



Pedro Miguel Sequeira Pinto

Licenciado em Engenharia do Ambiente

**Utilização de redes neuronais artificiais
como ferramenta de gestão de
processos de digestão anaeróbia.
Caso de estudo – ETAR do Seixal**

Dissertação para a obtenção do Grau de Mestre em
Engenharia do Ambiente – Perfil de Engenharia Sanitária

Orientadora: Professora Doutora Leonor Miranda
Monteiro de Amaral, Prof. Auxiliar, FCT-UNL
Co-orientadora: Engenheira Lisete Calado Epifâneo,
SIMARSUL, S.A.

Júri:

Presidente: Prof. Doutor Fernando José Pires Santana
Arguentes: Prof. Doutor António Pedro Macedo Coimbra Mano
Prof. Doutor José Manuel Ribeiro Matos da Fonseca
Vogal: Prof. Doutora Leonor Miranda Monteiro do Amaral



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Outubro, 2014



Pedro Miguel Sequeira Pinto

Licenciado em Engenharia do Ambiente

**Utilização de redes neuronais artificiais
como ferramenta de gestão de
processos de digestão anaeróbia.
Caso de estudo – ETAR do Seixal**

Dissertação para a obtenção do Grau de Mestre em
Engenharia do Ambiente – Perfil de Engenharia Sanitária

Orientadora: Professora Doutora Leonor Miranda
Monteiro de Amaral, Prof. Auxiliar, FCT-UNL
Co-orientadora: Engenheira Lisete Calado Epifâneo,
SIMARSUL, S.A.

Júri:

Presidente: Prof. Doutor Fernando José Pires Santana
Arguentes: Prof. Doutor António Pedro Macedo Coimbra Mano
Prof. Doutor José Manuel Ribeiro Matos da Fonseca
Vogal: Prof. Doutora Leonor Miranda Monteiro do Amaral

Utilização de redes neuronais artificiais como ferramenta de gestão de processos de digestão anaeróbia. Caso de estudo – ETAR do Seixal, “Copyright” Pedro Miguel Sequeira Pinto, da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

“A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor”.

Aos meus pais

AGRADECIMENTOS

A realização deste trabalho não teria sido possível sem o contributo de várias pessoas, e a elas desejo aqui expressar os meus sinceros agradecimentos:

À orientadora, a Professora Leonor Amaral, por me ter concedido a oportunidade de abordar este tema e pela confiança que depositou em mim desde o início. O apoio, o entusiasmo e a dedicação ímpares que sempre prestou constituíram aspetos fundamentais para a realização deste trabalho.

À co-orientadora, a Eng.^a Lisete Epifâneo, da SIMARSUL, pela paciência e pela disponibilidade que sempre demonstrou. As ideias, a experiência e os conselhos concedidos foram fulcrais para a definição do rumo que o trabalho tomou.

Ao Eng.^o Eduardo Ventura, pela disponibilidade e atenção prestadas, e ao restante grupo da SIMARSUL pelo ambiente e simpatia com que me acolheram.

À AdP e a todos os participantes das sessões de trabalho, pela cordialidade, pela partilha de resultados e pelo debate de ideias.

Ao Eng.^o José Pacheco, da ETAR do Seixal, pela prontidão com que disponibilizou os registos das medições.

Ao Professor José Manuel Fonseca, pelo tempo e esclarecimentos concedidos na temática das redes neuronais artificiais.

Por último, quero deixar uma palavra de agradecimento à minha família, aos meus amigos e aos meus colegas, pelo interesse, pela força e pela coragem com que me presentearam ao longo de todos estes meses.

RESUMO

As redes neuronais artificiais (RNA) têm sido apontadas como uma boa ferramenta de gestão da digestão anaeróbia, e o presente trabalho procurou explorar as funcionalidades de um aplicativo informático de criação de RNA, analisando a sua capacidade para modelar e otimizar os processos anaeróbios.

O trabalho inseriu-se num estudo mais amplo, desenvolvido pelo grupo Águas de Portugal, no qual se pretendeu obter uma visão mais alargada e robusta da aplicabilidade deste tipo de ferramentas em digestores anaeróbios de diferentes instalações. O caso de estudo foi a ETAR do Seixal, pertencente à SIMARSUL, e o software utilizado foi o NeuralTools®.

O desenvolvimento do estudo iniciou-se com a preparação dos dados referentes à ETAR do Seixal, tendo-se considerado esta como sendo a etapa determinante. A partir da caracterização das variáveis e de uma análise de correlações entre elas, foi possível seleccionar 20 variáveis a integrar nos ensaios de treino e de teste, cujos principais objetivos se prenderam com a identificação da RNA com maior capacidade para prever o biogás produzido e a seleção das variáveis mais adequadas para a modelação dos processos anaeróbios. O treino e teste de redes envolveu a realização de 266 ensaios, a partir dos quais se identificaram as cinco melhores redes para previsão.

A melhor RNA foi criada a partir dos dados de tempo de retenção hidráulico, pH, temperatura, ácidos gordos voláteis e alcalinidade total do digestor, e permitiu obter boas previsões do biogás produzido. Os resultados alcançados com esta rede ficaram, contudo, aquém dos valores de referência de uma previsão considerada “muito boa” e o reduzido número de casos usados para treinar a rede afigura-se como a principal causa. A escassez de dados constituiu, de resto, a principal limitação ao longo do estudo, permitindo realçar a importância da monitorização na gestão da digestão anaeróbia.

Palavras-chave: Digestão anaeróbia, Redes neuronais artificiais, Modelação, Biogás, Previsão.

ABSTRACT

Artificial neural networks (ANN) have been identified as a proper management tool for anaerobic digestion. Therefore, this study sought to explore the functionalities of a computer application designed for ANN creation, analyzing its ability to model and optimize anaerobic processes.

This work was part of a wider study, developed by Águas de Portugal group, which aims to obtain a broader and sturdier vision of the feasibility of such tools in anaerobic digesters of different facilities. The case study was the wastewater treatment plant of Seixal, belonging to SIMARSUL. The software used was NeuralTools®.

The development of the study began with the preparation of data on the WWTP of Seixal. This was considered the determinant step. From the characterization of the operating parameters and the analysis of correlations between them, it was possible to select 20 variables to be integrated into training and test trials, whose main objectives were held with the identification of the ANN with greater ability to predict gas production and with the selection of the most suitable variables for anaerobic processes modeling. The network training and testing involved 266 trials, from which the top five networks for prediction were identified.

The best ANN was developed from the hydraulic retention time, pH, temperature, volatile fatty acids and total digester alkalinity data, and yielded good predictions of the produced gas. The results achieved with this network were, however, below the reference values of a "very good" prediction, and the small number of cases used for neural network training appears to be the main cause for such. Moreover, the data scarcity was the main limitation throughout the study, emphasizing the importance of monitoring in anaerobic digestion processes management.

Keywords: Anaerobic digestion, Artificial neural networks, Modeling, Gas production, Prediction.

ÍNDICE DE MATÉRIAS

1. INTRODUÇÃO	1
1.1 Enquadramento.....	1
1.1.1 Aspetos gerais sobre o tratamento de águas residuais.....	1
1.1.2 Legislação	3
1.1.3 Linha de tratamento de uma ETAR.....	5
1.2 Caracterização dos processos de digestão anaeróbia	9
1.2.1 Aspetos gerais.....	9
1.2.2 Descrição das fases do processo	10
1.2.3 Fatores ambientais.....	14
1.2.4 Aspetos de dimensionamento dos digestores	18
1.2.5 Produtos	21
1.2.6 Monitorização e controlo do processo	22
1.3 Redes Neurais Artificiais	25
1.3.1 Aspetos gerais.....	25
1.3.2 Princípios de funcionamento.....	26
1.3.3 Topologias das RNA	27
1.3.4 Desenvolvimento de uma RNA	30
2. OBJETIVOS	33
3. METODOLOGIA	35
3.1 Descrição sumária do estudo.....	35
3.2 O caso de estudo	36
3.2.1 O grupo Águas de Portugal e a SIMARSUL	36
3.2.2 A ETAR do Seixal.....	39
3.3 Desenvolvimento do estudo	43
3.3.1 Software NeuralTools®	43
3.3.2 Preparação dos dados	44
3.3.3 Treino e Teste	47
3.3.4 Previsão	54
4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS OBTIDOS	57
4.1 Preparação dos dados	57
4.1.1 Descrição das variáveis	58
4.1.2 Parâmetros de funcionamento do digestor	62
4.1.3 Análise de correlações	63
4.2 Treino e Teste	66
4.2.1 Familiarização com o software NeuralTools®	66
4.2.2 Ensaios de redes sem desfasamento de variáveis.....	69

4.2.3 Ensaio de redes com desfasamento de variáveis.....	70
4.2.4 Ensaio com conjuntos de dados adicionais	73
4.2.5 Ensaio com as melhores redes	77
4.3 Previsão	81
4.4 Considerações finais	86
5. CONCLUSÕES	91
5.1 Conclusões.....	91
5.2 Sugestões e recomendações futuras.....	92
REFERÊNCIAS BIBLIOGRÁFICAS	93
APÊNDICE I.....	101
APÊNDICE II	107
APÊNDICE III	111
APÊNDICE IV.....	125
APÊNDICE V	131
APÊNDICE VI.....	137
APÊNDICE VII.....	141

ÍNDICE DE FIGURAS

Figura 1.1 – Percursos das águas, águas residuais e águas pluviais	2
Figura 1.2 – Operações e processos unitários usados no tratamento de lamas e respetivo envio para destino final	8
Figura 1.3 – Representação esquemática da digestão anaeróbia de macromoléculas complexas	11
Figura 1.4 – Taxa de remoção específica de substrato em função da temperatura	14
Figura 1.5 – Efeito da produção de ácidos voláteis no pH de um digestor	15
Figura 1.6 – Condições de pH e de concentração de bicarbonato para a digestão de lamas ...	16
Figura 1.7 – Influência da temperatura na extensão e taxa de digestão anaeróbia de lamas primárias	19
Figura 1.8 – Digestor anaeróbio de baixa carga	20
Figura 1.9 – Digestor anaeróbio de alta carga	20
Figura 1.10 – Digestão anaeróbia em dois estágios	21
Figura 1.11 – Representação esquemática de uma RNA típica	26
Figura 1.12 – Modelo básico de um neurónio	26
Figura 1.13 – Rede neuronal do tipo <i>multilayer feed-forward</i>	28
Figura 1.14 – Outras topologias aplicadas às redes neuronais artificiais	29
Figura 1.15 – Rede neuronal probabilística	30
Figura 3.1 – Estrutura organizacional do grupo AdP e território nacional abrangido pelos sistemas multimunicipais de abastecimento de água e saneamento de águas residuais	36
Figura 3.2 – Dados sobre a produção de energia renovável no grupo AdP	37
Figura 3.3 – Território da península de Setúbal abrangido pelas infraestruturas de drenagem e tratamento de efluentes da SIMARSUL	38
Figura 3.4 – Vista aérea da ETAR do Seixal	38
Figura 3.5 – Planta do tratamento da fase líquida da ETAR do Seixal	40
Figura 3.6 – Digestores anaeróbios da ETAR do Seixal	41
Figura 3.7 – Planta do tratamento da fase sólida da ETAR do Seixal	42
Figura 3.8 – Comando <i>Data Set Manager</i> do NeuralTools®	45
Figura 3.9 – Caixa de diálogo do comando <i>Data Set Manager</i> do NeuralTools®	46
Figura 3.10 – Caixa de diálogo do comando <i>Missing Data Utilities</i> do NeuralTools®	47
Figura 3.11 – Comando <i>Train</i> do NeuralTools®	47
Figura 3.12 – Caixa de diálogo do comando <i>Train</i> do NeuralTools®	48
Figura 3.13 – Separador <i>Net Configuration</i> da caixa de diálogo do comando <i>Train</i> do NeuralTools®	49
Figura 3.14 – Separador <i>Runtime</i> da caixa de diálogo do comando <i>Train</i> do NeuralTools®	50
Figura 3.15 – Janela de pré-visualização do treino no NeuralTools®	51
Figura 3.16 – Janela de visualização do progresso do treino no NeuralTools®	52
Figura 3.17 – Comando <i>Predict</i> do NeuralTools®	54

Figura 3.18 – Caixa de diálogo do comando <i>Predict</i> do NeuralTools®	54
Figura 3.19 – Janela de pré-visualização da previsão no NeuralTools®	55
Figura 4.1 – Histograma das temperaturas medidas no digestor entre 2012 e 2014 e série temporal dos valores desta variável e do caudal de biogás produzido.....	58
Figura 4.2 – Histograma dos caudais de lamas afluentes ao digestor entre 2012 e 2014 e série temporal dos valores desta variável e do caudal de biogás produzido	59
Figura 4.3 – Histograma dos tempos de retenção hidráulicos do digestor, correspondentes ao período entre 2012 e 2014	60
Figura 4.4 – Histograma das medições do caudal de biogás afluente ao digestor entre 2012 e 2014 e série temporal dos valores desta variável	61
Figura 4.5 – Indicadores de desempenho dos ensaios de treino e de teste com desfasamento, realizados a partir da rede “2-A”	71
Figura 4.6 – Histograma dos intervalos entre amostragens da variável “MV in”	72
Figura 4.7 – Comparação entre o número de casos utilizados para treinar e testar as melhores redes identificadas para cada número de variáveis de entrada.	79
Figura 4.8 – Comparação entre as percentagens de más previsões e os erros médios quadráticos obtidos no treino e teste das melhores redes identificadas para cada número de variáveis de entrada.	79
Figura 4.9 – Esquematização das fases de desenvolvimento do estudo	80
Figura 4.10 – Previsões dos caudais de biogás produzido efetuadas pelas cinco melhores redes neuronais e comparação com os valores reais e com a média das previsões.....	83
Figura 4.11 – Comparação entre os caudais de biogás previstos através da rede “1-A” e os valores reais.....	85
Figura I.1 – Exemplo de um relatório detalhado do treino e teste criado pelo NeuralTools® ...	103
Figura I.2 – Gráficos inseridos no relatório detalhado do treino e do teste do NeuralTools® ...	105
Figura II.1 – Localização dos pontos de amostragem de alguns parâmetros da digestão anaeróbia	109
Figura II.2 – Localização do ponto de amostragem da variável “Q biogás”.....	110
Figura III.1 – Histograma dos valores de pH medidos à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014, e séries temporais dos valores destas variáveis	113
Figura III.2 – Histograma das concentrações de CQO medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014, e séries temporais dos valores destas variáveis.....	114
Figura III.3 – Histograma das cargas de CQO afluentes e efluentes ao digestor, correspondentes ao período entre 2012 e 2014, e séries temporais dos valores destas variáveis	115
Figura III.4 – Histograma das cargas de CQO removida no digestor, correspondentes ao período entre 2012 e 2014	116

Figura III.5 – Histograma das concentrações de matéria seca medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014, e séries temporais dos valores destas variáveis.....	117
Figura III.6 – Histograma das cargas de matéria seca afluentes e efluentes ao digestor, correspondentes ao período entre 2012 e 2014	119
Figura III.7 – Histograma das concentrações de matéria volátil medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014, e séries temporais dos valores destas variáveis.....	119
Figura III.8 – Histograma das cargas de matéria volátil afluentes e efluentes ao digestor, correspondentes ao período entre 2012 e 2014	120
Figura III.9 – Histograma das cargas de LMV removida no digestor, correspondentes ao período entre 2012 e 2014	121
Figura III.10 – Histograma dos valores da carga volúmica orgânica, correspondentes ao período entre 2012 e 2014	121
Figura III.11 – Histograma das concentrações de alcalinidade total no digestor, correspondentes ao período entre 2012 e 2014	121
Figura III.12 – Histograma das concentrações de ácidos gordos voláteis no digestor, correspondentes ao período entre 2012 e 2014	122
Figura III.13 – Histograma dos quocientes entre a alcalinidade total e os ácidos gordos voláteis no digestor, correspondentes ao período entre 2012 e 2014, e séries temporais das variáveis “AGV reator” e “TAL reator”	123

ÍNDICE DE QUADROS

Quadro 1.1 – Requisitos para as descargas das ETAR, constantes do Anexo I do Decreto-Lei n.º 152/97, de 19 de junho	4
Quadro 1.2 – Componentes presentes nas águas residuais	6
Quadro 1.3 – Identificação e quantificação das populações bacterianas presentes nos digestores anaeróbios de lamas	13
Quadro 1.4 – Concentrações de toxicidade e de inibição de materiais inorgânicos na digestão anaeróbia.....	17
Quadro 1.5 – Conteúdo de metano no biogás e respetivos fatores de conversão em energia .	22
Quadro 1.6 – Alguns dos parâmetros operacionais a monitorizar nos processos anaeróbios e correspondentes variações normais admissíveis	23
Quadro 4.1 – Variáveis selecionadas a partir do histórico de monitorização da qualidade e dos relatórios diários de supervisão da ETAR do Seixal	57
Quadro 4.2 – Médias aritméticas dos valores dos parâmetros de funcionamento da digestão anaeróbia na ETAR do Seixal.....	62
Quadro 4.3 – Valores dos coeficientes de correlação de Pearson entre variáveis da digestão anaeróbia	65
Quadro 4.4 – Indicadores de desempenho dos ensaios de treino e de teste das redes “0-1” e “0-3”	67
Quadro 4.5 – Indicadores de desempenho do ensaio de treino e de teste da rede “0-2”	68
Quadro 4.6 – Melhores valores obtidos para cada um dos quatro indicadores de desempenho utilizados e respetivas redes.....	70
Quadro 4.7 – Indicadores de desempenho dos ensaios de treino e de teste com desfasamento, realizados a partir da rede “2-A”.....	71
Quadro 4.8 – Indicadores de desempenho do ensaio realizado com a rede “4-A-3”, da segunda etapa, e com outras quatro, com as mesmas variáveis desfasadas de forma diferente em relação ao caudal de biogás.....	73
Quadro 4.9 – Indicadores de desempenho da rede “5-F-3” e da rede “5-F-3-5”, com as mesmas variáveis à exceção do “Q lamas”	74
Quadro 4.10 – Indicadores de desempenho dos ensaios realizados com as redes “9-G-1” e “9-G-2”, criadas a partir das variáveis com maior correlação com o caudal de biogás produzido.....	74
Quadro 4.11 – Indicadores de desempenho do ensaio realizado com a rede “9-G-3”, criada a partir de variáveis medidas à entrada do digestor.....	75
Quadro 4.12 – Indicadores de desempenho do ensaio realizado com a rede “9-G-4”, criada a partir de variáveis medidas no interior e à saída do digestor.....	75
Quadro 4.13 – Indicadores de desempenho da rede “BNTrTsRNA3.16”, treinada com dados da ETAR Sul da SIMRIA, e da rede “5-H-5-4-A”, criada a partir das mesmas variáveis medidas na ETAR do Seixal	76

Quadro 4.14 – Indicadores de desempenho das melhores redes neuronais identificadas para cada número de variáveis de entrada.	78
Quadro 4.15 – Indicadores estatísticos referentes às previsões dos valores do caudal de biogás produzido, efetuadas pelas cinco melhores redes neuronais	81
Quadro IV.1 – Indicadores de desempenho do treino/teste das melhores redes neuronais criadas sem desfasamento de variáveis	127
Quadro V.1 – Indicadores de desempenho do treino/teste de redes neuronais criadas com desfasamento da variável “Q biogás” em relação às restantes.....	133
Quadro VI.1 – Indicadores de desempenho das melhores redes neuronais criadas a partir de conjuntos de dados adicionais	139
Quadro VII.1 – Indicadores de desempenho das 13 melhores redes neuronais treinadas/testadas	143

LISTA DE ABREVIATURAS, ACRÓNIMOS, SIGLAS E SÍMBOLOS

Abreviaturas

e. p. – Equivalente de população

Acrónimos

AdP – Águas de Portugal, SGPS, S.A.

CCE – Centro para a Conservação da Energia

CVO – Central de valorização orgânica

ETAR – Estação de Tratamento de Águas Residuais

DQA – Diretiva Quadro da Água

INAG – Instituto Nacional da Água, I.P.

IWA – International Water Association

PEAASAR – Plano Estratégico de Abastecimento de Água e Saneamento de Águas Residuais

PENSAAR – Plano Estratégico Nacional para o Setor do Abastecimento de Águas e Saneamento de Águas Residuais

RNA – Redes Neurais Artificiais

SANEST – Saneamento da Costa do Estoril, S.A.

SIMARSUL – Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal, S.A.

SIMRIA – Sistema Multimunicipal de Saneamento da Ria de Aveiro, S.A.

SIMTEJO – Sistema Multimunicipal de Saneamento do Tejo e Trancão, S.A.

Siglas

M.O. – Matéria orgânica

Simbologia

CH₃COOH – Ácido acético

CH₄ – Metano

CO₂ – Dióxido de carbono

CaCO₃ – Carbonato de cálcio

H₂O – Água

H₂S – Sulfídrico

HCO₃⁻ – Ião bicarbonato

NH₃ – Amoníaco

NaCl – Cloreto de sódio

°C – Grau Celsius

g – Grama

h – Hora

J – Joule

L – Litro

M – Molar

m – Metro

Nm³ – Metro cúbico normal (medido em condições standard de pressão e temperatura)

V – Volt

W – Watt

AGV – Ácidos gordos voláteis

AGV reator – Concentração de ácidos gordos voláteis nas lamas digeridas (mg CH₃COOH/L)

AGV/TAL – Quociente entre a concentração de AGV no digestor e a alcalinidade total

CBO5 – Carência bioquímica de oxigénio ao fim de cinco dias (mg/L)

COT – Carbono orgânico total (mg/L)

CQO – Carência química de oxigénio (mg/L)

CQO in – CQO medida nas lamas afluentes ao digestor anaeróbio (mg/L)

CQO reator - CQO medida nas lamas digeridas (mg/L)

CTO – Carência total de oxigénio (mg/L)

CV – Carga orgânica volúmica (kg/(m³.dia))

LCQO in – Carga de CQO afluyente ao digestor anaeróbio (kg/dia)

LCQO out - Carga de CQO nas lamas digeridas (kg/dia)

LCQO removida – Carga de CQO removida na digestão anaeróbia (kg/dia)

LMS in - Carga de matéria seca afluyente à digestão anaeróbia (kg/dia)

LMS out - Carga de matéria seca nas lamas digeridas (kg/dia)

LMV in - Carga de matéria volátil afluyente à digestão anaeróbia (kg/dia)

LMV out - Carga de matéria volátil nas lamas digeridas (kg/dia)

LMV removida – Carga de matéria volátil removida na digestão anaeróbia (kg/dia)

MS in – Concentração de sólidos totais nas lamas afluentes à digestão anaeróbia (mg/L)

MS reator - Concentração de sólidos totais nas lamas digeridas (mg/L)

MV in - Concentração de sólidos voláteis nas lamas afluentes à digestão anaeróbia (mg/L)

MV reator - Concentração de sólidos voláteis nas lamas digeridas (mg/L)

pH – potencial de hidrogénio

pH in – pH das lamas afluentes ao digestor anaeróbio

pH reator – pH das lamas digeridas

Q biogás – Caudal de biogás produzido no digestor anaeróbio (m^3/dia)

Q lamas - Caudal de lamas elevadas para o digestor anaeróbio (m^3/dia)

SST – Sólidos suspensos totais (mg/L)

SSV – Sólidos suspensos voláteis (mg/L)

T reator – Temperatura no digestor anaeróbio ($^{\circ}\text{C}$)

TAL reator - Alcalinidade total nas lamas digeridas ($\text{mg CaCO}_3/\text{L}$)

T_{RH} – Tempo de retenção hidráulico no digestor anaeróbio (dias)

T_{RS} – Tempo de retenção de sólidos no digestor anaeróbio (dias)

V_{digestor} – Volume útil do digestor anaeróbio (m^3)

UV – Ultravioleta

ADM1 – Anaerobic Digestion Model No. 1

GRNN - Generalized Regression Neural Network

MLFN - Multilayer Feed-forward Network

MLP – Multilayer Perceptron

PNN – Probabilistic Neural Network

PLS – Partial Least Squares

RMSE – Root Mean Square Error

R^2 – Coeficiente de determinação

f - Função de transferência ou função de ativação

n – Número de conjuntos de inputs x_j

O – Output do neurónio

R - Sinal produzido pelo neurónio

x_j – Inputs de um neurónio

x_{ij} - Inputs de um nó

w_j - Peso sináptico do input

w_0 – Termo de viés

Θ – Valor limite da soma do resultado da função de input com o termo de viés

ρ – Coeficiente de correlação de Pearson

1. INTRODUÇÃO

1.1 Enquadramento

Das principais questões ambientais atuais, as relacionadas com a água são talvez as mais relevantes, e às quais maior importância é dada por governos e entidades dedicadas à proteção do património ecológico. A água constitui um elemento fundamental à vida e é também essencial para inúmeras atividades económicas, como a agricultura, pescas e transportes. Trata-se, contudo, de um recurso limitado e bastante sensível, o que conduz frequentemente a situações de escassez e de degradação dos parâmetros de qualidade, que põem em causa a proteção da saúde pública e dos ecossistemas (Quevauviller *et al.*, 2006; Metcalf & Eddy, 2003).

O tratamento de águas residuais constitui uma das etapas essenciais para a prevenção deste tipo de problemas, mas verificam-se frequentemente situações em que o tratamento é incompleto, desajustado ou até mesmo inexistente. É neste âmbito que se insere o papel da Engenharia Sanitária, cujas principais competências se prendem também com o planeamento e gestão de sistemas de recolha e tratamento de águas residuais de forma ambientalmente eficiente e economicamente viável (Metcalf & Eddy, 2003).

1.1.1 Aspetos gerais do tratamento de águas residuais

As águas residuais são essencialmente as águas de abastecimento de uma comunidade após terem sido usadas numa variedade de aplicações. Podem ser definidas como sendo a combinação dos resíduos líquidos e sólidos transportados nas águas removidas de residências, instituições e estabelecimentos comerciais e industriais, juntamente com águas superficiais e pluviais captadas nos coletores (Metcalf & Eddy, 2003).

Quando as águas residuais não tratadas se acumulam e se tornam séticas, a matéria orgânica nelas contida decompõe-se e ocorre a produção de gases fétidos e o desenvolvimento de microrganismos patogénicos comuns no trato intestinal humano. Alguns dos componentes são tóxicos e potencialmente mutagénicos ou cancerígenos, e a presença de determinados nutrientes tende a estimular o crescimento de algas e plantas aquáticas responsáveis por fenómenos de eutrofização do meio recetor. Como tal, a remoção imediata das águas residuais das suas fontes de produção, seguida do devido tratamento ou reutilização, é necessária para a protecção da saúde pública e ambiental (Figura 1.1) (Metcalf & Eddy, 2003).

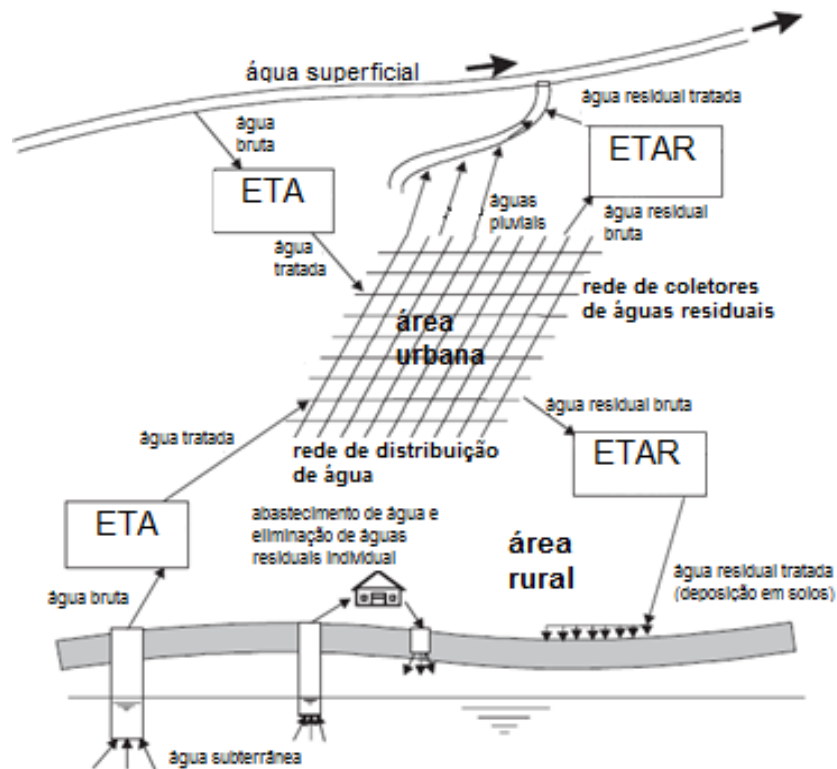


Figura 1.1 – Percursos das águas, águas residuais e águas pluviais (adaptado de Von Sperling, 2007).

Ao longo dos últimos 200 anos, tem-se verificado uma evolução dos sistemas destinados à recolha e tratamento de águas residuais. Inicialmente conduzidos pela necessidade de reduzir os focos de doença humana (Vesilind, 2003) e de assegurar condições de salubridade nos centros urbanos (Amado, 2009), estes sistemas eram apenas constituídos por redes de esgotos que drenavam as águas residuais e as descarregavam num meio recetor aquático, sem qualquer tipo de tratamento (Vesilind, 2003).

No século XX são desenvolvidos os primeiros sistemas de tratamento de águas residuais municipais, geradas essencialmente em fontes domésticas (Metcalf & Eddy, 2003). Os objetivos de tratamento prendiam-se principalmente com a remoção de nutrientes e de material coloidal, suspenso e fluviável, com o tratamento de matéria orgânica biodegradável e com a eliminação de organismos patogénicos (Metcalf & Eddy, 2003; Von Sperling, 2007). A criação deste tipo de sistemas ganhou expressão após a Segunda Guerra Mundial, como resultado do significativo desenvolvimento industrial que aumentou substancialmente a quantidade de efluentes descarregados nos sistemas coletores municipais (Metcalf & Eddy, 2003).

No entanto, os sistemas de recolha e tratamento de águas residuais eram ainda pouco abrangentes em meados do século XX, e os efeitos da poluição provocada pela descarga descontrolada no meio recetor eram cada vez mais gravosos para os ecossistemas aquáticos e para os recursos hídricos, no geral (Vesilind, 2003).

É só na década de 70 que a consciencialização da população para os problemas ambientais resulta na tomada de medidas conducentes à limitação da poluição gerada pelas águas residuais (Metcalf & Eddy, 2003). Começa a surgir legislação sobre os limites de emissão de determinados poluentes e sobre os parâmetros físicos, químicos e microbiológicos das descargas efetuadas em meio recetor. Os sistemas de recolha e tratamento de águas residuais passam a ter de cumprir um conjunto de normas impostas pela legislação em vigor, visando a necessidade de eliminar os efeitos da poluição e de atingir níveis de qualidade da água do meio recetor. Com a recuperação deste meio, possibilita-se o regresso dos organismos marinhos nativos aos padrões de crescimento normal e fomenta-se um uso humano recreativo da água sem consequências gravosas para a saúde pública (Vesilind, 2003).

A partir de 1990, os efeitos de uma globalização emergente começam a ser sentidos por todo o mundo, assistindo-se a um entrelaçamento de problemas, atores e estados (Wijen *et al.*, 2005). A necessidade de uma resposta integrada aos problemas ambientais dos recursos hídricos e às exigências inerentes ao desenvolvimento sustentável levou a que organismos como a União Europeia promovessem a definição de soluções de gestão das questões ambientais a uma escala multiterritorial (Quevauviller *et al.*, 2006).

1.1.2 Legislação

Em 1991, entra em vigor a Diretiva n.º 91/271/CEE, do Conselho, de 21 de maio, na qual são previstos os aspetos relativos à conceção dos sistemas de drenagem e de tratamento, bem como o regime de licenciamento das descargas de águas residuais urbanas e industriais nos Estados Membros da Comissão Europeia. A transposição das normas de descarga de águas residuais no meio aquático para o direito interno consta do Decreto-Lei n.º 152/97, de 19 de junho, mantendo-se atualmente como o documento legal de referência para os sistemas responsáveis pelo tratamento e descarga de efluentes em Portugal. No Quadro 1.1 são apresentados os requisitos para as descargas das ETAR, constantes deste documento.

Quadro 1.1 – Requisitos para as descargas das ETAR, constantes do Anexo I do Decreto-Lei n.º 152/97, de 19 de junho

Parâmetros	Concentração	Percentagem mínima de redução ⁽¹⁾
Carência bioquímica de oxigénio (CBO5 a 20 °C) sem nitrificação ⁽²⁾ .	25 mg/L O ₂	70-90
Carência química de oxigénio (CQO).	125 mg/L O ₂	75
Total de partículas sólidas em suspensão ⁽³⁾ .	35 mg/L ⁽³⁾	90 ⁽³⁾
	35 nos casos previstos no n.º 3 do artigo 5.º (e. p. > 10.000).	90 nos casos previstos no n.º 3 do artigo 5.º (e. p. > 10.000).
	60 nos casos previstos no n.º 3 do artigo 5.º (e. p. de 2.000 a 10.000).	70 nos casos previstos no n.º 3 do artigo 5.º (e. p. de 2.000 a 10.000).

⁽¹⁾ Redução em relação à carga do afluente.

⁽²⁾ O parâmetro pode ser substituído por outro: carbono orgânico total (COT) ou carência total de oxigénio (CTO), se for possível estabelecer uma relação entre a CBO5 e o parâmetro de substituição.

⁽³⁾ Este requisito é facultativo.

A Diretiva n.º 91/271/CEE, do Conselho, de 21 de maio, serviu também de base para a elaboração da Diretiva Quadro da Água, um documento legal criado pela União Europeia em 2000 para o estabelecimento de um quadro de ação comunitária no domínio da política da água. A DQA assenta numa abordagem abrangente e integrada de proteção e gestão dos recursos hídricos e em Portugal deu corpo à Lei da Água (Lei n.º 58/2005, de 29 de Dezembro) (INAG, 2010a). Para além de transpor para o direito nacional o normativo da DQA, a Lei da Água veio integrar, sistematizar, atualizar e harmonizar a legislação vigente em Portugal relativa à gestão dos recursos hídricos e à administração pública da água (INAG, 2010b).

No seguimento da aprovação da Lei da Água, estabeleceu-se como prioridade a realização de uma análise à situação deste setor e a definição do respetivo enquadramento estratégico e programático. É neste âmbito que foi publicado o PEAASAR II (2007-2013), um documento orientador dos objetivos e políticas do governo na procura de soluções nesta área, com vista a um desenvolvimento sustentável.

Um dos principais objetivos operacionais do PEAASAR II consistiu em servir cerca de 90% da população total do País com sistemas públicos de saneamento de águas residuais urbanas. O seu período de aplicação acabou em 2013 e os dados relativos ao saneamento de águas residuais desse ano apontam para 78% da população servida, um valor que ficou aquém do que tinha sido estabelecido como objetivo.

Atualmente, encontra-se já em elaboração o PENSAAR 2020, uma nova estratégia para o setor do abastecimento de águas e saneamento de águas residuais para o período 2014-2020. O PENSAAR 2020 terá como principais objetivos estratégicos (Vieira, 2014):

- A proteção do ambiente e melhoria da qualidade das massas de água;
- A melhoria da qualidade dos serviços prestados;
- A otimização e gestão eficiente dos recursos;
- A sustentabilidade económica e financeira.

As orientações estabelecidas para o cumprimento destes objetivos, no âmbito do saneamento de águas residuais, incluirão, à semelhança do PEAASAR II (2007-2013):

- O uso eficiente da água;
- A gestão das lamas;
- A eco-eficiência energética.

Os objetivos estratégicos estabelecidos para o PENSAAR 2020 permitirão aliar o cumprimento das disposições previstas no Decreto-Lei n.º 152/97, de 19 de junho, relativas ao tratamento e descarga de águas residuais no meio aquático, à integração das soluções técnicas mais sustentáveis relativas à água, às lamas e à energia produzida nas ETAR.

1.1.3 Linha de tratamento de uma ETAR

No planeamento de uma estação de tratamento de águas residuais, é importante ter em consideração todos os aspetos que influenciam o esquema de tratamento a escolher e o dimensionamento dos respetivos órgãos. O conhecimento do volume e da composição do caudal afluente constitui um dos aspetos indispensáveis para o correto funcionamento da ETAR, devendo, para tal, ser estimado o número de habitantes equivalentes no ano zero e no ano horizonte do projeto (Henze *et al.*, 2002).

De acordo com o Decreto-Lei n.º 152/97, de 19 de junho, considera-se que, em Portugal, a cada habitante equivalente (ou equivalente de população) corresponde a produção diária de 60 g de CBO5. Contudo, a restante composição das águas residuais apresenta uma grande variabilidade, sendo, por vezes, necessária a introdução na ETAR de tratamentos específicos para minimizar ou neutralizar os impactes negativos de um determinado componente para o meio recetor (Henze *et al.*, 2002). No Quadro 1.2 são disponibilizados os principais componentes presentes nas águas residuais e os respetivos efeitos no ambiente.

Quadro 1.2 – Componentes presentes nas águas residuais
(adaptado de Henze *et al.*, 2002).

Componente	De relevância	Efeito no ambiente
Microrganismos	Bactérias patogénicas, vírus e ovos de larvas	Risco de contacto no banho ou na ingestão de marisco
Materiais orgânicos biodegradáveis	Depleção de oxigénio em rios, lagos e fiordes	Redução da diversidade da vida aquática
Outros materiais orgânicos	Detergentes, pesticidas, gorduras, óleos, coloração, solventes, fenóis, cianeto	Efeitos tóxicos, inconvenientes estéticos, bioacumulação
Nutrientes	Azoto, fósforo, amoníaco	Eutrofização, depleção do oxigénio, efeitos tóxicos
Metais	Hg, Pb, Cd, Cr, Cu, Ni	Efeitos tóxicos, bioacumulação
Outros materiais inorgânicos	Ácidos, bases	Corrosão, efeitos tóxicos
Efeitos térmicos	Água quente	Alteração das condições de vida da fauna e da flora
Odor e sabor	Ácido sulfídrico	Inconvenientes estéticos, efeitos tóxicos
Radioatividade	Elementos radioativos	Efeitos tóxicos, acumulação

O tratamento ou a remoção destes componentes pode ser feito através de operações unitárias ou de processos unitários. Nas operações unitárias existe a predominância da aplicação de forças físicas, sendo a sedimentação, a flotação e a filtração alguns dos métodos correspondentes a este tipo de tratamento. Nos processos unitários o tratamento ocorre por meio de reações químicas ou biológicas (Qasim, 1999).

Se o processo for químico, dá-se a remoção ou conversão dos contaminantes através da adição de determinadas substâncias ou da indução de outras reações químicas, sendo a precipitação e a adsorção exemplos deste tipo de métodos. Nos processos biológicos é usada a atividade de microrganismos para remoção de nutrientes como o fósforo e o azoto, e para a conversão da fração coloidal ou dissolvida da matéria orgânica biodegradável em gases e sólidos inertes, sendo a digestão anaeróbia um dos métodos correspondentes a este tipo de tratamento (Metcalf & Eddy, 2003).

O esquema de tratamento aplicado numa ETAR resulta da combinação de várias operações e processos unitários, e para uma melhor compreensão do seu funcionamento, os métodos utilizados e os níveis de tratamento aplicados são usualmente esquematizados em duas linhas de tratamento principais – tratamento da fase líquida e tratamento da fase sólida (Metcalf & Eddy, 2003).

Tratamento da fase líquida

A componente líquida constitui 99,9% das águas residuais, e o seu tratamento é, geralmente, dividido em quatro níveis: preliminar, primário, secundário e terciário (Spellman, 2003).

O tratamento preliminar ou pré-tratamento inicia a linha de tratamento de uma ETAR, e compreende a remoção dos constituintes das águas residuais que possam alterar as operações e processos, ou danificar os órgãos e equipamentos da estação. Alguns dos métodos utilizados nesta fase incluem a gradagem, para remoção de detritos e objetos de grandes dimensões, a classificação de areias, para remoção do material grosseiro suspenso, e a flotação, para remoção de óleos e gorduras (Metcalf & Eddy, 2003).

Em seguida, as águas residuais são encaminhadas para o tratamento primário, o qual é definido no artigo 2.º do Decreto-Lei n.º 152/97, de 19 de junho, como sendo um tratamento em que a CBO5 é reduzida em 20% e o total das partículas sólidas em suspensão é reduzido em, pelo menos, 50%. Existem vários processos e operações que podem ser utilizados no tratamento primário, sendo o mais comum a decantação. As dimensões, o formato e o número de decantadores necessários são determinados pelas características do caudal afluyente, pelas limitações de espaço na ETAR e pela eficiência de remoção de sólidos pretendida (Metcalf & Eddy, 2003).

No seguimento do tratamento primário, surge o tratamento secundário, constituído por um processo químico ou biológico seguido de uma decantação. Os principais objetivos deste tipo de tratamento são a remoção da maior parte da matéria orgânica biodegradável remanescente e ainda a remoção de nutrientes, como o azoto e o fósforo (Metcalf & Eddy, 2003), para que a descarga das águas residuais tratadas no meio recetor respeite os valores constantes dos quadros n.º 1 e 2 do Decreto-Lei n.º 152/97, de 19 de junho. As lamas ativadas, as lagoas de estabilização e os reatores de filme fixo são exemplos de alguns dos métodos utilizados neste tipo de tratamento (Spellman, 2003).

Após o tratamento secundário, as águas residuais apresentam ainda, por vezes, valores de alguns parâmetros de qualidade diferentes dos pretendidos para descarga em meio recetor ou para reutilização na própria ETAR, sendo necessário um nível de tratamento de afinação. Este nível é denominado “avançado” ou “terciário” e destina-se, essencialmente, à remoção adicional de nutrientes, compostos tóxicos, material orgânico, sólidos suspensos e patogénicos. A separação por membrana e a desinfecção por cloragem ou por radiação UV são alguns dos processos utilizados (Metcalf & Eddy, 2003).

Tratamento da fase sólida

Apesar de as águas residuais afluentes possuírem uma componente sólida substancialmente inferior, em peso e em volume, à componente líquida, o seu tratamento é igualmente importante (Spellman, 2003). Ao longo dos vários níveis de tratamento da fase líquida, grande parte da componente sólida vai sendo removida sob a forma de gradados,

areias, escumas e lamas, sendo necessário um manuseio seguro e posterior envio para destino final dos vários resíduos produzidos (Qasim, 1999).

As areias e gradados removidos no tratamento preliminar são geralmente enviados para aterro sanitário. Contudo, as características das escumas e lamas produzidas no tratamento da fase líquida impossibilitam, na maioria dos casos, a sua deposição em aterro. A elevada concentração de sólidos (0,5% a 6%), o teor de água e o odor das escumas e lamas conferem uma complexidade ao processamento e eliminação deste tipo de resíduos que requer um conjunto de processos e operações próprios na linha de tratamento de uma ETAR (Qasim, 1999).

Apesar de existirem vários esquemas de tratamento possíveis, dependentes de fatores como a carga de sólidos, a carga hidráulica ou a existência de mais de um tipo de lamas, o tratamento convencional da fase sólida tem como principais objetivos a redução de volume e a estabilização microbiológica das lamas (Metcalf & Eddy, 2003).

A redução do volume é conseguida, primeiramente, através do espessamento dos sólidos, e depois, através da desidratação, e permite reduzir os custos de armazenamento, transporte e eliminação das lamas. Para o espessamento, são utilizados métodos como a flotação, a centrifugação ou o espessamento gravítico, e para a desidratação é comum a aplicação de filtros banda, filtros prensa ou leitos de secagem (Metcalf & Eddy, 2003)

A estabilização microbiológica tem como propósito a inibição da atividade de patogénicos, a eliminação de odores e a degradação de substratos orgânicos. A escolha do método de estabilização é bastante importante, uma vez que a respetiva eficiência terá influência no manuseamento, armazenamento e destino final das lamas tratadas (Qasim, 1999). Os métodos de estabilização das lamas produzidas numa ETAR podem ser biológicos, como a digestão aeróbia ou anaeróbia, químicos, como a oxidação química ou a adição de cal, ou térmicos (Von Sperling, 2007). Na Figura 1.2 são esquematizadas algumas das principais operações e processos unitários usados no tratamento de lamas e respetivo envio para destino final.

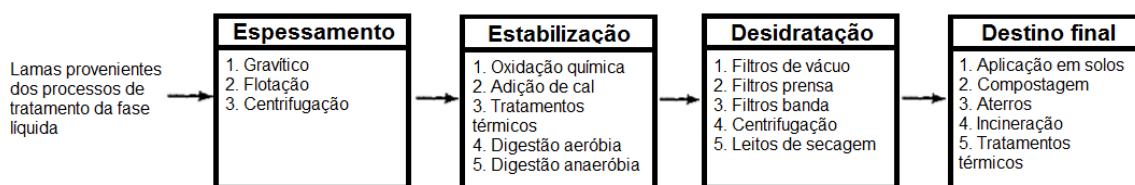


Figura 1.2 – Operações e processos unitários usados no tratamento de lamas e respetivo envio para destino final (adaptado de Qasim, 1999).

Em Portugal, estima-se que a quantidade de lamas produzidas em ETA e ETAR ascenda a duzentas mil toneladas por ano, e os custos associados ao seu tratamento, armazenamento e eliminação são cada vez mais significativos (PEAASAR II, 2007). Um dos

principais objetivos durante a concepção de uma ETAR consiste na escolha das opções que permitam, ao mesmo tempo, obter as melhores eficiências de tratamento das lamas e minimizar os gastos relativos à gestão deste tipo de resíduos (Qasim, 1999). Para tal, deve ser incentivada a procura das soluções técnicas que permitam, de forma energeticamente sustentável, uma maior redução do volume e uma estabilização mais eficiente das lamas.

Neste sentido, a digestão anaeróbia de lamas afigura-se como uma das soluções de maior potencial, pois permite aliar a redução do volume e a estabilização microbiológica das mesmas à produção de metano, que pode ser convertido em energia (Malina e Pohland, 1992).

1.2 Caracterização dos processos de digestão anaeróbia

1.2.1 Aspetos gerais

A digestão anaeróbia consiste num processo de decomposição de matéria orgânica na ausência de oxigénio que resulta na produção de biogás, constituído maioritariamente por gases como o metano (CH₄) e o dióxido de carbono (CO₂) (Casey, 1997). Apesar de ocorrer espontaneamente na natureza, a digestão anaeróbia é também um dos mais antigos processos aplicados no tratamento biológico de resíduos sólidos e de águas residuais (Metcalf & Eddy, 2003), e desde há muito tempo que são conhecidas as propriedades energéticas do metano produzido através deste processo (CCE, 2000).

A primeira situação conhecida de aproveitamento do gás metano remonta a 1859, em Bombaim, e em 1929, é construída na Dinamarca a primeira instalação de tratamento anaeróbio de águas residuais (CCE, 2000). Contudo, na maioria das situações, o gás metano produzido era apenas queimado e o processo não ganhou expressão em relação a outros já existentes, tendo o seu desenvolvimento sido muito lento (Henze *et al.*, 2002).

É só com a crise energética na década de 1970 que a dependência dos recursos naturais não renováveis levou os países desenvolvidos a procurar alternativas. A pesquisa sobre o aproveitamento do gás metano ganhou um novo fôlego, e os processos de digestão anaeróbia começaram a ser aplicados, não apenas no tratamento biológico de águas residuais, mas também no tratamento da fase sólida (CCE, 2000). Atualmente, este processo constitui uma das soluções mais utilizadas na redução do volume e na estabilização microbiológica de lamas, dadas as distintas vantagens que apresenta relativamente a outros métodos (Ramirez *et al.*, 2009).

Os processos anaeróbios ocorrem no manto de lamas de lagoas de estabilização e em reatores construídos para o efeito (Gray, 2004). Para além de reduzir os odores das lamas e de inibir a atividade dos patogénicos (Malina e Pohland, 1992), a digestão anaeróbia permite reduzir em 2 a 6 vezes a produção de biomassa em relação à digestão aeróbia, o que conduz a menores custos de processamento e de eliminação (Van Haandel e Lettinga, 1994). A recuperação do metano produzido nos digestores e a sua conversão em energia constitui também uma característica apelativa deste processo, podendo cobrir as necessidades

energéticas de aquecimento e mistura ou até, em alguns casos, da maioria das operações das instalações (Metcalf & Eddy, 2003).

Devido às suas características de sustentabilidade, a digestão anaeróbia é vista como uma solução para o futuro, e prevê-se para os próximos anos uma expansão da sua aplicabilidade (Ramirez *et al.*, 2009). Contudo, a forma como se processa e como é gerida continua associada a um certo grau de incerteza (Malina e Pohland, 1992). A sensibilidade do sistema e as várias interações entre as variáveis físicas, químicas e biológicas conferem à digestão anaeróbia uma complexidade que dificulta a otimização deste processo (Vesilind, 2003).

Assim, uma compreensão total dos fenómenos que ocorrem no interior do digestor anaeróbio poderá conduzir ao desenvolvimento de alternativas aos métodos convencionais de acompanhamento e gestão deste processo e à obtenção de melhores resultados em termos de redução de custos e de otimização da eficiência de tratamento (Van Haandel e Lettinga, 1994).

1.2.2 Descrição das fases do processo

A conversão microbiológica dos substratos orgânicos em biogás é um processo que envolve associações simbióticas entre vários grupos de microrganismos (Malina e Pohland, 1992). Estes substratos são constituídos essencialmente por macromoléculas complexas, como proteínas (25%), hidratos de carbono (25%) e lípidos (20%) (Gray, 2004, adaptado de Heukelekian, 1957), e são convertidos, em condições de anaerobiose, em metano e dióxido de carbono por diferentes populações de bactérias (Van Haandel e Lettinga, 1994).

O processo de conversão é complexo e envolve uma ação sequencial das várias populações bacterianas, a qual pode ser descrita por fases. Embora seja comum diferenciar a digestão anaeróbia em duas etapas, uma de fermentação ácida e outra de fermentação metanogénica, optou-se por caracterizar este processo em quatro fases distintas (Figura 1.3) (Van Haandel e Lettinga, 1994).

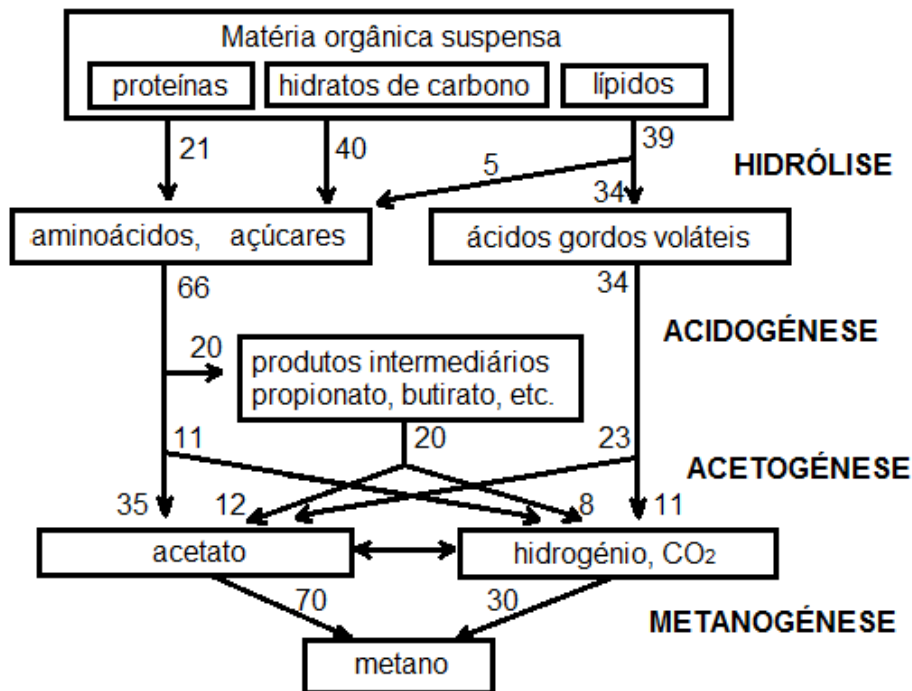


Figura 1.3 – Representação esquemática da digestão anaeróbia de macromoléculas complexas (adaptado de Van Haandel e Lettinga, 1994).

Hidrólise

Nesta fase, os sólidos voláteis de grandes dimensões são convertidos em compostos orgânicos intermediários por bactérias hidrolíticas facultativas (adaptam-se a ambientes com e sem oxigénio). As proteínas são degradadas em aminoácidos por bactérias proteolíticas, os hidratos de carbono são transformados em açúcares solúveis por bactérias celulíticas e os lípidos são convertidos em ácidos gordos de cadeia longa e em glicerol por bactérias lipolíticas (Gray, 2004).

As bactérias hidrolíticas são responsáveis pela excreção de enzimas extracelulares, denominadas exo-enzimas (Bitton, 1999). As exo-enzimas vão degradar as partículas orgânicas sólidas, e coloides e moléculas dissolvidos em compostos solúveis mais simples e de menor peso molecular, para que estes possam atravessar a membrana celular das bactérias fermentativas (Van Haandel e Lettinga, 1994; Wiesmann *et al.*, 2007; Amaral, 2012).

Acidogénese

Na acidogénese, os ácidos orgânicos solúveis formados na hidrólise atravessam a membrana celular das bactérias fermentativas e são convertidos em ácidos gordos voláteis (AGV), álcoois, ácido láctico, amoníaco (NH₃), CO₂, hidrogénio (H₂) e algum sulfídrico (H₂S) (Van Haandel e Lettinga, 1994). Durante esta fase, a quantidade de material orgânico presente no sistema mantém-se praticamente inalterada, ocorrendo apenas uma redistribuição dos vários compostos orgânicos simples (Malina e Pohland, 1992).

A acidogénese envolve a ação de um grupo heterogéneo de microrganismos, composto por bactérias acidogénicas facultativas e anaeróbias obrigatórias (Van Haandel e Lettinga, 1994). As acidogénicas anaeróbias obrigatórias não toleram a presença de oxigénio dissolvido, mesmo em concentrações de 0,01 mg/L (Casey, 1997), e utilizam a fermentação ácida para conversão dos ácidos orgânicos solúveis em produtos intermédios (Malina e Pohland, 1992). Por outro lado, as acidogénicas facultativas conseguem metabolizar a matéria orgânica usando uma via de conversão oxidativa, o que permite reduzir a presença de oxigénio dissolvido livre e proteger as bactérias anaeróbias obrigatórias (acidogénicas e metanogénicas) (Van Haandel e Lettinga, 1994).

Acetogénese

Nesta fase, os produtos da hidrólise e acidogénese são convertidos em acetato (70%) e hidrogénio (30%) (Van Haandel e Lettinga, 1994). Os ácidos gordos de cadeia curta, como o propionato e o butirato, e outros AGV são degradados por grupos de bactérias acetogénicas anaeróbias obrigatórias em acetato, hidrogénio e CO₂ (Wiesmann *et al.*, 2007).

Um outro grupo de bactérias, as homoacetogénicas, cataboliza monómeros de carbono e hidrolisa compostos com mais do que um carbono, havendo formação de acetato (Amaral, 2012). Este grupo possui também a capacidade de oxidar H₂ produzido pelas acidogénicas e acetogénicas, sendo responsável pela manutenção de baixas concentrações de hidrogénio no sistema (Rocha, 2013, adaptado de Hwang *et al.*, 2004).

Metanogénese

Apesar de não ser consensual a subdivisão dos processos que ocorrem durante a fermentação ácida, a separação da fermentação metanogénica em relação às outras fases é considerada unânime por diversos autores. Na metanogénese, os produtos finais da fermentação ácida são convertidos em metano e CO₂ por dois grupos de bactérias anaeróbias obrigatórias (Gray, 2004).

As bactérias metanogénicas acetotróficas são responsáveis pela produção de 60% a 70% do metano na digestão anaeróbia. Apesar de ser possível a descarboxilação de metanol e metilaminas no processo de conversão, o acetato constitui o principal substrato utilizado pelas acetotróficas (Chernicharo, 2007), sendo a equação de reação correspondente a seguinte (Bitton, 1999):



As bactérias metanogénicas hidrogenotróficas produzem 30% a 40% do metano na digestão anaeróbia, e utilizam o formato e, principalmente, o CO₂ e H₂ como substratos (Amaral, 2012). A redução do dióxido de carbono pelo hidrogénio é enunciada na seguinte equação de reação (Chernicharo, 2007):



Embora a hidrólise possa limitar a digestão anaeróbia para temperaturas baixas, a metanogénese constitui frequentemente o passo limitante deste processo (Van Haandel e Lettinga, 1994). A ação das bactérias metanogénicas acetotróficas é responsável pela maior parte do metano produzido, e o seu crescimento é inferior ao das hidrogenotróficas (Amaral, 2012, adaptado de Harremoes, 1983), o que se traduz na necessidade de manutenção de um número elevado de acetotróficas para que ocorra uma remoção eficiente de matéria orgânica na digestão anaeróbia (Van Haandel e Lettinga, 1994).

Por outro lado, a produção de AGV e de outros intermediários que se dissociam e libertam protões durante a acidogénese tende a diminuir o pH do sistema (Amaral, 2012). Apesar de as bactérias acidogénicas serem relativamente tolerantes a alterações no meio (Malina e Pohland, 1992), o crescimento das metanogénicas só ocorre em valores neutros de pH (Amaral, 2012). Assim, é necessária a existência de um equilíbrio dinâmico entre as duas populações microbianas, de forma a evitar situações de acidificação do sistema e de interrupção da produção de metano (Russell, 2006). No Quadro 1.3 são identificadas e quantificadas as principais populações bacterianas presentes nos digestores anaeróbios de lamas.

Quadro 1.3 – Identificação e quantificação das populações bacterianas presentes nos digestores anaeróbios de lamas (adaptado de Gray, 2004).

Grupo	Número (por mL)	Identidade genérica
Bactérias hidrolíticas		Maioria não identificada
Total	$10^8 - 10^9$	Gram-negativa
Proteolíticas	10^7	<i>Eubacterium</i>
Celulíticas	10^5	<i>Clostridium</i>
Produtoras de hidrogénio		Não identificada
Acetogénicas	10^6	Gram-negativa
Homoacetogénicas	$10^5 - 10^6$	<i>Clostridium</i> <i>Acetobacterium</i>
Metanogénicas	$10^6 - 10^8$	<i>Methanobacterium</i> <i>Methanospirillum</i> <i>Methanococcus</i> <i>Methanosarcina</i> ' <i>Methanothrix</i> '
Redutoras de sulfato	10^4	<i>Desulfovibrio</i> <i>Desulfotomaculum</i>

1.2.3 Fatores ambientais

O desenvolvimento do processo de conversão anaeróbia é pautado pela quantidade e qualidade do substrato disponível e pelas populações microbianas presentes. Contudo, os fatores ambientais do sistema condicionam sobremaneira o desenvolvimento da digestão anaeróbia e a produção de biogás (Malina e Pohland, 1992).

Os principais fatores ambientais da digestão anaeróbia de lamas são a temperatura, o pH, substâncias tóxicas ou inibidoras e nutrientes. A conversão de AGV em metano constitui, geralmente, a fase limitante dos processos anaeróbios, sendo por isso desejável a manutenção de condições ótimas para as bactérias metanogênicas no digestor (Malina e Pohland, 1992).

Temperatura

A temperatura constitui um importante fator a considerar no plano do tratamento de águas residuais, uma vez que condiciona a atividade microbiana e influencia a viscosidade do líquido e a solubilidade dos gases (Von Sperling, 2007). Na digestão anaeróbia, a temperatura é particularmente importante durante as fases de hidrólise e de metanogênese (Vesilind, 2003).

A taxa de conversão de substrato na metanogênese é variável com a temperatura e existem duas gamas ótimas de funcionamento (Malina e Pohland, 1992). A atividade das metanogênicas é alta nas gamas mesófila (30 a 38 °C) e termófila (50 a 57 °C) (Metcalf & Eddy, 2003), e é inibida no intervalo entre 40 e 50 °C (Malina e Pohland, 1992); a temperatura dos processos de digestão é, geralmente, regulada por estes intervalos (Figura 1.4).

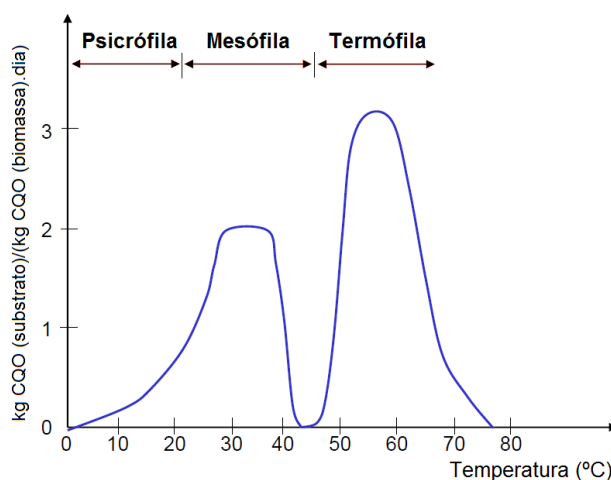


Figura 1.4 – Taxa de remoção específica de substrato em função da temperatura (adaptado de Henze *et al.*, 2002).

No geral, a diminuição de 1 °C em relação à gama mesófila ótima tende a diminuir a taxa de remoção de substrato em 11%, o que se traduz numa taxa de digestão de 35% num sistema a 20 °C. Para temperaturas do digestor inferiores a 20 °C, a taxa de conversão de lípidos é bastante reduzida, o que significa que a hidrólise pode constituir a fase limitante de todo o processo de digestão anaeróbia se as condições do meio forem desfavoráveis à

atividade das bactérias hidrolíticas e à ação das exo-enzimas lipídicas (Van Haandel e Lettinga, 1994).

A aplicação dos processos anaeróbios constitui uma solução atrativa para países tropicais, onde a temperatura das águas residuais é superior a 20 °C. Em países com um clima frio, onde as águas residuais rondam os 10 °C, as necessidades energéticas para elevar a temperatura até gamas ótimas de funcionamento podem não justificar o investimento (Van Haandel e Lettinga, 1994). No entanto, o aquecimento das águas residuais e lamas em gamas mesófilas e até termófilas, mesmo em países tropicais, poderá ser vantajoso em termos de eficiência de desidratação, remoção de substrato, digestão de escumas e destruição de patogénicos (Vesilind, 2003).

Embora a temperatura a que se processa a digestão anaeróbia seja importante, a manutenção de um valor estável é essencial. Os microrganismos envolvidos no processo anaeróbio, especialmente as bactérias metanogénicas, demonstram uma acentuada sensibilidade a alterações de temperatura, e existe um risco de falha no sistema se ocorrerem variações superiores a 1 °C por dia (Vesilind, 2003).

pH

As bactérias metanogénicas são também sensíveis a variações do pH do sistema (Vesilind, 2003). Embora a fermentação anaeróbia se possa processar com pH entre 5,0 e 9,0 (Russell, 2006), um digester normalizado possui, usualmente, um pH compreendido entre 6,2 e 7,8 (Gray, 2004). O intervalo de funcionamento ótimo da metanogénese ocorre para valores neutros de pH (6,6 a 7,6, segundo Metcalf & Eddy, 1991) e um aumento em relação a esta gama resulta numa produção e acumulação excessiva de produtos de conversão básicos, como o NH_3 (Malina e Pohland, 1992).

Para condições em que o meio apresenta um pH inferior a 6,2, a fermentação metanogénica diminui (Metcalf & Eddy, 1991). Como a sensibilidade das bactérias acidogénicas em relação a variações do pH é inferior à das metanogénicas, tende a haver uma acumulação de compostos ácidos, como os AGV, o que pode resultar numa acidificação do sistema (Figura 1.5) e conseqüente falha do digester (Van Haandel e Lettinga, 1994).

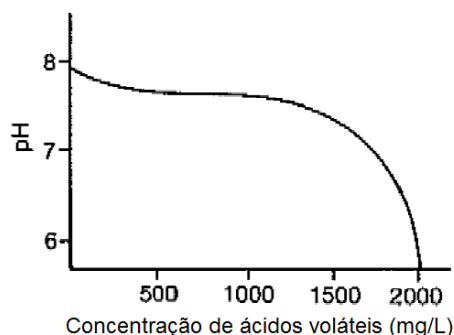


Figura 1.5 – Efeito da produção de ácidos voláteis no pH de um digester (adaptado de Gray, 2004).

Assim, é importante a presença de alcalinidade no sistema, essencialmente sob a forma de CO_2 , NH_3 e bicarbonato (HCO_3^-), para que a sua combinação com iões hidrogénio confira ao digestor um poder-tampão (Vesilind, 2003). A metanogénese aumenta a alcalinidade do meio e a acidogénese reduz a sua concentração, o que significa que um equilíbrio dinâmico entre as duas fases é essencial para a manutenção do poder-tampão do sistema (Henze *et al.*, 2002). Na Figura 1.6 são representadas as condições normais de pH e de alcalinidade na digestão anaeróbia de lamas.

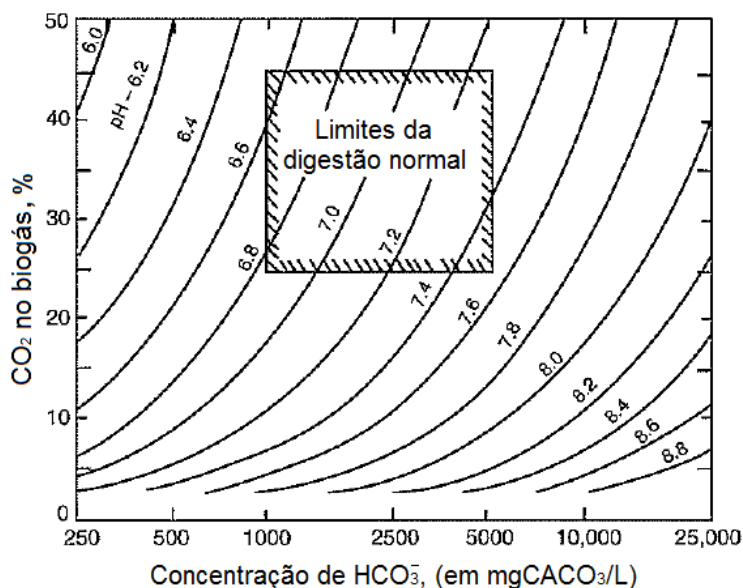


Figura 1.6 – Condições de pH e de concentração de bicarbonato para a digestão de lamas (adaptado de Grady *et al.*, 1999).

O planeamento de sistemas de mistura, aquecimento e alimentação do digestor são também importantes para minimizar o risco de falha do processo. No entanto, quando ocorre uma sobrecarga do digestor, ou quando a fermentação ácida prevalece sobre a metanogénica, poderá ser necessária a adição de reagentes, como a cal ou o bicarbonato de sódio, para que o excesso de ácidos seja neutralizado (Vesilind, 2003). Para o funcionamento normalizado de um digestor, deverá ser assegurada uma alcalinidade entre 2.000 a 5.000 $\text{mg CaCO}_3/\text{L}$ e uma concentração de AGV entre 250 e 1.000 mg/L (Gray, 2004).

Nutrientes

As exigências em macronutrientes devem ser suplantadas para que alguns microrganismos anaeróbios possam sintetizar vitaminas e aminoácidos essenciais. No geral, a relação entre a CQO e o azoto constitui um bom indicador das necessidades de nutrientes do sistema. Uma relação de CQO:N de 55:1 em sistemas com alta carga e 140:1 em sistemas de baixa carga (Malina e Pohland, 1992) permite assegurar um conteúdo adequado de azoto (80 a 120 g de N/kg SSV), fósforo (10 a 25 g de P/kg SSV), enxofre (10 a 25 g de S/kg SSV) e ferro (5 a 15 g de Fe/kg SSV) (Henze *et al.*, 2002).

O crescimento e metabolismo das populações de microrganismos envolvidas na digestão anaeróbia necessitam também da presença de micronutrientes, em particular, magnésio, cálcio, sódio, bário, molibdato e cobalto. O sistema enzimático das acidogênicas e metanogênicas requer ainda a presença de selênio, tungstênio e níquel. Usualmente, estes nutrientes são abundantes nos resíduos orgânicos, não sendo necessária a sua adição (Malina e Pohland, 1992).

Substâncias tóxicas

Para além da concentração de iões hidrogénio, outras substâncias afetam a taxa de remoção de substrato (Van Haandel e Lettinga, 1994), e quando a sua concentração ultrapassa um determinado valor, poderão ser criadas condições de instabilidade no digestor anaeróbio (Quadro 1.4). O efeito mais comum é a inibição da metanogénese e resulta da presença, em concentrações tóxicas, de compostos como o amoníaco, oxigénio, metais pesados, sulfuretos e alguns catiões (Vesilind, 2003).

Quadro 1.4 – Concentrações de toxicidade e de inibição de materiais inorgânicos na digestão anaeróbia (adaptado de Parkin e Owen, 1986, *apud* Vesilind, 2003).

Substância	Concentração de inibição moderada, mg/L	Concentração de inibição forte, mg/L
Na ⁺	3.500-5.500	8.000
K ⁺	2.500-4.500	12.000
Ca ⁺⁺	2.500-4.500	8.000
Mg ⁺⁺	1.000-1.500	3.000
Amoníaco-azoto (dependente do pH)	1.500-3.000	3.000
Sulfureto (gás não ionizado)	200	200
Cobre (Cu)	-	0,5 (solúvel)
	-	50-70 (total)
Crómio VI (Cr)	-	3,0 (solúvel)
	-	200-250 (total)
Crómio III	-	180-420 (total)
Níquel (Ni)	-	2,0 (solúvel)
	-	30,0 (total)
Zinco (Zn)	-	1,0 (solúvel)

No entanto, a presença da maioria destas substâncias em concentrações inibidoras da atividade microbiana não é comum em águas residuais. Algum oxigénio dissolvido poderá ser introduzido no sistema durante a alimentação do digestor, mas é depois consumido no metabolismo oxidativo das bactérias acidogênicas facultativas (Van Haandel e Lettinga, 1994).

Também o sulfídrico dissolvido poderá estar presente no sistema, como resultado da redução do sulfato, mas a concentração expectável deste composto em tratamentos anaeróbios de águas residuais é inferior a 50 mg/L (Rinzema, 1989, citado por Van Haandel e Lettinga, 1994). Nas situações em que a presença do sulfídrico dissolvido se aproxima de valores inibidores (> 200 mg/L), a concentração poderá ser controlada através da adição de sais de ferro, como, por exemplo, o cloreto férrico (Vesilind, 2003; Grady *et al.*, 1999).

Outros fatores ambientais

Alguns estudos têm sido realizados com o objetivo de analisar os efeitos de outros fatores na produção de metano e nos processos anaeróbios no geral. Valores de salinidade superiores a 0,2 M de NaCl possuem um efeito inibidor nos fluxos de metano, e a força iónica tende a afetar a atividade química do sistema e a inibir o efeito de outras espécies químicas. O potencial redox constitui também uma condicionante da atividade de bactérias metanogénicas, devendo ser mantidos valores entre -490 e -550 mV (Malina e Pohland, 1992).

1.2.4 Aspectos de dimensionamento dos digestores

O dimensionamento de digestores anaeróbios é importante na medida em que condiciona a quantidade de sólidos voláteis destruídos, que deve variar entre 40 a 60%. A sua destruição é função do tempo que os sólidos ficam retidos dentro do digestor, o qual pode ser expresso em termos de tempo de retenção de sólidos (T_{RS}) ou tempo de retenção hidráulico (T_{RH}) (Vesilind, 2003). O T_{RS} , também denominado *mean cell residence time* ou período de digestão (Qasim, 1999), é calculado a partir do quociente entre a quantidade de sólidos no digestor e os sólidos extraídos por dia, e o T_{RH} é obtido através da divisão do volume útil do digestor pelo caudal afluente diário (Vesilind, 2003). Para reatores sem recirculação, o T_{RS} é igual ao T_{RH} (Metcalf & Eddy, 2003).

O tempo de retenção está diretamente relacionado com o desenvolvimento das fases de digestão anaeróbia, e um acréscimo no seu valor corresponde a um aumento do tempo das reações que ocorrem em cada fase do processo. Se não for assegurado um tempo de retenção mínimo, o crescimento de algumas populações de bactérias pode não ser conseguido, e as reações por que estão responsáveis acabam por parar (Vesilind, 2003).

O aquecimento do digestor constitui outro aspeto importante a considerar na sua conceção (Figura 1.7). Para além de assegurar as devidas condições ótimas de temperatura para a atividade das populações microbianas envolvidas no processo, o aquecimento permite reduzir substancialmente o T_{RS} (Gray, 2004) e minimizar o volume útil do digestor (Vesilind, 2003). O aquecimento pode ser conseguido através da injeção de vapor, ou a partir de permutadores de calor internos ou externos, e deverá ter em consideração as perdas através das paredes, cobertura, chão e tubagens do digestor (Qasim, 1999).

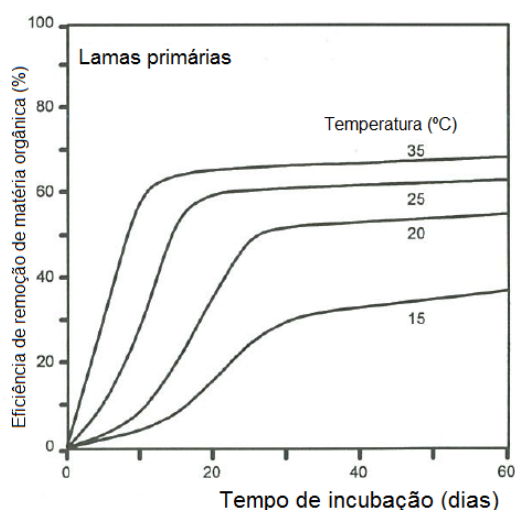


Figura 1.7 – Influência da temperatura na extensão e taxa de digestão anaeróbia de lamas primárias (adaptado de Van Haandel e Lettinga, 1994).

A mistura do digestor anaeróbio constitui também um aspeto bastante importante, uma vez que promove o contacto entre as lamas e a biomassa, dispersa os produtos metabólicos produzidos durante o processo (Qasim, 1999), impede a formação de escumas e a decantação de sólidos, previne a acumulação localizada de substâncias inibidoras (Malina e Pohland, 1992), mantém a mistura dos sólidos e temperatura no reator uniformes e estimula a libertação de gás nas zonas inferiores do digestor (Gray, 2004).

A agitação do conteúdo do reator pode ocorrer naturalmente, através da ascensão das bolhas de gás e das correntes de convecção térmica resultantes da adição de lamas aquecidas (Qasim, 1999). No entanto, em algumas situações a agitação natural não é suficiente, e é necessária a instalação de sistemas de mistura artificiais. A agitação pode ser feita por bombagem, por gás ou mecanicamente (Vesilind, 2003).

A configuração de um digestor anaeróbio de biomassa suspensa, em relação ao T_{RS} e aos sistemas de mistura e aquecimento, pode ser de três tipos: baixa carga, alta carga ou em dois estágios (Qasim, 1999).

Digestão em baixa carga

Nos digestores de baixa carga, também denominados digestores anaeróbios convencionais, não existe nenhum sistema de mistura artificial e o sistema de aquecimento pode ou não estar presente. O período de digestão varia entre 30 a 60 dias (Vesilind, 2003). A alimentação das lamas ao digestor é intermitente e não existe recirculação (Gray, 2004). A ausência de agitação determina a estratificação do conteúdo do digestor em quatro zonas: uma camada de escumas, uma camada líquida ou de sobrenadante, uma camada de digestão ativa e uma camada de sólidos digeridos e inertes (Figura 1.8) (Qasim, 1999).

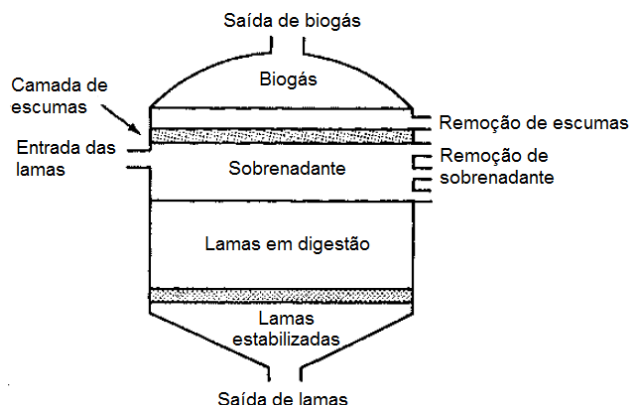


Figura 1.8 – Digestor anaeróbico de baixa carga (adaptado de Qasim, 1999).

A digestão anaeróbia em baixa carga é considerada um processo instável e ineficiente (Qasim, 1999), e a sua aplicação é usada apenas em estações de pequena dimensão (de caudal afluyente inferior a 4.000 m³/dia) e sem o propósito de obter uma digestão otimizada (Vesilind, 2003).

Digestão em alta carga

No processo de digestão em alta carga (Figura 1.9) o digestor é aquecido e mantido a temperaturas na gama mesófila ou termófila (Vesilind, 2003). O conteúdo do reator é homogeneizado através da manutenção de condições de mistura completa e não existe separação e recolha do sobrenadante (Qasim, 1999). O período de digestão varia entre 10 e 20 dias (Russell, 2006). A alimentação das lamas ao digestor é contínua, sendo por isso um processo com elevada estabilidade (Gray, 2004; Vesilind, 2003). Antes da digestão, é comum a existência de espessamento de lamas, o que resulta numa maior carga de sólidos afluyente ao digestor (Qasim, 1999).

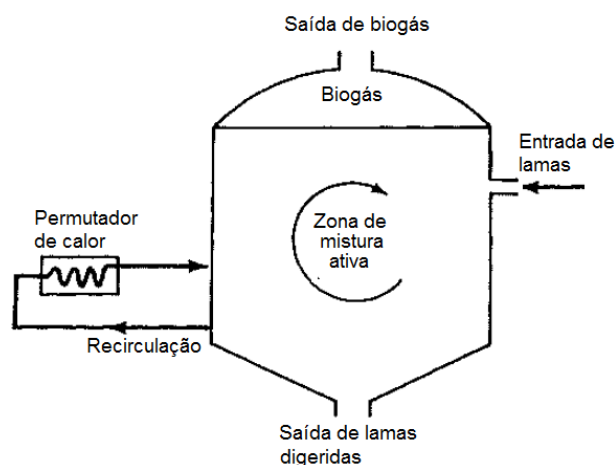


Figura 1.9 – Digestor anaeróbico de alta carga (adaptado de Qasim, 1999).

Digestão em dois estágios

A digestão anaeróbia em dois estágios consiste numa combinação dos processos de baixa e alta carga, havendo uma divisão entre a função de fermentação e a de separação sólido-líquido em dois tanques (Figura 1.10). O primeiro tanque assegura uma estabilização em alta carga e o segundo tanque armazena as lamas e separa a fração sólida da líquida, não possuindo sistemas de aquecimento ou mistura (Vesilind, 2003). Em alguns casos, o segundo tanque pode ainda permitir a digestão adicional de sólidos e a recuperação de gás (Qasim, 1999).

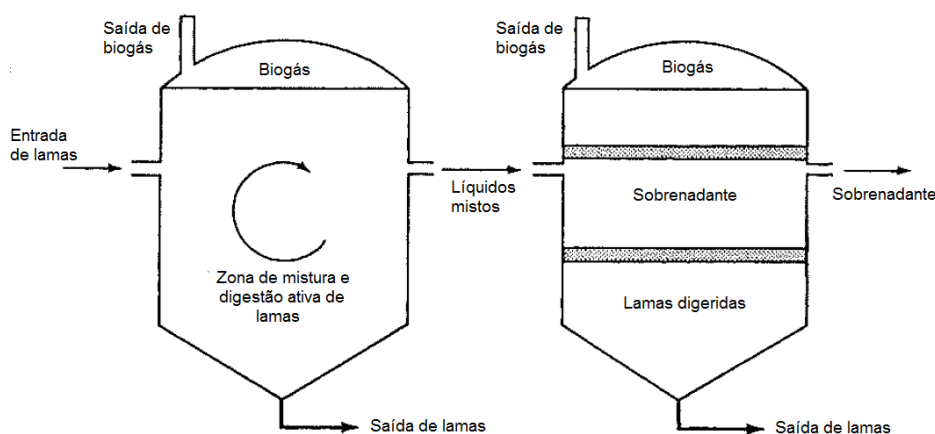


Figura 1.10 – Digestão anaeróbia em dois estágios (adaptado de Qasim, 1999).

1.2.5 Produtos

Os principais produtos da digestão anaeróbia de lamas são as lamas digeridas (praticamente inócuas) e gases (Malina e Pohland, 1992). As lamas retiradas do digestor são compostas por sólidos sedimentados, material orgânico estabilizado e material celular, e depois de passarem por uma etapa de desidratação, possuem características adequadas à sua aplicação em solos ou aterros. Nos digestores de baixa carga ou de dois estágios, é também produzido sobrenadante, o qual é geralmente enviado para a cabeça da ETAR, para ser tratado com as águas residuais (Qasim, 1999).

O conjunto de gases produzidos durante os processos anaeróbios é denominado biogás. Tipicamente, 60% a 70% do seu conteúdo (em volume) é constituído por CH_4 , e 25% a 30% por CO_2 (Qasim, 1999). Estão ainda presentes pequenas quantidades de N_2 , H_2 , H_2S , vapor de água e outros gases (Metcalf & Eddy, 2003). A composição do biogás depende essencialmente da composição do substrato (Casey, 1997), mas não apresenta, geralmente, alterações significativas (Malina e Pohland, 1992). A quantidade produzida é variável e depende do conteúdo em sólidos voláteis nas lamas afluentes e da atividade biológica no digestor. A produção específica de biogás varia entre $0,75 \text{ m}^3/\text{kg}$ a $1,12 \text{ m}^3/\text{kg}$ de sólidos voláteis destruídos (Metcalf & Eddy, 2003), e o poder calorífico entre 21.000 kJ/m^3 a 24.500

kJ/m^3 (Qasim, 1999). No Quadro 1.5 são apresentados os fatores de conversão em energia correspondentes a cada conteúdo de metano no biogás.

Quadro 1.5 – Conteúdo de metano no biogás e respetivos fatores de conversão em energia (adaptado de Henze *et al.*, 2002).

Conteúdo de CH_4 no biogás		%	50	60	70	80	100
Poder calorífico		kJ/m^3	17.500	21.000	24.500	28.000	35.000
Conversão de 1 m^3 de CH_4 em:	eletricidade (30%)	kWh/m^3	1,5	1,7	2,0	2,3	2,9
	calor (50%)	kWh/m^3	2,5	2,9	3,4	3,9	4,9
	perdas (20%)	kWh/m^3	1,0	1,1	1,3	1,5	1,9

O elevado conteúdo energético do biogás torna apelativa a sua utilização como combustível para o aquecimento das lamas afluentes e para a manutenção da temperatura adequada no digestor (Malina e Pohland, 1992). Cerca de 50% da energia é aplicada na produção de calor e 30% pode ser utilizada para cobrir outras necessidades energéticas da estação (Henze *et al.*, 2002). Assim, os digestores devem ser equipados com dispositivos que permitam captar os gases libertados na superfície do líquido e encaminhá-los para tanques de armazenamento (Vesilind, 2003).

A mistura do ar com o metano ou com o sulfídrico, mesmo em concentrações de aproximadamente 5%, é potencialmente explosiva (Amaral, 2012) e os sistemas de recolha, distribuição e armazenamento de biogás devem ser isolados e concebidos de modo a que a pressão seja mantida positiva, para prevenir o risco de explosão (Vesilind, 2003). Se o biogás for utilizado como combustível, a presença de sulfídrico, vapor de água e partículas constitui também um inconveniente, sendo por vezes necessária a introdução de reagentes ou a instalação de depuradores antes da sua utilização nos motores de combustão (Metcalf & Eddy, 2003).

1.2.6 Monitorização e controlo do processo

A operação de um digestor anaeróbio deve ser acompanhada pela medição de parâmetros de monitorização e controlo do processo. A evolução dos valores de cada parâmetro poderá indiciar a ocorrência de problemas operacionais futuros (Quadro 1.6) e a sua correta interpretação permitirá criar medidas de prevenção adequadas (Henze *et al.*, 2002).

Quadro 1.6 – Alguns dos parâmetros operacionais a monitorizar nos processos anaeróbios e correspondentes variações normais admissíveis (adaptado de Henze *et al.*, 2002).

Parâmetro	Propósito	Variação normal admissível
Temperatura	Manutenção da temperatura constante	$\pm 1,0$ °C/dia
pH	Controlo da instabilidade	$\pm 0,5$ /dia
Concentração da M.O. nas lamas digeridas	Controlo da eficiência do tratamento	$\pm 10\%$ /dia
Produção de biogás	Controlo das bactérias metanogénicas	$\pm 20\%$ /dia
Qualidade das lamas (% volátil)	Controlo da função de estabilização das lamas	$\pm 5\%$ /dia

A escolha do tipo e número de medições a efetuar deve ser feita de acordo com o modelo de digestão que é operado e com as condições ambientais mantidas no reator (Henze *et al.*, 2002). Contudo, alguns dos parâmetros a medir requerem a realização de ensaios químicos ou microbiológicos individuais, que tendem a ser morosos e dispendiosos. Assim, a medição rotineira destes parâmetros traduz-se numa solução bastante limitada, o que constitui uma das razões pela qual não existe unanimidade quanto à identificação dos métodos mais eficazes de monitorização e controlo da digestão anaeróbia (Malina e Pohland, 1992).

Apesar de a digestão anaeróbia ser um dos métodos mais antigos utilizados na estabilização de lamas (Ramirez *et al.*, 2009), tem prevalecido um certo desconhecimento sobre a microbiologia e bioquímica das reações de conversão e uma incompreensão do significado individual ou coletivo de alguns dos mecanismos e parâmetros das várias fases (Malina e Pohland, 1992). O grau de incerteza associado a este processo é comparável ao do funcionamento de uma caixa negra (Holubar *et al.*, 2002), sendo esta outra razão para que durante muito tempo tenha existido uma escassez de soluções em termos de controlo e até mesmo de otimização do desempenho dos digestores (Malina e Pohland, 1992).

Vários estudos têm sido realizados no sentido de averiguar o desempenho de determinadas ferramentas na gestão e otimização da digestão. Os testes piloto têm vindo a ser utilizados com este propósito, mas os extensos períodos que requerem na operação do processo condicionam a aplicabilidade deste método. Por essa razão, os modelos matemáticos de controlo e predição do comportamento da digestão anaeróbia assumem-se como a solução mais atrativa (Parker, 2005).

Os primeiros modelos matemáticos criados eram do tipo “modelação mecânica” e baseavam-se em equações de balanço de massa que simulavam os processos inerentes à digestão anaeróbia (Lauwers *et al.*, 2013). Contudo, a crescente complexidade das tecnologias avançadas de digestão culminou na necessidade de desenvolver um modelo generalizado, e

em 1997 é desenvolvido o ADM1 da IWA (Batstone *et al.*, 2002; Wayne, 2005). Este modelo integra 26 variáveis de estado dinâmicas, modela 19 processos químicos biocinéticos e abrange três reações de transferência cinética líquido-gás (Russell, 2006), e a simulação dos processos ocorridos na digestão anaeróbia tem vindo a obter bons resultados em termos de descrição do comportamento do digestor (Lee *et al.*, 2009).

No entanto, a utilização deste método e da modelação mecânica, no geral, apresenta algumas desvantagens (Lee *et al.*, 2009). O grau de complexidade dos mecanismos do processo (Schubert *et al.*, 1994, citado por Holubar *et al.*, 2002) e a dificuldade na medição de alguns parâmetros limitam substancialmente a aplicabilidade de equações de balanço de massa na simulação do comportamento dos digestores (Choi, 2000, Vanrolleghem *et al.*, 1999, citado por Lee *et al.*, 2009). Adicionalmente, a aplicação do modelo ADM1 em diferentes tecnologias avançadas de digestão anaeróbia implica a introdução de alguns ajustes (Blumensaata e Keller, 2005), uma tarefa largamente dificultada pela falta de experiência e conhecimento verificada neste tema (Lee *et al.*, 2009).

Para além da modelação mecânica, os modelos matemáticos podem ser também inferidos a partir da informação recolhida durante a operação do digestor (Strik *et al.*, 2005). Este tipo de modelação denomina-se *black-box* e é considerado mais vantajoso que a modelação mecânica na simulação de processos que não são compreendidos na totalidade ou em que é impossível a determinação de alguns parâmetros essenciais, como é o caso da digestão anaeróbia (Strik *et al.*, 2004, citado por Strik *et al.*, 2005).

A modelação do tipo *black-box* pode ser feita através da regressão por mínimos quadrados parciais (PLS) ou por redes neuronais artificiais (RNA). A modelação do tipo PLS corresponde a uma abordagem linear, e em simulações dinâmicas, como é o caso da digestão anaeróbia, a abordagem não-linear das RNA será a mais aconselhável para previsão de biogás ou para controlo do processo (Lauwers *et al.*, 2013).

Em comparação com métodos de estimação não-linear usados em processos biológicos, as RNA apresentam distintas vantagens, uma vez que não requerem um conhecimento prévio da estrutura das relações existentes entre variáveis importantes e podem ser usadas em sistemas de apoio à decisão. Estes sistemas permitem automatizar tarefas de gestão de informação e contribuir para a otimização da operação do sistema, tendo em vista a prevenção de instabilidade no digestor e o aumento da produção de metano (Holubar *et al.*, 2002). Talvez por esta razão, a utilização das RNA como ferramenta de apoio à decisão esteja já bastante difundida no seio da Engenharia Sanitária, sendo possível encontrar diversos estudos que abordam a sua aplicabilidade, não apenas na digestão anaeróbia, mas também na modelação de outros processos de tratamento e na previsão de valores de alguns compostos em ETA e ETAR.

1.3 Redes Neurais Artificiais

1.3.1 Aspectos gerais

O desejo de compreender o funcionamento do cérebro humano é antigo, e a informação obtida sobre as suas funções tem vindo a aumentar gradualmente ao longo dos anos. O maior entendimento sobre o comportamento das redes neuronais biológicas e o fascínio pela forma rápida e precisa com que conseguem processar informação têm constituído fontes de motivação para a procura de tecnologias artificiais que simulem o funcionamento do cérebro humano (Kartalopoulos, 1996).

Em 1943 é desenvolvido por McCulloch e Pitts o primeiro modelo matemático de neurónio artificial. Apesar de ter sido demonstrado que o neurónio tinha aptidão para resolver problemas lógicos complexos, o modelo carecia de capacidade de aprendizagem, ou adaptação (Lingireddy e Brion, 2005). Nos anos que se seguiram, muitos trabalhos pretenderam desenvolver métodos de treino automático de redes de neurónios, mas as limitações referentes aos processos de aprendizagem não possibilitaram a ocorrência de avanços significativos. É só na década de 1980 que é divulgado por Rumelhart e McClelland o algoritmo de retro propagação do erro ou *error-back-propagation*, introduzido por Paul Werbos em 1974, que veio permitir ultrapassar a maioria das dificuldades de aprendizagem identificadas nas décadas anteriores (Taylor, 1996).

Desde então, tem-se verificado um crescimento acentuado da investigação no campo das redes neuronais artificiais, mais precisamente na criação e aplicação de novos métodos de treino (Taylor, 1996). Atualmente, estes modelos matemáticos são capazes, quando devidamente treinados, de aprender, reconhecer, armazenar e generalizar padrões subjacentes a um determinado conjunto de dados (Lingireddy e Brion, 2005). De entre as principais características das RNA, destacam-se as seguintes (Haykin, 1999; Matos, 2008):

- ✓ Não-linearidade;
- ✓ Robustez;
- ✓ Capacidade de generalização;
- ✓ Desconhecimento prévio das relações existentes entre variáveis.

A simplicidade da utilização das redes neuronais artificiais na simulação, predição e modelação do desempenho de processos constitui ainda outra das razões pela qual esta ferramenta é usada no presente na previsão e simulação de uma variedade de processos dinâmicos de diversos ramos da ciência e da engenharia (Vemuri, 1992; Prakash *et al.*, 2008, citado por Elmolla e Chaudhuri, 2011).

1.3.2 Princípios de funcionamento

O funcionamento das RNA é análogo ao dos processos biológicos do cérebro humano (Lingireddy e Brion, 2005). Estas podem ser definidas como um conjunto de redes constituídas por vários processadores simples (neurónios), conectados por vias de comunicação (conexões). Os neurónios estão ordenados por camadas, sendo que os da primeira camada estão atribuídos às variáveis de input e os da última às variáveis de output (Lauwers *et al.*, 2013). Entre estas duas camadas podem ainda existir uma ou mais camadas ocultas (Figura 1.11) (Kartalopoulos, 1996).

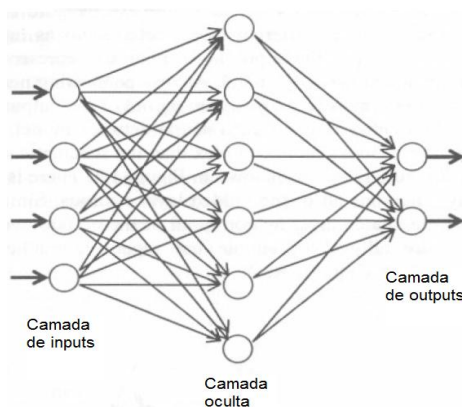


Figura 1.11 – Representação esquemática de uma RNA típica (adaptado de Lingireddy e Brion, 2005).

Cada neurónio possui uma quantidade de memória local que contém dados numéricos provenientes de n conjuntos de inputs x_j . Quando o neurónio integra uma rede neuronal, passa a ser denominado nó, contendo n conjuntos de inputs x_{ij} . A cada input é atribuído um peso sináptico w_j , sendo em seguida somados através de uma função de input (Lingireddy e Brion, 2005; Kartalopoulos, 1996; Strik *et al.*, 2005).

Posteriormente, é adicionado ao resultado da função de input um termo de viés w_0 e um valor limite Θ que tem de ser atingido ou excedido para que o neurónio produza um sinal R . O valor resultante é usado como argumento de uma função de transferência (ou ativação) f que atua no sinal (Lingireddy e Brion, 2005; Kartalopoulos, 1996; Strik *et al.*, 2005). Na Figura 1.12 é apresentado o modelo básico de um neurónio.

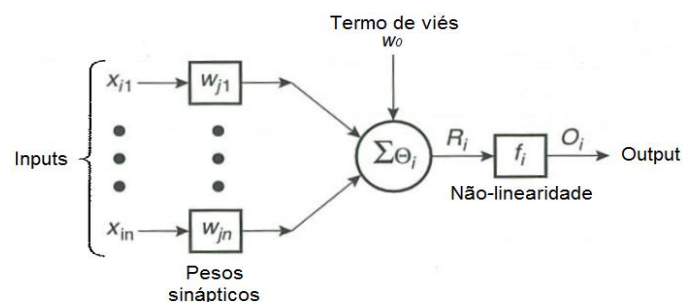


Figura 1.12 – Modelo básico de um neurónio (adaptado de Kartalopoulos, 1996).

A função de ativação é geralmente não-linear e pode ser do tipo *hard limiting*, *threshold logic* ou sigmoidal, variando entre -1 e 1 ou entre 0 e 1. O resultado é o output O do neurónio, que vai constituir o input de neurónios da camada seguinte (Vemuri, 1992; Kartalopoulos, 1996). A função de ativação pode ser descrita pela seguinte relação (Kartalopoulos, 1996):

$$O_i = f_i \left(\sum_{j=1}^n w_{ij} \cdot x_{ij} + w_0 \right) \quad (\text{Equação 1.3})$$

A condição de ativação do nó é dada pela seguinte expressão (Kartalopoulos, 1996):

$$\sum_{j=1}^n w_{ij} \cdot x_{ij} \geq \theta_i \quad (\text{Equação 1.4})$$

1.3.3 Topologias das RNA

Quando o output de um nó é distribuído por outros nós, a conexão correspondente pode ter várias topologias. As principais configurações topológicas incluem as redes de camada única, multicamada, *feed-forward*, *feed-back* e conectividade lateral (Vemuri, 1992). A forma como as camadas estão organizadas é denominada arquitetura da rede, e define, em conjunto com o sentido e peso das conexões, os fluxos de informação nas RNA.

Redes de camada única

O percetão desenvolvido por Rosenblatt em 1958 constitui a forma mais simples de uma RNA, e possui duas camadas (Lingireddy e Brion, 2005). No entanto, apenas a camada de output possui uma capacidade computacional, uma vez que a de entrada tem como único papel a receção dos inputs e a sua distribuição à camada seguinte (Haykin, 1999). Assim, as RNA constituídas por percetões de Rosenblatt são consideradas como sendo de camada única.

As redes de camada única conseguem apenas classificar padrões linearmente separáveis, o que constituiu uma séria limitação da aplicação do percetão de Rosenblatt na modelação de funções não-linearmente separáveis. Contudo, em 1969, Minsky e Papert demonstraram que uma rede com duas camadas poderia superar as restrições apontadas aos percetões de camada única (Lingireddy e Brion, 2005).

Redes de camada múltipla

Quando existe a presença de uma ou mais camadas ocultas entre as camadas de input e de output, a rede é considerada como sendo do tipo camada múltipla, multicamada ou *multilayer Perceptron* (MLP) (Kartalopoulos, 1996). À semelhança da camada de saída, a camada oculta é composta por nós computacionais, que intercedem entre o input externo e o output da rede. A presença de uma ou mais camadas ocultas possibilita a obtenção de estatísticas de complexidade superior, o que constitui uma distinta vantagem em situações em

que a camada de input é extensa (Haykin, 1999). A rede passa também a adquirir uma perspectiva global, em detrimento de uma conectividade local, como resultado do conjunto adicional de conexões sinápticas e de uma dimensão extra de interações neuronais (Churchland e Sejnowski, 1992, citado por Haykin, 1999).

A configuração dos MLP em termos do número de camadas ocultas e do número de nós por camada é variável. No geral, uma ou duas camadas ocultas são suficientes para a implementação de funções lógicas Booleanas arbitrárias na repartição de padrões espaciais, para a classificação de padrões ou para a implementação de transformações não-lineares em problemas de aproximação funcional. O número ótimo de nós por cada camada não é fixo e a sua determinação não encerra nenhum método formal, sendo comum a utilização do método da tentativa e erro (Kartalopoulos, 1996).

As conexões que se estabelecem numa RNA correspondem, geralmente, a interações entre nós de camadas adjacentes. Se todos os nós de cada camada de uma RNA estiverem interligados com os da camada adjacente, a rede é totalmente conectada. Se alguma das conexões sinápticas estiver ausente, a rede é considerada parcialmente conectada (Haykin, 1999).

Redes *feed-forward*

Qualquer RNA pode ser classificada como rede de camada única ou como rede de camada múltipla, tendo em consideração a sua arquitetura. Existem, contudo, outras topologias possíveis para este tipo de redes, e o sentido do fluxo de informação entre nós ou camadas é frequentemente utilizado para distinguir diferentes classes de RNA.

Quando a informação é transmitida no sentido da camada de inputs para a camada de outputs, a rede é denominada *feed-forward* ou acíclica. Os nós da camada de entrada recebem os inputs e enviam a informação para os nós da camada seguinte e nunca para os da camada anterior. As conexões que ocorrem neste tipo de redes são sempre estabelecidas entre nós de camadas adjacentes, e nunca entre nós da mesma camada, o que se traduz num fluxo unidirecional da informação (Haykin, 1999).

A topologia *feed-forward* pode ser aplicada em redes com uma única camada ou com múltiplas camadas, sendo as últimas denominadas redes *multilayer feed-forward* (MLFN) (Haykin, 1999). Na Figura 1.13 é representada uma rede neuronal do tipo MLFN.

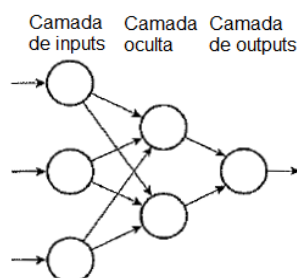


Figura 1.13 – Rede neuronal do tipo *multilayer feed-forward* (adaptado de Kartalopoulos, 1996).

Outras topologias

Para além das redes do tipo *feed-forward*, existem outras topologias que podem ser aplicadas às RNA, como resultado da combinação de diferentes arquiteturas e sentidos do fluxo de informação (Figura 1.14).

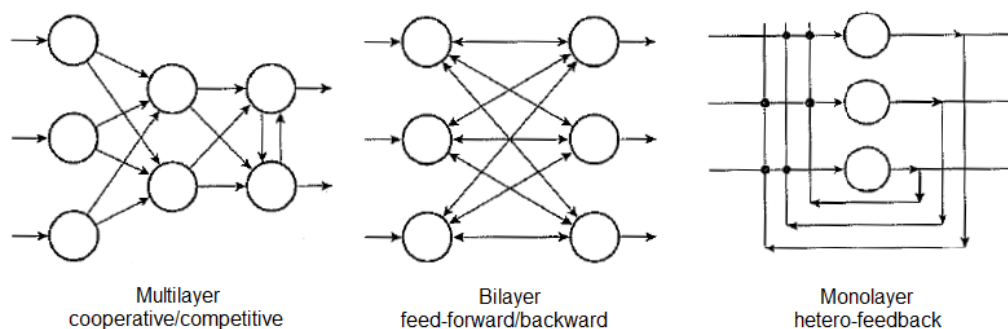


Figura 1.14 – Outras topologias aplicadas às redes neurais artificiais (adaptado de Kartalopoulos, 1996).

As redes neurais recorrentes ou *hetero-feedback neural networks* distinguem-se das *feed-forward* por possuírem um *feedback loop*, i.e. o envio de um sinal de output de um determinado nó para os inputs de outros nós da mesma camada. Este tipo de redes possibilita a existência de um ou mais trajetos fechados para a transmissão de sinais no sistema, o que conduz a uma melhoria do desempenho e capacidade de aprendizagem da rede (Haykin, 1999).

Para além da previsão numérica e simulação de processos, nas últimas décadas têm vindo a ser desenvolvidos vários modelos de redes neurais com vista à classificação e categorização de dados. A rede neuronal probabilística ou *probabilistic neural network* (PNN), de camada múltipla, constitui um dos modelos de RNA dirigidos para a classificação de dados. Possui uma primeira camada oculta, com um número de nós igual ao número de casos de treino, e uma segunda camada oculta, sendo o número de nós correspondente ao número de classes. Os múltiplos exemplos de cada classe contida na primeira camada estabelecem uma conexão apenas com o nó correspondente da segunda camada oculta (Kartalopoulos, 1996).

Durante a fase de aprendizagem, as PNN utilizam os padrões de treino para estimar a distribuição de probabilidades das classes. Estes padrões são armazenados na rede de forma a evitar o processo iterativo, o que determina uma velocidade de aprendizagem bastante rápida (Kartalopoulos, 1996). Na Figura 1.15 é apresentada uma rede neuronal probabilística.

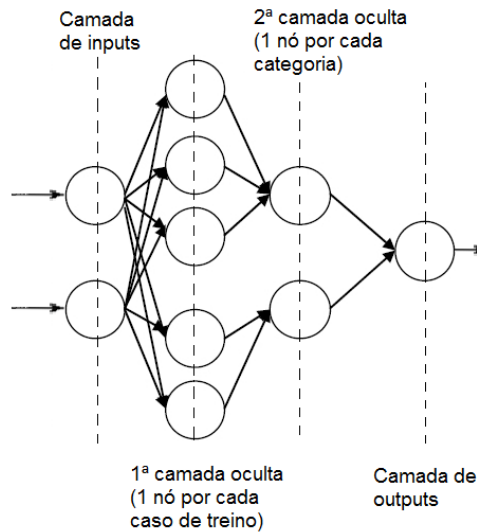


Figura 1.15 – Rede neural probabilística (adaptado de Palisade, 2012).

1.3.4 Desenvolvimento de uma RNA

O perceptron de McColluch e Pitts, divulgado em 1943, permitia modular um único neurónio artificial, mas sem qualquer processo de aprendizagem ou adaptação. A investigação sobre as aplicações deste modelo, e subseqüentes limitações, cedo demonstrou a importância de um processo de aprendizagem para um bom desempenho das RNA (Kartalopoulos, 1996).

O processo de aprendizagem, também denominado treino, consiste no modo como a rede se adapta a um estímulo e produz, eventualmente, a resposta desejada. Durante o treino, os pesos sinápticos das conexões (que contêm a informação ou conhecimento) são ajustados como resposta aos estímulos do input conhecido, para que o output real convirja para a resposta de output desejada. Quando o output real coincide com o pretendido, a fase de treino está completa, o que significa que a rede adquiriu conhecimento (Kartalopoulos, 1996; Strik *et al.*, 2005).

O treino descrito corresponde a um tipo de aprendizagem supervisionada, pois os valores de output desejados são concedidos à rede e atuam como um supervisor, que concebe um sinal de erro que irá ser usado para ajustar os pesos sinápticos. No sentido inverso, a aprendizagem não supervisionada, ou de auto-organização, não requer a presença de um supervisor, organizando padrões de input em categorias (Kartalopoulos, 1996). Considera-se ainda a existência de outros tipos de aprendizagem, como a reforçada, que corresponde a uma forma intermédia entre a supervisionada e a não supervisionada (Elmolla e Chaudhuri, 2011), ou a competitiva, que se distingue da supervisionada pelas características de operação e de arquitetura (Kartalopoulos, 1996).

Ao longo dos anos, a investigação em torno das RNA tem vindo a aumentar a criação de novos modelos de redes neuronais, ou paradigmas, que requerem a criação de variadas regras e equações de aprendizagem. A regra ou equação definida no desenvolvimento de uma RNA denomina-se algoritmo de aprendizagem e descreve matematicamente a forma como os pesos sinápticos evoluem durante o treino (Kartalopoulos, 1996).

O algoritmo de retro propagação do erro, ou *error-back-propagation*, foi o primeiro algoritmo de aprendizagem desenvolvido com sucesso, em 1974 (Kartalopoulos, 1996), e continua a ser um dos mais utilizados métodos de treino de redes (Strik *et al.*, 2005). Este algoritmo pode ser incluído na categoria de aprendizagem supervisionada (Vemuri, 1992) e aplica o método de otimização do gradiente descendente para minimização do erro. A utilização da retro propagação do erro é particularmente comum em RNA do tipo *feed-forward*, mas existem atualmente inúmeros algoritmos alternativos disponíveis para treino de redes (Lingireddy e Brion, 2005).

A fase de treino é fulcral para o desenvolvimento de uma RNA, mas existem ainda outros aspetos a considerar com vista a um bom desempenho da rede. A preparação dos inputs, e as fases de teste da rede, validação de resultados e previsão de dados constituem também importantes etapas a ter em conta.

A preparação dos dados integra a fase inicial do desenvolvimento de uma RNA e requer, geralmente, mais tempo e esforço do que as restantes etapas, incluindo o treino (Lingireddy e Brion, 2005). Apesar de não ser exigido o conhecimento prévio sobre a estrutura das relações existentes entre as variáveis de input e de output (Holubar *et al.*, 2002), as redes neuronais necessitam que os conjuntos de entrada contenham dados suficientes para que as relações encontradas possuam o grau de exatidão desejado. Contudo, o excesso de variáveis tende a aumentar a complexidade das RNA e, por isso, deve-se procurar minimizar o tamanho das redes, mas mantendo um desempenho satisfatório. Como tal, deve existir um conhecimento prévio sobre o processo que será modulado, para que sejam selecionados apenas os inputs com maior relevância. Ao excluir variáveis de entrada irrelevantes, é mantida a simplicidade do modelo, evita-se a introdução de “ruído” na rede e reduzem-se os custos relacionados com a recolha de dados (Lingireddy e Brion, 2005; Holubar *et al.*, 2002).

A fase de teste da rede decorre, geralmente, em simultâneo com o treino, e tem como principal objetivo avaliar o desempenho da generalização através de medições estatísticas do erro durante o processo de aprendizagem. O teste pode utilizar dados do conjunto usado no treino (avalia o desempenho da precisão) ou dados que não foram usados no treino (avalia o desempenho da generalização). A proporção entre os conjuntos de treino e de teste é variável, consoante a extensão de dados disponível, mas é usual considerar-se uma proporção entre 3:1 e 5:1. Se os resultados estatísticos obtidos no teste não forem satisfatórios, o treino deve continuar (Lingireddy e Brion, 2005; Palisade, 2012).

Quando o treino termina, pode ocorrer uma verificação dos resultados obtidos no teste para confirmação da generalização, sendo esta fase denominada validação de resultados. Quando a rede está treinada, a capacidade de generalização pode ser usada para prever valores de output. Esta fase corresponde à previsão de dados e permite obter resultados de output desconhecidos a partir da introdução de um conjunto de inputs na rede treinada (Lingireddy e Brion, 2005; Palisade, 2012). É esta capacidade preditiva das RNA, sem o conhecimento prévio sobre a estrutura das relações existentes entre variáveis de input e output, que lhes confere uma distinta vantagem sobre outras ferramentas de modelação tradicionais (Holubar *et al.*, 2002).

2. OBJETIVOS

A utilização de redes neuronais artificiais como ferramenta de apoio à gestão da digestão anaeróbia tem sido apontada por diversos autores como uma solução promissora, não apenas no tratamento de resíduos e de águas residuais, mas também noutros campos da engenharia e das ciências, no geral. No entanto, a introdução das RNA na gestão deste tipo de processos é recente, e a sua aplicação prática continua ainda envolta num certo grau de incerteza quanto às metodologias e potencialidades.

O presente trabalho enquadra-se num estudo mais amplo, desencadeado pelo grupo Águas de Portugal, no qual se pretende testar a aplicabilidade de redes neuronais artificiais como ferramentas de modelação do funcionamento de digestores anaeróbios a operar nas instalações sob a responsabilidade de entidades gestoras pertencentes ao grupo AdP. O caso de estudo será a ETAR do Seixal, pertencente à SIMARSUL, e o instrumento utilizado será o software NeuralTools[®]. Os principais objetivos do trabalho são definidos da seguinte forma:

- ✓ Analisar e identificar as variáveis intervenientes na digestão anaeróbia e selecionar as que exercem uma maior influência na quantidade e qualidade do biogás produzido no digestor;
- ✓ Criar conjuntos de dados de entrada que permitam o desenvolvimento de redes neuronais capazes de modelar de forma adequada o processo de digestão anaeróbia na ETAR do Seixal;
- ✓ Realizar previsões dos valores dos parâmetros de quantidade e, se possível, de qualidade do biogás produzido no digestor, mediante a seleção de variáveis que resultem numa minimização da monitorização do processo;
- ✓ Avaliar o desempenho das redes neuronais artificiais como ferramentas de modelação e de apoio à decisão na gestão dos processos de digestão anaeróbia.

Deste modo, pretende-se dar continuidade ao trabalho realizado previamente no âmbito das redes neuronais e abrir caminho a investigações futuras que abordem a temática da modelação dos processos de digestão anaeróbia.

3. METODOLOGIA

3.1 Descrição sumária do estudo

O presente trabalho visa avaliar algumas possibilidades de otimização da gestão do processo de digestão anaeróbia de lamas de uma estação de tratamento de águas residuais. Serão utilizadas redes neuronais artificiais como instrumento de modelação do desempenho dos processos anaeróbios e da produção de biogás no digestor.

Numa primeira fase, é essencial a realização de uma pesquisa bibliográfica direcionada para a recolha de toda a informação relevante para o estudo. Assim, no Capítulo 1 são enunciados os principais conceitos e noções sobre o funcionamento dos sistemas de digestão anaeróbia e sobre os fundamentos teóricos de modelos de redes neuronais artificiais. É importante averiguar que modelos de gestão existem para estes sistemas e quais as principais noções a reter, no sentido de perceber de que forma o presente trabalho poderá contribuir para a inovação, e também para ajudar a definir os objetivos centrais do estudo, sendo estes enunciados no Capítulo 2.

Como ferramenta base de trabalho será utilizado o NeuralTools[®], um aplicativo informático da Palisade[®] para criação de modelos de redes neuronais artificiais. No Capítulo 3 será feita uma apresentação do NeuralTools[®] e serão explicados os principais procedimentos referentes a cada uma das fases de desenvolvimento de uma RNA com este software: preparação, treino, teste e previsão de dados. Neste capítulo será ainda apresentado o caso de estudo, incluindo a enunciação das principais noções sobre o funcionamento e sobre o esquema de tratamento da ETAR do Seixal.

No Capítulo 4 serão apresentados e discutidos os resultados obtidos em cada uma das fases de desenvolvimento das redes neuronais. A sua análise permitirá averiguar o grau de afastamento dos resultados em relação aos valores expectáveis, e compará-los com os resultados obtidos noutros trabalhos realizados no âmbito desta temática.

No Capítulo 5 serão apresentadas as principais conclusões referentes ao estudo, determinando se os objetivos inicialmente estabelecidos foram cumpridos e quais as perspetivas de generalização e melhoria da ferramenta utilizada no âmbito dos processos de digestão anaeróbia. Serão ainda referidas algumas sugestões e recomendações para trabalhos futuros.

3.2 O caso de estudo

3.2.1 O grupo Águas de Portugal e a SIMARSUL

O grupo Águas de Portugal (AdP) integra o setor empresarial do Estado e incorpora mais de 40 empresas que atuam nos domínios do abastecimento de água, do saneamento de águas residuais e do tratamento e valorização de resíduos sólidos.

No domínio do saneamento de águas residuais, o grupo AdP é responsável pela gestão e exploração de 20 sistemas multimunicipais, que atuam ao nível da drenagem, tratamento e rejeição de águas residuais urbanas e industriais, incluindo a sua reciclagem e reutilização. Os serviços de saneamento de águas residuais permitem abranger 8,22 milhões de habitantes e são dotados de 899 ETAR e de uma rede de drenagem com 6.347 km, o que possibilitou, em 2013, tratar 505,1 milhões de m³ de águas residuais (Figura 3.1) (AdP, 2014).

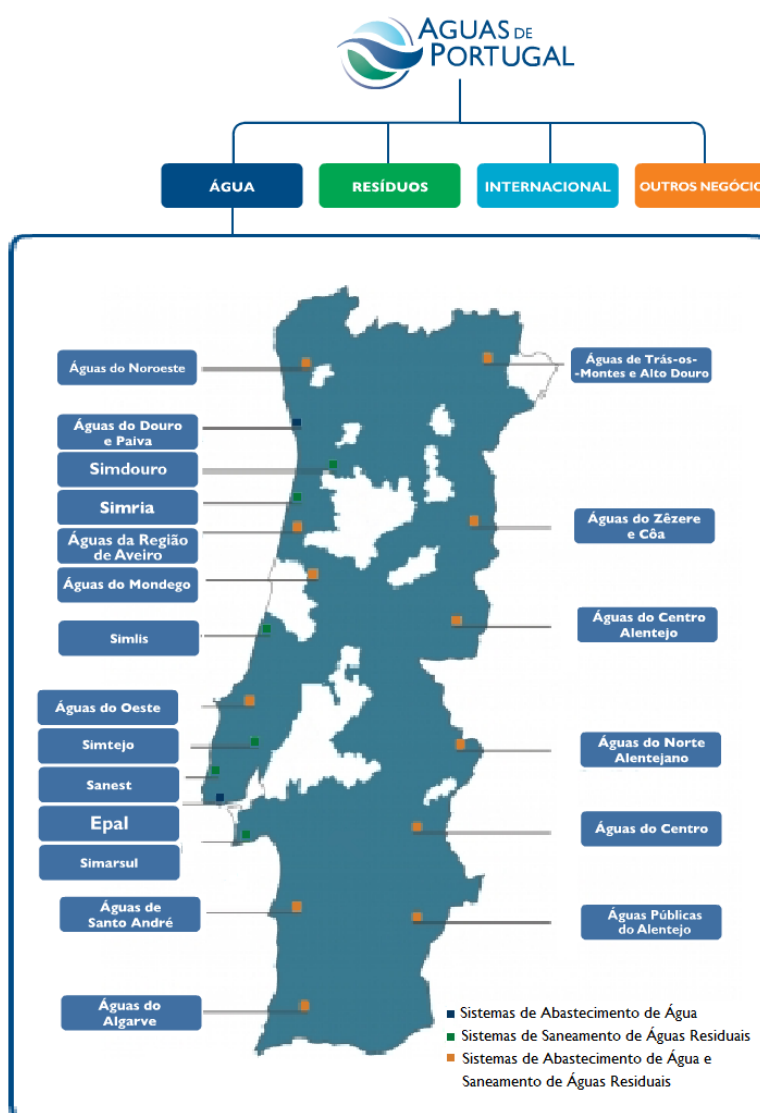


Figura 3.1 – Estrutura organizacional do grupo AdP e território nacional abrangido pelos sistemas multimunicipais de abastecimento de água e saneamento de águas residuais (adaptado de AdP, 2014).

Mais recentemente, o grupo AdP começou também a desenvolver projetos de produção de energia renovável com base nos seus ativos e nos seus recursos endógenos, representando 2,4% do consumo energético de Portugal. As orientações estratégicas no domínio da sustentabilidade incluem a maximização do aproveitamento dos subprodutos dos processos de tratamento de águas e de resíduos, a racionalização dos consumos energéticos e a redução ou compensação das emissões de gases com efeito de estufa.

O aproveitamento do potencial energético do biogás produzido a partir da digestão anaeróbia de resíduos de centrais de valorização orgânica (CVO) e de lamas de ETAR tem contribuído largamente para aumentar a sustentabilidade do setor e da empresa, e é visto como uma solução de futuro, com uma elevada margem de crescimento (Figura 3.2).

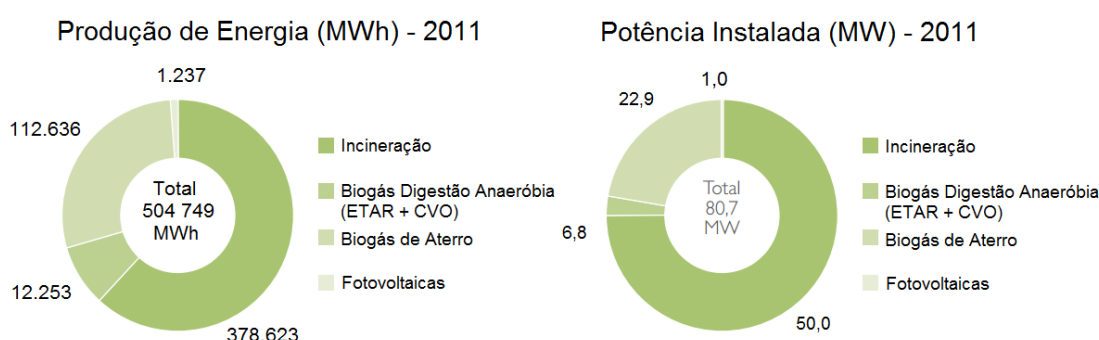


Figura 3.2 – Dados sobre a produção de energia renovável no grupo AdP (AdP, 2014).

Em 2013, o grupo AdP avançou com a realização de um estudo relativo à gestão dos processos de digestão anaeróbia na ETAR Sul, pertencente à SIMRIA. Utilizaram-se redes neuronais artificiais para modelar o desempenho de um digestor, e os resultados obtidos permitiram concluir que a sua utilização poderia contribuir para a otimização da produção de biogás. Neste sentido, o grupo AdP propôs a replicação do estudo em digestores anaeróbios de outras estações de tratamento, pertencentes à SANEST, SIMARSUL, SIMRIA e SIMTEJO.

Assim, o presente trabalho é realizado no âmbito do estudo proposto pela AdP numa ETAR da SIMARSUL. Esta empresa é responsável pelas atividades de recolha, tratamento e rejeição de efluentes em oito municípios da península de Setúbal: Alcochete, Barreiro, Moita, Montijo, Palmela, Seixal, Sesimbra e Setúbal. A SIMARSUL foi criada em 2003, e as infraestruturas de drenagem e tratamento de águas residuais incluem, no presente, 15 ETAR, 61 estações elevatórias e 211 km de sistema de emissários e condutas elevatórias (Figura 3.3). O sistema multimunicipal abrange uma área geográfica de 1.450 km² e em 2011 serviu 741 mil pessoas, o que corresponde a 23,4 milhões de m³ de águas residuais tratadas anualmente (SIMARSUL, 2014).



Figura 3.3 – Território da península de Setúbal abrangido pelas infraestruturas de drenagem e tratamento de efluentes da SIMARSUL (SIMARSUL, 2014).

Para a realização do estudo na SIMARSUL, foram considerados os digestores anaeróbios da ETAR de Sesimbra e da ETAR do Seixal (Figura 3.4). A escolha recaiu sobre a ETAR do Seixal, por estarem disponíveis dados em maior quantidade e em formato digital.



Figura 3.4 – Vista aérea da ETAR do Seixal (SIMARSUL, 2014).

3.2.2 A ETAR do Seixal

Situada no concelho do Seixal, freguesia de Aldeia de Paio Pires, a ETAR do Seixal constitui a segunda maior infraestrutura da SIMARSUL e está inserida no subsistema do Seixal, que inclui ainda sete estações elevatórias e 31 km de emissários e condutas elevatórias. A sua construção teve início em abril de 2007 e a receção provisória por parte da SIMARSUL deu-se em fevereiro de 2012 (FERROVIAL-AGROMAN/CADAGUA, 2007; SIMARSUL, 2014).

A ETAR do Seixal foi projetada para tratar águas residuais de 110 mil habitantes equivalentes no ano de arranque (2012) e de 156 mil habitantes equivalentes em horizonte de projeto (2034). O caudal médio diário no ano horizonte de projeto é de 44.000 m³/dia e o caudal de ponta é de 3.500 m³/h (FERROVIAL-AGROMAN/CADAGUA, 2007; SIMARSUL, 2014).

Caracterização da fase líquida

A linha de tratamento da fase líquida na ETAR do Seixal inicia-se com um pré-tratamento, no qual se processa a remoção de partículas sólidas de maior dimensão, areias e gorduras. O pré-tratamento possui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Dois tamisadores paralelos com autolimpeza (abertura inferior a 3 mm), e um canal de by-pass com grelha de limpeza manual e espaçamento entre grades de 10 mm. Os gradados são encaminhados por um parafuso transportador-compactador sem-fim para um contentor com capacidade para 5 m³.
- ✓ Dois desarenadores/desengorduradores em paralelo. As areias são lavadas e enviadas para um classificador de areias. As gorduras são concentradas e elevadas para o tanque de homogeneização de lamas, passando a integrar a linha de tratamento da fase sólida.

Em seguida, as águas residuais são submetidas a um tratamento primário, para remoção de matéria sólida em suspensão. O tratamento primário possui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Dois decantadores lamelares quadrados (15 x 15 m). As lamas são espessadas nos decantadores e extraídas por gravidade, sendo posteriormente elevadas para o tanque de homogeneização de lamas. O sobrenadante é extraído dos decantadores e enviado para o concentrador de gorduras.

Segue-se o tratamento secundário, no qual se promove a depuração biológica das águas residuais, seguida de um processo de decantação para separação das fases sólida e líquida. O tratamento secundário inclui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Três reatores biológicos em sistema de lamas ativadas com arejamento por ar difuso. Os reatores operam em contínuo e com uma configuração do tipo fluxo pistão. O regime é em média carga e o volume total é de 7.000 m³.

- ✓ Três decantadores secundários circulares com diâmetro de 36 m. Cada órgão está equipado com uma ponte raspadora com raspador de fundo e tubo de sucção para remoção de lamas secundárias. As lamas secundárias são encaminhadas para uma estação elevatória onde são depois enviadas por bombagem para os reatores biológicos, para recirculação externa. As lamas em excesso são enviadas por bombagem para espessamento mecânico. A ponte raspadora possui ainda um raspador de superfície para remoção de escumas, sendo estas encaminhadas para uma tremonha e, posteriormente, enviadas por bombagem para o concentrador de gorduras.

A fase líquida é concluída através de um tratamento terciário, para que o efluente tratado seja submetido a um processo de desinfecção antes da sua descarga no Estuário do Tejo. O tratamento terciário possui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Sistema de desinfecção por radiação UV. As lâmpadas são de baixa pressão e de alta intensidade, e estão orientadas horizontalmente relativamente ao fluxo de água tratada, o qual é posteriormente encaminhado para um emissário de descarga. O sistema de limpeza das lâmpadas é automático.

Na Figura 3.5 é apresentada a disposição dos órgãos e equipamentos que integram o tratamento da fase líquida na ETAR do Seixal.

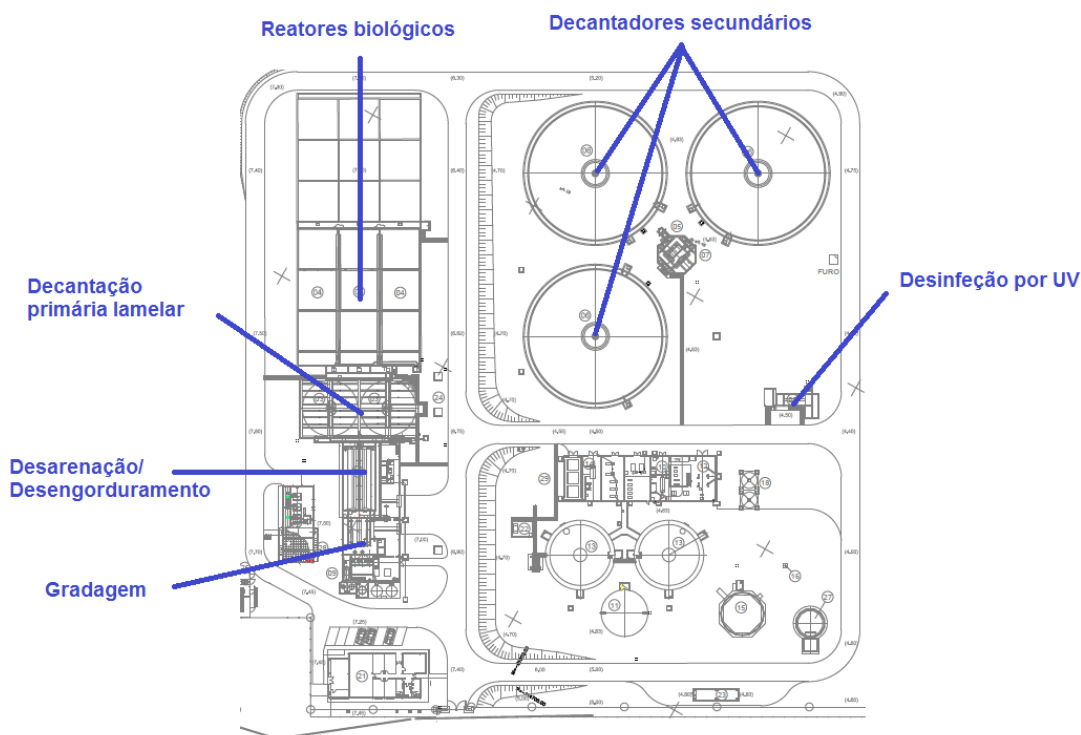


Figura 3.5 – Planta do tratamento da fase líquida da ETAR do Seixal (adaptado de SIMARSUL, 2014).

Caracterização da fase sólida

A linha de tratamento da fase sólida na ETAR do Seixal inicia-se com um processo de espessamento das lamas produzidas no tratamento da fase líquida, com vista à redução do seu volume. O espessamento de lamas é mecânico e inclui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Um espessador mecânico de tambor circular com tela filtrante, que assegura o espessamento de lamas biológicas em excesso e de escorrências. É adicionado polieletrólito às lamas a espessar, para que ocorra a formação de flocos. O líquido separado é enviado para a rede de escorrências e as lamas são encaminhadas para o tanque de homogeneização de lamas.
- ✓ Um tanque de homogeneização de lamas espessadas de 45 m³. O tanque recebe as lamas primárias espessadas nos decantadores lamelares, as lamas biológicas espessadas mecanicamente e as gorduras provenientes do pré-tratamento, e mistura-as através de um agitador de 2,8 kW. O tempo de retenção é de cerca de 4 horas para o caudal máximo. As lamas homogeneizadas são depois alimentadas por bombagem aos digestores anaeróbios durante 12 horas por dia.

Segue-se uma etapa de estabilização das lamas, que ocorre através de um processo de digestão anaeróbia com valorização energética do biogás produzido. A estabilização de lamas inclui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Três (2+1R) bombas de água de doseamento de cloreto férrico. O doseamento de cloreto férrico nas lamas homogeneizadas a montante da digestão anaeróbia permite eliminar a presença de H₂S no biogás produzido.
- ✓ Dois digestores anaeróbios com diâmetro de 18 m e volume unitário de 1.963 m³, concebidos para operar em alta carga na gama mesófila (Figura 3.6). Cada digestor possui duas unidades de aquecimento e agitação “HEATMIX”, que promovem a circulação de água quente numa camisa exterior e injeção de gás em pressão num tubo interior, provocando um efeito *air-lift*. Três (2+1R) permutadores-agitadores por cada digestor asseguram, também, a recirculação das lamas e a manutenção das partículas em suspensão no interior do digestor.



Figura 3.6 – Digestores anaeróbios da ETAR do Seixal (SIMARSUL, 2014).

- ✓ Um depósito de armazenamento das lamas digeridas com agitação. O depósito possui um tempo de retenção de 2 dias, 12 m de diâmetro e volume útil de 718 m³.
- ✓ Um sistema de aproveitamento do biogás produzido na digestão anaeróbia. Este sistema é constituído por três (2+1R) compressores de biogás, um gasómetro de dupla membrana com um volume útil de 970 m³, uma *flare* com capacidade para queimar 459 Nm³/h de biogás em excesso e um sistema de cogeração, constituído por um grupo motogerador GUASCOR FGLD 240/55 de 348 kW de potência, que produz energia elétrica e permite aproveitar a energia térmica para aquecimento dos digestores.

A fase sólida é concluída através de um processo de desidratação, que permite reduzir o volume das lamas antes do seu armazenamento. A desidratação de lamas inclui (FERROVIAL-AGROMAN/CADAGUA, 2007):

- ✓ Duas centrífugas. É adicionado polieletrólito às lamas provenientes do depósito de armazenamento de lamas digeridas, sendo depois elevadas para duas centrífugas. A força centrífuga separa a fase sólida da líquida, ficando os sólidos junto à parede interior. Estes são depois extraídos continuamente através de um parafuso transportador sem-fim, que os conduz para o sistema de elevação.
- ✓ Dois silos de armazenamento de lamas desidratadas com volume unitário de 120 m³.

Na Figura 3.7 é apresentada a disposição dos órgãos e equipamentos que integram o tratamento da fase sólida na ETAR do Seixal.

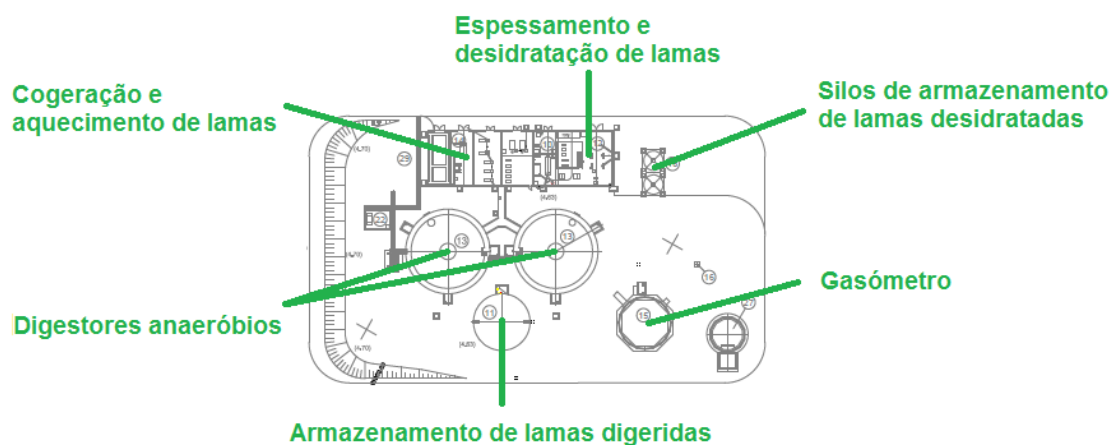


Figura 3.7 – Planta do tratamento da fase sólida da ETAR do Seixal
(adaptado de SIMARSUL, 2014).

Outros aspetos da ETAR

A ETAR do Seixal dispõe também de um sistema de reutilização da água de serviço, composto por filtração, desinfecção, armazenamento e sistema de pressurização, que assegura o abastecimento interno de água em quantidade e qualidade adequadas.

Existe ainda um sistema de desodorização composto por duas torres de lavagem química com diâmetro de 3 m, instaladas no edifício de pré-tratamento. A desodorização permite eliminar odores do edifício de pré-tratamento, depósito de armazenamento de lamas espessadas, edifício de desidratação e silos de armazenamento de lamas desidratadas, mediante lavagem por via húmida. A instalação inclui ainda os ventiladores, os tanques de armazenamento de reagentes e as respetivas bombas doseadoras.

3.3 Desenvolvimento do estudo

Após ter sido escolhida a ETAR para o desenvolvimento do estudo, foi necessário seleccionar a ferramenta a usar no trabalho com os dados. Atualmente, é possível criar modelos de redes neuronais artificiais em inúmeros aplicativos informáticos, tais como o Neural Networks Toolbox[®] para MATLAB[®], o Neural Networks Module[®] da biblioteca do Python[®], o ou o software R[®], entre outros. Apesar de alguns destes aplicativos terem sido utilizados nos casos de estudo de outras empresas do grupo AdP, no presente caso de estudo, a escolha recaiu sobre o NeuralTools[®] da Palisade[®].

3.3.1 Software NeuralTools[®]

O NeuralTools[®] consiste num aplicativo informático desenvolvido pela Palisade[®] para criação de modelos de redes neuronais artificiais. Este software permite modelar determinados conjuntos de dados e é capaz de aprender as suas relações complexas, sendo, por isso, de bastante utilidade na resolução de problemas de classificação e de previsão numérica de dados.

Para o desenvolvimento do estudo, foi instalada a versão 5.5.0 do NeuralTools[®], por apresentar as seguintes vantagens:

- ✓ Utilização fácil – a partir do NeuralTools[®] é possível aceder às funcionalidades das redes neuronais artificiais de forma simplificada;
- ✓ Linguagem acessível – todas as funcionalidades do software estão escritas num inglês simples e claro, facilitando a compreensão do seu modo de funcionamento;
- ✓ Trabalho a partir do Microsoft Excel[®] – o NeuralTools[®] é instalado sob a forma de suplemento do Excel, permitindo trabalhar os conjuntos de dados no ambiente familiar do Microsoft Office[®];
- ✓ Assistência técnica – a versão 5.5.0 do NeuralTools[®] data de 2009 e existem, por isso, inúmeros sites e plataformas online de apoio e assistência ao software, inclusivamente em português;

- ✓ Similaridade com o caso de estudo da ETAR Sul – no estudo da utilização de RNA na gestão de processos de digestão anaeróbia na ETAR Sul da SIMRIA foi também aplicado o NeuralTools[®] como ferramenta de trabalho, o que facilita a comparação dos procedimentos efetuados e dos resultados obtidos com os do presente caso de estudo.

As vantagens apontadas foram determinantes para a seleção do NeuralTools[®] como ferramenta de trabalho por se considerar que, de entre as várias hipóteses disponíveis, este era o software que melhor se ajustava a uma primeira abordagem às funcionalidades das redes neuronais artificiais.

3.3.2 Preparação dos dados

A preparação dos dados constitui a primeira fase do desenvolvimento de uma rede neuronal e muitos autores consideram-na como sendo a mais importante. As relações entre as variáveis de input e as de output, encontradas pelas RNA durante a fase de treino, são fortemente afetadas pela extensão das séries de dados e pela relevância das variáveis de entrada escolhidas. A procura de todos os dados importantes e a consolidação de séries suficientemente extensas e robustas constitui uma tarefa morosa e por vezes complicada, requerendo, usualmente, mais tempo e esforço por parte do utilizador do que as restantes fases de desenvolvimento de redes neuronais (Lingireddy e Brion, 2005; Holubar *et al.*, 2002).

Existem inúmeras variáveis com possível influência na digestão anaeróbia, e a seleção das mais importantes para a modelação dos seus processos constitui uma tarefa complexa, que poderá comprometer todo o estudo se for efetuada de forma errada. Assim, antes da seleção das variáveis e da recolha dos dados, foi necessário o estudo dos processos de digestão anaeróbia e a compreensão do impacto individual de cada variável no desempenho do digestor e na produção de biogás.

Em seguida, procedeu-se à análise dos registos das variáveis monitorizadas na ETAR do Seixal. Depois de terem sido identificadas as mais relevantes para os processos de digestão anaeróbia, foi efetuada a recolha e organização dos dados correspondentes, que compreendeu os seguintes passos:

1. Recolha das séries de dados – os dados relativos à monitorização de parâmetros do tratamento da fase sólida da ETAR do Seixal encontram-se armazenados em folhas de cálculo do Excel. Os parâmetros de qualidade das lamas afluentes e efluentes ao digestor anaeróbio possuem uma frequência de amostragem inferior a uma vez por semana e os valores correspondentes encontram-se já dispostos em séries nas folhas de cálculo, o que facilitou a sua recolha. Por outro lado, os dados sobre as variáveis com frequência de amostragem diária constam de relatórios diários de supervisão individuais, tendo a sua recolha sido, por isso, particularmente morosa.
2. Verificação da fiabilidade dos dados – os dados da monitorização de parâmetros da ETAR do Seixal são introduzidos manualmente nas folhas de registo em Excel e estão,

por essa razão, sujeitos a erros e enganos. Assim, foi verificada a fiabilidade das séries através da comparação dos dados com valores de referência encontrados na literatura. Procedeu-se também à criação de histogramas e de outros gráficos, para controlo da consistência das séries, para averiguação dos intervalos de valores normais de cada variável e ainda para procurar potenciais candidatos a *outliers* (i.e. valores que se afastam muito do padrão geral dos dados, devendo ser removidos das séries para evitar a introdução de “ruído” nas redes). As séries de dados foram ainda mostradas ao engenheiro responsável pela ETAR do Seixal para uma maior elucidação sobre as razões da presença de alguns valores “anormais” ou oscilações nas séries.

3. Cálculo de variáveis – algumas das variáveis escolhidas para modelar os processos de digestão anaeróbia são obtidas de forma indireta, através do cálculo com outros parâmetros monitorizados. É também importante a verificação das unidades de cada variável para que seja facilitado o cálculo e a sua visualização na folha de Excel.
4. Análise de correlações – apesar de serem identificadas apenas as variáveis consideradas relevantes para os processos de digestão anaeróbia e, em particular, para a produção de biogás, por vezes é ainda necessário reduzir a diversidade de inputs com que se trabalha, para facilitar a tomada de decisões metodológicas nas posteriores fases de desenvolvimento das RNA. Assim, foi calculado o coeficiente de correlação de Pearson entre cada variável, para identificar as que possuíam uma elevada dependência com a variável de saída ou entre elas. Desta forma foi possível excluir algumas variáveis de input e simplificar a criação de conjuntos de dados.

Depois de estarem definidos os dados que serão utilizados no desenvolvimento das redes, procede-se ao seu tratamento através do NeuralTools[®]. Para tal, o software disponibiliza o comando *Data Set Manager*, que possibilita a definição e gestão dos conjuntos de dados que serão usados para treinar, testar e prever (Figura 3.8).

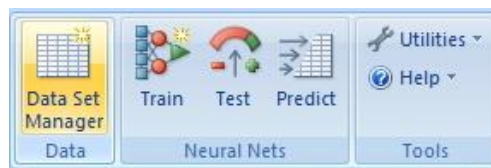


Figura 3.8 – Comando *Data Set Manager* do NeuralTools[®].

Os conjuntos de dados são estruturados a partir de variáveis e de casos. As séries de variáveis estatísticas são arrumadas em colunas contíguas e a primeira célula de cada coluna corresponde ao nome da variável. Cada linha do conjunto de dados corresponde a um caso, e cada caso contém um valor proveniente de cada uma das séries de variáveis de input independentes e um valor proveniente da série de uma variável de output dependente. Se o conjunto de dados for destinado ao treino ou teste de redes, os valores da variável de saída são conhecidos; se o conjunto de dados for destinado à previsão, os valores da variável de saída não são introduzidos.

Na Figura 3.9 é apresentada a janela que surge clicando no botão “Data Set Manager”.

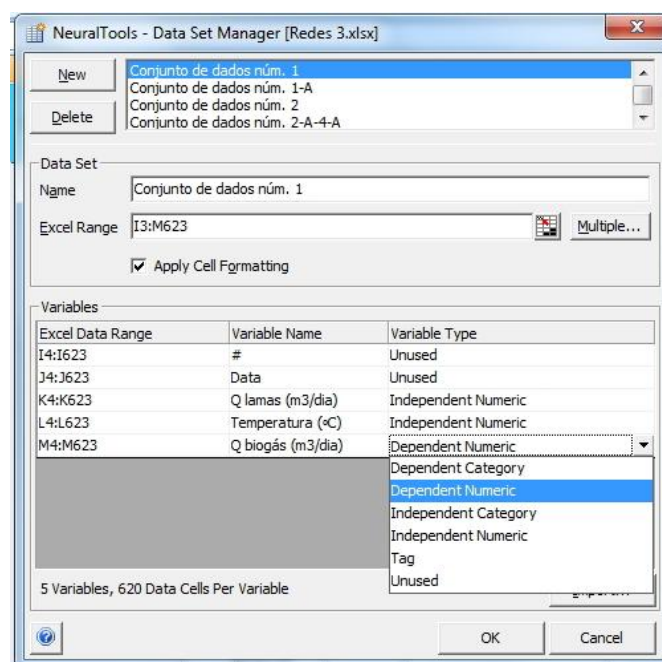


Figura 3.9 – Caixa de diálogo do comando *Data Set Manager* do NeuralTools.

O comando *Data Set Manager* permite realizar as seguintes ações:

- ✓ Criar ou eliminar um conjunto de dados.
- ✓ Mudar o nome de um conjunto de dados.
- ✓ Editar o intervalo do conjunto de dados – clicando no botão “Multiple...”, é possível definir vários intervalos de células num único conjunto de dados; esta opção é particularmente útil para o Microsoft Excel® 2003 ou versões anteriores, uma vez que cada coluna dispõe apenas de 65.536 células e assim é possível expandir o conjunto de dados indefinidamente.
- ✓ Formatar o conjunto de dados com cores e limites distintivos.
- ✓ Definir o tipo de variável – o comando permite distinguir cada variável entre independente (input) ou dependente (output), ou entre numérica ou categórica. As variáveis categóricas são utilizadas na resolução de problemas de classificação (por exemplo: “sim” ou “não”), e por isso não foram consideradas no presente estudo. Cada variável pode ainda ser classificada como “tag” ou “não-usada”: as variáveis “tag” são usadas para identificar os casos de um conjunto de dados que serão usados para treino, teste ou previsão; as variáveis “não usadas” identificam as que não serão utilizadas pela rede. As colunas destinadas à data e ao número do caso costumam ser classificadas como variáveis “não-usadas”.
- ✓ Importar o tipo de variável a partir de outro conjunto de dados ou de uma rede treinada.

Em algumas situações, os conjuntos de dados exibem algumas células das variáveis independentes em branco, com uma mensagem de erro ou com um valor não numérico. Assim, o NeuralTools® disponibiliza o comando *Missing Data Utilities*, que permite substituir este tipo de células (ou qualquer outra específica) por um valor numérico estatístico referente à série (Figura 3.10).

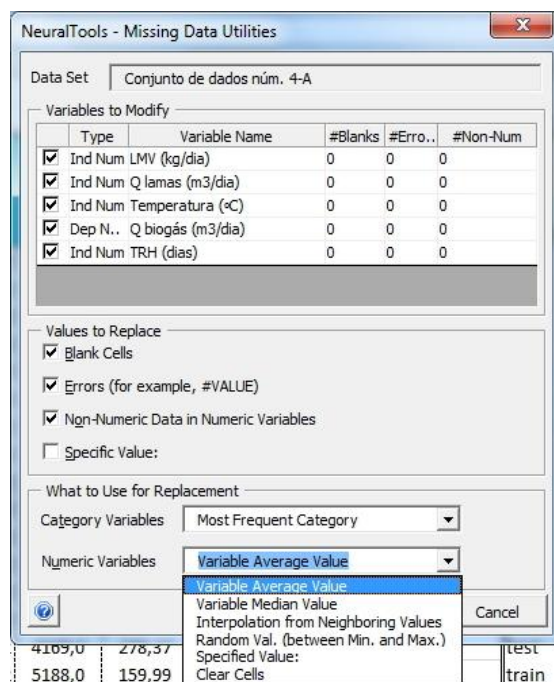


Figura 3.10 – Caixa de diálogo do comando *Missing Data Utilities* do NeuralTools.

Depois de serem devidamente configurados, os conjuntos de dados ficam aptos para as fases de treino, teste e previsão.

3.3.3 Treino e Teste

Depois de ser definido o conjunto de dados, os valores das variáveis são usados para treinar uma rede neuronal. O treino da rede pode ocorrer de forma isolada ou em simultâneo com a fase de teste e de previsão, usando o comando *Train* do NeuralTools® (Figura 3.11).

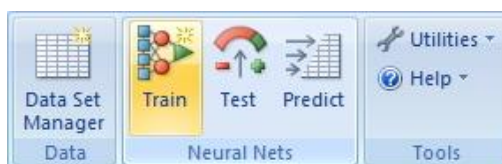


Figura 3.11 – Comando *Train* do NeuralTools.

No presente estudo, optou-se por acoplar as fases de treino e de teste, para que somente as redes com melhores resultados obtidos nestas duas fases fossem utilizadas para previsão de output. Na Figura 3.12 é apresentada a caixa de diálogo que surge clicando no botão “Train”.

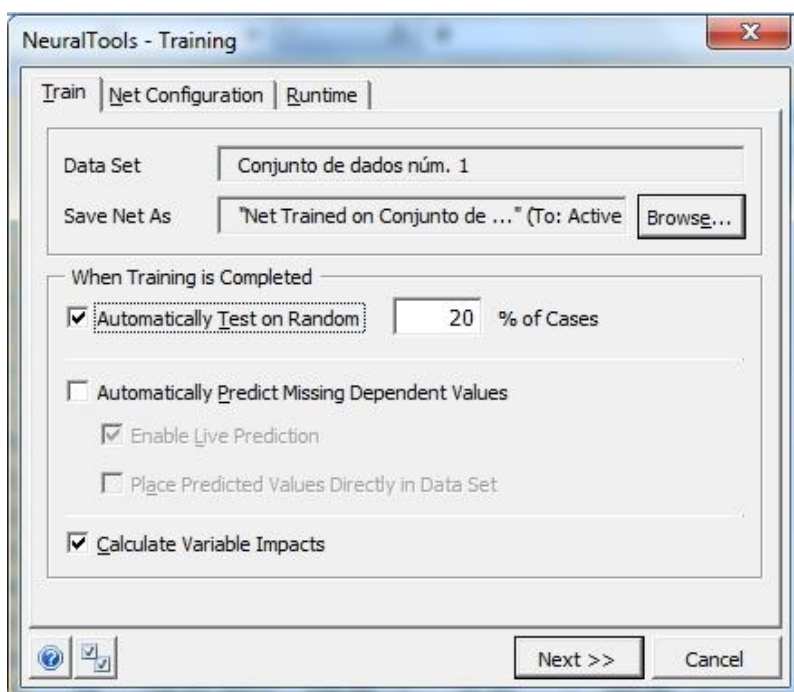


Figura 3.12 – Caixa de diálogo do comando *Train* do NeuralTools.

O separador “Train” da caixa de diálogo permite:

- ✓ Especificar o nome e a localização da rede treinada.
- ✓ Testar automaticamente a rede depois do treino. Neste caso o teste é efetuado a casos aleatórios do conjunto de dados, podendo a percentagem ser definida pelo utilizador. No presente estudo optou-se por testar de forma aleatória 20% dos casos do conjunto de dados, mas em situações em que as séries de dados disponíveis são reduzidas, será preferível diminuir o número de casos de teste para que estes possam ser destacados para o treino. A outra forma de testar a rede seria através da definição do tipo de variável como sendo “Tag”, no comando *Data Set Manager*.
- ✓ Prever automaticamente os valores dependentes das variáveis. Se os valores das variáveis dependentes do conjunto de dados não forem conhecidos, o NeuralTools® permite a sua previsão de forma automática, podendo os valores previstos ser colocados ou não diretamente no conjunto de dados. É ainda possível prever os valores através da definição do tipo de variável como sendo “Tag”, no comando *Data Set Manager*.

- ✓ Calcular o impacto das variáveis. Esta função permite medir a sensibilidade das previsões da rede em relação às alterações nas variáveis independentes, apenas para os dados usados no treino. É atribuído a cada variável independente um valor em percentagem, o qual é tanto menor quanto menor for o impacto da variável nas previsões. Os resultados de uma análise ao impacto relativo das variáveis independentes poderão ajudar na seleção de um novo conjunto de variáveis independentes, excluindo, por exemplo, a que revelou o menor impacto relativo. Contudo, se uma variável revelar um impacto de 99% e outra um impacto de 1%, a exclusão desta última poderá comprometer a precisão das previsões, mesmo considerando que a primeira variável é mais importante. É também importante referir que os resultados do impacto das variáveis são respeitantes apenas à rede treinada, o que não exclui a hipótese de uma variável que tenha pouco impacto numa rede contribuir de forma significativa para previsões precisas noutra tipo de rede. Assim, os resultados do impacto das variáveis independentes devem ser usados apenas para procurar o melhor conjunto de variáveis e não para suportar conclusões firmes sobre a sua relevância na produção de biogás.

Na Figura 3.13 é apresentado o separador que surge clicando em “Net Configuration”, na caixa de diálogo do comando *Train*.



Figura 3.13 – Separador *Net Configuration* da caixa de diálogo do comando *Train* do NeuralTools.

O separador “Net Configuration” permite escolher a rede neuronal que será usada para treinar o conjunto de dados, podendo ser do tipo PNN/GRNN ou MLFN.

As redes neuronais probabilísticas (PNN) são usadas para problemas de classificação ou previsão de categorias das variáveis dependentes. Se a variável dependente for numérica, será treinada uma rede neuronal de regressão generalizada ou *generalized regression neural*

network (GRNN), que opera de forma semelhante às PNN. Este tipo de redes não requer nenhuma configuração, possui um tempo de treino muito reduzido e permite obter uma precisão nas previsões comparável à das MLFN (Palisade, 2012).

As redes *multilayer feed-forward* (MLFN) podem ser usadas em conjuntos de dados em que a variável dependente é categórica ou numérica. Este tipo de redes permite previsões rápidas, é capaz de fazer generalizações a partir de conjuntos de treino muito reduzidos e é mais fiável em situações em que o valor de uma variável independente sai do intervalo do conjunto de treino (Palisade, 2012). Para as MLFN, o NeuralTools® permite definir uma ou duas camadas de nós ocultas e atribuir entre 2 a 100 nós por camada.

As redes possuem inúmeras combinações de configurações possíveis e, por vezes, não existe uma indicação prévia sobre a configuração que irá produzir os melhores resultados de previsão de variáveis. Assim, o separador “Net Configuration” disponibiliza ainda o comando *Best Net Search*, que permite identificar a rede que obteve os melhores resultados na previsão das variáveis dependentes. Quando a procura da melhor rede está ativa, é obrigatória a atribuição de uma percentagem de casos do conjunto de dados para teste, sendo a seleção da melhor rede efetuada automaticamente pelo software a partir da comparação do erro obtido nestes casos.

A “Net Configuration” possibilita ainda a realização de uma regressão linear com o conjunto de dados, para comparação posterior com os resultados das redes treinadas. Contudo, esta função apenas está disponível para previsão numérica e a sua ativação inibe o cálculo do impacto das variáveis.

Na Figura 3.14 é apresentado o terceiro separador da caixa de diálogo do comando *Train*.

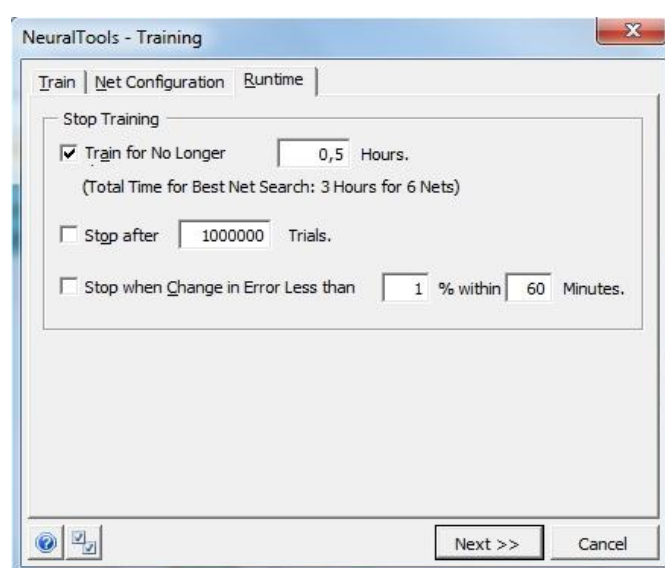


Figura 3.14– Separador *Runtime* da caixa de diálogo do comando *Train* do NeuralTools.

Neste separador é possível selecionar as condições de paragem do treino. O treino de uma rede é parado automaticamente quando o algoritmo de treino “estabiliza”, mas é difícil de definir previamente a duração desta fase. Assim, é preferível a configurar a paragem do treino segundo os seguintes critérios:

1. Tempo de treino de cada rede.
2. Número de ensaios realizados por rede.
3. Progresso mínimo admitido num determinado período. Ativando esta função, o utilizador pode parar o treino se uma determinada percentagem do erro não for alterada num certo intervalo de tempo.

Estes três critérios de paragem podem ser usados individualmente ou em conjunto. No entanto, se a função *Best Net Search* estiver ativada, é obrigatória a seleção do tempo de treino como critério de paragem.

Depois de efetuadas todas as configurações na caixa de diálogo do comando *Train*, clica-se no botão “Next>>”, surgindo a seguinte janela apresentada na Figura 3.15.

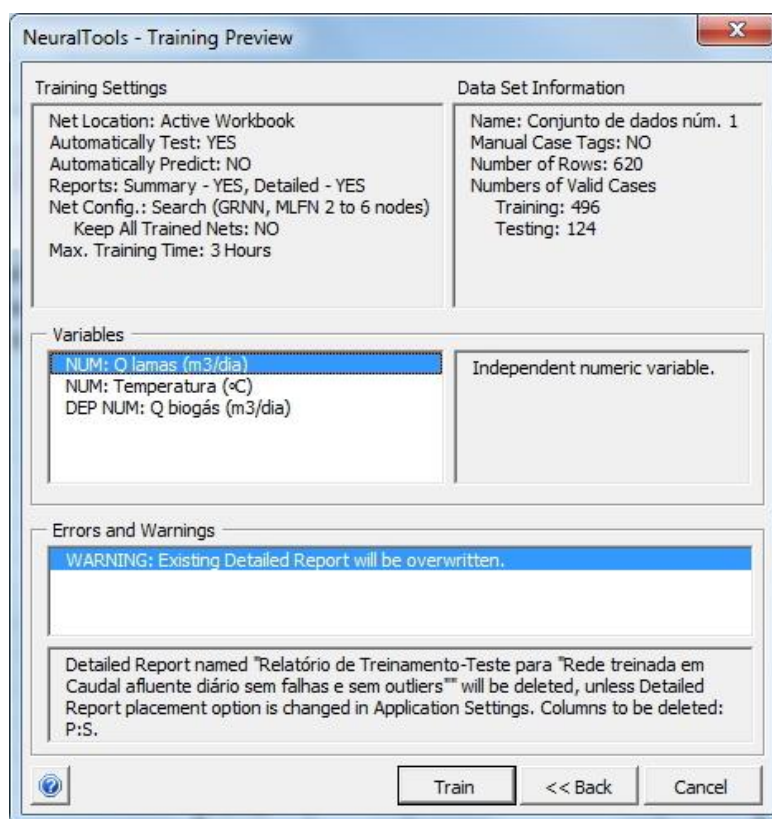


Figura 3.15 – Janela de pré-visualização do treino no NeuralTools.

Esta janela permite visualizar algumas definições gerais sobre o treino, o conjunto de dados e as variáveis, e alerta ainda para a existência de erros ou avisos relativos à rede que irá ser treinada, para possibilitar a correção ou reconfiguração antes do treino. Clicando no botão “Train” nesta janela, é dado início ao treino da rede, durante o qual surge a janela de progresso do treino apresentada na Figura 3.16.

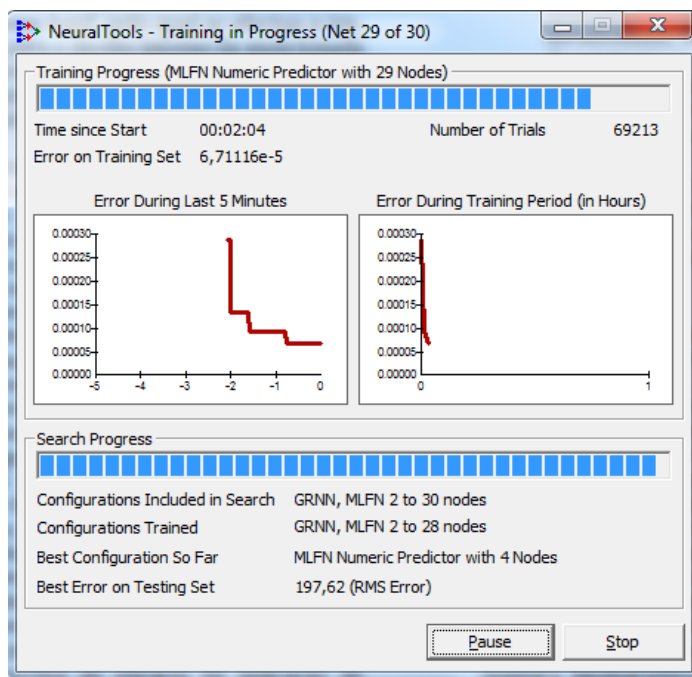


Figura 3.16 – Janela de visualização do progresso do treino no NeuralTools.

A janela do progresso do treino informa em tempo real sobre o estado do treino da rede e sobre a redução do erro. Nesta janela é possível visualizar:

- ✓ A duração total do treino.
- ✓ O número de ensaios realizados.
- ✓ O erro no treino durante os últimos 5 minutos e durante a última hora.
- ✓ Quais as configurações de redes que já foram treinadas de entre o total.
- ✓ Qual a melhor configuração de rede até ao momento.
- ✓ A melhor medida de erro obtida no conjunto de teste até ao momento.

A qualquer momento é possível pausar ou parar o treino da rede atual ou de todas as configurações de rede previstas. Depois do treino ser parado, através do utilizador, da estabilização do algoritmo ou dos critérios de paragem, é emitido um relatório – sumário ou detalhado - sobre o desempenho da configuração da melhor rede encontrada ou de todas as redes, caso tenha sido ativada a função correspondente no separador “Net Configuration” da

caixa de diálogo do comando *Train*. No Apêndice I é possível consultar um exemplo de um relatório detalhado de treino e de teste.

No presente estudo, o treino foi efetuado em simultâneo com o teste, as variáveis dependentes eram numéricas e a função de cálculo do impacto das variáveis foi ativada. Assim, os relatórios de treino e de teste emitidos ofereceram as seguintes informações estatísticas:

- ✓ Percentagem de más previsões – se o valor previsto não estiver contido dentro das margens definidas por uma tolerância em relação ao valor real, é considerada uma “má previsão”; por definição, a tolerância é de 30%, mas este valor pode ser alterado nas definições do NeuralTools®.
- ✓ Erro médio quadrático (RMSE) – este parâmetro mede o desvio das previsões em relação ao valor real e é definido da seguinte forma:

$$\text{Erro médio quadrático} = \sqrt{\frac{\sum_{i=1}^n (\text{valor previsto}_i - \text{valor real}_i)^2}{n}} \quad (\text{Equação 3.1})$$

- ✓ Erro médio absoluto – média do desvio das previsões em relação ao valor real.
- ✓ Desvio padrão do erro absoluto – desvio padrão das previsões em relação ao valor real.
- ✓ Impacto das variáveis em relação aos valores previstos.
- ✓ Histogramas das diferenças residuais – as diferenças residuais correspondem à diferença entre o valor real e o valor previsto.
- ✓ Gráficos de dispersão ou *scatter plots* – este tipo de gráficos mostra as relações existentes entre os valores reais, os valores previstos e as diferenças residuais.

As informações estatísticas contidas nos relatórios de treino e de teste serviram de base para a análise do desempenho das redes no treino, e ajudaram a selecionar as configurações de rede e os conjuntos de variáveis independentes que foram usados na previsão das variáveis dependentes.

3.3.4 Previsão

A última fase de desenvolvimento de uma rede neuronal consiste na previsão dos valores desconhecidos das variáveis dependentes dos conjuntos de dados, utilizando o comando *Predict* (Figura 3.17).

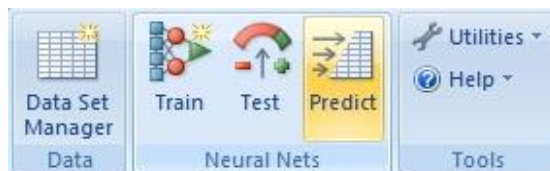


Figura 3.17 – Comando *Predict* do NeuralTools.

Na Figura 3.18 é apresentada a caixa de diálogo que surge clicando no botão “Predict”.

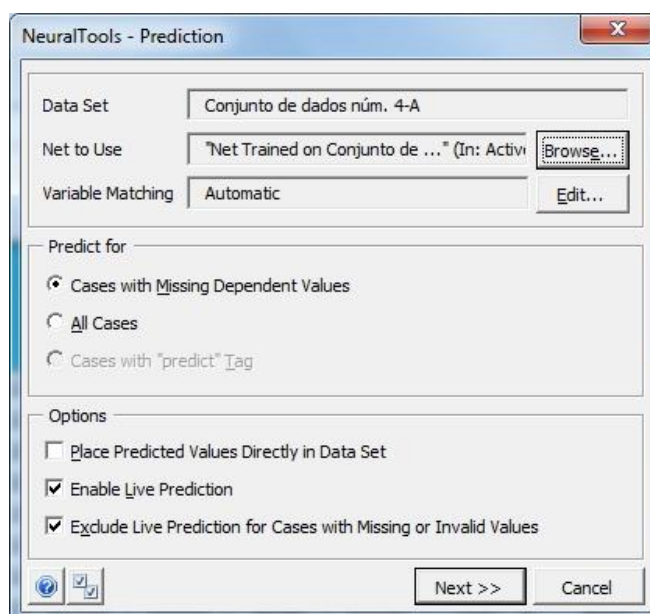


Figura 3.18 – Caixa de diálogo do comando *Predict* do NeuralTools.

As opções presentes nesta caixa de diálogo incluem:

- ✓ Escolha da rede treinada que será usada na previsão, especificando o seu nome e localização.
- ✓ Correspondência entre variáveis – esta opção permite especificar a correspondência entre as variáveis do conjunto de dados que será usado na previsão e as variáveis do conjunto de dados de treino da rede.
- ✓ Escolha dos casos usados na previsão – é possível selecionar os casos em que os valores das variáveis dependentes não são conhecidos, os casos marcados com variáveis “Tag” ou todos os casos.

- ✓ Introdução direta dos valores previstos no conjunto de dados.
- ✓ *Live Prediction* – é possível associar fórmulas às células onde o valor previsto da variável dependente é introduzido pelo NeuralTools®. Esta opção é muito importante, uma vez que permite alterar os valores das células das variáveis independentes e calcular instantaneamente o valor da variável dependente correspondente.

Na Figura 3.19 é apresentada a janela que surge clicando no botão “Next>”, na caixa de diálogo do comando *Predict*.

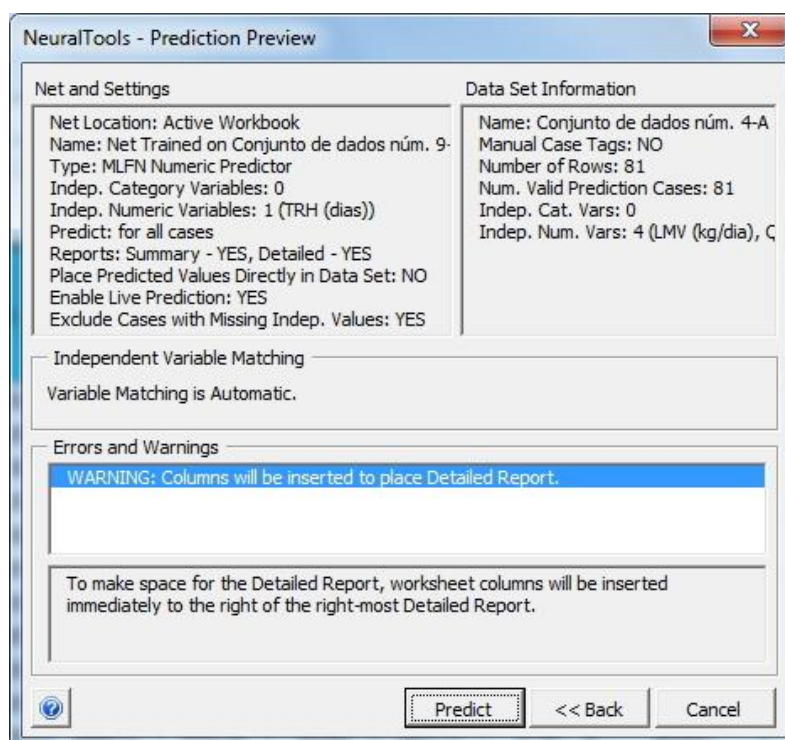


Figura 3.19 – Janela de pré-visualização da previsão no NeuralTools.

Nesta janela é possível visualizar informações sobre algumas definições gerais da rede, da previsão e do conjunto de dados. São também exibidos alertas para a existência de erros ou avisos relativos aos dados que serão previstos, para possibilitar a correção ou reconfiguração antes da previsão. Clicando no botão “Predict” nesta janela, é realizada a previsão dos valores, que serão posteriormente introduzidos no conjunto de dados ou num relatório detalhado, dependendo das definições escolhidas pelo utilizador.

4. APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS OBTIDOS

Ao longo das várias etapas do estudo foram sendo obtidos resultados que contribuíram para a definição da direção a tomar nas etapas posteriores. Neste capítulo serão apresentados todos os resultados relevantes para o desenvolvimento do estudo, incluindo valores numéricos, dados estatísticos ou outro tipo de informação pertinente sobre as variáveis e sobre as redes neuronais. Será ainda efetuada uma análise sobre a influência destes resultados no desenvolvimento do estudo e sobre o seu impacto no cumprimento dos objetivos propostos para o trabalho.

4.1 Preparação dos dados

As séries de variáveis destinadas à criação de redes neuronais foram recolhidas a partir do histórico de monitorização da qualidade e dos relatórios diários de supervisão da digestão anaeróbia da ETAR do Seixal (Quadro 4.1).

Quadro 4.1 – Variáveis selecionadas a partir do histórico de monitorização da qualidade e dos relatórios diários de supervisão da ETAR do Seixal.

Nome da variável	Unidades	Frequência da amostragem	Forma de obtenção
Q lamas	m ³ /dia	diária	medição
pH in	[-]	< 1/semana	análise lab.
CQO in	mg/L	< 1/semana	análise lab.
LCQO in	kg/dia	< 1/semana	cálculo
MS in	mg/L	< 1/semana	análise lab.
MV in	mg/L	< 1/semana	análise lab.
LMS in	kg/dia	< 1/semana	cálculo
LMV in	kg/dia	< 1/semana	cálculo
CV	kg/(m ³ .dia)	< 1/semana	cálculo
MS reator	mg/L	< 1/semana	análise lab.
MV reator	mg/L	< 1/semana	análise lab.
LMS out	kg/dia	< 1/semana	cálculo
LMV out	kg/dia	< 1/semana	cálculo
LMV removida	kg/dia	< 1/semana	cálculo
pH reator	[-]	< 1/semana	análise lab.
CQO reator	mg/L	< 1/semana	análise lab.
LCQO out	kg/dia	< 1/semana	cálculo
LCQO removida	kg/dia	< 1/semana	cálculo
TAL reator	mg/L	< 1/semana	análise lab.
AGV reator	mg/L	< 1/semana	análise lab.
T reator	°C	diária	medição
T _{RH}	dias	diária	cálculo
Q biogás	m ³ /dia	diária	medição

Os locais onde os parâmetros foram medidos ou de onde foram retiradas as amostras enviadas para análise laboratorial podem ser consultados no Apêndice II. As restantes variáveis foram calculadas a partir dos parâmetros de monitorização da digestão anaeróbia.

É também importante mencionar que a preparação dos dados, descrita no presente capítulo, recai apenas sobre as séries de valores utilizados no treino e no teste de redes neuronais. Devido à escassez de dados de algumas variáveis, pretendeu-se estender ao máximo as séries usadas para a previsão de valores e, como tal, não foi possível efetuar a sua caracterização em tempo útil.

4.1.1 Descrição das variáveis

Para uma melhor compreensão das variáveis selecionadas para descrição dos processos anaeróbios no digester, considerou-se importante a realização de uma caracterização detalhada dos respetivos valores. Em seguida, serão apresentadas as caracterizações efetuadas às variáveis que integram os relatórios diários de supervisão. No Apêndice III é possível consultar a descrição das variáveis que integram o histórico de monitorização de qualidade.

Temperatura

A temperatura (T reator) é medida diretamente no digester. Esta variável integra os relatórios diários de supervisão e por essa razão estão disponíveis os valores medidos diariamente de 31 de janeiro de 2012 até 14 de maio de 2014 (Figura 4.1)

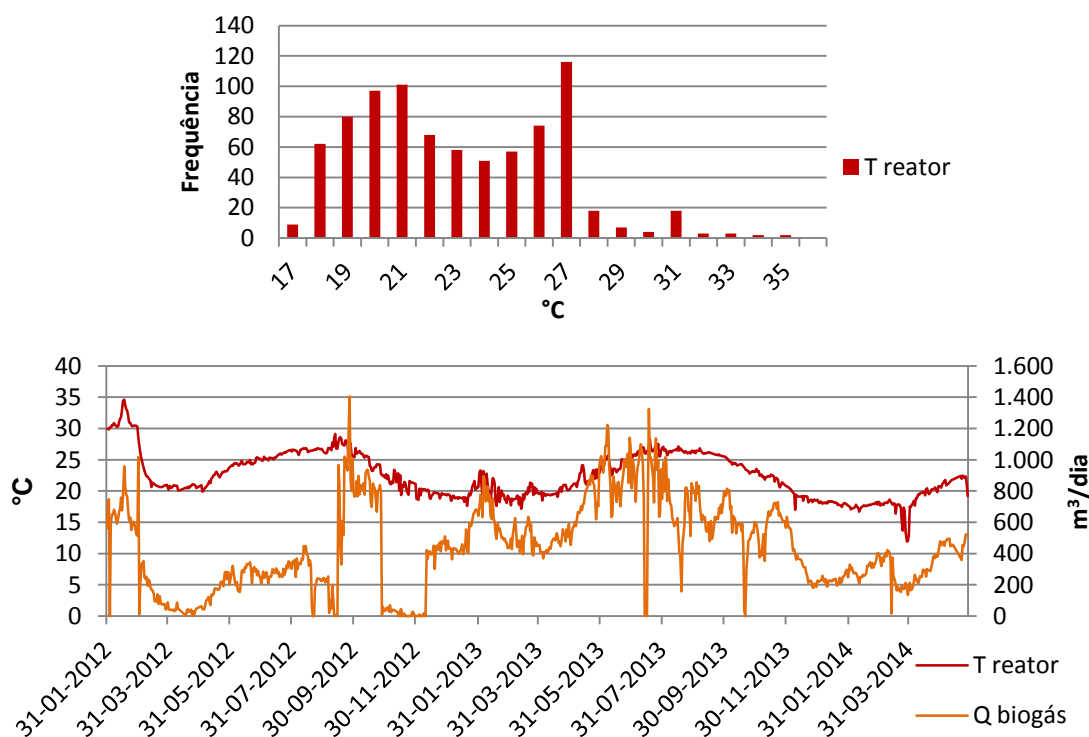


Figura 4.1 – Histograma das temperaturas medidas no digester entre 2012 e 2014 (em cima) e série temporal dos valores desta variável e do caudal de biogás produzido (em baixo).

Apesar de o digestor ter sido concebido para operar em alta carga na gama mesófila (30-38 °C), observa-se na Figura 4.1 que o aquecimento das lamas no digestor ocorreu apenas nos dois primeiros meses de 2012, verificando-se uma oscilação sazonal entre os 17 °C e os 29 °C no restante período. Este facto é devido à presença excessiva de sulfídrico no biogás produzido no processo, o que impediu a sua aplicação no motor de cogeração, responsável pelo aquecimento do digestor.

É também possível verificar que a série temporal das temperaturas no digestor acompanha, de certo modo, a tendência de evolução da série dos caudais de biogás produzido. Este facto permite realçar o papel que a temperatura desempenha nos processos anaeróbios e, apesar de não ser completamente evidente devido às oscilações dos valores de biogás produzido, poderá constituir um indicador da sua importância nos ensaios de previsão com as redes neuronais.

Caudal de lamas (Q lamas)

A variável “Q lamas” mede o caudal diário (m^3/dia) das lamas homogeneizadas que são elevadas para o digestor anaeróbio. Como o digestor funciona a nível constante, o caudal de lamas medido à entrada do digestor é igual ao caudal de saída. Esta variável integra os relatórios diários de supervisão e por essa razão estão disponíveis para treino/teste os valores medidos diariamente de 31 de janeiro de 2012 até 14 de maio de 2014 (Figura 4.2).

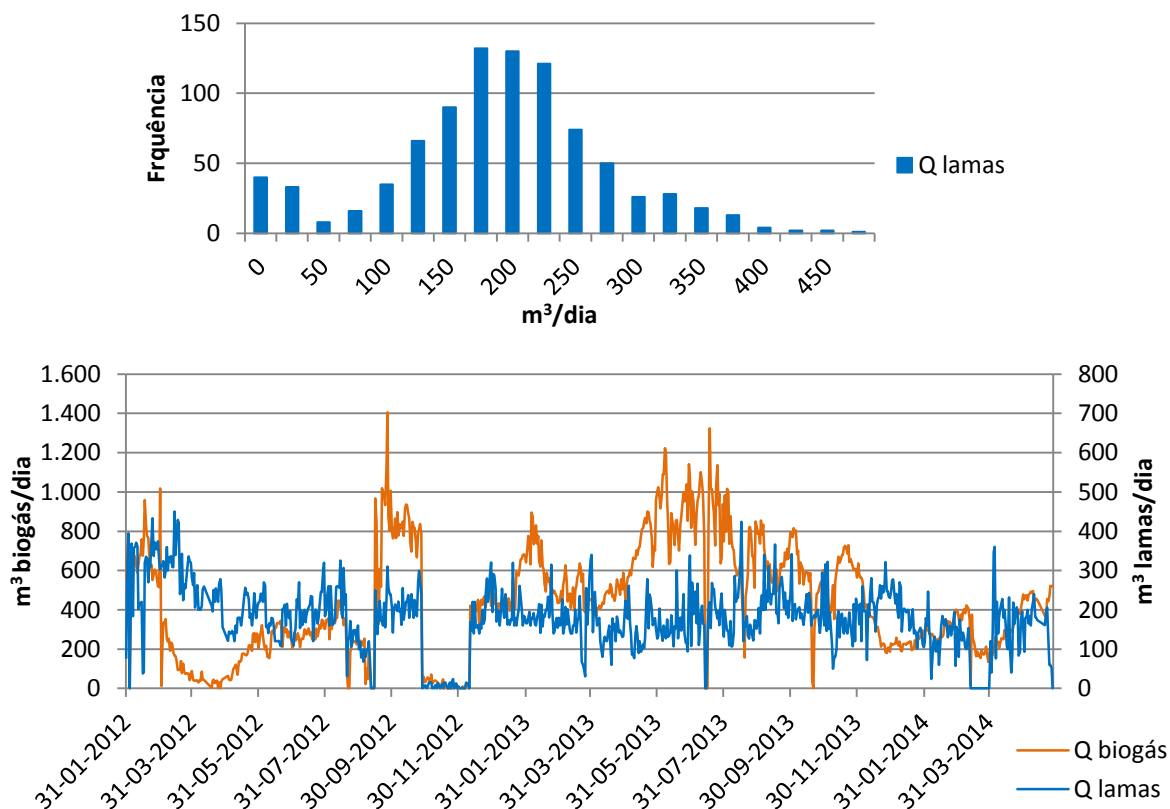


Figura 4.2 – Histograma dos caudais de lamas afluentes ao digestor entre 2012 e 2014 (em cima) e série temporal dos valores desta variável e do caudal de biogás produzido (em baixo).

O caudal de projeto das lamas afluentes à digestão anaeróbia na ETAR do Seixal é de 188 m³/dia no ano zero e de 253 m³/dia no ano horizonte, e na Figura 4.2 é possível verificar que a distribuição dos caudais se centra entre estes dois valores. No entanto, observa-se também a existência de inúmeros valores diários de caudal de lamas nulos ou próximos de zero, referentes aos dias em que não houve alimentação das lamas devido a problemas operacionais. Assim, criou-se uma segunda série com os dados do caudal de lamas à qual foram retirados todos os valores de caudal inferiores a 75 m³/dia e superiores a 375 m³/dia, tendo sido esta a série usada para o cálculo de outras variáveis. Uma análise comparativa entre as duas séries temporais apresentadas permitiu também verificar que os valores do caudal de lamas afluente não acompanham a tendência evolutiva do caudal de biogás produzido, excetuando nos períodos em que ocorreram problemas operacionais no digestor.

Tempo de retenção hidráulico (T_{RH})

O tempo de retenção hidráulico no digestor (T_{RH}) é calculado a partir do volume útil do digestor e do caudal de lamas elevado para digestão, segundo a seguinte equação:

$$T_{RH} = \frac{V_{digestor}}{Q_{lamas}} \quad (\text{Equação 4.1})$$

A dimensão da série de dados do T_{RH} é igual à série do caudal de lamas, estando disponíveis os valores diários de 31 de janeiro de 2012 até 14 de maio de 2014 (Figura 4.3). Apesar de o conceito “valor diário de T_{RH}” não ser inteiramente correto, considerou-se importante a sua inclusão no grupo de variáveis a utilizar na criação de redes neuronais, por conferir uma dimensão diferente ao caudal de lamas.

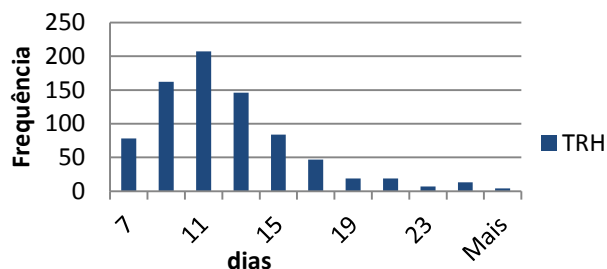


Figura 4.3 – Histograma dos tempos de retenção hidráulicos do digestor, correspondentes ao período entre 2012 e 2014.

Uma vez que foi utilizada a série de dados do caudal de lamas sem os caudais inferiores a 75 m³/dia e superiores a 375 m³/dia, os valores de T_{RH} calculados variam apenas entre 5,2 dias e 26,2 dias. A figura anterior corresponde a um histograma dos valores diários do T_{RH} e é possível verificar a existência de vários fora do intervalo aconselhado para digestão anaeróbia em alta carga, de 10 a 20 dias (Russell, 2006). Este facto é devido, essencialmente, aos elevados valores do caudal de lamas verificados nos primeiros meses de 2012, que correspondem ao único período durante o qual o digestor foi aquecido.

Caudal de biogás produzido (Q biogás)

Os parâmetros de quantidade e qualidade de biogás produzido constituem excelentes indicadores do desempenho de um digestor anaeróbio, e um dos principais objetivos do presente estudo consiste na modelação do processo de digestão para previsão dos valores destes parâmetros. No entanto, os registos da ETAR do Seixal possuem apenas dados relativos aos parâmetros de quantidade de biogás produzido e à sua aplicação na produção de energia na estação. Como tal, considerou-se apenas o caudal de biogás produzido (Q biogás), em m^3/dia , para modelar, em conjunto com as variáveis referidas anteriormente, o processo de digestão anaeróbia.

Esta variável integra os relatórios diários de supervisão e por essa razão estão disponíveis os valores medidos diariamente de 31 de janeiro de 2012 até 14 de maio de 2014. Como o caudal de biogás produzido irá constituir a variável dependente dos ensaios a realizar no presente estudo, é importante que a série correspondente seja extensa de forma a cobrir o período das séries das variáveis independentes (Figura 4.4).

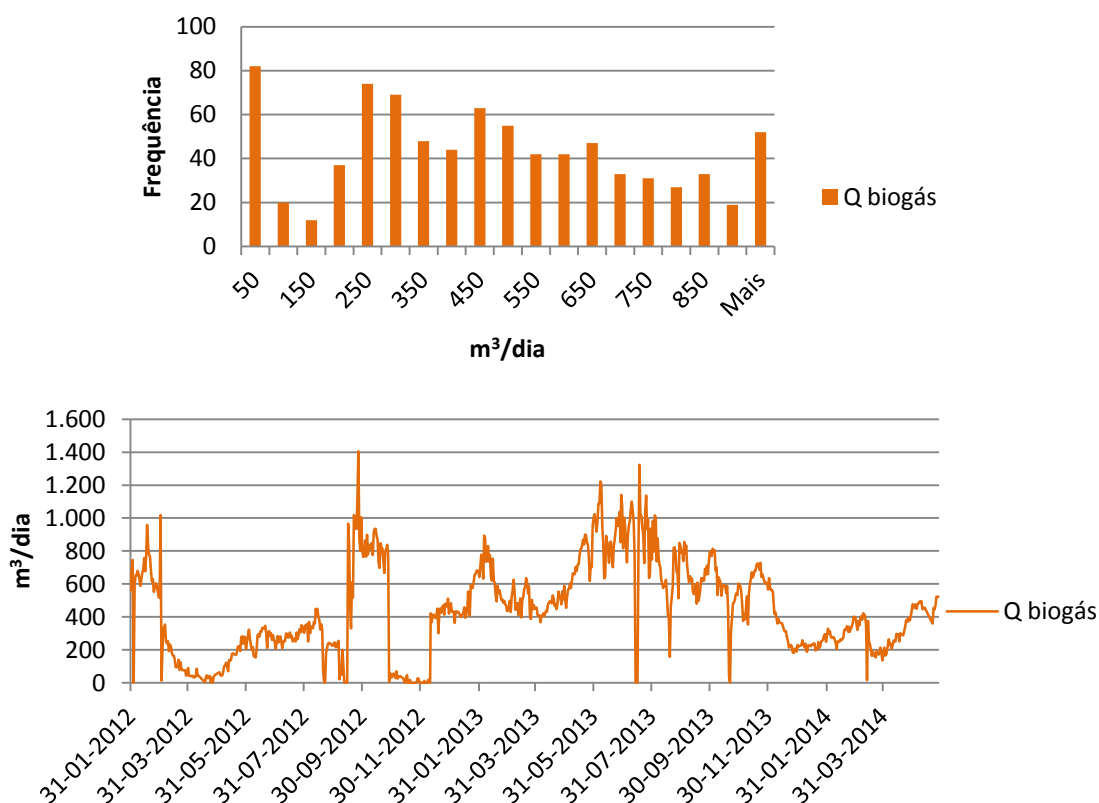


Figura 4.4 – Histograma das medições do caudal de biogás afluente ao digestor entre 2012 e 2014 (em cima) e série temporal dos valores desta variável (em baixo).

É possível verificar na Figura 4.4 que o histograma desta variável não segue uma distribuição normal. Este facto ajuda a compreender, em conjunto com as oscilações observadas na série temporal, a grande variabilidade de caudais de biogás produzido e a importância da existência de uma ferramenta de modelação que permita prever os seus

valores. É de notar também a existência de mais de 80 medições de caudais próximos de zero, nomeadamente entre outubro e dezembro de 2012, resultantes de problemas operacionais.

4.1.2 Parâmetros de funcionamento do digestor

Para o acompanhamento do desempenho da digestão anaeróbia, existem vários parâmetros que podem ser tidos em consideração. No Quadro 4.2 é feito um resumo dos parâmetros de funcionamento - medidos e calculados - do digestor anaeróbio da ETAR do Seixal e das correspondentes médias aritméticas.

Quadro 4.2 – Médias aritméticas dos valores dos parâmetros de funcionamento da digestão anaeróbia na ETAR do Seixal.

Parâmetro	Média aritmética	Unidades	Valores de referência / Fonte
Q lamas	177	m ³ /dia	188 a 253 (FERROVIAL-AGROMAN / CADAGUA, 2007)
pH in	6,04	[-]	5,0 a 8,0 (Vesilind, 2003)
CQO in	36.115	mg/L	-
LCQO in	6.668	kg/dia	-
MS in	2,6	%	4,0 a 7,0 (Vesilind, 2003)
MS in	24.876	mg/L	-
MV in	16.969	mg/L	-
LMS in	4.837	kg/dia	-
LMV in	3.319	kg/dia	-
LCQO in/LMV in	2,0	[-]	1,3 a 1,7 (Henze <i>et al.</i> , 2002)
CV	1,71	kg/(m ³ .dia)	1,6 a 3,2 (Vesilind, 2003)
MS reator	21.169	mg/L	-
MV reator	13.063	mg/L	-
LMV out	2.600	kg/dia	-
LMV removida	719	kg/dia	-
% MV removida	21,7	%	40 a 60 (Vesilind, 2003)
pH reator	6,48	[-]	6,2 a 7,8 (Gray, 2004)
CQO reator	29.473	mg/L	-
LCQO reator	5.261	kg/dia	-
LCQO removida	1.407	kg/dia	-
TAL reator	2.204	mg/L	2.000 a 5.000 (Metcalf & Eddy, 2003)
AGV reator	1.331	mg/L	250 a 1.000 (Gray, 2004)
AGV/TAL	0,7	[-]	0,1 a 0,3 (Spellman, 2003)
T reator	22,5	°C	30 a 38 (Metcalf & Eddy, 2003)
T _{RH}	11,2	dias	10 a 20 (Russell, 2006)
Q biogás	486,8	m ³ /dia	-
Q biogás/m ³ digestor	0,25	m ³ /(m ³ .dia)	1,0 a 2,0 (Amaral, 2012)
Q biogás/LMV removida	0,32	m ³ /(kg)	0,75 a 1,12 (Metcalf & Eddy, 2003)
Q biogás/LMV in	0,15	m ³ /(kg)	≈ 0,50 (Gray, 2004)
Q biogás/LMS in	0,10	m ³ /(kg)	≈ 0,38 (Gray, 2004)

Conforme se observa no Quadro 4.2, a carga orgânica volúmica, o pH das lamas afluentes, o pH das lamas digeridas, a alcalinidade total e o tempo de retenção hidráulico apresentam valores, em média, dentro dos intervalos de funcionamento de referência. Contudo, é preciso ter em atenção que estes resultados correspondem apenas a médias aritméticas, tendo sido possível verificar nos histogramas das séries de dados que estas variáveis revelaram alguns valores diferentes dos intervalos aconselháveis.

Segundo Vesilind (2003), a concentração de sólidos totais nas lamas (primárias + secundárias) espessadas varia, usualmente, entre 4,0% e 7,0%, e é possível verificar no Quadro 4.2 que a sua concentração média é de apenas 2,6%. Uma causa provável para o desvio dos valores de concentração das lamas espessadas será a existência de infiltrações de água do estuário do Tejo na rede de drenagem da ETAR do Seixal. O excesso de salinidade nas águas residuais afluentes à estação condiciona a sedimentação de sólidos no tratamento primário e no tratamento secundário, o que conduz a uma concentração de lamas inferior ao desejável para este tipo de processos.

Por outro lado, os parâmetros de produção específica de biogás demonstram que o caudal de biogás produzido tem sido inferior ao esperado para este tipo de digestão anaeróbia, como resultado da reduzida remoção de matéria volátil. A percentagem de matéria volátil removida na digestão anaeróbia varia, usualmente, entre 40% e 60% (Vesilind, 2003), e no digestor anaeróbio da ETAR do Seixal este valor foi, em média, de 22%. No entanto, as médias das concentrações de AGV (1.331 mg/L) e das temperaturas verificadas no digestor (22,5 °C) afastaram-se dos intervalos de referência para o funcionamento em alta carga na gama mesófila, o que poderá também constituir uma das possíveis causas para a reduzida remoção da matéria volátil e conseqüente redução no caudal de biogás produzido.

De um modo geral, os parâmetros de funcionamento apresentaram alguns desvios em relação aos valores desejáveis para uma digestão anaeróbia a operar em alta carga na gama mesófila, deixando patente a instabilidade verificada durante o período considerado no estudo.

4.1.3 Análise de correlações

A partir do histórico de supervisão e monitorização da qualidade da digestão anaeróbia da ETAR do Seixal foram recolhidos os dados de 13 variáveis, tendo sido posteriormente calculadas mais 11. Para a criação de redes neuronais, a utilização de 24 séries de variáveis constitui um número excessivo, na medida em que existem inúmeras combinações possíveis que obrigariam à criação de vários conjuntos de dados. Assim, para facilitar a tomada de decisões metodológicas nas posteriores fases de desenvolvimento do estudo, procedeu-se à análise das correlações entre as séries das 24 variáveis, utilizando, para tal, o coeficiente de correlação de Pearson (ρ).

Este coeficiente determina a existência de uma estrutura linear num diagrama de dispersão de duas variáveis, permitindo identificar as que possuem uma maior dependência com o caudal de biogás produzido e excluir as que apresentem uma forte correlação com

outras variáveis (Pestman, 1998). Deste modo, o coeficiente de correlação de Pearson pode ser definido, para duas variáveis x e y , da seguinte forma (Wilcox, 2001):

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X).var(Y)}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (\text{Equação 4.2})$$

Em que:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (\text{Equação 4.3})$$

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i \quad (\text{Equação 4.4})$$

O sinal de ρ constitui um indicador da direção da correlação, isto é, se a correlação entre duas variáveis é positiva ou negativa, e o seu valor, que varia entre +1 e -1, permite determinar a força da correlação (Wilcox, 2001). Se o módulo de ρ é inferior a 0,30, considera-se que a correlação linear entre as duas variáveis é fraca; entre 0,30 e 0,70, a correlação linear é moderada; se ρ for superior a 0,70, a correlação linear existente é forte. No Quadro 4.3, apresentam-se os valores dos coeficientes de correlação determinados entre todas as variáveis.

Analisando o Quadro 4.3, verifica-se a existência de uma correlação linear forte entre:

- O caudal de lamas e o tempo de retenção hidráulico (-0,900);
- O pH das lamas elevadas para digestão e a concentração de ácidos gordos voláteis nas lamas digeridas (0,725);
- A concentração da matéria seca e da matéria volátil nas lamas elevadas para digestão (0,931);
- A carga da matéria seca e da matéria volátil nas lamas elevadas para digestão (0,931);
- A carga da matéria volátil nas lamas elevadas para digestão e a carga orgânica volúmica (1,000);
- A concentração da matéria seca e da matéria volátil nas lamas digeridas (0,910).

Deste modo, optou-se por excluir dos conjuntos de dados a utilizar nos ensaios com as RNA as variáveis “MS in”, “LMS in”, “MS reator” e “CV”. Mantiveram-se as variáveis “pH in” e “AGV reator” por apresentarem uma correlação próxima de 0,70 e por se tratar de parâmetros medidos à entrada e à saída do digestor, respetivamente. Optou-se ainda por manter as séries do caudal de lamas e do tempo de retenção hidráulico, uma vez que o T_{RH} é determinado de forma não linear a partir da variável “Q lamas”; o elevado valor do coeficiente de correlação entre as duas variáveis deve-se à exclusão dos valores de caudal de lamas inferiores a 75 m³/dia e superiores a 375 m³/dia, os quais originariam valores extremos de T_{RH} que reduziriam o valor do coeficiente de correlação.

Foi ainda identificada a existência de várias correlações moderadas entre variáveis, em particular, entre o “Q biogás” e as restantes. Uma análise destes valores permitiu definir as variáveis às quais deveria ser dada primazia aquando da criação dos conjuntos de dados a utilizar em fases posteriores do estudo. Assim, as variáveis com uma correlação linear moderada com o caudal de biogás produzido são: “Q lamas”, “pH in”, “MV in”, “LMV in”, “LMV removida”, “pH reator”, “AGV reator”, “T reator” e “ T_{RH} ”.

4.2 Treino e Teste

Depois de terem sido preparados os conjuntos de dados destinados à criação de redes neuronais, procedeu-se à fase seguinte do desenvolvimento do estudo, o treino/teste de RNA. No sentido de organizar o trabalho a realizar e de simplificar a tomada de decisões inerentes ao treino e teste de redes neuronais, optou-se por dividir esta fase em várias etapas.

4.2.1 Familiarização com o software NeuralTools[®]

Apesar de o software NeuralTools[®] permitir uma utilização simples e acessível, a aplicação das suas funcionalidades no desenvolvimento de redes neuronais requer ainda um período de adaptação. Assim, a primeira etapa do treino e teste de RNA consistiu na familiarização com o software por parte do utilizador, tendo como principais propósitos o estabelecimento de um primeiro contacto com os comandos e funcionalidades do NeuralTools[®]

no Microsoft Excel[®] e ainda a realização de ensaios de treino e de teste informais para percepção do comportamento das variáveis selecionadas no desempenho das RNA.

Ao longo desta etapa foram realizados 69 ensaios, tendo a maior parte deles visado fornecer alguma elucidação sobre as configurações possíveis no comando *Train* do NeuralTools[®], nomeadamente sobre a definição do tipo de rede, do número de nós por camada e do critério de paragem do treino. A experiência adquirida nesta etapa permitiu depreender que as melhores opções perante as características do presente estudo seriam:

- ✓ A ativação da função *Best Net Search* para a realização de treinos com vários tipos de redes, com vista à procura da que melhor se ajustasse ao conjunto de dados.
- ✓ A atribuição de 2 a 6 nós por camada para as redes do tipo MLFN. Posteriormente, a atribuição do número de nós veio a ser debatida numa reuniões com os responsáveis pelos estudos idênticos que foram realizados em simultâneo noutras empresas do grupo AdP, tendo sido sugerido o alargamento deste intervalo. No seguimento desta reunião, as redes treinadas passaram a ser configuradas com a atribuição de 2 a 30 nós por camada.
- ✓ A utilização do tempo (1h no máximo) como critério de paragem do treino. Para além de ser obrigatório quando a função *Best Net Search* está ativada, este critério de paragem permite ainda ao utilizador flexibilizar o período de treino consoante o tempo disponível.

É também conhecido que o método de procura das relações entre variáveis utilizado pelas redes neuronais não requer o conhecimento prévio da estrutura dessas relações, assentando numa abordagem não linear. Neste sentido, alguns dos ensaios realizados nesta etapa (exemplificados no Quadro 4.4) recaíram sobre conjuntos de dados constituídos por variáveis de input e de output cujas relações eram conhecidas pelo utilizador mas não pelo NeuralTools[®].

Quadro 4.4 – Indicadores de desempenho dos ensaios de treino e de teste das redes “0-1” e “0-3”.

Rede	Variáveis de input		Impacto	Variáveis de output		Resultados	
0-3	MV in	mg/L	56,35%	CV	kg/(m ³ /dia)	Melhor rede	MLFN (21 nós)
	Q lamas	m ³ /dia	43,65%			% más previsões (treino - 30% toler.)	0,00%
						% más previsões (teste - 30% toler.)	0,00%
						Erro médio quadrático (treino)	0,00
						Erro médio quadrático (teste)	0,00
0-1	Q lamas	m ³ /dia	100,00%	TRH	dias	Melhor rede	MLFN (6 nós)
						% más previsões (treino - 30% toler.)	0,00%
						% más previsões (teste - 30% toler.)	0,00%
						Erro médio quadrático (treino)	0,00
						Erro médio quadrático (teste)	0,00

Na análise de correlações entre variáveis, constatou-se que existia uma correlação linear perfeita entre a carga de matéria volátil afluente e a carga orgânica volúmica. O ensaio “0-3” teve como objetivo averiguar se a abordagem não linear do NeuralTools[®] permitia treinar

uma rede de forma a prever a relação linear existente entre as duas variáveis usadas para calcular a carga de matéria volátil afluyente (inputs) e a carga orgânica volúmica (outputs). No Quadro 4.4 é possível verificar que o NeuralTools[®] previu com precisão os valores da variável “CV”, o que demonstra a capacidade deste software em prever relações lineares.

O ensaio “0-1” visou perceber de que forma o NeuralTools[®] reagia à previsão de valores de uma variável determinada de forma não linear apenas a partir de outra. A variável de input utilizada foi o “Q lamas” e a variável de output foi o “T_{RH}”. Os indicadores do desempenho da rede “0-1” constantes do Quadro 4.4 demonstram que o NeuralTools[®] previu na totalidade os valores de output, o que confirma a capacidade deste software de aprender relações não lineares sem o conhecimento prévio da sua estrutura.

A partir da análise de correlações entre variáveis, foi também possível verificar a existência de algumas correlações fortes entre variáveis que não haviam sido calculadas, como a “MS in” e a “MV in”; este facto motivou a realização de um ensaio com o propósito de avaliar a capacidade do NeuralTools[®] para prever os valores de uma variável a partir dos dados de outra. Os resultados obtidos são apresentados no Quadro 4.5.

Quadro 4.5 – Indicadores de desempenho do ensaio de treino e de teste da rede “0-2”.

Rede	Variáveis de input		Impacto	Variáveis de output		Resultados	
0-2	MS in	mg/L	100,00%	MV in	mg/L	Melhor rede	MLFN (8 nós)
						% más previsões (treino - 30% toler.)	6,72%
						% más previsões (teste - 30% toler.)	0,00%
						Erro médio quadrático (treino)	2282,28
						Erro médio quadrático (teste)	1725,14

Na rede “0-2” foram utilizados os valores da variável “MS in” como dados de input e os valores da variável “MV in” como dados de output para o ensaio. Os indicadores de desempenho da rede revelaram 6,72% de más previsões no treino, o que se enquadra no resultado obtido através do coeficiente de correlação de Pearson, isto é, para um ρ de 0,931, tem-se que:

$$(1 - 0,931) \times 100\% = 6,90\% \quad \text{(Equação 4.5)}$$

A rede “0-2” acusou também 0,00% de más previsões e um erro médio quadrático de 1.725,14 no teste. Considerando que a média aritmética da série de output era de 18.763 (aproximadamente dez vezes superior ao erro médio quadrático) e que foi utilizada apenas uma variável de input, o erro das previsões no teste foi bastante favorável. Para além de confirmarem a capacidade do NeuralTools[®] de entender relações entre variáveis, os resultados obtidos neste ensaio permitiram também concluir que a análise de correlações constitui uma opção válida no apoio à seleção das combinações de variáveis que devem ser usadas para o treino de redes neuronais.

4.2.2 Ensaios de redes sem desfasamento de variáveis

A segunda etapa da fase de treino e de teste de redes neuronais teve como principal objetivo a realização de ensaios com conjuntos de dados em que as variáveis de input não possuíam um desfasamento temporal em relação às de output, isto é, que cada caso integrava apenas os valores das variáveis correspondentes ao mesmo dia de amostragem. De entre as várias etapas definidas para a fase de treino/teste, esta foi a que envolveu a realização de um maior número de ensaios (112), tendo-se revelado uma tarefa algo morosa.

Nesta etapa, o caudal de biogás produzido foi sempre definido como sendo a variável dependente. A abordagem adotada para a criação dos conjuntos de dados compreendeu a inclusão do maior número possível de combinações das variáveis com maior correlação com os valores do caudal de biogás. Procurou-se ainda integrar em alguns conjuntos de dados as séries de valores das restantes variáveis.

Embora nesta etapa do desenvolvimento do estudo fosse ainda prematura uma identificação das RNA com melhor capacidade para prever com exatidão o biogás produzido, foi possível, ainda assim, excluir as que apresentaram resultados de treino e de teste menos satisfatórios. Os critérios de exclusão de redes basearam-se em alguns dos indicadores de desempenho constantes dos relatórios de treino e de teste. Assim, foram excluídas as RNA que se enquadraram em pelo menos duas das seguintes condições:

- ✓ Más previsões no treino (30% tolerância) > 50%
- ✓ Más previsões no teste (30% tolerância) > 50%
- ✓ Erro médio quadrático no treino > 250
- ✓ Erro médio quadrático no teste > 250

Das redes treinadas e testadas nesta etapa, 35 não se enquadraram nos critérios de exclusão estabelecidos. Por se considerar que eram viáveis face aos objetivos do estudo definidos, estas redes foram selecionadas para a realização de novos ensaios nas etapas seguintes de treino e de teste. No Apêndice IV, é possível consultar um quadro com alguns indicadores de desempenho das redes que apresentaram melhores resultados nesta etapa; no Quadro 4.6 são apresentados os melhores valores obtidos para cada um dos indicadores de desempenho considerados.

Quadro 4.6 – Melhores valores obtidos para cada um dos quatro indicadores de desempenho utilizados e respetivas redes.

Indicadores de desempenho	Rede	
Nº ensaios	112	-
Nº redes "viáveis"	35	-
% más previsões (treino - 30% toler.) mais baixa	0,00%	8-E/8-F/ 8-G/ 8-H /9-G-5/9-G-7
% más previsões (teste - 30% toler.) mais baixa	10,00%	8-D
Erro médio quadrático (treino) mais baixo	0,04	8-H
Erro médio quadrático (teste) mais baixo	106,36	8-D

Observa-se no Quadro 4.6 que, apesar dos bons resultados obtidos pelas redes “8-D” e “8-H”, não foi possível identificar a rede com o melhor desempenho nesta fase.

4.2.3 Ensaios de redes com desfasamento de variáveis

Quando as lamas homogeneizadas são elevadas para o digestor anaeróbio, permanecem no seu interior durante um período suficientemente longo para que possam ser devidamente estabilizadas. Como este período corresponde, geralmente, a vários dias, considerou-se a hipótese de as características das lamas afluentes não surtirem um efeito imediato nos parâmetros medidos no interior ou à saída do digestor, nomeadamente no caudal de biogás produzido. Assim, a terceira etapa do treino/teste de redes consistiu na realização de ensaios com conjuntos de dados em que a série de valores do caudal de biogás está desfasada em relação às variáveis de input.

Nesta etapa foram efetuados 40 ensaios de treino e de teste, tendo sido utilizados conjuntos de dados de variáveis com frequência de amostragem inferior a uma vez por semana e conjuntos com frequência de amostragem diária. A maioria dos ensaios realizados baseou-se nas redes da segunda etapa com melhores indicadores de desempenho, mas a função de procura da melhor rede do NeuralTools[®] manteve-se ativada para estes casos. No Apêndice V, é possível consultar um quadro com os indicadores de desempenho de algumas das redes treinadas e testadas nesta etapa.

Em relação aos conjuntos de dados diários, estipulou-se a realização de ensaios com as mesmas variáveis da rede “2-A”. Como o número de variáveis com séries diárias era muito reduzido, a escolha do conjunto de dados da rede “2-A” prendeu-se com os bons resultados obtidos na segunda etapa e com o facto de estarem incluídas todas as variáveis diárias. Devido à disponibilidade de dados correspondentes a dias seguidos, optou-se também pela execução de ensaios em que o valor do caudal estava desfasado, num mesmo caso do conjunto de dados, 3 dias, 7 dias, 10 dias, 12 dias e 15 dias em relação aos valores das restantes variáveis (Quadro 4.7).

Quadro 4.7 – Indicadores de desempenho dos ensaios de treino e de teste com desfaseamento, realizados a partir da rede “2-A”.

Rede			2-A	2-A-4-A	2-A-4-B	2-A-4-C	2-A-4-D	2-A-4-E
Nº dias de desfaseamento			0	3	7	10	12	15
Nº casos			620	617	613	610	608	605
Variáveis entrada	T reator	°C	62,59%	52,53%	56,58%	63,87%	30,69%	51,13%
	TRH	dias	32,89%	28,36%	0,43%	4,60%	28,49%	10,82%
	Q lamas	m ³ /dia	4,52%	19,11%	42,99%	31,53%	40,82%	38,05%
Resultados	Melhor rede		GRNN	GRNN	GRNN	GRNN	MLFN (3 nós)	GRNN
	% más previsões (treino - 30% toler.)		49,60%	46,56%	42,65%	42,21%	51,65%	43,80%
	% más previsões (teste - 30% toler.)		53,23%	46,34%	49,59%	57,38%	55,74%	56,20%
	Erro médio quadrático (treino)		221,04	223,37	206,07	202,57	236,63	206,18
	Erro médio quadrático (teste)		232,57	230,64	244,61	258,62	266,14	257,71

Na Figura 4.5 é efetuada uma comparação dos resultados obtidos nestes ensaios.

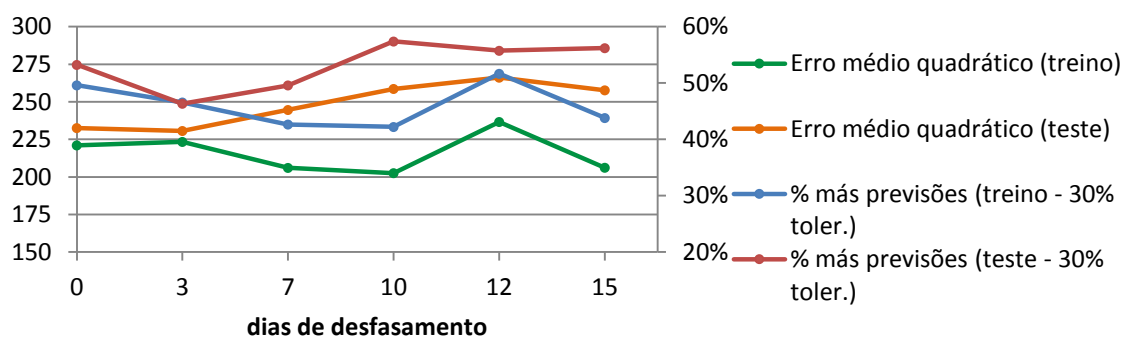


Figura 4.5 – Indicadores de desempenho dos ensaios de treino e de teste com desfaseamento, realizados a partir da rede “2-A”.

Comparando os resultados apresentados no Quadro 4.7 e na Figura 4.5, é possível verificar que o desfaseamento de alguns dias da variável “Q biogás” não produz alterações muito significativas nos valores dos indicadores de desempenho. Uma análise ao impacto relativo de cada variável permite também averiguar a inexistência de um padrão claro nestes indicadores, mantendo-se, no geral, a temperatura como a variável mais influente e uma alternância entre a influência do caudal de lamas e do T_{RH} .

Apesar de os resultados obtidos não terem sido particularmente conclusivos, a hipótese de desfasear variáveis de conjuntos de dados diários não deverá ser excluída em trabalhos futuros. Refere-se que a rede “2-A” foi criada a partir de três variáveis apenas e que uma delas - a temperatura - possui uma taxa de variação muito reduzida, razões que poderão ter contribuído para que o desfaseamento destas variáveis em relação ao “Q biogás” pouco ou nada alterasse os resultados obtidos. Assim, seria interessante a realização de ensaios com conjuntos de dados constituídos por séries diárias de vários parâmetros de qualidade, tanto das lamas afluentes à digestão como das lamas presentes no reator, para que pudessem ser efetuadas diversas combinações de desfaseamento de variáveis nas redes criadas. No presente estudo, a indisponibilidade de dados diários de qualidade impossibilitou a adoção da referida metodologia.

A maioria dos ensaios efetuados na terceira etapa recaíram sobre conjuntos de dados com frequência de amostragem inferior a uma vez por semana. Considerou-se que o procedimento mais correto a adotar neste tipo de ensaios seria a introdução dos valores desfasados da variável “Q biogás” diretamente nos conjunto de dados. No entanto, a incerteza quanto ao número ideal de dias de desfasamento que deveria ser escolhido e o elevado tempo necessário para a criação de cada conjunto de dados conduziram à procura de outras soluções.

Como se pretendia maximizar o número de ensaios a realizar nesta etapa, a alternativa adotada passou pelo recuo da série de caudais de biogás uma célula na folha de cálculo. Ao introduzir um valor da variável “Q biogás” no caso imediatamente anterior, era esperado que fosse introduzido um desfasamento correspondente ao período entre os dois casos. Contudo, a periodicidade entre amostragens da mesma variável revelou-se algo irregular. Na Figura 4.6 observa-se um histograma dos intervalos entre amostragens da variável “MV in”.

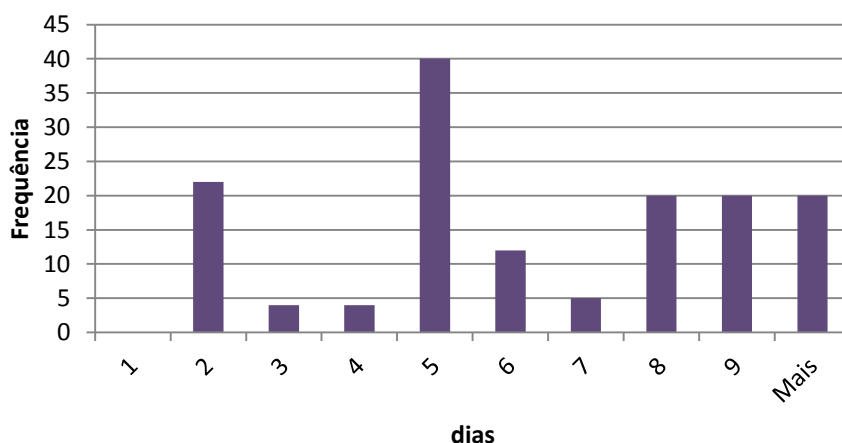


Figura 4.6 – Histograma dos intervalos entre amostragens da variável “MV in”.

O parâmetro “MV in” constitui a variável de qualidade com o maior número de dados disponível. Embora a frequência de amostragem média seja de 7,2 dias, é possível verificar na Figura 4.6 que a periodicidade das análises é bastante variável. Como tal, a análise dos ensaios efetuados deste modo deverá apenas contemplar a evolução das redes com a introdução de desfasamento, e não a reação dos indicadores de desempenho a um número específico de dias de desfasamento.

No Apêndice V, são apresentados os indicadores de desempenho de ensaios realizados com conjuntos de dados desfasados uma célula na folha de cálculo. Ao comparar estes valores com os resultados das redes da segunda etapa que lhes serviram de base, é possível constatar que, no geral, este tipo de desfasamento não surtiu um efeito positivo no treino e teste de redes neuronais. A redução do número de casos usados para treinar as redes poderá ter contribuído para aumentar o erro nesta etapa. Contudo, a irregularidade na

frequência de amostragem afigura-se como a causa mais provável para a diminuição da qualidade das previsões, e, por essa razão, considerou-se importante a realização de alguns ensaios com a variável “Q biogás” desfasada de forma mais regular. Os resultados destes ensaios são apresentados no Quadro 4.8.

Quadro 4.8 – Indicadores de desempenho do ensaio realizado com a rede “4-A-3”, da segunda etapa, e com outras quatro, com as mesmas variáveis desfasadas de forma diferente em relação ao caudal de biogás.

Rede			4-A-3	4-A-3-4	4-A-4-A	4-A-4-B	4-A-4-C
Desfasamento			-	1 célula	3 dias	7 dias	10 dias
Nº casos			94	91	93	93	91
Variáveis entrada	TRH	dias	32,83%	26,38%	44,25%	22,08%	55,55%
	Q lamas	m ³ /dia	31,10%	17,41%	0,19%	26,50%	0,65%
	T reator	°C	19,82%	26,43%	25,77%	27,28%	2,65%
	LMV in	kg/dia	16,25%	29,77%	29,79%	24,13%	41,15%
Resultados	Melhor rede		MLFN (3 nós)	MLFN (2 nós)	GRNN	MLFN (5 nós)	GRNN
	% más previsões (treino - 30% toler.)		40,00%	36,99%	40,54%	28,38%	54,79%
	% más previsões (teste - 30% toler.)		47,37%	50,00%	42,11%	52,63%	61,11%
	Erro médio quadrático (treino)		162,34	177,07	220,06	124,62	250,88
	Erro médio quadrático (teste)		187,67	253,24	232,52	177,85	264,35

No Quadro 4.8, observa-se que, na rede “4-A-3-4”, o desfasamento resultou do recuo de uma célula na folha de cálculo por parte da série dependente e à semelhança do que havia sucedido com praticamente todos os ensaios realizados nestes moldes, os indicadores de desempenho acusaram piores resultados. Nas restantes três redes foi aplicado o desfasamento através da substituição direta dos caudais de biogás pelos valores desfasados 3 dias, 7 dias e 10 dias. Apenas os indicadores de desempenho da rede “4-A-4-B”, com desfasamento de 7 dias, apresentaram resultados favoráveis, tendo, por isso, sido selecionada para a realização de mais ensaios nas etapas seguintes do estudo.

De um modo geral, o desfasamento dos caudais de biogás em relação às variáveis de input não obteve melhores resultados que as redes da segunda etapa, sem desfasamento. Porém, o desfasamento de variáveis deverá continuar a ser considerado uma opção válida no desenvolvimento de redes neuronais, e a realização de mais ensaios com séries de dados diários de um maior número de variáveis permitirá obter resultados mais conclusivos.

4.2.4 Ensaios com conjuntos de dados adicionais

Na fase de preparação dos dados, verificou-se a existência de um elevado número de variáveis com as quais seria possível criar conjuntos de dados de treino e de teste. O número de combinações possíveis (x) de n variáveis é dado pela equação:

$$x = 2^n - 1 \quad (\text{Equação 4.6})$$

Se forem consideradas 20 variáveis para criação de conjuntos de dados, então existem 1.048.575 combinações possíveis, o que torna impraticável a execução manual do mesmo número de ensaios. Como tal, os ensaios realizados nas etapas de treino e de teste anteriores representam apenas uma reduzida amostra do que seria possível face ao número de variáveis

utilizadas. Por essa razão, os ensaios desta etapa incluem o treino e teste de conjuntos de dados definidos segundo critérios específicos, não experimentados em etapas anteriores.

Analisando os resultados obtidos em ensaios anteriores, é possível identificar algumas variáveis de input com um impacto no treino da rede bastante inferior ao das restantes. Neste sentido, considerou-se importante a realização de alguns ensaios complementares com conjuntos de dados em que a variável menos influente era excluída (Quadro 4.9).

Quadro 4.9 – Indicadores de desempenho da rede “5-F-3” e da rede “5-F-3-5”, com as mesmas variáveis à exceção do “Q lamas”.

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
5-F-3	T reator	°C	44,64%	81	Melhor rede	GRNN
	pH in	-	22,30%		% más previsões (treino - 30% toler.)	20,00%
	TRH	dias	21,88%		% más previsões (teste - 30% toler.)	56,25%
	CV	kg/(m ³ .dia)	10,92%		Erro médio quadrático (treino)	107,44
	Q lamas	m ³ /dia	0,26%		Erro médio quadrático (teste)	158,59
5-F-3-5	T reator	°C	41,79%	81	Melhor rede	GRNN
	H in	-	24,05%		% más previsões (treino - 30% toler.)	29,23%
	CV	kg/(m ³ .dia)	17,95%		% más previsões (teste - 30% toler.)	50,00%
	TRH	dias	16,20%		Erro médio quadrático (treino)	158,97
	-	-	-		Erro médio quadrático (teste)	302,04

A rede “5-F-3-5” foi criada a partir da “5-F-3”, tendo sido apenas excluída a variável “Q lamas” devido ao seu reduzido impacto. No Quadro 4.9 é possível verificar que a exclusão da variável com menos influência agravou a percentagem de más previsões no treino e o erro médio quadrático do treino e do teste. Outros ensaios idênticos foram realizados ao longo da quarta etapa, e, de um modo geral, os resultados obtidos não melhoraram, o que leva a crer que a exclusão da variável menos influente pode comprometer a precisão das previsões.

A partir da análise de correlações efetuada na preparação dos dados, concluiu-se que as variáveis com uma maior correlação linear com o caudal de biogás produzido eram: “Q lamas”, “pH in”, “MV in”, “LMV in”, “LMV removida”, “pH reator”, “AGV reator”, “T reator” e “T_{RH}”. Como tal, foram criados os conjuntos de dados “9-G-1” e “9-G-2” a partir das séries de dados destas variáveis (Quadro 4.10).

Quadro 4.10 – Indicadores de desempenho dos ensaios realizados com as redes “9-G-1” e “9-G-2”, criadas a partir das variáveis com maior correlação com o caudal de biogás produzido.

Rede	Variáveis entrada	Impacto	Nº casos	Resultados		
9-G-1	pH reator	-	31,03%	41	Melhor rede	GRNN
	pH in	-	29,71%		% más previsões (treino - 30% toler.)	0,00%
	T reator	°C	21,73%		% más previsões (teste - 30% toler.)	37,50%
	MV in	mg/L	7,88%		Erro médio quadrático (treino)	14,25
	TRH	dias	5,23%		Erro médio quadrático (teste)	167,22
	LMV in	kg/dia	3,95%			
	AGV reator	mg/L	0,32%			
	Q lamas	m ³ /dia	0,16%			
9-G-2	T reator	°C	82,35%	54	Melhor rede	GRNN
	MV in	mg/L	11,02%		% más previsões (treino - 30% toler.)	0,00%
	pH in	-	5,39%		% más previsões (teste - 30% toler.)	63,64%
	LMV in	kg/dia	1,23%		Erro médio quadrático (treino)	0,20
	pH reator	-	0,01%		Erro médio quadrático (teste)	276,79
	TRH	dias	0,00%			
	Q lamas	m ³ /dia	0,00%			
	LMV removida	kg/dia	0,00%			

No Quadro 4.10, é possível constatar que a variável “LMV removida”, uma das que possuía uma correlação linear moderada com o caudal de biogás produzido, não foi incluída no primeiro conjunto de dados, assim como a variável “AGV reator” não foi incluída no segundo. A razão para tal prende-se com a escassez de casos correspondentes aos dias em que ambas as variáveis foram amostradas, tendo sido necessária a sua separação em dois conjuntos de dados. Embora os dois ensaios demonstrem bons resultados no treino, a percentagem de más previsões e o erro médio quadrático agravaram-se no teste. A reduzida dimensão das séries para conjuntos de dados com oito variáveis assume-se como a causa provável dos elevados valores dos indicadores de desempenho do teste. Os resultados obtidos nestes ensaios não são, todavia, negativos, o que leva a crer que uma extensão das séries de dados poderá contribuir para a redução do número de más previsões e do erro médio quadrático no teste.

Outra abordagem à criação de conjuntos de dados foi o agrupamento de variáveis consoante o seu local de amostragem. Como tal, criou-se o conjunto “9-G-3”, com as variáveis medidas à entrada do digestor (Quadro 4.11), e o conjunto “9-G-4”, com as variáveis medidas no interior e à saída do digestor (Quadro 4.12).

Quadro 4.11 – Indicadores de desempenho do ensaio realizado com a rede “9-G-3”, criada a partir de variáveis medidas à entrada do digestor.

Rede	Variáveis entrada	Impacto	Nº casos	Resultados		
9-G-3	LCQO in	kg/dia	99,77%	67	Melhor rede	GRNN
	Q lamas	m ³ /dia	0,14%		% más previsões (treino - 30% toler.)	38,89%
	TRH	dias	0,08%		% más previsões (teste - 30% toler.)	76,92%
	CQO in	mg/L	0,01%		Erro médio quadrático (treino)	184,34
	pH in	-	0,00%		Erro médio quadrático (teste)	235,57
	MV in	mg/L	0,00%			
	LMV in	kg/dia	0,00%			

Apesar de a variável “LCQO in” ter tido uma maior influência na previsão do caudal de biogás que as restantes variáveis, os indicadores de desempenho no treino e no teste da rede “9-G-3” não se revelaram totalmente desfavoráveis. Assim, segundo os critérios de seleção de redes viáveis utilizado na segunda etapa, a rede “9-G-3” prossegue para a etapa seguinte, para realização de novos ensaios.

Quadro 4.12 – Indicadores de desempenho do ensaio realizado com a rede “9-G-4”, criada a partir de variáveis medidas no interior e à saída do digestor.

Rede	Variáveis entrada	Impacto	Nº casos	Resultados		
9-G-4	T reator	°C	39,75%	36	Melhor rede	GRNN
	pH reator	-	27,06%		% más previsões (treino - 30% toler.)	3,45%
	TAL reator	mg/L	17,16%		% más previsões (teste - 30% toler.)	14,29%
	TRH	dias	7,16%		Erro médio quadrático (treino)	36,54
	AGV reator	mg/L	5,23%		Erro médio quadrático (teste)	157,91
	AGV/TAL	-	1,68%			
	LMV out	kg/dia	0,59%			
	Q lamas	m ³ /dia	0,56%			
	MV reator	mg/L	0,39%			
	CQO reator	mg/L	0,23%			
LCQO reator	kg/dia	0,20%				

No Quadro 4.12, é possível verificar que o número de variáveis de input utilizadas é bastante elevado, principalmente considerando que a extensão das séries é de apenas 36 dados; no entanto, os indicadores de desempenho do treino e do teste foram favoráveis, o que sugere que a rede “9-G-4” possui algum potencial para a previsão do caudal de biogás produzido.

Como foi referido no capítulo três, o grupo AdP avançou em 2013 com a realização de um estudo referente à utilização de RNA na gestão de processos de digestão anaeróbia na ETAR Sul da SIMRIA, e os resultados obtidos motivaram a realização de trabalhos idênticos noutras empresas do grupo, entre as quais, a SIMARSUL. Nesse estudo, foram identificadas as três redes neuronais que apresentaram os melhores resultados na previsão do caudal de biogás produzido na ETAR Sul, e considerou-se que seria importante a replicação desse ensaio com os dados da digestão anaeróbia da ETAR do Seixal (Quadro 4.13).

Quadro 4.13 – Indicadores de desempenho da rede “BNTrTsRNA3.16”, treinada com dados da ETAR Sul da SIMRIA (Rocha, 2013), e da rede “5-H-5-4-A”, criada a partir das mesmas variáveis medidas na ETAR do Seixal.

Rede	Variáveis entrada	Impacto	Nº casos	Resultados		
BNTrTsRNA3.16	T reator	°C	28,61%	242	Melhor rede	GRNN
	MV in	%	26,94%		% más previsões (treino - 15% toler.)	17,01%
	CV	kg/(m ³ .dia)	22,39%		% más previsões (teste - 15% toler.)	27,08%
	Q lamas	m ³ .dia	22,06%		Erro médio quadrático (treino)	143,19
					Erro médio quadrático (teste)	179,77
5-H-5-4-A	Q lamas	m ³ .dia	44,93%	97	Melhor rede	MLFN (3 nós)
	T reator	°C	41,16%		% más previsões (treino - 30% toler.)	38,46%
	MV in	mg/L	7,82%		% más previsões (teste - 30% toler.)	52,63%
	CV	kg/(m ³ .dia)	6,99%		Erro médio quadrático (treino)	157,58
				Erro médio quadrático (teste)	249,48	

No Quadro 4.13, é possível verificar que os resultados obtidos no ensaio com os dados da ETAR do Seixal foram agravados em relação ao ensaio com os dados da ETAR Sul, mesmo com a ativação da função *Best Net Search* e com a duplicação da tolerância no treino e no teste. A diminuição do número de dados utilizados no treino poderá ter contribuído para o agravamento dos indicadores de desempenho no ensaio da rede “5-H-5-4-A”, mas o facto de os dois digestores estarem a operar em condições distintas assume-se como a causa provável das diferenças observadas. Refere-se que estas duas redes foram treinadas com algoritmos de aprendizagem diferentes, não tendo sido possível a previsão de valores de caudal de biogás da ETAR do Seixal com a rede treinada na ETAR Sul.

No final desta quarta etapa de treino e de teste, procedeu-se à identificação das redes com melhores resultados, usando os critérios estabelecidos para os ensaios realizados na segunda etapa. Assim, das 45 redes treinadas e testadas nesta etapa, transitaram 16 para a seguinte. No Apêndice VI é possível consultar os indicadores de desempenho dessas 16 redes neuronais.

4.2.5 Ensaio com as melhores redes

A quinta e última etapa do treino e teste das redes neuronais consistiu na realização de ensaios com as melhores redes encontradas anteriormente, tendo como principal objetivo a identificação das redes que serão utilizadas na previsão dos caudais de biogás produzido. Como os critérios de seleção estabelecidos anteriormente foram algo amplos, transitaram para esta etapa 52 das 266 redes treinadas e testadas. Considerou-se este número excessivo para integrar nos ensaios previstos, e, por isso, procedeu-se à sua triagem. Os critérios utilizados basearam-se novamente nos indicadores de desempenho do treino e teste das RNA, tendo sido excluídas as que satisfizeram pelo menos duas das seguintes condições:

- ✓ Más previsões no treino (30% tolerância) > 20%
- ✓ Más previsões no teste (30% tolerância) > 40%
- ✓ Erro médio quadrático no treino > 100
- ✓ Erro médio quadrático no teste > 200

A partir deste processo de triagem obtiveram-se 13 redes que não se enquadraram nos critérios de exclusão estabelecidos, tendo sido, deste modo, seleccionados os conjuntos de dados que integrariam os ensaios desta etapa. O passo seguinte consistiu na repetição do treino e do teste das redes identificadas. Inicialmente, havia sido considerada a hipótese de reduzir a tolerância às más previsões no treino e no teste de 30% para 15%, mas a curta extensão das séries correspondentes e as elevadas percentagens de más previsões, principalmente no teste, conduziram à rejeição desta abordagem.

Assim, optou-se pela realização de três ensaios, com as mesmas configurações, a cada um dos 13 conjuntos de dados, para confirmação dos resultados obtidos em etapas anteriores. Foi ainda calculada a média aritmética dos indicadores de desempenho dos quatro

ensaios realizados a cada um dos 13 conjuntos de dados para que pudesse ser efetuado um último processo de triagem. Deste modo, foram excluídas as redes que se enquadraram em pelo menos um dos seguintes critérios:

- ✓ Soma da média aritmética das más previsões no treino e no teste > 55%
- ✓ Soma do erro médio quadrático no treino e no teste > 250

A partir deste processo de triagem foram selecionados cinco conjuntos de dados para a previsão dos caudais de biogás produzido: “8-F”, 8-H”, “9-G-1”, “9-G-8” e “10-A”. Os resultados dos ensaios realizados com os 13 melhores conjuntos de dados e com as redes que transitaram para a fase de previsão podem ser consultados no Apêndice VII.

Embora a metodologia de seleção das melhores redes neuronais adotada tenha possibilitado a obtenção de resultados conclusivos, considerou-se que seria importante, numa perspetiva de sistematização da informação recolhida, efetuar uma comparação entre as melhores redes identificadas segundo critérios diferentes daqueles que haviam sido enunciados anteriormente. Neste sentido, optou-se pela elaboração do Quadro 4.14, no qual são apresentados os indicadores de desempenho de cada uma das melhores RNA identificadas para um determinado número de variáveis de entrada.

Quadro 4.14 – Indicadores de desempenho das melhores redes neuronais identificadas para cada número de variáveis de entrada.

Melhor rede	9-E-1	9-E-3	9-E-7	8-E	8-H	9-G-7	9-G-1	9-G-4
Nº variáveis	1	2	3	4	5	7	8	11
T reator	8 em 8	T reator	T reator	T reator	T reator	T reator	T reator	T reator
LMV in	3 em 8	LMV in	LMV in				LMV in	
pH in	2 em 8		pH in				pH in	
pH reator	5 em 8			pH reator	pH reator	pH reator	pH reator	pH reator
AGV reator	5 em 8			AGV reator	AGV reator	AGV reator	AGV reator	AGV reator
TAL reator	4 em 8			TAL reator	TAL reator	TAL reator		TAL reator
CQO reator	3 em 8				CQO reator	CQO reator		CQO reator
MV reator	2 em 8					MV reator		MV reator
LCQO reator	2 em 8					LCQO reator		LCQO reator
TRH	2 em 8						TRH	TRH
Q.lamas	2 em 8						Q.lamas	Q.lamas
MV in	1 em 8						MV in	
AGV/TAL	1 em 8							AGV/TAL
LMV out	1 em 8							LMV out
Nº casos	618	87	76	46	38	36	41	36
Melhor rede	MLFN (29 nós)	MLFN (6 nós)	GRNN	MLFN (3 nós)	MLFN (14 nós)	GRNN	GRNN	GRNN
% más previsões (treino - 30% toler.)	49,39%	31,43%	27,87%	24,32%	0,00%	6,67%	0,00%	3,45%
% más previsões (teste - 30% toler.)	50,81%	41,18%	26,67%	33,33%	37,50%	42,86%	37,50%	14,29%
Erro médio quadrático (treino)	240,89	144,47	126,55	89,78	0,04	62,92	14,25	36,54
Erro médio quadrático (teste)	243,50	242,49	226,38	173,48	192,97	173,14	167,22	157,91

A partir do Quadro 4.14, verifica-se que a variável “T reator” está presente em todos os conjuntos de dados de cada uma das melhores redes, o que permite realçar a importância da sua inclusão num modelo de previsão do caudal de biogás produzido. Destacam-se também as variáveis “pH reator” e “AGV reator”, cuja presença em cinco dos oito conjuntos de dados das melhores redes evidencia a sua relevância na previsão do biogás.

Analisando as oito melhores redes, constantes do Quadro 4.14, é ainda possível retirar algumas conclusões referentes ao número de variáveis de entrada de cada conjunto de dados.

Na Figura 4.7 é efetuada uma comparação entre o número de casos utilizados para treinar e testar as melhores redes identificadas para cada número de variáveis de entrada.

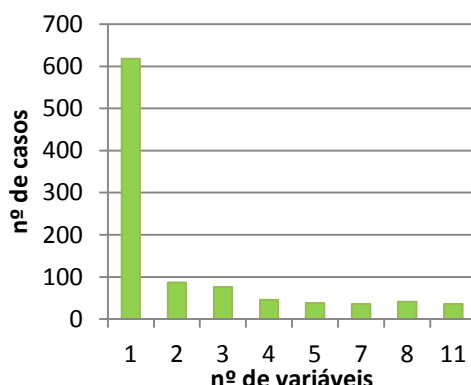


Figura 4.7 – Comparação entre o número de casos utilizados para treinar e testar as melhores redes identificadas para cada número de variáveis de entrada.

Observa-se na Figura 4.7 que, com o aumento do número de variáveis de entrada, o número de casos utilizados para treinar e testar a rede é reduzido. Assume-se que este facto esteja relacionado com a frequência de amostragem das variáveis, tanto devido à existência de poucas variáveis medidas diariamente, como pela irregularidade na medição das restantes, o que dificulta a identificação de casos com todos os dados necessários disponíveis.

Relativamente aos indicadores de desempenho das oito melhores redes identificadas no Quadro 4.14, foi ainda possível comparar as percentagens de más previsões e os erros médios quadráticos, tanto do treino como do teste, obtidos para cada número de variáveis de entrada (Figura 4.8).

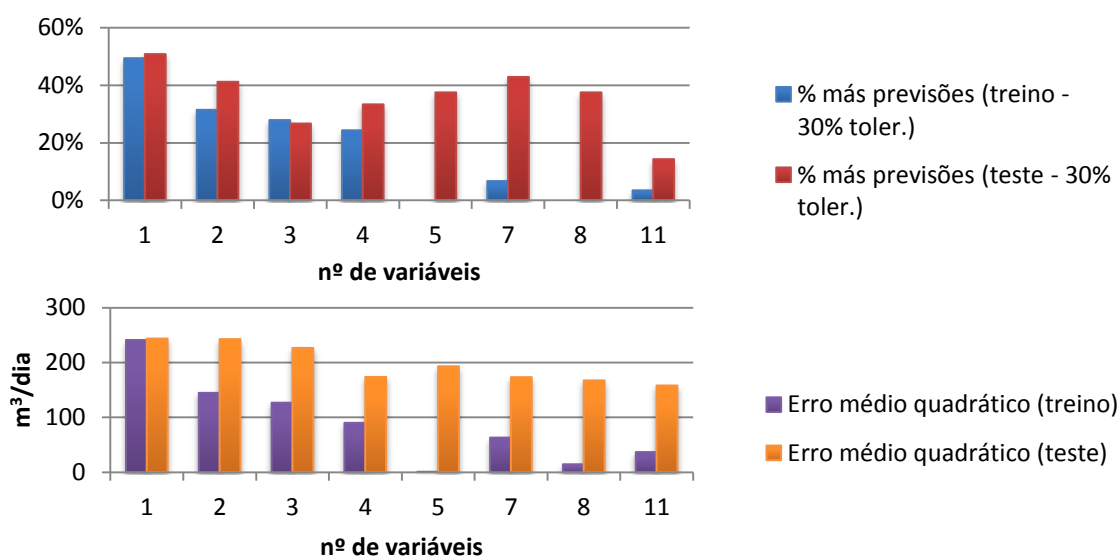


Figura 4.8 – Comparação entre as percentagens de más previsões (em cima) e os erros médios quadráticos (em baixo) obtidos no treino e teste das melhores redes identificadas para cada número de variáveis de entrada.

Verifica-se na Figura 4.8 que, de um modo geral, o aumento do número de variáveis de entrada tende a diminuir as percentagens de más previsões e os erros médios quadráticos do

treino e do teste. Contudo, observa-se que a partir de cinco variáveis de entrada não ocorre uma melhoria significativa dos indicadores de desempenho. Este facto sugere que, para a criação de um modelo sólido de previsão do biogás produzido e também numa perspetiva de minimização da quantidade de parâmetros a monitorizar com este propósito, o número de variáveis de entrada aconselhável será cinco.

A realização das várias fases consideradas necessárias ao desenvolvimento do estudo envolveu a execução de diversas metodologias e a obtenção de inúmeros resultados intermédios. Na Figura 4.9, é apresentada uma representação esquemática das várias etapas completadas ao longo do estudo, com o intuito de simplificar a compreensão dos principais resultados obtidos e dos procedimentos algo complexos adotados ao longo do percurso que conduziu à seleção das melhores redes destinadas à previsão.

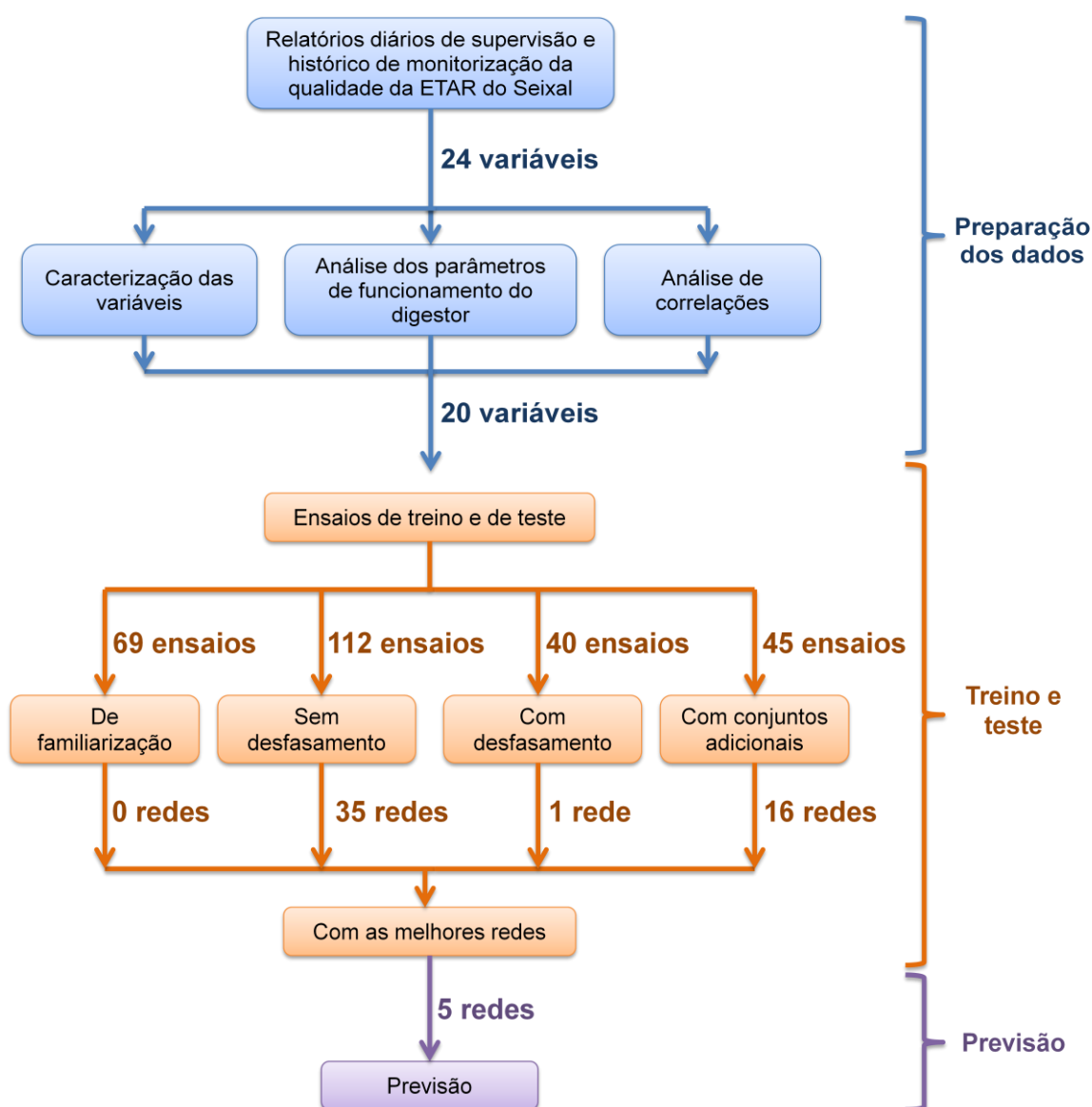


Figura 4.9 – Esquematização das fases de desenvolvimento do estudo.

4.3 Previsão

Estando reunidas as condições para prosseguir com o estudo, procedeu-se à terceira e última fase do seu desenvolvimento, a previsão. Optou-se por realizar ensaios de previsão a cada uma das RNA obtidas nos ensaios de repetição do treino/teste dos cinco melhores conjuntos de dados da etapa anterior. Devido à escassez de dados de algumas variáveis, procurou-se estender ao máximo as séries destinadas à previsão e, como tal, não foi possível efetuar o seu tratamento em tempo útil. As séries utilizadas correspondem ao período entre 20 de maio de 2014 e 15 de setembro de 2014.

No sentido de simplificar a análise dos resultados, selecionou-se a melhor rede neuronal de cada um dos cinco conjuntos de dados, tendo o critério de escolha sido o erro médio quadrático entre os valores reais e previstos da variável “Q biogás”. No Quadro 4.15 são apresentados alguns indicadores estatísticos relativos às previsões efetuadas.

Quadro 4.15 – Indicadores estatísticos referentes às previsões dos valores do caudal de biogás produzido, efetuadas pelas cinco melhores redes neuronais.

	8-F-T1	8-H-T3	9-G-8-T3	9-G-1-T1	10-A-T1	Média das previsões
Variáveis entrada	T reator	T reator	T reator	T reator	T reator	-
	pH reator	pH reator	pH reator	pH reator	pH reator	-
	AGV reator	AGV reator	-	AGV reator	AGV reator	-
	TRH	-	TRH	TRH	TRH	-
	-	-	LMV in	LMV in	LMV in	-
	TAL reator	TAL reator	-	-	-	-
	-	-	-	MV in	-	-
	-	CQO reator	-	-	-	-
	-	-	LCQO in	-	-	-
	-	-	-	pH in	-	-
-	-	-	Q lamas	-	-	
Casos treino/teste	42	38	66	41	44	46
Casos previsão	11	11	12	11	11	12
RMS	105,62	123,51	124,30	182,81	180,79	109,55
R ²	0,445	0,120	0,051	0,071	0,009	0,199

Uma análise ao Quadro 4.15 permite verificar que as variáveis “T reator” e “pH reator” integraram os conjuntos de dados das cinco melhores redes, o que confirma a importância do papel que desempenham nos processos anaeróbios. Este facto deixa também patente a relevância destas duas variáveis na criação de um modelo de RNA para previsão da produção de biogás, às quais deverá ser dada prioridade aquando do estabelecimento dos principais parâmetros a monitorizar na digestão anaeróbia. Também as variáveis “AGV reator” e “T_{RH}”, utilizadas em quatro das cinco melhores redes, e ainda a “LMV in”, utilizada em três das cinco melhores redes, integram o conjunto das principais variáveis a ter em consideração na modelação dos processos anaeróbios com vista à otimização da produção de biogás.

O desempenho das RNA na previsão de valores de uma variável dependente pode ser avaliado de diversas formas, sendo o erro médio quadrático e o coeficiente de determinação (R^2) alguns dos indicadores mais comumente utilizados com este propósito. Para a previsão de dados, é desejável que o erro médio quadrático seja o mais reduzido possível, não havendo, contudo, um valor de referência relativo ao caudal de biogás produzido; quando $R^2 \geq 0,8$, considera-se que a RNA estabeleceu um modelo de previsão “muito bom” (Kasabov, 1998). É possível constatar no Quadro 4.15 que os resultados obtidos para os coeficientes de determinação foram bastante inferiores a 0,8. Porém, considera-se que estes valores poderiam ser consideravelmente superiores se o número de casos utilizados na previsão fosse maior. As previsões efetuadas pelas cinco melhores redes e as médias aritméticas dessas previsões são apresentadas na Figura 4.10.

Para além da análise dos indicadores de desempenho das redes na previsão, são ainda apresentadas no Quadro 4.15 as médias das previsões. Como os casos utilizados nos cinco ensaios correspondem ao mesmo dia (excetuando um caso utilizado nas previsões com a rede “9-G-8-T3”), considerou-se importante efetuar uma média aritmética das cinco previsões de cada dia, com o intuito de perceber se as redes produzem melhores resultados em conjunto ou individualmente. O erro médio quadrático foi de 109,55, tendo sido inferior ao valor obtido em quatro das cinco redes, e o coeficiente de determinação foi de 0,199, sendo superior ao valor de quatro das cinco redes. As médias das previsões são apresentadas na Figura 4.10, em conjunto com as previsões efetuadas pelas cinco melhores redes.

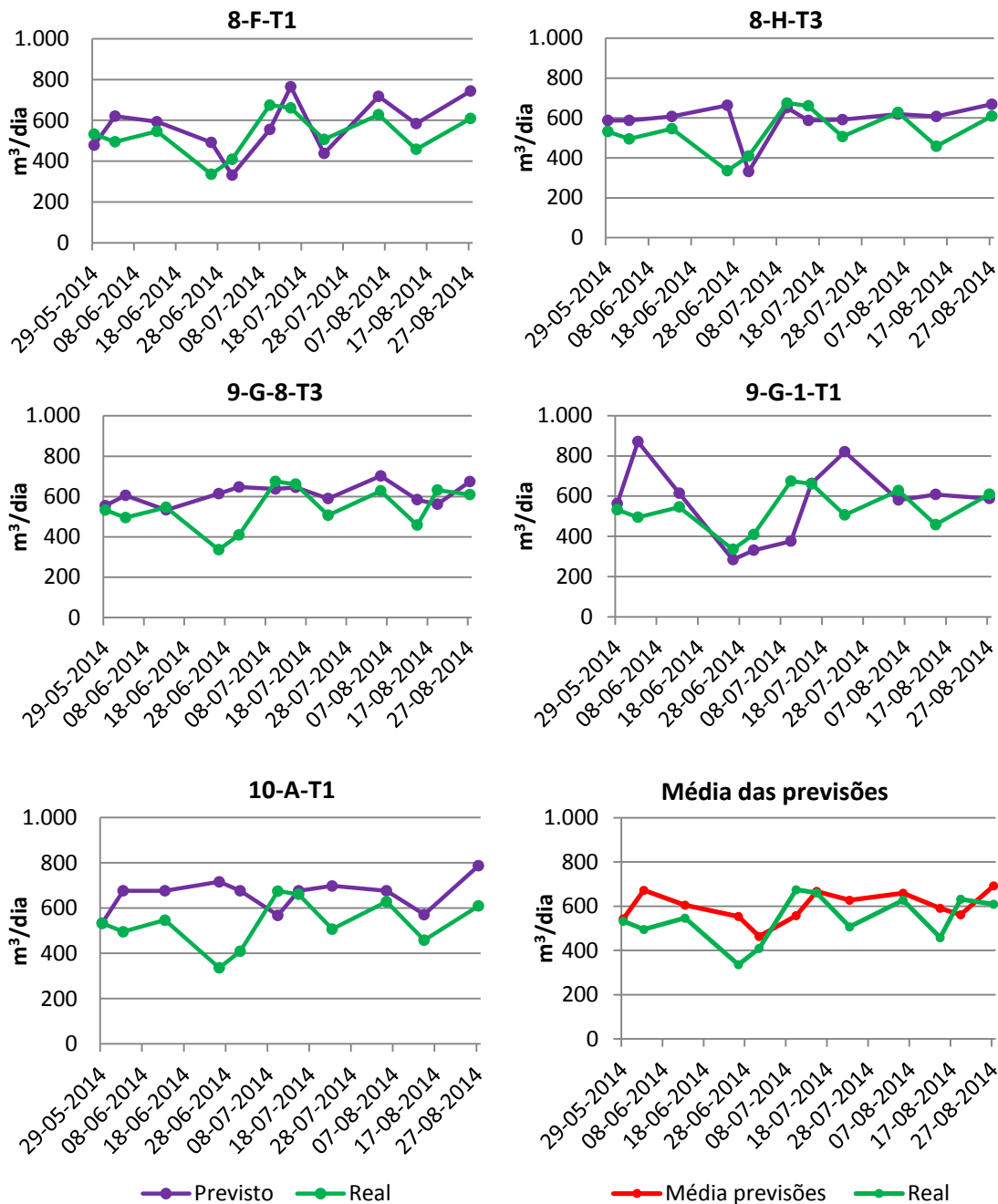


Figura 4.10 – Previsões dos caudais de biogás produzido efetuadas pelas cinco melhores redes neuronais e comparação com os valores reais e com a média das previsões.

Apesar de os resultados obtidos através das médias das previsões terem sido positivos, mais até do que quatro das cinco redes selecionadas, a sua aplicação num modelo de otimização da produção de biogás é inviável, uma vez que seria necessária a monitorização de 11 variáveis.

É possível constatar no Quadro 4.15 que a rede “8-F-T1” possui os valores de erro médio quadrático e de R^2 mais favoráveis de todas as cinco redes. As variáveis utilizadas na criação da rede “8-F-T1” foram o “pH reator”, “ T_{RH} ”, “TAL reator”, “AGV reator” e “T reator”,

cujas séries de dados permitiram a seleção de 11 casos para previsão. Apesar de ser ainda um número de casos muito reduzido, é possível comprovar, na Figura 4.10, a existência de uma “boa” aproximação entre todos os valores previstos e os valores reais. O maior erro nas previsões foi de 156, o que não deixa de ser um indicador interessante, considerando que foram utilizados apenas 47 casos para treinar esta rede. É de notar, também, que quatro das cinco variáveis que integram o conjunto de dados “8-F-T1” apresentaram um coeficiente de correlação linear moderado com o “Q biogás”, podendo este facto ter contribuído para as previsões favoráveis verificadas neste ensaio.

Assim, é possível concluir que a rede neuronal “8-F-T1” possui, de entre todas as redes desenvolvidas neste estudo, o maior potencial para prever o biogás produzido e para modelar a digestão anaeróbia, podendo o erro averiguado na previsão vir a ser reduzido com a extensão do número de casos usados para treinar a rede. Refere-se ainda que como todas as variáveis usadas na sua criação podem ser controladas – com maior ou menor facilidade - na ETAR do Seixal, esta rede poderá ser utilizada como ferramenta de apoio na operação e gestão dos processos de digestão anaeróbia.

Como todas as redes neuronais seleccionadas para a fase de previsão do caudal de biogás foram treinadas e testadas com um número reduzido de casos (entre 38 e 66), considerou-se que seria relevante a execução de ensaios de previsão com conjuntos de dados mais extensos. Deste modo, optou-se pela realização de um ensaio de previsão com a rede “1-A”, com 620 casos de treino/teste e 251 casos destacados para previsão. Os resultados deste ensaio podem ser consultados na Figura 4.11.

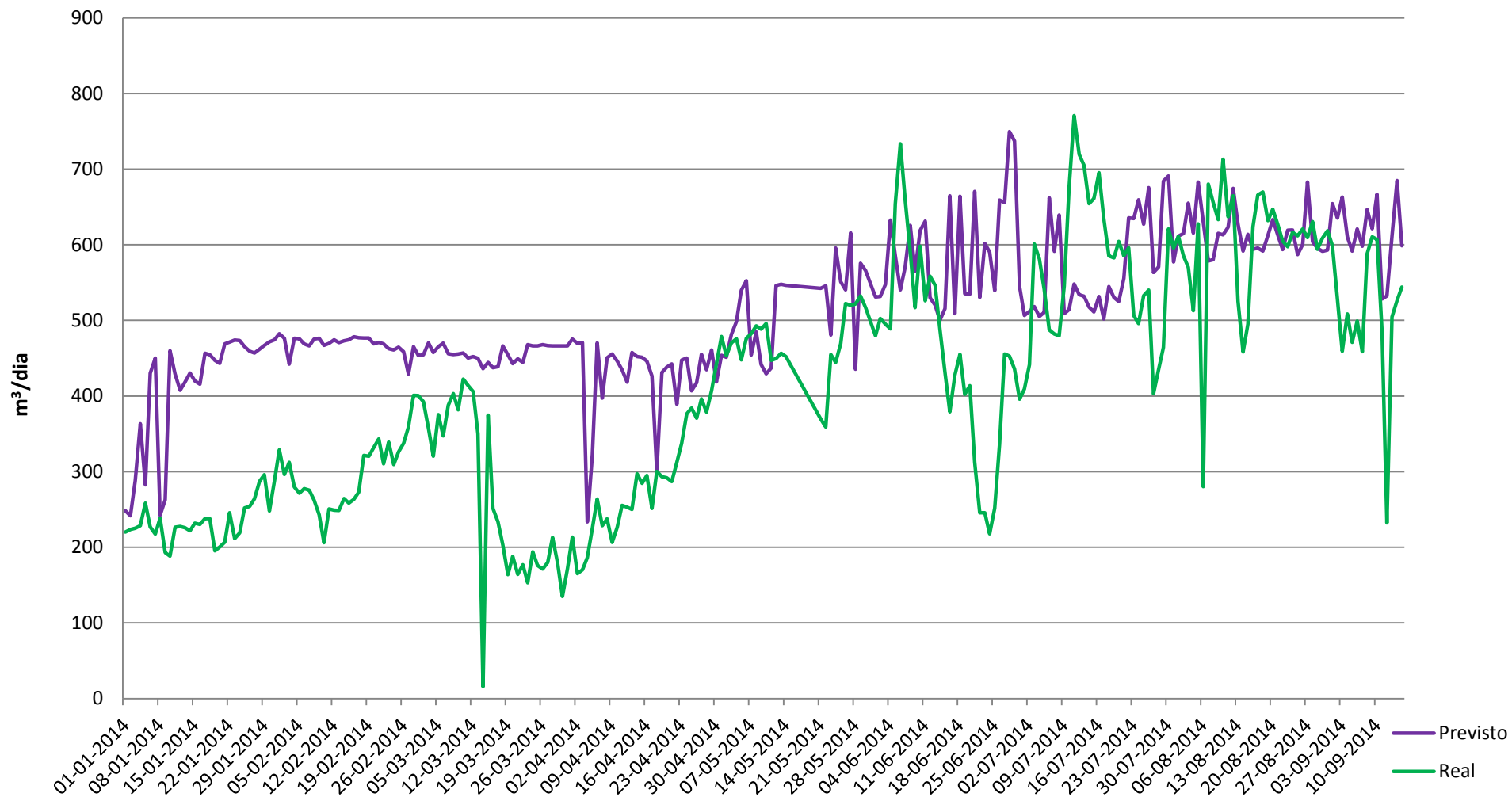


Figura 4.11 – Comparação entre os caudais de biogás previstos através da rede “1-A” e os valores reais.

O erro médio quadrático do ensaio de previsão com a rede “1-A” foi de 163,19 e o coeficiente de determinação foi de 0,422; apesar de se tratar de resultados desfavoráveis relativamente aos valores de referência encontrados na literatura e aos resultados obtidos com a rede “8-F-T1”, é de assinalar que a rede “1-A” conseguiu, ainda assim, aproximar vários valores previstos dos caudais reais de biogás. Embora o número de casos de treino fosse, teoricamente, suficiente para que se conseguisse reduzir substancialmente o erro nas previsões, é preciso ter em consideração que esta rede foi criada a partir de duas variáveis apenas, o “Q lamas” e a “T reator”.

Este facto sugere que, se fosse possível criar um conjunto de dados com a extensão do conjunto “1-A” e combinar o caudal de lamas e a temperatura no digestor com duas variáveis da rede “8-F-T1”, por exemplo, seria expectável que os resultados obtidos no ensaio de previsão fossem bastante favoráveis, presumivelmente, mais até do que os que se verificaram nos ensaios com a “1-A” e “8-F-T1”.

No histórico de monitorização da digestão anaeróbia da ETAR do Seixal é possível encontrar relatórios diários de supervisão com medições da temperatura no digestor e do caudal de lamas, a partir do qual é calculado o T_{RH} . Contudo, alguns parâmetros de qualidade, como o pH das lamas digeridas, a concentração de AGV nas lamas digeridas ou a concentração de matéria volátil afluyente, são determinados laboratorialmente, possuindo uma frequência de amostragem inferior a uma vez por semana. Neste sentido, será importante o alargamento do número de amostras recolhidas por semana para que a produção de biogás na digestão anaeróbia possa ser otimizada, bem como a realização de estudos com o intuito de averiguar se o acréscimo na produção de biogás viabiliza o aumento do número de análises laboratoriais aos parâmetros de qualidade ou mesmo a introdução de instrumentação de monitorização na ETAR que permita uma medição direta dos seus valores.

4.4 Considerações finais

O presente trabalho pretendeu explorar as funcionalidades de um aplicativo informático de criação de modelos de redes neuronais e estudar as possibilidades de otimização da gestão da digestão anaeróbia na ETAR do Seixal, decorrentes da sua aplicação ao histórico de dados.

O software utilizado foi o NeuralTools[®] da Palisade[®], cujas funcionalidades simplificadas e linguagem acessível permitiram abordar a temática das redes neuronais artificiais com comodidade, mesmo do ponto de vista de um utilizador inexperiente. Por outro lado, o NeuralTools[®] é instalado sob a forma de suplemento do Excel, o que possibilitou a execução de todas as fases de desenvolvimento de modelos de redes neuronais no ambiente familiar do Microsoft Office[®].

A preparação dos dados constituiu a primeira e mais demorada fase do desenvolvimento das RNA, e considera-se ter sido esta a etapa determinante de todo o estudo. Os histogramas criados a partir dos dados recolhidos em bruto do histórico de monitorização

da ETAR do Seixal permitiram obter uma perspectiva geral sobre a distribuição dos valores de cada variável e excluir valores adulterados, mal transcritos ou *outliers*, passíveis de introduzir “ruído” nos resultados dos ensaios realizados com as redes neuronais.

A determinação da média aritmética dos parâmetros de funcionamento do digestor e a sua comparação com valores de referência encontrados na literatura contribuiu para avaliar de que forma a instabilidade no digestor e os problemas operacionais e de salinidade verificados na ETAR do Seixal afetam o desempenho da digestão anaeróbia. A análise dos parâmetros de funcionamento permitiu ainda identificar o desvio de algumas variáveis em relação aos valores aconselháveis, tendo esta situação sido verificada com a concentração de matéria seca nas lamas afluentes, com a percentagem de matéria volátil removida, com a temperatura no digestor, com a concentração de AGV nas lamas digeridas, e com o caudal de biogás produzido, o qual se revelou inferior ao que seria expectável.

Ainda na preparação dos dados, considerou-se bastante útil a realização de uma análise aos coeficientes de correlação entre variáveis. Este processo permitiu a identificação de relações de linearidade muito fortes entre variáveis independentes, como a concentração de matéria seca e matéria volátil nas lamas afluentes, e a exclusão de uma delas, para que o número de variáveis a incluir nos ensaios com as RNA fosse reduzido. A análise de correlações possibilitou também encontrar a existência de correlações lineares moderadas entre o caudal de biogás produzido e as variáveis “Q lamas”, “pH in”, “MV in”, “LMV in”, “LMV removida”, “pH reator”, “AGV reator”, “T reator” e “T_{RH}”, tendo, deste modo, sido facilitada a tomada de decisões metodológicas nas fases posteriores de desenvolvimento de redes neuronais.

É importante mencionar que, devido à escassez de dados de algumas variáveis, se procurou estender ao máximo as séries destinadas à previsão de output, e que, por isso, não foi possível efetuar o seu tratamento em tempo útil. Este facto pode ter contribuído para aumentar ligeiramente o erro dos resultados obtidos na fase de previsão.

Neste estudo, optou-se também por acoplar a fase de teste à fase de treino das redes neuronais, funcionalidade essa facultada pelo software NeuralTools[®]. Porém, a reduzida dimensão de alguns conjuntos de dados acabou por determinar a seleção de um pequeno número de casos de treino de RNA, que se revelou, por vezes, insuficiente face aos objetivos propostos. Assim, conclui-se que, em situações de escassez de dados, é preferível reduzir a percentagem de casos destinados para teste, aumentando o número de casos destacados para o treino das redes neuronais.

Refere-se, também, que a metodologia de seleção das melhores redes neuronais adotada neste estudo se revelou algo inadequada. Face ao reduzido número de casos disponível em alguns conjuntos de dados, os critérios de seleção – baseados nos indicadores de desempenho das RNA no treino e no teste – poderão ter excluído algumas redes com potencial para prever o caudal de biogás. Como tal, considera-se que, em situações de

escassez de dados, será preferível acoplar a fase de previsão às fases de treino e de teste e realizar um ensaio “completo” a cada conjunto de dados.

A partir dos ensaios realizados na fase de treino/teste, verificou-se que, de um modo geral, o desfaseamento não favoreceu a obtenção de melhores resultados. Porém, a aplicação deste tipo de tratamento aos conjuntos de dados deverá continuar a ser considerada uma opção válida no desenvolvimento de RNA, e a realização de mais ensaios com séries de dados diários, com um maior número de variáveis de qualidade, permitirá obter resultados mais conclusivos.

Outros ensaios realizados na fase de treino e de teste permitiram concluir que a repetição de um ensaio a um conjunto de dados ao qual foi excluída a variável menos influente prejudica a previsão do biogás, podendo mesmo comprometer o desempenho da rede neuronal de acordo com os objetivos propostos. A partir de uma das melhores redes neuronais concebidas nos ensaios realizados na ETAR Sul, da SIMRIA, criou-se, também, um conjunto de dados com as mesmas variáveis, medidas na ETAR do Seixal. Nos resultados obtidos no ensaio de treino/teste, foi possível verificar um agravamento dos indicadores de desempenho da nova rede, e considera-se que o facto de os dois digestores terem sido concebidos para operar em condições distintas será a causa mais provável para as diferenças observadas.

Ao longo da fase de treino e de teste, foram sendo selecionadas as redes neuronais que apresentaram os melhores indicadores de desempenho nos ensaios realizados, tendo como principal objetivo a sua “passagem” à última fase de desenvolvimento das RNA – a previsão. Cinco redes neuronais integraram este grupo e uma análise aos respetivos conjuntos de dados permitiu identificar as cinco principais variáveis utilizadas na previsão do biogás:

1. Temperatura no digestor;
2. pH das lamas digeridas;
3. T_{RH} ;
4. Concentração de AGV nas lamas digeridas;
5. Carga de MV nas lamas afluentes à digestão.

Uma análise aos indicadores estatísticos das várias previsões efetuadas possibilitou também a identificação da RNA que previu com maior precisão os valores dos caudais de biogás produzido. Utilizando as variáveis “pH reator”, “ T_{RH} ”, “TAL reator”, “AGV reator” e “T reator”, a rede neuronal “8-F-T1” conseguiu prever o “Q biogás” com um erro médio quadrático de 105,62 e com um coeficiente de determinação de 0,445, tendo, por isso, sido considerada a RNA com maior potencial para prever o biogás produzido e para modelar a digestão anaeróbia na ETAR do Seixal. Refere-se, ainda, que quatro das cinco variáveis que integram o conjunto de dados “8-F-T1” apresentaram uma correlação linear moderada com o “Q biogás”, o que ajuda a reforçar a importância do contributo dado pela análise de correlações efetuada na fase de preparação de dados.

Embora a rede “8-F-T1” tenha obtido uma boa aproximação das previsões aos valores reais de biogás produzido, o erro correspondente ficou, ainda assim, aquém dos valores de referência de uma previsão considerada “muito boa”. O reduzido número de casos usados para treinar a rede (47) afigura-se como a principal causa para tal, e será expectável que a extensão das séries de dados usadas na criação desta RNA consiga reduzir o erro das previsões. A escassez de dados referentes aos parâmetros de qualidade da digestão anaeróbia constituiu, de resto, a principal limitação, não só na fase de previsão, mas também ao longo de todo o estudo. A importância da monitorização nos processos anaeróbios ficou bem patente neste estudo, e uma maior e regular frequência de amostragem dos vários parâmetros de qualidade assume-se como a principal medida a implementar na ETAR com vista a uma modelação mais sólida da digestão anaeróbia.

Relativamente às funcionalidades do NeuralTools[®], os diversos ensaios realizados ao longo das fases de desenvolvimento das RNA permitiram verificar que, tanto as redes do tipo GRNN, como as do tipo MLFN alcançaram bons indicadores de desempenho. Mesmo nas cinco redes neuronais destacadas para previsão, a respetiva configuração de rede foi alternando entre estes dois tipos, o que demonstra que, no treino de RNA, a ativação da função *Best Net Search* constitui a melhor opção. No que respeita ao número de nós por camada nas redes do tipo MLFN, não foi encontrado um padrão mais favorável, considerando-se, por isso, a atribuição de 2 a 30 nós como sendo o procedimento mais indicado.

A utilização do NeuralTools[®] como ferramenta de trabalho ao longo deste estudo serviu, também, para confirmar a capacidade deste software de aprender relações não lineares entre variáveis sem o conhecimento prévio da sua estrutura. Os resultados obtidos nos diversos ensaios efetuados com os dados da ETAR do Seixal demonstraram que as redes neuronais artificiais possuem potencial para modelar a digestão anaeróbia e que a sua utilização como ferramenta de apoio na operação e gestão deste tipo de processos se afigura como uma solução promissora, não só no tratamento de águas residuais, mas também noutros campos da engenharia e das ciências, no geral.

5. CONCLUSÕES

5.1 Conclusões

Este trabalho envolveu uma extensa análise estatística dos dados referentes à digestão anaeróbia na ETAR do Seixal e a realização de inúmeros ensaios com as RNA no NeuralTools[®], o que conduziu à recolha de uma significativa quantidade de informação. No sentido de sistematizar o vasto conjunto de resultados obtidos, são apresentadas em seguida as principais conclusões que podem ser retiradas das diferentes fases de desenvolvimento do estudo:

- A preparação dos dados constituiu a fase determinante de todo o estudo.
- A análise das variáveis e a determinação dos parâmetros de funcionamento contribuíram para realçar os problemas de instabilidade verificados no digestor.
- A análise de correlações possibilitou uma diminuição do número de variáveis utilizadas na realização dos ensaios e permitiu identificar as que apresentam uma correlação linear moderada com o biogás produzido: “Q lamas”, “pH in”, “MV in”, “LMV in”, “LMV removida”, “pH reator”, “AGV reator”, “T reator” e “T_{RH}”.
- Quando as séries de dados são muito curtas, será mais vantajoso o aumento do número de casos destacados para treino, reduzindo a percentagem de casos destinados para teste. Em situações de escassez de dados, será também preferível acoplar a fase de previsão às fases de treino e de teste.
- O NeuralTools[®] constitui uma ferramenta ajustada a uma primeira abordagem à temática das redes neuronais artificiais.
- Os resultados dos ensaios realizados com as RNA do tipo GRNN e MLFN foram idênticos, o que demonstra que a ativação da função *Best Net Search* no NeuralTools[®] constitui a melhor opção para o treino e teste.
- A realização de ensaios de treino/teste com desfasamento de variáveis não obteve melhores resultados que os ensaios efetuados sem desfasamento.
- A exclusão da variável menos influente em cada ensaio de treino/teste prejudica o desempenho das RNA na previsão do biogás.
- Não foi possível reproduzir, com os dados da ETAR do Seixal, os resultados de treino/teste obtidos num dos melhores ensaios realizados na ETAR Sul, da SIMRIA.
- As cinco principais variáveis utilizadas na previsão do biogás foram:
 1. Temperatura no digestor;
 2. pH das lamas digeridas;
 3. T_{RH};
 4. Concentração de AGV nas lamas digeridas;
 5. Carga de MV nas lamas afluentes ao digestor.

- A melhor RNA conseguiu prever o caudal de biogás produzido com um erro médio quadrático de 105,62 e com um R^2 de 0,445, utilizando as variáveis “pH reator”, “ T_{RH} ”, “TAL reator”, “AGV reator” e “T reator”.
- A escassez de dados referentes aos parâmetros de qualidade da digestão anaeróbia constituiu a principal limitação ao longo de todo o estudo, e o aumento da frequência de amostragem dos vários parâmetros de qualidade assume-se como a principal medida a implementar na ETAR com vista a uma modelação mais sólida da digestão anaeróbia.
- As RNA possuem potencial para modelar a digestão anaeróbia e a sua utilização como ferramenta de apoio na operação e gestão deste tipo de processos afigura-se como uma solução promissora, não apenas no tratamento de águas residuais, mas também noutros campos da engenharia e das ciências, no geral.

5.2 Sugestões e recomendações futuras

Uma maior e mais regular frequência de amostragem de parâmetros relacionados com a qualidade das lamas afluentes e das lamas no interior do digestor poderá contribuir, não apenas para uma modelação mais sólida e precisa dos processos anaeróbios, mas também para monitorizar as condições no interior do digestor e alertar quanto a possíveis problemas que possam surgir. A medição de parâmetros de qualidade do biogás produzido constitui uma medida que se enquadra nesta abordagem de gestão dos processos anaeróbios, e seria interessante a realização de estudos direccionados para a optimização do metano produzido através da utilização de ferramentas de modelação, como o NeuralTools[®].

Também o desfasamento de variáveis se assume como uma opção válida a considerar na modelação através de redes neuronais. Embora no presente estudo os resultados alcançados com esta abordagem não tenham sido conclusivos, admite-se que a utilização de séries de dados mais sólidas e com maior frequência de amostragem possa contribuir para a redução do erro nas previsões e para a obtenção de resultados mais favoráveis. O número de combinações de desfasamento possíveis entre variáveis leva a crer que esta temática terá ainda muita matéria por abordar.

Resta ainda fazer uma referência à microbiologia da digestão anaeróbia e à sua importância em todo o processo. Neste trabalho, foram abordados os aspetos microbiológicos teóricos fundamentais deste tipo de digestão e relacionados com alguns parâmetros físicos e químicos, não tendo, contudo, sido feita referência à sua possível integração na modelação dos processos através das redes neuronais. Segundo Lauwers. *et al.* (2013), acredita-se que os desenvolvimentos futuros da modelação dos processos anaeróbios se irão centrar no aspeto microbiológico, relacionando matematicamente o desempenho da digestão com a diversidade e atividade microbiana. Como tal, investigações futuras que pretendam abordar a temática da modelação deste tipo de processos poderão incluir os aspetos microbiológicos e relacioná-los com o desempenho do digestor e com as restantes variáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

AdP (2014), Águas de Portugal, site disponível online em: <http://www.adp.pt/>; consultado em 15/08/2014.

Amado, M. P. (2009), *Planeamento Urbano Sustentável*, 3ª edição, Caleidoscópico, p.14.

Amaral, L. (2012), Digestão Anaeróbia, documento de apoio à cadeira de Drenagem e Tratamento de Águas Residuais, 2012/2013, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

Batstone, D. J., Keller, J., Angelidaki, I., Kalyuzhnyi, S.V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T.M., Siegrist H., Vavilin, V.A. (2002), The IWA Anaerobic Digestion Model No 1 (ADM1), *Water Science and Technology*, Vol 45 (nº 10), pp. 65-73.

Bitton, G. (1999), *Wastewater Microbiology*, Wiley Series in Ecological and Applied Microbiology, 2ª edição, Wiley-Liss, EUA, pp. 282-286.

Blumensaata, F., Keller, J. (2005), Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No. 1 (ADM1), *Water Research*, 39, pp. 171-183.

Casey, T. J. (1997), *Unit Treatment Processes in Water and Wastewater Engineering*, John Wiley & Sons, Inglaterra, pp. 207/209/216.

CCE - Centro para a Conservação da Energia (2000), *Guia Técnico do Biogás*, Centro para a Conservação da Energia, Lisboa, pp. 9-11.

Chernicharo, C. A. L. (2007), *Anaerobic Reactors*, Biological Wastewater Treatment Series, Vol. 4, IWA Publishing, Inglaterra, pp. 2/8/13.

Elmolla, E. S., Chaudhuri, M. (2011), The Use of Artificial Neural Network (ANN) for Modelling, Simulation and Prediction of Advanced Oxidation Process Performance in Recalcitrant Wastewater Treatment, 6, in: Hui, C. L. P. (Ed), *Artificial Neural Networks: Application*, InTech, EUA, pp. 105-122, disponível online em: <http://www.intechopen.com/books/artificial-neural-networks-application>; consultado em 13/08/2014.

FERROVIAL-AGROMAN/CADAGUA (2007), *Memória Descritiva e Justificativa do Projeto de Execução da ETAR do Seixal*, Vol. 1 – Processo de Tratamento, SIMARSUL, pp. 15.

Grady, Jr. C. P. L., Daigger, G. T., Love, N. G., Filipe, C. D. M. (1999), *Biological Wastewater Treatment*, Taylor & Francis Group, EUA, pp. 634.

Gray, N. F. (2004), *Biology of Wastewater Treatment*, Series on Environmental Science and Management, Vol. 4, 2ª edição, Imperial College Press, EUA, pp. 754-777.

Haykin, S. (1999), *Neural Networks: A Comprehensive Foundation*, 2ª edição, Prentice Hall, EUA, pp. 1-23.

Henze, M., Harremoës, P., Jansen, E. A. L. C. (2002), *Wastewater Treatment: Biological and Chemical Processes*, 3ª edição, Springer-Verlag, Alemanha, pp. 116-125/299-324.

Holubar, P., Zani, L., Hager, M., Fröschl, W., Radak, Z., Braun, R. (2002), Advanced controlling of anaerobic digestion by means of hierarchical neural networks, *Water Research*, 36, pp. 2582-2588.

INAG (2010a), Portal da Água, disponível online em:

<http://www.portaldagua.org/PT/InfoTecnica/Directiva/DQA/Pages/Default.aspx>; consultado em 7/07/2014.

INAG (2010b), Portal da Água, disponível online em:

<http://www.portaldagua.org/PT/InfoTecnica/Legislacao/APRH/Pages/default.aspx>; consultado em 10/07/2014.

Kartalopoulos, S. V. (1996), *Understanding Neural Networks and Fuzzy Logic: Basic Concepts and Applications*, Institute of Electrical and Electronics Engineers, EUA, pp. 37-53/61-70/75-77.

Kasabov, N. K. (1998), *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, 2ª edição, Massachusetts Institute of Technology, EUA.

Lauwers, J., Appels, L., Thompson, I. P., Degève, J., Van Impe, J. F., Dewil, R. (2013), Mathematical modelling of anaerobic digestion of biomass and waste: Power and limitations, *Progress in Energy and Combustion Science*, 39, pp. 383-402.

Lee, M. Y., Suh, C. W., Ahn, Y. T., Shin, H. S. (2009), Variation of ADM1 by using temperature-phased anaerobic digestion (TPAD) operation, *Bioresource Technology*, 100, pp. 2816-2822.

Lingireddy, S.; Brion, G. M. (2005), *Artificial Neural Networks in Water Supply Engineering*, Water Quality Measurements Series, American Society of Civil Engineers, EUA, pp. 1-22/36-46.

Malina, Jr. J. F., Pohland, F. G. (1992), *Design of Anaerobic Processes for the Treatment of Industrial and Municipal Wastes*, Water Quality Management Library, Vol. 7, Technomic Publishing Company, EUA, pp. 1-33/167-210.

Matos, J. P. (2008), *Utilização de Redes Neurais Artificiais para a Modelação da Degradação de Sistemas de Distribuição de Água*, Dissertação de Mestrado em Engenharia Civil, Instituto Superior Técnico da Universidade Técnica de Lisboa, Lisboa, pp. 22-23.

Metcalf & Eddy (1991), *Wastewater engineering – Treatment, Disposal and Reuse*, 3ª edição, McGraw-Hill, EUA, pp. 1-12/47-94/420-427/813-834.

Metcalf & Eddy (2003), *Wastewater engineering – Treatment, Disposal and Reuse*, 4ª edição, McGraw-Hill, EUA, pp. 1-12/629-635/984-999/1505-1525.

Ministério do Ambiente, do Ordenamento do Território e do Desenvolvimento Regional (2007), *PEAASAR II: Plano Estratégico de Abastecimento de Água e de Saneamento de Águas residuais 2007-2013*, p. 124.

Palisade (2012), *Guide to Using NeuralTools: Neural Network Add-In for Microsoft Excel*, versão 6, Palisade Corporation, EUA.

Parker, W. J. (2005), Application of the ADM1 model to advanced anaerobic digestion, *Bioresource Technology*, 96, pp. 1832-1842.

Pestman, W. R. (1998), *Mathematical Statistics: An Introduction*, Walter de Gruyter GmbH, EUA, pp. 28-29/201-203.

Qasim, S. R. (1999), *Wastewater Treatment Plants: Planning, Design, and Operation*, 2ª edição, CRC Press LLC, EUA.

Quevauviller, P., Thomas, O., Beken, A. V. D. (2006), *Wastewater Quality Monitoring and Treatment*, John Wiley & Sons, Inglaterra, pp 2-4.

Ramirez, I., Volcke, E. I. P., Rajinikanth, R., Steyer, J. P. (2009), Modeling microbial diversity in anaerobic digestion through an extended ADM1 model, *Water Research*, 43, pp. 2787-2800.

Rocha, H. F. (2013), *Utilização de Redes Neurais Artificiais na gestão de processos de Digestão Anaeróbia*, Dissertação de Mestrado em Engenharia Civil – Ramo Gestão da Construção, ISEP, Porto.

- Russell, D.L. (2006), *Practical Wastewater Treatment*, John Wiley & Sons, EUA, pp. 203-210.
- SIMARSUL (2014), Sistema Integrado Multimunicipal de Águas Residuais da Península de Setúbal, site disponível online em: <http://www.simarsul.pt/>; consultado em 14/07/2014.
- Spellman, F. R. (2003), *Handbook of Water and Wastewater Treatment Plant Operations*, CRC Press LLC, EUA, pp 620.
- Strik, D. P. B. T. B., Domnanovich, A. M., Zani, L., Braun, R., Holubar, P. (2005), Prediction of trace compounds in biogas from anaerobic digestion using the MATLAB Neural Network Toolbox, *Environmental Modelling & Software*, 20, pp. 803–810.
- Taylor, J. G. (1996), *Neural Networks and Their Applications*, John Wiley & Sons, Inglaterra, pp. 1-4/24.
- Van Haandel, A. C., Lettinga, G. (1994), *Anaerobic Sewage Treatment: A Practical Guide for Regions with a Hot Climate*, John Wiley & Sons, Inglaterra, pp. 1-47/121-146.
- Vemuri, V. R. (1992), *Artificial Neural Networks: Concepts and Control Applications*, Institute of Electrical and Electronics Engineers, EUA, pp. 1-6.
- Vesilind, P. A. (2003), *Wastewater Treatment Plant Design*, Water Environment Federation, Inglaterra.
- Vieira, M. C. (2014), A Internacionalização do Setor Português da Água, in: *Oportunidades de Negócio nos Mercados de Croácia, Polónia e Sérvia para as Empresas Portuguesas do Setor da Água*, Águas do Douro e Paiva, Vila Nova de Gaia, 25 de fevereiro de 2014, AEP em colaboração com Parceria Portuguesa para a Água, pp. 22-23/43.
- Von Sperling, M. (2007), *Wastewater Characteristics, Treatment and Disposal*, Biological Wastewater Treatment Series, Vol. 1, IWA Publishing, Inglaterra, pp. 4/30-33/256-258.
- Wiesmann, U., Choi, I. S., Dombrowski, E. M. (2007), *Fundamentals of Biological Wastewater Treatment*, Wiley-VCH, Alemanha, pp. 169-172.
- Wijen, F., Zoeteman, K., Pieters, J., Van Seters, P. (2005), *A Handbook of Globalisation and Environmental Policy: National Government Interventions in a Global Arena*, Edward Elgar Publishing, Inglaterra, pp. 5-6/20.

Wilcox, R. R. (2001), *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, Springer-Verlag, EUA, pp. 110/179-180.

APÊNDICES

APÊNDICE I

Relatório detalhado do treino e do teste

NeuralTools (Report: Neural Net Training and Auto-Testing)	
Created for: user	
Date: quinta-feira, 10 de Julho de 2014	
Summary	
Net Information	
Name	Net Trained on Conjunto de dados núm. 1-A
Configurations Included in Search	GRNN, MLFN 2 to 30 nodes
Best Configuration	GRNN Numeric Predictor
Location	This Workbook
Independent Category Variables	0
Independent Numeric Variables	2 (Q lamas (m ³ /dia), T reator (°C))
Dependent Variable	Numeric Var. (Q biogás (m ³ /dia))
Training	
Number of Cases	496
Training Time (h:min:sec)	00:00:02
Number of Trials	67
Reason Stopped	Auto-Stopped
% Bad Predictions (30% Tolerance)	46,3710%
Root Mean Square Error	222,16
Mean Absolute Error	176,10
Std. Deviation of Abs. Error	135,44
Testing	
Number of Cases	124
% Bad Predictions (30% Tolerance)	53,2258%
Root Mean Square Error	236,33
Mean Absolute Error	193,48
Std. Deviation of Abs. Error	135,70
Data Set	
Name	Conjunto de dados núm. 1-A
Number of Rows	620
Manual Case Tags	NO
Variable Impact Analysis	
T reator (°C)	61,1927%
Q lamas (m ³ /dia)	38,8073%

Figura I.1 – Exemplo de um relatório detalhado do treino e do teste criado pelo NeuralTools®
(continua)

<i>Best Net Search</i>	RMS Error	Training Time (h:min:sec)	Reason Training Stopped
GRNN	236,33	00:00:02	Auto-Stopped
MLFN 2 nodes	249,43	00:01:30	Auto-Stopped
MLFN 3 nodes	247,22	00:01:30	Auto-Stopped
MLFN 4 nodes	254,18	00:01:30	Auto-Stopped
MLFN 5 nodes	253,46	00:01:30	Auto-Stopped
MLFN 6 nodes	260,48	00:01:30	Auto-Stopped
MLFN 7 nodes	255,04	00:01:30	Auto-Stopped
MLFN 8 nodes	252,40	00:01:30	Auto-Stopped
MLFN 9 nodes	257,66	00:01:30	Auto-Stopped
MLFN 10 nodes	242,84	00:01:30	Auto-Stopped
MLFN 11 nodes	257,02	00:01:30	Auto-Stopped
MLFN 12 nodes	245,08	00:01:30	Auto-Stopped
MLFN 13 nodes	264,80	00:01:30	Auto-Stopped
MLFN 14 nodes	254,25	00:01:30	Auto-Stopped
MLFN 15 nodes	261,89	00:01:30	Auto-Stopped
MLFN 16 nodes	265,91	00:01:30	Auto-Stopped
MLFN 17 nodes	266,14	00:01:30	Auto-Stopped
MLFN 18 nodes	264,12	00:01:30	Auto-Stopped
MLFN 19 nodes	266,00	00:01:30	Auto-Stopped
MLFN 20 nodes	269,84	00:01:30	Auto-Stopped
MLFN 21 nodes	273,33	00:01:30	Auto-Stopped
MLFN 22 nodes	262,71	00:01:30	Auto-Stopped
MLFN 23 nodes	263,81	00:01:30	Auto-Stopped
MLFN 24 nodes	272,60	00:01:30	Auto-Stopped
MLFN 25 nodes	261,12	00:01:30	Auto-Stopped
MLFN 26 nodes	264,63	00:01:30	Auto-Stopped
MLFN 27 nodes	271,00	00:01:30	Auto-Stopped
MLFN 28 nodes	265,62	00:01:30	Auto-Stopped
MLFN 29 nodes	267,57	00:01:30	Auto-Stopped
MLFN 30 nodes	267,71	00:01:30	Auto-Stopped

Figura I.1 – (continuação)

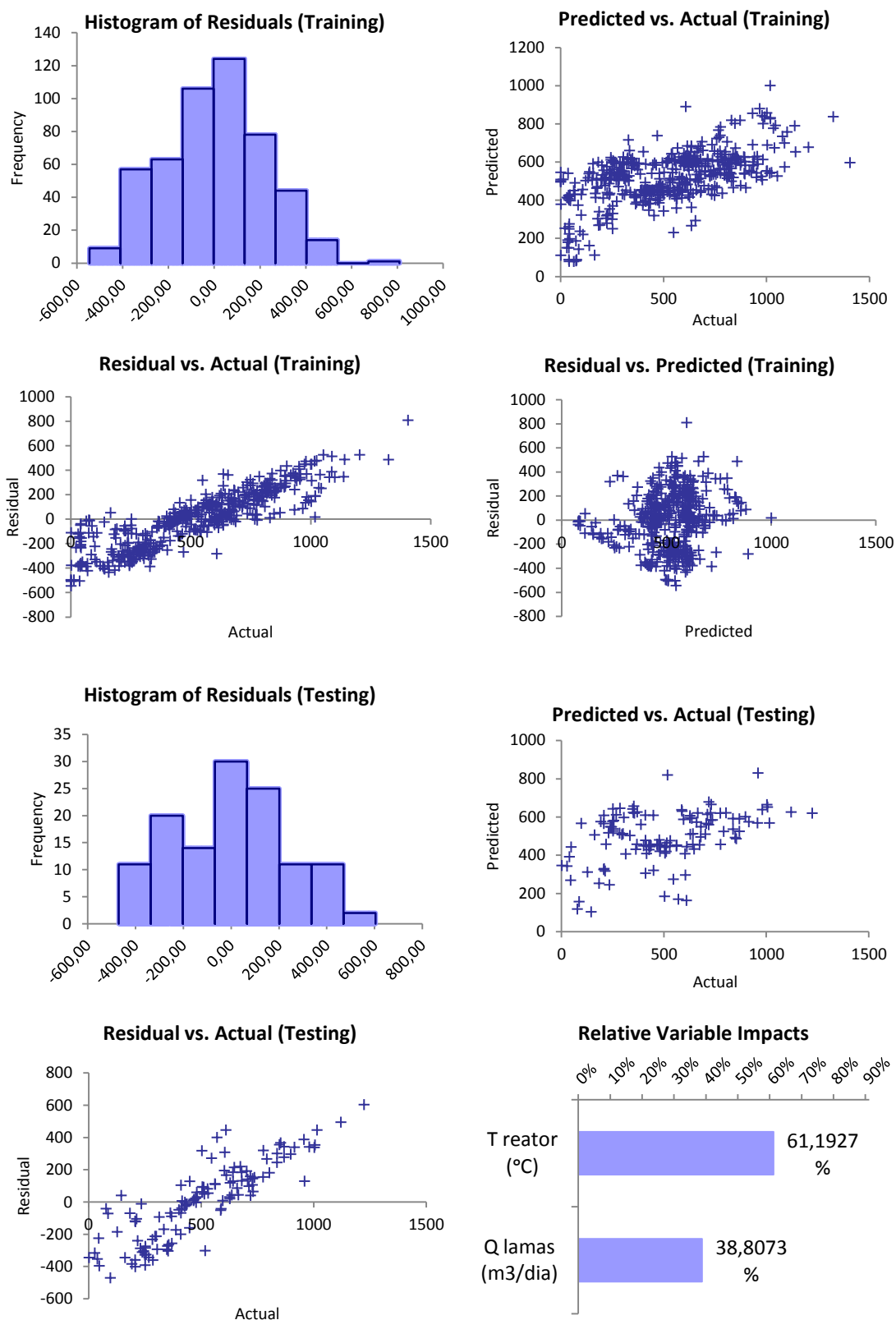


Figura I.2 – Gráficos inseridos no relatório detalhado do treino e do teste do NeuralTools®.

APÊNDICE II

Localização dos pontos de amostragem das variáveis

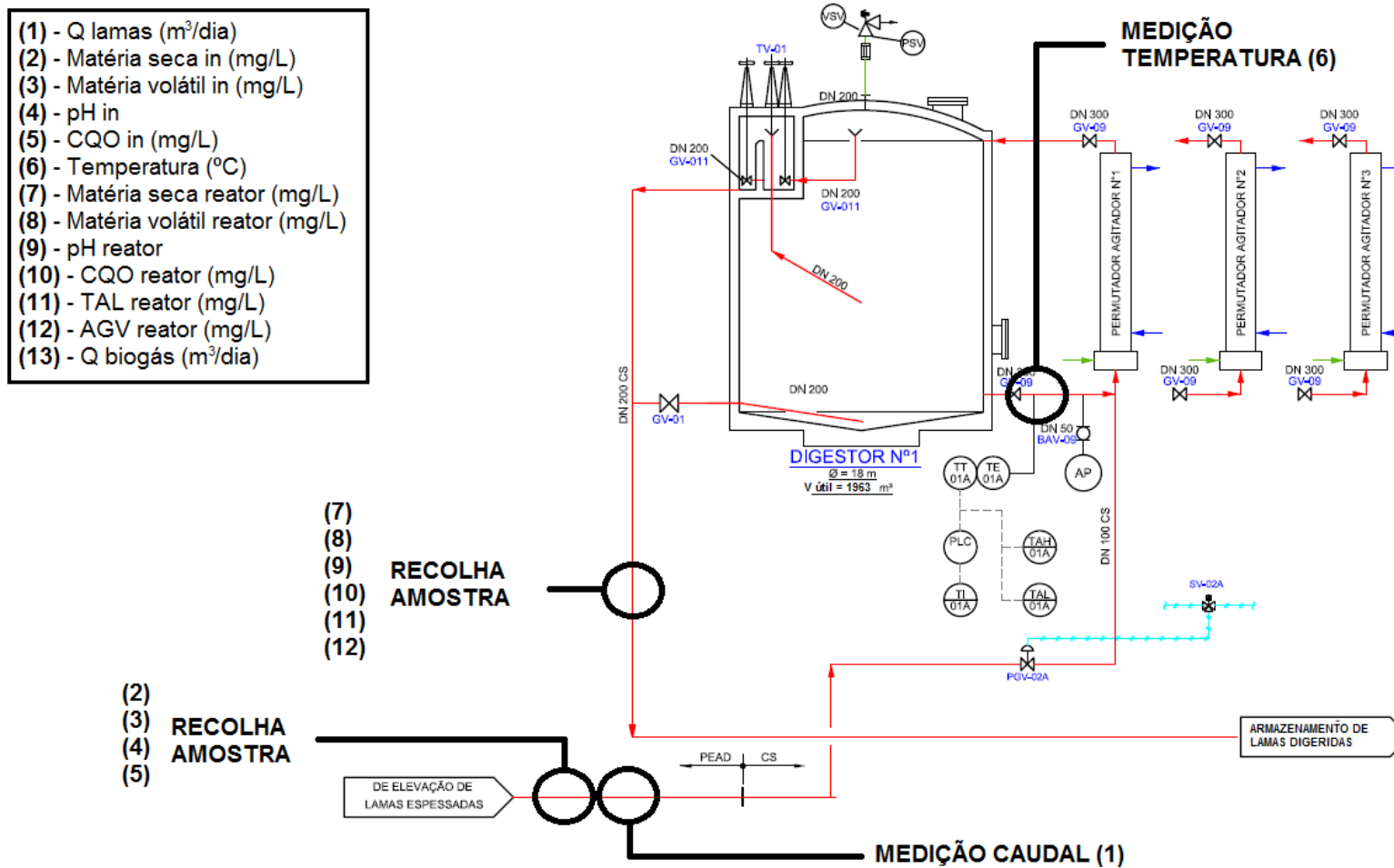


Figura II.1 – Localização dos pontos de amostragem de alguns parâmetros da digestão anaeróbia.

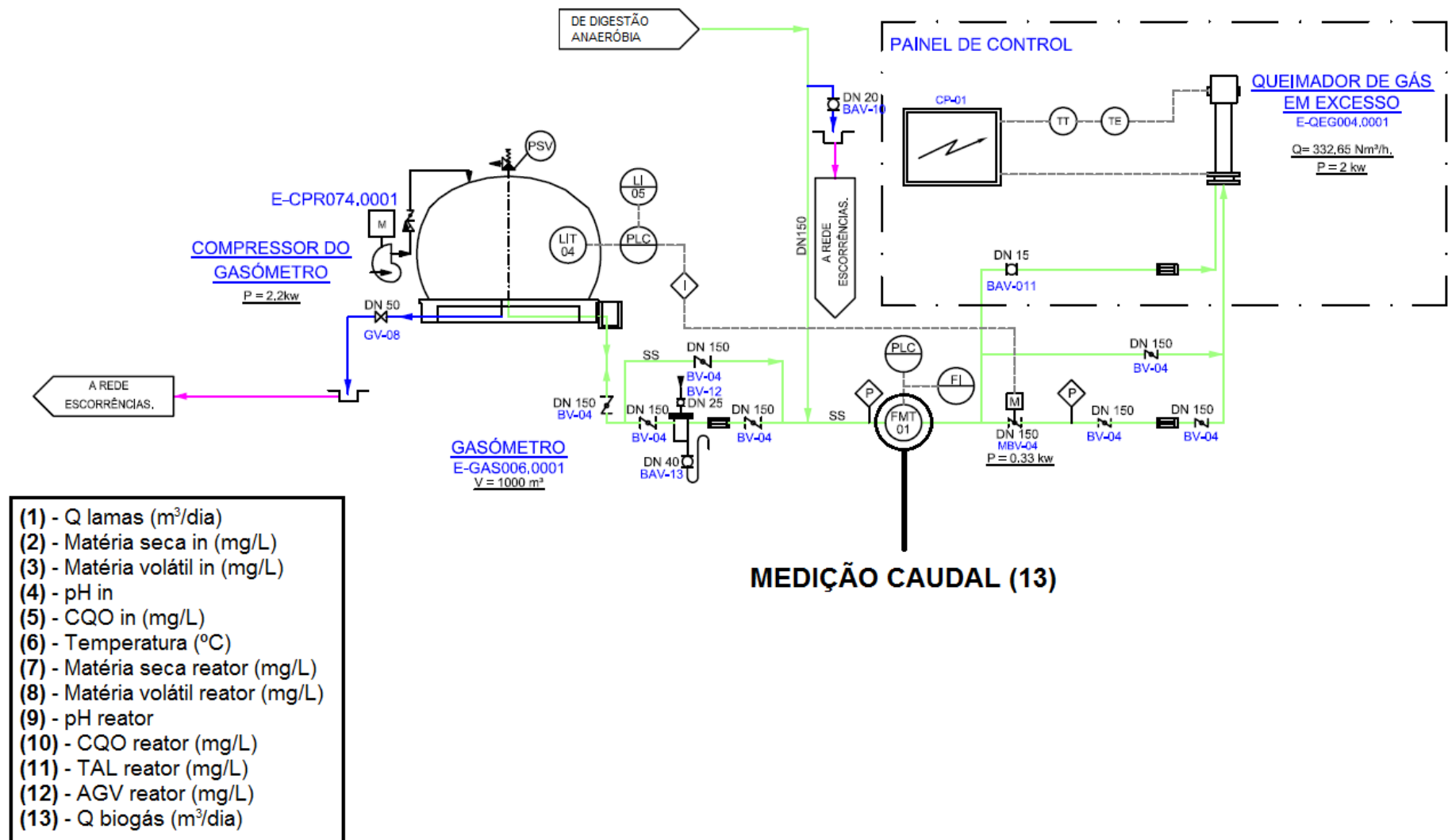


Figura II.2 – Localização do ponto de amostragem da variável “Q biogás”.

APÊNDICE III

Descrição das variáveis

pH

O pH é medido a partir da recolha de amostras à entrada do digestor (pH in) e à saída (pH reator). Como se trata de um parâmetro de qualidade das lamas, a frequência de amostragem é inferior a uma vez por semana. A série de dados do pH das lamas afluentes ao digestor selecionada para treino/teste corresponde ao período de 29 de março de 2012 até 20 de maio de 2014, e a série de dados do pH das lamas digeridas corresponde ao período de 15 de março de 2012 até 20 de maio de 2014 (Figura III.1).

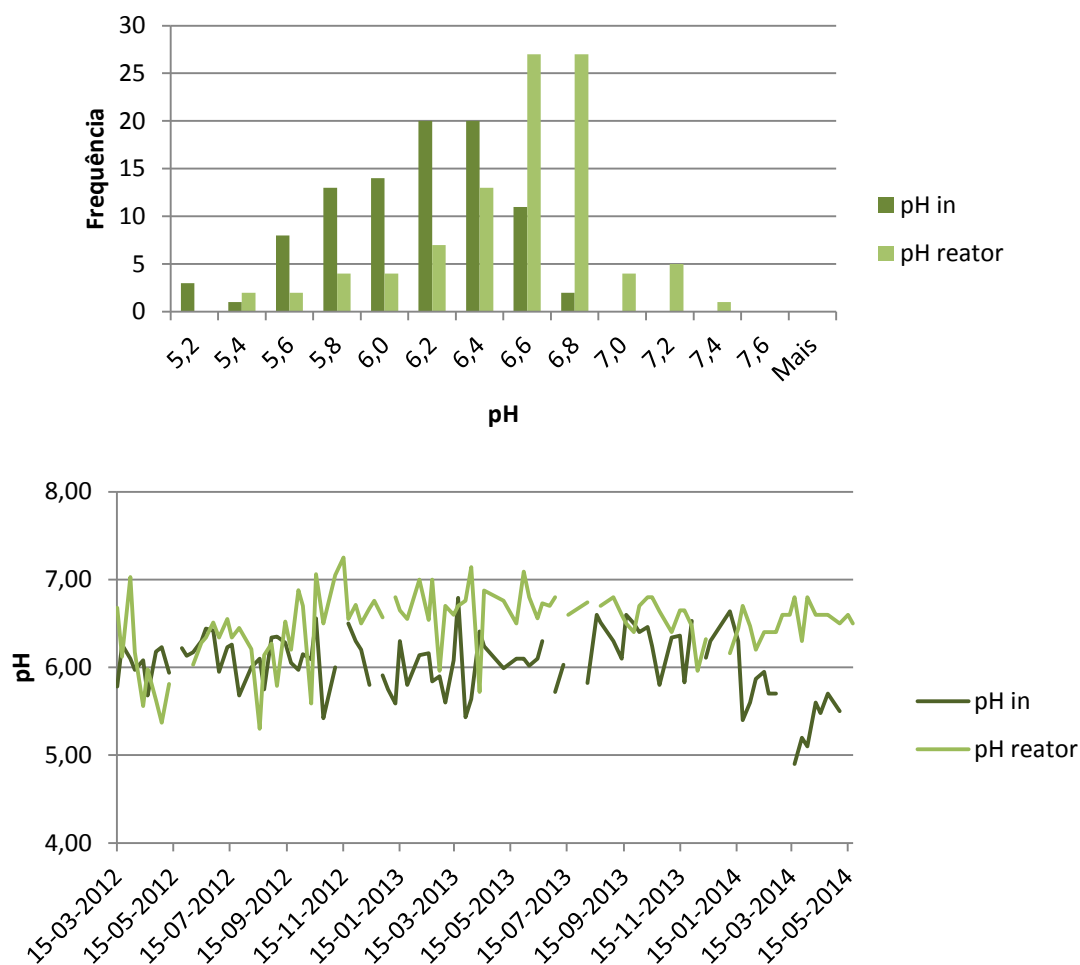


Figura III.1 – Histograma dos valores de pH medidos à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais dos valores destas variáveis (em baixo).

Embora o intervalo ótimo da metanogénese ocorra para valores de pH compreendidos entre 6,6 e 7,6 (Metcalf & Eddy, 1991), o funcionamento normalizado de um digestor é conseguido com valores de pH entre 6,2 e 7,8 (Gray, 2004). Na Figura III.1 é possível verificar, que o digestor opera usualmente nesta gama, mesmo com a elevação frequente de lamas ácidas. No entanto, cerca de 20% das amostragens revelaram valores de pH no digestor

inferiores a 6,2, o que poderá ter influenciado negativamente o seu desempenho nos períodos correspondentes.

Carência química de oxigênio (CQO)

Este parâmetro representa a quantidade de oxigênio necessária para oxidar a matéria orgânica (Qasim, 1999), e é medido nas lamas afluentes à digestão anaeróbia (CQO in) e nas lamas digeridas (CQO reator). Como se trata de um parâmetro de qualidade, a frequência de amostragem é inferior a uma vez por semana. As séries de dados da CQO das lamas afluentes ao digestor e das lamas digeridas correspondem ao período de 19 de janeiro de 2012 até 20 de maio de 2014 (Figura III.2).

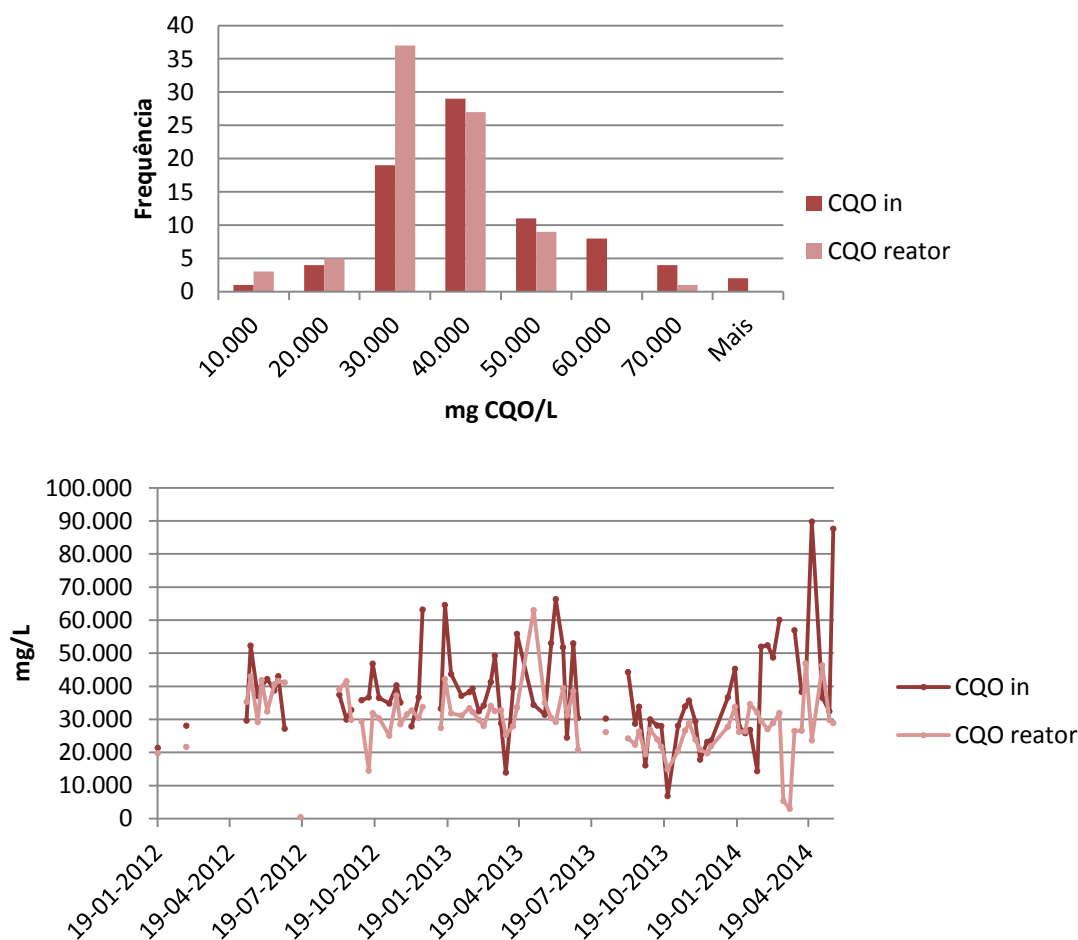


Figura III.2 – Histograma das concentrações de CQO medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais dos valores destas variáveis (em baixo).

Segundo Metcalf & Eddy (2003), concentrações de CQO superiores a 2.000 mg/L são necessárias para produzir metano em quantidade suficiente para que o aquecimento das lamas ocorra sem uma fonte de combustível externa. Na Figura III.2 é visível que, no período em estudo, a CQO foi consideravelmente superior a 2.000 mg/L.

Uma vez que as variáveis “CQO in” e “CQO reator” representam a concentração da carência química de oxigênio nas lamas, em mg/L, procedeu-se ainda ao cálculo das respetivas cargas. Assim, multiplicando o caudal de lamas pela concentração da CQO nas lamas elevadas para digestão e nas lamas digeridas, obtém-se a carga da CQO afluente à digestão (LCQO in) e a carga da CQO nas lamas digeridas (LCQO out), respetivamente. As séries das cargas de CQO possuem a mesma extensão que as séries das respetivas concentrações, e os valores são expressos em kg/dia (Figura III.3).

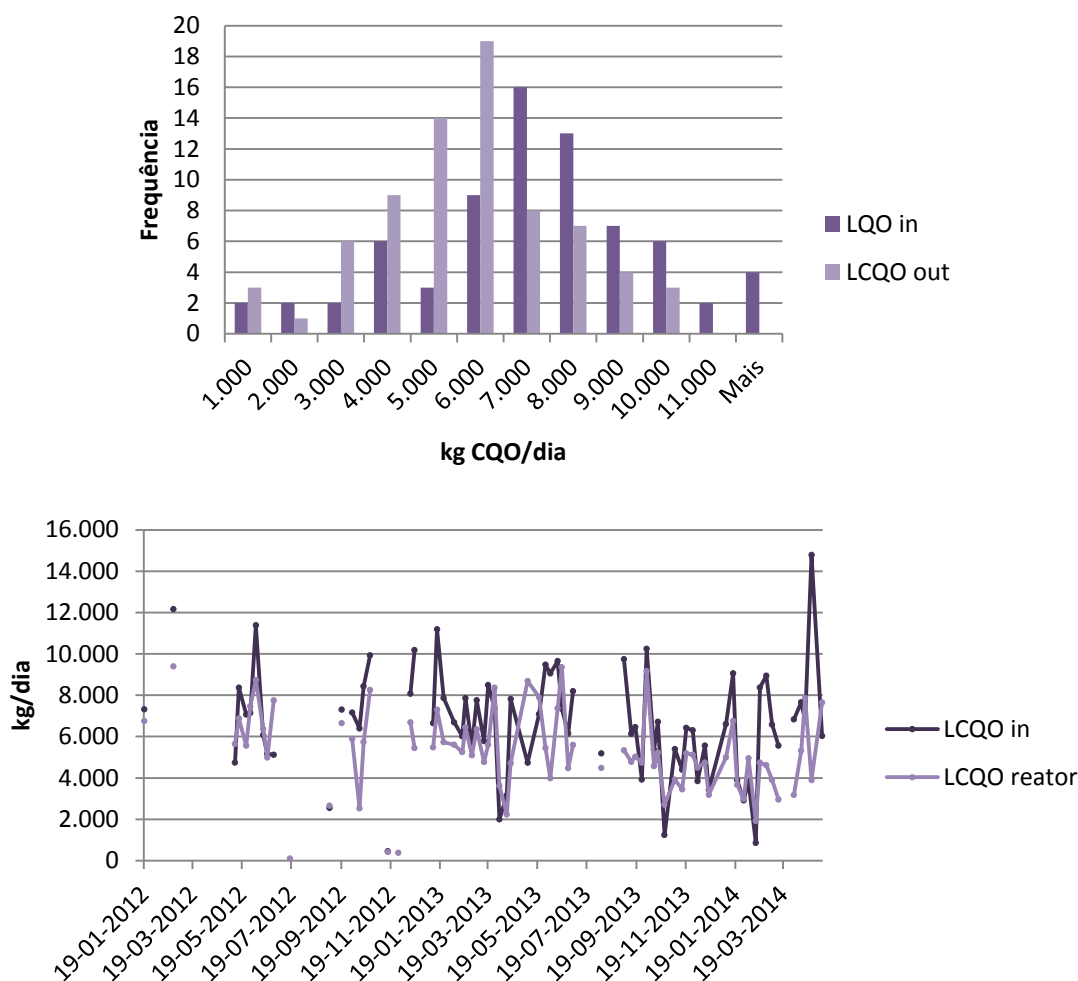


Figura III.3 – Histograma das cargas de CQO afluente e efluente ao digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais dos valores destas variáveis (em baixo).

A partir das séries das cargas de CQO nas lamas elevadas para digestão e nas lamas digeridas, foi também possível criar uma série com as cargas de CQO removida no digestor (LCQO removida), subtraindo à variável “LCQO in” os valores da variável “LCQO out” do mesmo dia. É importante mencionar que esta variável não representa a carga de CQO removida durante os processos anaeróbios de um determinado dia, mas sim o balanço das cargas de CQO no digestor nesse mesmo dia. A sua integração no estudo sob a forma de

variável teve como principal propósito contribuir para avaliar o impacto da CQO no desempenho do digester. A dimensão da série criada é inferior à das séries a partir das quais foi determinada devido à exclusão dos valores negativos. Na Figura III.4 é apresentado um histograma das cargas de CQO removida.

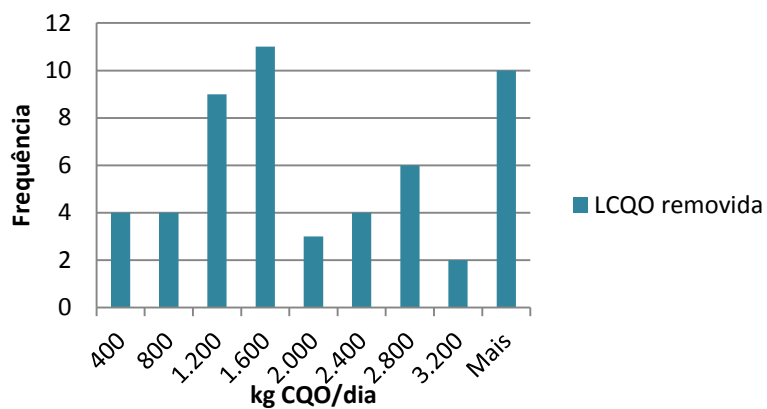


Figura III.4 – Histograma das cargas de CQO removida no digester, correspondentes ao período entre 2012 e 2014.

Matéria seca (MS)

Esta variável representa a concentração dos sólidos totais nas lamas e é medida a partir da recolha de amostras das lamas elevadas para digestão (MS in) e das lamas digeridas (MS reator). A frequência de amostragem da matéria seca das lamas elevadas para digestão e das lamas digeridas é de uma vez por semana entre 3 de janeiro de 2012 e 20 de maio de 2014 (Figura III.5).

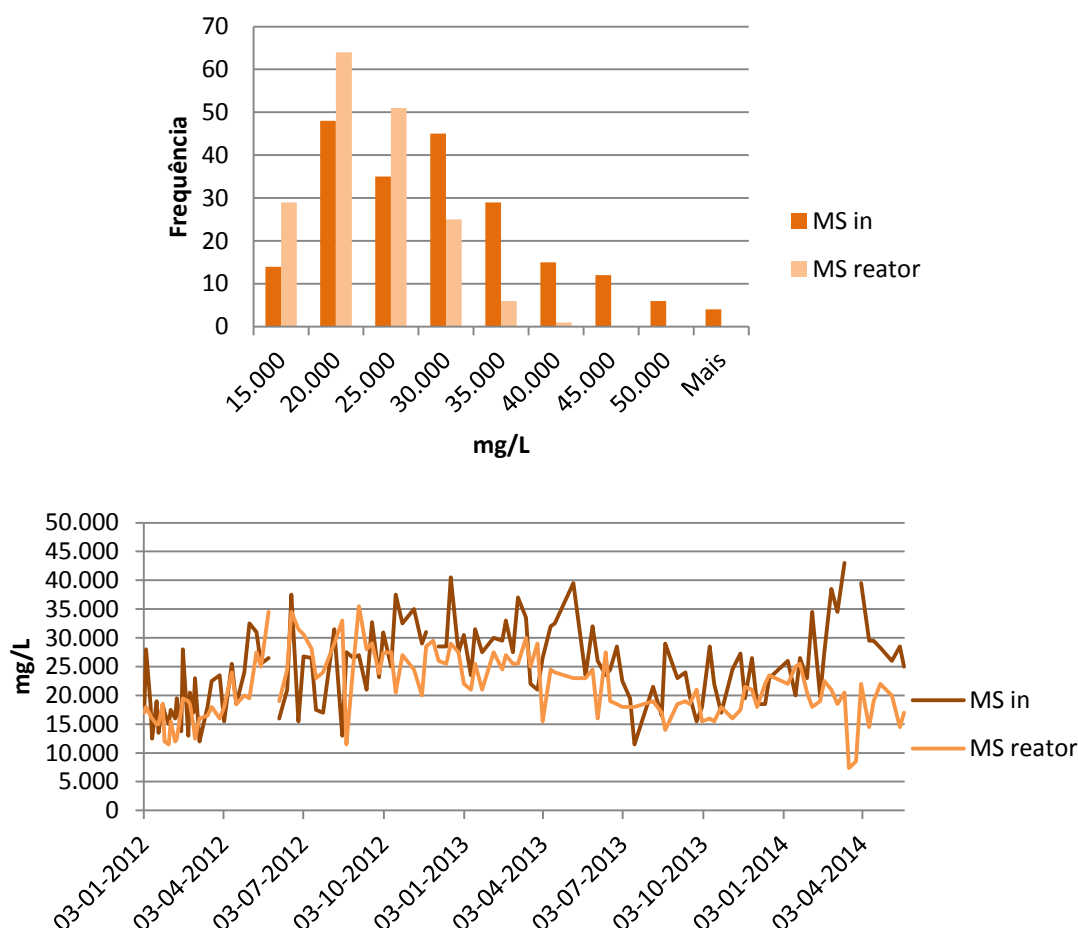


Figura III.5 – Histograma das concentrações de matéria seca medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais dos valores destas variáveis (em baixo).

Uma vez que as variáveis “MS in” e “MS reator” representam a concentração da matéria seca nas lamas, em mg/L, procedeu-se ainda ao cálculo das respetivas cargas. Assim, multiplicando o caudal de lamas pela concentração da matéria seca nas lamas elevadas para digestão e pelas lamas digeridas, obtém-se a carga da matéria seca afluyente à digestão (LMS in) e a carga da matéria seca nas lamas digeridas (LMS out), respetivamente. As séries das cargas de matéria seca possuem a mesma extensão que as séries das respetivas concentrações, e os valores são expressos em kg/dia (Figura III.6).

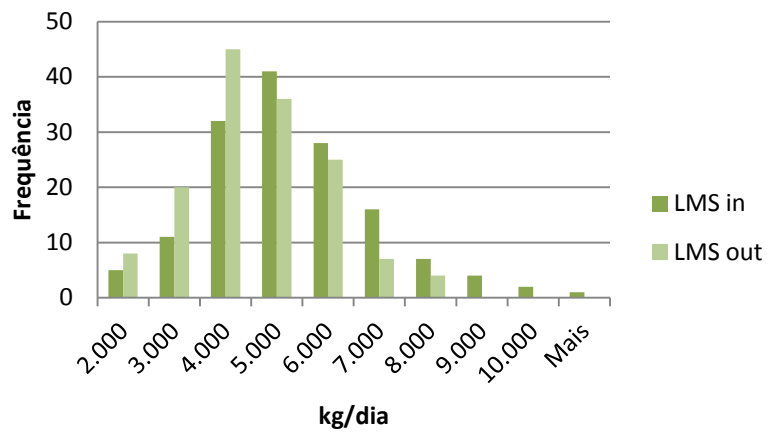


Figura III.6 – Histograma das cargas de matéria seca afluentes e efluentes ao digestor, correspondentes ao período entre 2012 e 2014.

Matéria volátil (MV)

Este parâmetro mede a concentração dos sólidos voláteis nas lamas e é determinado a partir da recolha de amostras das lamas elevadas para digestão (MV in) e das lamas digeridas (MV reator). A frequência de amostragem da matéria volátil das lamas elevadas para digestão e nas lamas digeridas é de uma vez por semana entre 3 de janeiro de 2012 e 20 de maio de 2014 (Figura III.7).

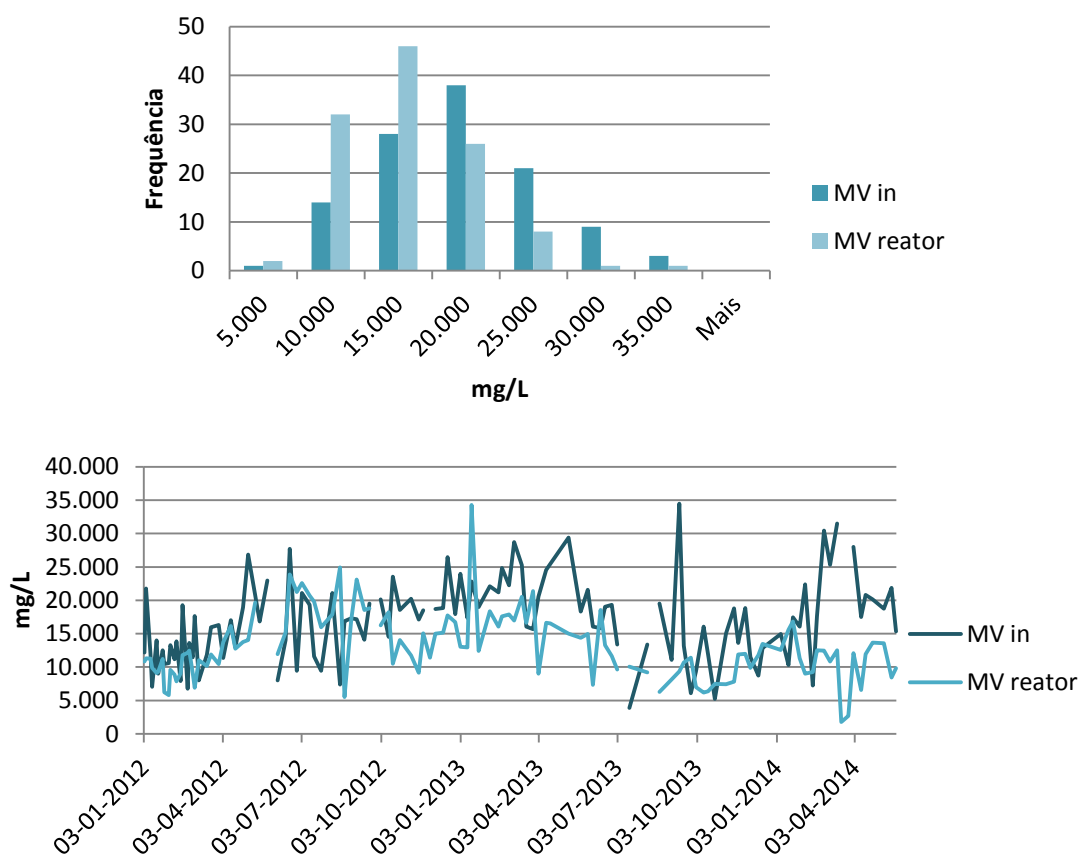


Figura III.7 – Histograma das concentrações de matéria volátil medidas à entrada e à saída do digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais dos valores destas variáveis (em baixo).

Como as variáveis “MV in” e “MV reator” representam a concentração da matéria volátil nas lamas, em mg/L, procedeu-se ainda ao cálculo das respetivas cargas. Multiplicando o caudal de lamas pela concentração da matéria volátil nas lamas elevadas para digestão e nas lamas digeridas, obtêm-se a carga da matéria volátil afluente à digestão (LMV in) e a carga da matéria volátil nas lamas digeridas (LMV out), respetivamente. As séries das cargas de matéria volátil possuem a mesma extensão que as séries das respetivas concentrações, e os valores são expressos em kg/dia. Na Figura III.8 é apresentado um histograma das cargas de matéria volátil.

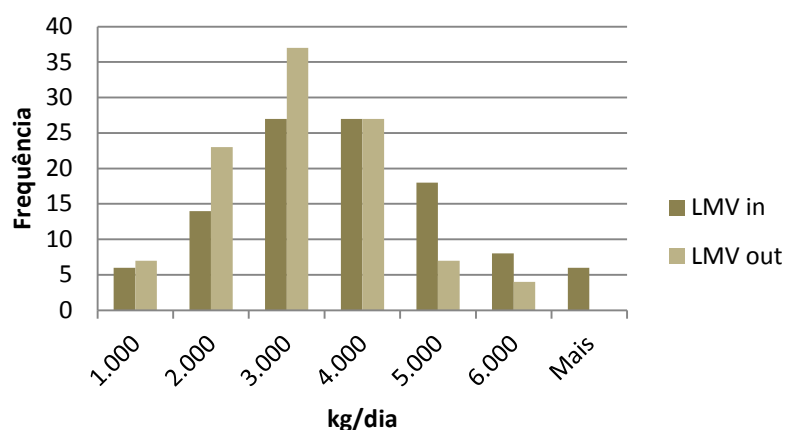


Figura III.8 – Histograma das cargas de matéria volátil afluentes e efluentes ao digestor, correspondentes ao período entre 2012 e 2014.

A partir das séries das cargas de matéria volátil nas lamas elevadas para digestão e nas lamas digeridas, foi também possível criar uma série com as cargas de matéria volátil removida na digestão (LMV removida), subtraindo à variável “LMV in” os valores da variável “LMV out” do mesmo dia. É importante mencionar que, à semelhança da “CQO removida”, esta variável não representa a carga de LMV removida durante os processos anaeróbios de um determinado dia, mas sim o balanço das cargas de LMV no digestor nesse mesmo dia. A sua integração no estudo sob a forma de variável teve como principal propósito contribuir para avaliar o impacto da LMV no desempenho do digestor. A dimensão da série criada é inferior à das séries a partir das quais foi determinada devido à exclusão dos valores negativos. Na Figura III.9 é apresentado um histograma das cargas de LMV removida.

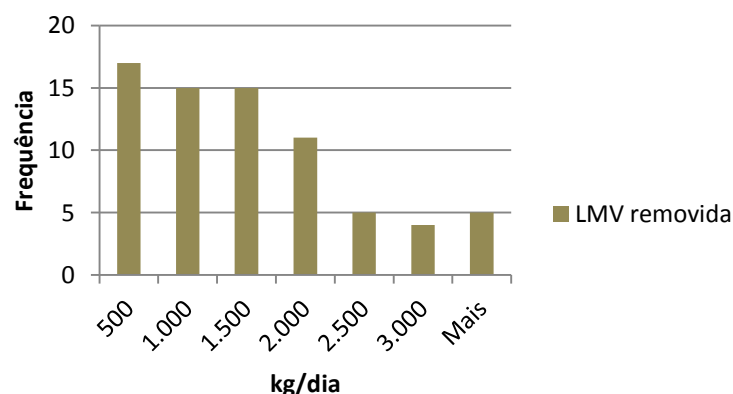


Figura III.9 – Histograma das cargas de LMV removida no digestor, correspondentes ao período entre 2012 e 2014.

Para além das variáveis “LMV in”, “LMV out” e “LMV removida”, foi ainda possível calcular a carga orgânica volúmica (CV). Esta variável é determinada a partir dos valores de “LMV in” e do volume útil do digestor (1963 m³), segundo a seguinte equação:

$$CV (kg \cdot m^{-3} \cdot d^{-1}) = \frac{LMV \text{ in}}{V_{\text{digestor}}} \quad (\text{Equação III.1})$$

Na Figura III.10, apresenta-se o histograma dos valores da carga volúmica orgânica.

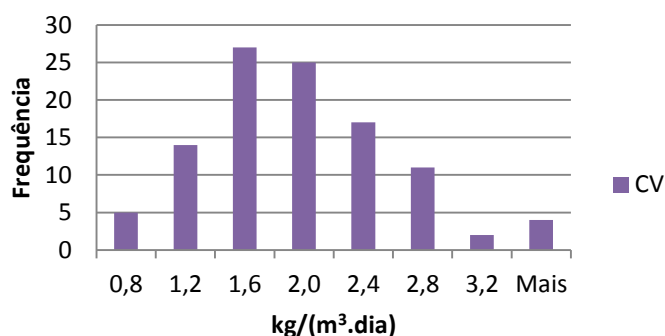


Figura III.10 – Histograma dos valores da carga volúmica orgânica, correspondentes ao período entre 2012 e 2014.

Segundo Vesilind (2003), os valores da carga volúmica ótima para digestores anaeróbios a funcionarem na gama mesófila devem situar-se entre 1,6 e 3,2 kg/(m³.dia). Na Figura III.10, observa-se a existência de alguns valores inferiores a este intervalo, o que sugere que a alimentação de sólidos voláteis verificada em alguns dias ficou aquém do que seria desejável para as dimensões do digestor.

Alcalinidade total (TAL)

Este parâmetro representa a alcalinidade total no digestor (TAL reator) sob a forma de CaCO₃. Como se trata de um parâmetro de qualidade, a frequência de amostragem é de uma vez a cada duas semanas entre 20 de março de 2012 e 20 de maio de 2014 (Figura III.11).

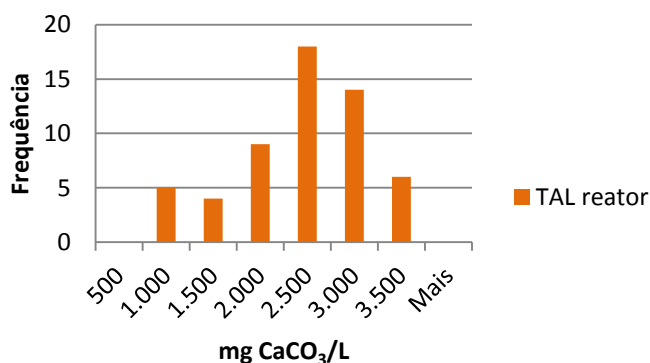


Figura III.11 – Histograma das concentrações de alcalinidade total no digestor, correspondentes ao período entre 2012 e 2014.

De acordo com Metcalf & Eddy (2003), deverá ser assegurada num digestor anaeróbio uma alcalinidade entre 2.000 e 5.000 mg CaCO₃/L. Na Figura III.11 é possível verificar que a maioria dos valores de alcalinidade total medidos no digestor se insere neste intervalo de referência, apesar de se observar a existência de alguns valores inferiores a 2.000 mg CaCO₃/L.

Ácidos gordos voláteis (AGV)

Este parâmetro representa a concentração de ácidos gordos voláteis no digestor (AGV reator) sob a forma de CH₃COOH. A frequência de amostragem é de uma vez a cada duas semanas entre 20 de março de 2012 e 20 de maio de 2014 (Figura III.12).

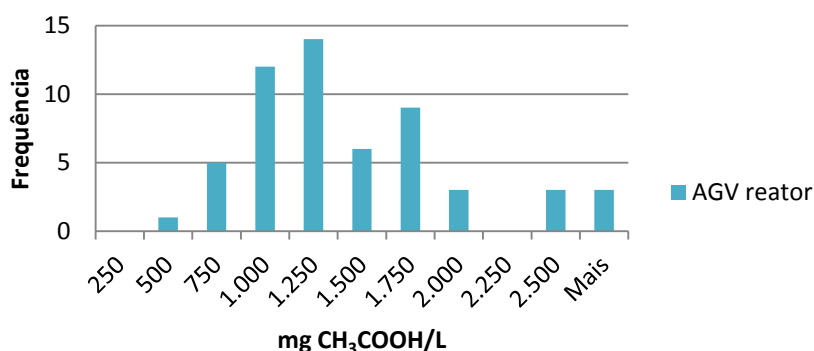


Figura III.12 – Histograma das concentrações de ácidos gordos voláteis no digestor, correspondentes ao período entre 2012 e 2014.

Para o funcionamento de um digestor em estado estacionário, deverá ser assegurada uma concentração de AGV entre 250 e 1.000 mg/L, sendo que concentrações superiores a 1.800-2.000 mg/L constituem, geralmente, indicadores da existência de problemas no digestor (Gray, 2004). Na Figura III.12, observa-se que 6 das 56 amostragens excederam os 2.000 mg/L, o que poderá indiciar a existência de problemas no digestor nos dias correspondentes.

Uma vez que a alcalinidade é largamente condicionada pela concentração dos ácidos gordos voláteis no digestor, procedeu-se ainda à criação de uma série de valores calculados através do quociente entre os AGV e a TAL. Segundo Spellman (2003), o quociente entre estas duas variáveis, num digestor a operar normalmente, varia entre 0,1 e 0,3. Deste modo, a concentração de ácidos gordos voláteis é reduzida e o risco de acidificação do digestor é minimizado. Na Figura III.13, observa-se em que o valor deste parâmetro é superior ao intervalo aconselhável, o que sugere que deveria ter sido adicionada mais alcalinidade, sob a forma de cal, às lamas elevadas para digestão.

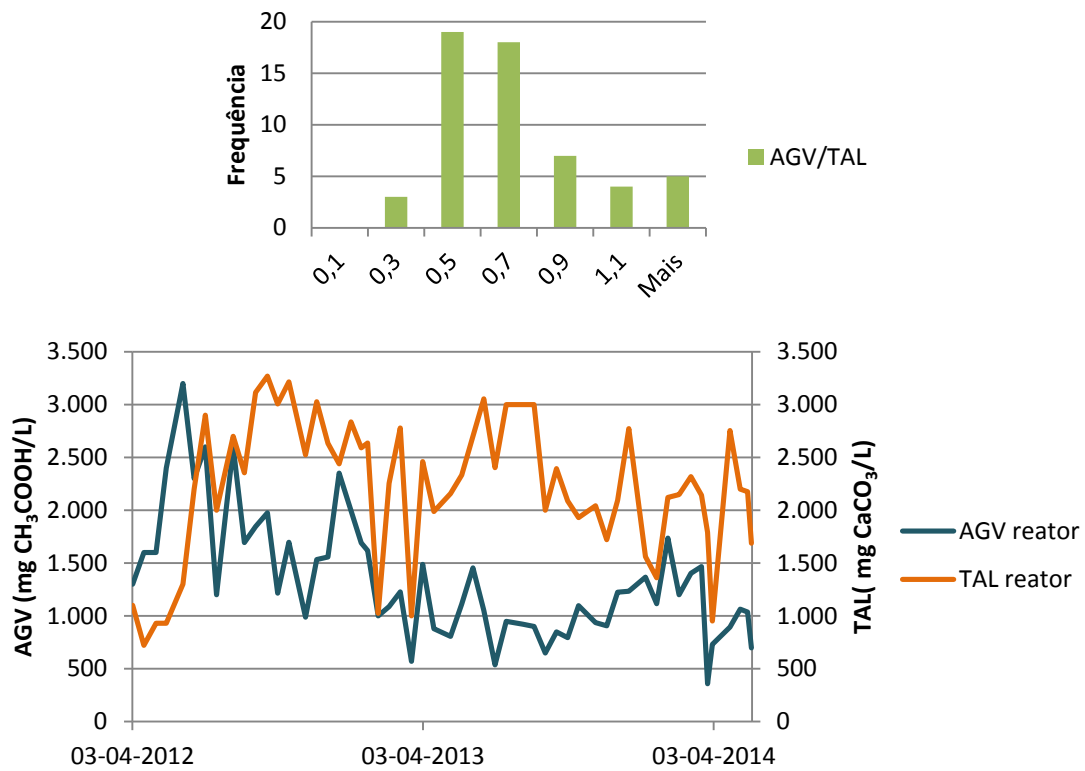


Figura III.13 – Histograma dos quocientes entre a alcalinidade total e os ácidos gordos voláteis no digestor, correspondentes ao período entre 2012 e 2014 (em cima), e séries temporais das variáveis “AGV reator” e “TAL reator” (em baixo).

APÊNDICE IV

Melhores redes neuronais criadas sem desfasamento

Quadro IV.1 – Indicadores de desempenho do treino/teste das melhores redes neurais criadas sem desfaseamento de variáveis (continua)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
1-A	T reator	°C	61,19%	620	Melhor rede	GRNN
	Q lamas	m ³ /dia	38,81%		% más previsões (treino - 30% toler.)	46,37%
					% más previsões (teste - 30% toler.)	53,23%
					Erro médio quadrático (treino)	222,16
				Erro médio quadrático (teste)	236,33	
2-A	T reator	°C	62,59%	620	Melhor rede	GRNN
	TRH	dias	32,89%		% más previsões (treino - 30% toler.)	49,60%
	Q lamas	m ³ /dia	4,52%		% más previsões (teste - 30% toler.)	53,23%
					Erro médio quadrático (treino)	221,04
				Erro médio quadrático (teste)	232,57	
3-C-3	T reator	°C	75,07%	69	Melhor rede	GRNN
	TRH	dias	24,30%		% más previsões (treino - 30% toler.)	25,45%
	LMV removida	kg/dia	0,59%		% más previsões (teste - 30% toler.)	42,86%
	Q lamas	m ³ /dia	0,04%		Erro médio quadrático (treino)	126,65
				Erro médio quadrático (teste)	240,69	
3-L-3	T reator	°C	58,03%	58	Melhor rede	MLFN (2 nós)
	TRH	dias	19,98%		% más previsões (treino - 30% toler.)	23,91%
	LMV removida	kg/dia	13,50%		% más previsões (teste - 30% toler.)	50,00%
	pH in	-	8,50%		Erro médio quadrático (treino)	140,48
				Erro médio quadrático (teste)	228,53	
3-N	pH reator	-	26,47%	58	Melhor rede	MLFN (4 nós)
	pH in	-	24,84%		% más previsões (treino - 30% toler.)	15,22%
	TRH	dias	24,15%		% más previsões (teste - 30% toler.)	50,00%
	Q lamas	m ³ /dia	14,20%		Erro médio quadrático (treino)	74,62
	LMV removida	kg/dia	10,34%		Erro médio quadrático (teste)	187,58
4-A-3	T reator	°C	35,90%	94	Melhor rede	GRNN
	TRH	dias	28,85%		% más previsões (treino - 30% toler.)	42,67%
	LMV in	kg/dia	34,51%		% más previsões (teste - 30% toler.)	68,42%
	Q lamas	m ³ /dia	0,73%		Erro médio quadrático (treino)	206,91
				Erro médio quadrático (teste)	245,6	
5-A-3	T reator	°C	54,02%	94	Melhor rede	GRNN
	TRH	dias	27,64%		% más previsões (treino - 30% toler.)	36,00%
	CV	kg/(m ³ .dia)	15,76%		% más previsões (teste - 30% toler.)	31,59%
	Q lamas	m ³ /dia	2,59%		Erro médio quadrático (treino)	172,43
				Erro médio quadrático (teste)	186,01	
7-l-3	MV in/MS in	%	31,17%	110	Melhor rede	MLFN (3 nós)
	TRH	dias	17,98%		% más previsões (treino - 30% toler.)	38,64%
	Q lamas	m ³ /dia	13,89%		% más previsões (teste - 30% toler.)	50,00%
	MS in	kg/dia	12,69%		Erro médio quadrático (treino)	192,22
	pH in	-	12,13%		Erro médio quadrático (teste)	335,77
	LMV in	kg/dia	12,13%			
8-A	pH reator	-	40,13%	72	Melhor rede	GRNN
	T reator	°C	38,36%		% más previsões (treino - 30% toler.)	25,86%
	Q lamas	m ³ /dia	21,50%		% más previsões (teste - 30% toler.)	64,29%
					Erro médio quadrático (treino)	125,78
				Erro médio quadrático (teste)	239,04	

Quadro IV.1 – (continuação; continua)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
8-B	T reator	°C	45,16%	69	Melhor rede	MLFN (2 nós)
	MV in	mg/L	29,53%		% más previsões (treino - 30% toler.)	30,91%
	LMV removida	kg/dia	19,15%		% más previsões (teste - 30% toler.)	21,43%
	TRH	dias	6,17%		Erro médio quadrático (treino)	144,57
					Erro médio quadrático (teste)	119,84
8-B-2	T reator	°C	70,42%	94	Melhor rede	GRNN
	MV in	mg/L	29,04%		% más previsões (treino - 30% toler.)	44,00%
	LMV in	kg/dia	0,31%		% más previsões (teste - 30% toler.)	47,37%
	TRH	dias	0,23%		Erro médio quadrático (treino)	189,39
					Erro médio quadrático (teste)	265,51
8-C	T reator	°C	40,77%	59	Melhor rede	MLFN (3 nós)
	pH reator		27,86%		% más previsões (treino - 30% toler.)	17,02%
	TRH	dias	22,33%		% más previsões (teste - 30% toler.)	41,67%
	LMV removida	kg/dia	9,04%		Erro médio quadrático (treino)	79,39
					Erro médio quadrático (teste)	192,23
8-C-2	TRH	dias	26,74%	80	Melhor rede	MLFN (4 nós)
	LMV in	kg/dia	26,45%		% más previsões (treino - 30% toler.)	15,63%
	pH reator		25,63%		% más previsões (teste - 30% toler.)	62,50%
	T reator	°C	21,19%		Erro médio quadrático (treino)	100,38
					Erro médio quadrático (teste)	192,56
8-D	T reator	°C	100,00%	51	Melhor rede	GRNN
	CQO in	mg/L	0,00%		% más previsões (treino - 30% toler.)	24,39%
	TRH	dias	0,00%		% más previsões (teste - 30% toler.)	10,00%
	LMV removida	kg/dia	0,00%		Erro médio quadrático (treino)	161,69
					Erro médio quadrático (teste)	106,36
8-D-2	TRH	dias	37,69%	70	Melhor rede	GRNN
	T reator	°C	28,75%		% más previsões (treino - 30% toler.)	26,79%
	CQO in	mg/L	17,02%		% más previsões (teste - 30% toler.)	35,71%
	LMV in	kg/dia	16,54%		Erro médio quadrático (treino)	148,29
					Erro médio quadrático (teste)	244,72
8-E	TAL reator	mg/L	41,79%	46	Melhor rede	MLFN (3 nós)
	T reator	°C	27,64%		% más previsões (treino - 30% toler.)	24,32%
	AGV reator	mg/L	15,29%		% más previsões (teste - 30% toler.)	33,33%
	pH reator	-	15,28%		Erro médio quadrático (treino)	89,78
					Erro médio quadrático (teste)	173,48
8-F	AGV reator	mg/L	25,84%	42	Melhor rede	GRNN
	TAL reator	mg/L	24,45%		% más previsões (treino - 30% toler.)	2,94%
	pH reator	-	23,05%		% más previsões (teste - 30% toler.)	37,50%
	T reator	°C	17,54%		Erro médio quadrático (treino)	45,05
	TRH	dias	9,13%		Erro médio quadrático (teste)	175,32
8-G	AGV reator	mg/L	39,26%	38	Melhor rede	GRNN
	T reator	°C	26,28%		% más previsões (treino - 30% toler.)	3,33%
	CQO in	mg/L	14,47%		% más previsões (teste - 30% toler.)	62,50%
	TAL reator	mg/L	11,13%		Erro médio quadrático (treino)	53,33
	pH reator	-	8,87%		Erro médio quadrático (teste)	246,62
8-H	T reator	°C	23,36%	38	Melhor rede	MLFN (14 nós)
	CQO reator	mg/L	22,20%		% más previsões (treino - 30% toler.)	0,00%
	pH reator	-	18,74%		% más previsões (teste - 30% toler.)	37,50%
	AGV reator	mg/L	18,51%		Erro médio quadrático (treino)	0,04
	TAL reator	mg/L	17,20%		Erro médio quadrático (teste)	192,97

Quadro IV.1 – (continuação; continua)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
9-E-1	T reator	°C	100,00%	618	Melhor rede	MLFN (29 nós)
					% más previsões (treino - 30% toler.)	49,39%
					% más previsões (teste - 30% toler.)	50,81%
					Erro médio quadrático (treino)	240,89
					Erro médio quadrático (teste)	243,50
9-E-2	T reator	°C	59,74%	618	Melhor rede	MLFN (14 nós)
	Q lamas	m ³ /dia	40,26%		% más previsões (treino - 30% toler.)	48,58%
					% más previsões (teste - 30% toler.)	54,03%
					Erro médio quadrático (treino)	225,34
					Erro médio quadrático (teste)	241,72
9-E-3	T reator	°C	50,24%	87	Melhor rede	MLFN (6 nós)
	LMV in	kg/dia	49,76%		% más previsões (treino - 30% toler.)	31,43%
					% más previsões (teste - 30% toler.)	41,18%
					Erro médio quadrático (treino)	144,47
					Erro médio quadrático (teste)	242,49
9-E-4	pH in	-	53,56%	81	Melhor rede	MLFN (11 nós)
	T reator	°C	46,44%		% más previsões (treino - 30% toler.)	15,38%
					% más previsões (teste - 30% toler.)	50,00%
					Erro médio quadrático (treino)	79,24
					Erro médio quadrático (teste)	283,29
9-E-5	T reator	°C	99,84%	87	Melhor rede	GRNN
	LMV in	kg/dia	0,13%		% más previsões (treino - 30% toler.)	50,00%
	Q lamas	m ³ /dia	0,03%		% más previsões (teste - 30% toler.)	35,29%
					Erro médio quadrático (treino)	213,35
					Erro médio quadrático (teste)	200,82
9-E-6	T reator	°C	99,93%	76	Melhor rede	GRNN
	pH in	-	0,07%		% más previsões (treino - 30% toler.)	49,18%
	Q lamas	m ³ /dia	0,00%		% más previsões (teste - 30% toler.)	46,67%
					Erro médio quadrático (treino)	213,09
					Erro médio quadrático (teste)	232,27
9-E-7	pH in	-	52,64%	76	Melhor rede	GRNN
	LMV in	kg/dia	27,57%		% más previsões (treino - 30% toler.)	27,87%
	T reator	°C	19,78%		% más previsões (teste - 30% toler.)	26,67%
					Erro médio quadrático (treino)	126,55
					Erro médio quadrático (teste)	226,38
9-F-2	pH in	-	52,35%	79	Melhor rede	MLFN (8 nós)
	CV	kg/(m ³ .dia)	47,65%		% más previsões (treino - 30% toler.)	19,05%
					% más previsões (teste - 30% toler.)	50,00%
					Erro médio quadrático (treino)	103,62
					Erro médio quadrático (teste)	270,83
9-F-4	T reator	°C	99,49%	94	Melhor rede	GRNN
	CV	kg/(m ³ .dia)	0,51%		% más previsões (treino - 30% toler.)	49,33%
					% más previsões (teste - 30% toler.)	42,11%
					Erro médio quadrático (treino)	211,64
					Erro médio quadrático (teste)	218,40
9-F-6	pH in	-	46,45%	79	Melhor rede	GRNN
	CV	kg/(m ³ .dia)	32,99%		% más previsões (treino - 30% toler.)	31,75%
	T reator	°C	20,56%		% más previsões (teste - 30% toler.)	62,50%
					Erro médio quadrático (treino)	178,18
					Erro médio quadrático (teste)	184,40

Quadro IV.1 – (continuação)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
9-F-7	T reator	°C	97,83%	87	Melhor rede	GRNN
	CV	kg/(m ³ .dia)	2,09%		% más previsões (treino - 30% toler.)	51,43%
	Q lamas	m ³ /dia	0,09%		% más previsões (teste - 30% toler.)	17,65%
					Erro médio quadrático (treino)	217,36
					Erro médio quadrático (teste)	165,14
9-G-5	pH reator	-	45,66%	44	Melhor rede	GRNN
	T reator	°C	34,89%		% más previsões (treino - 30% toler.)	0,00%
	LCQO removida	kg/dia	19,21%		% más previsões (teste - 30% toler.)	44,44%
	TRH	dias	0,23%		Erro médio quadrático (treino)	17,51
					Erro médio quadrático (teste)	233,92
9-G-6	T reator	°C	46,31%	45	Melhor rede	GRNN
	LCQO in	kg/dia	24,35%		% más previsões (treino - 30% toler.)	17,14%
	LCQO removida	kg/dia	15,31%		% más previsões (teste - 30% toler.)	22,22%
	TRH	dias	14,03%		Erro médio quadrático (treino)	108,85
					Erro médio quadrático (teste)	176,06
9-G-7	T reator	°C	45,45%	36	Melhor rede	GRNN
	pH reator	-	45,30%		% más previsões (treino - 30% toler.)	6,67%
	TAL reator	mg/L	7,96%		% más previsões (teste - 30% toler.)	42,86%
	AGV reator	mg/L	1,00%		Erro médio quadrático (treino)	62,92
	MV reator	mg/L	0,17%		Erro médio quadrático (teste)	173,14
	CQO reator	mg/L	0,07%			
	LCQO out	kg/dia	0,05%			
9-G-8	pH reator	-	37,32%	66	Melhor rede	GRNN
	T reator	°C	33,51%		% más previsões (treino - 30% toler.)	5,66%
	TRH	dias	21,05%		% más previsões (teste - 30% toler.)	61,54%
	LCQO in	kg/dia	7,72%		Erro médio quadrático (treino)	60,91
	LMV in	kg/dia	0,39%		Erro médio quadrático (teste)	177,03
10-A	pH reator	-	30,00%	44	Melhor rede	MLFN (5 nós)
	T reator	°C	21,85%		% más previsões (treino - 30% toler.)	2,86%
	TRH	dias	19,81%		% más previsões (teste - 30% toler.)	33,33%
	LMV in	kg/dia	15,03%		Erro médio quadrático (treino)	23,04
	AGV reator	mg/L	13,31%		Erro médio quadrático (teste)	214,82

APÊNDICE V

Redes neuronais criadas com desfasamento

Quadro V.1 – Indicadores de desempenho do treino/teste de redes neuronais criadas com desfasamento da variável “Q biogás” em relação às restantes (continua)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados		Desfasamento
2-A	T reator	°C	62,59%	620	Melhor rede	GRNN	s.d.
	TRH	dias	32,89%		% más previsões (treino - 30% toler.)	49,60%	
	Q lamas	m ³ /dia	4,52%		% más previsões (teste - 30% toler.)	53,23%	
					Erro médio quadrático (treino)	221,04	
					Erro médio quadrático (teste)	232,57	
2-A-4-A	T reator	°C	52,53%	617	Melhor rede	GRNN	3 dias
	TRH	dias	28,36%		% más previsões (treino - 30% toler.)	46,56%	
	Q lamas	m ³ /dia	19,11%		% más previsões (teste - 30% toler.)	46,34%	
					Erro médio quadrático (treino)	223,37	
					Erro médio quadrático (teste)	230,64	
2-A-4-B	T reator	°C	56,58%	613	Melhor rede	GRNN	7 dias
	Q lamas	m ³ /dia	42,99%		% más previsões (treino - 30% toler.)	42,65%	
	TRH	dias	0,43%		% más previsões (teste - 30% toler.)	49,59%	
					Erro médio quadrático (treino)	206,07	
					Erro médio quadrático (teste)	244,61	
2-A-4-C	T reator	°C	63,87%	610	Melhor rede	GRNN	10 dias
	TRH	dias	4,60%		% más previsões (treino - 30% toler.)	42,21%	
	Q lamas	m ³ /dia	31,53%		% más previsões (teste - 30% toler.)	57,38%	
					Erro médio quadrático (treino)	202,57	
					Erro médio quadrático (teste)	258,62	
2-A-4-D	Q lamas	m ³ /dia	40,82%	608	Melhor rede	MLFN (3 nós)	12 dias
	T reator	°C	30,69%		% más previsões (treino - 30% toler.)	51,65%	
	TRH	dias	28,49%		% más previsões (teste - 30% toler.)	55,74%	
					Erro médio quadrático (treino)	236,63	
					Erro médio quadrático (teste)	266,14	
2-A-4-E	T reator	°C	51,13%	605	Melhor rede	GRNN	15 dias
	Q lamas	m ³ /dia	38,05%		% más previsões (treino - 30% toler.)	43,80%	
	TRH	dias	10,82%		% más previsões (teste - 30% toler.)	56,20%	
					Erro médio quadrático (treino)	206,18	
					Erro médio quadrático (teste)	257,71	
3-C-3	T reator	°C	58,67%	69	Melhor rede	GRNN	s.d.
	TRH	dias	34,30%		% más previsões (treino - 30% toler.)	25,45%	
	LMV removida	kg/dia	3,97%		% más previsões (teste - 30% toler.)	21,43%	
	Q lamas	m ³ /dia	3,06%		Erro médio quadrático (treino)	130,68	
					Erro médio quadrático (teste)	226,51	
3-C-3-4	T reator	°C	64,79%	66	Melhor rede	GRNN	1 célula
	Q lamas	m ³ /dia	34,85%		% más previsões (treino - 30% toler.)	39,62%	
	TRH	dias	0,33%		% más previsões (teste - 30% toler.)	7,69%	
	LMV removida	kg/dia	0,04%		Erro médio quadrático (treino)	192,77	
					Erro médio quadrático (teste)	160,31	
3-C	TRH	dias	51,33%	57	Melhor rede	GRNN	s.d.
	T reator	°C	25,94%		% más previsões (treino - 30% toler.)	13,04%	
	LMV removida	kg/dia	21,90%		% más previsões (teste - 30% toler.)	54,55%	
	Q lamas	m ³ /dia	0,83%		Erro médio quadrático (treino)	150,03	
					Erro médio quadrático (teste)	334,61	
3-C-4	Q lamas	m ³ /dia	25,00%	56	Melhor rede	GRNN	1 célula
	TRH	dias	25,00%		% más previsões (treino - 30% toler.)	48,84%	
	T reator	°C	25,00%		% más previsões (teste - 30% toler.)	54,55%	
	LMV removida	kg/dia	25,00%		Erro médio quadrático (treino)	233,2	
					Erro médio quadrático (teste)	337,39	
3-L	pH in	-	28,89%	46	Melhor rede	MLFN (6 nós)	s.d.
	T reator	°C	24,99%		% más previsões (treino - 30% toler.)	5,41%	
	LMV removida	kg/dia	23,42%		% más previsões (teste - 30% toler.)	77,78%	
	TRH	dias	22,71%		Erro médio quadrático (treino)	9,49	
					Erro médio quadrático (teste)	288,31	
3-L-4	TRH	dias	25,00%	43	Melhor rede	GRNN	1 célula
	T reator	°C	25,00%		% más previsões (treino - 30% toler.)	50,00%	
	LMV removida	kg/dia	25,00%		% más previsões (teste - 30% toler.)	44,44%	
	pH in		25,00%		Erro médio quadrático (treino)	234,41	
					Erro médio quadrático (teste)	255,92	

Quadro V.1 – (continuação; continua)

Rede	Variáveis entrada	Impacto	Nº casos	Resultados		Desfasamento
3-L-3	T reator	°C	27,65%	58	Melhor rede	MLFN (5 nós)
	TRH	dias	26,58%		% más previsões (treino - 30% toler.)	10,87%
	LMV removida	kg/dia	23,43%		% más previsões (teste - 30% toler.)	50,00%
	pH in	-	22,34%		Erro médio quadrático (treino)	58,81
					Erro médio quadrático (teste)	220,73
3-L-3-4	LMV removida	kg/dia	100,00%	55	Melhor rede	GRNN
	T reator	°C	0,00%		% más previsões (treino - 30% toler.)	47,73%
	pH in	-	0,00%		% más previsões (teste - 30% toler.)	18,18%
	TRH	dias	0,00%		Erro médio quadrático (treino)	217,95
					Erro médio quadrático (teste)	119,8
3-N	pH reator	-	36,05%	46	Melhor rede	GRNN
	pH in	-	32,00%		% más previsões (treino - 30% toler.)	21,62%
	Q lamas	m ³ /dia	21,46%		% más previsões (teste - 30% toler.)	11,11%
	TRH	dias	10,09%		Erro médio quadrático (treino)	102,67
	LMV removida	kg/dia	0,40%		Erro médio quadrático (teste)	168,17
3-N-4	pH reator	-	49,46%	43	Melhor rede	GRNN
	TRH	dias	33,02%		% más previsões (treino - 30% toler.)	29,41%
	pHin	-	17,24%		% más previsões (teste - 30% toler.)	66,67%
	Q lamas	m ³ /dia	0,17%		Erro médio quadrático (treino)	198,51
	LMV removida	kg/dia	0,10%		Erro médio quadrático (teste)	214,86
4-A	T reator	°C	26,60%	81	Melhor rede	MLFN (4 nós)
	Q lamas	m ³ /dia	25,18%		% más previsões (treino - 30% toler.)	33,85%
	LMV in	kg/dia	24,80%		% más previsões (teste - 30% toler.)	43,75%
	TRH	dias	23,43%		Erro médio quadrático (treino)	140,45
					Erro médio quadrático (teste)	202,43
4-A-4	TRH	dias	79,33%	78	Melhor rede	GRNN
	T reator	°C	19,10%		% más previsões (treino - 30% toler.)	53,23%
	Q lamas	m ³ /dia	0,93%		% más previsões (teste - 30% toler.)	62,50%
	LMV in	kg/dia	0,64%		Erro médio quadrático (treino)	252,16
					Erro médio quadrático (teste)	247,60
4-A-4-A	TRH	dias	44,25%	93	Melhor rede	GRNN
	LMV in	kg/dia	29,79%		% más previsões (treino - 30% toler.)	40,54%
	T reator	°C	25,77%		% más previsões (teste - 30% toler.)	42,11%
	Q lamas	m ³ /dia	0,19%		Erro médio quadrático (treino)	220,06
					Erro médio quadrático (teste)	232,52
4-A-4-B	T reator	°C	27,28%	93	Melhor rede	MLFN (5 nós)
	Q lamas	m ³ /dia	26,50%		% más previsões (treino - 30% toler.)	28,38%
	LMV in	kg/dia	24,13%		% más previsões (teste - 30% toler.)	52,63%
	TRH	dias	22,08%		Erro médio quadrático (treino)	124,62
					Erro médio quadrático (teste)	177,85
4-A-4-C	TRH	dias	55,55%	91	Melhor rede	GRNN
	LMV in	kg/dia	41,15%		% más previsões (treino - 30% toler.)	54,79%
	T reator	°C	2,65%		% más previsões (teste - 30% toler.)	61,11%
	Q lamas	m ³ /dia	0,65%		Erro médio quadrático (treino)	250,88
					Erro médio quadrático (teste)	264,35
4-A-3	TRH	dias	32,83%	94	Melhor rede	MLFN (3 nós)
	Q lamas	m ³ /dia	31,10%		% más previsões (treino - 30% toler.)	40,00%
	T reator	°C	19,82%		% más previsões (teste - 30% toler.)	47,37%
	LMV in	kg/dia	16,25%		Erro médio quadrático (treino)	162,34
					Erro médio quadrático (teste)	187,67
4-A-3-4	LMV in	kg/dia	29,77%	91	Melhor rede	MLFN (2 nós)
	T reator	°C	26,43%		% más previsões (treino - 30% toler.)	36,99%
	TRH	dias	26,38%		% más previsões (teste - 30% toler.)	50,00%
	Q lamas	m ³ /dia	17,41%		Erro médio quadrático (treino)	177,07
					Erro médio quadrático (teste)	253,24

Quadro V.1 – (continuação)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados		Desfasamento
5-A-3	TRH	dias	38,05%	94	Melhor rede	MLFN (3 nós)	s.d.
	Q lamas	m ³ /dia	31,71%		% más previsões (treino - 30% toler.)	34,67%	
	CV	kg/(m ³ .dia)	15,36%		% más previsões (teste - 30% toler.)	47,37%	
	T reator	°C	14,88%		Erro médio quadrático (treino)	156,59	
					Erro médio quadrático (teste)	230,48	
5-A-3-4	T reator	°C	35,19%	91	Melhor rede	GRNN	1 célula
	Q lamas	m ³ /dia	34,57%		% más previsões (treino - 30% toler.)	45,21%	
	TRH	dias	19,02%		% más previsões (teste - 30% toler.)	66,67%	
	CV	kg/(m ³ .dia)	11,22%		Erro médio quadrático (treino)	199,20	
					Erro médio quadrático (teste)	264,97	
5-F-3	T reator	°C	44,64%	81	Melhor rede	GRNN	s.d.
	pH in	-	22,30%		% más previsões (treino - 30% toler.)	20,00%	
	TRH	dias	21,88%		% más previsões (teste - 30% toler.)	56,25%	
	CV	kg/(m ³ .dia)	10,92%		Erro médio quadrático (treino)	107,44	
	Q lamas	m ³ /dia	0,26%		Erro médio quadrático (teste)	158,59	
5-F-3-4	T reator	°C	98,71%	78	Melhor rede	GRNN	1 célula
	CV	kg/(m ³ .dia)	1,11%		% más previsões (treino - 30% toler.)	56,45%	
	TRH	dias	0,13%		% más previsões (teste - 30% toler.)	56,25%	
	Q lamas	m ³ /dia	0,04%		Erro médio quadrático (treino)	238,20	
	pH in		0,01%		Erro médio quadrático (teste)	228,66	
7-G	Q lamas	m ³ /dia	76,21%	111	Melhor rede	GRNN	s.d.
	MV in/MS in	%	23,79%		% más previsões (treino - 30% toler.)	0,00%	
					% más previsões (teste - 30% toler.)	77,27%	
					Erro médio quadrático (treino)	1,63	
					Erro médio quadrático (teste)	1256,71	
7-G-4	Q lamas	m ³ /dia	54,20%	108	Melhor rede	GRNN	1 célula
	MV in/MS in	%	34,81%		% más previsões (treino - 30% toler.)	66,28%	
					% más previsões (teste - 30% toler.)	63,64%	
					Erro médio quadrático (treino)	763,82	
					Erro médio quadrático (teste)	815,63	

APÊNDICE VI

Redes neuronais criadas a partir de conjuntos de dados adicionais

Quadro VI.1 – Indicadores de desempenho das melhores redes neuronais criadas a partir de conjuntos de dados adicionais (continua)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
3-N-5	TRH	dias	35,39%	43	Melhor rede	MLFN (2 nós)
	Q lamas	m ³ /dia	33,58%		% más previsões (treino - 30% toler.)	20,59%
	pH reator	-	18,99%		% más previsões (teste - 30% toler.)	66,67%
	pH in	-	8,64%		Erro médio quadrático (treino)	115,24
	LMV removida	kg/dia	3,40%		Erro médio quadrático (teste)	238,23
3-N-5-2	pH reator	-	30,55%	46	Melhor rede	MLFN (5 nós)
	Q lamas	m ³ /dia	23,76%		% más previsões (treino - 30% toler.)	2,70%
	pH in	-	23,46%		% más previsões (teste - 30% toler.)	77,78%
	TRH	dias	22,22%		Erro médio quadrático (treino)	49,70
					Erro médio quadrático (teste)	222,10
3-N-5-3	TRH	dias	34,58%	43	Melhor rede	MLFN (2 nós)
	Q lamas	m ³ /dia	33,87%		% más previsões (treino - 30% toler.)	23,53%
	pH reator	-	24,02%		% más previsões (teste - 30% toler.)	66,67%
	pH in	-	7,53%		Erro médio quadrático (treino)	134,01
					Erro médio quadrático (teste)	246,26
3-P-5-2	T reator	°C	56,45%	80	Melhor rede	MLFN (3 nós)
	LMV out	kg/dia	33,09%		% más previsões (treino - 30% toler.)	26,56%
	CV	kg/(m ³ .dia)	10,45%		% más previsões (teste - 30% toler.)	50,00%
					Erro médio quadrático (treino)	150,73
					Erro médio quadrático (teste)	227,33
3-Q-5	pH in	-	38,02%	80	Melhor rede	GRNN
	LMV out	kg/dia	32,16%		% más previsões (treino - 30% toler.)	35,94%
	T reator	°C	27,39%		% más previsões (teste - 30% toler.)	50,00%
	TRH	dias	2,24%		Erro médio quadrático (treino)	186,88
	LMV in	kg/dia	0,18%		Erro médio quadrático (teste)	236,29
3-R-5	LMV in	kg/dia	68,30%	92	Melhor rede	GRNN
	TRH	dias	31,55%		% más previsões (treino - 30% toler.)	52,70%
	LMV out	kg/dia	0,11%		% más previsões (teste - 30% toler.)	50,00%
	Q lamas	m ³ /dia	0,04%		Erro médio quadrático (treino)	235,26
					Erro médio quadrático (teste)	208,65
3-R-5-2	Q lamas	m ³ /dia	44,23%	92	Melhor rede	GRNN
	CV	kg/(m ³ .dia)	36,72%		% más previsões (treino - 30% toler.)	0,00%
	LMV out	kg/dia	28,41%		% más previsões (teste - 30% toler.)	55,56%
					Erro médio quadrático (treino)	0,00
					Erro médio quadrático (teste)	191,90
3-S-5	T reator	°C	37,00%	78	Melhor rede	GRNN
	Q lamas	m ³ /dia	36,17%		% más previsões (treino - 30% toler.)	38,71%
	LMV in	kg/dia	24,77%		% más previsões (teste - 30% toler.)	62,50%
	LMV out	kg/dia	1,53%		Erro médio quadrático (treino)	172,56
	TRH	dias	0,54%		Erro médio quadrático (teste)	233,25
5-H-5-A	T reator	°C	65,88%	97	Melhor rede	GRNN
	LMV in	kg/dia	17,06%		% más previsões (treino - 30% toler.)	39,74%
	CV	kg/(m ³ .dia)	17,06%		% más previsões (teste - 30% toler.)	52,63%
					Erro médio quadrático (treino)	172,91
					Erro médio quadrático (teste)	224,43
5-H-5-2	T reator	°C	75,76%	80	Melhor rede	GRNN
	Q lamas	m ³ /dia	17,03%		% más previsões (treino - 30% toler.)	43,75%
	LMV in	kg/dia	7,01%		% más previsões (teste - 30% toler.)	62,50%
	CV	kg/(m ³ .dia)	0,19%		Erro médio quadrático (treino)	214,18
					Erro médio quadrático (teste)	249,23

Quadro VI.1 – (continuação)

Rede	Variáveis entrada		Impacto	Nº casos	Resultados	
5-H-5-4-A	Q lamas	m ³ /dia	44,93%	97	Melhor rede	MLFN (3 nós)
	T reator	°C	41,16%		% más previsões (treino - 30% toler.)	38,46%
	MV in	mg/L	7,82%		% más previsões (teste - 30% toler.)	52,63%
	CV	kg/(m ³ .dia)	6,99%		Erro médio quadrático (treino)	157,58
					Erro médio quadrático (teste)	249,48
5-H-5-5-A	Q lamas	m ³ /dia	41,03%	97	Melhor rede	MLFN (2 nós)
	T reator	°C	32,65%		% más previsões (treino - 30% toler.)	39,74%
	MV in	mg/L	16,70%		% más previsões (teste - 30% toler.)	47,37%
	LMV in	kg/dia	9,60%		Erro médio quadrático (treino)	175,52
					Erro médio quadrático (teste)	221,19
5-H-5-6-A	T reator	°C	47,93%	81	Melhor rede	MLFN (2 nós)
	Q lamas	m ³ /dia	35,15%		% más previsões (treino - 30% toler.)	35,38%
	MV in	%	10,74%		% más previsões (teste - 30% toler.)	43,75%
	LMV in	kg/dia	9,18%		Erro médio quadrático (treino)	183,41
					Erro médio quadrático (teste)	240,39
9-G-1	pH	-	31,03%	41	Melhor rede	GRNN
	pH in	-	29,71%		% más previsões (treino - 30% toler.)	0,00%
	T reator	°C	21,73%		% más previsões (teste - 30% toler.)	37,50%
	MV in	mg/L	7,88%		Erro médio quadrático (treino)	14,25
	TRH	dias	5,23%		Erro médio quadrático (teste)	167,22
	LMV in	kg/dia	3,95%			
	AGV reator	mg/L	0,32%			
	Q lamas	m ³ /dia	0,16%			
9-G-3	LCQO in	kg/dia	99,77%	67	Melhor rede	GRNN
	Q lamas	m ³ /dia	0,14%		% más previsões (treino - 30% toler.)	38,89%
	TRH	dias	0,08%		% más previsões (teste - 30% toler.)	76,92%
	CQO in	mg/L	0,01%		Erro médio quadrático (treino)	184,34
	pH in	-	0,00%		Erro médio quadrático (teste)	235,57
	MV in	mg/L	0,00%			
	LMV in	kg/dia	0,00%			
9-G-4	T reator	°C	39,75%	36	Melhor rede	GRNN
	pH reator	-	27,06%		% más previsões (treino - 30% toler.)	3,45%
	TAL reator	mg/L	17,16%		% más previsões (teste - 30% toler.)	14,29%
	TRH	dias	7,16%		Erro médio quadrático (treino)	36,54
	AGV reator	mg/L	5,23%		Erro médio quadrático (teste)	157,91
	AGV/TAL	-	1,68%			
	LMV out	kg/dia	0,59%			
	Q lamas	m ³ /dia	0,56%			
	MV reator	mg/L	0,39%			
	CQO reator	mg/L	0,23%			
	LCQO out	kg/dia	0,20%			

APÊNDICE VII

Melhores redes neuronais treinadas/testadas

Quadro VII.1 – Indicadores de desempenho das 13 melhores redes neuronais treinadas/testadas (continua)

Rede	Variáveis entrada	Impacto	Nº casos	Resultados	Original	1ª repetição	2ª repetição	3ª repetição	Média	Soma	Usar na previsão	
3-N	pH reator	-	26,47%	58	Melhor rede	MLFN (4 nós)	MLFN (2 nós)	GRNN	MLFN (3 nós)	-	-	Não
	pH in	-	24,84%		% más previsões (treino - 30% toler.)	15,22%	30,43%	23,91%	26,09%	23,91%	73,91%	
	TRH	dias	24,15%		% más previsões (teste - 30% toler.)	50,00%	58,33%	41,67%	50,00%	50,00%	299,69	
	Q lamas	m ³ /dia	14,20%		Erro médio quadrático (treino)	74,62	141,00	123,19	92,27	107,77		
	LMV removida	kg/dia	10,34%		Erro médio quadrático (teste)	187,58	165,17	177,22	237,71	191,92		
8-C	T reator	°C	40,77%	59	Melhor rede	MLFN (3 nós)	GRNN	GRNN	GRNN	-	-	Não
	pH reator	-	27,86%		% más previsões (treino - 30% toler.)	17,02%	0,00%	0,00%	14,89%	7,98%	49,65%	
	TRH	dias	22,33%		% más previsões (teste - 30% toler.)	41,67%	66,67%	25,00%	33,33%	41,67%	241,25	
	LMV removida	kg/dia	9,04%		Erro médio quadrático (treino)	79,39	25,81	0,02	90,24	48,87		
					Erro médio quadrático (teste)	192,23	215,99	162,65	198,66	192,38		
8-C-3	pH reator	-	31,90%	42	Melhor rede	MLFN (19 nós)	GRNN	GRNN	MLFN (19 nós)	-	-	Não
	TRH	dias	24,26%		% más previsões (treino - 30% toler.)	0,00%	0,00%	11,76%	0,00%	2,94%	52,94%	
	T reator	°C	22,09%		% más previsões (teste - 30% toler.)	37,50%	50,00%	37,50%	75,00%	50,00%	210,56	
	LMV removida	kg/dia	21,75%		Erro médio quadrático (treino)	0,06	26,9	63,23	0,03	22,56		
					Erro médio quadrático (teste)	138,41	178,81	202,34	232,44	188,00		
8-E	TAL reator	mg/L	41,79%	46	Melhor rede	MLFN (3 nós)	MLFN (2 nós)	GRNN	MLFN (2 nós)	-	-	Não
	T reator	°C	27,64%		% más previsões (treino - 30% toler.)	24,32%	29,73%	21,62%	27,03%	25,68%	72,89%	
	AGV reator	mg/L	15,29%		% más previsões (teste - 30% toler.)	33,33%	44,44%	33,33%	77,77%	47,22%	275,96	
	pH reator	-	15,28%		Erro médio quadrático (treino)	89,78	122,29	88,53	105,99	101,65		
					Erro médio quadrático (teste)	173,48	117,09	114,9	291,78	174,31		
8-F	AGV reator	mg/L	25,84%	42	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Sim
	TAL reator	mg/L	24,45%		% más previsões (treino - 30% toler.)	2,94%	8,82%	8,82%	2,94%	5,88%	52,76%	
	pH reator	-	23,05%		% más previsões (teste - 30% toler.)	37,50%	50,00%	37,50%	62,50%	46,88%	228,56	
	T reator	°C	17,54%		Erro médio quadrático (treino)	45,05	50,29	61,32	39,01	48,92		
	TRH	dias	9,13%		Erro médio quadrático (teste)	175,32	200,81	130,25	212,18	179,64		
8-H	T reator	°C	23,36%	38	Melhor rede	MLFN (14 nós)	GRNN	GRNN	GRNN	-	-	Sim
	CQO reator	mg/L	22,20%		% más previsões (treino - 30% toler.)	0,00%	6,67%	10,00%	0,00%	4,17%	51,04%	
	pH reator	-	18,74%		% más previsões (teste - 30% toler.)	37,50%	62,50%	50,00%	37,50%	46,88%	233,55	
	AGV reator	mg/L	18,51%		Erro médio quadrático (treino)	0,04	31,39	62,52	0,20	23,54		
	TAL reator	mg/L	17,20%		Erro médio quadrático (teste)	192,97	287,63	172,94	186,52	210,02		
9-G-6	T reator	°C	46,31%	45	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Não
	LCQO in	kg/dia	24,35%		% más previsões (treino - 30% toler.)	17,14%	0,00%	2,86%	0,00%	5,00%	52,23%	
	LCQO removida	kg/dia	15,31%		% más previsões (teste - 30% toler.)	22,22%	55,56%	55,56%	55,56%	47,23%	265,74	
	TRH	dias	14,03%		Erro médio quadrático (treino)	108,85	1,65	47,25	24,97	45,68		
					Erro médio quadrático (teste)	176,06	275,73	203,13	225,32	220,06		

Quadro VII.1 – (continuação; continua)

Rede	Variáveis entrada	Impacto	Nº casos	Resultados	Original	1ª repetição	2ª repetição	3ª repetição	Média	Soma	Usar na previsão	
9-G-7	T reator	°C	45,45%	36	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Não
	pH reator	-	45,30%		% más previsões (treino - 30% toler.)	6,67%	0,00%	3,33%	0,00%	2,50%	48,93%	
	TAL reator	mg/L	7,96%		% más previsões (teste - 30% toler.)	42,86%	57,14%	57,14%	28,57%	46,43%		
	AGV reator	mg/L	1,00%		Erro médio quadrático (treino)	62,92	24,79	41,78	16,23	36,43	266,76	
	MV reator	mg/L	0,17%		Erro médio quadrático (teste)	173,14	322,25	270,34	155,59	230,33		
	CQO reator	mg/L	0,07%									
	LCQO out	kg/dia	0,05%									
9-G-8	pH reator	-	37,32%	66	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Sim
	T reator	°C	33,51%		% más previsões (treino - 30% toler.)	5,66%	9,43%	0,00%	28,30%	10,85%	51,23%	
	TRH	dias	21,05%		% más previsões (teste - 30% toler.)	61,54%	46,15%	38,46%	15,38%	40,38%		
	LCQO in	kg/dia	7,72%		Erro médio quadrático (treino)	60,91	71,68	26,46	140,39	74,86	248,94	
	LMV in	kg/dia	0,39%		Erro médio quadrático (teste)	177,03	170,77	201,57	146,93	174,08		
3-R-5-2	Q lamas	m ³ /dia	44,23%	92	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Não
	CV	kg/(m ³ .dia)	36,72%		% más previsões (treino - 30% toler.)	0,00%	59,46%	56,76%	59,46%	43,92%	102,26%	
	LMV out	kg/dia	28,41%		% más previsões (teste - 30% toler.)	55,56%	55,56%	66,67%	55,56%	58,34%		
					Erro médio quadrático (treino)	0,00	243,76	244,86	239,89	182,13	409,31	
					Erro médio quadrático (teste)	191,90	233,99	231,32	251,51	227,18		
9-G-1	pH reator	-	31,03%	41	Melhor rede	GRNN	GRNN	GRNN	GRNN	-	-	Sim
	pH in	-	29,71%		% más previsões (treino - 30% toler.)	0,00%	0,00%	0,00%	0,00%	0,00%	50,00%	
	T reator	°C	21,73%		% más previsões (teste - 30% toler.)	37,50%	62,50%	75,00%	25,00%	50,00%		
	MV in	mg/L	7,88%		Erro médio quadrático (treino)	14,25	14,61	11,47	15,95	14,07	216,62	
	TRH	dias	5,23%		Erro médio quadrático (teste)	167,22	245,07	260,95	136,95	202,55		
	LMV in	kg/dia	3,95%									
	AGV reator	mg/L	0,32%									
	Q lamas	m ³ /dia	0,16%									