

II-437 - PREDIÇÃO DA REMOÇÃO DE MATÉRIA ORGÂNICA EM LAGOA AERADA USANDO REDE NEURAL ARTIFICIAL**Jácina Tábita Gurgel Morais**⁽¹⁾

Engenheira Química pela Universidade Federal do Rio Grande do Norte. Mestranda em Engenharia Industrial pela Universidade Federal da Bahia, Programa de Pós-Graduação em Engenharia Industrial (UFBA/PEI).

Karla Patrícia Oliveira Esquerre⁽²⁾

Engenheira Química – Universidade Federal de Alagoas (UFAL). Mestre e Doutora em Engenharia Química – Universidade Estadual de Campinas (UNICAMP). Pós-doutora em Engenharia Sócio-Ambiental – Hokkaido University (Hokudai, Japão). Professora do Departamento de Engenharia Química e do Programa de Pós-Graduação em Engenharia Industrial – Escola Politécnica, Universidade Federal da Bahia (EP/UFBA).

Asher Kiperstok⁽³⁾

Engenheiro Civil – TECHNION, Israel Institute of Technology. MSc. e PhD em Engenharia Química. Tecnologias Ambientais – University of Manchester Institute of Science and Technology (UMIST, Inglaterra). Coordenador da Rede de Tecnologias Limpas e Minimização de Resíduos (TECLIM) e do Programa de Pós-Graduação em Produção Limpa – Departamento de Engenharia Ambiental, (EP/UFBA).

Luciano Matos Queiroz⁽⁴⁾

Engenheiro Civil pela Universidade Federal da Bahia (1997). Doutor em Engenharia Civil pela Universidade de São Paulo (2009). Professor Adjunto da Escola Politécnica da UFBA.

Endereço⁽¹⁾: Rua Afonso Pena, 58 1º andar - Paredões - Mossoró - RN - CEP: 59618-020 - Brasil- Tel: +55 (71) 9275-3978 - e-mail: tabitagurgel@yahoo.com.br

RESUMO

As redes neurais artificiais, diferentes de outros tipos de modelagem tradicionais, possuem a capacidade de representar relações, lineares e não lineares, de um conjunto de dados, pelo processo de aprendizagem. Muitas técnicas avançadas vêm sendo adotada em plantas de tratamento de efluentes industriais com o propósito de melhorar o controle operacional da planta a fim de garantir a qualidade do efluente tratado antes de descartá-lo. Neste trabalho são utilizadas redes neurais artificiais para prever a quantidade de matéria orgânica dissolvida, representada pela demanda química de oxigênio, DQO, em um efluente tratado em lagoa aerada. Para tanto, três conjuntos de dados foram formados para construção dos modelos de predição. O desempenho dos modelos foi avaliado pelo erro quadrático médio (EQM) e pelo coeficiente de correlação linear (R^2). A DQO como carga orgânica foi útil para melhorar o desempenho do modelo de predição, sendo obtido menor EQM e maior R^2 .

PALAVRAS-CHAVE: Demanda Química de Oxigênio, Lagoa aerada, Redes Neurais Artificiais, Tratamento de efluentes industriais.

1. INTRODUÇÃO

Um problema ambiental relevante devido ao aumento das atividades industriais é a disposição e o tratamento de resíduos líquidos (efluentes). O tratamento de efluentes é uma atividade complexa cuja operação incorreta pode acarretar sérios problemas de saúde pública e ambiental.

A lagoa aerada é um exemplo típico de sistema de tratamento de efluentes cujo principal objetivo é reduzir a quantidade de matéria orgânica dissolvida e promover a nitrificação através da atividade de grupos específicos de bactérias. A estabilidade do processo de tratamento, no que concerne à remoção de matéria orgânica, é usualmente avaliada através das concentrações de DQO e DBO do efluente final.

Como as normas ambientais são cada vez mais rígidas, muitas indústrias tem adotado técnicas avançadas para melhorar o desempenho das estações de tratamento de efluentes (ETEs), a fim de minimizar a geração de resíduos e efluentes. Modelos são utilizados para prever qual e quantitativamente variações do efluente, a fim de atingir as normas padrões de descarga (Hamed, Khalafallah e Hassanien, 2004; Mjalli, Al-Asheh e Alfadala, 2007; Akrotos, Papaspyros e Tsihrintzis, 2008).

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

A aplicação de RNA para melhorar os processos de tratamento biológicos de efluentes através de modelos de predição de parâmetros chaves, como concentração de DQO ou DBO, úteis para avaliação do desempenho da planta, tem apresentado resultados satisfatórios, como pode ser observado nos trabalhos de Hamed, Khalafallah e Hassanien, 2004; Oliveira-Esquerre *et. al.*, 2004; Grieu, Polit, e Colprim, 2005; Akrotos, Papaspyros e Tsihrintzis, 2008; May e Sivakumar, 2009 e Basant *et. al.*, 2010.

Há vários tipos de RNAs na literatura, a seleção da topologia neural apropriado à aplicação, isto é, qual configuração da rede neural a ser utilizada, ainda é feita de forma empírica, embora existam algumas heurísticas que conduzam a opção mais acertada. Em geral o procedimento requer alguma experiência dos projetistas, o que pode ser obtido com variadas aplicações. As redes perceptrons de múltiplas camadas tem sido utilizada com sucesso na modelagem de sistema de tratamento biológicos (Grieu, Polit, e Colprim, 2005; Akrotos, Papaspyros e Tsihrintzis, 2008; May e Sivakumar, 2009).

O presente trabalho tem por objetivo desenvolver um modelo de predição baseado na construção de redes neurais para avaliação da remoção da matéria orgânica, medida pela DQO, em um efluente industrial tratado em lagoa aerada.

2. METODOLOGIA

2.1 DESCRIÇÃO DO PROCESSO

O sistema de tratamento de efluentes da empresa de produção de papel e celulose em estudo consiste de duas etapas, um tratamento primário seguido de um tratamento biológico. O tratamento primário consiste de uma decantação quimicamente assistida que ocorre em dois tanques paralelos providos de agitadores mecânicos e câmaras de floculação. O tratamento biológico é realizado em uma lagoa aerada e complementado em um conjunto de cinco lagoas de estabilização. Os sólidos removidos nos tanques de decantação seguem para um tanque de secagem e, após a remoção da umidade, são apropriadamente descartados.

2.2 ANÁLISE DESCRITIVA DOS DADOS

Os dados experimentais disponibilizados pela empresa referem-se a um período de 1427 dias consecutivos de medição diários. Onze parâmetros de monitoração da lagoa aerada e dois referentes ao processo produtivo da empresa foram escolhidos através do julgamento de profissionais da área baseados na importância dessas variáveis para avaliação da eficiência global do processo de tratamento do efluente. Na Tabela 1 são apresentadas algumas estatísticas básicas dos dados referentes aos parâmetros de caracterização da água residuária estudada.

Tabela 1: Estatística básica dos parâmetros monitorados pela empresa

PARÂMETROS	MÉDIA	DESVIO PADRÃO	MÍNIMO	MÁXIMO	SKEWNESS	CURTOSE	LACUNAS (%)
Q (m³/dia)	67.363,81	11.588,51	4.474,00	97.850,00	-1,53	4,90	0
DBO _{in} (ppm)	245,05	46,33	41,00	449,00	0,08	1,54	6,24
DQO _{in} (ppm)	561,49	104,19	136,00	925,00	-0,16	1,04	6,24
SST (ppm)	149,20	85,74	12,00	591,00	1,59	4,13	60,4
pH	7,45	1,21	0,85	12,53	1,79	4,18	3,71
NAM (ppm)	2,45	1,77	0	20,00	2,42	16,41	54,2
NN (ppm)	1,43	0,88	0,03	7,39	2,42	11,05	80,5
Cor (ppm)	464,36	123,59	41,00	1.317,00	0,51	3,39	3,57
T (°C)	45,46	3,05	28,00	50,50	-2,32	8,74	32,6
EC (μS.cm ⁻¹ 20°C)	1.530,96	378,05	379,00	5.810,00	2,66	17,74	3,92
Q _{celulose} (ton/dia)	886,14	155,24	0	1.112,09	-3,31	14,16	7,21
Q _{Papel} (ton/dia)	1.042,72	94,2	382,40	1.304,80	-1,57	5,77	6,45
DQO _{out} (ppm)	315,48	73,58	105,00	865,00	1,21	6,58	5,81

in equivale a entrada; out equivale a saída.

Através do coeficiente de assimetria (*skewness*) e da curtose é possível analisar a proximidade de cada parâmetro em relação à gaussiana. Os valores de *skewness* estimados indicam que apenas a DBO_{in} e a Cor apresentam distribuições simétricas se aproximando da curva normal. Os valores da curtose indicam que apenas a DBO_{in} e DQO_{in}, apresentam achatamento mais próximo da distribuição normal.

Os dados foram retirados de um conjunto de 1427 amostras dos quais, foram excluídos, inicialmente, parâmetros que apresentavam quantidade de lacunas superior a 50%, em seguida descartado os dados com prováveis equívocos de medição. A DBO não foi escolhida como variável de entrada devido ao tempo requerido de cinco dias para sua medição em laboratório, o qual se torna impraticável para construção do modelo. Em adição a esses fatores o conjunto de dados foi reduzido a 786 amostras.

A coleta criteriosa dos dados é de fundamental importância para o futuro desempenho de uma rede neural. Esta tarefa requer uma avaliação para identificar características comuns e incomuns dos dados, desta forma, três diferentes conjuntos de dados foram testados: Dados 1 composto pelas 786 amostras; Dados 2 composto pela mesma quantidade de amostras, mas, o aporte de matéria orgânica foi quantificado como carga orgânica (kg DQO.dia⁻¹), calculada pelo produto da vazão por DQO; e o conjunto de Dados 3 composto por 778 amostras, onde os valores dos dias em que não houve produção de celulose foram excluídos, e o aporte de matéria orgânica foi medido como carga orgânica afluente com o objetivo de analisar o significado da quantidade de matéria orgânica dissolvida no efluente para o desempenho do modelo de predição por RNA.

A avaliação da eficiência de remoção de matéria orgânica na lagoa aerada foi realizada com base nos resultados da DQO da água residuária bruta e do efluente final desse tratamento.

2.3 ESTRUTURA DA REDE NEURAL ARTIFICIAL

A rede neural utilizada é a Perceptron de Múltiplas Camadas (MLP), conhecida como rede supervisionada, pois requer uma saída desejada para comparar suas respostas durante o treinamento. Em seguida os pesos sinápticos e níveis de *bias* são ajustados de forma a reduzir a diferença entre a saída calculada pela rede e a saída desejada. O algoritmo de aprendizagem aplicado para ajuste desses parâmetros livres (pesos sinápticos e *bias*) é o Levenberg-Marquardt, este é uma adaptação do *backpropagation*, e se destaca por acelerar o tempo de treinamento como também melhorar seu desempenho na classificação de padrões (Alexandre de França, Morais França e Koyama, 2009).

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

celulose. As funções de transferência utilizadas são a sigmoideal e linear, para a camada intermediária e de saída, respectivamente. O critério de parada utilizado foi encerrar o treino quando o menor erro quadrático médio do conjunto de validação foi atingido.

Os modelos de predição de valores de DQO obtidos foram simulados por diversas vezes utilizando-se 1 ou 2 camadas intermediárias e variando-se o número de neurônios intermediários de 1 a 5. Foram utilizados dois métodos de aleatorização dos dados e números de taxa de aprendizagem de 0,1, 0,01 e 0,05. Os dados de cada conjunto foram aleatorizados e divididos em três conjuntos, treino, validação e teste.

Os demais parâmetros da rede como função de ativação e método de treinamento não foram modificados ao longo do treinamento e receberam as mesmas funções e métodos para todas as configurações das redes.

2.4 DESEMPENHO DOS MODELOS

O desempenho do modelo de predição para uma rede com aprendizado supervisionado é avaliado pelo erro quadrático médio (EQM). Outro critério de avaliação discutido na literatura é o coeficiente de correlação linear (R^2), cuja idéia é avaliar a relevância de um dado de entrada de uma RNA, em relação à saída desejada (Oliveira-Esquerre *et. al.*, 2004; Akratos, Papaspyros e Tsihrintzis, 2008; Hamed, Khalafallah e Hassanien 2004). Este coeficiente assume apenas valores entre 0 e 1, em que altos valores de R^2 significa boa correlação entre as variáveis de entrada com a de saída, e R^2 próximo de zero significa que as duas variáveis não dependem linearmente uma da outra.

Para examinar a adequação do modelo, analistas requerem que os erros gerados pelo modelo sejam normalmente distribuídos. Para verificação aproximada da normalidade, pode-se construir gráfico de probabilidade normal dos resíduos, histograma de frequência, e também padronizar os resíduos.

Para padronização dos resíduos, os resíduos são escalonados de modo que seus desvios-padrão sejam aproximadamente iguais a um. Assim, os resíduos grandes (que podem indicar possíveis *outliers*, ou observações não usuais) serão mais óbvios a partir da inspeção dos gráficos residuais. Os resíduos são padronizados a partir da Equação (2.1):

$$d_i = e_i / \sqrt{\sigma^2}, i = 1, 2, \dots, n \quad \text{Equação (2.1)}$$

Onde, e_i é o erro, calculados pela diferença da saída da rede pela saída desejada, e σ é o desvio-padrão da amostra.

Os erros apresentam distribuição normal, se aproximadamente 95% dos resíduos padronizados estiverem dentro do intervalo (-2,2) (Montgomery, Runger e Hubele, 2004).

3. RESULTADOS

Diferentes topologias dos modelos foram testadas, e, considerando o EQM e R^2 obtidos, verificou-se que, para todas MLP uma camada oculta foi suficiente. A taxa da aprendizagem adotada foi de 0,05, e a divisão dos dados foi de 70%, 20% e 10% para os conjuntos de treino, validação e teste, respectivamente.

A Tabela 2 mostra os melhores resultados obtidos para o conjunto de teste de cada modelo descritos na seção 2.2..

Tabela 2: Resultados do desempenho dos modelos

Parâmetros comparativos	Modelos		
	M1 ^a	M2 ^b	M3 ^b
Número de neurônios na camada intermediária	3	1	1
Número de interações	11	103	78
EQM	2,1E-03	4,6E-08	6,5E-08
R^2	0.69	1	1

^a (mg DQO.L⁻¹) e ^b (kg DQO.dia⁻¹)

Os resultados encontrados do EQM e R^2 , mostrados na Tabela 2, mostra que a arquitetura da rede foi satisfatória para o aprendizado dos três conjuntos de dados, e é capaz de prever valores a partir de novos dados gerados na estação de tratamento.

Nota-se que o modelo M1 obtido pelo conjunto Dados 1 foi o único que apresentou baixo desempenho, avaliando EQM e R^2 . Isso comprova que o aporte de matéria orgânica ao sistema quantificado como carga orgânica ($\text{kg DQO} \cdot \text{dia}^{-1}$) ao invés da concentração ($\text{mg DQO} \cdot \text{L}^{-1}$) representa melhor a quantidade de matéria orgânica dissolvida no efluente, o que influenciou uma significativa melhora no desempenho dos modelos M2 e M3, obtidos pelos conjuntos de dados 2 e 3, respectivamente.

Os modelos M2 e M3 apresentam a mesma topologia e uma diferença insignificante quanto ao EQM, ou seja, ambos apresentam o mesmo desempenho, portanto a exclusão dos dados para os dias em que não houve produção de celulose, considerado na construção do modelo M3, foi desnecessária; desta forma, durante a preparação de dados tal informação foi irrelevante. Considerando essas hipóteses, o modelo M2 é o que apresenta melhor desempenho em termos de EQM e R^2 , sem perda de informação. A Figura 1 mostra a relação entre a DQO de saída medida versus a predita.

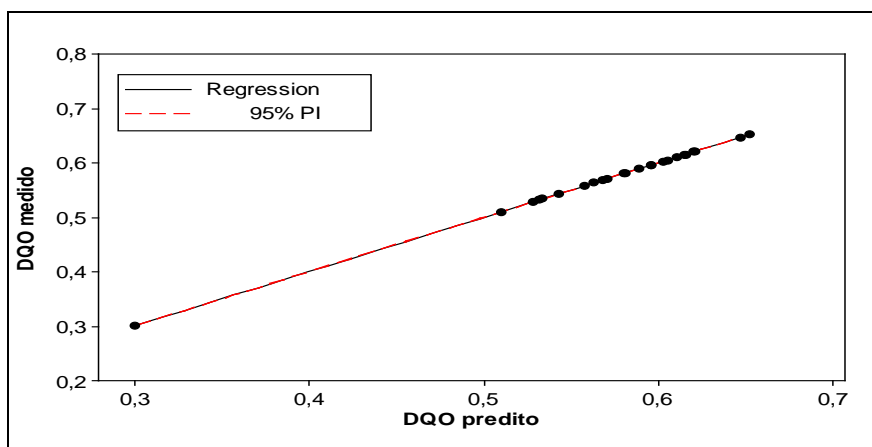


Figura 3.1: Relação entre a DQO de saída medida versus a predita em um intervalo de predição de 95% de confiança.

O gráfico da DQO medida versus a predita, mostra o bom desempenho do modelo M2. Isto também pode ser observado na Figura 3.2, onde é apresentada uma representação gráfica de série temporal dos resíduos padronizados deste modelo, considerando um intervalo de confiança de 95%, os resíduos deste.

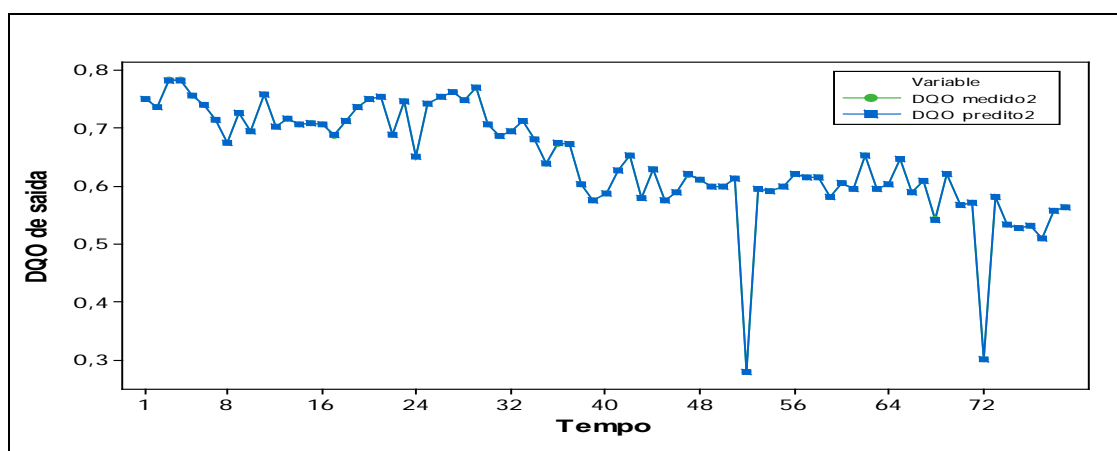


Figura 3.2: Gráfico de série temporal das DQO medida e predita.

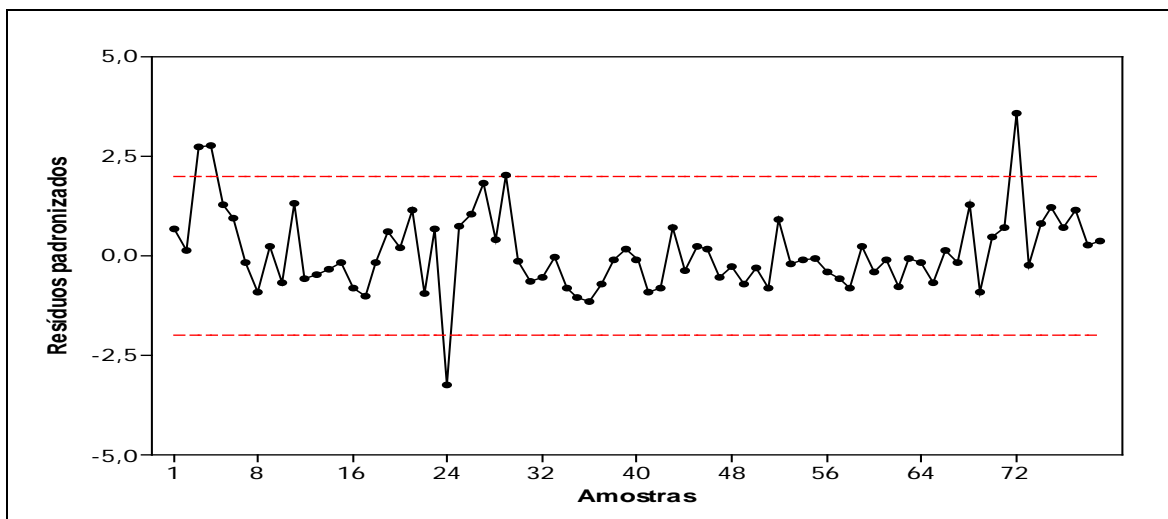


Figura 3.3: Gráfico de série temporal dos resíduos padronizados- linhas superior e inferior indicam o intervalo onde os erros devem cair.

A metodologia apresentada nesse trabalho pode ser utilizada para outros dados e processos. O procedimento básico de regressão feito através de redes neurais segue uma rotina bem definida. A variação maior ocorre com o pré-tratamento de dados.

4. CONCLUSÕES

Esse estudo sugeriu a construção de um modelo de predição de matéria orgânica medido pela DQO de saída de uma lagoa aerada para facilitar a operação e otimizar os mecanismos de gestão desse sistema de tratamento, por uso de redes neurais. Esta técnica facilita o processo de remoção da poluição através do aprendizado das relações existentes entre as variáveis preditoras e a ser predita.

A fim de se obter um modelo mais confiável, três conjuntos de dados foram cuidadosamente analisados sob os seguintes quesitos: medição de DQO como concentração, medição de DQO como carga orgânica e se os dados dos dias em que não houve produção de celulose eram divergentes. Formando assim os respectivos modelos, M1, M2 e M3.

As diferenças entre os modelos foram: o aporte de matéria orgânica quantificado como carga orgânica ($\text{kg DQO} \cdot \text{dia}^{-1}$) melhora o desempenho do modelo de predição; já exclusão de dados identificados como divergentes não foi eficiente para construção de um modelo mais confiável.

Em adição a esses fatores, o modelo M2 é mais satisfatório e eficiente no monitoramento dos parâmetros de entrada da lagoa aerada quando predito a DQO. Deve-se ressaltar que os compostos químicos específicos em efluentes podem atuar de outras formas estimulantes e satisfatórias influenciando a atividade microbiana na lagoa aerada, e a qualidade do efluente pode ser altamente influenciada pelas condições ambientais. Por se tratar de um modelo empírico, extrapolações devem ser utilizadas.

O bom desempenho deste modelo por uso de RNAs poderá promover a aplicação de modelos de predição no controle operacional de plantas de tratamento de efluentes fornecendo uma avaliação sistemática para a decisão de parâmetros considerados no modelo.

5. REFERÊNCIA BIBLIOGRÁFICA

1. Akrotos, C.S., Papaspyros, J.N.E., Tsihrintzis, V.A. An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands. *Chemical Engineering Journal* v.143, p.96–110. 2008.
2. Basant, N., Gupta, S., Malik, A., Singh, K.P. Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical oxygen demand of the surface water — A case study. *Chemometrics and Intelligent Laboratory Systems*, v.104, p.172–180. 2010.
3. May, D.B., Sivakumar, M. Prediction of urban stormwater quality using artificial neural networks. *Environmental Modelling & Software*, v.24, p. 296–302. 2009.
4. Mjalli, F.S., Al-Asheh, S., Alfadala, H.E. Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management*, v.83, p. 329–338. 2007
5. Chen, H., Ning, S., Yu, R., Hung, M. Optimizing the Monitoring Strategy of Wastewater Treatment Plants by Multiobjective Neural Networks Approach. *Environ Monit Assess*, v.125, p.325–332. 2007.
6. França, J.A., França, M.B.M., Koyama, M.H., Silva, T.P. Uma implementação do algoritmo Levenberg-Marquardt dividido para aplicações em visão computacional. *Semina: Ciências Exatas e Tecnológicas*, v. 30, n. 1, p. 51-62. 2009.
7. Oliveira-Esquerre, K.P., Seborg, D.E., Bruns, R.E., Mori, M. Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill Part I. Linear approaches. *Chemical Engineering Journal* v.104, p.73–81. 2004.
8. Hamed, M.M., Khalafallah, M.G., Hassanien, E.A. Prediction of wastewater treatment plant performance using artificial neural networks. *Environmental Modelling & Software*, v.19, p.919–928. 2004.
9. Montgomery, D.C., Runger, G.C., Hubele, N.F. Estatística aplicada à engenharia. Construindo modelos empíricos. Segunda edição, LTC - Livros Técnicos e Científicos Editora S.A, Rio de Janeiro, p. 203-204, 2004.
10. Grieu, S., Traore, A., Polit, M., Colprim, J. Prediction of parameters characterizing the state of a pollution removal biologic process. *Engineering Applications of Artificial Intelligence*, v.18, p.559–573. 2005.