

Boletim Técnico da Escola Politécnica da USP
Departamento de Engenharia de Telecomunicações
e Controle

ISSN 1517-3550

BT/PTC/0405

técnicas de Seleção de Variáveis de
Entrada para Identificação da
Concentração de Amônia de uma Planta
de Tratamento de Efluentes Utilizando
Modelos NARX Polinomiais

Rosimeire Aparecida Jerônimo
Claudio Garcia

São Paulo – 2004

1442527

O presente trabalho é um resumo da tese de doutorado apresentada por Rosimeire Aparecida Jerônimo, sob orientação do Prof. Dr. Claudio Garcia: "Técnicas de Seleção de Variáveis de Entrada para Identificação da Concentração de Amônia de uma Planta de Tratamento de Efluentes Utilizando Modelos NARX Polinomiais", defendida em 26/03/04, na Escola Politécnica.

A íntegra da tese encontra-se à disposição com o autor e na Biblioteca de Engenharia de Eletricidade da Escola Politécnica/USP.

FICHA CATALOGRÁFICA

Jerônimo, Rosimeire Aparecida

Técnicas de seleção de variáveis de entrada para identificação da concentração de amônia de uma planta de tratamento de efluentes utilizando modelos NARX polinomiais / Rosimeire Aparecida Jerônimo, Claudio Garcia. -- São Paulo : EPUSP, 2004.

21 p. -- (Boletim Técnico da Escola Politécnica da USP, Departamento de Engenharia de Telecomunicações e Controle ; BT/PTC/0405)

1. Sistemas a várias entradas e saídas 2. Identificação (Teoria de sistemas e controle) 3. Efluentes (Tratamento) I. Garcia, Claudio II. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle III. Título IV. Série
ISSN 1517-3550

CDD 629.83

003.1

628.362

Técnicas de Seleção de Variáveis de Entrada para Identificação da Concentração de Amônia de uma Planta de Tratamento de Efluentes Utilizando Modelos NARX Polinomiais

Rosimeire Aparecida Jerônimo¹, Claudio Garcia¹

¹Laboratório de Automação e Controle
Departamento de Telecomunicações e Controle

Resumo. Neste trabalho é proposta uma ferramenta para selecionar variáveis de entrada em identificação de sistemas. Tal ferramenta se baseia em análise de Funções de Correlação (FC). O método aqui proposto é comparado com outros dois métodos já apresentados na literatura, sendo a saber, Efetividade de Entrada e Projeção (EEP) e Análise de Componentes Principais (PCA). Estes métodos são aplicados para selecionar as variáveis de entrada relevantes dentre um conjunto de 15 variáveis candidatas a entrada de uma planta simulada de tratamento biológico de efluentes. Esta planta é identificada utilizando-se modelos MISO para a variável de saída, a saber, concentração de amônia (S_{NH_3}). O procedimento de identificação e estimação de parâmetros foi realizado utilizando-se modelos NARX polinomiais.

Abstract. This work puts forward a tool to select input variables in system identification. Such tool is based on analysis of Correlation Functions (FC). The proposed method is compared with two other methods that have already been presented in literature, which are: Input Effectiveness and Projection (EEP) and Principal Component Analysis (PCA). These methods are applied to select the relevant variables among a set of 15 candidates for input of a biological effluent treatment plant. One output variable MISO models - concentration of ammonia (S_{NH_3}) - were used to identify this plant. The procedure of identification and estimation of parameters was accomplished using polynomial NARX models.

1. Introdução

É comum ouvir que a *Identificação de Sistemas* ainda é uma *arte*. De fato, vários aspectos da identificação de um sistema real envolvem decisões e escolhas subjetivas. Alguns desses aspectos dizem respeito ao projeto de testes, à escolha da estrutura do modelo e à determinação da taxa de amostragem (Aguirre, 2000).

Em problemas do mundo real, para obtenção de modelos não-lineares, pode haver várias entradas disponíveis para uso. Porém, um problema comum é que não se sabe qual ou quais entradas são necessárias para o modelo, de forma que ele possa representar, dinamicamente bem, o processo em estudo. O aumento de variáveis de entrada no procedimento de identificação faz com que se obtenha um modelo muito complexo tanto na dimensionalidade como também no processo computacional.

A principal proposta deste trabalho se baseia no desenvolvimento de métodos estatísticos para auxílio na seleção de variáveis de entrada em identificação de sistemas. Neste trabalho são apresentadas três técnicas para selecionar variáveis de entrada: *Efetividade de Entrada e Projeção (EEP)*, *Análise de Componentes Principais (PCA)* e *Funções de Correlação (FC)*. As duas primeiras são técnicas apresentadas na literatura. A terceira, pouco descrita na literatura, aparece, por exemplo, em (Aguirre, 2000), o qual menciona, de forma pouco detalhada, a seleção de variáveis de entrada usando funções de correlação. Nesse sentido, o presente trabalho teve como principal objetivo a busca de requisitos de seleção de variáveis de entrada baseando-se em funções de correlação.

Foram consideradas, neste trabalho, uma técnica de identificação, a fim de avaliar a eficiência dos modelos obtidos pelos critérios de seleção de variáveis de entrada. A técnica considerada se baseou em funções polinomiais (NARX). A técnica de identificação implementada nesta pesquisa será aplicada a um modelo de referência ("*Benchmark*") de uma planta de tratamento biológico de efluentes implementado em *SIMULINK/MATLAB* (Sotomayor *et al.*,

1999), (Sotomayor *et al.*, 2001). O "Benchmark" representa um processo de lodo ativado, em uma configuração com pré-desnitrificação, incluindo os processos de remoção de matéria orgânica, nitrificação e desnitrificação de efluentes domésticos. As simulações serão realizadas empregando-se programas a serem desenvolvidos em *MATLAB* e fazendo-se uso também do *SIMULINK*.

2. O Processo ASWWTP-USP Benchmark

O modelo de referência utilizado na aplicação deste trabalho é denominado (*ASWWTP-USP (Activated Sludge Wastewater Treatment Plant - University of São Paulo) benchmark*), implementado por Oscar Sotomayor. Este modelo de referência, desenvolvido para a simulação dinâmica de um processo de tratamento biológico de esgotos domésticos por lodo ativado foi implementado em ambiente *MATLABTM/Simulink* (<http://www.mathworks.com>).

Este modelo de referência representa um processo de lodo ativado em uma configuração de fluxo contínuo com pré-desnitrificação, para remoção de matéria orgânica e nitrogênio de esgotos domésticos, operando a uma temperatura constante de 15°C e pH neutro. Este modelo de referência é uma descrição geral de um protocolo de simulação padrão que inclui o arranjo físico da estação, os modelos fenomenológicos usados, os parâmetros destes modelos, as características do esgoto, as condições iniciais de simulação, um procedimento para teste em malha aberta do comportamento dinâmico do processo e critérios para avaliação do desempenho do sistema controlado.

Na figura 1 é mostrado um diagrama esquemático de um ASP (Processo de Lodo Ativado).

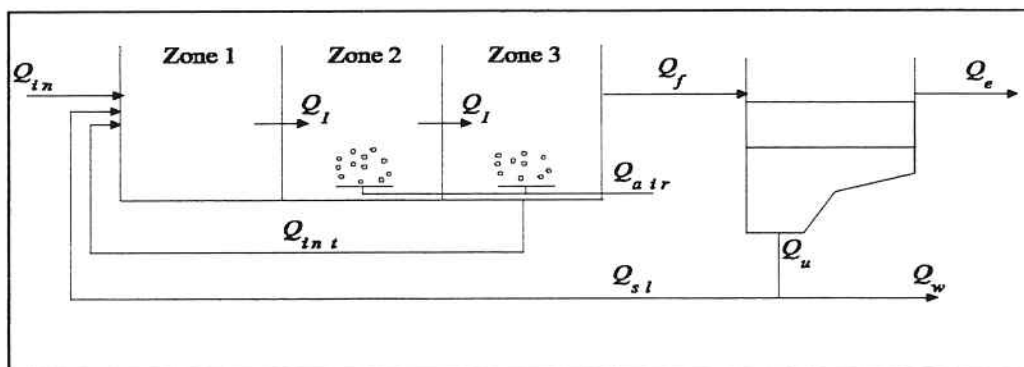


Figura 1: Processo de Lodo Ativado composto de um Bioreator e um Sedimentador

A configuração do processo considerada no *ASWWTP-USP benchmark*, conforme foi apresentada na figura 1, é formada por um bioreator em série com um sedimentador secundário. O bioreator é composto por uma zona anóxica (zona 1 com volume de 13 m³) e duas zonas aeróbicas (zonas 2 e 3 com volumes de 18 m³ e 20 m³, respectivamente), enquanto que o sedimentador secundário tem um volume de 20 m³. Em condições de operação nominal e de estado estacionário, a vazão afluyente de esgoto é $Q_{in} = 4,17 \text{ m}^3/\text{h}$ com um valor médio de *DQO* (Demanda Química de Oxigênio) total de 284 mg/l e *N* total de 45,88 mg/l. O tempo de retenção hidráulica é de 17 horas (baseado no volume total, isto é, bioreator + sedimentador secundário). A vazão de reciclo interno ou retorno de licor misto (lodo mais água tratada, vazão rica em nitrato que vem da zona 3 à zona 1 do bioreator) é $Q_{int} = 2Q_{in}$, a vazão de reciclo externo ou retorno de lodo (vazão de lodo que vem do sedimentador secundário à zona 1 do bioreator) é $Q_{sl} = 0,5Q_{in}$, e a vazão de descarte de lodo é $Q_w = 25,8 \text{ l/h}$. As vazões de injeção de ar para as zona 2 e 3 são $Q_{ar2} = 44 \text{ l/h}$ e $Q_{ar3} = 33 \text{ l/h}$, respectivamente. Na zona 1, por ser uma zona anóxica, não há aeração (Sotomayor, 2002).

As variáveis de estado consideradas no modelo de referência *IAWQ¹ N° 1* são dadas na tabela 1.

¹A *IAWQ* (International Association on Water Quality) no passado era conhecida como *IAWPRC* (International Association on Water Pollution Research and Control). Maiores informações sobre a *IAWQ* podem ser acessadas no site: <http://www.iawq.org.uk/>

Tabela 1: Estados incluídos no Modelo IAWQ N° 1

Concentração	Z
Elementos Particulados	X
Biomassa heterotrófica ativa	X_{BH}
Biomassa autotrófica ativa	X_{BA}
Substrato biodegradável	X_S
Nitrogênio orgânico biodegradável	X_{ND}
Produtos particulados inerte	X_{IP}
Matéria orgânica inerte	X_I
Elementos solúveis	S
Oxigênio dissolvido	S_O
Nitrogênio (Amônio + Amônia)	S_{NH}
Nitrogênio orgânico biodegradável	S_{ND}
Nitrogênio (Nitrato e Nitrito)	S_{NO}
Substrato biodegradável	S_S
Matéria orgânica inerte	S_I
Alcalinidade	S_{ALK}

A tabela 2 apresenta os parâmetros dos valores médios de fluxos, considerados no *benchmark*. Enquanto, que a tabela 3 apresenta os parâmetros dos valores médios de fluxos, relacionados à caracterização da água residual, considerados no *benchmark*.

Tabela 2: Fluxos - Valores Médios

Fluxo	Valor	Unidade
Q_{in}	4,17	m^3/h
Q_{sl}	$0.5 * Q_{in}$	m^3/h
Q_{int}	$2 * Q_{in}$	m^3/h
Q_w	0,0258	m^3/h

Tabela 3: Caracterização da água residual - Valores Médios

Concentração	Valor	Unidade
X_{BH}	0	$mg\ COD/l$
X_{BA}	0	$mg\ COD/l$
X_S	160	$mg\ COD/l$
X_{ND}	18,28	$mg\ COD/l$
X_P	40	$mg\ COD/l$
S_O	0	$mg(-COD)/l$
S_{NH}	12,5	$mg\ N/l$
S_{ND}	10,1	$mg\ N/l$
S_{NO}	1	$mg\ N/l$
S_S	64	$mg\ COD/l$
S_{ALK}	7	$m\ mol/l$
X_{SS}	166,452	$mg\ SS/l$

3. Seleção de Entradas Utilizando Conceito de Efetividade de Entrada e Projeção (EEP)

Nesta seção apresenta-se um critério de seleção de variáveis de entrada baseado no artigo de (Cao *et al.*, 1997). Neste artigo o procedimento de seleção de variáveis de entrada usa o conceito de medida de efetividade de entrada e projeção com a relação entre o valor singular mínimo (SV). O valor singular mínimo (SV) é uma medida de controlabilidade útil que quantifica o efeito das variáveis de entrada relevantes em um sistema (Cao *et al.*, 1997). O objetivo de se escolher tal técnica é para poder ter outro critério de comparação para com o método proposto neste trabalho.

Efetividade de Entrada e Projeção. Considere um sistema, dado por:

$$y = Gu \quad (1)$$

sendo que $y \in \mathbb{C}^m$ e $u \in \mathbb{C}^n$ são vetores de saída e entrada respectivamente, e a matriz $G \in \mathbb{C}^{m \times n}$ tem *rank* r . G representa a matriz de funções de transferência do sistema. Especialmente, em

estado nominal, $G(0)$ é uma matriz real. A matriz de projeção linha é:

$$P_{\mathbb{R}^\perp} = G^+ G \quad (2)$$

sendo que G^+ é a matriz inversa generalizada, ou seja, pseudo-inversa. A parte efetiva de um vetor de entrada u de dimensão n é a projeção dentro do espaço linha, $P_{\mathbb{R}^\perp} u$. Assim, a efetividade de entrada (IE), o qual foi primeiramente introduzida em (Cao and Biss, 1996), é definida como se segue:

A efetividade de entrada j é a norma-2 da projeção do j -ésimo vetor unitário natural, e_j , dentro do espaço linha de G , sendo:

$$\eta_{Ij} = \| P_{\mathbb{R}^\perp} e_j \|_2 \quad (3)$$

$$\eta_{Ij}^2 = [P_{\mathbb{R}^\perp}^\perp]_{jj} \quad (4)$$

Portanto, η_{Ij}^2 corresponde exatamente aos elementos da diagonal principal de $P_{\mathbb{R}^\perp}^\perp$.

3.1. Propriedade Recursiva

Na seleção das entradas, várias linhas e/ou colunas da matriz G necessitam ser eliminadas. Após a eliminação, a efetividade das entradas restantes variará. Porém, as mudanças da efetividade da entrada (IE) pode ser recursivamente obtida usando-se a fórmula recursiva da matriz inversa generalizada:

$$G = [G_{:,n-1} \quad g_{:,n}] = \begin{bmatrix} G_{m-1,:} \\ g_{m,:} \end{bmatrix} \quad (5)$$

$$G^+ = [\hat{G}_{:,m-1} \quad \hat{g}_{:,m}] = \begin{bmatrix} \hat{G}_{n-1,:} \\ \hat{g}_{n,:} \end{bmatrix} \quad (6)$$

respectivamente. O rank da matriz de transferência é um importante parâmetro do sistema. **Lema .1.**

$$\eta_{In} \neq 1 \iff \text{rank}(G) = \text{rank}(G_{:,n-1}) \quad (7)$$

$$\eta_{In} = 0 \iff g_{:,n} = 0 \iff \hat{g}_{n,:} = 0 \quad (8)$$

A prova deste lema pode ser encontrada em (Cao *et al.*, 1997).

3.2. Mínimo Valor Singular (SV) e Seleção de Entrada

O mínimo valor singular é uma importante medida de controlabilidade. Nesta seção a implicação da seleção de entrada é considerada em termos da medida do mínimo SV.

Usando-se a notação da equação (5) e (6) e fazendo-se:

$$G_1 = [G_{:,n-1} \quad 0] \text{ e } G_2 = \begin{bmatrix} G_{m-1,:} \\ 0 \end{bmatrix},$$

o seguinte teorema descreve a "distância" entre G^+ e as sub-matrizes,

$$\| G_{:,n-1}^+ \|_2 \text{ e } \| G_{m-1,:}^+ \|_2.$$

Teorema .1. Se $\text{rank}(G) = \text{rank}(G_1) = \text{rank}(G_2)$, então

$$\| G^+ - G_1^+ \|_2^2 = \frac{\| \hat{g}_{n,:} \|_2^2}{1 - \eta_{In}^2} \quad (9)$$

Além disso, o seguinte limite pode ser obtido:
Se $\text{rank}(G) = \text{rank}(G_1) = \text{rank}(G_2)$, então

$$\|G^+\|_2 \leq \|G_1^+\|_2 \leq \|G^+\|_2 + \frac{\|\hat{g}_n\|_2}{\sqrt{1 - \eta_{In}^2}} \quad (10)$$

Para selecionar um subsistema o mais próximo possível do sistema original, a seguinte medida de distância pode ser usada como um critério de seleção de entrada, definida como:

$$\alpha_{Ij} = \frac{\|\hat{g}_j\|_2}{\sqrt{1 - \eta_{In}^2}} \quad (11)$$

Desta forma, o critério se baseia-se em eliminar a entrada ou entradas que tenham o menor valor de α_{Ij} , e o subsistema resultante tem o maior mínimo valor singular (Cao *et al.*, 1997).

4. Seleção das Variáveis de Entrada da Planta de Tratamento de Efluentes Utilizando Efetividade de Entrada e Projeção

Nesta seção é feita a seleção das variáveis de entrada da planta de tratamento de efluentes, utilizando-se os critérios e conceitos apresentados na seção 3. A variável de saída considerada é a concentração de amônia (S_{NH_3}). Os dados de amostras coletados foram excitados com sinais de amplitudes aleatórias utilizando a função "uniform random" (sinais randomicamente uniformes) do *Simulink / MATLAB*. Na saída considerada do processo, S_{NH_3} (concentração de amônia), foi adicionada sinais de ruído branco. O processo foi simulado durante 6.000 horas, considerando-se um período de amostragem T_s de 1,5 horas, coletando-se 4.001 amostras de dados de entradas e saída. Os dados foram coletados considerando-se o processo em malha aberta.

• Seleção das Variáveis Candidatas a Entrada para a Saída S_{NH_3} :

Seja u o vetor representando o conjunto das 15 variáveis candidatas a entrada do processo de tratamento de efluentes e y a saída de concentração de amônia (S_{NH_3}). Com os dados de saída e entrada coletados da planta de tratamento de efluentes, obtém-se G através da equação (1):

$$G = \begin{bmatrix} 1,0960 & -0,0004 & -0,0261 & -10,4785 & -44,5172 & -49,3680 & -0,0052 \\ -0,0062 & 0,1200 & 0,0054 & 0,1259 & 0,1570 & -0,1582 & -0,0099 & -0,0315 \end{bmatrix} \quad (12)$$

A tabela 4 apresenta os resultados obtidos de α_{Ij} , e η_{Ij}^2 para os critérios de seleção de entrada, em ordem decrescente em relação a η_{Ij}^2 , considerando-se a variável de saída S_{NH_3} .

Tabela 4: Resultados da seleção de entrada da planta de tratamento de efluentes para a saída S_{NH_3}

u_j	α_{Ij}^2	η_{Ij}^2
$u_6 = Q_{ar3}$	0,0160	0,5380
$u_5 = Q_{ar2}$	0,0131	0,4375
$u_4 = Q_w$	0,0023	0,0242
$u_1 = Q_{in}$	0,0002	0,0003
$u_2 = Q_{int}$	0,0000	0,0000
$u_3 = Q_{si}$	0,0000	0,0000
$u_7 = X_{BH}$	0,0000	0,0000
$u_8 = X_S$	0,0000	0,0000
$u_9 = X_{ND}$	0,0000	0,0000
$u_{10} = X_{IP}$	0,0000	0,0000
$u_{11} = S_{NH}$	0,0000	0,0000
$u_{12} = S_{ND}$	0,0000	0,0000
$u_{13} = S_{NO}$	0,0000	0,0000
$u_{14} = S_S$	0,0000	0,0000
$u_{15} = S_{ALK}$	0,0000	0,0000

Observa-se, pelos resultados apresentados, que para a maioria das entradas ambos os critérios obtidos possuem valores iguais a zero. Neste caso, ambos os critérios, η_{Ij} e α_{Ij} , indicam que as entradas selecionadas são $[Q_{ar3} \ Q_{ar2} \ Q_w \ Q_{in}]$. Desta forma, será analisado o sistema caso a caso, em relação às 4 variáveis a entrada, as quais obtiveram valores de η_{Ij} e α_{Ij} diferentes de zero.

O sistema referente às 4 entradas (1×4) citadas anteriormente, é dado por:

$$G_1 = [\ 1,0960 \ -10,4785 \ -44,5172 \ -49,3680 \] \quad (13)$$

Como o sistema analisado é em relação a uma saída (S_{NH3}), neste caso G_1 tem valor singular (SV) igual a 67,3051.

Para um sistema de 3 entradas, e observando-se que a variável Q_{in} apresentou menor valor de $\eta_{Ij}^2 = 0,0003$, então Q_{in} é eliminada nesta etapa, e o novo subsistema obtido torna-se:

$$G_2 = [\ -10,4785 \ -44,5172 \ -49,3680 \] \quad (14)$$

O sistema G_2 tem valor singular (SV) igual a 67,2962.

Para um sistema de 2 entradas, nesta etapa será eliminada a variável Q_w por apresentar menor valor de $\eta_{Ij}^2 = 0,0242$, sendo que o novo subsistema torna-se:

$$G_3 = [\ -44,5172 \ -49,3680 \] \quad (15)$$

O sistema G_3 tem valor singular (SV) igual a 66,4754.

Portanto, comparando-se os valores singulares dos sistemas calculados anteriormente (1×4 , 1×3 e 1×2) o sistema de 4 entradas apresenta o maior valor singular com $\sigma = 67,3051$, sendo que neste caso opta-se por manter as 4 variáveis de entrada selecionadas para a variável de saída S_{NH3} , a saber: Q_{in} , Q_w , Q_{ar2} e Q_{ar3} .

5. Análise de Componentes Principais (PCA)

Um dos métodos utilizados para tratamento de dados multivariados é o PCA (em Inglês - *Principal Component Analysis*), que consiste numa manipulação da matriz de dados com o objetivo de representar as variações presentes em muitas variáveis, através de um número menor de "fatores". Constrói-se um novo sistema de eixos (denominados rotineiramente de fatores, componentes principais, variáveis latentes ou ainda autovetores) para representar as amostras, no qual a natureza multivariada dos dados pode ser visualizada em poucas dimensões.

Matematicamente, o processo consiste na decomposição em valores singulares (SVD) de uma matriz X , não quadrada, em que cada amostra é representada numa linha, e as variáveis são as colunas. A matriz passa a ser representada pelo produto de três outras matrizes, a matriz V denominada matriz de "loadings" e o produto US , matriz de "scores". Um algoritmo comumente empregado é o de decomposição de valores singulares (SVD), sendo que a matriz de dados é escrita como:

$$X = USV^t \quad (16)$$

A matriz V é a matriz dos "loadings", sendo que a primeira coluna contém os "loadings" de PC1 e assim por diante. $U \cdot S$ corresponde à matriz T dos "scores". S é uma matriz diagonal cujos elementos (valores singulares) contém informação sobre a quantidade de variância que cada componente principal descreve. A matriz S é importante na determinação da dimensionalidade intrínseca da matriz de dados X (posto de X). Os autovalores que forem pequenos serão excluídos, e as informações relevantes podem, de alguma maneira, ser separadas, eliminando-se assim os ruídos experimentais. A expressão abaixo mostra a relação entre o valor singular e a variância contida na j -ésima componente principal.

$$VAR\%PCj = \frac{s_{jj}^2}{\sum_{j=1}^p s_{jj}^2} \times 100 \quad (17)$$

sendo que o denominador dá o valor da variância total e p é o número de valores singulares do conjunto de dados (o menor valor dentre n e m), sendo que n é o número de amostras e m o número de variáveis.

5.1. Análise de Componentes Principais Aplicada à Planta de Tratamento de Efluentes

A Análise de Componentes Principais (*PCA*) foi utilizada com o objetivo de se ter uma idéia do número necessário de componentes principais para descrever o conjunto de dados, relacionados à planta de tratamento de efluentes. A matriz de dados X consiste, neste caso, em 16 colunas ($m = 16$, sendo as 15 primeiras colunas relacionadas às variáveis candidatas a entrada e a última coluna representando a variável de saída) nas 4001 amostras ($n = 4001$). Os dados experimentais originais podem não ter uma distribuição adequada para análise, dificultando a extração de informações úteis e interpretação dos mesmos. Nestes casos, um pré-processamento nos dados originais pode ser de grande valia. Medidas em diferentes unidades e variáveis com diferentes variâncias são algumas das razões que levam a estes problemas. Os métodos de pré-processamento mais utilizados consistem basicamente em centrar na média ou autoescalar os dados. O primeiro tipo de pré-processamento consiste em calcular a média dos dados de determinada coluna (variável) e, em seguida, subtraí-la de cada dado da referida coluna. Assim, as médias das variáveis são igualadas a zero. No segundo tipo, são calculadas as médias e os desvios padrão, em seguida subtrai-se cada valor da média e divide-se o resultado pelo desvio padrão. Portanto, a média de cada variável será zero e o desvio padrão será igual a 1.

O pré-processamento dos dados realizado foi o de autoescalar. Foi utilizado o *software MATLAB* para análise dos dados.

A análise foi feita, considerando-se a variável de saída com as possíveis variáveis candidatas a entrada.

Os resultados da porcentagem de variância explicada em cada Componente Principal (*PC*) estão na tabela 5.

Tabela 5: Porcentagem de variância obtida com a análise de componentes principais (*PCA*) considerando-se a variável de saída S_{NH_3}

Porcentagem de variância explicada pela análise de <i>PCA</i>		
(<i>PC</i>)	% Variância desta <i>PC</i>	% Variância total
1	10,4582	10,4582
2	9,2381	19,6963
3	7,6387	27,3350
4	7,0634	34,3984
5	6,9037	41,3021
6	6,5442	47,8463
7	6,4114	54,2577
8	6,1094	60,3671
9	6,0648	66,4319
10	5,9907	72,4226
11	5,7103	78,1329
12	5,3541	83,4870
13	5,1623	88,6493
14	5,0025	93,6518
15	3,3189	96,9707
16	3,0294	100,0001

Analisando-se a porcentagem de variância de cada Componente Principal (*PC*) e a variância total, observa-se que seria necessário no mínimo 7 componentes principais para poder representar acima de 50% da variância total.

A tabela 6 ilustra os resultados da variância explicada pelas duas primeiras componentes principais, em relação a cada variável candidata a entrada. Estes resultados foram obtidos do seguinte modo:

- A proporção de variância calculada para cada variável é dada, elevando-se o primeiro elemento ao quadrado e somando-o com o segundo, também elevado ao quadrado, sendo que estes elementos são obtidos da matriz de "loadings". Esta análise, se baseou conforme apresentado em (Dunteman, 1989).

Tabela 6: Proporção de variância calculada para a saída S_{NH_3} pelas duas primeiras componentes principais

Proporção de Variância	
(Variável)	Proporção de Variância Calculada
Q_{in}	0,3103
Q_{int}	0,0022
Q_{sl}	0,1938
Q_w	0,0172
Q_{ar2}	0,0444
Q_{ar3}	0,0029
X_{BH}	0,0045
X_S	0,4138
X_{ND}	0,0918
X_{IP}	0,0015
S_{NH}	0,0578
S_{ND}	0,0162
S_{NO}	0,0129
S_S	0,3500
S_{ALK}	0,0008

Desta forma, as 7 variáveis de maior influência representadas pela proporção de variância explicada pelas duas primeiras componentes principais para a variável de saída S_{NH_3} , são: X_S , S_S , Q_{in} , Q_{sl} , X_{ND} , S_{NH} e Q_{ar2} .

Os "loadings" em relação à primeira componente e à segunda componente, são representados na figura a seguir. A primeira (PC1) e a segunda (PC2) componentes principais estão relacionadas respectivamente com a primeira e com a segunda coluna obtidos da matriz de "loadings".

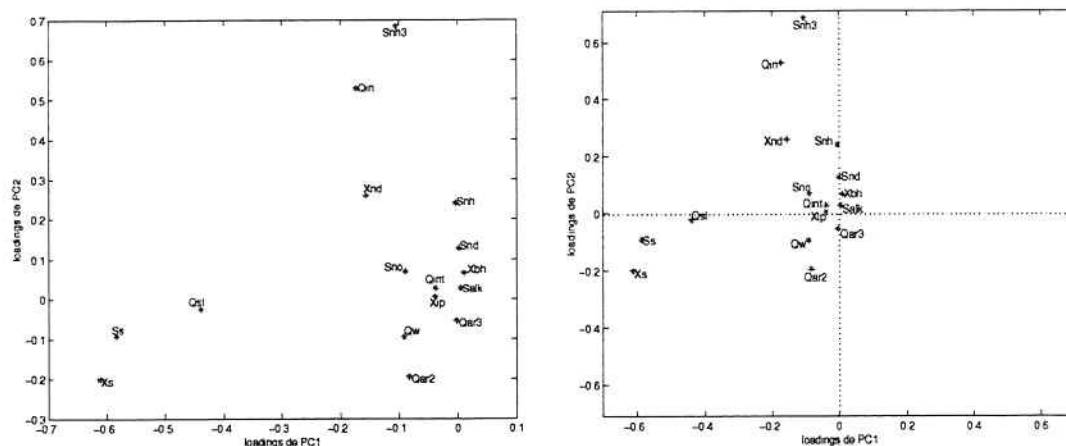


Figura 2: "Loadings" da primeira e segunda componentes em relação à variável de saída S_{NH_3}

6. Funções de Correlação

Quando se pode demonstrar que existe associação entre duas variáveis quantitativas, isto é, quando se constata que uma variável afeta a outra, diz-se que as variáveis estão correlacionadas.

Pode-se escrever a *Função de Correlação Cruzada (FCC)* entre dois sinais $u(t)$ e $y(t)$ como:

$$r_{uy}(\tau, t) = E[u(t)y(t + \tau)] \quad (18)$$

sendo $E[\cdot]$ a esperança matemática, τ o atraso e t o tempo.

A *função de autocorrelação (FAC)* de um sinal $u(t)$ pode ser estimada usando-se:

$$r_{uu}(\tau, t) = E[u(t)u(t + \tau)] \quad (19)$$

7. Metodologia Proposta para Seleção das Variáveis de Entrada

A seguir são descritos alguns procedimentos para seleção de variáveis de entrada, a partir da função de correlação. Em (Aguirre, 2000), a função de correlação é utilizada como um procedimento para auxiliar na seleção de variáveis de entrada. Após se verificar a existência ou não de correlação entre as variáveis candidatas a entrada e a saída desejada, verifica-se qual ou quais variáveis candidatas a entrada estão correlacionadas entre si e, caso duas variáveis candidatas a entrada estejam correlacionadas entre si, despreza-se uma delas.

Todavia, por se tratar de um procedimento com aspecto subjetivo, nesta tese, após a realização de alguns testes utilizando-se uma análise com critérios estatísticos, o seguinte procedimento para a seleção de variáveis candidatas a entrada a partir de função de correlação é adotado:

Passo 1: Calculam-se as funções de correlação cruzada entre cada uma das variáveis candidatas a entrada e a saída desejada. Verifica-se qual ou quais das variáveis candidatas a entrada têm correlação com a saída, verificando-se o valor de correlação apresentando na *FCC* de cada uma delas. É importante ressaltar que as variáveis candidatas a entrada que na função de correlação cruzada não estiverem correlacionadas com a saída desejada, são eliminadas nas demais análises dos passos seguintes.

Passo 2: Após verificar qual ou quais variáveis candidatas a entrada estão correlacionadas com a saída, calcula-se as funções de correlação cruzada entre todas as possíveis variáveis candidatas a entrada.

Passo 3: Dentre os pares de variáveis candidatas a entrada que estão correlacionadas entre si, analisando-se par a par, seleciona-se dentre cada par a variável que apresentar maior *FCC* com a saída desejada. E caso não se constate evidência de correlação entre duas variáveis candidatas a entrada, ambas devem ser incluídas inicialmente no critério de seleção da saída desejada.

Passo 4: Verifica-se em cada grupo de variáveis candidatas a entrada, a variável que apresentar maior *FCC*.

7.1. Análise da Função de Correlação Cruzada Entre Entradas e Saídas

O objetivo principal de se analisar a função de correlação cruzada entre as variáveis candidatas a entrada e a saída desejada, é de verificar a existência ou não de correlação entre elas. Se uma variável candidata a entrada do modelo, não afeta a saída, então a presença da mesma não se justifica.

Em um primeiro momento, com os dados coletados realizou-se o teste de correlação cruzada, para verificar se as variáveis candidatas a entrada estavam correlacionadas com a saída desejada. As relações existentes, entre as variáveis candidatas a entrada e a saída desejada da planta de tratamento de efluentes, serão explicadas através dos gráficos de correlação cruzada. Antes de se apresentar alguns resultados dos gráficos obtidos da função de correlação entre as variáveis candidatas a entrada e a saída desejada é importante introduzir o intervalo de confiança a ser considerado, pois resalta-se que os pontos pontilhados nos gráficos da função de correlação cruzada, indicam um intervalo de confiança de 95%. Como os sinais usados são aleatórios, e considerando-se uma variável aleatória x cuja distribuição é *gaussiana* com média $\bar{x}$ e desvio padrão σ_x , afim de satisfazer o intervalo de confiança de 95%, então para uma determinada realização de x_i de x , a probabilidade de x_i estar compreendida no intervalo de 95% é:

$$\bar{x} - Z\sigma_x < x_i < \bar{x} + Z\sigma_x \quad (20)$$

Portanto, para se obter a estimativa do intervalo de confiança, teremos

$$\bar{x} \pm Z \frac{\sigma}{\sqrt{N}} \quad \text{ou} \quad \bar{x} - Z \frac{\sigma}{\sqrt{N}} \leq x_i \leq \bar{x} + Z \frac{\sigma}{\sqrt{N}} \quad (21)$$

sendo que $\frac{\sigma}{\sqrt{N}}$ representa o erro padrão da média aritmética de σ_x , N indica o comprimento do registro dos dados e Z representa o valor correspondente a uma área por $(1 - \alpha) \times 100\%$. α é a proporção de *caudas da distribuição gaussiana* que estão fora do intervalo de confiança. Um *nível de confiança* de 95% conduz a um valor de Z igual a $\pm 1,96$. Então para satisfazer o intervalo de confiança, os sinais aleatórios coletados foram convertidos para Z unidades padronizadas

$$Z = \frac{\bar{x} - x_i}{\sigma_x} \quad (22)$$

Em identificação de sistemas, devido à forma com que os testes e coleta de dados são efetuados, normalmente trabalha-se com sinais de média nula. Neste caso, considerando-se

uma distribuição gaussiana de média zero ($\bar{x} = 0$) e variância ($\sigma_x^2 = 1$), tem-se o desvio padrão ($\sigma = 1$). Portanto, para o intervalo de confiança de 95%, os limites inferiores e superiores tornam-se, desta forma, $\pm \frac{1,96}{\sqrt{N}}$, conforme equação (21) e (7.1). Os dados após serem convertidos para Z unidades padronizadas, foram normalizados antes da análise da função de correlação cruzada (FCC). As figuras a seguir ilustram graficamente as funções de correlação cruzada (FCC) entre a saída desejada e algumas das variáveis candidatas a entrada da planta de tratamento de efluentes.

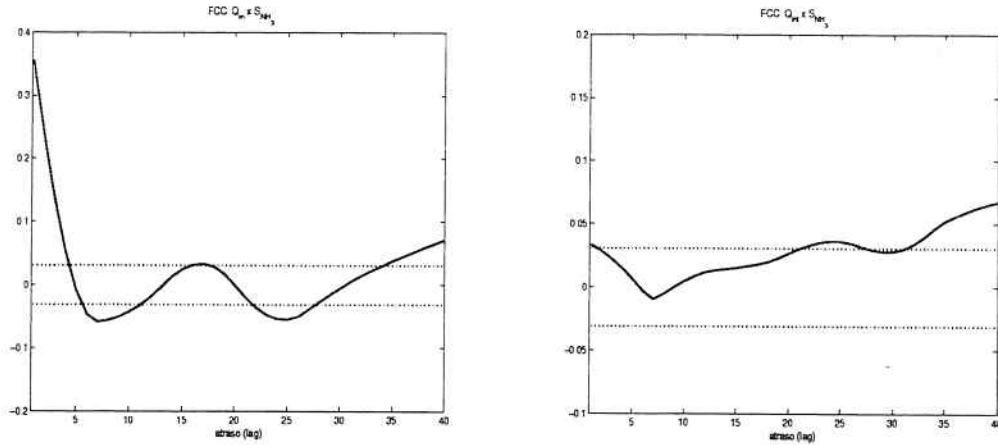


Figura 3: FCC das variáveis candidatas a entrada Q_{in} , Q_{int} e a variável de saída S_{NH_3}

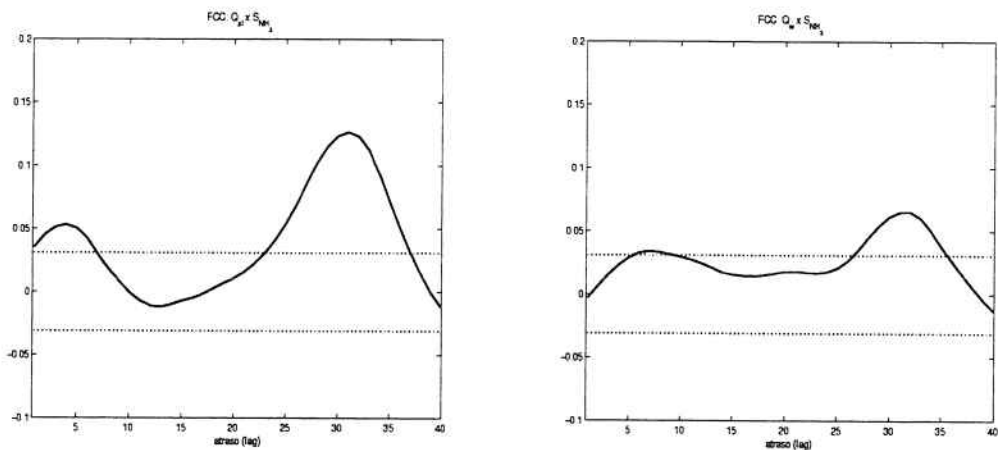


Figura 4: FCC das variáveis candidatas a entrada Q_{sl} , Q_w e a variável de saída S_{NH_3}

Em análise da variável de saída com as variáveis candidatas a entrada, foi calculado o maior valor absoluto da FCC . A tabela a seguir ilustra os resultados apresentados.

Tabela 7: FCC entre as variáveis candidatas a entrada e a saída desejada

E/S	Q_{in}	Q_{int}	Q_{sl}	Q_w	Q_{ar2}	Q_{ar3}	X_{BH}	X_S
S_{NH_3}	0,3556	0,0677	0,1267	0,0652	0,1021	0,0604	0,0528	0,0516

E/S	X_{ND}	X_{IP}	S_{NH}	S_{ND}	S_{NO}	S_S	S_{ALK}
S_{NH_3}	0,1832	0,0566	0,1485	0,1253	0,0449	0,0518	0,0364

Ao se analisar os gráficos obtidos de FCC das variáveis de saída, S_{NH_3} (concentração de amônia) com as variáveis candidatas a entrada, pode-se verificar que Q_{in} , Q_{int} , Q_{sl} , Q_w , Q_{ar2} , Q_{ar3} , X_{BH} , X_S , X_{ND} , X_{IP} , S_{NH} , S_{ND} , S_S e S_{ALK} estão ambas correlacionadas com a variável de saída, S_{NH_3} . A tabela a seguir ilustra a existência de correlação entre cada variável de entrada e a saída desejada.

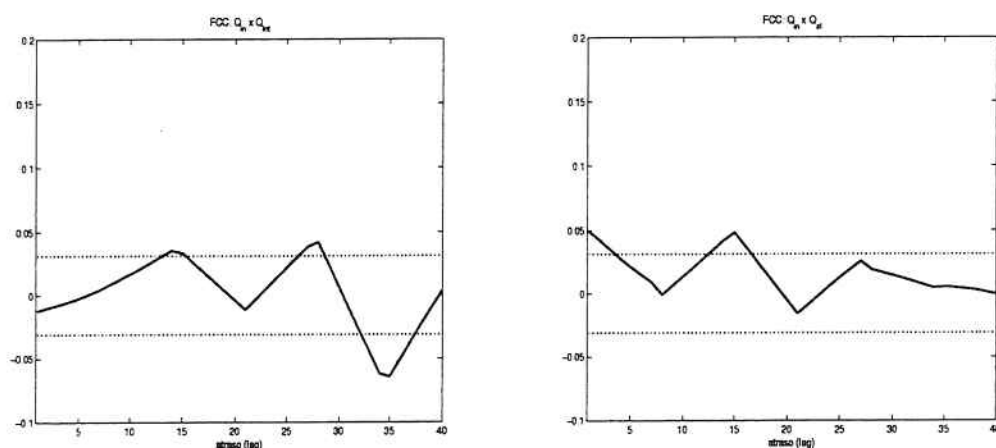
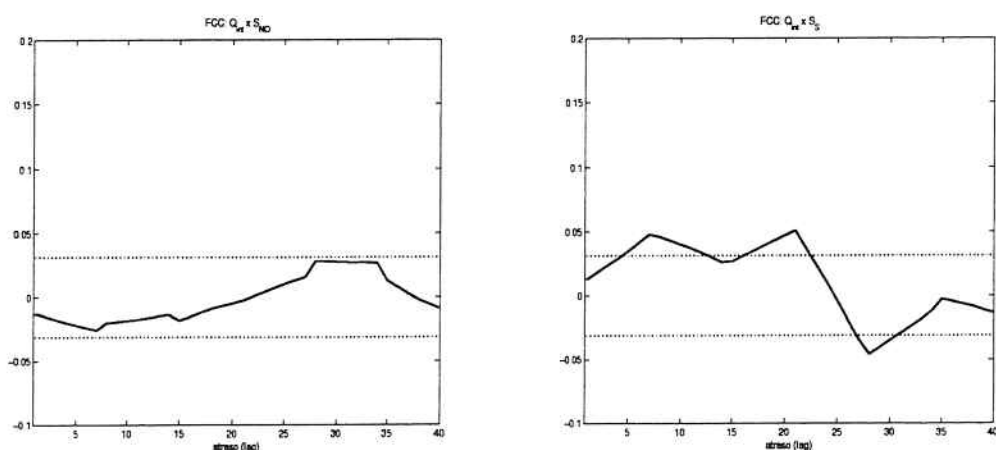
Tabela 8: FCC entre as variáveis candidatas a entrada e a saída desejada

E/S	Q_{in}	Q_{int}	Q_{sl}	Q_w	Q_{ar2}	Q_{ar3}	X_{BH}	X_S
S_{NH_3}	C	C	C	C	C	C	C	C

E/S	X_{ND}	X_{IP}	S_{NH}	S_{ND}	S_{NO}	S_S	S_{ALK}
S_{NH_3}	C	C	C	C	C	C	C

7.2. Análise de Função de Correlação Cruzada Entre Entradas

Após verificar a existência de correlação entre as variáveis candidatas a entrada e a saída desejada, aplica-se então a função de correlação cruzada entre todas as possíveis variáveis candidatas a entrada, que estão correlacionadas com a saída. Caso verifique que um par de variáveis de entrada estão correlacionadas, uma delas é candidata a ser eliminada, o qual o primeiro critério de eliminação, parte-se da análise de função de correlação cruzada calculada entre as variáveis de entrada e a saída desejada. Os gráficos a seguir indicam as respostas dos testes de correlação cruzada entre algumas das candidatas a variáveis de entrada.

Figura 5: FCC das variáveis candidatas a entrada $Q_{in} \times Q_{int}$, $Q_{ar2} \times Q_{sl}$ Figura 6: FCC das variáveis candidatas a entrada $Q_{int} \times S_{NO}$, $Q_{int} \times S_S$

Analisando-se as funções de correlação cruzada das variáveis candidatas a entradas, pode-se observar na tabela a seguir a existência de correlação ou não entre cada uma delas. Esta análise é feita através dos gráficos da função de correlação, sendo que caso a FCC se mantenha dentro do intervalo de confiança de 95%, não apresentando picos significativos fora deste

intervalo ² conclui-se que não existe correlação entre o par de variáveis analisadas. As tabelas a seguir ilustram os valores calculados da *FCC* entre as variáveis candidatas a entrada.

Tabela 9: *FCC* entre as variáveis candidatas a entrada

<i>E/E</i>	<i>Q_{in}</i>	<i>Q_{int}</i>	<i>Q_{sl}</i>	<i>Q_w</i>	<i>Q_{ar2}</i>	<i>Q_{ar3}</i>	<i>X_{BH}</i>	<i>X_S</i>
<i>Q_{in}</i>	-	0,0646	0,0494	0,0910	0,0668	0,0540	0,0529	0,0617
<i>Q_{int}</i>		-	0,0737	0,0628	0,0936	0,0418	0,1211	0,0678
<i>Q_{sl}</i>			-	0,0837	0,0610	0,0686	0,1059	0,2766
<i>Q_w</i>				-	0,0622	0,0411	0,0341	0,0482
<i>Q_{ar2}</i>					-	0,1128	0,0602	0,0437
<i>Q_{ar3}</i>						-	0,0750	0,0326
<i>X_{BH}</i>							-	0,0867
<i>X_S</i>								-

<i>E/E</i>	<i>X_{ND}</i>	<i>X_{IP}</i>	<i>S_{NH}</i>	<i>S_{ND}</i>	<i>S_{NO}</i>	<i>S_S</i>	<i>S_{ALK}</i>
<i>Q_{in}</i>	0,0752	0,0625	0,0464	0,0600	0,3556	0,0845	0,0369
<i>Q_{int}</i>	0,0592	0,0685	0,0592	0,0860	0,0282	0,0510	0,0706
<i>Q_{sl}</i>	0,0741	0,0562	0,0371	0,0601	0,0429	0,1598	0,0751
<i>Q_w</i>	0,0573	0,0754	0,0981	0,0679	0,0565	0,0741	0,0484
<i>Q_{ar2}</i>	0,0963	0,0441	0,0512	0,0528	0,0861	0,0662	0,0748
<i>Q_{ar3}</i>	0,0809	0,0599	0,1018	0,0700	0,0570	0,0435	0,0598
<i>X_{BH}</i>	0,0752	0,0619	0,0601	0,0566	0,0686	0,0825	0,0678
<i>X_S</i>	0,0378	0,0437	0,0689	0,0391	0,0545	0,4745	0,0548

<i>E/E</i>	<i>X_{ND}</i>	<i>X_{IP}</i>	<i>S_{NH}</i>	<i>S_{ND}</i>	<i>S_{NO}</i>	<i>S_S</i>	<i>S_{ALK}</i>
<i>X_{ND}</i>	-	0,0615	0,0521	0,0366	0,0504	0,0704	0,0685
<i>X_{IP}</i>		-	0,0855	0,0509	0,0779	0,0693	0,0689
<i>S_{NH}</i>			-	0,0607	0,0654	0,0711	0,0544
<i>S_{ND}</i>				-	0,0669	0,0539	0,0539
<i>S_{NO}</i>					-	0,0741	0,0364
<i>S_S</i>						-	0,0513
<i>S_{ALK}</i>							-

Conforme os resultados apresentados, na tabela a seguir, "C" significa correlacionada e "NC" significa não correlacionada.

Tabela 10: *FCC* - Entrada x Entrada

<i>E/E</i>	<i>Q_{in}</i>	<i>Q_{int}</i>	<i>Q_{sl}</i>	<i>Q_w</i>	<i>Q_{ar2}</i>	<i>Q_{ar3}</i>	<i>X_{BH}</i>	<i>X_S</i>
<i>Q_{in}</i>	-	C	C	C	C	C	C	C
<i>Q_{int}</i>		-	C	C	C	C	C	C
<i>Q_{sl}</i>			-	C	C	C	C	C
<i>Q_w</i>				-	C	C	C	C
<i>Q_{ar2}</i>					-	C	C	C
<i>Q_{ar3}</i>						-	C	C
<i>X_{BH}</i>							-	C
<i>X_S</i>								-

²O valor calculado do intervalo de confiança foi igual a 0,0310, portanto, todo valor de *FCC* acima do intervalo de confiança, foi considerado como existência de correlação entre duas variáveis candidatas a entradas

E/E	X_{ND}	X_{IP}	S_{NH}	S_{ND}	S_{NO}	S_S	S_{ALK}
Q_{in}	C	C	C	C	C	C	C
Q_{int}	C	C	C	C		C	C
Q_{sl}	C	C	C	C	C	C	C
Q_w	C	C	C	C	C	C	C
Q_{ar2}	C	C	C	C	C	C	C
Q_{ar3}	C	C	C	C	C	C	C
X_{BH}	C	C	C	C	C	C	C
X_S	C	C	C	C	C	C	C

E/E	X_{ND}	X_{IP}	S_{NH}	S_{ND}	S_{NO}	S_S	S_{ALK}
X_{ND}	-	C	C	C	C	C	C
X_{IP}		-	C	C	C	C	C
S_{NH}			-	C	C	C	C
S_{ND}				-	C	C	C
S_{NO}					-	C	C
S_S						-	C
S_{ALK}							-

8. Variáveis Seleccionadas para a Planta de Tratamento de Efluentes

Analisando-se a tabela 7 da função de correlação cruzada entre as variáveis candidatas a entrada, e levando-se em consideração a FCC entre cada variável candidata a entrada e a saída desejada, e conforme o procedimento de seleção de variáveis de entrada descrito na seção 7, foram primeiramente rastreadas as variáveis em cada combinação de entradas, com a respectiva saída desejada. A tabela 11 ilustra os resultados obtidos para a saída S_{NH_3} (concentração de amônia).

Tabela 11: Rastreamento das variáveis candidatas a entrada (par a par) para a saída S_{NH_3}

Grupo de variáveis	Variáveis rastreadas
Grupo 1: Q_{in}	Q_{in} (14 grupos)
Grupo 2: Q_{int}	Q_{sl}, Q_{int} (6 grupos), $Q_{ar2}, X_{ND}, S_{NH}, S_{ND}, S_{NO}$
Grupo 3: Q_{sl}	Q_{sl} (10 grupos), X_{ND}, S_{NH}
Grupo 4: Q_w	Q_{ar2}, Q_w (7 grupos), X_{ND}, S_{NH}, S_{ND}
Grupo 5: Q_{ar2}	Q_{ar2} (7 grupos), X_{ND}, S_{NH}, S_{ND}
Grupo 6: Q_{ar3}	Q_{ar3} (6 grupos), X_{ND}, S_{NH}, S_{ND}
Grupo 7: X_{BH}	X_{BH} (4 grupos), $X_{ND}, X_{IP}, S_{NH}, S_{ND}$
Grupo 8: X_S	$X_{ND}, X_{IP}, S_{NH}, S_{ND}, X_S$ (2 grupos), S_S
Grupo 9: X_{ND}	X_{ND} (6 grupos)
Grupo 10: X_{IP}	S_{NH}, S_{ND}, X_{IP} (3 grupos)
Grupo 11: S_{NH}	S_{NH} (4 grupos)
Grupo 12: S_{ND}	S_{ND} (3 grupos)
Grupo 13: S_{NO}	S_S, S_{NO}
Grupo 14: S_S	S_S

Observa-se pela tabela 11 que a variável Q_{in} foi mais significativa com respeito à influência que todas as outras variáveis, sendo neste caso uma variável que não será eliminada. Percebe-se, pela tabela 11 que a variável S_{ALK} não apareceu nenhuma vez em nenhum dos grupos, o que implica que esta variável já está eliminada para a saída S_{NH_3} . Propõe-se então adotar o seguinte critério: em cada grupo citado nas tabela 11, seleccionar a variável de entrada que apresentar maior FCC em valor absoluto com a saída. A ideia aqui é privilegiar apenas as variáveis que tenham maior influência sobre a saída.

A tabela 12 ilustra a quantidade de ocorrência nos grupos, considerando-se todas as variáveis candidatas a entrada com relação à respectiva saída desejada.

Tabela 12: Número de ocorrências nos grupos de cada variável candidata a entrada para a saída S_{NH_3}

Variáveis candidata a entrada	Número de ocorrências das variáveis nos grupos
Q_{in}	14 vezes (em relação à todas as outras variáveis)
X_{ND}	13 vezes
S_{NH}	12 vezes
Q_{sl}, S_{ND}	10 vezes
Q_{ar2}	9 vezes
Q_w	7 vezes
Q_{int}, Q_{ar3}	6 vezes
X_{IP}	5 vezes
X_{BH}	4 vezes
S_S	3 vezes
X_S, S_{NO}	2 vezes
S_{ALK}	Nenhuma vez

Desta forma, na tabela 13 tem-se para cada grupo de variáveis descrito anteriormente, em relação à saída, a variável de entrada selecionada, respectivamente para a saída S_{NH_3} .

Tabela 13: Seleção de variáveis candidatas a entrada de maior FCC em cada grupo para a saída S_{NH_3}

Grupo de variáveis	Variáveis rastreadas
Grupo 1: Q_{in}	Q_{in}
Grupo 2: Q_{int}	X_{ND}
Grupo 3: Q_{sl}	X_{ND}
Grupo 4: Q_w	X_{ND}
Grupo 5: Q_{ar2}	X_{ND}
Grupo 6: Q_{ar3}	X_{ND}
Grupo 7: X_{BH}	X_{ND}
Grupo 8: X_S	X_{ND}
Grupo 9: X_{ND}	S_{ND}
Grupo 10: X_{IP}	S_{NH}
Grupo 11: S_{NH}	S_{NH}
Grupo 12: S_{ND}	S_{ND}
Grupo 13: S_{NO}	S_S
Grupo 14: S_S	S_S

Portanto, tomando-se as variáveis escolhidas para cada grupo na tabela 13 e considerando-se os procedimentos e critérios descritos anteriormente foram selecionadas as seguintes variáveis de entradas para a saída desejada (S_{NH_3}): Q_{in} , X_{ND} , S_{NH} , S_{ND} , S_S . Observa-se portanto, que dentro de um conjunto inicial de 15 variáveis candidatas a entrada, foram selecionadas 5 variáveis para a saída S_{NH_3} (concentração de amônia).

9. Identificação da Planta de Tratamento de Efluentes

Nesta seção é realizada a identificação da planta de tratamento de efluentes, utilizando-se modelos NARX polinomiais. A estrutura do modelo NARX ("Non-Linear Auto-Regressive with eXogenous inputs- não-linear, auto-regressivo, com entradas externas), o qual pode representar os modelos observáveis e controláveis de uma ampla classe de sistemas não-lineares em tempo discreto (Leontaritis and Billings, 1985); (Chen and Billings, 1989) é dada seguir.

$$y(t) = f^l[y(t-1), y(t-2), \dots, y(t-n_y+1), u(t-d), u(t-d-1), u(t-d-2), \dots, u(t-d-n_u+1)] + e(t) \quad (23)$$

sendo que $t = 1 \dots N$, f^l representa um vetor de função não-linear qualquer, com grau de não-linearidade l ; $y(t)$, $u(t)$, e $e(t)$ são respectivamente, saída, entrada e ruído do sistema. n_y e n_u representam atrasos máximos, respectivamente nos dados de saída e entrada do sistema. d é o atraso puro de tempo medido em intervalos de amostragem τ_s . O modelo NARX é uma simplificação do modelo NARMAX ("Non-Linear Auto-Regressive Moving with eXogenous inputs- não-linear, auto-regressivo, com média móvel e entradas externas), obtido pela não incorporação dos erros atrasados ou residuais no processo de identificação. Uma descrição mais completa desta classe de modelos não-lineares pode ser vista em classe de modelos NARMAX polinomiais.

A determinação da estrutura ou quais termos a serem incluídos no modelo é essencial na estimação de parâmetros não-lineares, visto que o modelo **NARMAX** pode facilmente tornar-se sobreparametrizado. Uma estimação direta baseada em uma expansão polinomial pode envolver um número excessivo de termos, sendo que o número de termos possíveis em modelos polinomiais cresce bastante com o aumento do grau de não-linearidade l e com a ordem dos termos dinâmicos (n_u, n_y, n_e) no caso de modelos **NARMAX** e (n_u, n_y) no caso de modelos **NARX**. Sendo que n_y, n_u, n_e , representam, respectivamente, os atrasos máximos da saída, entrada e ruído do sistema.

O número de todos os termos possíveis em modelos polinomiais pode ser determinado através das expressões que serão apresentadas a seguir, tanto para o caso de modelos **SISO** como **MIMO**. O número máximo de coeficientes da classe do modelo **NARMAX**, dado pela equação (23), representando um modelo **SISO**, conforme (Korenberg *et al.*, 1988), é dado por:

$$n_\theta = M + 1 \quad (24)$$

no qual n_θ é considerado o número de termos do processo e do ruído do modelo, sendo que:

$$M = \sum_{i=1}^l n_i$$

$$n_i = \frac{\{n_{i-1}(n_y + n_u + n_e = i - 1)\}}{i}, \quad n_0 = 1 \quad (25)$$

O termo $M + 1$, dado na equação (24), indica que está sendo considerado o termo *bias* no modelo. Em (Jerônimo, 1998) são apresentados exemplos numéricos de como se determinar o número de todos os termos possíveis em modelos polinomiais. A detecção de estrutura é considerada a tarefa mais difícil e mais importante no processo de identificação; devido ao fato de que as características dinâmicas e estáticas do sistema estão intimamente ligadas à sua estrutura. No entanto, alguns modelos podem incluir termos que não são necessários quando se procura um melhor ajuste dos dados de identificação para representar dinamicamente o sistema, sendo que estes modelos tendem a se tornar excessivamente complexos. Em grande parte, a principal inconveniência de considerar estruturas de modelos com um grande número de parâmetros está relacionada com o tempo computacional e problemas numéricos. Pela razão do número de termos possíveis em modelos **NARMAX** polinomiais ser muito grande (tipicamente centenas), o algoritmo ortogonal é uma das técnicas mais eficientes para solucionar o problema de *detecção de estrutura*, visto que os termos significativos do modelo podem ser avaliados baseados em "taxa de redução de erro" (*ERR*).

O algoritmo de estimação de parâmetros dos modelos **NARX** polinomiais para o processo de identificação desta tese, é uma versão modificada do método "Clássico de Gram-Schmidt" (*CGS*) intitulada método de "Gram-Schmidt Modificado" (*MGS*), apresentada em (Chen *et al.*, 1989). Este algoritmo é denominado de *algoritmo ortogonal de regressão em avanço* usando o procedimento de ortogonalização "Gram-Schmidt Modificado" (*MGS*). Este algoritmo é uma extensão do algoritmo de estimação ortogonal proposto por (Korenberg *et al.*, 1988), (Billings *et al.*, 1988). O algoritmo de estimação ortogonal tem provado ser eficiente na determinação dos termos significativos e provê estimativa dos parâmetros correspondentes. Em algoritmo (*MGS*) para identificação de uma processo de neutralização de pH. Todavia, uma descrição mais profunda deste algoritmo pode ser encontrada por exemplo, além das literaturas anteriormente citadas, em (Jerônimo, 1998), (Aguirre, 2000). O modelo de regressão linear é dado por:

$$z(t) = \sum_{j=0}^{n_\theta} p_j(t)\theta_j + \xi(t), \quad t = 1, \dots, N, \quad (\forall j = 0, \dots, n_\theta) \quad (26)$$

sendo que na análise de regressão linear, $z(t)$ é a variável dependente, na qual representa a saída $y(t)$; $p_j(t)$ são os regressores ou preditores, os quais são obtidos dos diferentes termos do polinômio representando os monômios até o grau l , ressaltando que $p_0(t) = 1$; $\xi(t)$ é algum erro de modelamento (também chamado de erro de predição); θ_j são os parâmetros não conhecidos a serem estimados, no qual θ_0 representa o termo *bias* ($\theta_0 = dc$), n_θ é o número total de termos (de processo e de ruído) no modelo e N é o comprimento dos dados.

A taxa de redução de erro (*error reduction ratio*) ou *ERR* (Billings *et al.*, 1989) associa a cada termo candidato um índice correspondente à contribuição deste na explicação da variância dos dados de saída. A taxa de redução de erro (*error reduction ratio*) ou *ERR* devida à componente w_j é então definida conforme (Chen *et al.*, 1989) e (Billings and Chen, 1989):

$$[ERR]_j = \frac{g_j^2 \langle w_j, w_j \rangle}{\langle z, z \rangle}, \quad 1 \leq j \leq n_\theta \quad \text{ou} \quad [ERR] = \frac{g_j^2 w_j^T w_j}{z^T z}, \quad 1 \leq j \leq n_\theta \quad (27)$$

Portanto, todos os regressores $p_j(t)$ são considerados na determinação de cada $w_j(t)$, considerando-se como termo escolhido aquele que apresentar o maior w_j . w_j indica o termo incluído e g_j o seu respectivo parâmetro. A "taxa de redução de erro" (ERR) oferece um simples e eficaz meio de selecionar um subconjunto de termos significativos de um amplo número de termos candidatos em um modo de regressão em avanço. A seleção prossegue até que:

$$1 - \sum_{j=0}^{n_{\theta_s}} [ERR]_j < \rho, \quad n_{\theta_s} < n_{\theta} \quad (28)$$

sendo que ρ é a tolerância desejada, sendo limitada respectivamente em $(0 < \rho < 1)$, (Billings and Chen, 1989) ou então (Chen *et al.*, 1989) e esta conduz a um subconjunto de termos do modelo de n_{θ_s} (Billings and Chen, 1989), sendo que n_{θ_s} é o número de termos escolhidos, ou seja, os termos significativos. Tem-se que $(1 - \sum [ERR]_j)$ é a porção da variância da saída explicada pela inclusão de um novo termo no modelo (Billings and Chen, 1989).

Além da "taxa de redução de erro" (ERR) foram utilizado no algoritmo de regressão em avanço dois critérios de parada baseados no Critério de Informação de Akaike (AIC) e no Critério de Informação de Bayes (BIC). No entanto, o critério da "taxa de redução de erro" (ERR) foi auxiliado pelos critérios de Akaike e de Bayes. A combinação do procedimento MGS auxiliado pelos critérios (AIC) e (BIC) foi utilizada em (Pröll, 1993), (Pröll and Karim, 1994), em que os resultados apresentados foram considerados consistentes. Ambos os critérios levam em conta o desempenho e a complexidade do modelo. A utilização destes critérios auxiliam na determinação do número de termos, no entanto, seus procedimentos não distinguem quais termos devem ser incluídos no modelo, sendo isto portanto, realizado pelo critério da "taxa de redução de erro" (ERR), que propõe a discriminação de acordo com o seu grau de representatividade no modelo.

9.1. Identificação com modelo NARX Polinomial - Sem Seleção de Variáveis de Entrada (SSVE)

Apresentam-se os resultados obtidos com o modelo NARX polinomial, sem a seleção de variáveis de entrada, ou seja, utilizando-se todas as variáveis candidatas a entrada.

A tabela 14 ilustra os resultados dos parâmetros e os regressores obtidos na identificação do modelo da concentração de amônia (S_{NH_3}). O modelo NARX polinomial foi obtido utilizando-se grau de não-linearidade ($l = 2$), $n_y = 4$, $n_u = 4$. De um total de 2145 termos do modelo ($n_{\theta} = 2145$), com a utilização do ERR foram selecionados 73 termos ($n_{\theta_s} = 73$).

Tabela 14: Modelo NARX polinomial para a saída S_{NH_3} - Modelo sem seleção de variáveis de entrada

-0,001942; 0,042266$Q_{in}(t-1)$; 0,015161$Q_{int}(t-1)$; 0,003303$Q_{sl}(t-1)$; 0,000365$Q_w(t-1)$; -0,012952$Q_{ar2}(t-1)$; -0,014249$Q_{ar3}(t-1)$; 0,000014$X_{BH}(t-1)$; -0,009968$X_S(t-1)$; 0,019574$X_{ND}(t-1)$; -0,000151$X_{IP}(t-1)$; 0,020663$S_{NH}(t-1)$; 0,017091$S_{ND}(t-1)$; -0,000007$S_{NO}(t-1)$; -0,005926$S_S(t-1)$; -0,000027$S_{ALK}(t-1)$; 0,218998$S_{NH_3}(t-1)^2$; -0,029064$S_{NH_3}(t-1)Q_{in}(t-1)$; -0,003060$S_{NH_3}(t-1)Q_{int}(t-1)$; 0,002443$S_{NH_3}(t-1)Q_{sl}(t-1)$; 1,382635$S_{NH_3}(t-1)$; -0,302121$S_{NH_3}(t-2)$; -0,029285$Q_{int}(t-2)$; -0,131303$S_{NH_3}(t-3)$; 0,005604$Q_{int}(t-3)$; 0,008690$Q_{in}(t-2)X_{ND}(t-2)$; 0,012303$Q_{ar3}(t-2)$; -0,009119$S_{NH}(t-3)$; -0,010630$Q_{sl}(t-2)$; 0,006656$X_S(t-3)$; -0,232822$S_{NH_3}(t-1)S_{NH_3}(t-3)$; -0,007284$S_{ND}(t-4)$; 0,006365$Q_{ar2}(t-2)$; -0,009826$X_{ND}(t-4)$; -0,013403$Q_{in}(t-3)$; 0,014370$X_{ND}(t-2)$; -0,011991$X_{ND}(t-3)$; -0,008891$Q_{in}(t-4)$; 0,001613$S_S(t-4)$; 0,006808$Q_{in}(t-1)S_{NH}(t-1)$; 0,005125$Q_{in}(t-1)S_{ND}(t-1)$; 0,001004$Q_{sl}(t-3)$; 0,031634$S_{NH_3}(t-1)S_{NH_3}(t-4)$; 0,005616$Q_{int}(t-3)$; -0,004085$S_{NH}(t-4)$; -0,003443$Q_{in}(t-3)S_{NH}(t-3)$; 0,007558$X_{ND}(t-1)X_{ND}(t-3)$; -0,003602$S_{ND}(t-4)$; 0,003217$Q_{ar2}(t-3)$; -0,005299$Q_{in}(t-4)X_{ND}(t-4)$; -0,003824$Q_{in}(t-3)S_{ND}(t-3)$; -0,006242$Q_{sl}(t-2)Q_{ar2}(t-1)$; 0,004341$Q_{sl}(t-1)Q_{ar2}(t-1)$; 0,001687$Q_{sl}(t-4)$; 0,006405$Q_{in}(t-2)Q_{in}(t-4)$; -0,003872$Q_{in}(t-1)Q_{in}(t-4)$; -0,005118$X_{ND}(t-2)X_{ND}(t-3)$; 0,002606$Q_{in}(t-4)S_{ND}(t-1)$; -0,005212$Q_{sl}(t-1)Q_{ar3}(t-1)$; 0,003315$Q_{sl}(t-2)Q_{ar3}(t-2)$; -0,010620$S_{NH_3}(t-1)Q_{sl}(t-2)$; 0,003796$Q_{in}(t-2)$; -0,003441$X_S(t-2)$; 0,003102$Q_{in}(t-4)X_{ND}(t-1)$; 0,003598$X_S(t-4)$; 0,002911$S_S(t-3)$; -0,003726$Q_{in}(t-1)S_S(t-1)$; 0,002672$Q_{in}(t-4)S_S(t-4)$; -0,002031$S_{NH}(t-1)S_{NO}(t-1)$; 0,027392$S_{NH_3}(t-2)Q_{in}(t-1)$; -0,001772$Q_{in}(t-2)X_{IP}(t-3)$; -0,009363$S_{NH_3}(t-1)Q_{ar3}(t-1)$; 0,006557$S_{NH_3}(t-3)Q_{ar3}(t-2)$

9.2. Identificação com Modelos NARX Polinomiais - Com Seleção de Variáveis de Entrada (CSV E)

Foram apresentados 3 critérios para a seleção das variáveis candidatas a entrada, sendo:

- 1) *Efetividade de Entrada e Projeção (EEP)*;
- 2) *Análise de Componentes Principais (PCA)*;
- 3) *Funções de Correlação (FC)*.

1) Identificação Utilizando-se as Variáveis de Entrada Seleccionadas pelo Critério de Efetividade de Entrada e Projeção (EEP):

Para a concentração de amônia (S_{NH_3}), as variáveis de entrada seleccionadas foram: Q_{in} , Q_w , Q_{ar2} e Q_{ar3} . A tabela 15 mostra o modelo obtido para a concentração de amônia (S_{NH_3}). O modelo NARX polinomial foi obtido utilizando-se grau de não-linearidade ($l = 2$), $n_y = 4$, $n_u = 4$. De um total de 231 termos do modelo ($n_\theta = 231$), com a utilização do *ERR* foram seleccionados 22 termos ($n_{\theta_s} = 22$).

Tabela 15: Modelo NARX polinomial para a saída S_{NH_3} - Modelo com seleção de variáveis de entrada pelo critério de efetividade de entrada e projeção

$$\begin{aligned} &-0,000874; 0,035743Q_{in}(t-1); 0,000136Q_w(t-1); -0,009821Q_{ar2}(t-1); \\ &-0,008892Q_{ar3}(t-1); 0,005866S_{NH_3}(t-1)^2; 0,040034S_{NH_3}(t-1)Q_{in}(t-1); \\ &0,000838S_{NH_3}(t-1)Q_w(t-1); -0,002806S_{NH_3}(t-1)Q_{ar2}(t-1); 1,934544S_{NH_3}(t-1); \\ &-1,180194S_{NH_3}(t-2); -0,010342Q_{in}(t-3); 0,213353S_{NH_3}(t-3); -0,013358Q_{in}(t-2); \\ &0,016661Q_{ar3}(t-2); 0,008351Q_{ar2}(t-2); -0,007874Q_{ar3}(t-3); \\ &-0,024873S_{NH_3}(t-3)Q_{in}(t-3); -0,014398Q_{in}(t-1)Q_{in}(t-3); \\ &0,010347Q_{in}(t-2)Q_{in}(t-3); -0,021108S_{NH_3}(t-1)Q_w(t-3); \\ &0,018194S_{NH_3}(t-1)Q_w(t-2) \end{aligned}$$

2) Identificação Utilizando-se as Variáveis de Entrada Seleccionadas pela Análise de Componentes Principais (PCA):

As variáveis de entrada seleccionadas para a concentração de amônia (S_{NH_3}) foram: X_S , S_S , Q_{in} , Q_{sl} , X_{ND} , S_{NH} e Q_{ar2} .

A tabela 16 mostra o modelo NARX obtido para a concentração de amônia (S_{NH_3}), o qual foi gerado utilizando-se grau de não-linearidade ($l = 2$), $n_y = 4$, $n_u = 4$. De um total de 561 termos do modelo ($n_\theta = 561$), com a utilização do *ERR* foram seleccionados 41 termos ($n_{\theta_s} = 41$).

Tabela 16: Modelo NARX polinomial para a saída S_{NH_3} - Modelo com seleção de variáveis de entrada pela análise de componentes principais

$$\begin{aligned} &-0,001347; 0,038970Q_{in}(t-1); 0,001999Q_{sl}(t-1); -0,010831Q_{ar2}(t-1); \\ &-0,010195X_S(t-1); 0,018620X_{ND}(t-1); 0,020360S_{NH}(t-1); -0,005244S_S(t-1); \\ &0,255907S_{NH_3}(t-1)^2; -0,008450S_{NH_3}(t-1)Q_{in}(t-1); -0,005592S_{NH_3}(t-1)Q_{sl}(t-1); \\ &-0,002963S_{NH_3}(t-1)Q_{ar2}(t-1); 1,649766S_{NH_3}(t-1); -0,757609S_{NH_3}(t-2); \\ &0,062856S_{NH_3}(t-3); -0,009661S_{NH}(t-1); -0,015323X_{ND}(t-3); -0,010861Q_{in}(t-3); \\ &0,006976X_S(t-3); 0,010786X_{ND}(t-2); -0,392254S_{NH_3}(t-1)S_{NH_3}(t-2); \\ &0,007659Q_{ar2}(t-2); -0,009638Q_{sl}(t-2); 0,005347Q_{in}(t-2)X_{ND}(t-1); \\ &0,003852S_S(t-3); 0,006655Q_{in}(t-1)S_{NH}(t-1); -0,008493Q_{in}(t-1)S_{NH}(t-2); \\ &0,004037Q_{sl}(t-3); 0,146167S_{NH_3}(t-1)S_{NH_3}(t-3); -0,004818Q_{in}(t-2); \\ &-0,004224S_{NH}(t-2); -0,003863X_{ND}(t-4); 0,005108Q_{in}(t-2)S_{NH}(t-1); \\ &-0,008859X_{ND}(t-2)S_S(t-2); 0,006871X_{ND}(t-1)S_S(t-1); \\ &0,017375S_{NH_3}(t-1)Q_{in}(t-2); 0,007973X_{ND}(t-1)X_{ND}(t-3); \\ &-0,006182X_{ND}(t-2)X_{ND}(t-3); -0,005344Q_{sl}(t-2)Q_{ar2}(t-1); \\ &0,004064Q_{sl}(t-3)Q_{ar2}(t-2); -0,002809Q_{in}(t-4) \end{aligned}$$

3) Identificação Utilizando-se as Variáveis de Entrada Seleccionadas pelo Método de Funções de Correlação (FC):

Para a concentração de amônia (S_{NH_3}), as variáveis de entrada seleccionadas foram: Q_{in} , X_{ND} , S_{NH} , S_{ND} e S_S . A tabela 17 mostra o modelo NARX polinomial obtido utilizando-se grau de não-linearidade ($l = 2$), $n_y = 4$, $n_u = 4$. De um total de 325 termos do modelo ($n_\theta = 325$), com a utilização do *ERR* foram seleccionados 35 termos ($n_{\theta_s} = 35$).

Tabela 17: Modelo NARX polinomial para a saída S_{NH_3} - Modelo com seleção de variáveis de entrada pelo método de funções de correlação

-0,001312; 0,040395 $Q_{in}(t-1)$; 0,017574 $X_{ND}(t-1)$; 0,020162 $S_{NH}(t-1)$;
 0,014215 $S_{ND}(t-1)$; -0,010368 $S_S(t-1)$; 0,139523 $S_{NH_3}(t-1)^2$;
 -0,011423 $S_{NH_3}(t-1)Q_{in}(t-1)$; 0,003134 $S_{NH_3}(t-1)X_{ND}(t-1)$;
 0,008731 $S_{NH_3}(t-1)S_{NH}(t-1)$; 1,644849 $S_{NH_3}(t-1)$; -0,796853 $S_{NH_3}(t-2)$;
 0,141134 $S_{NH_3}(t-3)$; -0,009888 $S_{NH}(t-3)$; 0,012553 $Q_{in}(t-2)X_{ND}(t-2)$;
 -0,010215 $Q_{in}(t-3)$; -0,013977 $X_{ND}(t-3)$; -0,008364 $S_{ND}(t-3)$; 0,011484 $X_{ND}(t-2)$;
 0,006322 $S_S(t-3)$; -0,128281 $S_{NH_3}(t-1)S_{NH_3}(t-2)$; -0,007940 $Q_{in}(t-3)X_{ND}(t-3)$;
 0,009111 $Q_{in}(t-1)S_{ND}(t-1)$; -0,009793 $S_{NH_3}(t-2)S_{NH_3}(t-3)$;
 -0,006627 $Q_{in}(t-2)S_{ND}(t-2)$; -0,004170 $Q_{in}(t-2)$; 0,014845 $S_{NH_3}(t-1)Q_{in}(t-2)$;
 -0,002489 $Q_{in}(t-1)S_S(t-2)$; -0,004011 $X_{ND}(t-4)$; -0,004461 $Q_{in}(t-4)$;
 -0,037578 $S_{NH_3}(t-4)$; 0,002371 $Q_{in}(t-1)S_{NH}(t-1)$; -0,007533 $X_{ND}(t-1)S_{ND}(t-1)$;
 0,006197 $X_{ND}(t-2)S_{ND}(t-2)$; -0,003366 $S_{NH}(t-2)$

10. Critérios de Desempenho do Modelo

Nesta seção faz-se uma comparação dos modelos identificados *com e sem seleção de variáveis de entrada*.

Os indicadores aqui apresentados são usados para avaliar os modelos identificados como também os modelos validados, conforme será apresentado no capítulo seguinte.

- **Cálculo da Variância:**

A variância é dada por:

$$C(\hat{\Theta}_s) = \frac{1}{N} \sum_{t=1}^N \varepsilon^T(t) \varepsilon(t) \quad (29)$$

sendo que N é o número de dados disponíveis, ε é o erro de modelamento.

- **RMSE (Relative Mean Square Error):**

O RMSE é dado por:

$$RMSE = \frac{\sqrt{\sum_{k=1}^N (y(k) - \hat{y}(k))^2}}{\sqrt{\sum_{k=1}^N (y(k) - \bar{y})^2}} \quad (30)$$

sendo que $\hat{y}(k)$ é o valor estimado do sinal e $\bar{y}$ é o valor médio do sinal medido $y(k)$, sendo que a média é calculada na janela de identificação (Aguirre, 2000).

10.1. Modelos NARX Polinomiais

Uma análise dos modelos **NARX** polinomiais para a variável de saída S_{NH_3} é apresentada a seguir.

A tabela 18 apresenta os resultados comparativos dos modelos **NARX** polinomiais *com e sem seleção de variáveis de entrada* para a concentração de amônia (S_{NH_3}).

Tabela 18: Comparação dos modelos NARX polinomiais para a saída S_{NH_3}

$l = 2; n_y = 4;$ $n_y = 4$	SSVE	EEP	PCA	FC
n_θ	2145	231	561	325
n_{θ_s}	73	22	41	35
$C(\hat{\Theta}_s)$	$1,394114 \times 10^{-4}$	$3,811021 \times 10^{-4}$	$2,431215 \times 10^{-4}$	$2,438434 \times 10^{-4}$
RMSE	0,0475	0,0785	0,0627	0,0628

Observa-se, conforme era de se esperar, que o melhor modelo **NARX** foi obtido para o caso em que todas as variáveis de entrada foram empregadas. Os métodos *PCA* e *FC*

apresentaram desempenho bastante similar. Embora, os resultados dos índices de desempenho sejam próximos, observa-se que os métodos *PCA* e *FC* foram melhores do que o método de efetividade de entrada e projeção.

No entanto, o modelo identificado com seleção de variáveis de entrada através da análise de componentes principais (*PCA*) apresenta um número de termos maior que o modelo obtido aplicando-se as funções de correlação. Desta forma, pode-se concluir que, dentre os modelos **NARX** polinomiais identificados, o modelo com seleção de variáveis de entrada através do método de funções de correlação é uma boa opção para representar a concentração de amônia (S_{NH_3}).

11. Validação dos Modelos NARX Polinomiais

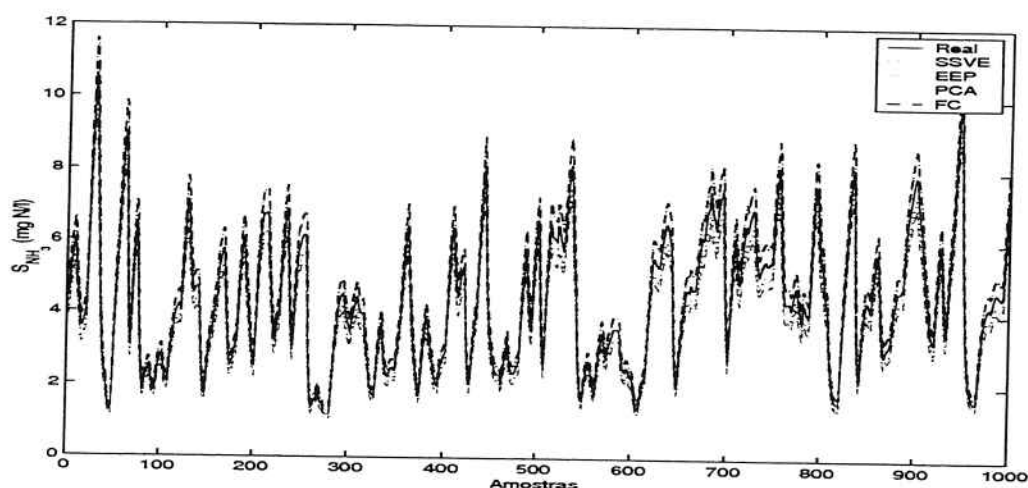
Para o procedimento de validação foi coletado um novo conjunto de dados, correspondendo 2001 amostras. O processo foi simulado por 3000 horas, mantendo-se o período de amostragem utilizado no procedimento de identificação, o qual foi de 1,5 horas. Os sinais foram aleatórios com distribuição uniforme. Aqui se apresentam os resultados de validação dos modelos **NARX** polinomiais, utilizando-se o modelo sem seleção de variáveis de entrada e modelos com seleção de variáveis de entrada, pelo critério de *efetividade de entrada e projeção*, pela *análise de componentes principais* (*PCA*) e pelo método proposto neste trabalho, baseado em *funções de correlação*.

Uma análise dos modelos **NARX** polinomiais para a variável de saída S_{NH_3} é apresentada a seguir. A tabela 19 apresenta os resultados dos índices de desempenho relacionados à validação dos modelos **NARX** polinomiais *com e sem seleção de variáveis de entrada* para a concentração de amônia (S_{NH_3}).

Tabela 19: Comparação dos índices de desempenho dos modelos NARX polinomiais para a saída S_{NH_3}

$l = 2; n_y = 4;$ $n_u = 4$	<i>SSVE</i>	<i>EEP</i>	<i>PCA</i>	<i>FC</i>
n_θ	2145	231	561	325
n_{θ_s}	73	22	41	35
$C(\theta_s)$	0,1261	0,2947	0,1815	0,1823
<i>RMSE</i>	0,2005	0,3066	0,2406	0,2412

Analisando-se a tabela 19 verifica-se que os modelos *PCA* e *FC* apresentaram índices de desempenho melhores do que o modelo *EEP*. Os resultados para o modelos *PCA* e *FC* são similares e apenas pouco inferiores ao modelo sem seleção de variáveis de entrada. A título de facilitar a comparação, a figura 7 mostra a saída do processo e dos 4 modelos obtidos.



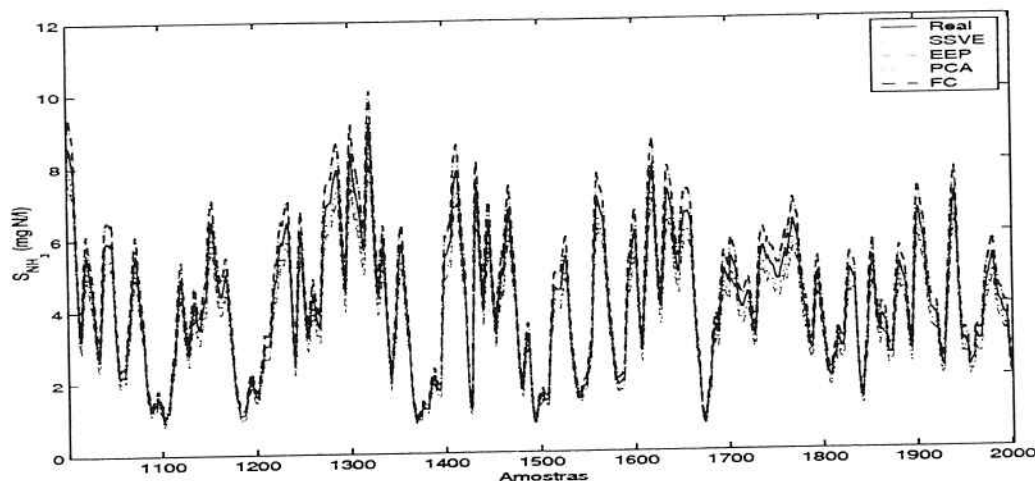


Figura 7: Validação dos modelos NARX polinomiais - Saída S_{NH_3}

De um modo geral pode-se observar que os modelos com *seleção de variáveis de entrada* tiveram um desempenho comparável ao do modelo *sem seleção de variáveis de entrada*.

12. Conclusões

O estudo de seleção de variáveis de entrada apresentado neste trabalho, teve sua aplicação em uma planta de tratamento de efluentes, um sistema altamente complexo, devido ao fato de incorporar um grande número de processos biológicos e físico-químicos. As ETE's (Estações de Tratamento de Esgotos) são consideradas grandes sistemas não-lineares sujeitas a grandes perturbações no fluxo e na carga, ao mesmo tempo com incertezas na composição do efluente. Portanto, a sua modelagem é complicada.

Foram apresentadas neste trabalho três técnicas para seleção de variáveis de entrada em identificação de sistemas, sendo duas delas já apresentadas na literatura, a saber, *efetividade de entrada e projeção* e *análise de componentes principais (PCA)*, sendo esta segunda, bastante divulgada. A terceira técnica, a qual emprega *funções de correlação*, é proposta neste trabalho. A idéia inicial de se trabalhar com funções de correlação partiu de (Aguirre, 2000), que menciona a seleção de variáveis de entrada usando funções de correlação, mas de uma forma pouca detalhada. Neste aspecto, este trabalho procurou estabelecer alguns requisitos para selecionar as variáveis de entrada, com base em funções de correlação. A proposta foi implementada e então aplicada a uma planta de tratamento de efluentes, a qual possui um conjunto de 15 variáveis candidatas a entrada.

Essas 15 variáveis foram usadas na análise para compor modelos MISO, para a concentração de amônia (S_{NH_3}).

A identificação dos modelos foi realizada usando-se uma técnica, convencional, através de modelos NARX polinomiais.

Os modelos com *seleção de variáveis de entrada* foi comparado com o modelo *sem seleção de variáveis de entrada*. Nesta análise, foi verificado que o modelo com seleção de variáveis de entrada pelo critério de *efetividade de entrada e projeção (EEP)* foi o que apresentou pior resultado quando comparado com os modelos com variáveis de entrada selecionadas pela *análise de componentes principais (PCA)* e *funções de correlação (FC)*. Para os modelos com seleção de variáveis de entrada através de *análise de componentes principais (PCA)* e *funções de correlação (FC)*, os resultados de validação foram semelhantes. Por outro lado, o modelo com seleção de variáveis de entrada selecionadas pela *análise de componentes principais (PCA)* apresentou um número maior de variáveis de entrada selecionadas em relação ao modelo com seleção de variáveis de entrada através de *funções de correlação*. Desta forma, pôde-se verificar que o modelo obtido através de variáveis selecionadas pelo método de *funções de correlação* incorporou menos termos em relação ao modelo obtido com a seleção de variáveis de entrada pela *análise de componentes principais (PCA)*. Portanto, pode-se concluir que a ferramenta proposta de seleção de variáveis de entrada baseada em *funções de correlação* foi considerada satisfatória na identificação de um processo altamente complexo e não linear. A técnica proposta é relativamente simples de ser usada se comparada com outras existentes na literatura, o que

facilita seu emprego por eventuais usuários.

Referências

- Aguirre, L. A. (2000). *Introdução à Identificação de Sistemas*. Belo Horizonte: Editora UFMG.
- Billings, S. A. and S. Chen (1989). Extended Model Set, Global and Threshold Model Identification of Severely Non-Linear Systems. *International Journal of Control* **50**(5), 1897–1923.
- Billings, S. A., M. J. Korenberg and S. Chen (1988). Identification of Non-Linear Output-Affine Systems Using an Orthogonal Least-Squares Algorithm. *International Journal of Systems Science* **49**, 1559–1568.
- Billings, S. A., S. Chen and M. J. Korenberg (1989). Identification of MIMO Non-Linear Systems Using a Forward-Regression Orthogonal Estimator. *International Journal of Control* **49**(6), 2157–2189.
- Cao, Y. and D. Biss (1996). New Screening Techniques for Choosing Manipulated Variables. in *Proceedings of IFAC'96, 13th World Congress M*, 103–108.
- Cao, Y., D. Rossiter and D. Owens (1997). Screening criteria for input and output selection. http://www.cds.caltech.edu/conferences/related/ECC97/proceeds/751_1000/ECC773.PDF.
- Chen, S. and S. A. Billings (1989). Representation of Nonlinear System: The NARMAX Model. *International Journal of Control* **49**, 1013–1032.
- Chen, S., S. A. Billings and W. Luo (1989). Orthogonal Least Squares Methods and Their Applications to Non-Linear System Identification. *International Journal of Control* **50**(5), 1873–1896.
- Dunteman, G. H. (1989). *Principal Components Analysis*. Sage Publications, Inc.
- Jerônimo, R. A. (1998). Identificação Não-Linear de um Processo de Neutralização de pH Multivariável Utilizando Modelos Polinomiais. Dissertação de Mestrado. Escola Politécnica da Universidade de São Paulo, São Paulo, SP, Brasil.
- Korenberg, M. J., S. A. Billings, Y. P. Liu and P. J. McIlroy (1988). Orthogonal Parameter Estimation Algorithm for Non-Linear Stochastic Systems. *International Journal of Control* **48**(1), 193–210.
- Leontaritis, I. J. and S. A. Billings (1985). Input-Output Parametric Models for Nonlinear System. *International Journal of Control* **41**, 303–344.
- Pröll, T. (1993). Model Predictive Control Based on Nonlinear Autoregressive and Neural Network Models. Ph.D. Thesis. Colorado State University.
- Pröll, T. and M. N. Karim (1994). Model Predictive pH Control Using Real-Time NARX Approach. *AIChE Journal*.
- Sotomayor, O. A. Z. (2002). Modelagem, Identificação e Controle de Sistemas de Tratamento de Lodo Ativado com Remoção de Nitrogênio. Tese de Doutorado. Escola Politécnica da Universidade de São Paulo, São Paulo, SP, Brasil.
- Sotomayor, O. A. Z., S. W. Park and Claudio Garcia (2001). A Simulation Benchmark to Evaluate the Performance of Advanced Control Techniques in Biological Wastewater Treatment Plants. *Brazilian Journal of Chemical Engineering* **18**(1), 81–101.
- Sotomayor, O. A. Z., S. W. Park and Garcia C. (1999). Um Modelo de Referência para o Controle do Processo de Lodo Ativado. Boletim Técnico da Escola Politécnica da USP, Departamento de Engenharia Eletrônica, BT/PEE/9918, São Paulo, SP, Brasil.

BOLETINS TÉCNICOS - TEXTOS PUBLICADOS

- BT/PTC/9901 – Avaliação de Ergoespirômetros Segundo a Norma NBR IEC 601-1- MARIA RUTH C. R. LEITE, JOSÉ CARLOS TEIXEIRA DE B. MORAES
- BT/PTC/9902 – Sistemas de Criptofonia de Voz com Mapas Caóticos e Redes Neurais Artificiais – MIGUEL ANTONIO FERNANDES SOLER, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/9903 – Regulação Sincronizada de Distúrbios Senodais – VAIDYA INÉS CARRILLO SEGURA, PAULO SÉRGIO PEREIRA DA SILVA
- BT/PTC/9904 – Desenvolvimento e Implementação de Algoritmo Computacional para Garantir um Determinado Nível de Letalidade Acumulada para Microorganismos Presentes em Alimentos Industrializados – RUBENS GEDRAITE, CLÁUDIO GARCIA
- BT/PTC/9905 – Modelo Operacional de Gestão de Qualidade em Laboratórios de Ensaio e Calibração de Equipamentos Eletromédicos – MANUEL ANTONIO TAPIA LÓPEZ, JOSÉ CARLOS TEIXEIRA DE BARROS MORAES
- BT/PTC/9906 – Extração de Componentes Principais de Sinais Cerebrais Usando Karhunen – Loève Neural Network – EDUARDO AKIRA KINTO, EUVALDO F. CABRAL JR.
- BT/PTC/9907 – Observador Pseudo-Derivativo de Kalman Numa Coluna de Destilação Binária – JOSÉ HERNANDEZ LÓPEZ, JOSÉ JAIME DA CRUZ, CLAUDIO GARCIA
- BT/PTC/9908 – Reconhecimento Automático do Locutor com Coeficientes Mel-Cepstrais e Redes Neurais Artificiais – ANDRÉ BORDIN MAGNI, EUVALDO F. CABRAL JÚNIOR
- BT/PTC/9909 – Análise de Estabilidade e Síntese de Sistemas Híbridos – DIEGO COLÓN, FELIPE MIGUEL PAIT
- BT/PTC/0001 – Alguns Aspectos de Visão Multiescalas e Multiresolução – JOÃO E. KOGLER JR., MARCIO RILLO
- BT/PTC/0002 – Placa de Sinalização E1: Sinalização de Linha R2 Digital Sinalização entre Registradores MFC- PHILLIP MARK SEYMOUR BURT, FERNANDA CARDOSO DA SILVA
- BT/PTC/0003 – Estudo da Técnica de Comunicação FO-CDMA em Redes de Fibra Óptica de Alta Velocidade – TULIPA PERSO, JOSÉ ROBERTO DE A. AMAZONAS
- BT/PTC/0004 – Avaliação de Modelos Matemáticos para Motoneurônios – DANIEL GUSTAVO GOROSO, ANDRÉ FÁBIO KOHN
- BT/PTC/0005 – Extração e Avaliação de Atributos do Eletrocardiograma para Classificação de Batimentos Cardíacos – ELDER VIEIRA COSTA, JOSÉ CARLOS T. DE BARROS MORAES
- BT/PTC/0006 – Uma Técnica de Imposição de Zeros para Auxílio em Projeto de Sistemas de Controle – PAULO SÉRGIO PIERRI, ROBERTO MOURA SALES
- BT/PTC/0007 – A Connected Multireticulated Diagram Viewer – PAULO EDUARDO PILON, EUVALDO F. CABRAL JÚNIOR
- BT/PTC/0008 – Some Geometric Properties of the Dynamic Extension Algorithm – PAULO SÉRGIO PEREIRA DA SILVA
- BT/PTC/0009 – Comparison of Alternatives for Capacity Increase in Multiple-Rate Dual-Class DS/CDMA Systems – CYRO SACARANO HESI, PAUL ETIENNE JESZESKY
- BT/PTC/0010 – Reconhecimento Automático de Ações Faciais usando FACS e Redes Neurais Artificiais – ALEXANDRE TORNICE, EUVALDO F. CABRAL JÚNIOR
- BT/PTC/0011 – Estudo de Caso: Tornando um Projeto Testável Utilizando Ferramentas Synopsys – REINALDO SILVEIRA, JOSÉ ROBERTO A. AMAZONAS
- BT/PTC/0012 – Modelos Probabilísticos para Rastreamento em Carteiras de Investimento – HUGO G. V. DE ASSUNÇÃO, OSWALDO L. V. COSTA
- BT/PTC/0013 – Influência de um Controle Imperfeito de Potência e Monitoramento da Atividade Vocal na Capacidade de Sistemas DS/CDMA – MÁRCIO WAGNER DUARTE ROLIM, PAUL JEAN ETIENNE JESZESKY
- BT/PTC/0014 – Canceladores de Interferência Sucessivo e Paralelo para DS/CDMA – TAUFIK ABRÃO, PAUL JEAN E. JESZESKY
- BT/PTC/0015 – Transmissão de Serviços de Multimídia num Sistema Móvel Celular CDMA de Banda Larga – EDUARDO MEIRELLES MASSAUD, PAUL JEAN ETIENNE JESZESKY
- BT/PTC/0016 – Disseminação do HIV em uma População Homossexual Heterogênea – MARCOS CASADO CASTÑO, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PTC/0017 – Implementação e Avaliação em Laboratório de um Monitor Cardíaco Portátil para Três Derivações – RAISA FERNÁNDEZ NUNEZ, JOSE CARLOS TEIXEIRA DE BAROS MORAES
- BT/PTC/0018 – Projeto de Filtros Recursivos de N-ésima Banda – IRINEU ANTUNES JÚNIOR, MAX GERKEN
- BT/PTC/0019 – Relative Flatness and Flatness of Implicit Systems – PAULO SÉRGIO PEREIRA DA SILVA, CARLOS CORRÊA FILHO
- BT/PTC/0020 – Estimativa de Fluxo Sangüíneo nas Artérias Coronárias Usando Imagens de Cineangiocardiografia – ANA CRISTINA DOS SANTOS, SÉRGIO SHIGUEMI FURUIE
- BT/PTC/0021 – Modelos Populacionais para AIDS e Análise do Equilíbrio sem Epidemia – ELIZABETH FERREIRA SANTOS, JOSÉ ROBERTO CASTILHO PIQUEIRA

- BT/PTC/0101 – Model-Based Soft-Sensor Design for On-Line Estimation of the Biological Activity in Activated Sludge Wastewater Treatment Plants – OSCAR A. Z. SOTOMAYOR, SONG WON PARK, CLAUDIO GARCIA
- BT/PTC/0102 – Reconhecimento Automático do Locutor Utilizando a Rede Neural Artificial *Field Distributed Memory* – FDM – MARCELO BLANCO, EUVALDO F. CABRAL JR.
- BT/PTC/0103 – Algoritmos de Filtragem e Previsão em Modelos de Volatilidade – FERNANDO LOVISOTTO, OSWALDO L. V. COSTA
- BT/PTC/0104 – Método de Diferenças Temporais Aplicado às Equações de Riccati Acopladas entre Si – OSWALDO L. V. COSTA, JULIO C. C. AYA
- BT/PTC/0105 – Método de Diferenças Finitas e de Monte Carlo em Derivativos – ANDRÉ CURY MAIALI, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0106 – Resolução de um Problema Inverso de Eletromagnetismo por Meio de Redes Neurais Artificiais – ARNALDO MEGRICH, JORGE MIECZYSLAW JANISZEWSKI
- BT/PTC/0107 – Projeto de Controlador de Temperatura para Perfusão Peritoneal com Hipertermia e Quimioterapia – GIANCARLO ANTONIO BERZACOLA, FUAD KASSAB JÚNIOR
- BT/PTC/0108 – O Papel de Diferentes Grupos Populacionais na Transmissão Sexual do HIV – ELIZABETH FERREIRA SANTOS, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PTC/0109 – Terapias Ótimas Anti-HIV para a Redução da Transmissão Vertical – RENATO BEVILACQUA, LUIZ HENRIQUE ALVES MONTEIRO
- BT/PTC/0110 – Brain Signal Analysis Using Non-Linear ARIMA Models – ERNANE J. X. COSTA, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/0111 – Cancelamento de Eco Acústico Estéreo: Análise de Algoritmos Adaptativos e um novo Método de Redução do Desalinhamento – ROBERTO ROSCHEL BELLI, PHILLIP MARK SEYMOUR BURT
- BT/PTC/0112 – Natural Gas Flow Computer With Open Architecture Using Intelligent Instrumentation And Field Bus – OSMEL REYES VAILLANT, CLAUDIO GARCIA
- BT/PTC/0113 – Aplicação de Métodos de Inteligência Artificial em Inteligência de Negócios – ROGÉRIO GARCIA DUTRA, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/0114 – Detectores Multiusuário para DS/CDMA – Canceladores de Interferência – TAUFIK ABRÃO, PAUL JEAN E. JESZENSKY
- BT/PTC/0115 – Reconhecimento Automático do Locutor Usando Pré-Processamento em Sons Nasalizados com Diversos Classificadores Neurais – ROBERTO AMILTON BERNARDES SÓRIA, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/0116 – Aplicando a Técnica de Times Assíncronos na Otimização de Problemas de Empacotamento Unidimensional – REINALDO DE BERNARDI, TSEN CHUNG KANG
- BT/PTC/0117 – A Simplified Implementation of the Theory of Emotions for Emotots – RODRIGO DE TOLEDO CAROPRESO, RODRIGO DIAS MORGADO, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/0118 – Modelos Dinâmicos Aplicados à Propagação de Vírus em Redes de Computadores – BETYNA FERNÁNDEZ NAVARRO, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PTC/0119 – Técnicas para Melhorar o Desempenho do Método dos Análogos Modificado em Predição de Séries Temporais Caóticas – LUIZ MONZÓN BENITEZ, ADEMAR FERREIRA
- BT/PTC/0120 – Um Modelo Microscópico para a Simulação de uma Intersecção Semaforizada – EDUARDO ULI NELLAR, ADEMAR FERREIRA
- BT/PTC/0121 – Subspace Identification Methods Applied to Activated Sludge Processes: A Performance Comparison Study – OSCAR A. Z. SOTOMAYOR, SONG WON PARK, CLAUDIO GARCIA
- BT/PTC/0122 – Controle Avançado de Temperatura para Perfusão Peritoneal com Hipertermia e Quimioterapia – ARNALDO LAGANARO JUNIOR, FUAD KASSAB JUNIOR
- BT/PTC/0123 – Análise de Sinais Biológicos Utilizando *Wavelets* – FRANCO BELTRAME RUNZA, FUAD KASSAB JUNIOR
- BT/PTC/0123 – Certificação de Estimuladores Neuromusculares no Sistema Brasileiro de Certificação de Equipamentos Médicos – EDUARDO RODRIGUES MANDL, JOSÉ CARLOS TEIXEIRA DE BARROS
- BT/PTC/0201 – Aplicação de Transformadas Tempo – Frequência e de Redes Neurais Artificiais no Processamento Digital de Sinais Cerebrais – MARCOS TEIXEIRA DE FREITAS FILHO, EUVALDO FERREIRA CABRAL JR.
- BT/PTC/0202 – Aspectos de Identificação para Controle de Sistemas NARX usando Redes Neurais RBF – RITA DE CÁSSIA PAVANI LAMAS, ADEMAR FERREIRA
- BT/PTC/0203 – A Equação de Riccati Estacionária na Estimação Linear em Sistemas Lineares Discretos no Tempo com Saltos Markovianos – SUSSET GUERRA JIMÉNEZ, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0204 – Comparação entre Algoritmos de Identificação Não-Assistida de Canais de Comunicação Baseados em Estatísticas de Segunda Ordem – CLAUDIO JOSE BORDIN JUNIOR, LUIZ ANTONIO BACCALA
- BT/PTC/0205 – Desenvolvimento de um Simulador de uma Sub-Rede da Medula Espinhal – ROGÉRIO RODRIGUES LIMA CISI, ANDRÉ FABIO KOHN
- BT/PTC/0206 – Novos Modelos de Motoneurônios de Vertebrados – MARCOS FRAGA VIEIRA, ANDRÉ FABIO KOHN

- BT/PTC/0207 – Modelagem Auto-Regressiva e Memórias Associativas na Classificação de Sinais Eletromiográficos de Agulhas – IARA KAZUYO IKE, CINTHIA ITIKI
- BT/PTC/0208 – Análise da Classificação e da Síntese de Expressões Faciais com Redes Neurais Artificiais – OLIMPO MURILO CAPELI, EUVALDO FERREIRA CABRAL JUNIOR
- BT/PTC/0209 – Guiagem Automática de um Veículo Autônomo Baseada em Sistema de Visão – NICOLÁS CÉSAR LAVINIA, JOSÉ JAIME DA CRUZ
- BT/PTC/0210 – Some Geometric Properties of Differential-Algebraic Equations – PAULO SÉRGIO PEREIRA DA SILVA, CARLOS JUTTI WATANABE
- BT/PTC/0211 – Modelos de Média-Variância de Período Simples e Multi-Períodos na Análise de Carteiras de Investimento – MARGARETH AP. DE SOUZA BUENO, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0212 – Avaliação da Utilização de Centrais Públicas de Informações de Crédito num Modelo de Previsão para Risco de Crédito – GUILHERME GONZALEZ C. PARENTE, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0213 – Propagação de Erros de Frequência em Redes Mestre Escravo em Estrela Dupla – SANTOS ANDRÉS CASTILHO VARGAS, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PTC/0214 – Avaliação Prática de um Algoritmo de Alta Resolução na Determinação de Frequências de Ressonância em Estruturas de Grande Porte – LUIZ ANTONIO BARBOSA COELHO, LUIZ ANTONIO BACCALÁ
- BT/PTC/0301 – Finite Horizon Quadratic Optimal Control Problem of Markovian Jump Linear Systems With Partial Information – ESTEBAN FERNANDEZ TUESTA, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0302 – A Framework for Admission Control Based on Aggregate Traffic Measurements and Network Characterization – ALEXANDRE BARBOSA DE LIMA, JOSE ROBERTO DE ALMEIDA AMAZONAS
- BT/PTC/0303 – Identificação Nebulosa da Concentração de Oxigênio Dissolvido do Tanque de Aeração de uma Estação de Tratamento de Esgotos por Lodos Ativados – ALEXANDRE MAGNO PARENTE DA ROCHA, CLAUDIO GARCIA
- BT/PTC/0304 – Cálculo de Balanço de Potência em Sistemas DWDM com Spans Variáveis e Amplificadores com Curva de Ganho não Plana – WALMARA DE PAULA HERMAN, JOSÉ ROBERTO DE ALMEIDA AMAZONAS
- BT/PTC/0305 – Robust Model Tracking and 2-D Control Design – FABRIZIO LEONARDI, JOSÉ JAIME DA CRUZ
- BT/PTC/0306 – Classificação de Ações do Mercado Financeiro Utilizando Técnicas Fuzzy –RENATO APARECIDO AGUIAR, ROBERTO MOURA SALES
- BT/PTC/0307 – Esquemas de Multitaxa para Sistemas CDMA de Banda Larga – ANDRÉ FAGUNDES DA ROCHA, PAUL JEAN ETIENNE JESZENSKY
- BT/PTC/0308 – A Time-Invariant Flat System Always Admits a Time-Invariant Flat Output – PAULO SERGIO PEREIRA DA SILVA, PIERRE ROUCHON
- BT/PTC/0309 – Otimização Robusta de Carteiras Utilizando Desigualdades Matriciais Lineares – RODRIGO BARROS NABHOLZ, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0310 – Desenvolvimento de um Controlador Automático para Sistemas de Guiagem de Tiras de Aço em Linhas de Processos Siderúrgicos – AMAURI DIAS DE CARVALHO, JOSÉ JAIME DA CRUZ
- BT/PTC/0311 – Sincronismo em Redes Mestre-Escravo de Via Única: Estrela Simples, Cadeia Simples e Mista – CARLOS NEHEMY MARMO, JOSÉ ROBERTO CASTILHO PIQUEIRA
- BT/PTC/0312 – Aprendizado por Reforço Aplicado a Sistemas Neuro-Difusos para o Controle de Sinais de Trânsito - LEONCIO ZÁRATE GAMARRA, ADEMAR FERREIRA
- BT/PTC/0313 – Métodos de Decomposição em Problemas de Estoque e Roteirização – PATRÍCIA PRADO BELFIORE, OSWALDO LUIZ DO VALLE COSTA
- BT/PTC/0314 – Reconhecimento do Sinal Mioelétrico Aplicado à Engenharia de Reabilitação – LEONARDO PINHEIRO DA SILVA, FUAD KASSAB JR.
- BT/PTC/0401 – Comparação Entre Receptores Espaço-Temporais Semicegos de Formação de Feixe e Diversidade para SistemaS W-CDMA – IVAN ROBERTO SANTANA CASELLA, PAUL JEAN ETIENNE JESZENSKY
- BT/PTC/0402 – Mapas Acoplado e Aplicações: Processamento de Imagens, Auto-Organização e Processamento Simbólico – ROGÉRIO DE OLIVEIRA, LUIZ HENRIQUE ALVES MONTEIRO
- BT/PTC/0403 – Modelagem de FCC Usando Métodos de Identificação por Predição de Erro e por Sub-Espaços – SANTOS DEMETRIO MIRANDA BORJAS, CLAUDIO GARCIA
- BT/PTC/0404 – Implementação em Hardware / Firmware de um Sensor Virtual Utilizando Algoritmos de Identificação Fuzzy – CASSIO DE CARVALHO BERNI, CLAUDIO GARCIA

