matlab, resultado dos testes usando rede Feedforward, um passo à frente e um dia anterior como entrada (NNCO=3)



Os resultados da Figura 103, mostram que existem uma pequena diferença menor que 1 % nos testes realizados em duplicata com as interações diferentes, se apresentando um MAPE na ordem de 39,77 a 39,15% e um RMSE entre 159 a 163 para a DQO de saída usando a rede *feedforward*

Na Figura 104, se encontra o resultado para o período de teste, de uma rede *feedforward* alimentada com dados de dois dias anteriores da DQO de saída alimentando a rede com cinco neurônios na camada oculta (NNCO=5), duas entradas e uma saída.

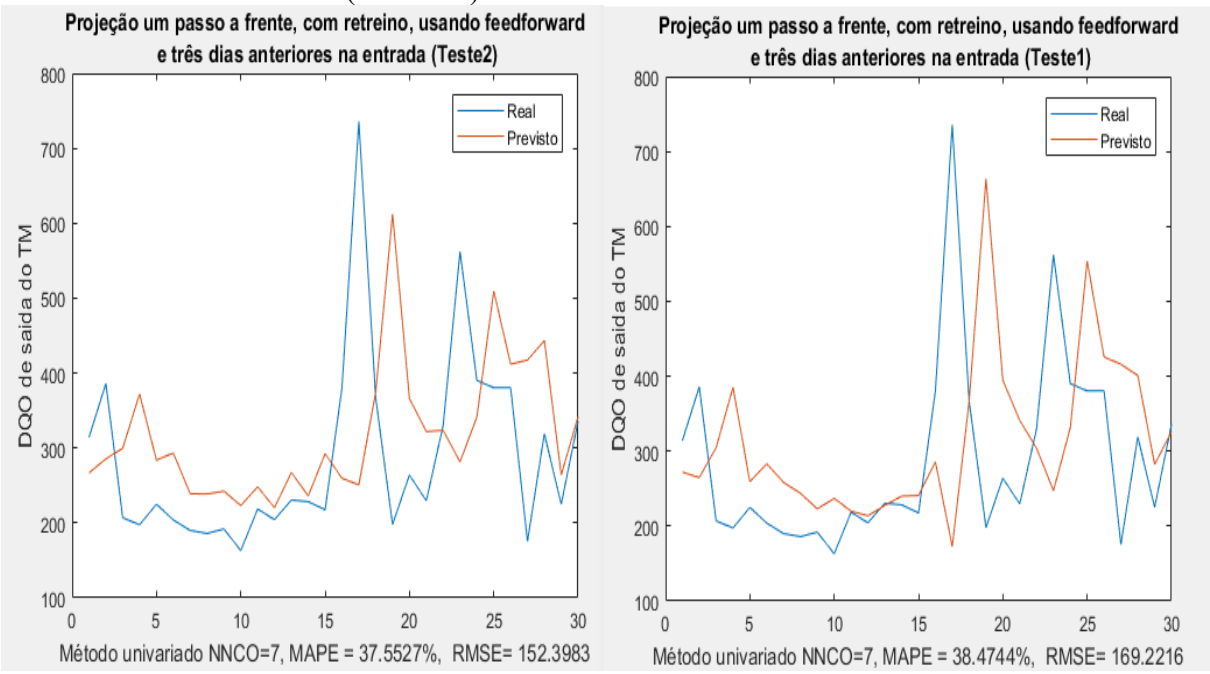
Figura 104 -Tela do Matlab, resultado dos testes usando rede *Feedfrward*, um passo à frente e dois dias anteriores como entrada.(NNCO=5)



Os resultados da Figura 104, mostram que existem uma pequena diferença menor que 1,5 % nos testes realizados entre as interações diferentes, se apresentando um MAPE na ordem de 38 a 41% e um RMSE entre 159 a 164 para a DQO de saída usando a rede *feedforward,* observa-se que dois dias anteriores, neste caso, não apresentou ganho significante em relação aos testes com apenas um dia descrito anteriormente

Na Figura 105, se encontra o resultado para o período de teste, de uma rede feedforward alimentada com dados de três dias anteriores da DQO de saída, alimentando a rede com sete neurônios na camada oculta (NNCO=7), duas entradas e uma saída.

Figura 105 -Tela do Matlab, resultado dos testes usando rede Feedforward, um passo à frente e três dias anteriores como entrada (NNCO=7)



Os resultados da Figura 105, mostram que existem uma pequena diferença menor que 1 % nos testes realizados entre as interações diferentes, se apresentando um MAPE na ordem de 37,55 a 38,47% e um RMSE entre 152 a 169 para a DQO de saída usando a rede *feedforward.* Observa-se que três dias anteriores, neste caso, não apresentou ganho significante em relação aos testes com apenas um dia descrito anteriormente

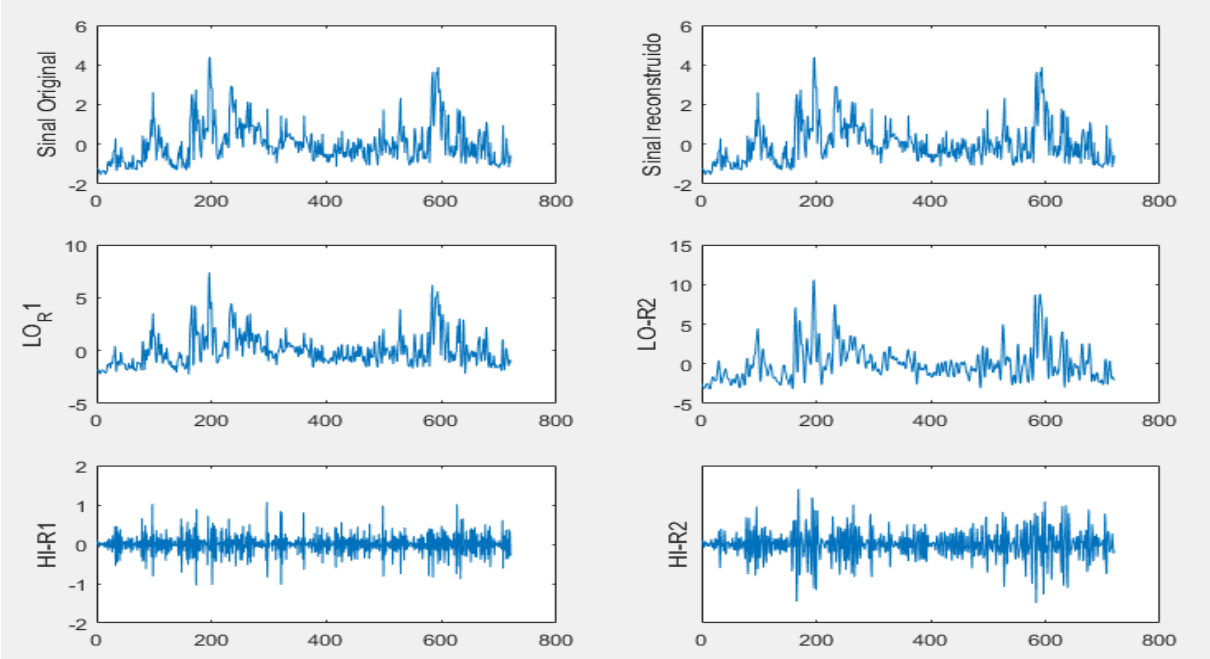
No **Apêndice C,** se encontra o programa usado para previsão um passo à frente, com retreino usando *feedforward* univariado.

4.5.10 Previsão univariada usando a transformada wavelet e rede neural feedforward

Para se verificar se existia ganho na previsão, se realizou dois testes com a transformada wavelet, um teste com dois filtros de baixa e dois filtros de alta frequencia (sinais LO_R1, LO_R2, H1_R1 e H1_R2, da Figura 106, transformando a variável DQO de saída, em uma transformada discreta estacionária, com quatro sinais de entrada na rede feedforward , nove

neuronios na camada oculta e quatro saídas, com a mesma trasformada para a variável um passo à frente.

Figura 106 -Decomposição e reconstrução do sinal original usando a transformada wavelet aplicada a DQO de saída do TM, usando 2 filtros de alta frequência e 2 filtros de baixa frequência



Os sinais de saída um passo à frente foram recompostos e se obteve os resultados da Figura 108.

Se usou o *soft* MATLAB, ao qual se conseguiu uma correlação maior na rede *feedforward* (Figura 107), do que com as usadas com método, sem usar nenhuma transformada, o programa em MATLAB usado se encontra descrito no **Apêndice D**.

Após treinada e retreinada a rede para cada dia de teste, a saída da rede obtida foi recomposta e comparado os resultados entre o real e o previsto, conforme demonstrado na Figura 108.

Figura 107-Correlação e histograma de rede Feedforward, com as variáveis transformadas wavelet com dois filtros de baixa e dois filtros de alta frequência

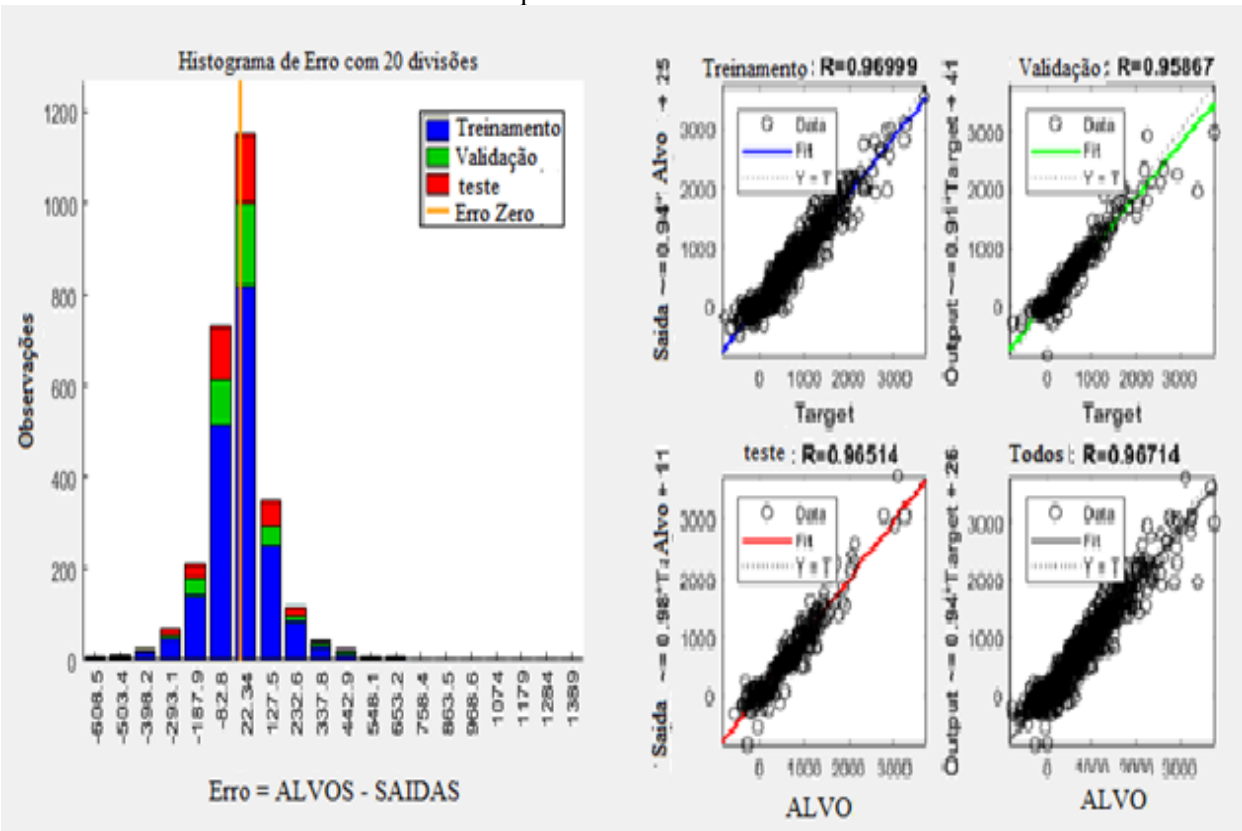
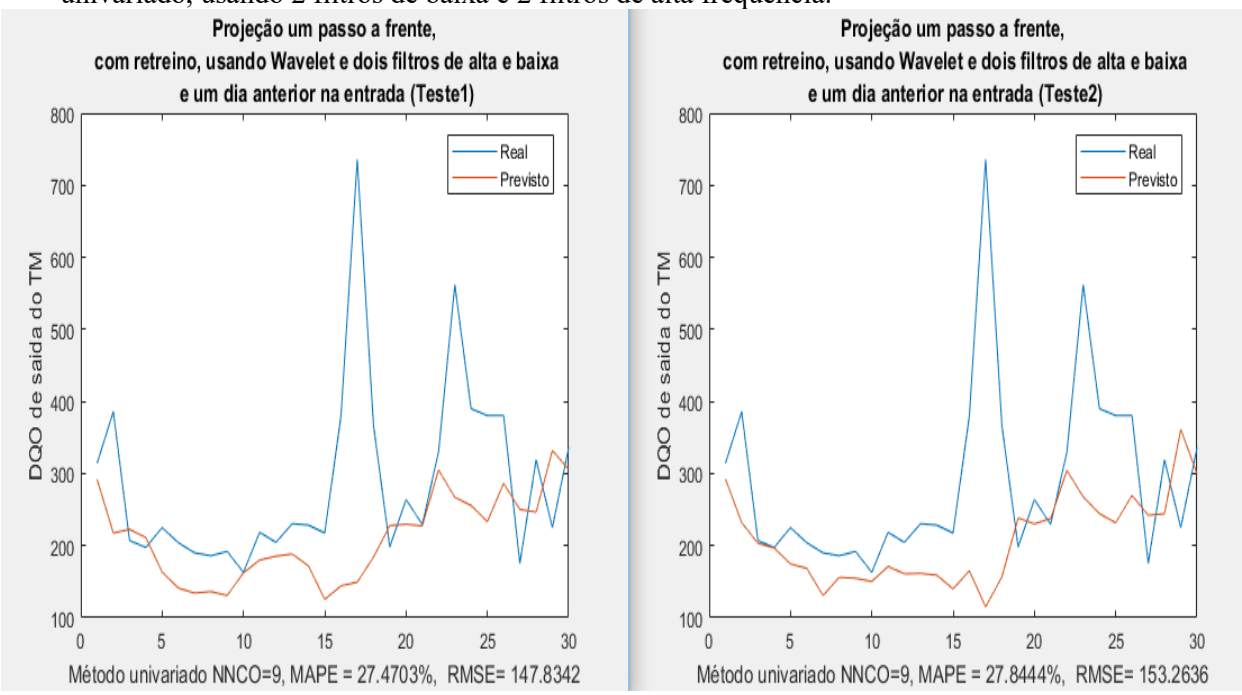


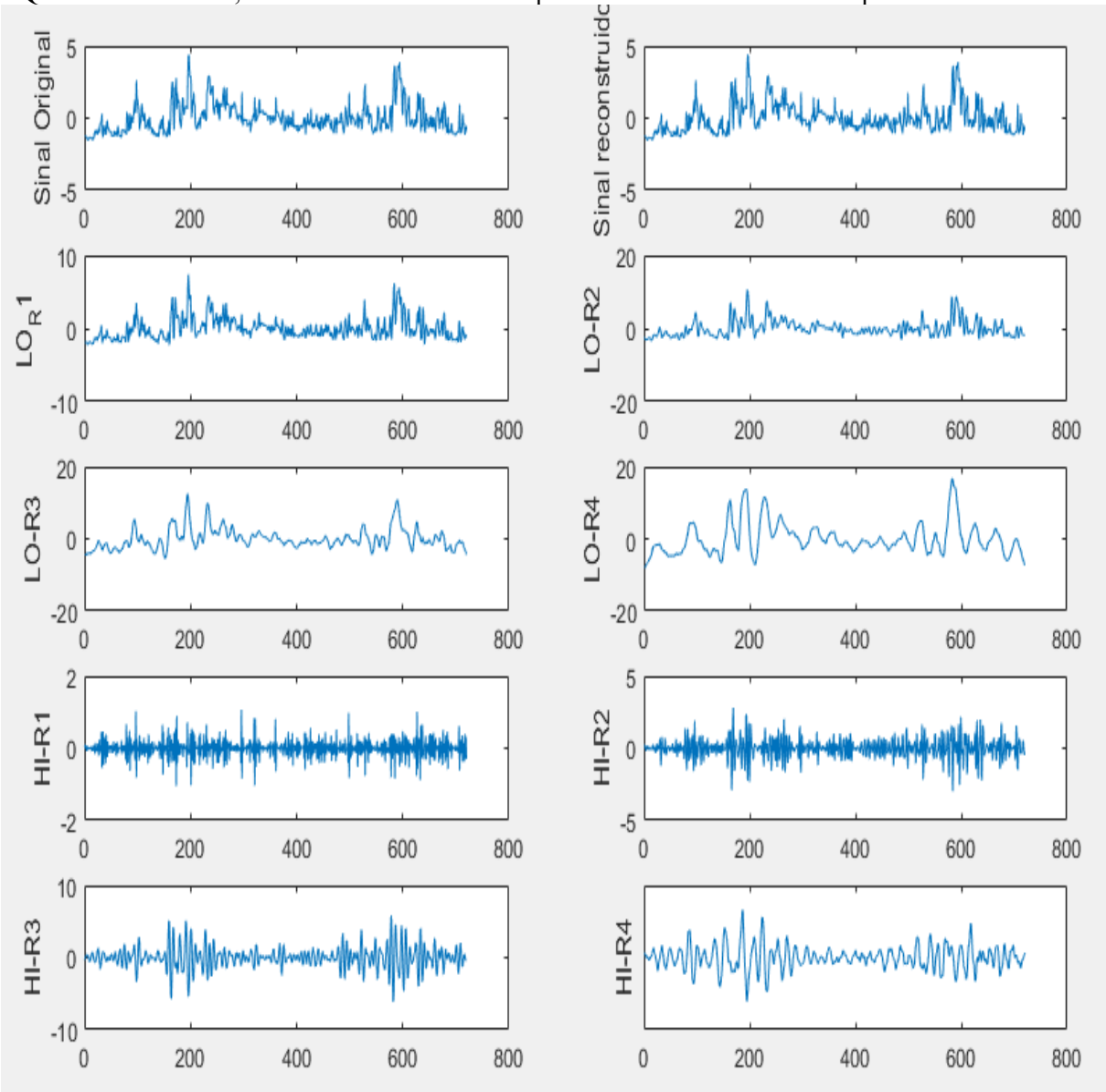
Figura 108 -Resultado dos testes um passo à frente, usando transformada wavelet, em método univariado, usando 2 filtros de baixa e 2 filtros de alta frequência.



Observa-se na Figura 108, que os resultados usando a transformada wavelet associados a rede neural apresentaram um valor de MAPE (27,47 e 27,84%) de previsão melhor que que os da rede neural sem uso de wavelet e também melhor que o SVM, considerando este período de teste.

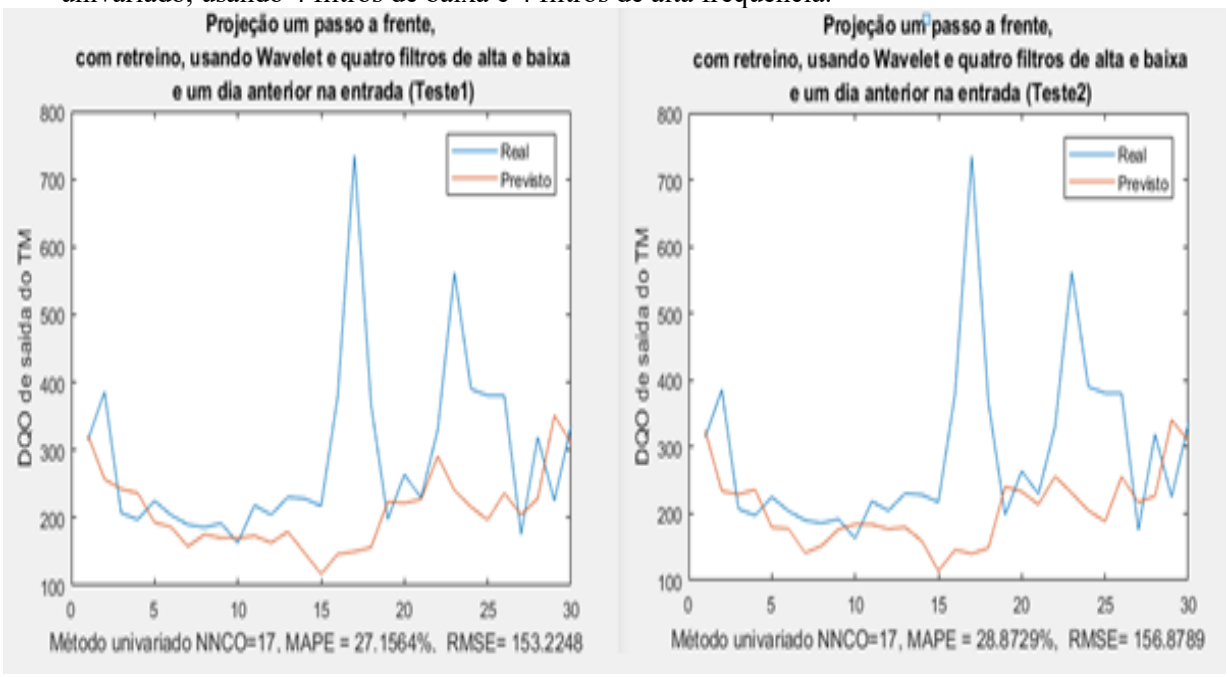
Para se verificar se haveria ganho na previsão, se realizou mais um teste da mesma forma que o anterior, só que usando quatro filtros de baixa e quatro filtros de alta frequência (conforme demonstrado na Figura 109), se obtendo os resultados da Figura 109.

Figura 109 -Decomposição e reconstrução do sinal original usando a transformada wavelet aplicada a DQO de saída do TM, usando 4 filtros de alta frequência e 4 filtros de baixa frequência.



Observa-se na Figura 110, que os resultados usando a transformada *Wavelet* (4 filtros de alta e 4 filtros de baixa frequência), associados a rede neural apresentaram um valor de MAPE (27,15 e 28,87%) de previsão. Este resultado se mostra melhor, que os da rede neural sem uso de wavelet e também melhor que o SVM, considerando este período de teste, porém na mesma faixa dos resultados usando *Wavelet* com apenas 2 filtros de alta e 2 filtros de baixa frequência.

Figura 110 -Resultado dos testes um passo à frente, usando transformada wavelet, em método univariado, usando 4 filtros de baixa e 4 filtros de alta frequência.



4.6 PREVISÃO MULTIVARIADA, UM PASSO À FRENTE, DA DQO DE SAÍDA.

Neste método as variáveis sofrem normalização, como descrito na seção 2.7.1, com o uso de todas as variáveis como dados entrada (DQO_EEQ($mg.L^{-1}$), Vazão_TM(m^3/h), Carga Orgânica Aplicada _TM ($kgDQO.dia^{-1}$), pH da Alimentação_TM, Alcalinidade_TM no Ponto - 02_TM, AGV no Ponto _TM- 02($mg.L^{-1}$), Relação de AGV x Alcalinidade_TM, p H do Ponto - 02_TM, Temperatura do Reator_TM (ºC), DQO de Saída_TM(mgO_2. L^{-1}), relacionando com a variável objeto do valor futuro que se deseja determinar (neste caso se escolheu a DQO de saída do tanque de metanização), podendo estas ser suavizadas ou transformadas por métodos diversos como *Wavelet*. A variável objeto, um dia a frente, é calculada pelo método e

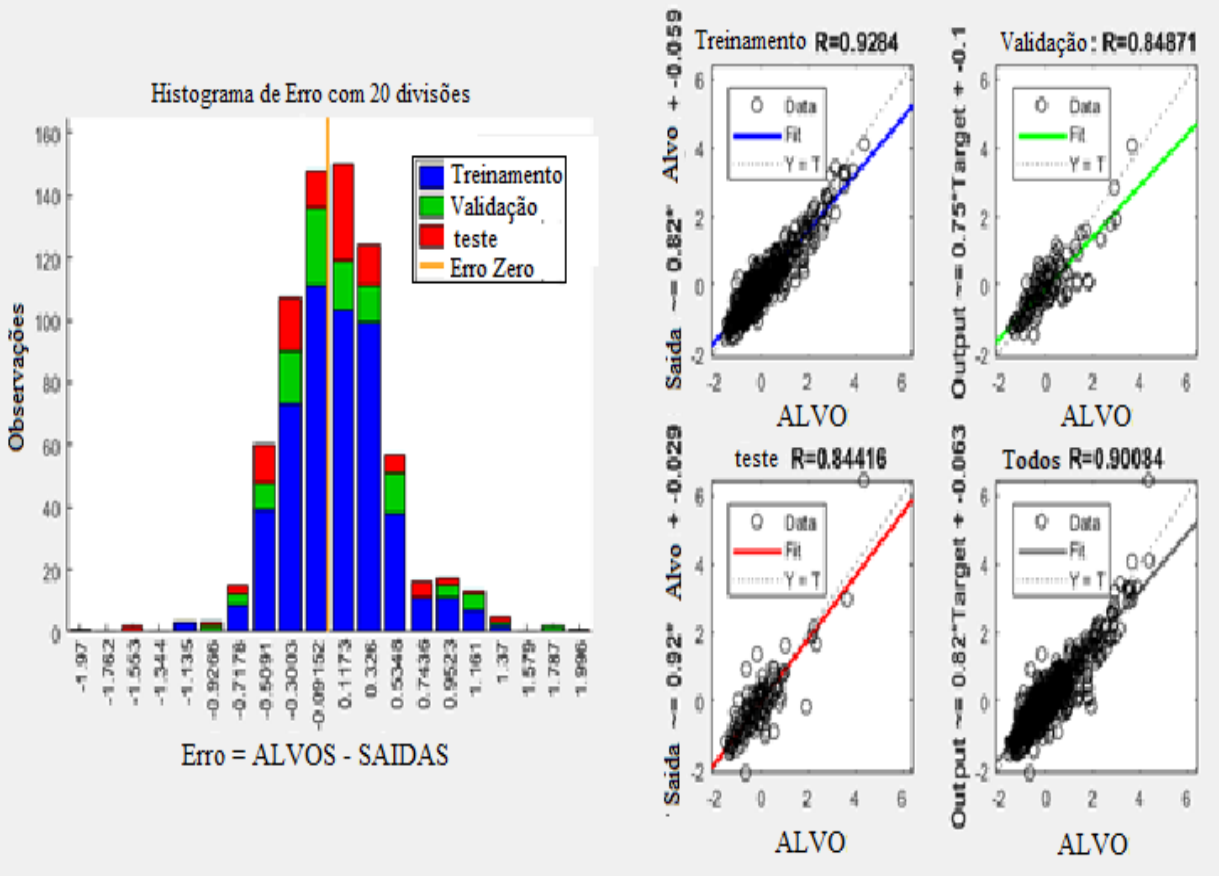
retransformada para seu valor original, ao qual se determina de forma idêntica o RMSE e o MAPE para comparação dos métodos.

4.6.1 Previsão Multivariada usando apenas feedforward e um dia anterior.

Todas as variáveis de controle sofreram normalização pelo método Z-score e foi testado uma rede neural *feedforward (10,21,1),* com 10 valores das análises do dia anterior, 2N+1 neuronios na camada oculta (NNCO) ou seja 21 neurônios e uma saída para previsão da DQO um dia à frente, para se comparar se existia ganho de correlação e redução dos erros em relação ao método feedforward com apenas uma dimensão, se obteve as curvas de correlação e histograma de erro da rede (Figura 111). O programa elaborado em MATLAB para previsão está no **Apêndice C**.

Na Figura 111, se observa uma correlação acima de 90% entre os resultados e o período de treinamento, se obtendo correlações de 84% nos dados usados para teste e validação da rede neural, assim como, o histograma de erro com 20 divisões, apresenta uma variação normal dos resultados, oscilando entre valores positivos e negativos das diferenças encontradas entre o real e o método, porém apresentando resultado de dispersão maiores que o método univariado usando *wavelet*, mostrado na Figura 107.

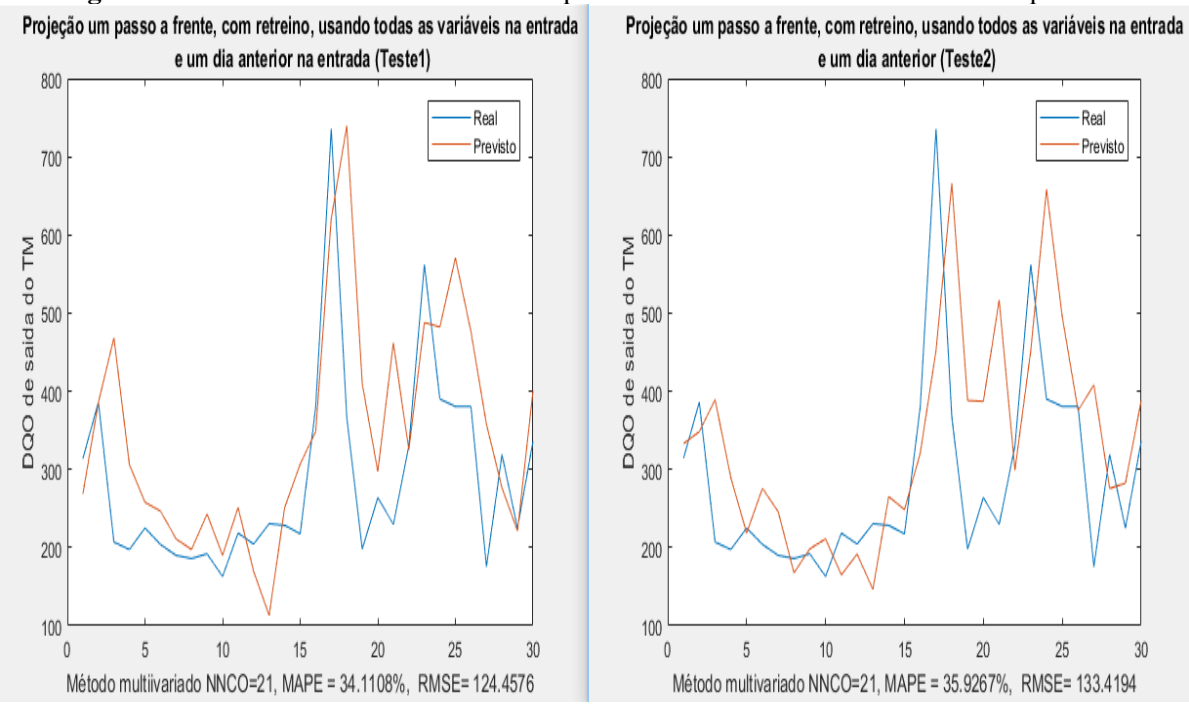
Figura 111 -Curvas de correlação e histograma de erro da rede neural



Na Figura 112, está o resultado deste teste com a previsão da DQO de saída e seus respectivos erros.

Observa-se na Figura 112, que usando um dia antes em redes *feedforward* em método multivariado com um dia antes e todas as variáveis na entrada, que o MAPE (34,11 a 35,92%) e o RMSE (124,45 a 133,41), valores que se aproximam dos resultados univariados com a vantagem dos métodos multivariados, poder associar as variáveis de entradas aos resultados de saída, permitindo deste modo realizar simulações do que se espera em resultados de eficiência de remoção da DQO, com as mudanças dessas variáveis.

Figura 112 -Resultado dos testes usando apenas feedforward e um dia anterior na previsão



4.6.2 Previsão multivariada usando transformada wavelet.

Neste método, cada uma das 10 variáveis de entrada foi decomposta utilizando a transformada wavelet. Para efeito de comparação foram feitos dois testes, um usando dois filtros de alta frequência e dois filtros de baixa frequência (o que leva a uma decomposição com 40 entradas na rede neural, 81 neurônios na camada oculta e 4 saidas da DQO de saída decomposta) , e o outro quatro filtros de alta e quatro filtros de baixa (o que leva a uma decomposição com 80 dados na entrada, 161 neuronios na camada oculta (NNCO) e oito dados de saída da DQO decomposta), sendo transformadas entradas e saídas da rede neural, que foram recompostas após a rede neural processar os dados e converte-los no período de teste aos valores dimensionais reais para comparação. Na Figura 113 e 114 estão os resultados, para o teste com 2 filtros de alta e baixa e na Figura 115 e 116, os resultados para 4 filtros de alta e 4 filtros de baixa. No **Apêndice E**, se encontra o programa usado em MATLAB, para previsão com *wavelet* e rede neural *Feedforward*

Figura 113 -Resultado do período de teste da rede neural suavizada com transformada *wavelet*, com 2 filtros de baixa e 2 filtros de alta, e 81 neurônios na camada oculta (teste1)

Figura 114 -Resultado do período de testes da rede neural suavizada com transformada wavelet com 2 filtros de baixa e 2 filtros de alta e 81 neurônios na camada oculta (teste2)

Observa-se nos testes da Figura 113 e 114, que o uso de filtros *Wavelets* associados a redes neurais reduzem o MAPE (30,03% a 32,64), mas o RMSE (125,18 a 127,18) praticamente não se altera, em relação ao uso de redes neurais, sem o uso dos filtros *Wavelets*.

Figura 115 -Resultado do período de teste da rede neural, suavizada com transformada wavelet com 4 filtros de baixa e 4 filtros de alta frequência e 161 neurônios na camada oculta (teste 1)

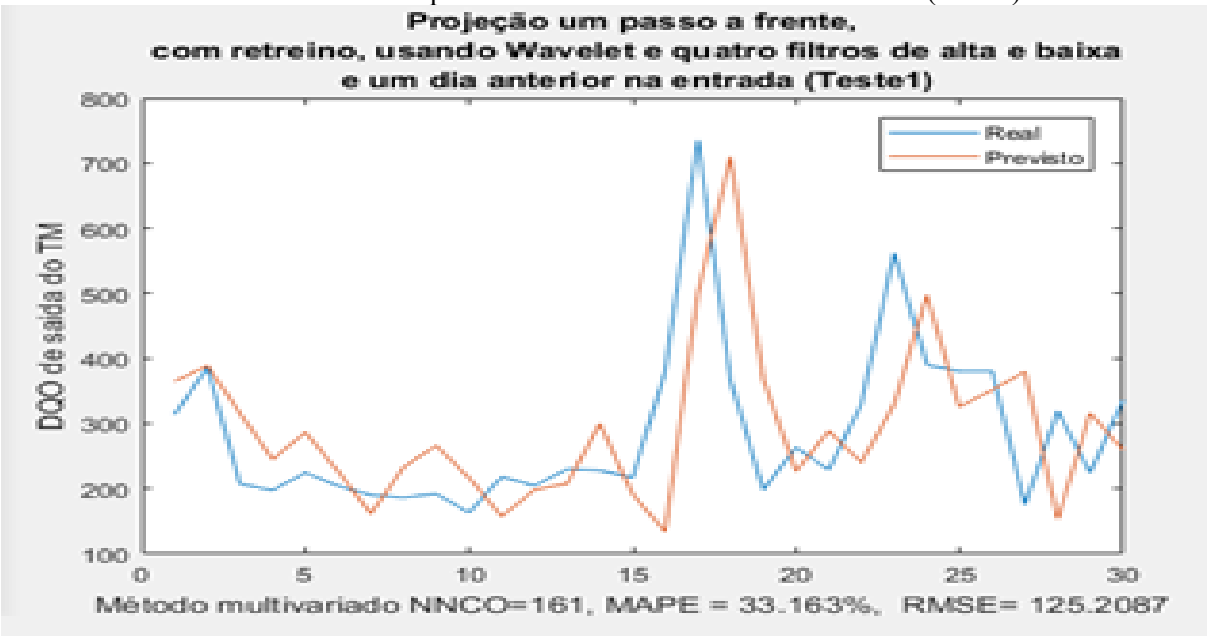
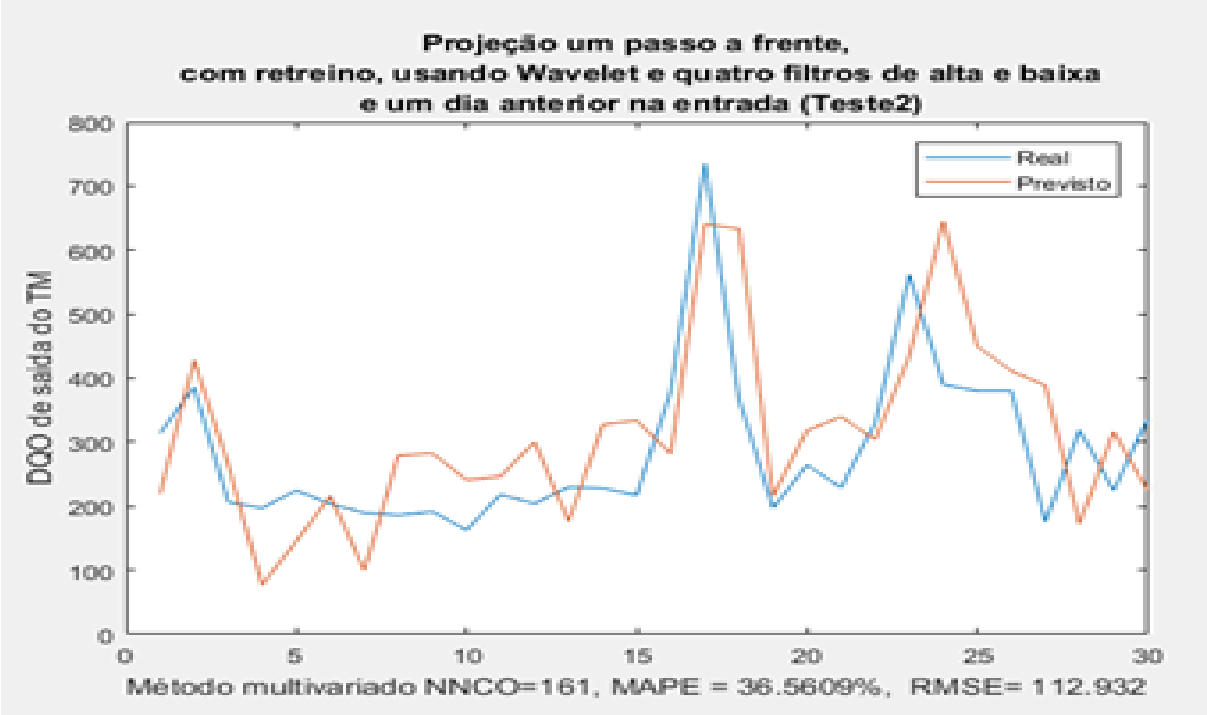


Figura 116 -Resultado da regressão e período de teste da rede neural, suavizada com transformada wavelet com 4 filtros de baixa e 4 filtros de alta frequência e 161 neurônios na camada oculta (teste 2)



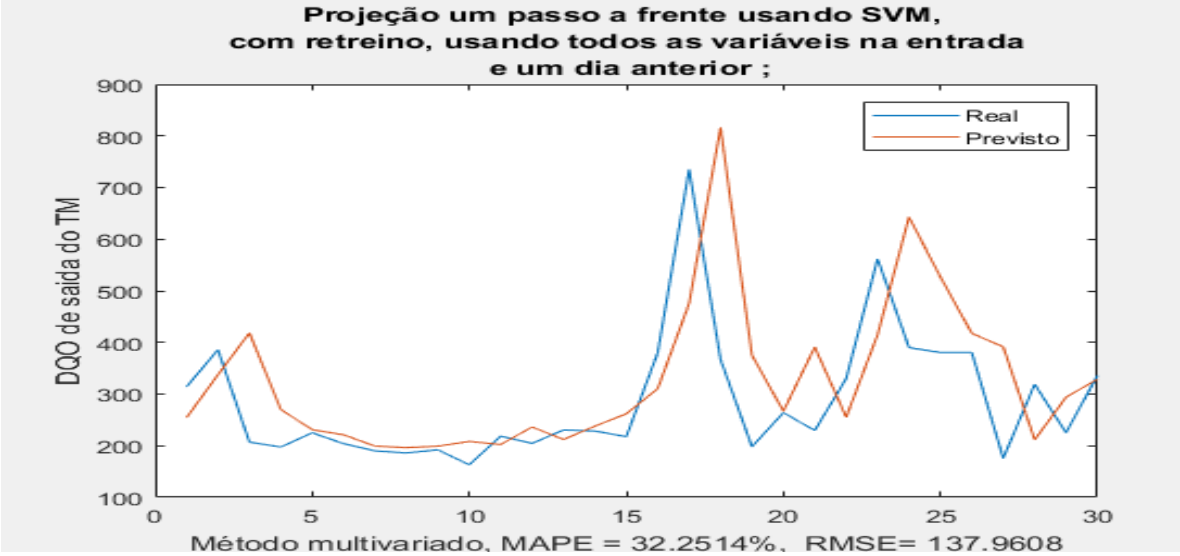
Observa-se nos testes 115 e 116, que o uso de mais filtros *Wavelets* (4 de baixa e 4 de alta e o aumento dos números de neurônios) associados a redes neurais não melhoram os

resultados encontrados para o MAPE (33,16 a 36,56%), e o RMSE (112,93 a 125,20) praticamente não se altera, em relação ao uso de redes neurais com 2 filtros *Wavelets.*

4.6.3 Previsão multivariada um passo à frente, usando SVM

Na Figura 117, temos uma previsão, um passo à frente, usando todas as dez variáveis na entrada, e usando o vetor suporte de máquina. Observa-se que este método teve erro menor que a rede feedforward, com os mesmos dados de entrada com a vantagem de reprodução dos resultados, sem precisar testar os resultados para número de neurônios diversos, e também resultados equivalentes ao das previsões usando filtros wavelets com MAPE de 32,25 e RMSE de 137,96)

Figura 117 -Previsão, um passo à frente, usando todas as dez variáveis na entrada, e usando o vetor suporte de maquina



No **Apêndice F,** esta apresentado o programa usado na previsão um passo à frente com retreino usando modelo *Support Vector Machine Regression* , dados normalizados.

4.7 RESUMO DOS TESTES UM PASSO À FRENTE UNIDIMENSIONAL E MULTIDIMENSIONAL

Na Tabela 17, tem o resumo dos testes realizados, utilizando os métodos univariados discutidos e suporte vetor de máquina, observando-se que todos os resultados deram valores abaixo de um desvio padrão da DQO de saída, exceto o método da média acumulada, com melhor resultado de previsão, para o método da média móvel simples para N=10, N=2 o método AED de Holt e o ARIMA(2,2,2).

Tabela 17-Métodos univariados clássicos que não usam rede neural

Para o período de testes de 30 dias	**Erros**	
Métodos univariados clássicos que não usam rede neural	**MAPE %**	**RMSE**
Média Acumulada	95,07	233,03
Média Móvel Simples N=2 (MMS N=2)	29,33	138,70
Média Móvel Simples N=3 (MMS N=3)	30,42	137,10
Média Móvel Simples N=4 (MMS N=4)	31,46	138,50
Média Móvel Simples N=10 (MMS N=10)	26,53	126,56
Média Móvel Simples N=15 (MMS N=15)	30,04	132,89
Média Móvel Dupla (MMD)	45,33	193,12
Amortecimento Exponencial Simples (AES)	29,33	131,58
Amortecimento Exponencial Duplo (AED Holt)	27,74	123,62
Amortecimento Exponencial Duplo (AED Brown)	31,90	138,78
Suporte Vetor de Maquina (SVM)-1 dia antes	31,31	130,13
Suporte Vetor de Maquina (SVM)-2 dias antes	42,56	154,20
Suporte Vetor de Maquina (SVM)-3 dias antes	39,78	148,48
ARIMA (1,1,1)	31,00	123,96
ARIMA (2,1,1)	30,31	123,37
ARIMA (2,1,2)	29,98	122,50
ARIMA (2,2,2)	28,42	122,36

Na Tabela 18, estão os resultados dos testes com rede neural *feedforward*, para um, dois e três dias de atraso na série temporal, observa-se que neste caso os valores são maiores que os métodos clássicos e o uso de um, dois ou três dias não altera significativamente os resultados, atingindo a rede seus limites, considerando a estrutura de neurônios usada.

Em função de ser um processo interativo, os resultados das redes neurais não possuem reprodutibilidade 100%, mas a repetição dos testes, tem resultado em valores próximos ao observados nos outros testes.

Na Tabela 19, temos os testes realizados com transformada wavelet e rede neural feedforward, com dois filtros de alta e dois filtros de baixa e quatro filtros de alta e quatro filtros de baixa, em métodos univariados. Os resultados da Tabela 19, mostram que a previsão fica em uma faixa de variação entre 27,15 a 28,87 % para o MAPE e 147,83 a 156,87 para o RMSE usando estas técnicas, com resultados superiores aos obtidos nos métodos feedforward considerando um dia, dois e três de atraso, listados na Tabela 18.

Tabela 18 - Métodos univariados com redes neurais feedforward

Métodos univariados com redes neurais feedforward	**Erros**		**Características da rede neural**		
	MAPE %	**RMSE**	Entradas	NNCO	saida
1 dia anterior de entrada	39,77	163,16	1	3	1
1 dia anterior de entrada	39,15	159,38	1	3	1
2 dias anteriores de entrada	40,77	164,45	2	5	1
2 dias anteriores de entrada	38,54	159,56	2	5	1
3 dias anteriores de entrada	37,55	152,39	3	7	1
3 dias anteriores de entrada	38,47	169,22	3	7	1

Tabela 19 - Métodos univariados com transformada wavelet e feedforward

Métodos univariados combinando	**Erros**		**Características da rede neural**		
Feedforward +Wavelet	**MAPE %**	**RMSE**	Entradas	NNCO	saida
2 filtros de alta e 2 de baixa	27,47	147,83	4	9	4
2 filtros de alta e 2 de baixa	27,84	153,26	4	9	4
4 filtros de alta e 4 de baixa	27,15	153,22	8	17	8
4 filtros de alta e 4 de baixa	28,87	156,87	8	17	8

Na Tabela 20, se encontram os resultados dos testes multivariados.

Os resultados dos testes multivariados, apresentaram previsões com margem de erro um pouco maior que o dos testes unidimensionalizados, possivelmente por incluir na previsão variáveis de entrada com oscilações maiores que a variável alvo. Porém os valores da transformada wavelet em apenas dois filtros de alta e dois filtros de baixa frequência, chegam

próximos aos valores dos métodos unidimensionalizados, com a vantagem da associação do resultado as outras variáveis de controle da ETE.

Tabela 20 –Métodos multivariados

	Erros		Caracteristicas da rede neural		
Métodos multivariados	**MAPE %**	**RMSE**	Entradas	NNCO	saida
Feedforward e um dia anterior.	34,11	124,45	10	21	1
Feedforward e um dia anterior.	35,92	133,41	10	21	1
Feedforward+wavelet+ 2 filtros de alta e 2 de baixa para cada variável	32,64	127,28	40	81	4
Feedforward+wavelet+ 2 filtros de alta e 2 de baixa para cada variável	30,03	125,18	40	81	4
Feedforward+wavelet+ 4 filtros de alta e 4 de baixa para cada variável	33,16	125,20	80	161	8
Feedforward+wavelet+ 4 filtros de alta e 4 de baixa para cada variável	36,56	112,93	80	161	8
SVM e um dia anterior	32,25	137,96	NA	NA	NA

Nota: em todos os métodos usando rede neural (univariado ou multivariado) se utilizou nas camadas ocultas a tangente sigmoide como função de ativação e a função linear na saída.

4.8 CLASSIFICAÇÃO DE DADOS UM PASSO A FRENTE, USANDO REDES NEURAIS CONVOLUTIVAS

O programa usado para desenvolver esta classificação em MATLAB, versão 2017a, está no **Apêndice B**. Se testou matrizes com um, dois, cinco, dez, quinze e vinte e cinco dias de atraso para se verificar qual apresentava melhor resultado de classificação, ou acurácia. Além da aferição de acurácia dos pequenos lotes feito no MATLAB, foi testado em cada simulação, lotes a parte, com trinta previsões para ver a acurácia, com o resultado da Tabela 21

Tabela 21 - Resultado dos testes com as redes CNN

Dias de atraso usado na figura	Acuracidade %
1	56,67
2	70
15	86,67
20	73,3
35	70

Observou-se durante os testes que quanto maior o número de dias de atraso mais rápido havia convergência da rede neural convolutiva, porém um número maior de dias reduz o número disponível para treinamento, todos os testes chegaram a acurácia de 100% para os pequenos lotes de teste da rede (pequenos testes realizados pelo programa que reserva lotes menores para testar acurácia), e foram feitas quinhentas interações

Observou-se também que pequenos lotes têm probabilidade de chegar a 100% de acurácia, porém resultados com a acurácia acima de 80%, podem se tornar muito próximo dos valores reais se houver mais testes consecutivos, para a mesma figura representante de dados transformados (ver seção 3.6, exemplo na Figura 46), retreinando a rede, ao qual a probabilidade, seria dada pelo número maior de testes realizados com a mesma figura representando os dados transformados.

A probabilidade de acerto se realizando três analises consecutivas e tomando-se o número maior de resultados iguais, com um modelo que responde a acurácia de 80%, seria de 99.2% de acerto, se realizando mais analises teremos valores próximos de 100%.

Assim, a previsão por rede convolutiva pode indicar valores de classificação com grande precisão.

A Tabela 22, faz um resumo dos resultados gerados no MATLAB, para previsão de 15 dias, observa-se uma acurácia de 86,67%, considerando quinze dias anteriores gerando a Figura com 15 colunas e 10 análises.

Tabela 22-Tabela com resumo dos resultados gerados no MATLAB, para previsão de 15 dias

Desce 408
Sobe 302

Treinamento em CPU única
Inicializando a normalização da imagem

Epoca	Iteracão	Tempo decorrido (segundos)	Lote pequeno perda	Lote pequeno acurácia	taxa de aprendizado base
1	1	0.08	0.7083	47.66%	1.00e-04
13	50	2.71	0.4491	77.34%	1.00e-04
25	100	5.22	0.3633	82.81%	1.00e-04
38	150	7.72	0.2465	91.41%	1.00e-04
50	200	10.22	0.2242	92.97%	1.00e-04
63	250	12.89	0.1615	94.53%	1.00e-04
75	300	15.39	0.1482	95.31%	1.00e-04
88	350	18.58	0.1130	96.88%	1.00e-04
100	400	21.11	0.1023	96.88%	1.00e-04
113	450	23.61	0.0798	99.22%	1.00e-04
125	500	26.12	0.0728	98.44%	1.00e-04
138	550	28.63	0.0578	99.22%	1.00e-04
150	600	31.14	0.0531	100.00%	1.00e-04
163	650	33.65	0.0431	99.22%	1.00e-04
175	700	36.15	0.0399	100.00%	1.00e-04
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
450	1800	96.79	0.0067	100.00%	1.00e-04
463	1850	100.04	0.0064	100.00%	1.00e-04
475	1900	102.96	0.0061	100.00%	1.00e-04
488	1950	105.50	0.0059	100.00%	1.00e-04
500	2000	108.01	0.0056	100.00%	1.00e-04

acurácia = **0,8667**

4.8 CLASSIFICAÇÃO DE DADOS, UM PASSO A FRENTE, USANDO VETOR SUPORTE DE MÁQUINA.

Uma classificação usando vetor suporte de máquina (SVM), empregando uma função Kernel linear, foi realizada para se comparar os resultados de classificação. O resultado está descrito na Tabela 23, ao qual os dados da previsão, que não tiveram valores iguais aos reais,

estão com um asterisco, a classificação usada foi a mesma da CNN, ou seja, classifica se a variável futura estará acima (‘sobe’) ou abaixo da média (‘desce’). Observou-se que os valores obtidos de acurácia se igualavam aos obtidos na CNN de 86,67%, usando uma Figura de 15 dias de atraso para as dez variáveis usadas. Nos testes com a SVM se usou apenas os resultados dos valores unidimensionalizados de variáveis com um dia de atraso.

Tabela 23 -Resultado dos testes de classificação usando SVM, com um dia de atraso.

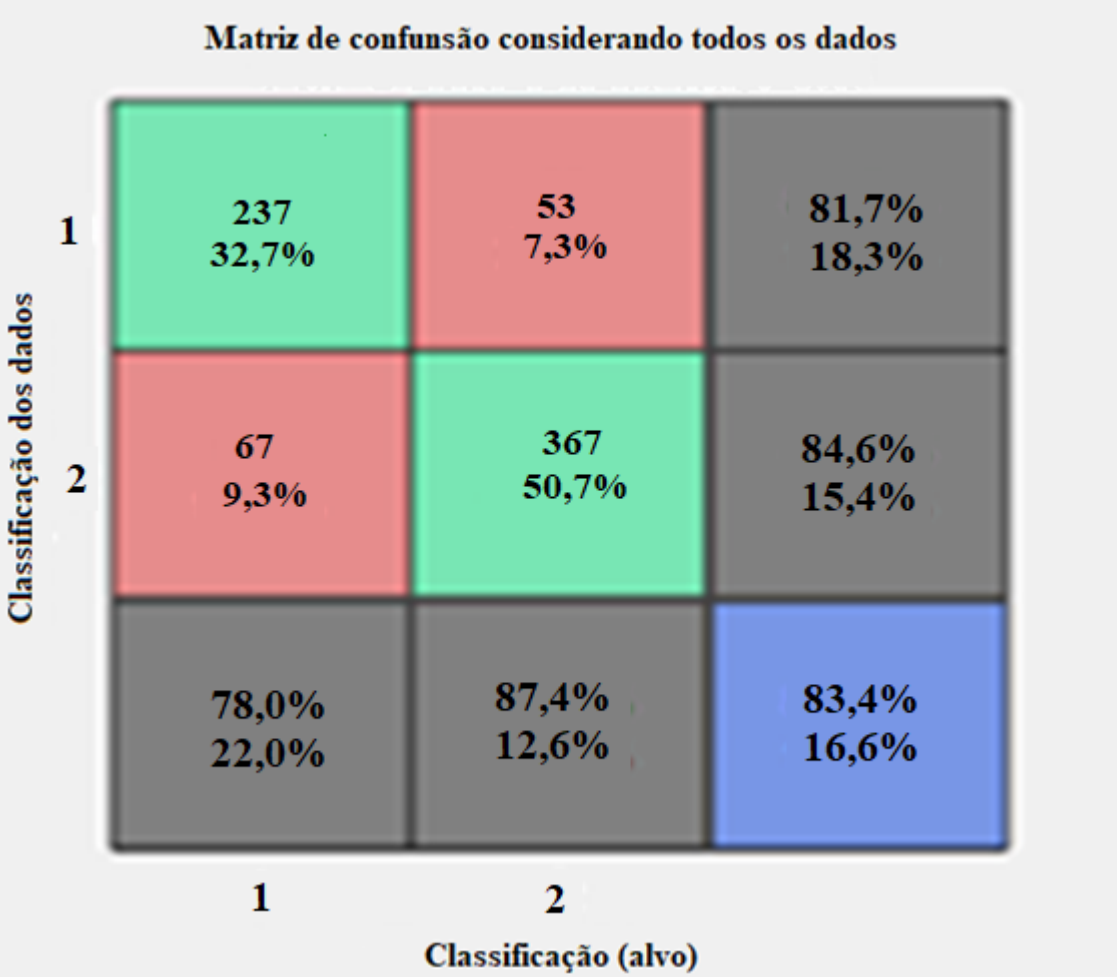
teste	Real	previsão	teste	Real	previsão
1	'desce'	'desce'	16	'desce'	'desce'
2	'desce'	'desce'	17	'sobe'	'sobe'
3	'desce'	'desce'	• 18	'desce'	'sobe'
4	'desce'	'desce'	19	'desce'	'desce'
5	'desce'	'desce'	20	'desce'	'desce'
6	'desce'	'desce'	21	'desce'	'desce'
7	'desce'	'desce'	22	'desce'	'desce'
8	'desce'	'desce'	• 23	'sobe'	'desce'
9	'desce'	'desce'	• 24	'desce'	'sobe'
10	'desce'	'desce'	• 25	'desce'	'sobe'
11	'desce'	'desce'	26	'desce'	'desce'
12	'desce'	'desce'	27	'desce'	'desce'
13	'desce'	'desce'	28	'desce'	'desce'
14	'desce'	'desce'	29	'desce'	'desce'
15	'desce'	'desce'	30	'desce'	'desce'
Acurácia =	86,67%				

4.8 CLASSIFICAÇÃO DE DADOS, UM PASSO A FRENTE, USANDO REDE NEURAL FEEDFORWARD.

Na Figura 118, está a matriz de confusão e a tela do MATLAB com a configuração de uma rede neural, para classificar se a previsão futura um passo à frente, estaria acima ou abaixo da média. Para isto a rede foi configurada com 10 neurônios na entrada, com as informações relativas as análises unidimensionalizadas, 21 neurônios na camada oculta e 2 neurônios na saída ajustados, com saída [1 0], para valores acima da média e [0 1], para valores abaixo da média. Obteve-se como resultado dos testes a matriz de confusão, que mostra um resultado de 83,4% de

acertos ao se utilizar o método. Nos quadros verdes estão representados o numero de amostra com acertos e o percentual deste numero em relação ao tamanho total das amostras (ou seja 237+57+67+367= 724 dados totais), nos vermelhos os erros encontrados na previsão, no marrom nas linhas e nas colunas o percentual de acertos e erros por lotes e no azul representa o percentual de acerto e erros total computado (acertos = 237+367, erros = 53+67, amostra total=724)

Figura 118-Matriz de confusão e configuração da rede neural para classificação dos dados



Observa-se que apesar de não ter 100% de confiabilidade, os métodos de classificação podem em conjunto ou de forma repetida aumentar o diagnóstico de previsão, considerando as probabilidades alcançadas nos resultados das redes CNN, SVM e NN maior que 80%.

4.9 COMPARAÇÃO DOS RESULTADOS COM OS DADOS ENCONTRADOS EM LITERATURA

Cada estação de Tratamento de efluentes tem suas características diferenciadas das outras estações, a metodologia pode ser usada em todas as estações, porém os resultados vão variar em função da variável alvo envolvida, dos métodos usados e dos processos.

Com relação aos erros e desvios padrão, pode se ter uma ordem de grandeza comparando os resultados do trabalho de Bousfield, Silva e Chapuis (2011) da Tabela 2 (seção 2.4) , ao qual os resultados de uma ETE de lodos ativados mostra parâmetros com desvio padrão variando entre 2,41 a 88,71%, com o da Tabela 13 (resultado deste trabalho) , ao qual os resultados variam de 2,6% a 67,79%.

Verifica-se que a ordem de grandeza varia de acordo com a variável envolvida e o processo de tratamento de efluentes.

Pinto (2014), selecionou 14 de 20 variáveis em uma estação de tratamento de efluentes com reator anaeróbio de fluxo ascendente, cujos principais objetivos se prenderam com a identificação da RNA com maior capacidade para prever o biogás produzido e a seleção das variáveis mais adequadas para a modelação dos processos anaeróbios. O treino e teste de redes envolveu a realização de 266 ensaios, a partir dos quais se identificaram as cinco melhores redes para previsão, encontrando resultados do percentual de dados não desejados na previsão nos testes entre 14,29 a 50,81% e erro médio quadrático de 157,91 a 243,50. Quanto maior o numero de variáveis na entrada da rede menor foi seu erro.

Neto (2015), avaliou um sistema de rede neural acompanhando o coeficiente de correlação (R), do treinamento de uma rede neural. Este coeficiente apesar de dar valores de correlação alto, especialmente aumentando o número de neurônios da rede, pode dar extrapolação de dados ao se realizar uma previsão, conforme descrito na seção 2.4.6.6, a avaliação dos erros absolutos e relativos da previsão em testes independentes pode gerar resultados de maior confiabilidade.

Se considerarmos o trabalho de Vida et.al (2016) que usou 140 amostras em treinamento de rede neural para prever as análises de Carbono Organico Total (TOC), obtendo resultados, de RMSE entre 0,01 e 0,045, em em um reator piloto operado durante 250 dias, em previsão usando

redes neurais, veremos que se usou apenas o RMSE como parâmetro de avaliação e não o MAPE. Valores medidos em escalas baixas, podem resultar em RMSE em iguais proporções, porém se faz necessário períodos maiores de análises, para se evitar coletas em períodos tendenciosos e usar o erro relativo (MAPE), para se ter uma noção melhor do erro encontrado.

Uma variável de grandeza maior, normalmente terá um erro relativo menor, que a mesma variável com grandezas menores, usando os mesmos métodos analíticos, pois os ruídos dos métodos são próximos. Assim se espera que uma DQO de entrada tenha menor erro relativo do que a DQO de saída.

Uma variável controlada, como os sólidos totais de uma ETE com lodos ativados, que pode ser regulada pelo descarte de lodo, tem variações menores no processo e apresentam uma previsão melhor do que a DBO_5 de saída, pois esta variável depende de muito mais ajustes, que as outras.

Fernandes (2014) estudou as RNA como uma técnica de previsão da produção de metano nos digestores anaeróbios da ETAR da Guia localizada em Cascais, Portugal. Os dados operacionais da ETAR durante um período de 12 meses foram recolhidos e utilizados na análise. O estudo considerou o efeito dos seguintes parâmetros operacionais dos digestores: vazão de entrada nos digestores, vazão de entrada de lamas que chegam à Fase Sólida dos tratamentos da ETAR em estudo, a percentagem e a carga de sólidos totais de entrada nos digestores. Para a previsão da produção de metano foi construído um modelo com RNA, com uma camada escondida, 30 neurônios e com um limite máximo de 600 iterações. O treino e teste do modelo com RNA foi efectuado com dados relativos aos primeiros nove meses. Na construção do modelo, o desempenho obtido no conjunto de teste foi de 9,84% de erro normalizado médio e um coeficiente de determinação (R2) médio de 0,86. Posteriormente, o modelo foi validado com dados que não foram utilizados durante as fases de treino e teste da construção do modelo (últimos três meses dos dados), demonstrando a eficácia do modelo para prever a produção de metano, com um R2 de 0;79 e um erro normalizado de 11,6%.

Se compararmos os resultados das previsões com ordem de grandeza apresentado por Wei (2013), mostrados nas Tabelas 8 e 10 (seção 2.4), verificamos que as previsões um dia a frente para os sólidos suspensos voláteis de uma ETE (variável controlada), mostraram

resultados que variam de 18,29 até 26,46%, e na Tabela 10, uma previsão um passo à frente usando redes neurais e a mesma metodologia usada neste trabalho, encontrou resultados de erro relativo de 21,85%. Os resultados obtidos nas Tabelas 16, 17, 18 e 19 (seção 4.7) deste trabalho, estão na mesma ordem de grandeza.

5 CONCLUSÕES E PERSPECTIVAS.

Segue as conclusões e perspectivas para trabalho futuro

5.1 CONCLUSÕES

As técnicas estatísticas convencionais podem indicar variação em relação à média, desvio padrão, moda e outras informações como previsão em série temporal, muito úteis ao controle de uma estação de tratamento de efluentes líquidos, possibilitando informações que ajudam na tomada de decisão e otimizam o uso das variáveis de controle como ferramenta de previsão, assim também como preenchimento de dados faltantes.

O desenvolvimento dos métodos de inteligência artificial, associados a métodos estatísticos de controle aumenta o poder de decisão e as possibilidades de controles de automação de uma estação de tratamento de efluentes líquidos. Com os métodos abordados verificou-se que é possível se ter quantitativamente à previsão futura, um passo à frente, do que ocorrerá em uma estação de tratamento de efluentes, funcionando de forma estável, dentro de padrões de operações normais. Um resultado real fora das previsões, pode apontar uma mudança no processo, possibilitando ações na operação da planta, antes que venha a causar problemas ambientais maiores.

Os métodos de previsão univariada, apresentam a vantagem de poder ser reproduzido dando o mesmo resultado, porém não conseguem associar uma variável de controle a outra, o que torna os métodos de inteligência artificial, mais indicado para simulação de dados e automação de uma planta, o método associando redes neurais a transformada wavelet apresentam resultados de uma previsão mais indicada para o método multivariado.

No caso em estudo, pode ser observado a previsão de valores numéricos da DQO de saída do Tanque de Metanização (TM), representando erros relativos na faixa entre 24,5% até 30%, em variável que apresenta variação de 57,7% de desvio padrão em relação a média e a classificação dos dados nos permite ter uma acurácia de mais de 80%, na previsão qualitativa desta variável, indicando se o valor esperado estará acima ou abaixo da média registrada.

Também foi possível, usando métodos de classificação, se ter uma análise qualitativa da previsão, neste estudo resultados maiores que 80% de acerto, usando as redes convolutivas, redes neurais *feedforward*, e o vetor suporte de máquinas.

O uso das técnicas dos componentes principais pode indicar o comportamento das variáveis desejadas, quanto ao ajustes de outras variáveis e simulações na planta. Esta metodologia em conjunto com a classificação e previsão univariada e multivariada, aplicados à operação de uma planta, seriam de grande ajuda à tomada de decisão, controle de automação e prevenção de resultados indesejados.

5.2 SUGESTÃO PARA TRABALHOS FUTUROS.

Segue algumas sugestões para trabalhos de pesquisas futuros:

-Uso das técnicas empregadas neste trabalho, associados ao controle biológico da estação de tratamento de efluentes líquidos, com uso de Inteligência artificial para reconhecimento de espécies de microrganismos e ações de controle da ETE.

-Uso das técnicas de controle empregadas neste trabalho, associados ao controle biológico de reatores contínuos de fermentação alcoólica.

REFERÊNCIAS

ALFARO, C.A.; AYDIN, B.; VALENCIA, C.E.; BULLITT, E.; LADHA, A. **Dimension reduction in principal component analysis for trees.** Computation Statistics & data analysis, Volume 74, June 2014, Pages 157-179, 2014.

BADO, CRISTIANE; PERCIO, JANAINA ELIZA; LINDINO, CLEBER ANTÔNIO; **A Demanda Quimica de Oxigênio: Questionamentos**; Laboratório de Estudos em Química Analítica Limpa – LEQAL Grupo Interdisciplinar de Pesquisa em Fotoquímica e Eletroquímica Ambiental. Universidade Estadual do Oeste do Paraná, *Revista Analytica*, Nº 62, Dezembro 2012/Janeiro 2013.

BEZERRA, EDUARDO. **Introdução à Aprendizagem Profunda**. Disponível em: <http://sbbd2016.fpc.ufba.br/sbbd2016/minicursos/minicurso3.pdf>. Acesso em: 20 jan. 2018.

BOGGESS, A.; NARCOWICH, F. J. **A first course in wavelets with fourier analysis**, 2ª ed. ISBN 978-0-470-43117-7, Estados Unidos da America, 2009

BOUSFIELD, PAULO; SILVA , JÉSSICA C. S. SILVA; CHAPUIS, LORIVAL. **Uso de uma rede neural artificial para previsão do volume de lodo gerado em estações de efluente têxtil: um estudo de caso.** XXXI Congresso da sociedade brasileira de computação, Rio Grande do Norte, NT. 2011

BRAGA, A. P. ; CARVALHO, A. P. L. F. ; LUDERMIR, T. B.; **Redes Neurais Artificiais: Teorias e aplicações**. 2ª Edição. Rio de Janeiro. LTC. 2007

BRANDÃO, JOANA F.M. COSTA. **Otimização da produção de biogás com recurso a redes neuronais artificiais**. Dissertação de mestrado. Instituto Superior de Engenharia do Porto (2014).

CAMPANA, ANA C. M. **Uma proposta de transformação de dados para análise de componentes principais.** Universidade Federal de Viçosa - UFV, Departamento de Economia Aplicada. *Rev. Bras. Biom*., São Paulo, v.28, n.2, p.103-115, 2010

CARVALHO F. O.; GARCIA, C. A. B.; GARCIA, H. L; COELHO, F. A.; WANDERLEY, H.S.; SANTOS, R.R.C.; CAVALCANTI, J.R.A. "**Sistemas Inteligentes (Redes Neurais) e Estatística Multivariada (PCA) aplicados a qualidade da água**" in XVIII Simpósio Brasileiro de Recursos Hídricos, Campo Grande – MS, 2009.

CARVALHO, F. O.; GARCIA, C. A. B.; GARCIA, H. L.; ALVES, J. P. H.; PIMENTEL, W. R. O.; SILVA, J. L.. **Modelagem do processo de eutrofização no reservatório Jacarecica II – Sergipe/Brasil utilizando redes neurais artificiais e estatística multivariada** In: Congresso Ibero Latino-Americano sobre Métodos Computacionais em Engenharia - CILAMCE 2007, Porto, 2007.

CHAPRA, STEVEM C. R. **Métodos numéricos aplicados com MATLAB para engenheiros e cientistas**, 3ª ed, p 336 – 379, Porto Alegre: AMGH editora LTDA, 2011.

CHO, K. H.; KIM, J.;KANG, S.; PARK, H.; KIM, S.; KIM, Y. M. **Achieving enhanced nitrification in communities of nitrifying bacteria in full-scale wastewater treatment plants via optimal temperature and pH. Separation and Purification Technology**, v. 132, p. 697–703, 2014

COSTA, F. B.; SILVA, K. M.; DANTAS, K. M. C.; SOUZA, B. A.; BRITO, N. S. D. **A wavelet-based algorithm for disturbances detection using oscillographic data**, International Conference on Power Systems Transients pp. Lyon, France. 2007

COSTA, F. B.; SOUZA, B. A.; BRITO, N. S. D. (2008). **A wavelet-based algorithm to analyze oscillographic data with single and multiple disturbances**, IEEE PES General Meeting pp. Pittsburgh USA. 2008

COSTA, F. B.; SOUZA, B. A.; BRITO, N. S. D. **Detecção de faltas evolutivas e múltiplos distúrbios em registros oscilográficos baseada na transformada wavelet discreta.***Sba Controle & Automação* [online], vol.21, n.2, pp.173-184. ISSN 0103-1759. 2010

DENG, NAiYANG; TIAN, YINGJIE; ZHANG, CHUNHUA. **Support Vector Machines Optimization Based Theory, Algorithms, and Extensions.** CRC Press. A CHAPMAN & HALL BOOK. 345 p .2013

DONOSO-BRAVO, A.; MAILIER, J.; MARTIN, C.; RODRÍGUEZ J, ;ACEVES-LARA C.A.; VANDE WOUWER A. **Model selection, identification and validation in anaerobic digestion**: A review. Water research, v. 45, p.5347-5364, 2011.

EHLERS, R.S. (2009) **Análise de Séries Temporais,** Disponível em: ¸<http://www.icmc.usp.br/ ehlers/stemp/stemp.pdf>. Acessado em 19/11/2017.

ENDERS, WALTER, **Applied econometric time séries**, Universidade do Alabama. – 4ª edition. 498 páginas. Capitulo 1. ISBN 978-1-118-80856-6. 2014.

ENSUS ENGENHARIA, **Análise de Vibração – Tipos de Sinais, Transformada de Fourier e PSD.** Disponível em: <http://ensus.com.br/analise-de-vibracao-tipos-de-sinais-transformada-de-fourier-e-psd>Acesso em: 19 dez. 2017

FAEDO, ANDREIA M., **Tecnologias convencionais e novas alternativas para o tratamento de efluentes domésticos,** TCC. Universidade do Sul de Santa Catarina, Florianopolis, 2010.

FAN, J.; JI, F.; XU, X;, WANG, Y.;YAN, D.; XU, X.;CHEN,Q.; XIONG, Q.; XIONG, j.; HE, Q. **Prediction of the effect of fine grit on the MLVSS/MLSS ratio of activated sludge**. *Bioresource Technology*, v. 190, p. 51–56, 2015.

FERNANDES, L.M F. **Modelação da digestão anaeróbia da ETAR da Guia com Redes Neuronais Artificiais**. Dissertação de mestrado. Instituto Superior Técnico-Lisboa (2014)

FERNANDEZ, MARFIZA N. **Preenchimento de falhas em séries temporais.** Dissertação de mestrado. Universidade Federal do Rio Grande do Norte, RN. 2007.

FERREIRA, FABIANA D.; CORAIOLA, MARCIO. **Eficiencia do lodo ativado em fluxo continuo para tratamento de esgoto,** Revista Acadêmica de Ciencias Agrarias e Ambientais, Curitiba, v 6, n 2, p 259-279, abr/jun 2008.

FRANCO, GLAURA C. **Modelagem ARIMA.** Depto. Estatística – UFMG. Belo Horizonte. 2011.

FREITAS, THAISA DE; OLIVEIRA, FERNANDO LUIZ CYRINO. **Modelos e aplicações para séries temporais hierárquicas: abordagens de reconciliação ótima e proporções de previsão**. Dissertação de Mestrado. Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, Rio de Janeiro 2016.

FREITAS, ANTÔNIO A. C. **Previsão de Séries Temporais via Seleção de Variáveis, Reconstrução Dinâmica, ARMA-GARCH e Redes Neurais Artificiais,** tese de doutorado, Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica, SP. 2007.

GHOSH, KOUSHIK; RAYCHAUDHURI, PROBHAS **An adaptive approach to filter a time séries datas.** ArXiv Astrophysics e-prints, jan 2007.

GONÇALVES, ANDRÈ R. **Máquina de Vetores Suporte.** Computer Science & Engineering User Home Pages Disponível em < https://www-users.cs.umn.edu/~agoncalv/arquivos/pdfs/svm.pdf> acessado em 22/02/2018.

GOPI, E.S. **Multi-Disciplinary Digital Signal Processing A Functional Approach Using Matlab,** Springer International Publishing . ISBN 978-3-319-57429-5. Suiça, 2018.

GURGEL, VALCIANO CAMILO **Aplicação de Técnicas Wavelets em análise de séries temporais para detecção de correlações**, TCC Universidade Federal Rural do Semiárido RN. 2013.

HAMED, M. M.; KHALAFALLAH,M. G.; & HASSANIEN, E. A.**Prediction of wastewater treatment plant performance using artificial neural networks**. Environmental Modelling and Software, 19, 919–928. 2004.

HANBAY, DAVUT; TURKOGLU, IBRAHIM; DEMIR, YAKUP. **Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks**; Expert systems with applications. An International journal, volume 34, numero2 , fevereiro de 2008.

HEIKKINEM, M.; HEIKNEM, T.; HILTUNEM, Y. **Modelling of Activated Sludge Treatment Process in a Pulp Mill Using Neural Networks.** The 6th International Conference on Computing, Communications and Control Technologies: CCCT, Orlando, Florida, USA, 2008.

HONGYU, K.**Comparação do GGE-biplot ponderado e AMMI-ponderado com outros modelos de interação genótipo × ambiente**. 2015. 155p. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2015.

HREIZ, R.; LATIFI, M. A.; ROCHE, N. **Optimal design and operation of activated sludge processes: State-of-the-art.** *Chemical Engineering Journal*, v. 281, p. 900–920, 2015.

HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. O. **Análise de Componentes Principais: resumo teórico, aplicação e interpretação**. E&S - Engineering and Science ISSN: 2358-5390DOI: 10.18607/ES20164053 Volume 1, Edição 5 -2016.

ICMC , **Redes neurais artificiais**, O Instituto de Ciências Matemáticas e de Computação (ICMC). Universidade de São Paulo. disponível em < http://conteudo.icmc.usp.br/pessoas/andre/research/neural/ > acesso em 12/01/2018.

KHASHEI, MEHDI; BIJARI, MEHDI. **A new class of hybrid models for time séries forecasting**. Journal Expert Systems with Applications, V39, p 4344-4357, 2012.

KINTO, EDUARDO A. **Otimização e análise das maquinas de vetores de suporte aplicadas à classificação de documentos.** Tese de doutorado. Universidade de São Paulo. São Paulo, 2011.

KOMOROWSKA-KAUFMAN, M.; MAJCHEREK, H.; KLACZYN, **E. Factors affecting the biological nitrogen removal from wastewater.** *Process Biochemistry*, v. 41, n. 3, p. 1015–1021, 2006.

LAUWERS,JOOST; APPELS LISE; THOMPSON IAN P; JAN DEGRÈVE; VAN IMPE, JAN F, RAF DEWIL .**Mathematical modelling of anaerobic digestion of biomass and waste: Power and limitations.** Progress in Energy and Combustion Science 39, p.383-402. 2013.

LEME, E. J. A. **Manual prático de tratamento de águas residuária**. 2ª edição. São Carlos: Ed.UFSCar, 599 p, SP. 2014..

LUCHESA, CLAUDIO J.;CHAVES NETO, ANSELMO. **Tamanho da Amostra em Pesquisas.** Disponível em: <http://www.unicuritiba.edu.br/images/calculo_do_tamanho_da_amostra_-_texto_final_para_impressapso1.pdf >. Acesso em: 20 fev. 2018.

LIANG, YIZENG; XU, QING-SONG; LI, HONG-DONG; CAO, DONG-SHENG, **Support Vector Machines and Their Application in Chemistry**and **Biotechnology,** CRC Press, Taylor & Francis Group, USA, pag 22-23, 2011.

LOPES, ALCIDES B.O. **Desenvolvimento de estratégias no monitoramento da cor do açúcar utilizando máquinas de aprendizado baseados em ferramentas de sistemas inteligentes e estatística.** Trabalho de conclusão de curso Enga. Química. Universidade Federal de Alagoas. Maceió, 2016.

LUDWING JR., OSWALDO; MONTGOMERY, EDUARD. **Redes Neurais, Fundamentos e Aplicações com Programas em C**. Editora Ciência Moderna. Rio de Janeiro. 2007.

MALLAT, S. **A Wavelet Tour of Signal Processing: The Sparse Way**. Third Edit ed.2009.

MARQUES, L. P. **Utilização de redes neurais artificiais e análise de componentesprincipais no monitoramento da qualidade da água**, Dissertação de Mestrado, Recife, PE, UFPE 2011.

MARTÍNEZ, F. C.; FRAIRE, T. E.; CHÁVEZ, J. R. B. **Prediction of wastewater organic matter with time séries.** International Journal of Engineering and Technical Research (IJETR) ISSN: 2321-0869 (O) 2454-4698 (P), Volume-7, Issue-2, February 2017

MISITI, M.; MISITI,Y.; OPPENHEIN,G.; POGGI,M. **Wavelet Toolbox User's Guide**: for use with matlab. The MathWorks, version 4.19, 2017b. Disponível em: <http://www.mathworks.com/access/helpdesk/help/pdf_doc/wavelet/wavelet_ug.pdf>. Acesso em: 19 dez. 2017.

MATHWORKS. **Wavelet Toolbox User's Guide**:. Disponível em: < https://www.mathworks.com/help/wavelet/gs/introduction-to-the-wavelet-families.html?searchHighlight=Daubechies&s_tid=doc_srchtitle>. Acesso em: 19 dez. 2017.

MORETTIN, P. A.; TOLOI, CLÉLIA, M. C. **Análise de Séries Temporais**. São Paulo. Edgard Blucher.415-434, 2006.

MULLER, T. G., NOYKOVA, N., GYLLENBERG, M., & TIMMER, J. (2002). **Parameter identification in dynamical models of anaerobic waste water treatment.** Mathematical Biosciences, 177–178, 147–160, 2002.

NETO, PEDRO N. S. **Modelagem e simulação de um sistema de tratamento de esgoto sanitário usando rede neural artificial**. Dissertação de mestrado. Universidade Estadual da Paraiba, Campina Grande. PB. 2015

OLIVEIRA-JUNIOR, G. **Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado.** Ciência da Computação, Universidade Federal de Pernambuco, [S.l.], 2010.

OLIVEIRA, MAURI A, **Aplicação de redes neurais artificiais na análise de séries temporais econômico-financeiras.** Tese de doutorado, Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. São Paulo, 2007.

OLIVEIRA, S. **Um estudo em séries temporais na análise da receita nominal de vendas de veículos e motos.** Conexão ci.: r. cient. UNIFOR-MG, Formiga, v. 7, n. 2, p. 64-73, jul./dez. 2012.

OPPENHEIM, A. V.; WIILSKY, A. S. **Sinais e Sistemas**. 2. ed. São Paulo: Pearson Prentice Hall, 2010.

PEIXOTO, LEONARDO DE AZEVEDO. **Redes Neurais Artificiais na Predição do Valor Genético.** UFV, Viçosa-MG, 2013.

PINTO, P. M. SEQUEIRA. **Utilização de redes neuronais artificiais como ferramenta de gestão de processos de digestão anaeróbia. Caso de estudo – ETAR do Seixal.** Dissertação de mestrado. Universidade Nova de Lisboa – Portugual -Lisboa(2014).

REBELO, L. D. T**. Avaliação automática do resultado estético do tratamento conservador do cancro da mama.** Dissertação de Mestrado — FEUP, Portugal. Porto, 2008.

RAJALAHTI,T.; KVALHEIM, O.M.; **Multivariate data analysis in pharmaceutics: a tutorialreview.** International Journal of pharmaceutics, v. 417, p. 280 – 290, 2011.

REAL STATISTICS . **Real Statistics Using Excel, time-séries-analysis.** Disponível em < real-statistics.com/time-séries-analysis/stochastic-processes/dickey-fuller-test/> acessado em 22/02/2018.

ROCHA, RAFAEL; MILIDI_U, RUY LUIZ. **Reconhecimento de Objetos por Redes Neurais Convolutivas**. Dissertação de Mestrado . Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro. Rio de Janeiro, 2015.

ROCHA, W.F.C.; BORGES, C.N., **Aplicação de MARS- "Multivariate Adaptive Regression Splines" em metrologia para seleção de variáveis na determinação de biodiesel**, Sociedade Brasileira de Química (SBC) disponível em/; http://sec.sbq.org.br/cdrom/35ra/resumos/T0899-1.pdf>, acessado em: 05/03/2017.

RUSTUM, R.; ADELOYE, A. J. (2006) **Features extraction from primary clarifier data using unsupervised neural networks (Kohonen Self Organising Map)**. In: Seventh International Conference on Hydroinformatics (HIC 2006), Nice, France. 2006.

SANTOS, ROZIANE S. **Homogeneidade e reconstrução de séries climatológicas para localidades no Estado de Minas Gerais.** tese de doutorado, Universidade Federal de Viçosa, Viçosa, MG, 2012.

SILVA, IVAN NUNES DA; SPATTI, DANILO HERNANE; FLAUZINO, ROGÉRIO ANDRADE. **Redes neurais artificiais: para engenharia e ciências aplicadas**. São Paulo: Artliber, 2010.

SILVA, MURILO; OLESKOVICZ , MÁRIO; COURY, DANIS V. **Uma nova ferramenta baseada na transformada wavelet para localização digital de faltas. Sba Controle & Automação** vol.16 no.3 Campinas July/Sept. 2005.

SINDHU. J. NAIR; HOTA H.S.; GHOSH P.K.; AGRAWAL M.L. **Artificial Neural Network Based Model for the Prediction of Effluent from Lab-Scale Upward Flow Hybrid Anaerobic Sludge Blanket (UHASB) Reactor.** International Journal of Scientific & Engineering Research, Volume 3, Issue 2, February-2012.

SIRIPONG, S.; RITTMANN, B. E. **Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants**. Water Research, v. 41, p. 1110–1120, 2007.

SOARES, A.S., **Predição por séries temporais econômicas por meios de redes neurais artificiais e transformada Wavelet: Combinando modelo técnico e fundamentalista.** Dissertação de mestrado. Escola de engenharia de São Carlos. Universidade de São Paulo. São Carlos, 2008.

SOUZA, A. D. de. **Estudo de perda de carga em escoamento multifásico utilizando técnicas de inteligência artificial com ênfase no escoamento de petróleo**. Dissertação de Mestrado. Universidade Federal de Sergipe/Programa de Pós-Graduação em Engenharia Química. São Cristovão.Sergipe. Brasil. 2011.

SOUZA, A. M.; BREITKREITZ, M.C.; FILGUEIRAS, P. R.; ROHWEDDER, J. R.; POPPI,R. J., **Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II**, Química Nova, v. 36, p. 1058., 2013.

SRIDEVI, K., SIVARAMAN, E. , MULLA, P., **Back propagation neural network modelling of biodegradation and fermentative biohydrogen production using di stillery wastewater ina hybrid upflow anaerobic sludge blanket reactor**,In BIORESOUR CE TECHNOLOGY; 165; 233-240; by Elsevier Science B.V., Amsterdam, 2014.

STOCK, J. H.; WATSON, M. W. **Econometria.** São Paulo: Addison Wesley, 2004.

SIRIPONG, S.; RITTMANN, B. E. **Diversity study of nitrifying bacteria in full-scale municipal wastewater treatment plants.** *Water Research*, v. 41, p. 1110–1120, 2007.

SPERLING, M. VON **Princípios do tratamento biológico de águas residuárias**. **volume 4**. ***Lodos Ativados***. 3a edição ed. Belo Horizonte, 2012. 428p

TEIXEIRA, JÚLIO C.; OLIVEIRA, GUILHERME S.; VIALI, AMANDA M,; MUNIZ, SAMUEL S. **Estudo do impacto das deficiências de saneamento básico sobre a saúde pública no Brasil no período de 2001 a 2009.** Engenharia Sanitaria Ambiental | v.19 n.1 | | 87-96. jan/mar 2014.

TEIXEIRA JÚNIOR, L. A.; SOUZA, R. M.; MENEZES, L. M.; et al., **Artificial Neural Network and Wavelet decomposition in the Forecast of Global Horizontal Solar Radiation.** Sobrapo, v.35, n. 1, p. 1–16, 2015.

TÜMER, ABDULLAH ERDAL; EDBALI, SERPIL, **Prediction of Wastewater Treatment Plant Performance Using Multilinear Regression and Artificial Neural Networks**, Innovations in Intelligent SysTems and Applications (INISTA), International Symposium , **ISBN:** 978-1-4673-7751-5. Madrid. Espanha 2015.

TURNER, E.; EDEBALI A. **-Prediction of wastewater treatment plant performance using multilinear regression and artificial neural networks**. Innovations in Intelligent SysTems and Applications (INISTA), International Symposium. Madrid, Espanha. 2015.

VAPNIK, V.N. **Statistical Learning Theory.** John Wiley & Sons, Inc., 1998.

VIDA ALVANI; RAMIN NABIZADEH; MOHAMMAD ANSARIZADEH; AMIR HOSSEIN MAHVI E HASAN RAHMANI. **Prevendo a eficiência de remoção de TOC em filtro aerado biológico híbrido usando rede neural artificial,** Dessalinização e Tratamento de Água, 57:43, 20283-20291, DOI: 10.1080 / 19443994.2015 .1112839. (2016).

VILLALOBOS, CRISTIAN M. **Deep learning, estudo orientado**. Universidade Catolica do Rio de Janeiro, 2016

WEI, XIUPENG ; KUSIAK, ANDREW; SADAT, HOSSEINI RAHIL **Prediction of Influent Flow Rate: Data-Mining Approach** – Journal of energy engineeirng © ASCE, p 118 - 123 . Junho, 2013.

ZAN, GUSTAVO L.; SELLITO, M.A. **Técnicas de previsão de demanda: um estudo de caso triplo com dados de venda de materiais eletro-mecânicos.** GEPROS. Gestão da Produção, Operações e Sistemas – Ano 2, vol. 4, jul-set/07, p. 95-106. 2007.

APÊNDICE A - PROGRAMA QUE PREENCHE OS DADOS FALTANTES E NORMALIZA OS DADOS USANDO Z-SCORE

==========

Programa que preenche os dados faltantes e unidimensionaliza os dados usando z-score

```
%limpa as variaveis da memoria e carrega os dados brutos
clc;clear;
[NN TT] = xlsread('metanizadororiginal.xls');
%%
%Preenche os dados brutos faltantes na série temporal por dados interpolados usando a media móvel 30 dias
for i=1:size(NN,2);
x =1:size(NN,1);
A(:,i)=NN(:,i) ;
%F(:,i) = fillmissing(A(:,i),'spline');
F(:,i) = fillmissing(A(:,i),'movmean',30);
end
%% Calcula os itens que são dependentes usando como entrada as interpolações
    for K=1:size(NN,1);
      F(K,3)=F(K,1).*F(K,2).*0.024;
      F(K,7)=F(K,6)./F(K,5);
      end

%% Gera os gráficos com os dados brutos mais as interpolações

for i=1:size(NN,2);
   figure(i);
x =1:size(NN,1);
plot(x,F(:,i),'r.-',x,A(:,i),'b.-');
legend('Dados interpolados','Dados originais');
title('Preenchimento de dados em falta');
ylabel(TT(1,i+1));
xlabel('Dias corridos');
end
%%
%gera uma variável(valormaior) com os valores maiores de cada série
valormedio= mean(F);
valordesvio=std(F);
%%
% unidimensionaliza os dados dividindo os valores pelo maior valor anotado
% em cada  variável e cria a variável dadosunid

dadosunid=(F-valormedio)./valordesvio;
%% Gera os gráficos com os dados unidimensionalizados

 for i=1:size(dadosunid,2);

   x =1:size(NN,1);
   figure(i+size(dadosunid,2));
   p=polyfit(x',dadosunid(:,i),1);
```

```
    ynew=polyval(p,x');
    p1= plot(x,dadosunid(:,i),'-b','LineWidth',0.05);
    hold on
     p2=  plot(x,0,':g.','LineWidth',0.05);
      p3= plot(x,ynew,'-m.','LineWidth',2);
     p4=  plot(x,1,':g.','LineWidth',0.05);
       plot(x,-1,':g.');
     p5=  plot(x,2,':y.','LineWidth',0.05);
       plot(x,-2,':y.');
     p6=  plot(x,3,':r.','LineWidth',0.05);
       plot(x,-3,':r.')
       hold off
    legend([p1 p3],'Dados','Tendência Linear')
      % legend([p1 p2 p3 p4 p5 p6],'Dados','média','tendência linear','+/- 1 desvio','+/- 2 desvios','+/- 3
desvios')
        title('Dados unidimensionalizados por Z-score');
        ylabel(['Desvio padrão unitário da ',TT(1,i+1)]);
        xlabel('Dias corridos');

end
%% acumula os dados para testes
dadosunidacum=cumsum(dadosunid);
dadosbrutoscomburaco=NN;
dadosbrutossemburaco=F;
cabecalhos=TT;
%%
clearvars -except dadosunidacum valormedio valordesvio cabecalhos dadosunid dadosbrutocomburaco
dadosbrutossemburaco
%% salva os dados principais para trabalho
save 'metanizador.mat'
%%
filename = 'dadosbrutossemburaco.xlsx';
xlswrite(filename,dadosbrutossemburaco);
```

APÊNDICE B - PROGRAMA QUE TRANSFORMA OS DADOS EM FIGURA E USA A REDE CONVOLUTIVAS.

==========

Programa que transforma os dados em Figura e usa a rede convolutiva.

```
clc;clear;
delete c:\Matlab\CNN\Teste\*.*
delete c:\Matlab\CNN\Sobe\*.*
delete c:\Matlab\CNN\Estabiliza\*.*
delete c:\Matlab\CNN\Desce\*.*
delete c:\Matlab\CNN\*.png
delete c:\Matlab\CNN\*.jpg
load('metanizador.mat')
%%
%em K o numero de dias anteriores envolvidos na Figura K=10 significa 10
%dias e 10 variaveis K=1 siginifica 1 dia e 10 variáveis
K=20;
%em V a posição da variavel na Matriz que queremos associar ou prever exemplo --> V=10 significa
DQO de
%saída
V=10;
%A matriz tem numeros negativos, e as Figuras pedem numeros positivos,  para resolver o problema
% somaremos o valor 10 a todos os numeros dos dados
%dadosunid=dadosunid+10
dadosunid=dadosunid';
%%divide os dados unitários retirando trinta dias para teste da rede
Xis=dadosunid(:,1:end-31);
Xteste=dadosunid(:,end-30:end-1);
Ypsolon=dadosunid(V,2:end-30);
Yteste=dadosunid(V,end-29:end);
%retira a média e o desvio padrão dos últimos 50 valores
MXis= 0;
desvioP=1;
regiaominimo=-0.8;
regiaomaximo=0,8;
% entre 0,7 e -0,7 Desvios ficará a faixa de estabilidade da planta
%classifica os dados em saídas com valores de 0 (abaixo de -0,7
%desvio)(diminue)
%0 +/- 0,7 o valor está entre a media e 0,8 desvio a mais ou a menos(estabiliza).
% 0,7 o valor se situa acima da média e mais um desvio (sobe)
%%
Xisconv=Xis;
%% para cada valor de Xis terei um valor convertido (Xisconv) em tres áreas indicando se o resultado
%futuro esperado é acima de um devio padrão da média, dentro de um desvio da médio
```

```
%para baixo ou para cima ou ainda abaixo um desvio padrão da média.

for Ji=1:size(Xis,2);
for i=1:size(Xis,1);
if Xis(i,Ji)>=0.7;
   Xisconv(i,Ji)=1.001;
elseif Xis(i,Ji)>=-0.7;
   Xisconv(i,Ji)=.498;
else
   Xisconv(i,Ji)=0.001;
end
end
end
%convertendo os dados de teste
for Ji=1:size(Xteste,2);
for i=1:size(Xteste,1);
if Xteste(i,Ji)>=0.7;
   Xtesteconv(i,Ji)=1.001;
elseif Xteste(i,Ji)>=-0.7;
   Xtesteconv(i,Ji)=.498;
else
   Xtesteconv(i,Ji)=0.001;
end
end
end
Ypsolonconv=[Xisconv(V,K+1:end) Xtesteconv(V,1)];
%%
for Ji=1:size(Yteste,2);
if Yteste(1,Ji)>=0.7;
   Ytesteconv(1,Ji)=1.01;
elseif Yteste(1,Ji)>=-0.7;
   Ytesteconv(1,Ji)=.498;
else
   Ytesteconv(1,Ji)=0.01;
end
end
%transforma os dados em Figuras (10 variáveis x K) numeradas .png
n=0;

   for i=1:size(Xisconv,2)-K+1;
  n=n+1;
   A=Xisconv(:,i:i+K-1);
    %imshow(A)
    B=Ypsolonconv(i);
    if B>=0.9
    imwrite((A),['C:\Matlab\CNN\Sobe\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    n=n+1;
    imwrite((A),['C:\Matlab\CNN\Sobe\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
     n=n+1;
    imwrite((A),['C:\Matlab\CNN\Sobe\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    elseif B>= 0.3
```

```
    imwrite((A),['C:\Matlab\CNN\Estabiliza\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    else
    imwrite((A),['C:\Matlab\CNN\Desce\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    n=n+1;
    imwrite((A),['C:\Matlab\CNN\Desce\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    n=n+1;
    imwrite((A),['C:\Matlab\CNN\Desce\',num2str(Ypsolonconv(i)),'-',num2str(n),'.png']);
    end
    %B = imread([num2str(n),'.png']);
   end
   %%
Diretorio = fullfile('c:\','matlab','CNN')
digitData                                                                                                    =
imageDatastore(Diretorio,'IncludeSubfolders',true,'FileExtensions','.png','LabelSource','foldernames')
%digitData = imageDatastore(Diretorio,'foldernames')
 %%
 % Ypsolonconv=Ypsolonconv'
%digitData.Labels=categorical(Ypsolonconv');

%%
T = countEachLabel(digitData)
%%
%O numero de arquivos de treinamento deve ser menor que o menor arquivos das %classes  definido na
variável T calculada acima, ou dará erro.
trainingNumFiles =380;
rng(1) % For reproducibility
[trainDigitData,testDigitData] = splitEachLabel(digitData,...
            trainingNumFiles,'randomize');


        % Define a arquitetura da rede neural convolucional.


layers = [imageInputLayer([10 K 1]);
        convolution2dLayer(1,10);
        reluLayer();
        maxPooling2dLayer(1,'Stride',1);
        fullyConnectedLayer(size(T,1));
        softmaxLayer();
        classificationLayer()];

    % Set the options to default settings for the stochastic gradient descent
% with momentum. Set the maximum number of epochs at 20, and start the
% training with an initial learning rate of 0.001.
options = trainingOptions('sgdm','MaxEpochs',350,...
   'InitialLearnRate',0.0001);

%%
% Train the network.
convnet = trainNetwork(trainDigitData,layers,options);
```

```
%%
% Em Yteste vai estar o valor de resposta da rede treinada que deve ser
% comparada com valores reais observados (TTest)
YTest = classify(convnet,testDigitData);
TTest = testDigitData.Labels;

%%
% Calculo da acuracia do treinamento (mede o numero de acertos das
% variáveis separadas no treinamento e testa a resposta com valores reais
% observados (em Ytest os valores de previsão e em TTest os valores reais)
accuracy = sum(YTest == TTest)/numel(TTest)

%%
%Transforma os dados dos futuros 30 dias dos testes em  Figuras para
%alimentar a rede neural.
%%
Xtesteconv=[Xisconv(:,end-K+1:end) Xtesteconv];
%%
 for i=1:size(Xtesteconv,2)-K;
  n=n+1;
   A=Xtesteconv(:,i:i+K-1);
    %imshow(A)
    B=Ytesteconv(i);
    imwrite((A),['C:\Matlab\CNN\Teste\',num2str(Ytesteconv(i)),'-',num2str(n),'.png']);
    end

   Diretorio = fullfile('c:\','matlab','CNN','Teste')
testeData                                                                    =
imageDatastore(Diretorio,'IncludeSubfolders',true,'FileExtensions','.png','LabelSource','foldernames');

YTeste30 = classify(convnet,testeData);
%%
Yteste=Yteste'
```

APÊNDICE C- PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO *FEEDFWARD* UNIVARIADO.

==========

Previsão um passo à frente com retreino usando *feedfward* univariado

```
%% Previsão um passo à frente com retreino usandofeedfward
clear;clc
load('univariado.mat')
%%
%tbl = dadosunitTabela;
%%
for I=30:-1:1
X=ffw1input(1:end-(I+1),:);
Y=alvo(1:end-(I+1),:);
Yteste=alvo(end-29:end,:);

Xteste=ffw1input(end-I,:);
X=X'
Y=Y'
net = feedforwardnet(3);
net = train(net,X,Y);
Xteste=Xteste'
YFit(31-I) = net(Xteste);
end

YFit=YFit'
table(Yteste,YFit,'VariableNames',...
    {'REAL','PREVISTO'})

erro_abs=(mean(abs(YFit-Yteste)./Yteste))*100;
raizerroquadraticomedio=(mean((YFit-Yteste).^2))^.5;
figure
x=1:30;
plot(x,Yteste);
hold on
plot(x,YFit);
title({'Projeção um passo à frente, com retreino, usando Feedforward';' e um dia anterior na entrada (Teste1)'});
legend('Real','Previsto');
xlabel(['Método univariado NNCO=3, MAPE = ',num2str(erro_abs),'%, RMSE= ',num2str(raizerroquadraticomedio)]);
ylabel('DQO de saída do TM')
hold off;

%ylabel(['Desvio padrão da ',cab(1,I+1)
```

APÊNDICE D - MÉTODO UNIVARIADO USANDO WAVELET COM QUATRO FILTROS DE BAIXA E QUATRO FILTROS DE ALTA FREQUÊNCIA.

==========

Método univariado usando wavelet com quatro filtros de baixa e quatro filtros de alta frequência, no arquivo univariado. mat deve ter as variáveis da DQO de saída que ficará no Banco de dados com o nome "ffw1input" e outro arquivo chamado "alvo", neste caso teremos um banco com a variável DQO de saída um passo à frente.

```
clear; clc
load ('univariado.mat')
%%
Yteste=alvo(end-29:end,:);
entrada=ffw1input;
[Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('bior3.5');
%usando dois filtros
%[swain,swdin] = swt(entrada,2,Lo_D,Hi_D);
%[swaout,swdout] = swt(alvo,2,Lo_D,Hi_D);
% usando quatro filtros
[swain,swdin] = swt(entrada,4,Lo_D,Hi_D);
[swaout,swdout] = swt(alvo,4,Lo_D,Hi_D);
entrada=[swain; swdin];
alvo=[swaout;swdout];
clearvars -except entrada alvo Yteste Lo_D Hi_D Lo_R Hi_R
%%recon = iswt(swa,swd,Lo_R,Hi_R);

%% Previsão um passo à frente com retreino usando modelo Feedforward
%dados unidimensionalizados
%%
for I=30:-1:1
X=entrada(:,1:end-I);
Y=alvo(:,1:end-I);
%X=X'
%Y=Y'
Xteste=entrada(:,31-I)
net = feedforwardnet(17);
net = train(net,X,Y);
%Xteste=Xteste'
YFit(:,31-I) = net(Xteste);
end

%% Refaz a wavelet
swa=YFit(1:4,:)
swd=YFit(5:8,:)
YFit = iswt(swa,swd,Lo_R,Hi_R);
YFit=YFit'
table(Yteste,YFit,'VariableNames',...
    {'REAL','PREVISTO'})
```

```
erro_abs=(mean(abs(YFit-Yteste)./Yteste))*100;
raizerroquadraticomedio=(mean((YFit-Yteste).^2))^.5;
figure
x=1:30;
plot(x,Yteste);
hold on
plot(x,YFit);
title({'Projeção um passo à frente,';' com retreino, usando Wavelet e quatro filtros de alta e baixa';' e um dia anterior na entrada (Teste1)'});
legend('Real','Previsto');
xlabel(['Método univariado NNCO=9, MAPE = ',num2str(erro_abs),'%, RMSE= ',num2str(raizerroquadraticomedio)]);
ylabel('DQO de saída do TM')
hold off;

%ylabel(['Desvio padrão da ',cab(1,I+1)])
```

APÊNDICE E - PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO MODELO *WAVELET/FEEDFORWARD* DADOS NORMALIZADOS, MULTIVARIADO

==========

Previsão um passo à frente com retreino usando modelo wavelet/Feedforward
%dados unidimensionalizados, multivariado

```
clear; clc
Nfiltros= 4 ;%Numero de filtros de alta e de baixa
load ('metanizador.mat');
Ncomp= int16(size(dadosunid,1)/(2^Nfiltros))*(2^Nfiltros);%numero de dados compativeis na amostra para o numero de filtros
%%
alvo=dadosunid(end-(Ncomp-1):end,10);%DQO um dia a frente unidimensionalizada %sem buracos
Entrada=dadosunid(end-Ncomp:end-1,:);%todos os dados antes de um dia a frente sem buracos unidimensionalizadas.

 i=1;
    [Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('bior3.5');
    [swain,swdin] = swt(Entrada(:,i),Nfiltros,Lo_D,Hi_D);
    input=[swain; swdin];
for i=2:size(Entrada,2);
    [Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('bior3.5');
    [swain,swdin] = swt(Entrada(:,i),Nfiltros,Lo_D,Hi_D);
    input=[input; swain; swdin];
end
%%
Yteste=alvo(end-29:end,:);
[Lo_D,Hi_D,Lo_R,Hi_R] = wfilters('bior3.5');
[swaout,swdout] = swt(alvo,Nfiltros,Lo_D,Hi_D);
alvo=[swaout; swdout]
%%

%%recon = iswt(swa,swd,Lo_R,Hi_R);

%%
for I=30:-1:1
X=input(:,1:end-I);
Y=alvo(:,1:end-I);
%X=X'
%Y=Y'
Xteste=input(:,end-I)
net = feedforwardnet(2*size(input,1)+1);
net = train(net,X,Y);
%Xteste=Xteste'
YFit(:,31-I) = net(Xteste);
I
```

```
end

%% Refaz a wavelet
swa=YFit(1:Nfiltros,:)
swd=YFit(Nfiltros+1:2*Nfiltros,:)
YFit = iswt(swa,swd,Lo_R,Hi_R);

%% desfazendo a normalização
YFit=YFit*(valordesvio(10))+valormedio(10)
Yteste=Yteste*(valordesvio(10))+valormedio(10)
%%
YFit=YFit'
table(Yteste,YFit,'VariableNames',...
    {'REAL','PREVISTO'})
erro_abs=(mean(abs(YFit-Yteste)./Yteste))*100;
raizerroquadraticomedio=(mean((YFit-Yteste).^2))^.5;
figure
x=1:30;
plot(x,Yteste);
hold on
plot(x,YFit);
title({'Projeção um passo à frente,';' com retreino, usando Wavelet e quatro filtros de alta e baixa';' e um dia anterior na entrada (Teste1)'});
legend('Real','Previsto');
xlabel(['Método multivariado NNCO=',num2str(2*size(input,1)+1),', MAPE = ',num2str(erro_abs),'%, RMSE= ',num2str(raizerroquadraticomedio)]);
ylabel('DQO de saída do TM')
hold off;

%ylabel(['Desvio padrão da ',cab(1,I+1)])
```

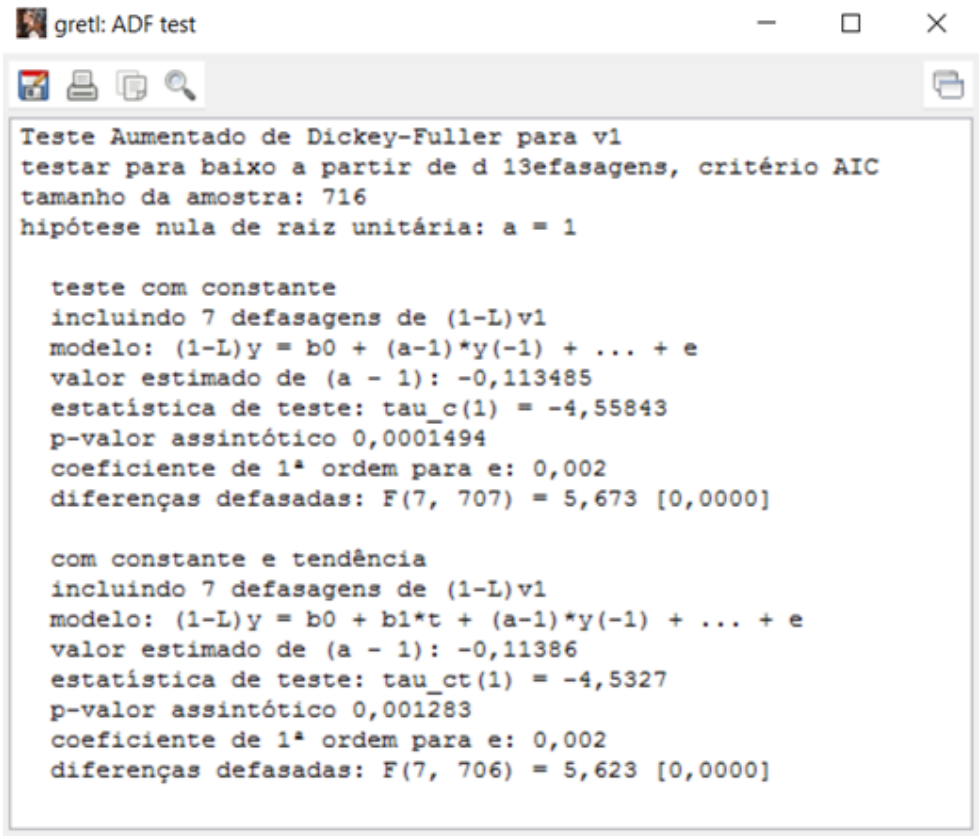
APÊNDICE F - PREVISÃO UM PASSO À FRENTE COM RETREINO USANDO MODELO SUPPORT VECTOR MACHINE REGRESSION, DADOS NORMALIZADOS.

==========

```
%% Previsão um passo à frente com retreino usando modelo Support Vector Machine Regression
%dados unidimensionalizados
clear;clc
%%
load('metanizador.mat')
%%
%tbl = dadosunitTabela;
%%
for I=30:-1:1
X=dadosunid(1:end-(I+1),10);
Xteste=dadosunid(end-I,10);
Y=dadosunid(2:end-I,10);
Yteste=dadosunid(end-29:end,10);
%%
%X=X'
%Xteste=Xteste'
%Y=Y';
%Yteste=Yteste';
N = size(X,1);
%%
% Partition the data into training and test sets. Hold out 10% of the data
% for testing.
rng(10); % For reproducibility
cvp = cvpartition(N,'Holdout',0.1);
idxTrn = training(cvp); % Training set indices
idxTest = test(cvp);    % Test set indices
%%
% Train a linear SVM regression model. Standardize the data.
Mdl = fitrsvm(X,Y,'Standardize',true);
%%
% |Mdl| is a |RegressionSVM| model.
%%
% Predict responses for the test set.
YFit(31-I) = predict(Mdl,Xteste);
end

%%
% Refaz os dados unidimensionalizados para valores reais de previsão e
% resposta certa
Yteste=Yteste*valordesvio(10)+valormedio(10)
YFit=YFit*valordesvio(10)+valormedio(10)
% Create a table containing the observed response values and the predicted
% response values side by side.
YFit=YFit'
table(Yteste,YFit,'VariableNames',...
    {'REAL','PREVISTO'})
```

```
erro_abs=(mean(abs(YFit-Yteste)./Yteste))*100;
raizerroquadraticomedio=(mean((YFit-Yteste).^2))^.5;
figure
x=1:30;
plot(x,Yteste);
hold on
plot(x,YFit);
title('Projeção um passo à frente, com retreino, usando SVM modo regressão');
legend('Real','Previsto');
xlabel(['Método multivariado MAPE = ',num2str(erro_abs),'%  RMSE= ',num2str(raizerroquadraticomedio)]);
ylabel('DQO de saída do TM')
hold off;

%ylabel(['Desvio padrão da ',cab(1,I+1)])
```

APÊNDICE G - TELA DO SOFTWARE GRETL COM TESTE DE ESTACIONARIEDADE DE DICKEY-FULLER