



# UFBA

UNIVERSIDADE FEDERAL DA BAHIA  
ESCOLA POLITÉCNICA  
PROGRAMA DE PÓS GRADUAÇÃO EM  
ENGENHARIA INDUSTRIAL - PEI

MESTRADO EM ENGENHARIA INDUSTRIAL

JÁCINA TÁBITA GURGEL MORAIS

"ANÁLISE DE COMPONENTES PRINCIPAIS INTEGRADA A  
REDES NEURAS ARTIFICIAIS PARA PREDIÇÃO  
DE MATÉRIA ORGÂNICA".



**SALVADOR**  
**2011**



ANÁLISE DE COMPONENTES PRINCIPAIS INTEGRADA A REDES  
NEURAS ARTIFICIAIS PARA PREDIÇÃO DE MATÉRIA ORGÂNICA

JÁCINA TÁBITA GURGEL MORAIS

Dissertação de Mestrado apresentada ao  
Programa de Pós-graduação em Engenharia  
Industrial, da Universidade Federal da Bahia,  
como parte dos requisitos necessários a  
obtenção do título de Mestre em Engenharia  
Industrial.

Orientadores: Prof<sup>a</sup>. Dra. Karla Patrícia S.O.R.  
Esquerre

Prof. Dr. Asher Kiperstok

Salvador  
Junho de 2011

---

M827 Morais, Jácina Tábita Gurgel

Análise de componentes principais integrada a redes neurais artificiais para predição de matéria orgânica / Jácina Tábita Gurgel Morais. – Salvador, 2011.

67 f. : il. color.

Orientador: Profa. Karla Patrícia Oliveira Rodriguez Esquerre  
Co-orientador: Prof. Asher Kiperstok

Dissertação (mestrado) – Universidade Federal da Bahia.  
Escola Politécnica, 2011.

1. Matéria orgânica. 2. Demanda bioquímica de oxigênio. 3. Redes neurais (computação). I. Esquerre-Oliveira, Karla Patrícia. II. Kiperstok, Asher. III. Título.

CDD: 660.63

---

**“ANÁLISE DE COMPONENTES PRINCIPAIS INTEGRADA A REDES NEURAIS  
ARTIFICIAIS PARA PREDIÇÃO DE MATÉRIA ORGÂNICA”.**

**JÁCINA TABITA GURGEL MORAIS**

Dissertação submetida ao corpo docente do programa de pós-graduação em Engenharia Industrial da Universidade Federal da Bahia como parte dos requisitos necessários para a obtenção do grau de mestre em Engenharia Industrial.

Examinada por:

KARLA PATRÍCIA SANTOS O. R. ESQUERRE Karla Patricia S-Oliveira R-Esquerre  
Doutorado em Engenharia Química pela Universidade Estadual de Campinas, Brasil  
(2003).

LUCIANO MATOS QUEIROZ Luciano Matos Queiroz  
Doutorado em Engenharia Civil pela Universidade de São Paulo (2009).

WAGNER ROBERTO DE OLIVEIRA PIMENTEL Wagner Roberto de Oliveira Pimentel  
Doutorado em Engenharia Química pela Universidade Estadual de Campinas, Brasil  
(2005).

Salvador, BA - BRASIL  
Julho/2011

"É impossível avaliar a força que possuímos sem medir o tamanho do obstáculo que podemos vencer, nem o valor de uma ação sem sabermos o sacrifício que ela comporta."(H. W. Beecher)

---

## Agradecimentos

---

A Deus pela realização de mais uma conquista na minha vida.

Aos meus pais João Batista e Nadja Maria, e minhas irmãs Jadna e Jacinara pelo amor incondicional, apoio e incentivo a minha formação profissional e pessoal.

Aos meus avós, em especial a minha avó paterna, Maria Silva (em memória) pela dedicação a minha educação e pelos momentos de carinho, amor e felicidade.

A Érico Britto pelo amor, compreensão e incentivo durante toda jornada dessa pesquisa, e a sua família pelo suporte e carinho durante minha estadia nesta cidade.

Ao exemplo de profissional, minha orientadora Karla Esquerre, pelas valiosas discussões, confiança, orientação, acessibilidade, incentivo e amizade. Agradeço ainda por me proporcionar realizar outras atividades que contribuíram para o meu desenvolvimento profissional e pessoal

Ao meu co-orientador Asher Kiperstok pelas críticas e sugestões.

Ao prof. Dr. Luciano Queiroz por contribuir com valiosas orientações e conhecimentos para elaboração deste trabalho.

Aos colegas do Programa de Engenharia Industrial, em especial, Aize Anne, Carlos Mendes, Fábio Menezes, Geiza Lima, Ingrid Barberino e Rosana Galvão, pelos conhecimentos compartilhados, momentos de alegria, convivência, apoio e incentivo. Agradeço também a Anderson Sisnando pelo suporte na modelagem e sua disponibilidade em tirar minhas dúvidas.

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo suporte financeiro a este trabalho. Agradeço ainda à Coordenadoria da Pós-graduação do PEI por todo apoio.

Enfim a todos aqueles que participaram desta realização direta ou indiretamente, na amizade, no auxílio e nas críticas, contribuindo assim para a conclusão dessa pesquisa.

Resumo da Dissertação apresentada ao PEI/UFBA como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

## ANÁLISE DE COMPONENTES PRINCIPAIS INTEGRADA A REDES NEURAIS ARTIFICIAIS PARA PREDIÇÃO DE MATÉRIA ORGÂNICA

Jácina Tábita Gurgel Morais

Junho/2011

Orientadores: Prof<sup>a</sup>. Dra. Karla Patricia S. Oliveira R. Esquerre  
Prof. Dr. Asher Kiperstok

Programa: Engenharia Industrial

Diversas técnicas avançadas vêm sendo adotada em plantas de tratamento de efluentes industriais com o propósito de melhorar o monitoramento e controle operacional da planta a fim de garantir a qualidade do efluente tratado antes de descartá-lo. Dentre estas técnicas, as redes neurais artificiais (RNAs) têm sido utilizadas com sucesso na modelagem destes sistemas. Entretanto, um passo importante e fundamental para um desempenho satisfatório das RNAs é o pré-processamento de dados. Esta tarefa requer uma análise cuidadosa dos dados a fim de obter-se um conjunto de variáveis e dados que melhor represente o sistema. O objetivo desta pesquisa consiste em construir modelos de predição da quantidade de matéria orgânica, medida por DQO, de uma lagoa aerada de uma empresa de produção de papel e celulose, e comparar o desempenho dos modelos construídos quando utilizada a técnica de análise de componentes principais (PCA) para pré-processamento dos dados. PCA foi utilizada nesta pesquisa para reduzir dimensionalmente o conjunto de dados através da seleção de componentes principais, descarte de variáveis originais e exclusão de possíveis *outliers*. Cinco conjuntos de dados foram formados para comparar o desempenho das RNAs com e sem aplicação de PCA, além do uso da DQO como unidade de concentração ( $\text{mg de DQO.L}^{-1}$ ) e como carga orgânica ( $\text{kg de DQO.dia}^{-1}$ ). Esse desempenho foi avaliado pelo erro quadrático médio (EQM), índice de correlação ( $R^2$ ), índice de correlação ajustado ( $R^2_{\text{ajustado}}$ ) e a complexidade da rede (Cn). A verificação da adequação do modelo é feita através da análise residual. A PCA foi capaz de facilitar o processo de aprendizagem da rede neural e reduzir os custos operacionais pelo descarte de variáveis originais. A DQO como carga orgânica também ajudou a melhorar o desempenho da rede PCA-RNA.

Abstract of Dissertation presented to PEI/UFBA as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

## PRINCIPAL COMPONENT ANALYSIS INTEGRATED ARTIFICIAL NEURAL NETWORK TO PREDCTION OF ORGANIC MATTER

Jácina Tábita Gurgel Morais

June/2011

Supervisors: Prof<sup>a</sup>. Dr. Karla Patricia S. Oliveira R. Esquerre

Prof. Dr. Asher Kiperstok

Department: Industrial Engineering Program

Advanced techniques have been developed at industrial wastewater treatment plants in order to improve their operational control and to ensure the quality of treated wastewater before discard it. Among this techniques, the neural network artificial (ANN) has been successfully applied to model these systems. However, an important and fundamental step for satisfactory performance of the ANNs is data preprocessing. This task requires a careful analysis of the data to define which variables and data best represent the system. The purpose this search is to build prediction models of the amount organic matter, measured by COD, of an aerated lagoon of a pulp and paper company, and to compare the performance of these models when applied the principal component analysis (PCA) technique for preprocessing the data. PCA was used is this research to reduce the dimension of the data set through selection of the principal components, discard of the original variables and exclusion of the possible outliers. Five data sets were used to evaluate the performance of the models with and without the application of PCA, in addition of the use of the COD as concentration ( $\text{mg of the COD.L}^{-1}$ ) and COD as organic load ( $\text{kg of the COD.day}^{-1}$ ). The performance of the each model was evaluated by computing the mean square error (MSE), the linear correlation index ( $R^2$ ) and adjusted linear correlation index (adjusted  $R^2$ ). Model adequacy was assessed by residual analysis. PCA was able to facilitate the learning process of the neural network and for reduce costs by variable discard. The use of COD as organic load also helped to improve the performance of the PCA-ANN model.

# Sumário

Lista de Figuras e Tabelas.....	x
Lista de abreviaturas.....	xi
Lista de símbolos.....	xii
Capítulo 1.....	1
INTRODUÇÃO.....	1
1.1. Motivação e Justificativa da escolha do tema.....	1
1.2. Objetivos do trabalho.....	3
1.2.1. Objetivo geral.....	3
1.2.2. Objetivos específicos.....	3
1.3. Organização do trabalho.....	3
Capítulo 2.....	5
DESCRIZAÇÃO DO PROCESSO.....	5
2.1. Sistema de tratamento.....	5
2.2. Estatística descritiva dos dados.....	5
2.3. Preparação dos dados.....	6
2.4. Considerações finais sobre o capítulo.....	7
Capítulo 3.....	8
FUNDAMENTAÇÃO TEÓRICA.....	8
3.1 Análise de componentes principais.....	8
3.1.1 Introdução.....	8
3.1.2 Desenvolvimento matemática.....	8
3.2 Redes neurais artificiais.....	11
3.2.1 Introdução.....	11
3.2.2 Conceitos básicos.....	11
3.2.3 Arquitetura das redes neurais.....	13
3.2.4 Treinamento.....	14
3.2.5 Treinamento com <i>backpropagation</i> .....	15
3.3 Análise dos modelos de RNAs.....	18
3.3.1 Parâmetros de avaliação do desempenho de RNAs.....	19
3.4 Considerações finais sobre o capítulo.....	20
Capítulo 4.....	21
MODELAGEM.....	21

4.1	Metodologia PCA.....	21
4.1.1	Seleção das componentes principais.....	21
4.1.2	Descarte de variáveis .....	22
4.1.3	Gráfico dos <i>scores</i> .....	23
4.2	Projeto de uma rede neural .....	24
4.2.1	Coleta e seleção dos dados.....	24
4.2.2	Configuração da rede .....	24
4.2.3	Treinamento .....	26
4.2.4	Teste.....	27
4.3	Avaliação da eficiência dos modelos MLP .....	27
4.4	Considerações finais sobre o capítulo .....	28
Capítulo 5.....		30
RESULTADOS E DISCUSSÕES.....		30
5.1	Pré-processamento dos dados .....	30
5.2	Modelos de predição.....	33
5.1	Considerações finais.....	38
Capítulo 6.....		40
CONSIDERAÇÕES FINAIS.....		40
6.1	Conclusões.....	40
6.	Sugestões para Trabalhos Futuros .....	41
REFERÊNCIAS .....		42
PUBLICAÇÕES DESENVOLVIDAS DURANTE A DISSERTAÇÃO.....		45
ANEXO 1: Gráficos de séries temporais .....		46
ANEXO 2: Teste de Anderson-Darling para verificação da normalidade dos parâmetros. ....		49
ANEXO 3: Gráficos com as <i>performances</i> dos modelos de predição para o conjunto de validação.....		51
ANEXO 4: Teste de Anderson-Darling para verificação da normalidade dos resíduos.....		53

## Lista de Figuras e Tabelas

Figura 2.1: Desenho esquemático do sistema de tratamento de efluentes .....	5
Figura 3.1: Fluxograma da obtenção das CPs. ....	10
Figura 3.2: Estrutura de um neurônio artificial.....	11
Figura 3.3: Estrutura da rede MLP com uma única camada intermediária.....	13
Figura 3.4: Representação gráfica para a função $E$ versus $w$ . ....	16
Figura 5.1: Gráfico de barras das CPs .....	31
Figura 5.2: Gráfico dos scores das duas primeiras CPs .....	32
Figura 5.3: Gráfico dos scores das duas primeiras componentes principais....	33
Figura 5.4: Relação entre a DQO de saída medida versus a predita para os modelos (a) M1, (b) M2, (c) M3, (d) M4 e (e) M5, considerando um intervalo de predição de 95% de confiança. ....	36
Figura 5.5: Gráfico de série temporal dos resíduos padronizados do modelo M4 – Linha superior e inferior indica um intervalo de confiança de 95%.....	37
Figure 5.6: Gráfico de série temporal da DQO de saída medida e predita.....	38
Tabela 2.1: Resultados da estatística básica dos parâmetros monitorados pela empresa .....	6
Tabela 2.2: Estrutura dos modelos.....	7
Tabela 5.1: Variâncias, variância explicada e acumulada das CPs .....	30
Tabela 5.2: Loadings das CPs selecionadas .....	31
Tabela 5.3: Resultados dos modelos de predição para o conjunto de teste ....	34
Tabela 5.4: Resultados dos pesos sinápticos dos modelos .....	35

## Lista de abreviaturas

Cor	Cor [mg/L]
CP	Componente principal
Dados 1	Conjunto de dados contendo 8 variáveis e 786 amostras com DQO (mg/L)
Dados 2	Conjunto de dados contendo 8 variáveis e 786 amostras com DQO(kg/dia)
Dados 3	Conjunto de dados transformados por PCA contendo 5 variáveis sintéticas e 786 amostras com DQO(kg/dia)
Dados 4	Conjunto de dados contendo 5 variáveis originais e 786 amostras com DQO(kg/dia)
Dados 5	Conjunto de dados contendo 8 variáveis originais e 719 amostras com DQO(kg/dia)
DQO	Demanda química de oxigênio [mg de O <sub>2</sub> /L]
DBO	Demanda bioquímica de oxigênio [mg de O <sub>2</sub> /L]
EC	Condutividade [ $\mu$ S/cm a 20°C]
EQM	Erro quadrático médio
MLP	Perceptron de múltiplas camadas
N.NH <sub>3</sub>	Nitrogênio amoniacal [mg/L]
N.NO <sub>3</sub> <sup>-</sup>	Nitrogênio nitrato [mg/L]
PCA	<i>Principal Component Analysis</i> ou Análise de componentes Principais
pH	Potencial hidrogênio iônico
Q	Vazão do afluente na lagoa aerada [ m <sup>3</sup> /dia]
Q <sub>papel</sub>	Produção de papel [ton/dia]
Q <sub>celulose</sub>	Produção de celulose [ton/dia]
R <sup>2</sup>	Coefficiente de correlação
R <sup>2</sup> <sub>ajustado</sub>	Coefficiente de correlação ajustado
RNA	Rede neural artificial
SQE	Soma quadrática dos resíduos
SQR	Soma quadrática da regressão
SQT	Soma quadrática total
SST	Sólidos em suspensão [mg/L]
T	Temperatura [°C]

## Lista de símbolos

$C(n)$	Complexidade dos modelos
$C_z$	Matriz de covariância
$D$	Matriz diagonal
$d_j(n)$	Resposta desejada do neurônio $j$
$e$	Resíduo
$E_{\text{médio}}$	Erro quadrado médio ou soma dos erros quadrados
$E(n)$	valor instantâneo da soma de erros quadrados
$E_{\text{médio}}$	Erro global médio da rede para todo o conjunto de treinamento
$E_{\text{total}}$	Soma total de erros quadrados
$L$	Camada intermediária ou oculta
$inputs$	Número de entradas da rede
$\mu$	Média amostral
$N$	Número de amostras
$n_o$	Número de neurônios de saída
$n_i$	Número de inputs
$n_h$	Número de neurônios na camada intermediária ou oculta $p$
$p$	Número de parâmetros ajustados pelo modelo
$v_j$	Função soma do neurônio $j$
$v_n$	Autovetores matriz diagonal $D$
$w_{ij}$	Peso sináptico da entrada $i$ pertencente ao neurônio $j$
$x_i$	Variáveis preditoras ou independentes
$y$	Variável predita ou dependente
$\bar{y}$	Valor médio da variável predita
$\hat{y}$	Valor estimado da variável predita
$y_j(n)$	Resposta calculada para o neurônio $j$ .
$\delta_j(n)$	Gradiente local de neurônio $j$ no tempo $n$
$Z$	Matriz de dados padronizados
$Z^T$	Matriz transposta de $Z$
$\Delta w$	pequena variação aplicada ao peso $w$
$\varphi$	Função de ativação do neurônio $j$
$\lambda$	Variâncias (autovalores)

# Capítulo 1

## INTRODUÇÃO

### 1.1. Motivação e Justificativa da escolha do tema

Um grande problema ambiental, bastante comum devido o aumento das atividades industriais, é a disposição e o tratamento de resíduos líquidos (efluentes). O tratamento de efluentes é uma atividade muito importante para que se possa garantir sua qualidade antes de ser descartado. Pois uma operação incorreta pode levar a sérios problemas de saúde pública e ambiental.

A lagoa aerada é um exemplo típico de unidade de tratamento de efluentes, seu principal objetivo é reduzir a quantidade de matéria orgânica dissolvida pelo uso de bactérias aeróbias. Essa quantidade de matéria orgânica presente é medida principalmente através dos respectivos parâmetros: demanda química de oxigênio (DQO) e demanda bioquímica de oxigênio (DBO). Estas são variáveis chaves do processo, indicadores da qualidade do efluente tratado.

Muitas técnicas avançadas vem sendo adotadas pelas indústrias para melhorar sua competitividade global, devido aos altos custos de energia e as normas ambientais, que são cada vez mais rígidas. O setor ambiental tem várias propostas em longo prazo de modelos de controle de predição com propósito de melhorar a eficiência energética, como também o desempenho das estações de tratamento de efluentes (ETEs) a fim de minimizar resíduos (HAMED, KHALAFALLAH e HASSANIEN, 2004; OLIVEIRA-ESQUERRE *et al.*, 2004; GRIEU, POLIT e COLPRIM, 2005; AL-ASHEH e ALFADALA, 2007; AKRATOS, PAPASPYROS E TSIHRINTZIS, 2008; MAY E SIVAKUMAR, 2009; BASANT *et al.*, 2010). Esses modelos de predição de parâmetros chaves, como concentração de DQO ou DBO, são construídos por redes neurais artificiais e utilizados para prever quali e quantitativamente variações do efluente, a fim de atingir as normas padrões de descarga.

As redes neurais artificiais (RNAs), diferentes de outras técnicas de modelagem tradicionais, tem a capacidade de representar relações lineares e não lineares, de um conjunto de dados, pelo processo de aprendizagem, mesmo quando pouco se sabe sobre o processo, desde que a estrutura de correlação do processo não mude; desta forma, a modelagem de sistemas ambientais, sem a prévia especificação das relações entre variáveis algébricas é possível. (SARASWATHI e SASEETHARAN, 2011). Um dos principais objetivos da RNA é de prever com precisão o comportamento do estado estacionário ou dinâmico, a fim de monitorar e melhorar o desempenho de um processo.

Há vários tipos de RNAs na literatura, a seleção de um modelo padrão neural apropriado à aplicação, isto é, qual configuração da rede neural a ser utilizada, ainda é feita de forma empírica, embora existam alguns métodos heurísticos que conduzam a opção mais acertada. Em geral o procedimento requer grande

experiência dos projetistas, o que pode ser obtido com variadas aplicações. As redes perceptrons de múltiplas camadas tem sido utilizadas com sucesso na modelagem de sistema de tratamento biológicos (GRIEU et al., 2005; AKRATOS, PAPASPYROS e TSIHRINTZIS, 2008; MAY e SIVAKUMAR, 2009). Entretanto, as principais desvantagens de sua utilização é que elas requerem uma quantidade considerável de dados para a construção dos modelos e o seu mapeamento não-linear pode levar a mínimos locais (OLIVEIRA-ESQUERRE, 2003). Outra desvantagem é o fato delas, normalmente, serem uma "caixa preta" assim mencionada por Ludwig Jr. e Montgomery (2007), tal nomeação se deve à dificuldade de saber qual a relevância de um peso sináptico para um determinado resultado, ou qual o seu significado físico.

Então como as RNAs tipo MLP podem utilizar grande quantidade de dados, e na prática o número de dados de treinamento tem limites. É recomendável evitar vetores de entrada de grandes dimensões, para isso um tratamento dos dados é útil para avaliar quais os que apresentam relevância e não redundâncias e assim, assegurar a qualidade dos resultados do modelo. (LUDWIG JR. e MONTGOMERY, 2007).

Yu (2006) propõem um método de preparar dados para RNAs, classificando-o em três partes: pré-análise, pré-processamento e pós-análise. O objetivo é de integrar essa preparação de dados às RNAs. Os resultados são satisfatórios utilizando o método proposto. Alguns métodos estatísticos clássicos foram mencionados para preparação dos dados, dentre eles a análise de componentes principais (PCA), que atualmente, vem sendo empregada como uma importante técnica na análise de dados multivariados em estudos de tratamento de efluentes, sua principal aplicação é na redução dimensional de um conjunto de dados com grande número de variáveis correlacionadas como pode ser observado nos estudos de Akrotos, Papaspyros e Tsihrintzis (2008), Grieu et al. (2005) e Hamed, Khalafallah e Hassanien (2004). A PCA também pode ser empregada para descarte de variáveis originais e exclusão de possíveis *outliers*.

A redução dimensional do conjunto de dados é útil para diminuir o número de parâmetros a serem ajustados (pesos sinápticos e *bias*) pela RNA durante o aprendizado. A PCA trata no geral de uma decomposição ortogonal adequada para encontrar um espaço de menor dimensão que capture maior variabilidade dos dados, em outras palavras, permite que variáveis altamente correlacionadas sejam transformadas em novas variáveis ortogonais, as chamadas componentes principais (CPs), que são combinações lineares das variáveis originais. Por fim, para reduzir o número de entradas da RNA, selecionam-se as principais CPs a partir de métodos encontrados na literatura.

Stanimirovic (2008) e Cadima, Cerdeira e Minhoto (2004) mencionam em seu trabalho, um método chamada método B4 desenvolvido por Jolliffe (1972) que é útil na seleção das principais CPs e na identificação das variáveis originais que estão mais fortemente correlacionadas a elas.

Há muitas situações, onde os pesquisadores se deparam com um grande conjunto de variáveis independentes para eventual inclusão em uma análise estatística multivariada. Na maioria dos casos, a inclusão de todas as variáveis

da análise estatística é, na melhor das hipóteses, desnecessária e, pior, um sério obstáculo para a correta interpretação dos dados (SILVA, 2001; CUMMING e WOOFF, 2007; CADIMA, CERDEIRA e MINHOTO, 2008; FUEDA, IIZUKA e MORI, 2009; MATTEAU, ASSANI e MESFIOUI, 2009).

A identificação de *outliers* é uma aplicação importante na modelagem empírica que trata com conjunto de dados de alta dimensão. Através dos *outliers* é possível encontrar informações que podem levar a incidentes graves. Portanto, problemas com identificação de *outliers* tem levado uma atenção particular como é o caso de Oliveira-Esquerre et al., (2004b) e Kristianse et al., (2010). Em ambas as pesquisas, a PCA foi o método adotado para identificar possíveis *outliers*, a qual se mostrou bastante eficiente apresentando resultados satisfatórios.

Nesta dissertação pretende-se realizar uma análise comparativa do desempenho dos modelos construídos com e sem a técnica de PCA integrada à RNA, para avaliação da remoção da matéria orgânica, medida pela DQO, em um efluente industrial tratado em lagoa aerada.

## **1.2. Objetivos do trabalho**

### **1.2.1. Objetivo geral**

Construir modelos por redes neurais artificiais para predição de matéria orgânica de um efluente tratado em lagoa aerada de uma unidade de tratamento de efluentes de uma empresa de produção de papel e celulose, e comparar o desempenho dos mesmos quando utilizada a técnica de pré-processamento de dados multivariados PCA para a redução dimensional do conjunto de dados de entrada da MLP através da seleção das componentes principais, descarte de variáveis originais e exclusão de possíveis *outliers*.

### **1.2.2. Objetivos específicos**

- Avaliar o desempenho dos modelos utilizando como parâmetro de entrada a DQO como concentração ( $\text{mg de DQO.L}^{-1}$ ) e como carga orgânica ( $\text{kg de DQO. dia}^{-1}$ );
- Definir critérios para uso da técnica de PCA para redução do número de entradas das RNAs e identificação de possíveis *outliers*;
- Comparar o desempenho dos modelos utilizando os dados originais e os tratados por PCA;
- Avaliar critérios de comparação de desempenho dos modelos construídos.

## **1.3. Organização do trabalho**

O presente trabalho está estruturado em cinco capítulos.

Capítulo 1 (presente capítulo) aborda a motivação e justificativa para elaboração do trabalho por um estudo de revisão da literatura, junto a eles os objetivos da pesquisa.

Capítulo 2 apresenta uma descrição da estação de tratamento de efluentes cujos dados foram oriundos, a preparação dos dados e a estrutura dos modelos formados.

Capítulo 3 apresenta uma breve introdução às redes neurais e a análise de componentes principais, seu desenvolvimento matemático e sua terminologia.

Capítulo 4 descreve a metodologia empregada para as RNAs, PCA, a verificação da adequação do modelo,

Capítulo 5 apresenta os resultados, bem como uma breve discussão a cerca dos resultados provenientes da aplicação da PCA aos dados de entrada da RNA.

Capítulo 6 encontra-se as conclusões e contribuições do trabalho desenvolvido, além de sugestões para o desenvolvimento de trabalhos futuros e as publicações desenvolvidas durante a dissertação.

Por fim, encontram-se as referências bibliográficas e os anexos.

## Capítulo 2 DESCRIÇÃO DO PROCESSO

### 2.1. Sistema de tratamento

O sistema de tratamento de efluentes da empresa de produção de papel e celulose considerada nesta pesquisa consiste de duas etapas, um tratamento primário seguido de um tratamento biológico. O tratamento primário contempla dois tanques de decantação em paralelo providos com misturadores e câmaras de floculação. O tratamento biológico é realizado em uma lagoa aerada de igual a 169 000 m<sup>3</sup> e complementado em um conjunto de cinco lagoas de sedimentação cujo volume total gira em torno de 1 500 000 m<sup>3</sup>. Os sólidos removidos dos tanques de decantação seguem para um tanque de secagem onde são apropriadamente descartados. Um desenho esquemático desse processo é apresentado na Figura 2.1.

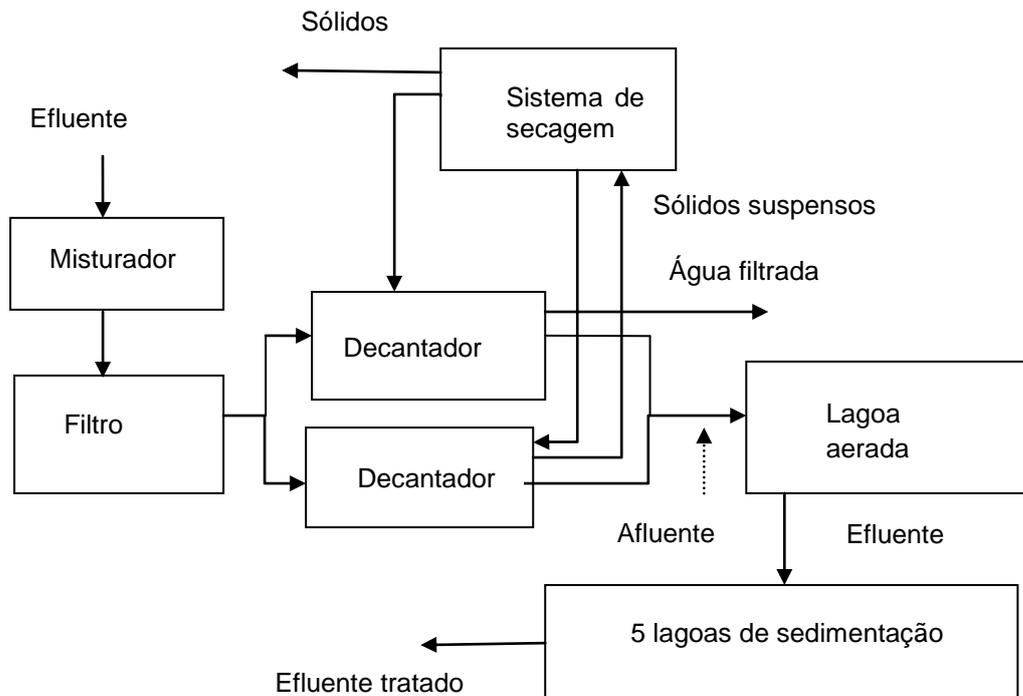


Figura 2.1: Desenho esquemático do sistema de tratamento de efluentes

### 2.2. Estatística descritiva dos dados

Os dados experimentais utilizados foram disponibilizados pela empresa, e referem-se a um período de 1427 dias consecutivos de medição, aproximadamente quatro anos amostrados. Oito parâmetros de monitoramento da lagoa aerada e dois referentes à produção da empresa foram escolhidos através do julgamento de profissionais da área baseados na importância dessas variáveis sobre a quantidade de matéria orgânica de saída da lagoa,

sendo eles: demanda química de oxigênio, demanda bioquímica de oxigênio, pH, sólidos em suspensão, cor, temperatura, condutividade, vazão do efluente, produção de papel e produção de celulose. As concentrações de nitrogênio (amoniacoal e nitrato), fósforo e sulfato eram medidas semanalmente.

Gráficos de séries temporais dos parâmetros são apresentados no ANEXO 1. A Tabela 2.1 apresenta os resultados da estatística básica dos parâmetros disponibilizados pela empresa.

**Tabela 2.1: Resultados da estatística básica dos parâmetros monitorados pela empresa**

Parâmetros	Média	Desvio padrão	Mínimo	Máximo	Skewness	Curtose	Lacunas (%)
<b>Q</b>	67.363,81	11.588,51	4.474	97.850	-1,53	4,90	0
<b>DBO<sub>in</sub></b>	245,05	46,33	41	449	0,08	1,54	6,24
<b>DQO<sub>in</sub></b>	561,49	104,19	136	925	-0,16	1,04	6,24
<b>SST</b>	149,20	85,74	12	591	1,59	4,13	60,4
<b>pH</b>	7,45	1,21	0,85	12,53	1,79	4,18	3,71
<b>NAM</b>	2,45	1,77	0	20	2,42	16,41	54,2
<b>NN</b>	1,43	0,88	0,03	7,385	2,42	11,05	80,5
<b>Cor</b>	464,36	123,59	41	1.317	0,51	3,39	3,57
<b>T</b>	45,46	3,05	28	50,50	-2,32	8,74	32,6
<b>EC</b>	1.530,96	378,05	379	5.810	2,66	17,74	3,92
<b>Q<sub>celulose</sub></b>	886,14	155,24	0	1.112,09	-3,31	14,16	7,85
<b>Q<sub>papel</sub></b>	1.042,72	94,16	382,40	1.304,80	-1,57	5,77	6,45
<b>DQO<sub>out</sub></b>	315,48	2,01	105	865	1,21	6,58	5,81

\*in equivale à entrada

\*out equivale à saída

Através do coeficiente de assimetria (*skewness*) e da curtose é possível analisar a proximidade de cada parâmetro em relação à gaussiana. Os valores de *skewness* estimados indicam que apenas a DBO<sub>in</sub> e a Cor apresentam distribuições aproximadamente simétricas. Os valores da curtose indicam que todos os parâmetros apresentam cauda mais pesada que a normal, com exceção da DBO<sub>in</sub> e DQO<sub>in</sub>, que apresentam achatamento mais próximo da distribuição normal. Nenhuma das variáveis passou no testes de normalidade de Anderson-Darling considerando um nível de confiança de 95%. Os resultados do teste de Anderson-Darling para verificação da normalidade dos parâmetros são mostrados no ANEXO 2.

### 2.3. Preparação dos dados

De forma a minimizar a perda de informação com a exclusão das amostras que contêm lacunas, os conjuntos de dados construídos para predição de DQO de saída da lagoa aerada contêm apenas as variáveis medidas que apresentam o percentual de lacunas inferior a 10 %, a saber: Q, DQO<sub>in</sub>, pH, Cor, T, EC, Q<sub>celulose</sub>, Q<sub>papel</sub>. A DBO não foi escolhida como variável de entrada devido o

tempo significativo de medição, aproximadamente cinco dias, o qual se torna impraticável para construção do modelo, e também ao grau de imprecisão de sua análise. Após exclusão dos dias que contem lacunas e prováveis equívocos de medição, o conjunto amostral foi reduzido de 1427 para 786 amostras.

Entretanto, cinco conjuntos de dados foram formados para construção dos modelos de predição. Dados 1 composto pelas 786 amostras, com a DQO como concentração ( $\text{mg de DQO.L}^{-1}$ ). Dados 2 composto pelas 786 amostras, mas com DQO como carga orgânica ( $\text{kg de DQO. dia}^{-1}$ ), calculada pelo produto da vazão por concentração de DQO. Esta variável é útil na representação da quantidade de matéria orgânica no efluente. Os conjuntos de dados 3 ao 5 são formados após a aplicação da técnica de pré-processamento de dados multivariados, PCA, sobre os Dados 2 em ordem para seleção das componentes principais (CPs), descarte de variáveis originais e exclusão de possíveis *outliers*. Na Tabela 2.2 é apresentada a estrutura dos modelos formados a partir dos conjuntos de dados especificados.

**Tabela 2.2: Estrutura dos modelos**

	<b>M1</b>	<b>M2</b>	<b>M3</b>	<b>M4</b>	<b>M5</b>
<b>DQO</b>	mg de DQO.L <sup>-1</sup>	Carga orgânica (kg de DQO.dia <sup>-1</sup> )			
<b>PCA</b>	-	-	Seleção das CPs	Descarte de variáveis originais	Exclusão de outliers

#### **2.4. Considerações finais sobre o capítulo**

O principal objetivo deste capítulo é apresentar o estudo de caso que será utilizado nesse estudo para avaliação da técnica de pré-processamento de dados, PCA, a qual será aplicada ao modelo de predição de matéria orgânica construído por redes neurais artificiais. Cinco conjuntos de dados são formados para construção dos modelos de predição, os quais serão utilizados para comparar o desempenho da RNA com e sem a aplicação de PCA, além da avaliação da DQO como concentração ( $\text{mg de DQO.L}^{-1}$ ) e como carga orgânica ( $\text{kg de DQO.dia}^{-1}$ ).

Os conjuntos de dados 1 e 2 foram formados para avaliar o efeito da quantidade de matéria orgânica como concentração e carga, respectivamente. O conjunto de dados 3 ao 5 foram formados a fim de avaliar individualmente cada aplicação que a PCA disponibiliza na literatura, nestes a DQO manteve-se como carga orgânica, por ser considerada desta forma como mais representativa da quantidade de matéria orgânica dissolvida no efluente a ser tratado.

## Capítulo 3

# FUNDAMENTAÇÃO TEÓRICA

### 3.1 Análise de componentes principais

#### 3.1.1 Introdução

A Análise de componentes principais – ACP ou PCA (do inglês *Principal component analysis*) foi primeiro introduzido por Pearson em 1901 e desenvolvida de forma independentemente por Hotelling em 1933.

A PCA trata-se de uma técnica estatística capaz de representar dados multivariados em um subespaço de dimensão reduzido, de modo que a distância entre os pontos, que representam as amostras nesse subespaço, corresponda o tanto quanto possível às dissimilaridades entre os pontos no espaço de dimensão original. (SILVA et al., 2005).

A utilização dos componentes principais para definir um espaço de fatores que englobe os dados, não modifica os dados em si, mas simplesmente encontra um sistema de coordenadas mais conveniente, capaz de remover ruídos dos dados sem distorcê-los e de diminuir sua dimensionalidade sem comprometer seu conteúdo de informações. (NETO e MOITA, 1998).

#### 3.1.2 Desenvolvimento matemática

A PCA consiste essencialmente em uma transformação de coordenadas de um conjunto de variáveis originais correlacionadas em um conjunto de variáveis sintéticas descorrelacionadas, as chamadas componentes principais (CPs). Cada componente é uma combinação linear das variáveis originais. Todas as CPs são ortogonais entre si, portanto não há informações redundantes.

“PCA tenta, simultaneamente, encontrar no espaço dimensional transformado a direção ao longo do qual os pontos se encontrem espalhados com variabilidade máxima e preservar a informação dos dados originais”. (SILVA et al., 2005). De um ponto de vista prático, isto é feito através dos seguintes passos: obtenção de uma matriz que represente o conjunto de dados, centralização dos dados em torno da média; cálculo da matriz de covariância; cálculo dos autovalores e autovetores e diagonalização da matriz de covariância.

No primeiro passo, os dados são organizados em uma matriz  $X_{n \times m}$ , onde  $n$  representa o número de observações e  $m$  o número de variáveis independentes.

O segundo passo, a centralização dos dados em torno da média, tem como objetivo prevenir que os pontos mais distantes do centro dos dados tenham maior influência que os mais próximos, o que é apropriado para dados em

diferentes unidades, assim mencionado por Luz (2003), essa transformação é feita pela aplicação da seguinte equação:

$$z_i = \frac{(x_i - \mu)}{\sigma} \quad \text{Equação (3.1)}$$

onde  $z_i$  são os valores transformados,  $x_i$  são as amostras para  $i = 1, 2, \dots, n$ ,  $\mu$  a média amostral da variável aleatória  $x_i$ ,  $\sigma$  é o desvio padrão amostral. Desta forma, todas as variáveis aleatórias são distribuídas com média zero e desvio padrão unitário.

O terceiro passo consiste no cálculo da matriz de covariância  $C_z$ :

$$C_z = Z \cdot Z^T \quad \text{Equação (3.2)}$$

Onde,  $Z^T$  é a transposta de  $Z$ . Os elementos da diagonal principal desta matriz se referem às variâncias ( $\lambda$ ) das colunas (variáveis independentes). Já os elementos fora da diagonal principal representam a covariância entre as variáveis. Quando a covariância é nula significa que as variáveis aleatórias são descorrelacionadas.

O método PCA permite a eliminação da covariância entre as coordenadas de um vetor de variáveis aleatórias por meio de uma mudança de base. Neste caso, as variáveis aleatórias são as entrada da rede neural, assim, a covariância implica nas informações redundantes e desnecessárias destas. A base formada pelos autovetores da matriz de covariância permite a diagonalização da matriz de covariância, ou seja, a eliminação da covariância entre as coordenadas do vetor de entrada. (LUDWIG JR. e MONTGOMERY, 2007).

A matriz de covariância é real e simétrica, o que significa ser possível encontrar um conjunto de autovalores e correspondentes autovetores ortogonais.

O quarto passo consiste então na determinação dos autovalores  $\lambda$  e autovetores  $v_n$  correspondentes da matriz  $C_z$ . Os autovetores são arranjados de modo decrescentes de acordo com os valores dos autovalores.

Encontrados os autovetores  $v_n$ , estes formarão as colunas de uma matriz  $P$ :

$$P = \{v_1, v_2, \dots, v_n\} \quad \text{Equação (3.3)}$$

Em PCA, os autovalores da matriz de covariância correspondem à variância das variáveis transformadas. Assim, se um autovetor possui autovalor grande, significa que este fica em uma direção em que há uma grande variância dos padrões. A importância desse fato está na classificação, pois, em geral, é mais fácil distinguir padrões usando uma base em que seus vetores não são correlacionados e que apontam para a direção da maior variância dos dados.

Os elementos dos autovetores, chamados de "*loadings*" na terminologia PCA, representam os cossenos diretores, ou seja, a contribuição com que cada um dos eixos originais entra na composição dos novos eixos, as CPs. Os

autovalores, por sua vez, representam a quantidade de variância original descrita pelos respectivos autovetores, os *scores*. (SILVA et al., 2005).

“Os *scores* são as projeções das amostras na direção das componentes principais. Suas coordenadas são obtidas a partir do produto da matriz de dados pela matriz de autovetores”. (SILVA et al., 2005).

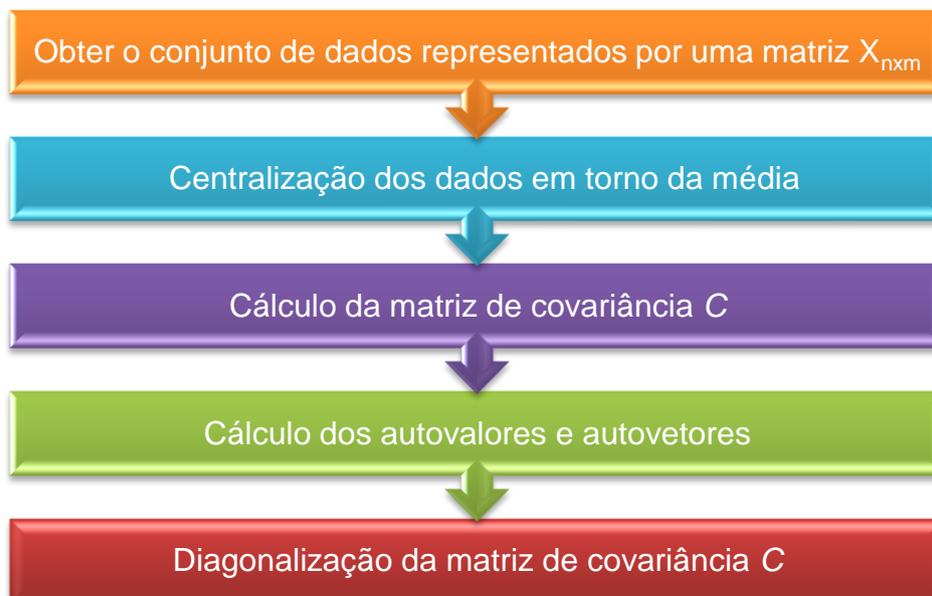
O último passo é a diagonalização. A matriz  $P$  é empregada para mudar a base de  $C_z$  obtendo uma matriz diagonal  $D$  de autovalores de  $C_z$ .

$$D = P^{-1}C_zP \quad \text{Equação (3.4)}$$

“A matriz  $D$  apresenta elementos iguais aos autovalores na diagonal principal, ou seja, não apresenta covariância, conseqüentemente não tem nenhuma informação redundante”. (LUDWIG JR. e MONTGOMERY, 2007).

Na transformada de Hotelling diferente do processo de diagonalização de matrizes, os autovetores são arranjados de modo decrescente de acordo com os valores dos  $n$  autovalores. A importância disso está na representação ordenada das componentes principais, onde a primeira componente ( $CP_1$ ) é representada no eixo de maior variância original dos dados, um eixo ao longo do qual os pontos se encontram espalhados de forma máxima. A segunda componente principal ( $CP_2$ ), ortogonal a primeira, representa o segundo eixo de maior variância, ou seja, o eixo com máxima quantidade de variância não explicada pelo primeiro autovetor, e assim sucessivamente, até a última componente principal. Resumindo, como cada CP captura o máximo de variância possível, a CP seguinte resta à variância residual, que se torna cada vez menor a cada CP sucessivo.

O fluxo de informação dos passos descritos para obtenção das CPs é ilustrado de forma resumida na Figura 3.1.



**Figura 3.1: Fluxograma da obtenção das CPs.**

## 3.2 Redes neurais artificiais

### 3.2.1 Introdução

“A rede neural artificial (RNA) é uma técnica de Inteligência Artificial (IA) especialmente projetada para processar informações e aquisição de conhecimento do cérebro humano através de simulação computacional”. (MELLO, 2004).

As RNAs são formadas por neurônios e conexões entre eles. O neurônio representa uma região onde informações são processadas e as conexões entre eles, denominadas pesos sinápticos, são responsáveis pelo armazenamento das informações. Além disso, elas definem o efeito que a saída de um neurônio exerce sobre a entrada do neurônio seguinte.

Para compreender a lógica de funcionamento das RNAs, alguns conceitos básicos referentes ao funcionamento do cérebro humano e seus componentes são de fundamental importância.

### 3.2.2 Conceitos básicos

O neurônio artificial, similarmente ao biológico, recebe um ou mais sinais de entrada e devolve um único sinal de saída, que pode ser distribuído como sinal de saída da rede ou como sinal de entrada para um ou vários outros neurônios da camada posterior. Seus três elementos básicos são: os pesos sinápticos, a função de soma, e a função de transferência também conhecida como função de ativação. (MELLO, 2004). A Figura 3.2 representa o modelo de um neurônio artificial.

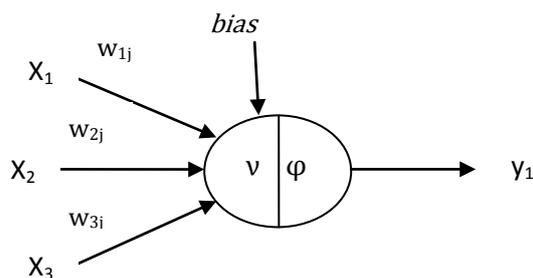


Figura 3.2: Estrutura de um neurônio artificial

Os pesos sinápticos, como descrito anteriormente, são as conexões entre os neurônios, responsáveis pelo armazenamento das informações.

A função de soma processa todos os estímulos ponderados pelos respectivos pesos, ou seja:

$$v_j = \sum_{i=1}^n w_{ij}x_i + b \quad \text{Equação (3.5)}$$

Onde,  $x_i$  corresponde às entradas,  $w_{ij}$  os pesos sinápticos e  $b$  o *bias*.

A função de transferência, ou função de ativação, tem a finalidade de evitar o acréscimo progressivo dos valores de saída ao longo das camadas da rede, visto que tais funções possuem valores máximos e mínimos contidos em intervalos determinados para cada tipo de função. (LUDWIG JR. e MONTGOMERY, 2007).

$$y_j = \varphi(x) \quad \text{Equação (3.6)}$$

Dentre as principais funções de transferência utilizadas, têm-se:

Função degrau	$\varphi(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x < 0 \end{cases}$	Equação (3.6a)
---------------	--	----------------

Função linear	$\varphi(x) = \begin{cases} 1, & x \geq \frac{1}{2} \\ x, & -\frac{1}{2} < x < \frac{1}{2} \\ 0, & x \leq -\frac{1}{2} \end{cases}$	Equação (3.6b)
---------------	---	----------------

Função sigmóide	$\varphi(x) = \frac{1}{1+e^{-\alpha x}}$	Equação (3.6c)
-----------------	--	----------------

Função hiperbólica	$\varphi(x) = \frac{e^{\alpha x} - e^{-\alpha x}}{e^{\alpha x} + e^{-\alpha x}}$	Equação (3.6d)
--------------------	--	----------------

O modelo neuronal matemático também pode incluir uma polarização ou *bias* de entrada. Esta variável é incluída ao somatório da função de ativação, com o intuito de aumentar o grau de liberdade desta função, e, conseqüentemente, a capacidade de aproximação da rede. O valor do *bias* é ajustado da mesma forma que os pesos sinápticos. O *bias* possibilita que um neurônio apresente saída não nula ainda que todas as suas entradas sejam nulas. (LUDWIG JR. e MONTGOMERY, 2007).

O tipo de conexão, número de camadas de neurônios e o tipo de treinamento são os aspectos que diferem os tipos de redes neurais. Cada um é mais adequado para determinado tipo de tarefa. Os diferentes tipos de conexões entre os neurônios de uma rede determinam a topologia (ou arquitetura) desta rede.

### 3.2.3 Arquitetura das redes neurais

As conexões entre as camadas podem gerar  $n$  números de estruturas diferentes. A forma com que estas estruturas estão dispostas se relaciona intimamente com o algoritmo de aprendizagem utilizado para treinar a rede. Portanto, a arquitetura da rede neural é livre, podendo ser modificada de acordo com as necessidades do projetista. (LUDWIG JR. e MONTGOMERY, 2007).

A rede neural Perceptron de Múltiplas Camadas ou Multilayer Perceptron (MLP) é composta por uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída, onde as camadas intermediárias funcionam como extratoras de características, seus pesos são uma codificação das características apresentadas pelas variáveis de *input* e permitem que a rede crie sua própria representação.

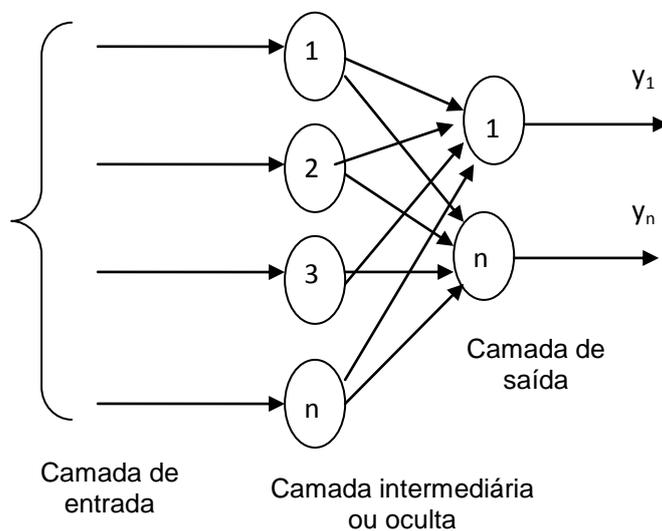


Figura 3.3: Estrutura da rede MLP com uma única camada intermediária.

O processo de aprendizagem da rede MLP é de forma supervisionada, ou seja, a rede é treinada com pares de conjuntos de entrada e de saída desejada. Quando é apresentada a rede um conjunto de entrada, esta retorna um conjunto de valores de saída, que é comparado ao conjunto de valores de saída desejada. Em seguida os pesos sinápticos e níveis de *bias* são ajustados de forma a minimizar a saída apresentada pela rede e a saída desejada. Este processo para todos os pares de entrada e saída que constituem o conjunto de treinamento da rede, até que a taxa de acerto seja considerada satisfatória.

Na MLP, o sinal de erro do neurônio de saída  $j$ , na interação  $n$ , é definido por:

$$e_j(n) = d_j(n) - y_j(n) \quad \text{Equação (3.7)}$$

Onde,  $d_j(n)$  é a resposta desejada do neurônio  $j$  e  $y_j(n)$  é a resposta calculada para o neurônio  $j$ .

O sinal do erro global e instantâneo da rede, para os  $J$  neurônios da camada de saída, na interação  $n$ , é definido por:

$$E(n) = \frac{1}{2} \sum_{j=1}^J e_j^2(n) \quad \text{Equação (3.8)}$$

De acordo com Ludwig Jr. e Montgomery, (2007) o erro de cada neurônio é elevado ao quadrado para evitar que um erro negativo oriundo de um neurônio compense um erro positivo de outro.

O erro global médio da rede para todo o conjunto de treinamento, ou seja, para um número  $N$  de interações necessárias para apresentar todo um ciclo de treinamento (uma época) em cada exemplo utilizado, é definido pela média aritmética dos erros globais instantâneos, conforme a Equação 3.9:

$$E_{\text{médio}} = \frac{1}{N} \sum_{n=1}^N E(n) \quad \text{Equação (3.9)}$$

“O erro global instantâneo médio, assim como, o erro global instantâneo são funções de todos os parâmetros livres (pesos sinápticos e *bias*)”. (LUDWIG JR. e MONTGOMERY, 2007).

### 3.2.4 Treinamento

Em síntese, o processo de treinamento da rede se constitui basicamente em um problema de otimização no ajuste de pesos das conexões e exige a apresentação da base de dados à rede por diversas vezes, até que o conjunto de pesos encontrados satisfaça a condição desejada para o erro apurado. Desta forma, as mesmas dificuldades associadas à otimização de processos também ocorrem na fase de aprendizado da rede, tais como problemas de convergência, existência de mínimos locais e tempo de computação indeterminado.

O algoritmo de aprendizagem é escolhido em função do tipo de rede utilizada e das características dos dados que serão ajustados. O mecanismo utilizado para aprendizagem no caso da MLP é o algoritmo de retropropagação do erro, conhecido como *backpropagation*.

De acordo com os parâmetros que eles atualizam, os algoritmos para treinamento desse tipo de rede podem ser classificados como: estáticos ou dinâmicos. (KOVÁCS, 2002).

Os algoritmos estáticos não alteram a estrutura da rede, variando apenas os valores de seus pesos, os algoritmos dinâmicos podem tanto reduzir quanto aumentar o tamanho da rede (número de camadas, número de nós nas camadas intermediárias e número de conexões).

Outro aspecto que precisa ser observado é a frequência de atualização dos pesos. A frequência de ajuste de pesos influencia o desempenho obtido durante o treinamento. “Duas abordagens diferentes têm sido utilizadas quanto à frequência (periodicidade) para o ajuste de pesos pelo algoritmo *backpropagation*: por padrão (*on line*) e por ciclo (*batch*)”. (KOVÁCS, 2002).

Na abordagem por padrão, os pesos são atualizados após a apresentação de cada padrão de treinamento. Esta abordagem é estável se a taxa de aprendizado for pequena (é aconselhável reduzir progressivamente esta taxa). Quando taxas elevadas são utilizadas, a rede geralmente se torna instável. A abordagem por padrão é geralmente mais rápida, sobretudo se o conjunto de treinamento for grande e redundante. Outra vantagem desta técnica é que ela requer menos memória.

Na abordagem por ciclo, por sua vez, os pesos são atualizados após um grupo de padrões terem sido apresentados. Esta técnica é geralmente mais estável, mas pode ser lenta se o conjunto de treinamento for grande e redundante. A escolha da abordagem a ser utilizada depende da aplicação e da distribuição estatística dos dados.

Um problema que pode ocorrer durante o treinamento de redes MLP é a ocorrência de *overffiting*. Este ocorre quando, após um determinado ciclo do treinamento, a rede, em vez de melhorar, começa a piorar a sua taxa de acertos para padrões diferentes daqueles utilizados para o ajuste dos pesos. Diz-se então que a rede memorizou os padrões de treinamento, gravando suas peculiaridades e ruídos. Isto tem como resultado uma piora na capacidade de generalização da rede. Assim como para os problemas de lentidão e mínimos locais, existem alternativas para lidar com o problema de *overffiting*: encerrar o treinamento da rede ou podar os pesos da rede é o que sugerem Demuth, Beale e Hagan (2010).

### **3.2.5 Treinamento com *backpropagation***

O treinamento do algoritmo *backpropagation* ocorre em duas fases, em que cada fase percorre a rede em um sentido. Estas duas fases são chamadas de fase *forward* e fase *backward*. A fase *forward* é utilizada para definir a saída da rede para um dado padrão de entrada. A fase *backward* utiliza a saída desejada e a saída fornecida pela rede para atualizar os pesos de suas conexões. (FARIA et al., 2008).

Na primeira fase, estímulos de entrada são apresentados à rede e as ativações fluem até chegarem à camada de saída, gerando um resultado. Já na segunda fase o resultado obtido pela rede é comparado com a saída desejada e o erro gerado é computado para as unidades de saída. Os pesos conectados às unidades de saída são então ajustados para reduzir este erro. Em seguida, o erro da camada de saída é utilizado para derivar estimativas de erro para as unidades da(s) camada(s) oculta(s), para que o erro seja então retropropagado até a conexão da camada de entrada. As equações deste processo são descritas pelas equações seguintes 3.10 a 3.20.

Mesmo conhecendo o erro global da rede, não é possível determinar os pesos exatos para poder corrigi-los. Entretanto, com base nesta informação, pode-se estabelecer a direção na qual os pesos devem ser ajustados para minimizar o erro quadrado total da saída da rede. Conhecida esta direção, é possível ajustar os pesos até que o menor erro global seja atingido.

Assim, vê-se que a formulação matemática do algoritmo de *backpropagation* consiste num problema de atualização dos pesos em função do erro. Para a formulação matemática deste problema é considerada uma rede contendo apenas uma camada intermediária.

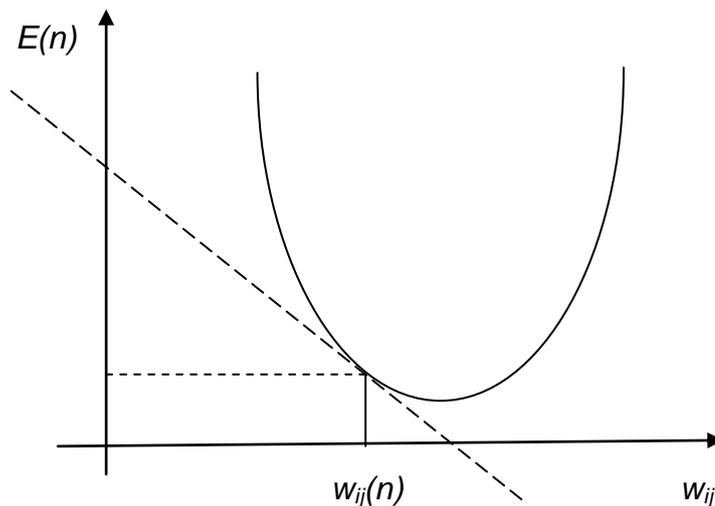
O ajuste de um peso  $w_{ij}$  que define seu valor para a próxima iteração é definido por:

$$w_{ij}(n + 1) = w_{ij}(n) + \Delta w_{ij}(n) \quad \text{Equação (3.10)}$$

A regra delta sugere que a variação dos pesos seja definida de acordo com o gradiente descendente do erro com relação ao peso, ou seja, que a variação do peso para um dado padrão seja definida pela Equação 3.11.

$$\Delta w_{ij}(n) \propto - \frac{\partial E(n)}{\partial w_{ij}(n)} \quad \text{Equação (3.11)}$$

Uma relação  $\partial E(n)/\partial w_{ij}(n)$  positiva significa que, se for aplicada uma correção positiva ao peso sináptico  $w_{ij}(n)$ , ter-se-ia um acréscimo no erro global  $E(n)$ , devendo, neste caso, ser aplicada uma correção negativa, e vice-versa. Caso o valor da relação  $\partial E(n)/\partial w_{ij}(n)$  seja pequeno, estaria próximo a um  $E(n)$  mínimo. Sendo assim deve ser aplicada uma correção a  $w_{ij}(n)$  pequena. Já se o valor desta derivada fosse grande, provavelmente, o valor de  $w_{ij}(n)$  estaria distante do valor que resultaria em  $E(n)$  mínimo, portanto seria necessário aplicar uma correção maior para  $w_{ij}(n)$ . (LUDWIG JR. e MONTGOMERY, 2007). A representação gráfica de uma forma similar a esse processo é apresentada na Figura 3.4.



**Figura 3.4: Representação gráfica para a função  $E$  versus  $w$ .**

Para determinar essa derivada parcial, é considerado um único neurônio da camada de saída, indexado pela letra  $j$ , que é alimentado por um conjunto de sinais funcionais  $y_i$  produzidos pela camada de neurônios anterior.

Aplicando-se a regra da cadeia na derivada parcial da Equação (3.11) tem-se:

$$\frac{\partial E(n)}{\partial w_{ij}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \cdot \frac{\partial e_j(n)}{\partial y_j(n)} \cdot \frac{\partial y_j(n)}{\partial v_j(n)} \cdot \frac{\partial v_j(n)}{\partial w_{ij}(n)} \quad \text{Equação (3.12)}$$

Desenvolvendo os fatores  $\partial E_j(n)/\partial e_j(n)$ ;  $\partial e_j(n)/\partial y_j(n)$ ;  $\partial y_j(n)/\partial v_j(n)$ , obtêm-se:

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j \quad \text{Equação (3.13)}$$

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad \text{Equação (3.14)}$$

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'_j(v_j(n)) \quad \text{Equação (3.15)}$$

A partir desse ponto já se pode escrever a equação do *gradiente local* do neurônio  $j$ ,  $\delta_j(n)$ , substituindo os fatores desenvolvidos:

$$\delta_j(n) = \frac{\partial E(n)}{\partial v_j(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} = -e_j(n) \varphi'_j(v_j(n)) \quad \text{Equação (3.16)}$$

Desenvolvendo o fator  $\partial v_j(n)/\partial w_{ij}(n)$ , em que:

$$v_j(n) = \sum w_{ij}(n) y_i(n) \quad \text{Equação (3.17)}$$

Tem-se

$$\frac{\partial v_j(n)}{\partial w_{ij}(n)} = y_i(n) \quad \text{Equação (3.18)}$$

Substituindo os fatores desenvolvidos na Equação (3.12) obtêm-se:

$$\frac{\partial E(n)}{\partial w_{ij}(n)} = -e_j(n) \varphi'_j(v_j(n)) y_i(n) \quad \text{Equação (3.19)}$$

Ou ainda:

$$\frac{\partial E(n)}{\partial w_{ij}(n)} = \delta_j(n) y_i(n) \quad \text{Equação (3.20)}$$

Com isto, já se pode aplicar correção a todos os pesos sinápticos dos neurônios da última camada, bem como dos *bias*, os quais devem ser somados aos valores atuais. A correção  $\Delta w_{ij}$  aplicada a  $\partial w_{ij}(n)$  é então definida por:

$$\Delta w_{ij} = -\eta \frac{\partial E(n)}{\partial w_{ij}(n)} = \eta e_j(n) \varphi'_j(v_j(n)) y_i(n) \quad \text{Equação (3.21)}$$

Ou,

$$\Delta w_{ij} = -\eta \delta_j(n) y_i(n) \quad \text{Equação (3.22)}$$

$$\Delta b_j = \eta e_j(n) \varphi'_j(v_j(n)) y_i(n) = -\eta \delta_j(n) y_i(n) \quad \text{Equação (3.23)}$$

Onde  $\eta$  a taxa de aprendizado.

A taxa de aprendizagem  $\eta$  é responsável pela velocidade com que se dá a busca no espaço de pesos, em direção aos valores que resultam em um erro global mínimo. Quanto menor for a taxa de aprendizagem, mais suave e precisa será a trajetória através do espaço de pesos, entretanto o aprendizado será lento. Em contraposição se for adotado um parâmetro muito grande para  $\eta$ , encontram-se modificações muito intensas nos pesos sinápticos e, conseqüentemente, uma busca oscilatória, ou seja, os valores dos pesos “passariam” do ponto ótimo e seriam remetidos de volta repetidamente, o que resultaria em uma rede instável. (LUDWIG JR. e MONTGOMERY, 2007).

Há diferença entre o ajuste de peso para os neurônios da camada de saída e da camada intermediária. Pois o cálculo do gradiente local para os neurônios da camada intermediária (ou oculta) é:

$$\delta_i(n) = \varphi'_i(v_i(n)) \sum_{j=1}^J (\delta_j(n) w_{ij}(n)) \quad \text{Equação (3.24)}$$

Então o cálculo do ajuste para todos os pesos desta camada, bem como dos *bias*, os quais devem ser somados aos valores atuais é:

$$\Delta w_{ki} = -\eta \varphi'_i(v_i(n)) y_k(n) \sum_{j=1}^J (\delta_j(n) w_{ij}(n)) = -\eta \delta_j(n) y_k(n) \quad \text{Equação (3.25)}$$

$$\Delta b_i = -\eta \varphi'_i(v_i(n)) y_k(n) \sum_{j=1}^J (\delta_j(n) w_{ij}(n)) = -\eta \delta_j(n) y_k(n) \quad \text{Equação (3.26)}$$

Várias alterações do algoritmo *backpropagation* têm sido propostas visando tanto a acelerar seu tempo de treinamento como a melhorar seu desempenho na classificação de padrões. Demuth, Beale e Hagan (2010) citam os algoritmos *backpropagation* com momentum, Levenberg-Marquardt e método de Newton como os mais utilizados.

### 3.3 Análise dos modelos de RNAs

Normalmente, a complexidade do modelo de redes neurais são referidas ao número de parâmetros, ou seja, o número de pesos e o número de *bias* a serem ajustados. Yu, Wang e Lai (2006) calculam a complexidade dos modelos de RNAs como:

$$C(n) = n_h(n_i + 1) + n_o(n_h + 1) \quad \text{Equação (3.27)}$$

$C(n)$  é a complexidade do modelo,  $n_i$  é o número de inputs,  $n_h$  é o número de neurônios na camada intermediária ou oculta, e  $n_o$  é o número de neurônios na camada de saída. Através da Equação (3.27), constata-se que a complexidade do modelo pode ser reduzido por um tratamento adequado de preparação de dados, como, por exemplo, redução de *inputs*.

Análises devem ser conduzidas para examinar a eficiência e adequação do modelo construído. Os parâmetros de avaliação abordados neste trabalho são: erro quadrático médio (EQM), o coeficiente de correlação múltipla ( $R^2$ ) e o coeficiente ajustado de determinação múltipla ( $R^2_{\text{ajustado}}$ ).

### 3.3.1 Parâmetros de avaliação do desempenho de RNAs

#### *Erro Quadrático Médio*

O Erro Quadrático Médio é definido por:

$$EQM = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad \text{Equação (3.28)}$$

onde,  $y_i$  é a saída desejada, e  $\hat{y}_i$  é a saída calculada pela rede.

Por se tratar de uma rede de aprendizagem supervisionada, cujo objetivo do processo de aprendizagem é ajustar os parâmetros livres (*pesos sinápticos e bias*) da rede para minimizar a energia do erro quadrado, o EQM é um dos principais parâmetros de avaliação utilizado.

#### *Coefficiente de Determinação Múltipla*

O coeficiente de determinação múltipla é definido como:

$$R^2 = \frac{SQR}{SQT} = 1 - \frac{SQE}{SQT} \quad \text{Equação (3.29)}$$

Onde,  $SQR$  é a soma quadrática da regressão, dada por  $\sum_{i=1}^n (\hat{y} - \bar{y})^2$ ,  $SQT$  a soma quadrática total corrigida de  $y$ , dada por  $\sum_{i=1}^n (y_i - \bar{y})^2$ ,  $SQE$  a soma quadrática do erro ou residual, dada por  $\sum_{i=1}^n (y_i - \hat{y})^2$  e  $n$  é o número de amostras.

A raiz quadrada positiva de  $R^2$  é chamada de coeficiente de correlação múltipla. Este coeficiente mede o grau de correlação e a direção dessa correlação entre duas variáveis, baseado na qualidade de um ajuste linear dos dados. A idéia é avaliar a relevância de um dado de entrada de uma RNA, em relação à saída desejada, por meio da verificação do coeficiente de correlação entre as ocorrências desta entrada e as respectivas saídas desejadas (LUDWIG JR. e MONTGOMERY, 2007).

Este coeficiente assume apenas valores entre 0 e 1, em que altos valores de  $R^2$  significa boa correlação entre as variáveis de entrada com a de saída, e  $R^2$  próximo de zero significa que as duas variáveis não dependem linearmente uma da outra. Entretanto, um valor alto de  $R^2$  não implica necessariamente que o modelo de regressão é bom. Adicionando uma variável ao modelo o valor de  $R^2$  sempre aumentará, indiferentemente se a variável adicionada é significativa ou não. Então, modelos que têm altos valores de  $R^2$  podem produzir pobres previsões de novas observações ou estimativas da resposta média (MONTGOMERY, RUNGER e HUBELE, 2004).

#### *Coefficiente ajustado de Determinação Múltipla*

O Coeficiente ajustado de Determinação Múltipla pode ser definido por:

$$R_{ajustado}^2 = 1 - \frac{SQE/(n-p)}{SQT/(n-1)} \quad \text{Equação (3.30)}$$

Onde,  $p$  o número de parâmetros a serem ajustados, no caso de RNA os pesos *sinápticos* e *bias*.

Diferente do  $R^2$  o  $R^2_{ajustado}$  nem sempre aumenta quando uma variável é adicionada ao modelo, o  $R^2_{ajustado}$  só aumenta se a adição da variável produzir uma redução na soma quadrática residual, que seja grande o suficiente para compensar a perda de um grau de liberdade no resíduo.

### 3.4 Considerações finais sobre o capítulo

O objetivo principal deste capítulo foi apresentar uma base teórica necessária para o entendimento das técnicas de análise de componentes principais e redes neurais artificiais.

A análise de componentes principais trata-se de uma técnica estatística de análise de dados multivariados. Uma das principais razões da sua aplicação é o fato dela poder representar dados multivariados em um sistema de menor dimensão. Neste capítulo foi apresentado seu desenvolvimento matemático seguido dos significados físicos de cada etapa para obtenção das CPs. Na literatura esta técnica é explorada com as seguintes aplicações: reduzir dimensionalmente um conjunto de dados; descartar variáveis e identificar possíveis *outliers*.

A rede neural artificial aplicada neste trabalho é do tipo MLP, que são redes que apresentam pelo menos uma camada intermediária. Inicialmente foram apresentadas as características básicas das redes MLP, o seu treinamento, o qual é de forma supervisionada, o principal algoritmo utilizado para treinar tais redes, o algoritmo *backpropagation*, outros algoritmos foram citados. Para formulação matemática do algoritmo *backpropagation* foi considerada uma rede contendo apenas uma camada intermediária e uma de saída.

Os modelos serão avaliados pelos respectivos parâmetros: EQM,  $R^2$ ,  $R^2_{ajustado}$ .

O software comercial Matlab foi utilizado para o desenvolvimento dos modelos MLP e também para obtenção das CPs. Os modelos MLP foram construídos baseados em um programa desenvolvido pelo mestre Anderson Sisnando, do Laboratório de Telecomunicações do departamento de Engenharia Elétrica da UFBA.

## Capítulo 4

### MODELAGEM

#### 4.1 Metodologia PCA

O pré-processamento de dados é uma tarefa muito importante para o futuro desempenho de uma RNA. Esta tarefa requer uma análise cuidadosa dos dados com a finalidade de identificar quais apresentam relevância e não redundâncias. (LUDWIG JR. e MONTGOMERY, 2007).

No pré-processamento, os dados são transformados para um novo sistema de coordenadas com eixos ortogonais. Em outras palavras, as n-variáveis originais geram, através de suas combinações lineares, n-CPs, cujas principais características são: ortogonalidade, onde cada CP pode ser analisada separadamente, e desta forma, ser usada para julgar as variáveis originais mais importantes em cada eixo; as CPs são obtidas em ordem decrescente de máxima variância, o que permite a redução da dimensionalidade dos pontos representativos das amostras; O gráfico da componente principal 1 versus a componente principal 2 fornece uma janela privilegiada (estatisticamente) para observação dos pontos no espaço n-dimensional. Enfim, a PCA é útil para reduzir a dimensão de um conjunto de dados através da seleção das CPs, descarte de variáveis originais e exclusão de possíveis *outliers*, os quais formam em ordem os conjuntos de Dados 3 ao 5.

##### 4.1.1 Seleção das componentes principais

Na nova base formada pela PCA um menor número de coordenadas é suficiente para conter grande parte da informação explicada pela distribuição dos dados, ou seja, as primeiras CPs apresentam grande variância dos dados, enquanto outras têm variância próxima de zero. Assim, quanto maior a porcentagem de variância nos dados puder ser explicada pelas primeiras CPs, uma redução de dimensionalidade mais eficiente pode ser obtida. A porcentagem de variância total explicada por cada CP é representada pela razão da variância (autovalor) correspondente e da soma total das variâncias. (STATISTICS TOOLBOX, 1999), ou seja:

$$\% \text{ explicada} = 100 \cdot \frac{\lambda_i}{\sum_{i=1}^n \lambda_i} \quad \text{Equação (4.1)}$$

Onde,  $\lambda_i$  corresponde à variância.

Assim, a porcentagem de variância nos dados explicada pelas primeiras CPs é igual ao acúmulo das variâncias explicadas por cada CP.

Um aspecto crítico da PCA é seleção do número de componentes principais. Não há uma solução ideal para esta seleção, e existem discrepâncias entre os diferentes métodos (FERRÉ, 1995). Assim diversas formas para seleção das componentes principais têm sido aplicadas na literatura:

- SŸKRBICÄ, URISŸICÄ-MLADENOVICÄ e CVEJANO (2005); MATTEAU, ASSANI e MESFIOUI (2009) incluíram apenas as CPs com autovalor maior que um;
- RAN, QIAO e YE (2004), SILVA et al. (2005) selecionam as CPs suficientes para explicar uma taxa de variância acumulada superior a 90% (JOLLIFFE, 1972; MARDIA et al. 1979); VALLE, LI e QIN (1999) dentre os métodos analisados para seleção de CPs, abordaram o procedimento empírico *Scree Plot*, proposto por CATTELL (1966). Este gráfico relaciona cada CP com o autovalor associado. Se essa curva evidenciar uma estabilização dos autovalores, podem-se reter apenas as CPs com números de ordem superiores àqueles que iniciam a estabilização.
- ELSHENAWY et al.(2010), GOOD, KOST e CHERRY (2010), KOUTSOGIANNIS e SORAGHAN (2002) utilizam outros critérios estatísticos mais complexos.

FERRÉ (1995) apresenta uma comparação detalhada de todos esses critérios de seleção de CPs abordados, e analisou que os métodos baseados na porcentagem da variância explicada em algumas ocasiões podem selecionar CPs insuficientes ou inadequadas, a depender da escolha da porcentagem da variação acumulada especificada, e no caso da escolha das CPs por corte do autovalor, apesar das soluções serem mais precisas, são bastante distintas a depender da escolha do autovalor, o que torna difícil determinar qual método de escolha de autovalor adotar.

Para JOLLIFFE (1972) o número  $p$  de componentes principais significativas, alternativamente, pode ser definido pelo número de CPs necessárias para explicar uma porcentagem superior a 90% da variação total dos dados, ou, pelo número do autovalor ( $\lambda$ ) associado maior que 0,7, sendo este considerado por ele como uma melhor decisão sobre as principais CPs, e, portanto o adotado neste estudo.

#### 4.1.2 Descarte de variáveis

Aqui as componentes principais são usadas para julgar a importância das variáveis originais. Vale ressaltar que CPs selecionadas permite reduzir a dimensão do conjunto de dados, e não o número de variáveis originais, já que as CPs são combinações lineares de todas as variáveis originais.

Muitos autores têm estudado seleção de variáveis por PCA. Estes estudos sugerem diferentes maneiras de selecionar ou descartar variáveis originais, sendo alguns métodos mais complexos (GUO et al., 2002; BRUSCO, SINGH e STEINLEY, 2009) e outros, adaptações de métodos clássicos (SILVA, 2000; CADIMA, CERDEIRA e MINHOTO, 2004; CUMMING e WOOFF, 2007; STANIMIROVIC et al., 2008).

Neste trabalho será adotado o método clássico de descarte de variáveis B4 proposto por Jolliffe (1972), baseado na preservação da variação da maioria dos dados. O método B4 envolve o uso das  $p$  primeiras componentes principais selecionadas, onde a variável de maior valor absoluto de *loading*

correspondente a primeira CP é selecionada, em seguida, a próxima variável a ser selecionada será a de maior valor absoluto de *loading* correspondente a segunda CP, e assim sucessivamente até a  $p$  CP selecionada. Caso a mesma variável apresente maior valor absoluto de *loading* em CPs diferentes, então a variável a ser selecionada será a de segundo maior valor absoluto da CP correspondente. As variáveis não selecionadas serão descartadas. Desta forma, o número de variáveis selecionadas é igual ao número  $p$  de CPs.

### 4.1.3 Gráfico dos scores

O método de análise multivariada PCA é uma ferramenta utilizada também para identificação de possíveis *outliers*. Como mencionado anteriormente, a presença de *outliers* nos dados pode atrapalhar o processo de aprendizagem da rede, já que esta se trata de um tipo de modelagem empírica, baseada em informações de dados experimentais, que pode ter seu modelo de predição enfraquecido devido à infiltração dessas amostras.

Como as CPs podem ser encaradas como eixos de máxima distribuição dos dados; é interessante visualizar a disposição dos dados nestes novos conjuntos de eixos. Esta figura, formada pela projeção dos objetos nas componentes principais, é denominada de gráfico dos *scores*. Suas coordenadas são obtidas a partir do produto da matriz de dados pela matriz de autovetores (CPs). Porém, através deste gráfico é possível visualizar um grupo de variáveis correlacionadas através de duas CPs. (SILVA et al., 2005).

Após a seleção das CPs, se as duas ou três primeiras CPs explicarem uma quantidade significativa da variância total dos dados, a elaboração do gráfico dos *scores* com essas coordenadas pode ser útil para identificação de possíveis *outliers*, os quais se diferem do restante dos dados, e por isso devem ser analisados separadamente, junto de especialistas do processo da respectiva planta de tratamento. Entretanto, não existe um consenso quanto ao uso de quantas e quais componentes devem ser avaliados para a identificação destas amostras. Estudos relacionados à interpretação de dados multivariados por PCA colocam grande atenção as duas primeiras CPs, as quais retêm na maioria das vezes maior variabilidade dos dados. Esquerre-Oliveira et al., (2004) e Kristiansen et al., (2010) encontram aplicações desta metodologia.

Esquerre-Oliveira (2003) utilizou outro método para identificar possíveis *outliers*, o parâmetro estatístico  $T^2$  de Hotelling, uma maneira analítica de medir a distância multivariada de cada observação do centro do conjunto de dados. Diferente do gráfico dos *scores*, este permite utilizar todas as CPs ao mesmo tempo para identificação dos *outliers*. Os possíveis *outliers* encontrados em seu trabalho não apresentaram nenhuma anormalidade de acordo com os especialistas do processo e permaneceram no conjunto de dados para construção do modelo.

Apesar do gráfico dos *scores* analisar apenas duas CPs ao mesmo tempo, esta parece ser uma alternativa satisfatória para identificação de possíveis *outliers*,

por uso das duas primeiras CPs, já que se trata das componentes com maior variabilidade explicada dos dados originais.

## **4.2 Projeto de uma rede neural**

Há quatro especificações básicas para um bom desempenho de um projeto de redes neurais: coleta de dados, configuração da rede, treinamento e teste.

### **4.2.1 Coleta e seleção dos dados**

A construção da rede começa pela identificação e coleta dos dados históricos relevantes para o problema. Em seguida é feita a preparação e adequação dos dados ao formato requerido pela rede neural em preparação, ou seja, a formatação dos dados. Neste processo os dados são padronizados em uma nova escala, a depender da função de ativação utilizada. Por exemplo, para redes MLP com camadas intermediárias com função sigmoideal e de saída com função linear o intervalo de padronização é  $[0,1]$  e  $[-1,1]$ , respectivamente. O objetivo da normalização é diminuir a influência causada por valores que se destacam excessivamente em relação aos demais, ou seja, diminuir a distância entre os valores de variáveis muito espaçadas

Os dados são separados em três conjuntos:

- (a) treinamento, utilizado para estimar os parâmetros do modelo;
- (b) validação, utilizada para verificar a habilidade de generalização do modelo frente a amostras independentes do conjunto de treinamento, e
- (c) teste, utilizado para validar o modelo usando novas amostras.

É recomendável a reordenação aleatória dos dados para prevenção de tendências associadas à ordem de apresentação dos dados. Alguns softwares computacionais trazem seus próprios algoritmos de aleatorização, como é o caso do Matlab versão 7.8.0.347(2009), utilizado para construção dos modelos.

Na prática, o número de dados de treinamento tem limites. Então, é recomendável evitar vetores de entrada de grandes dimensões. Por isso é importante aplicar técnica de pré-processamento de dados para auxiliar na determinação da composição do vetor de entrada (LUDWIG JR. e MONTGOMERY, 2007).

### **4.2.2 Configuração da rede**

O modelo escolhido para fazer previsão de matéria orgânica é um modelo de rede neural tipo MLP com algoritmo de aprendizagem Levenberg-Marquardt, que é uma adaptação do algoritmo *backpropagation*. O algoritmo Levenberg-

Marquardt não foi abordado anteriormente, entretanto este algoritmo será descrito brevemente. Os parâmetros da rede alterados ao longo do treinamento com a finalidade de se encontrar um modelo com um resultado mais satisfatório foram os seguintes:

- Quantidade de camadas intermediárias ou ocultas: foram adotadas arquiteturas com até duas camadas ocultas;
- Número de neurônios da primeira camada intermediária: variou entre 1 e 10;
- Número de neurônios da segunda camada oculta, mais precisamente: 2, 3, 4 e 5 neurônios;
- A taxa de aprendizagem variou entre 0,01, 0,05 e 0,5;
- Variação da porcentagem do conjunto de dados de treino, validação e teste;

Os demais parâmetros da rede como função de ativação e método de treinamento não foram modificados ao longo do treinamento e receberam as mesmas funções e métodos para todas as configurações de rede possíveis, ou seja, para a função de ativação na camada intermediária foi utilizada a função sigmoideal, e para camada de saída a função de ativação foi a linear.

Ressalta-se que a função de ativação sigmoideal aplicada nos neurônios da camada intermediária é útil para introduzir a não linearidade nas RNAs. Sem linearidade, as camadas intermediárias não fariam as RNAs mais poderosas, seria apenas um simples *perceptron*. A função linear na camada de saída é apropriada para variáveis contínuas, como é o caso da DQO de saída da lagoa aerada. (SARASWATHI e SASEETHARAN, 2011).

O algoritmo Levenberg-Marquardt é uma derivação do método de Newton de convergência mais rápida. Para problemas de aproximação de função (regressão) de redes com centenas de pesos, este é o algoritmo de mais rápida convergência. É normalmente usado para redes com pouco ou médio conjunto de treinamento, já que este requer uma grande memória de armazenamento para sua execução. Caso a memória seja um problema então há uma variedade de algoritmos rápidos disponíveis em Demuth, Beale e Hagan (2010). Em muitos casos esse algoritmo é útil para obter baixo EQM comparado aos demais. Sua desvantagem é o armazenamento de matrizes que podem ser muito grandes para determinados problemas.

As iterações do algoritmo Levenberg-Marquardt são regidas pela equação:

$$w[i + 1] = w[i] + \Delta w \quad \text{Equação (4.2)}$$

Onde,  $w$  é o vetor de parâmetros da RNA;  $i$  é a iteração;  $\Delta w = (J^T J)^{-1} J(y_d - y)$ ,  $y_d$  é a saída predita,  $y$  é a saída medida pela RNA, sendo  $e = y_d - y$ , o resíduo e  $J$  é a matriz *Jacobiana* dada por:

$$J = \begin{bmatrix} \frac{\partial y}{\partial w_1} | n = 0 & \frac{\partial y}{\partial w_2} | n = 0 & \dots & \frac{\partial y}{\partial w_k} | n = 0 \\ \frac{\partial y}{\partial w_1} | n = 1 & \frac{\partial y}{\partial w_2} | n = 1 & \dots & \frac{\partial y}{\partial w_k} | n = 1 \\ \frac{\partial y}{\partial w_1} | n = N & \frac{\partial y}{\partial w_2} | n = N & \dots & \frac{\partial y}{\partial w_k} | n = N \end{bmatrix} \quad \text{Equação (4.3)}$$

Cada gradiente é ponderada de acordo com sua curvatura. Assim há uma grande tendência de convergência na direção na qual o gradiente é menor.

Maiores detalhes sobre o desenvolvimento matemático deste algoritmo pode ser visto em França et al. (2009).

Assim, os modelos de RNAs foram estruturados em uma camada de entrada, onde o número de neurônios é igual ao número de variáveis de *input*, uma ou duas camadas ocultas com quantidades variadas ao longo do treinamento, e com apenas um neurônio na camada de saída, que representa a previsão diária de matéria orgânica de saída da uma lagoa aerada de uma estação de tratamento de efluentes de uma empresa de produção de papel e celulose.

Entretanto, o número de *inputs* pode ser reduzido através do uso de técnicas estatísticas de compressão de dados como é o caso da PCA, abordada nesta pesquisa, descrita em detalhes nas Seções 3.1, 4.1.1 e 4.1.2. Tal redução permite a eliminação de informações irrelevante tais como ruído e redundâncias presentes na matriz de dados. O sucesso da redução do número de variáveis de *input* da rede pode resultar no aumento da velocidade de treinamento, menor memória de armazenamento, melhor habilidade de generalização do modelo, obtenção de um modelo mais robusto com respeito aos ruídos nas medidas e representação de um modelo mais simples.

### 4.2.3 Treinamento

Uma vez que os pesos e os *bias* são inicializados, a rede já está pronta para o treinamento. No processo de treinamento são utilizados os conjuntos de dados de treinamento e validação que contém as variáveis selecionadas para a construção do modelo – *inputs* e saídas desejadas. Durante o treino, os pesos e os *bias* são iterativamente ajustados para minimizar a função de *performance* da rede. A função de *performance* padrão para redes *feedforward* é o erro quadrático médio entre a saída da rede e a saída desejada.

Valores do erro do treinamento usado na simulação foram da ordem de  $10^{-7}$ . Cada possível configuração utilizada foi apresentada para treinamento um total de 10 vezes, pois como a rede sempre inicia seu treinamento com pesos aleatórios, é possível de se obter um bom resultado com uma determinada configuração que não necessariamente fornece a melhor solução.

Quanto ao tempo de treinamento, podem ser adotados alguns indicadores, dentre os quais o número máximo de ciclos, a taxa de erro médio por ciclo, ou ainda, a capacidade de generalização da rede. Pode ocorrer que em um determinado instante do treinamento a generalização comece a se degenerar,

causando o *overfitting*. Então é sugerido que o treinamento seja interrompido quando o erro do conjunto de validação começar a subir, ou seja, quando a rede começar a incorporar o ruído presente nos dados, o que causa degradação na sua capacidade de generalização.

Há vários métodos para a determinação do momento em que o treinamento deve ser encerrado. Estes métodos são chamados de critérios de parada. Os critérios de parada sugeridos por Demuth, Beale e Hagan (2010) são:

- Encerrar o treinamento após N ciclos;
- Encerrar o treinamento até o EQM ficar abaixo de um determinado valor;
- Encerrar o treinamento quando o erro no conjunto de validação aumentar  $n$  vezes, em que  $n$  é o número de interações especificado pelo projetista (na linguagem de programação é o `Net.trainParam.max_fail`), entretanto, os pesos e *bias* do erro mínimo alcançado são retornados.

Se o resultado de treinamento da rede neural não for satisfatório, algumas medidas são sugeridas:

- aumento do número de neurônios na camada oculta;
- aumento do número dos *inputs*;
- tentar treinar a rede com outro algoritmo de treinamento.

#### 4.2.4 Teste

Durante esta fase o conjunto de teste é utilizado para avaliação do desempenho da rede com novos dados e também para comparar modelos.

Se o erro no conjunto de teste atingir o valor mínimo em um número de interação significativamente diferente que o erro do conjunto de validação, isso pode indicar uma pobre divisão do conjunto de dados. No ANEXO 3 estão os gráficos com as *performances* dos modelos de predição durante o treinamento, através deles é possível analisar se o conjunto de teste e validação tem características semelhantes.

Há outros testes como a análise dos pesos sinápticos e níveis de *bias*, pois se existirem valores muito pequenos, as conexões associadas podem ser consideradas insignificantes e assim serem eliminadas. De modo inverso, valores muito maiores que os outros indicam a possibilidade de *overfitting* da rede. (LUDWIG JR. e MONTGOMERY, 2007).

### 4.3 Avaliação da eficiência dos modelos MLP

#### *Parâmetros de avaliação*

Para avaliar o desempenho dos modelos de predição de matéria orgânica, valores reais desse parâmetro do conjunto de teste são comparados aos

valores preditos pelo modelo de RNA. Os métodos de avaliação do desempenho dos modelos de RNAs mais citados na literatura são o EQM e  $R^2$  (HAMED, KHALAFALLAH e HASSANIEN, 2004; OLIVEIRA-ESQUERRE et al., 2004; AKRATOS, PAPASPYROS e TSIHRINTZIS, 2008).

O  $R^2_{ajustado}$  foi adicionado como parâmetro de avaliação de desempenho de modelo MLP para avaliar a complexidade do modelo, em outras palavras, avaliar se a redução de parâmetros a serem ajustados influencia ou não desempenho dos modelos.

Uma diferença significativa entre  $R^2$  e  $R^2_{ajustado}$  indica que o modelo está com excesso de parâmetros, ou seja, termos que não contribuem significativamente ao ajuste foram incluídos.

Para Saraswathia e Saseetharan (2011) não há parâmetro estatístico semelhante ao  $R^2_{ajustado}$  para realizar análise comparativas dos desempenhos dos modelos de predição de RNAs, e os métodos propostos na literatura levam a resultados contraditórios.

#### *Análise dos resíduos*

Os resíduos são definidos por:

$$e_i = y_i - \hat{y}_i \quad \text{Equação (4.4)}$$

A análise dos resíduos é frequentemente útil na verificação da suposição de que erros sejam distribuídos de forma aproximadamente normal, com variância constante, assim como na determinação da utilidade dos termos adicionais no modelo.

Como verificação aproximada da normalidade, o experimentalista pode construir um histograma de frequência dos resíduos ou um gráfico de probabilidade normal dos resíduos. Outra forma de analisar a normalidade dos resíduos pode ser pela padronização dos mesmos, calculado por:

$$d_i = \frac{e_i}{\sqrt{\sigma^2}} \quad \text{Equação (4.5)}$$

Os resíduos padronizados são escalonados de modo que seus desvios-padrão sejam aproximadamente iguais a um, assim os resíduos grandes (que podem indicar possíveis *outliers* ou observações não usuais) serão mais óbvios a partir da inspeção dos gráficos residuais. Se os erros forem distribuídos normalmente, então aproximadamente 95% dos resíduos padronizados devem cair no intervalo (-2,+2). (MONTGOMERY, RUNGER e HUBELE, 2004).

#### **4.4 Considerações finais sobre o capítulo**

Neste capítulo foi apresentada a metodologia utilizada nesta pesquisa, que constitui no pré-processamento dos dados aplicado por PCA, nas etapas do projeto de uma RNA (coleta de dados, configuração da rede, treinamento e teste) e os métodos de avaliação dos modelos de predição.

A PCA é utilizada com três propósitos: redução dimensional do conjunto de dados; descarte de variáveis e exclusão de *outliers*.

Para redução dimensional foi utilizado o método proposto por Jolliffe (1972), onde as principais CPs devem ter autovalor (variância) maior que 0,7. Para descarte de variáveis o mesmo propôs um método baseado nas CPs selecionadas, o método B4, o qual procede em selecionar uma variável por componente selecionada, que deve ser a variável de maior valor absoluto de *loading*. E por fim, para exclusão de *outliers*, é proposta a análise do gráfico dos *scores* das duas primeiras CPs selecionadas.

Os dados de cada conjunto foram aleatorizados e divididos em três conjuntos, treino, validação e teste.

Os modelos MLP foram estruturados em uma camada de entrada, onde o número de neurônios é igual ao número de variáveis de *input*, uma ou duas camadas ocultas com quantidades variadas ao longo do treinamento, e com apenas um neurônio na camada de saída, matéria orgânica, medida por DQO de saída da lagoa aerada da estação de tratamento de efluentes de uma empresa de produção de papel e celulose. O algoritmo de aprendizagem aplicado para ajuste dos parâmetros livres (pesos sinápticos e *bias*) é o Levenberg-Marquardt, que é uma adaptação do *backpropagation*. As funções de transferência utilizadas são a sigmoide e linear, para a camada intermediária e de saída, respectivamente.

Os modelos MLP foram simulados por diversas vezes utilizando-se 1 ou 2 camadas intermediárias, variando-se também o número de neurônios intermediários e a taxa de aprendizagem de 0,1, 0,01 e 0,05.

O EQM e  $R^2$  são os parâmetros mais citados na literatura, o  $R^2_{ajustado}$  é utilizado para avaliar a complexidade do modelo, a qual será também avaliada a fim de obter um modelo menos complexo. Para examinar a adequação do modelo foi utilizada a análise dos resíduos padronizados.

## Capítulo 5

### RESULTADOS E DISCUSSÕES

#### 5.1 Pré-processamento dos dados

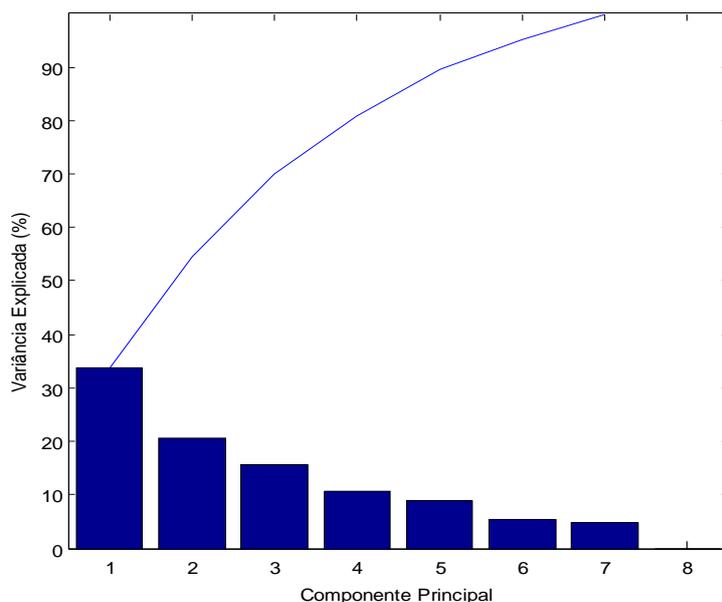
Inicialmente a PCA foi aplicada para reduzir a dimensão do conjunto de dados de entrada da RNA por seleção das CPs. Na Tabela 5.1 são apresentados os resultados das variâncias (autovalores) e das variâncias explicadas e acumuladas das componentes principais obtidas.

**Tabela 5.1: Variâncias, variância explicada e acumulada das CPs**

<b>Componentes Principais</b>	<b>Variância</b>	<b>Variância explicada (%)</b>	<b>Variância acumulada (%)</b>
PC <sub>1</sub>	<b>2,70</b>	33,79	33,79
PC <sub>2</sub>	<b>1,65</b>	20,62	54,41
PC <sub>3</sub>	<b>1,25</b>	15,68	70,09
PC <sub>4</sub>	<b>0,85</b>	10,65	80,74
PC <sub>5</sub>	<b>0,72</b>	9,01	89,75
PC <sub>6</sub>	0,44	5,48	95,24
PC <sub>7</sub>	0,38	4,76	100
PC <sub>8</sub>	0	0	100

De acordo com Jolliffe (1972) as principais CPs, aquelas com variância maior que 0,7, correspondem as cinco primeiras CPs, as quais expressam 89,75% da variância preservada dos dados originais, ou seja, as oito variáveis originais podem ser analisadas a partir dessas cinco CPs, o que garante uma reprodução da variabilidade de aproximadamente 90%. As CPs não selecionadas (CP<sub>6</sub> a CP<sub>8</sub>) expressam 5,48%, 4,76% e 0% da variância total, respectivamente. Por fim, com a seleção das principais CPs, o conjunto de dados com oito variáveis sintéticas (CPs) e 786 amostras foi reduzido a cinco variáveis sintéticas (CPs) e 786 amostras, formando assim o conjunto de Dados 3.

A Figura 5.1 apresenta os resultados em gráfico de barras mostrando a variância explicada de cada CPs.



**Figura 5.1: Gráfico de barras das CPs**

Através desse gráfico é possível verificar que a maior quantidade de variância explicada dos dados é representada pela primeira componente principal ( $CP_1$ ), a qual explica 34% da variância total dos dados originais. As cinco primeiras componentes selecionadas para representação dos dados apresentam 89,75% da variância explicada (JOLLIFFE, 1972), esse valor é observado pela curva no gráfico, que representa a variância acumulada das CPs.

Através das componentes selecionadas serão extraídas as principais variáveis originais pelo valor absoluto dos *loadings* como descrito pelo método B4. A Tabela 5.2 mostra os *loadings* das variáveis originais para as cinco CPs.

**Tabela 5.2: Loadings das CPs selecionadas**

Componentes Principais	Q	DQO <sub>carga</sub>	pH	Cor	T	EC	Q <sub>celulose</sub>	Q <sub>papel</sub>
CP <sub>1</sub>	<b>-0,49</b>	<b>-0,49</b>	0,27	-0,17	-0,28	0,38	-0,37	-0,24
CP <sub>2</sub>	-0,42	-0,42	-0,35	-0,05	0,37	0,07	0,39	<b>0,48</b>
CP <sub>3</sub>	0,06	0,06	0,53	<b>0,58</b>	0,32	0,49	0,16	0,10
CP <sub>4</sub>	-0,11	-0,11	-0,26	0,70	-0,35	-0,22	<b>-0,39</b>	0,31
CP <sub>5</sub>	-0,15	-0,15	-0,18	0,27	<b>0,58</b>	-0,29	-0,17	-0,64

De acordo com os resultados mostrados na Tabela 5.2, as variáveis selecionadas, ou seja, as variáveis com maior valor absoluto de *loading* correspondente a cada CP são: vazão (Q), DQO<sub>carga</sub>, Cor, temperatura (T), produção de celulose (Q<sub>celulose</sub>) e produção de papel (Q<sub>papel</sub>). Nota-se que a variável Cor apresenta maior valor absoluto de *loading* em duas CPs (CP<sub>3</sub> e CP<sub>4</sub>), porém, como a Cor já havia sido selecionada pela CP<sub>3</sub>, a Q<sub>celulose</sub>, próxima variável de maior valor absoluto de *loading* correspondente a CP<sub>4</sub>, foi

a variável selecionada. O mesmo procedimento foi utilizado para seleção da variável correspondente a CP<sub>5</sub>, onde a T foi selecionada.

Como a  $DQO_{carga}$  está em função da Q,  $DQO_{carga}=f(Q,DQO)$ , ambas variáveis apresentam o mesmo valor de *loading* nas CPs, desta forma, apenas a DQO é mantida como *input* para modelagem. A partir disso, as variáveis selecionadas para formar o conjunto de Dados 4 são:  $DQO_{carga}$ , Cor, T,  $Q_{celulose}$  e  $Q_{papel}$ .

Para entender como cada variável impacta em cada uma das componentes, uma estratégia comumente utilizada é o gráfico dos *scores* e *loadings*. Esse gráfico mostra quais variáveis contribuem para cada uma das componentes calculadas e qual é o grau dessa contribuição. O gráfico dos *scores* e *loadings* das componentes 1 e 2 é apresentado na Figura 5.2. Pode-se constatar que os valores dos *loadings* da  $DQO_{carga}$  e Q estão positivos, diferentes dos calculados pela PCA, isso deve ser pelo fato do comando do Matlab que permite gerar este gráfico (biplot) impor uma convenção de sinais, forçando o elemento com maior magnitude em cada CP ser positivo. Tal ação não interfere na interpretação dos resultados.

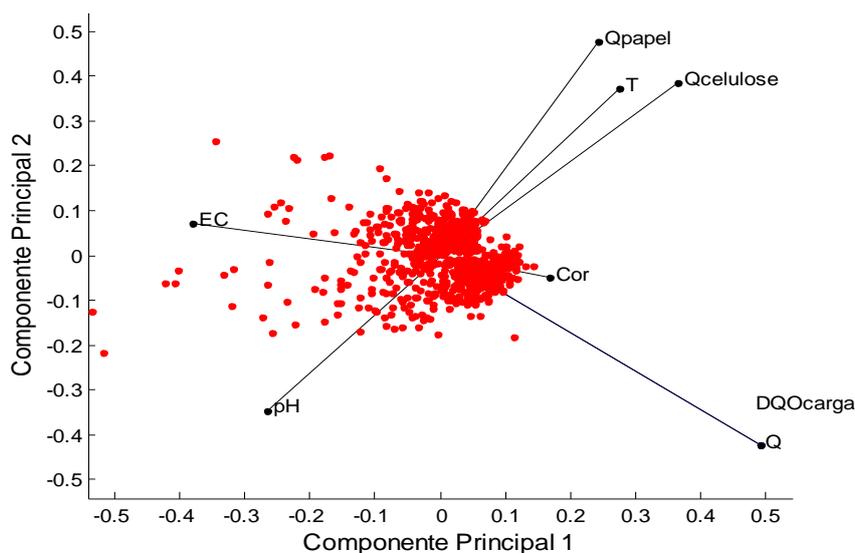


Figura 5.2: Gráfico dos scores das duas primeiras CPs

Através da Figura 5.2 pode-se observar que a  $Q_{papel}$ ,  $Q_{celulose}$  e T estão correlacionadas, assim como a Q e  $DQO_{carga}$ , como já era esperado. O pH e a EC são descorrelacionados, ou seja, são independentes.

O uso do gráfico dos *scores* é uma alternativa viável na identificação de *outliers*. A identificação dessas amostras envolve um alto grau de subjetividade, pois os dados considerados com *outliers* apresentam-se relativamente distantes da massa de dados, o que pode indicar a presença de amostras influentes ou *outliers* na matriz de Dados 2. Analisam-se apenas as primeiras CPs, já que estas retêm individualmente maior variabilidade explicada dos dados. Assim os dados questionáveis sobre possíveis *outliers* são identificados na Figura 5.3.



Tabela 5.3: Resultados dos modelos de predição para o conjunto de teste

Parâmetros comparativos	MODELOS				
	M1 <sup>1</sup>	M2 <sup>2</sup>	M3 <sup>2</sup>	M4 <sup>2</sup>	M5 <sup>2</sup>
Inputs	8 Variáveis originais	8 Variáveis originais	5 PCs	5 Variáveis originais	8 Variáveis originais
Número de neurônios na camada oculta	1	1	1	1	1
Número de interações	11	103	18	93	103
EQM	2,3E-03	4,59E-08	1,93E-05	2,98E-08	4,02E-05
C(n)	11	11	8	8	11
R <sup>2</sup>	0,45	1	0,99	1	1
R <sup>2</sup> <sub>ajustado</sub>	0,37	1	0,97	1	1

<sup>1</sup> DQO (mg.L<sup>-1</sup>) e <sup>2</sup> DQO (kg de DQO.dia-1)

Nota-se que o modelo M1 foi o único que apresentou baixo desempenho, avaliando EQM, R<sup>2</sup> e R<sup>2</sup><sub>ajustado</sub>. Isso é devido à utilização da DQO como carga em vez de concentração melhor representar a quantidade de matéria orgânica dissolvida no efluente, o que influenciou uma significativa melhora no desempenho dos modelos (M2 ao M5).

Outras pesquisas envolvendo modelos de predição de matéria orgânica com DQO como concentração apresentam valores razoáveis de R<sup>2</sup> (entre 0,4 a 0,7) como pode ser visto em Hamed, Khalafallah e Hassanien (2004), Oliveira-Esquerre et al. (2004), Grieu, Polit e Colprim (2005); Akrotos, Papaspyros e Tsihrintzis (2008), May e Sivakumar (2009), Basant et al. (2010).

Assim como o R<sup>2</sup>, os valores de R<sup>2</sup><sub>ajustado</sub> não mostraram nenhuma diferença significativa entre os modelos com pré-processamento para redução do número de variáveis de *input* da rede (M3 e M4), o que significa que o número de parâmetros a serem ajustados podem ser reduzidos sem afetar o desempenho do modelo de predição. Essa redução permite a eliminação de informações irrelevante tais como ruído e redundâncias presentes na matriz de dados. Desta forma, pode-se dizer que a PCA é útil para reduzir a complexidade de modelos MLP. Vale ressaltar que o modelo M4 é mais satisfatório, pois além de ser o modelo mais simples, apresenta o menor valor de erro (EQM).

A taxa de erro é geralmente mais significativa, por se tratar de uma rede do tipo supervisionada. Deve-se notar também, que apesar do modelo M3 apresentar apenas cinco *inputs* na RNA como variáveis preditoras, que são as CPs, estas requerem informações das oito variáveis originais.

Os resultados obtidos por PCA para exclusão de possíveis *outliers*, M5, mostrou resultados similares ao M2, o que representa que não houve melhora no desempenho do modelo, ou seja, o método de exclusão de *outliers* foi desnecessário nesse estudo de caso. Mesmo com aplicação sem sucesso, é importante excluir *outliers*, principalmente quando se trata de modelagem empírica, pois eles podem influenciar significativamente o desempenho do modelo, fornecendo resultados enganosos ou incorretos.

Diferente desse resultado, Wu, Chau e Fan (2010) obtiveram sucesso no desempenho do modelo de predição ao integrar PCA para exclusão de *outliers*.

Julgando-se todos esses fatores, pode-se considerar o melhor modelo de predição de DQO de saída de uma lagoa aerada o modelo M4, cuja DQO é medida como carga orgânica, tal modelo é mais simples e econômico, já que descarta a medição de duas variáveis do processo, pH e EC, e utiliza apenas cinco *inputs* na RNA. A vazão é considerada na obtenção da DQO<sub>carga</sub>.

A Tabela 5.4 apresenta os resultados dos pesos sinápticos relacionados a cada variável de entrada dos modelos.

**Tabela 5.4:** Resultados dos pesos sinápticos dos modelos

Modelos	DQO	Variáveis de entrada da MLP							
		Q	DQO	pH	Cor	T	EC	Q <sub>celulose</sub>	Q <sub>papel</sub>
M1	mg/L	-0,1514	-1,0645	0,1975	-0,0483	-0,0391	-0,3931	-0,0707	-0,0647
M2	Kg/dia	-0,1243	-0,3201	0,0001	0,0003	0,0001	-0,0003	0,0002	0,0001
M4	Kg/dia	-	-0,46	-	-0,003	0	-	-0,0002	-0,0001
M5	Kg/dia	-0,1243	-0,3201	0,0001	0,0003	0,0001	-0,0003	0,0002	0,0001
		<b>PC<sub>1</sub></b>	<b>PC<sub>2</sub></b>	<b>PC<sub>3</sub></b>	<b>PC<sub>4</sub></b>	<b>PC<sub>5</sub></b>			
M3	Kg/dia	-1,2451	-0,6821	0,1027	-0,1672	-0,1703	-	-	-

Através dos resultados mostrados na Tabela 5.4 observa-se que os pesos sinápticos das variáveis pH, Cor, T, EC, Q<sub>celulose</sub> e Q<sub>papel</sub> apresentam valores insignificantes se comparados aos pesos das variáveis Q e DQO dos modelos M2, M4 e M5. O mesmo não acontece para os modelos M1 e M3, nos quais, os pesos sinápticos apresentam valores na mesma ordem de grandeza. Este fato parece indicar que as informações da DQO como concentração (mg/L) e o uso de PCs como variáveis de entrada da rede MLP são satisfatórias para predição da matéria orgânica de saída da lagoa aerada. Entretanto, para melhor entendimento desses pesos sinápticos é necessário realizar uma análise de sensibilidade, como esta não foi realizada na presente pesquisa, considera-se todos para construção do modelo MLP.

Contudo, o melhor modelo julgado, M4, é representado pela Equação 5.1:

$$DQO_{saída} = \frac{8,67}{1 + e^{-0,46DQO_{entrada} - 0,003cor - 0,0002Q_{papel} - 0,0001Q_{celulose} + 0,0017}} - 4,3$$

Esse modelo também dispensa a medição da Temperatura, e desta forma, é reduzido a quatro variáveis dependentes para obtenção da quantidade de carga orgânica na saída da lagoa aerada, DQO<sub>saída</sub>

Na Figura 5.4 são apresentados gráficos da DQO medida versus a predita pelos modelos M1 a M5 para o conjunto de teste.

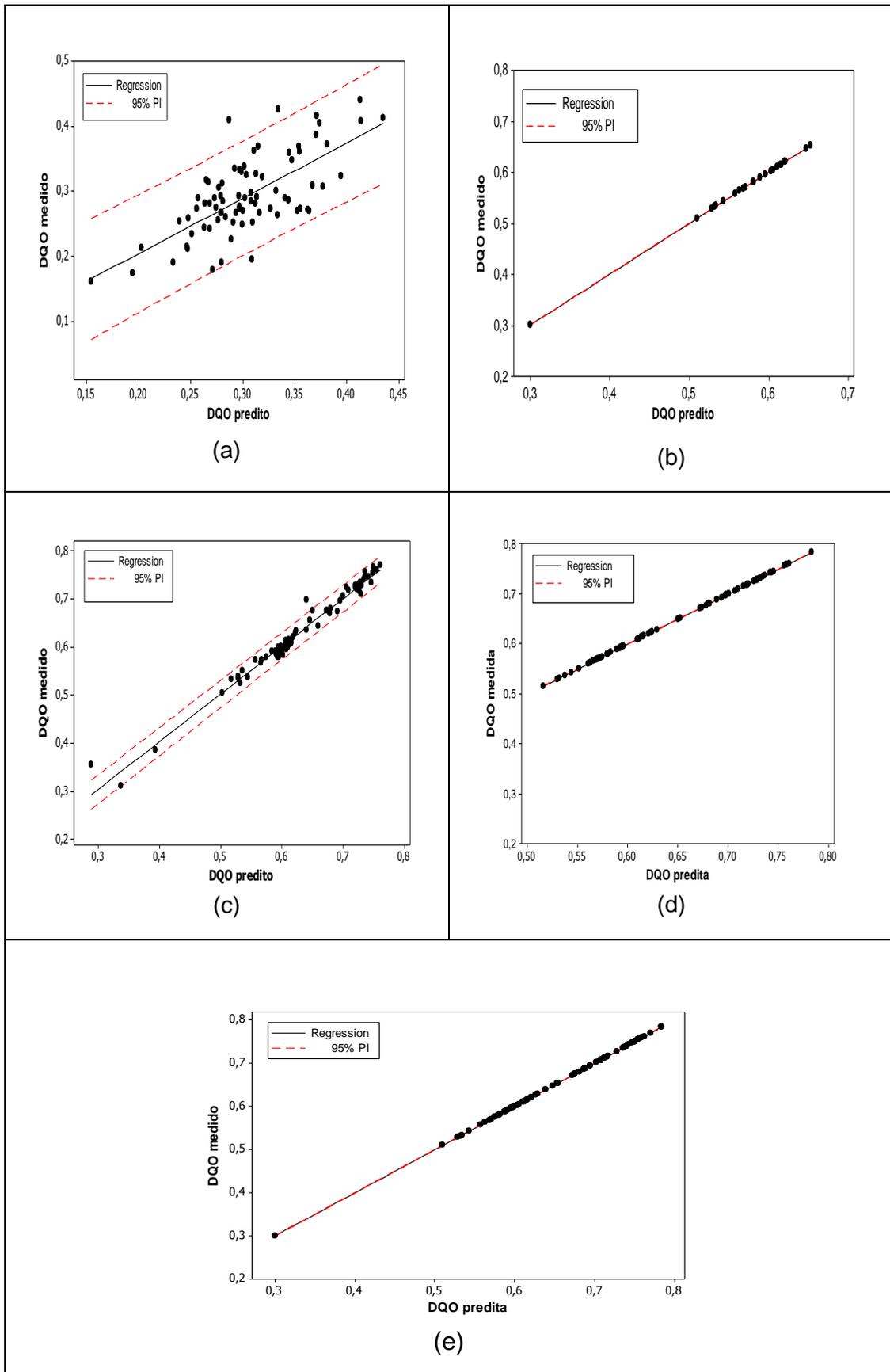
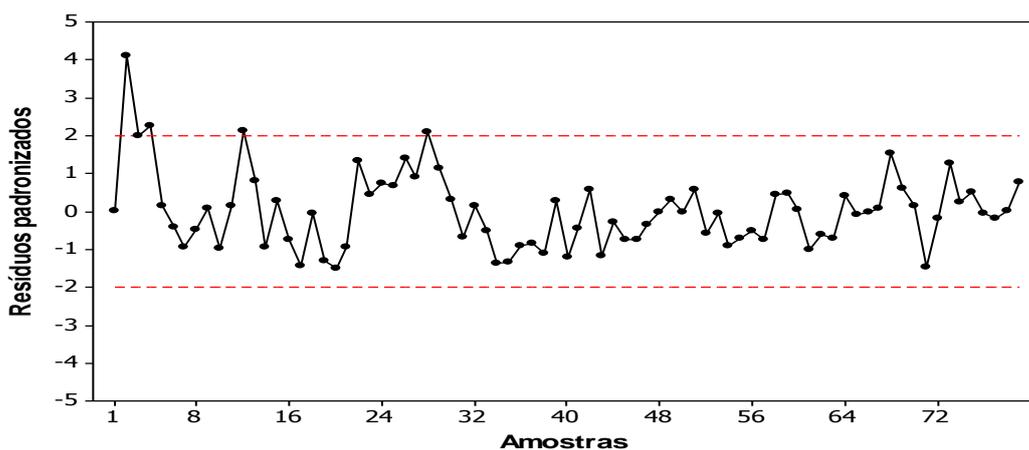


Figura 5.4: Relação entre a DQO de saída medida versus a predita para os modelos (a) M1, (b) M2, (c) M3, (d) M4 e (e) M5, considerando um intervalo de predição de 95% de confiança.

Observa-se que o modelo M1, embora seja o modelo de menor desempenho, é capaz de reproduzir bons resultados de predição da  $DQO_{carga}$ . O modelo M2 reproduz perfeitamente a variação total observada no bioprocesso. No modelo M3, 97% de futuras observações estão dentro da faixa de predição. O modelo M4, assim como o M2 também reproduz perfeitamente a predição de futuras observações. O modelo M5 como era esperado, devido à semelhança com o modelo M2, também reproduz perfeitamente a variação total observada no bioprocesso.

No teste de normalidade de Anderson-Darling para um nível de confiança de 95%, os modelos M2, M3 e M5 apresentam alguns resíduos extremamente distantes da linha reta desenhada pelos resíduos restantes, resultando assim a não normalidade dos resíduos e, portanto, a inadequação desses modelos. Os resultados do teste de Anderson-Darling para análise da normalidade dos resíduos estão no ANEXO 3. Os resíduos padronizados são mais úteis quando se verifica a magnitude residual. Para todos os modelos, 95% dos resíduos padronizados encontram-se dentro do intervalo de  $\pm 2$  desvios padrões, indicando a normalidade dos resíduos e, portanto, adequação dos modelos. O gráfico temporal dos resíduos padronizados do modelo M4 é mostrado na Figura 5.5.



**Figura 5.5: Gráfico de série temporal dos resíduos padronizados do modelo M4 – Linha superior e inferior indica um intervalo de confiança de 95%.**

Em síntese, a técnica de PCA neste estudo mostrou-se ineficaz apenas quando utilizada com a finalidade de exclusão de *outliers*. No caso do descarte de variáveis originais, foi obtido um modelo mais simples, sem perda de desempenho da rede. Para este modelo, um gráfico de série temporal da DQO medida e predita para o conjunto de teste é apresentado na Figura 5.6.

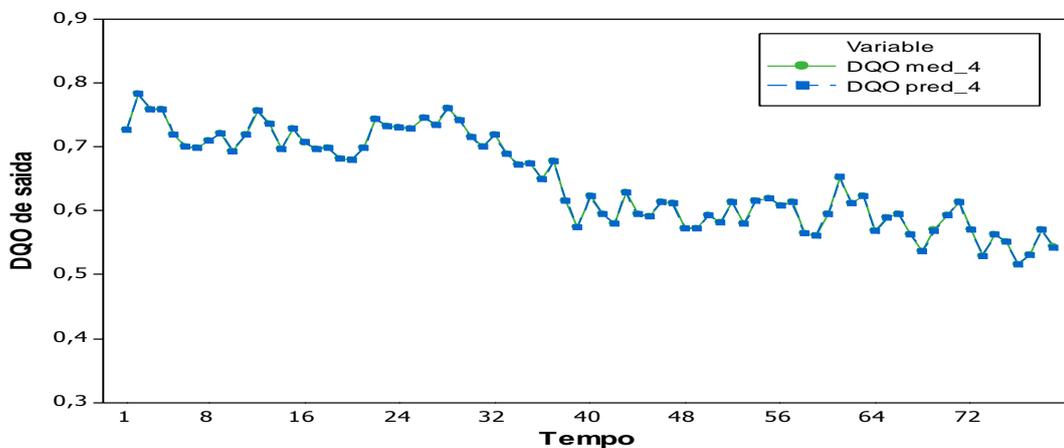


Figure 5.6: Gráfico de série temporal da DQO de saída medida e predita

## 5.1 Considerações finais

Este capítulo apresentou os resultados de cinco modelos de predição da DQO de saída da lagoa aerada de uma empresa de produção de papel e celulose, por uso de uma rede do tipo MLP. Os modelos são diferenciados por uso dos dados com e sem pré-processamento por PCA, além do uso da DQO como unidade de concentração e como carga orgânica. A PCA foi aplicada nesta pesquisa com a finalidade de redução da dimensão do conjunto de dados através da seleção de componentes principais (M3), descarte de variáveis originais (M4) e exclusão de possíveis *outliers* (M5).

Como resultado da redução dimensional do conjunto por PCA, obtido pelo método proposto por Jolliffe (1972), foi obtido um conjunto de dados transformados composto por cinco componentes principais que contêm 89,75% da variância preservada dos dados originais. O método B4 foi utilizado para descarte de variáveis, o qual considerou apenas cinco variáveis como necessárias para construção do modelo de predição,  $DQO_{carga}$ , Cor, T,  $Q_{papel}$ ,  $Q_{celulose}$ . Os *outliers* foram identificados pelo gráfico dos *scores* das duas primeiras CPs.

Concluído a etapa de pré-processamento, foram apresentados os melhores modelos de predição obtidos para o conjunto de teste. Com base nesses resultados os seguintes pontos devem ser evidenciados:

- De um modo geral, os resultados indicam que a representação da matéria orgânica como carga ( $kg\ DQO \cdot dia^{-1}$ ) ao invés da concentração ( $mg\ DQO \cdot L^{-1}$ ) representa melhor a quantidade de matéria orgânica dissolvida no efluente, o

que influenciou uma significativa melhora no desempenho dos modelos M2 ao M5, obtidos pelos Dados 2 ao 5, respectivamente.

- Os valores de  $R^2$  e  $R^2_{\text{ajustado}}$  não mostraram nenhuma diferença significativa para os modelos M2 ao M5;
- Os menores valores de EQM são obtidos nos modelos M2 e M4, os quais usam as variáveis originais como entrada;
- Os resultados do modelo construído com aplicação de PCA para exclusão de possíveis *outliers*, representada pelo modelo M5, não foram eficientes;
- O melhor modelo obtido foi o M4, mais simples e econômico, o qual descarta o uso de três variáveis para construção do modelo MLP, T, pH e EC.

## Capítulo 6

# CONSIDERAÇÕES FINAIS

### 6.1 Conclusões

Nesta pesquisa foram construídos modelos de predição da quantidade de matéria orgânica, representada pela DOQ de saída da lagoa aerada, de uma estação de tratamento de efluentes industrial. Para tanto, foram construídos modelos de redes neurais artificiais e, como técnica de pré-processamento de dados, a análise de componentes principais.

Para avaliar o desempenho dos modelos com aplicação desta técnica, cinco modelos foram comparados. M1 e M2 utilizando a demanda química de oxigênio como concentração e carga, respectivamente. M3 ao M5 com aplicação de PCA para redução dimensional do conjunto de dados de entrada da RNA com os respectivos propósitos: seleção das CPs, descarte de variáveis originais e exclusão de possíveis *outliers*.

De acordo com os resultados encontrados, apenas o modelo M1 apresentou desempenho inferior aos demais, apesar de ainda satisfatório considerando o EQM e o teste de normalidade de seus resíduos.

Os cinco modelos testados apresentam informações precisas, o que torna possível conhecer como o sistema está evoluindo (se, possivelmente, para um estado crítico que exige uma ação do operador). Esta informação de forma eficaz, contribuiu conhecimento de peritos no processo.

Embora exista muitas técnicas de pré-processamento de dados na literatura, a diferença da PCA em relação as demais é sua aplicação para diferentes propósitos.

Os resultados mostraram que a PCA quando aplicada para reduzir a dimensão do conjunto de dados por seleção das CPs (M3), embora não melhore o desempenho do modelo, é eficiente no aprendizado da rede, reduzindo o número de parâmetros a serem ajustados e desta forma contribuindo para formação de um modelo mais robusto. Porém quando aplicada para exclusão de possíveis *outliers* nenhuma informação extra foi identificada. A PCA aplicada para descarte de variáveis elimina duas medições de entrada da MLP (EC e pH sendo a vazão (Q) função da  $DQO_{entrada}$ ) resultando um modelo com quatro variáveis dependentes:  $DQO_{entrada}$ , Cor,  $Q_{papel}$  e  $Q_{celulose}$ . A temperatura (T) foi dispensada, pois, apresentou peso sináptico zero para a configuração adotada. Desta forma, o modelo M4 além de simples também é econômico, já que dispensa três medições desnecessárias de variáveis, duas determinada pela PCA e uma pelo próprio modelo MLP.

Esta pesquisa buscou avaliar a importância do pré-processamento de dados em redes neurais com diferentes objetivos. Embora a técnica aplicada neste estudo não tenha mostrado eficiência em todas as suas aplicações, existem outras técnicas na literatura que podem ser exploradas. Vale ressaltar que a

escolha do melhor modelo de redes neurais não deve ser feita de forma indiscriminada e descuidada, sendo necessário o uso de diversos parâmetros estatísticos para auxiliar na escolha e na comparação entre modelos de diferentes dimensões e estruturas, e que o pré-processamento dos dados para ser significativo deve ser acompanhado por um especialista do processo. Assim a integração das opiniões de um especialista e do projetista de RNAs é uma questão importante e relevante.

## 6. Sugestões para Trabalhos Futuros

Durante o desenvolvimento desta pesquisa foram identificados alguns pontos que necessitam de uma investigação mais profunda e servem como continuação deste trabalho, estes pontos são:

- Testar outros métodos de pré-processamento de dados existentes na literatura, tais como o proposto por: Olja Stanimirovic (2008) para seleção de CPs; Good *et al.* (2010) para seleção de variáveis originais; Yuan Li e Hiroyuki Kitagawa (2007) para identificar possíveis *outliers*.
- Utilizar o gráfico estatístico  $T^2$  de Hotelling para identificação de possíveis *outliers*, e comparar os resultados aos do gráfico dos *scores*;
- Avaliar a importância de cada peso sináptico dos modelos associando-os a aspectos fenomenológicos;
- Levantar técnicas para exclusão de pesos sinápticos;
- Testar outras variáveis de entrada para o modelo de predição, como a quantidade de oxigênio dissolvido introduzida na lagoa, no caso de lagoas aeradas, com objetivo de otimizar o consumo de energia durante o tratamento biológico;
- Comparar o desempenho do modelo para diferentes algoritmos de aprendizagem existentes na literatura;
- Construir um modelo de predição de DBO de entrada e saída através da DQO em função da exigência da legislação ambiental vigente para descarte de efluentes;
- Aplicar PCA com as três finalidades a um conjunto de dados e analisar o efeito dos três juntos.

## REFERÊNCIAS

- AKRATOS, C.S., PAPASPYROS, J.N.E., TSIHRINTZIS, V.A., 2008. An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands. *Chemical Engineering Journal* v. 143, pp. 96–110.
- BASANT, N., *et al.*, 2010. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water — A case study. *Chemometrics and Intelligent Laboratory Systems*, v.104, pp.172–180.
- BRUSCO, M., SINGH, R., STEINLEY, D., 2009. Variable neighborhood search heuristics for selecting a subset of variables in principal component analysis. *Psychometrika*, v. 74, n. 4, pp. 705–726.
- CADIMA, J. , CERDEIRA, J.O. , MINHOTO, M., 2008. Computational aspects of algorithms for variable selection in the context of principal components. *Computational Statistics & Data Analysis*, v. 47, pp. 225 – 236.
- CUMMING, J.A., WOOFF, D.A., 2007. Dimension reduction via principal variables. *Computational Statistics & Data Analysis*, v. 52, pp. 550 – 565.
- DEMUTH, H., BEALE, M., HAGAN, M., 2010. Neural Network Toolbox User's Guide. The MathWorks, Inc. 901 f. 2010.
- ELSHENAWY, L.M. *et al.*, 2010. Efficient Recursive Principal Component Analysis Algorithms for Process Monitoring. *Ind. Eng. Chem. Res.*, v. 49, pp. 252–259.
- FERRÉ, L., 1995. Selection of components in Principal Component Analysis: A comparison of methods. *Computational Statistics & Data Analysis*, v. 19, pp. 669 – 682.
- FARIA, E.L. *et al.*, 2008. Previsão do Mercado de Ações Brasileiro utilizando Redes Neurais Artificiais. Centro Brasileiro de Pesquisas Físicas – CBPF. Rio de Janeiro. 21 p.
- FUEDA, K., IIZUKA, M., MORI, Y., 2009. Variable selection in multivariate methods using global score estimation. *Comput Stat*, v. 24, pp. 127–144.
- FRANÇA, J.A., *et al.*, 2009. Uma implementação do algoritmo Levenberg-Marquardt dividido para aplicações em visão computacional. *Semina: Ciências Exatas e Tecnológicas*, v. 30, n. 1, pp. 51-62.
- GOOD, R.P., KOST, D., CHERRY, G.A., 2010. Introducing a Unified PCA Algorithm for Model Size Reduction. *IEEE Transactions on Semiconductor Manufacturing*, v. 23, n. 2, pp. 201-209.
- GRIEU *et al.*, 2005. Prediction of parameters characterizing the state of a pollution removal biological process. *Engineering Applications of Artificial Intelligence*, v. 18, pp. 559–573.
- GUO *et al.* 2002. Feature selection in principal component analysis of analytical data. *Chemometrics and Intelligent Laboratory Systems* v. 61 pp. 123– 132.
- HAMED, M.M., KHALAFALLAH, M.G., HASSANIEN, E.A., 2004. Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling & Software*, v. 19, pp. 919–928.
- JOLLIFFE I.T., 1972. Discarding Variables in a Principal Component Analysis. I: Artificial Data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 2, n. 21, pp. 160-173.
- KOUTSOGIANNIS, G.S., SORAGHAN, J.J., 2002. Selection of number of principal components for de-noising signals. *Electronics Letters*, v. 38, n. 13, pp. 664-666.

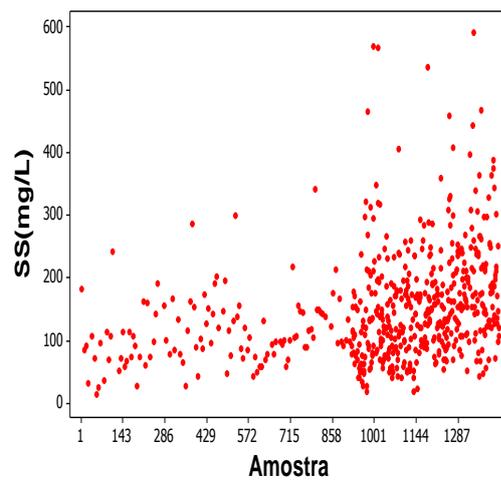
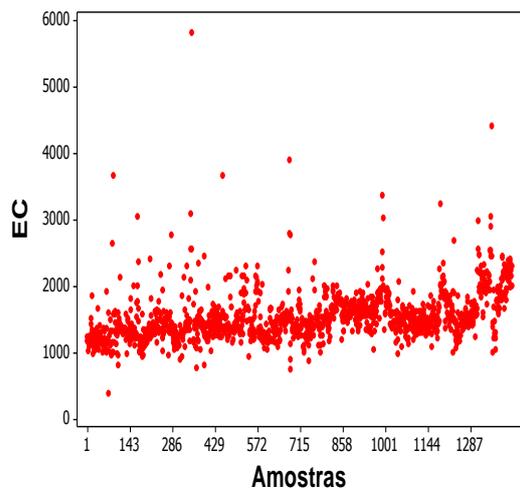
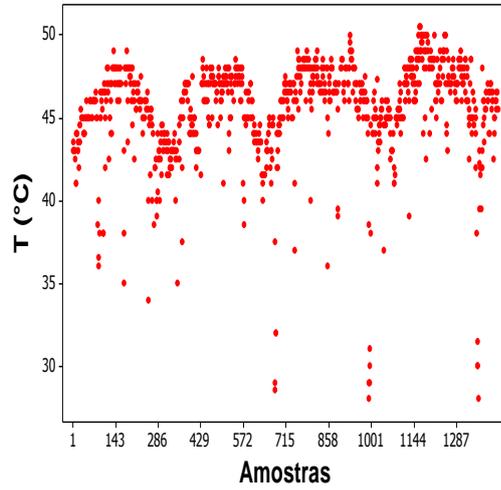
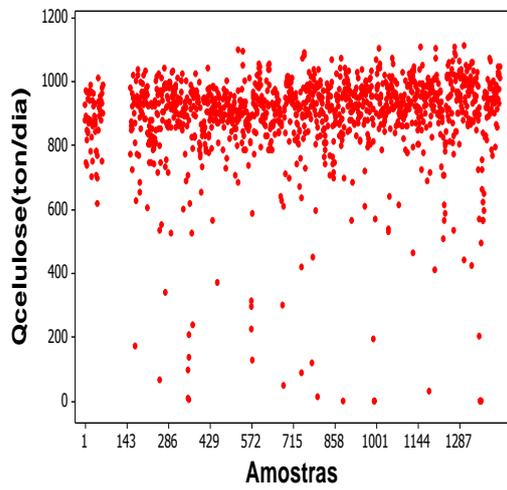
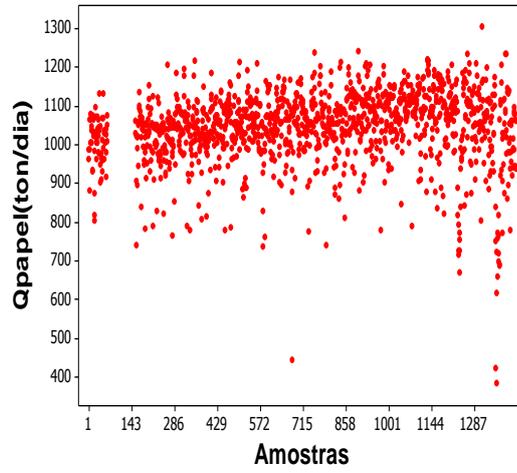
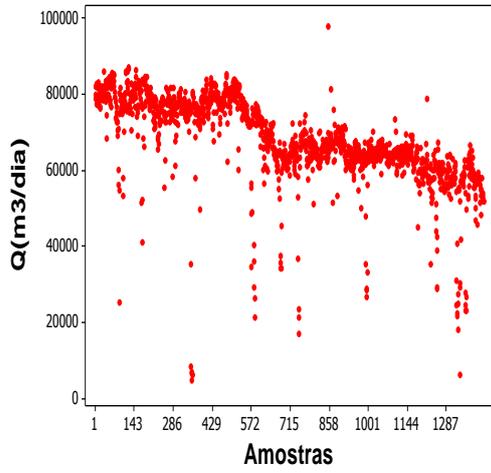
- KOVÁCS, Z.L., 2002. Redes Neurais Artificiais: Fundamentos e Aplicações. 3ª ed.: São Paulo: Livraria da Física Editora, 2002, 174 paginas.
- KRISTIANSEN, L.C., *et al.*, 2010. Using a cross-model loadings plot to identify protein spots causing 2-DE gels to become outliers in PCA. *Proteomics*, v. 10, pp. 1721–1723.
- LI, Y, KITAGAWA, H., 2007. DB-Outlier Detection by Example in High Dimensional Datasets. *IEE*, pp. 73-78.
- LUDWIG JR., O., MONTGOMERY, E., 2007. Redes Neurais: Fundamentos e Aplicações com Programas em C. Ed. 1. Editora Ciência moderna. 2007. 186 f.
- LUZ, E. R. Predição de propriedades de gasolinas usando espectroscopia FTIR e regressão por mínimos quadrados parciais. 2003. Dissertação (Mestre em Química). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2003.
- MAY, D.B., SIVAKUMAR, M., 2009. Prediction of urban stormwater quality using artificial neural networks. *Environmental Modelling & Software*, v.24, pp. 296–302.
- MARDIA, K. V., KENT, J. T. e BIBBY, J. M., 1979. Multivariate Analysis, Academic Press.
- MATTEAU, M., ASSANI, A.A., MESFIOUI, M., 2009. Application of multivariate statistical analysis methods to the dam hydrological impact studies. *Journal of Hydrology*, v. 371, pp. 120–128.
- MELLO, MARÍLIA TERRA. Aplicação de redes neurais artificiais no processo de precificação de ações. 2004. 50 f. Monografia (Bacharel em Ciência da computação). Instituto de Física e Matemática, Universidade Federal de Pelotas, Pelotas, 2004.
- MONTGOMERY, D.C., RUNGER, G.C., HUBELE, N.F. Estatística Aplicada à Engenharia. 2 ed.: Rio de Janeiro: Livros Técnicos e Científicos - LTC, 2004, pp. 203-204.
- NETO, J. M. M., MOITA, G.C., 1998. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, v. 21, n. 4, pp. 467- 469.
- OLIVEIRA-ESQUERRE, K.P., 2003. Aplicação de técnicas estatísticas multivariadas e de redes neurais na modelagem de um sistema de tratamento de efluentes industriais. Tese (Doutorado em engenharia química). Faculdade de engenharia química. Universidade Estadual de Campinas, São Paulo, 2003.
- OLIVEIRA-ESQUERRE, K.P., SEBORG, D.E., BRUNS, R.E., MORI, M. 2004a. Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill Part I. Linear approaches. *Chemical Engineering Journal*, v. 104, pp.73–81.
- OLIVEIRA-ESQUERRE, K.P., *et al.*, 2004b. Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill Part II. Nonlinear approaches. *Chemical Engineering Journal*, v. 104, pp. 73–81.
- OLIVEIRA-ESQUERRE, K.P., MORI, M., BRUNS, R.E., 2002. Simulation of an industrial wastewater treatment plant using artificial neural network and principal component analysis. *Brazilian Journal of Chemical Engineering*, v. 19, n.4, pp. 365 – 370.
- YU, L., WANG, S., LAI, K.K., 2006. An Integrated Data Preparation Scheme for Neural Network Data Analysis. *IEE transactions on knowledge and data engineering*, v. 18, pp. 217-230.
- RAN, W., QIAOL, J., YE, X. Soft-measuring Approach to On-line Predict BOD Based on PCA Time-Delay Neural Network. Proceedings of the 5<sup>th</sup> World Congress on Intelligent Control and Automation, 2004, Hangzhou, P.R. China.
- SANTOS, ANDERSON RODRIGO. Identificação de faces humanas através de PCA – LDA e redes neurais som. Dissertação(Mestrado em Engenharia elétrica). Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2005.

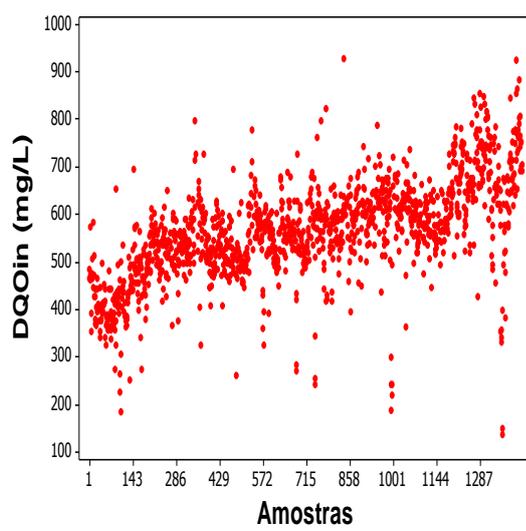
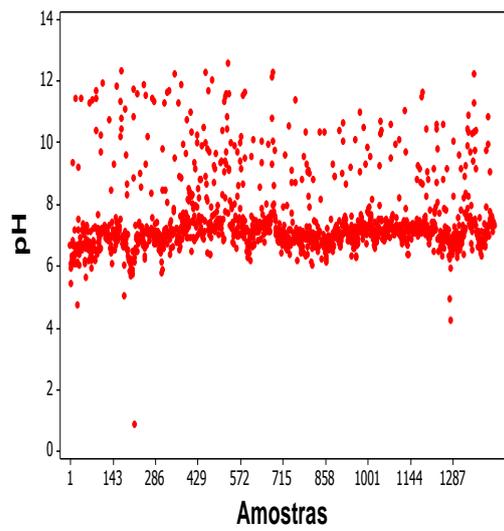
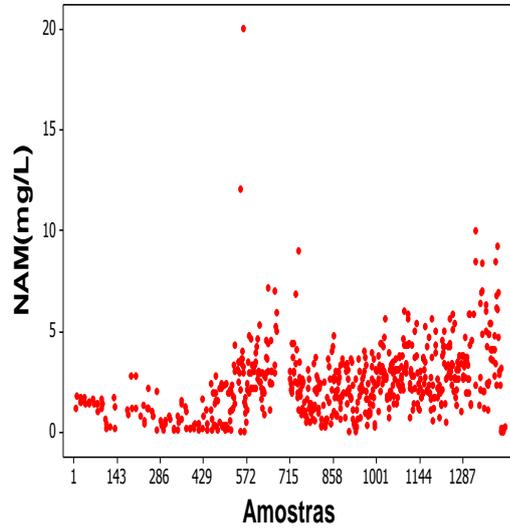
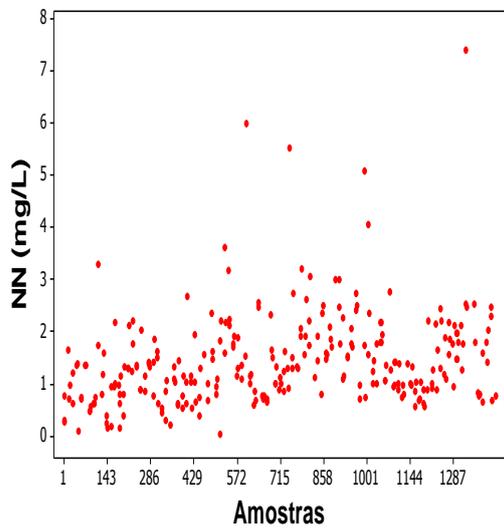
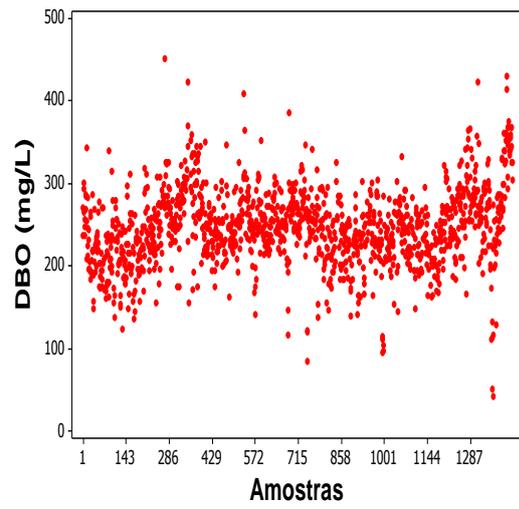
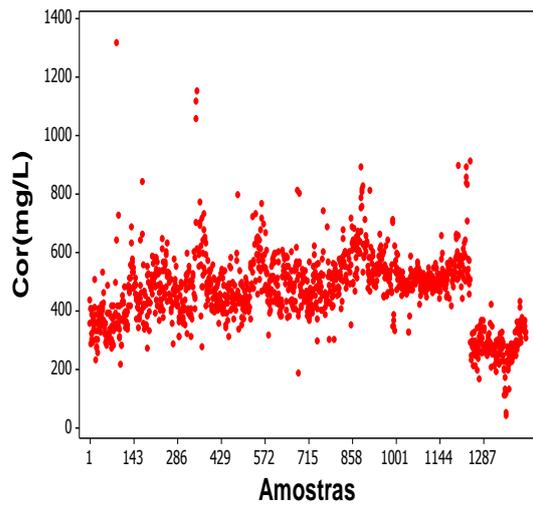
- SARASWATHI, R., SASEETHARAN, M.K., 2011. Performance prediction of a paper and pulp mill wastewater treatment plant using an artificial neural network–based model. Submetido ao periódico *Journal of Environmental Management*.
- SILVA, A.P.D., 2001. Efficient Variable Screening for Multivariate Analysis. *Journal of Multivariate Analysis*, v. 76, pp. 35-62.
- SŸKRBICÄ, B., URISŸICÄ-MLADENOVICÄ, N., CVEJANOV, J., 2005. Principal Component Analysis of Trace Elements in Serbian Wheat. *Journal of Agricultural and Food Chemistry*, v.53, pp. 2171-2175.
- STANIMIROVIC, O., *et al*, 2008. Variable Selection Methods as a Tool To Find Sensor Locations for Distributed Parameter Systems. *Ind. Eng. Chem. Res.*, v. 47, pp. 1184-1191.
- STATISTICS TOOLBOX. User's Guide. Version 2. 1999. 420 p. Disponível em: <[www.maths.lth.se/matstat/staff/krys/Program/stats\\_tb.pdf](http://www.maths.lth.se/matstat/staff/krys/Program/stats_tb.pdf)>. Acesso em: 10 de outubro de 2010.
- WU, C.L., CHAU, K.W., FAN, C., 2010. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. *Journal of Hydrology*, v. 389, pp. 146–167.
- VALLE, S., LI, W., QIN, S.J., 1999. Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods. *Ind. Eng. Chem. Res.*, v. 38, pp. 4389-4401.

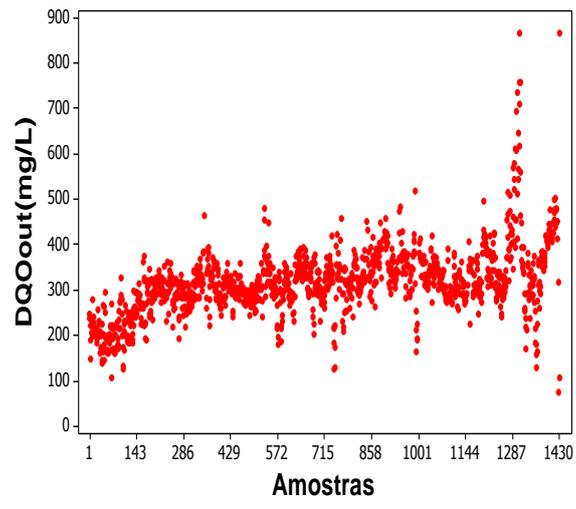
## PUBLICAÇÕES DESENVOLVIDAS DURANTE A DISSERTAÇÃO

1. MOTTA, A.R.P., MORAIS, J.T.G., OLIVEIRA-ESQUERRE, K.P., KIPERSTOK, A, 2010. O *fouling* no tratamento de água e efluentes através de processos de membranas: o estado-da-arte em estudos do fenômeno e tecnologias e estratégias para superação do problema. I Congresso Baiano de Engenharia Sanitária e Ambiental, Salvador. Universalização do saneamento, urbanização e meio ambiente: desafios tecnológicos e de gestão.
2. MORAIS, J.T.G., OLIVEIRA-ESQUERRE, K.P., KIPERSTOK, A., QUEIROZ, L.M., 2011. Predição da remoção de matéria orgânica em lagoa aerada usando rede neural artificial. 26° congresso Brasileiro de Engenharia Sanitária e Ambiental, Porto Alegre. Efluentes sanitários e ambientais.
3. MORAIS, J.T.G., OLIVEIRA-ESQUERRE, K.P., KIPERSTOK, A., QUEIROZ, L.M. DOURADO, A., 2011. Principal component analysis and artificial neural networks for organic load prediction. In.: Anais do 8<sup>th</sup> *Symposium on Systems Analysis and Integrated Assessment*, Watermatex 2011, 656-663, San Sebastian, 2011. (Em análise para publicação em periódico).

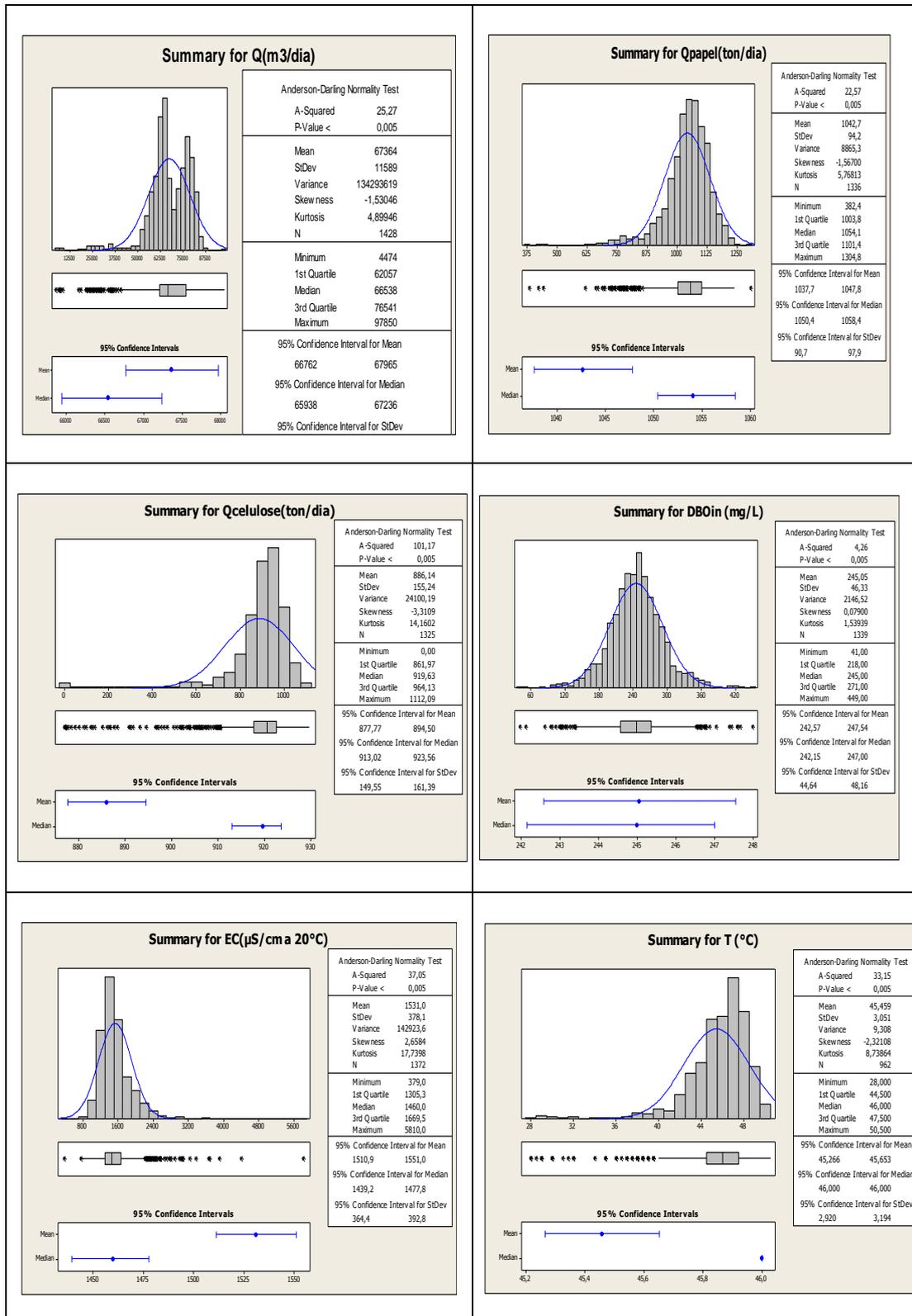
## ANEXO 1: Gráficos de séries temporais

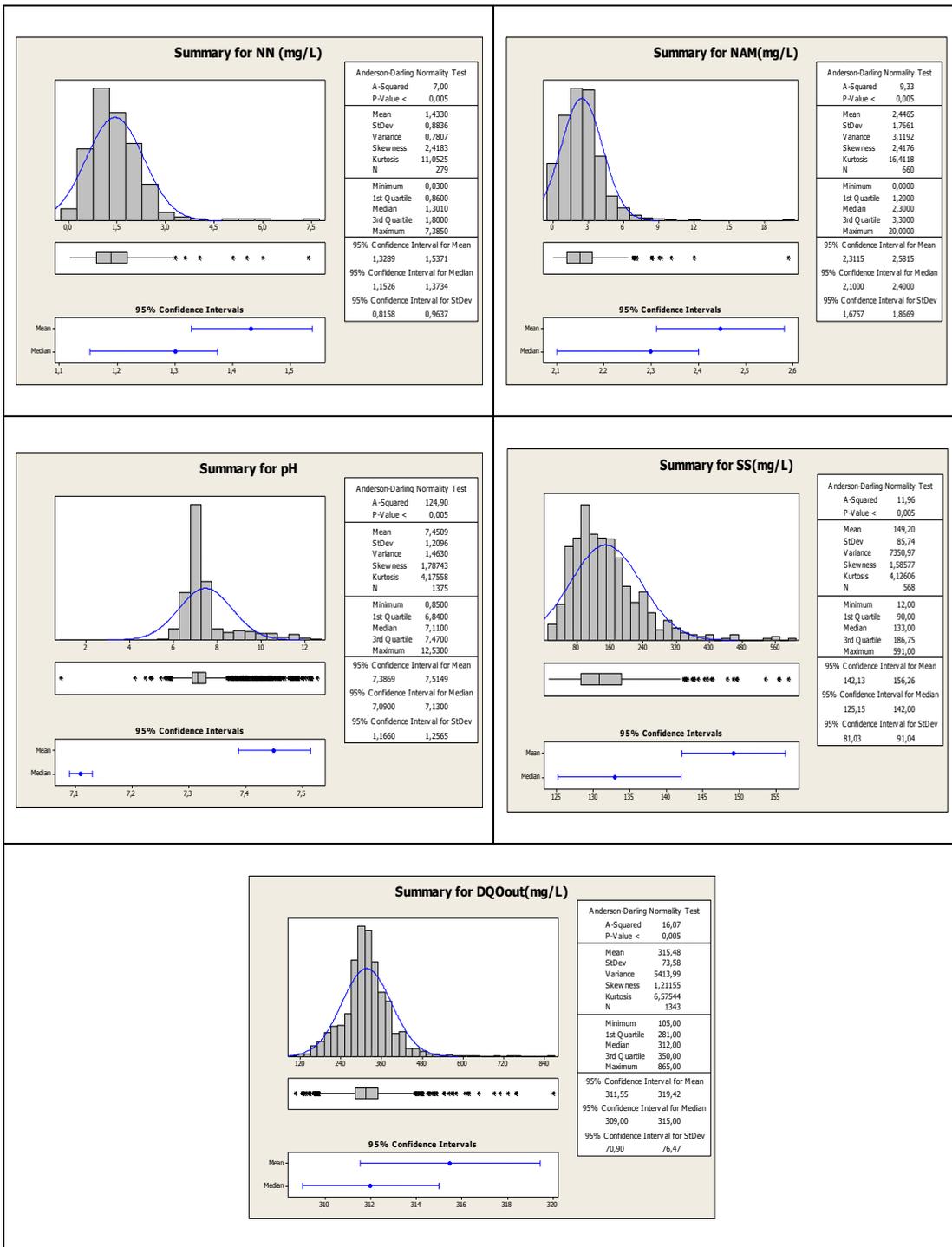






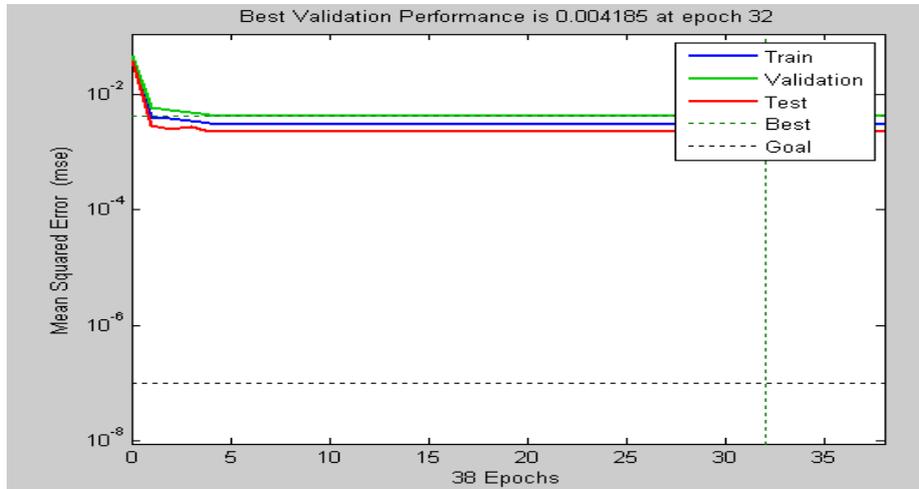
## ANEXO 2: Teste de Anderson-Darling para verificação da normalidade dos parâmetros.



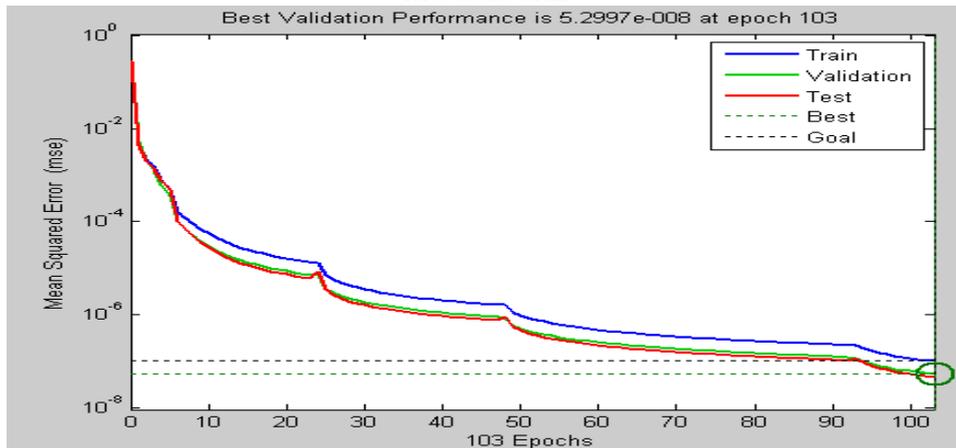


### ANEXO 3: Gráficos com as performances dos modelos de predição para o conjunto de validação.

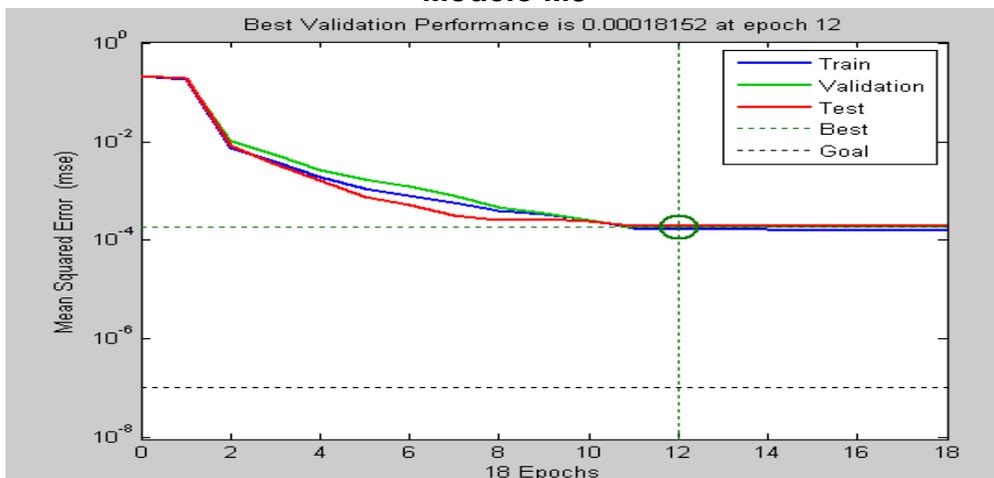
#### Modelo M1



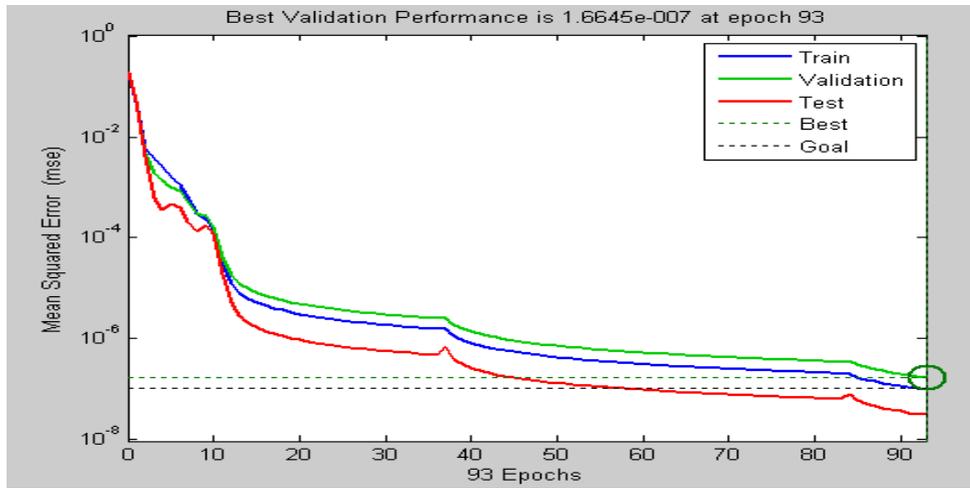
#### Modelo M2



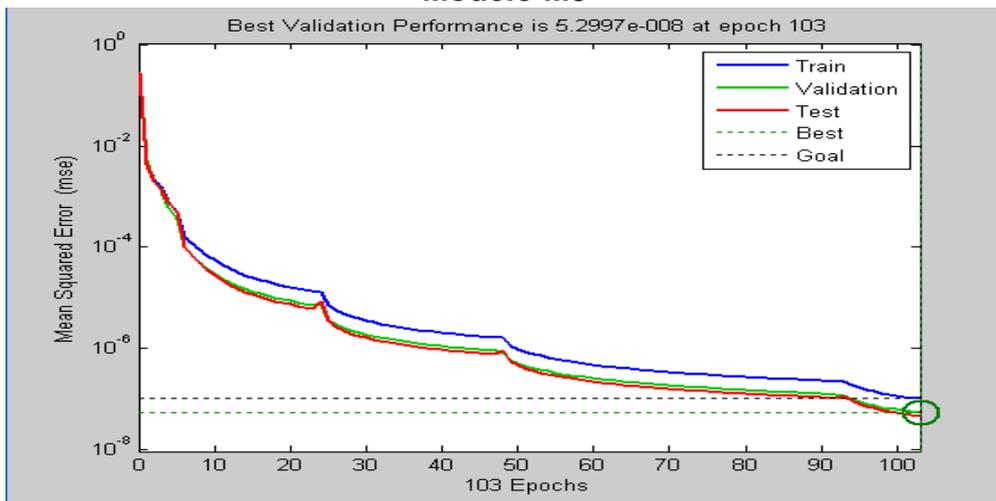
#### Modelo M3



### Modelo M4

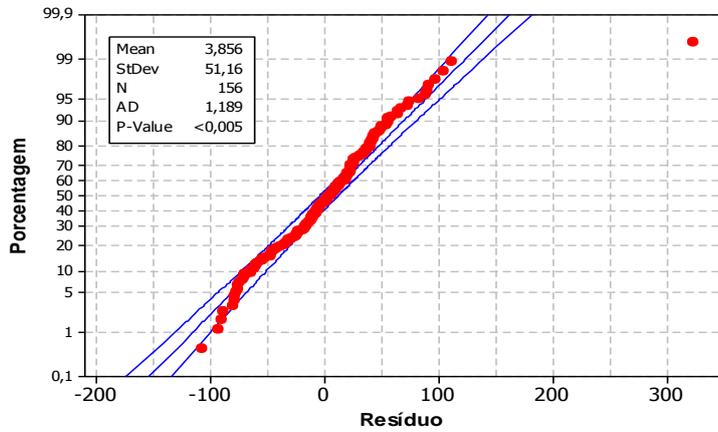


### Modelo M5

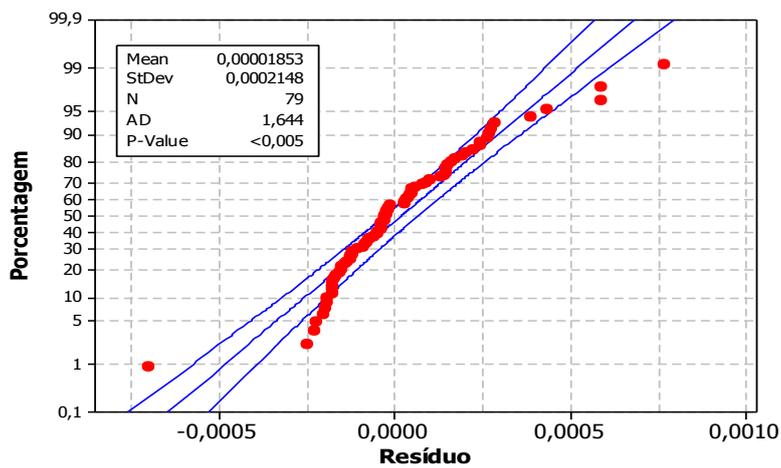


## ANEXO 4: Teste de Anderson-Darling para verificação da normalidade dos resíduos.

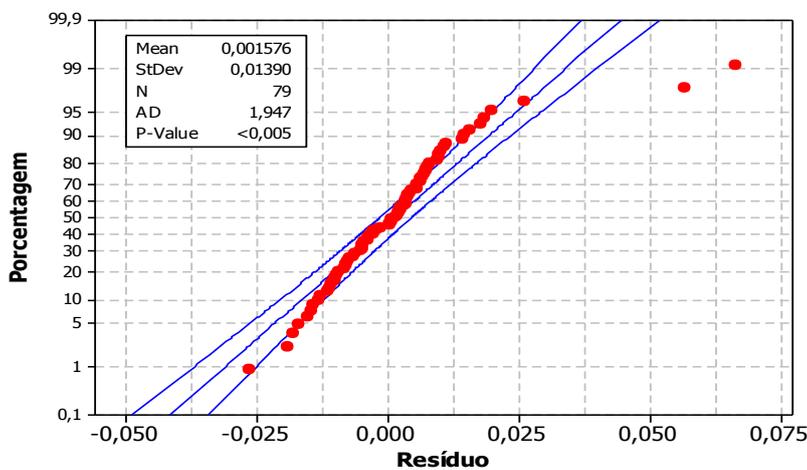
**Gráfico de probabilidade dos resíduos do modelo M1**  
Normal - 95% CI



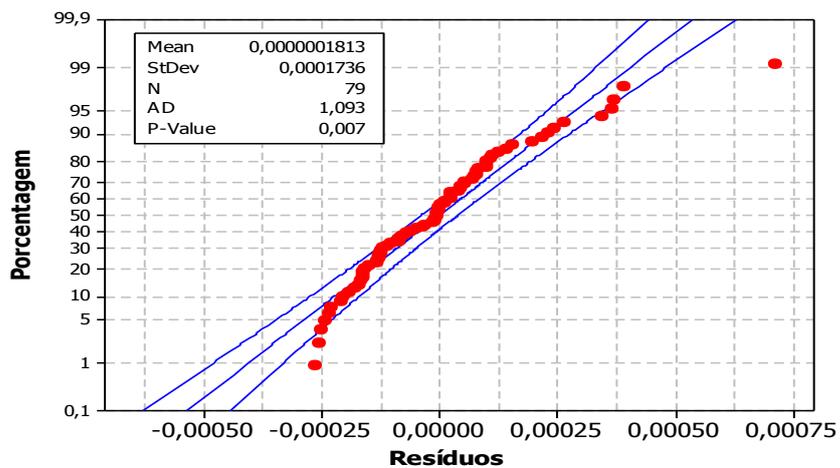
**Gráfico de probabilidade dos resíduos do modelo M2**  
Normal - 95% CI



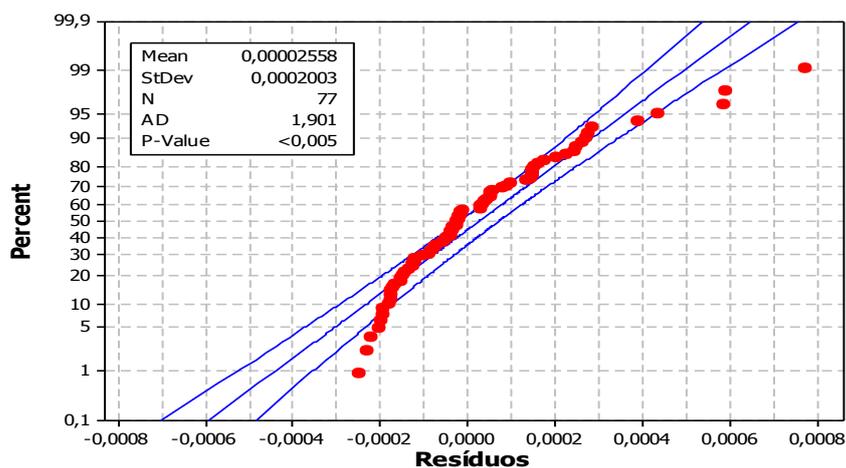
**Gráfico de probabilidade dos resíduos do modelo M3**  
Normal - 95% CI



**Gráfico de probabilidade dos resíduos do modelo M4**  
Normal - 95% CI



**Gráfico de probabilidade dos resíduos do modelo M5**  
Normal - 95% CI



**UFBA**  
**UNIVERSIDADE FEDERAL DA BAHIA**  
**ESCOLA POLITÉCNICA**

**PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA INDUSTRIAL - PEI**

Rua Aristides Novis, 02, 6º andar, Federação, Salvador BA  
CEP: 40.210-630  
Telefone: (71) 3283-9800  
E-mail: [pei@ufba.br](mailto:pei@ufba.br)  
Home page: <http://www.pei.ufba.br>

