

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

**Nova Metodologia para Desenvolvimento
de Inferências Baseadas em Dados**

DISSERTAÇÃO DE MESTRADO

Thiago Dantas Fleck

Porto Alegre

2012

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

Nova metodologia para o desenvolvimento de inferências baseadas em dados

Thiago Dantas Fleck

Dissertação de Mestrado apresentada como
requisito parcial para obtenção do título de
Mestre em Engenharia

Área de concentração: Controle de Processos

Orientadores:

Prof. Dr. Jorge Otávio Trierweiler

Prof. Dr. Marcelo Farenzena

Porto Alegre

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

A Comissão Examinadora, abaixo assinada, aprova a Dissertação *Nova metodologia para o desenvolvimento de inferências baseadas em dados*, elaborada por Thiago Dantas Fleck, como requisito parcial para obtenção do Grau de Mestre em Engenharia.

Comissão Examinadora:

Prof. Dr. Michel José Anzanello – Eng. Produção / UFRGS

Dr. Ricardo Guilherme Duraiski – TriSolutions

Dr. Luciane Ferreira Trierweiler – DEQUI / UFRGS

*Toda a nossa ciência, comparada com a realidade,
é primitiva e infantil – e, no entanto, é a coisa mais
preciosa que temos.*

Albert Einstein (1879 – 1955)

*Nada é mais digno de nosso patrocínio que o fomen-
to da ciência e da literatura. O conhecimento é, em
todo e qualquer país, a base mais segura da felici-
dade pública.*

George Washington, discurso no Congresso, 8 de
Janeiro de 1790

Agradecimentos

Agradeço à UFRGS e ao Departamento de Engenharia Química pela oportunidade de realização do mestrado e a todos os professores que tive pelo conhecimento transmitido. Agradeço em especial aos meus orientadores Jorge e Farezena pela pelas ideias, discussões e auxílio.

Agradeço a toda equipe da TriSolutions pelos anos de convivência e por terem me incentivado e apoiado na realização desse trabalho. Aos engenheiros e operadores da Petrobras com quem convivi durante o trabalho na unidade industrial. Aos colegas de CENPRO que tornaram melhor os últimos meses de trabalho nessa dissertação e compartilharam a rotina de provas semanais e fins de semana de estudos no curso de formação da Petrobras.

Agradeço a meu pai, Eduardo, pelo exemplo e apoio durante toda a vida e a minha mãe, Cristina, por andar sempre comigo. Agradeço a minha irmã, Fernanda, pela alegria e companheirismo e às minhas avós, Maria Helena e Vera, por serem as mulheres fortes que são. Agradeço a todos os meus tios, tias, primos, primas e familiares que fazem minha vida mais feliz.

Agradeço a todos os amigos que fiz durante a graduação pelos bons momentos que passamos juntos. Ao RPGneto, amigos de décadas, por todas as discussões, polêmicas e encontros. Se tudo correr como esperado, nossa amizade vai durar pra sempre!

Agradeço à Camila pelo apoio e companheirismo em todas as horas. Obrigado por tornar minha vida melhor.

Resumo

As inferências têm diversas aplicações na indústria de processos químicos, sendo essenciais no sucesso de projetos de controle avançado. O desempenho do controle será sempre ligado ao desempenho da inferência, sendo importante a manutenção da sua qualidade ao longo do tempo. Neste trabalho, uma nova metodologia é sugerida para o desenvolvimento de inferências baseadas em dados seguindo uma abordagem segmentada com o objetivo de facilitar a sua manutenção. A nova proposta consiste em modelar a parte estacionária separada da parte dinâmica, diferentemente do que é feito na metodologia tradicional, onde o modelo dinâmico é gerado diretamente dos dados de processo. O modelo estacionário é obtido através de uma regressão PLS (Partial Least Squares), enquanto as dinâmicas são inseridas posteriormente utilizando-se um algoritmo de otimização. A técnica é aplicada a uma coluna de destilação e o resultado obtido é semelhante ao de inferências dinâmicas e estáticas desenvolvidas com métodos tradicionais. Outras etapas do desenvolvimento de inferências também são investigadas. Na seleção de variáveis, métodos estatísticos são comparados com a busca exaustiva e se conclui este último deve ser usado como padrão, visto que custo computacional não é mais um problema. Também são apresentadas boas práticas no pré-tratamento de dados, remoção do tempo morto do cromatógrafo modelado e detecção de estados estacionários.

Palavras-chave: Inferência, analisador virtual, modelos empíricos, PCA, PLS, processamento de gás natural

Abstract

Soft-sensors have several applications in the chemical processes industry and are essential for the success of advanced control projects. Its performance will always be linked to the performance of the soft-sensor, so it is important to maintain its quality over time. In this paper, a new methodology is suggested for the development of data-based soft-sensors following a segmented approach in order to facilitate its maintenance. The new proposal is to model the stationary part separated from the dynamic, unlike the traditional methodology where the dynamic model is generated directly from process data. The stationary model is obtained by a PLS (Partial Least Squares) regression, while the dynamics are inserted using an optimization algorithm. The technique is applied to a distillation column and its performance is similar to dynamic and static soft-sensors developed using traditional methods. Other steps in the development of soft-sensors are also investigated. In variable selection issue, statistical methods are compared with the testing of all possibilities; the latter should be used as default, since computational cost is no longer a problem. We also present best practices in data pre-processing, gas chromatograph dead-time removal and steady state detection.

Keywords: Soft-sensor, virtual analyser, data-based model, PCA, PLS, natural gas processing

Simbologia e nomenclatura

| | |
|---------------------|---|
| AIC | Akaike Information Criteria |
| ARX | Autoregressive with exogenous input |
| BE | Busca exaustiva |
| β | Vetor dos coeficientes de regressão |
| $\hat{\beta}$ | Vetor dos coeficientes de regressão estimados |
| c_i | Vetor de constantes do modelo quadrático |
| ε | Erro do modelo MLR |
| E | Matriz residual da redução de X |
| F | Matriz residual da redução de Y |
| i, j | Índices das linhas e colunas das matrizes |
| $I_{DER, i}$ | Índice de estacionaridade para a variável i |
| ICA | Independent Component Analysis |
| MLR | <i>Multivariable Linear Regression</i> |
| n | Número de linhas/amostra em uma matriz |
| p | Número de variáveis auxiliares |
| P | Matriz dos vetores de projeção de X |
| PCA | <i>Principal Components Analysis</i> |
| PLS | <i>Partial Least Squares</i> |
| PRESS | Predictive sum of squares |
| PVR | Pressão de Vapor Reid |
| Q^2 | Coeficiente de correlação na predição |
| Q | Matriz dos vetores de projeção de Y |
| R^2 | Coeficiente de correlação |
| SR | Stepwise regression |
| SS_{Erro} | Soma dos quadrados do erro |
| SS_{Total} | Soma dos quadrados total |
| t | Vetor de coordenada de x |
| T | Matriz dos vetores de coordenada de X |
| τ_i | Constante de tempo da variável i |
| U | Matriz de vetores de projeção de Y |
| $\hat{\dot{x}}$ | Primeira derivada de um sinal suavizado da variável i |
| $\hat{\ddot{x}}$ | Segunda derivada de um sinal suavizado da variável i |
| x | Vetor de dados de uma variáveis auxiliares |
| X | Matriz de dados das variáveis auxiliares |
| y | Vetor com variável de saída |
| \hat{y} | Variável de saída estimada pelo modelo |
| \bar{y} | Média da variável de saída |
| Y | Matriz das variáveis de saída |

Sumário

| | |
|---|-----------|
| Simbologia e nomenclatura | 8 |
| Sumário | 9 |
| Lista de figuras | 11 |
| Lista de tabelas | 14 |
| Introdução | 17 |
| 1.1 Motivação | 17 |
| 1.2 Inferências e analisadores virtuais | 18 |
| 1.3 Objetivo | 19 |
| Revisão bibliográfica..... | 21 |
| 2.1 Tratamento de dados | 22 |
| 2.1.1 Compressão de dados | 22 |
| 2.1.2 Outliers..... | 23 |
| 2.2 Identificação de estados estacionários..... | 23 |
| 2.2.1 Método de Cao-Rhinehart..... | 24 |
| 2.2.2 Método da derivada | 24 |
| 2.3 Desenvolvimento de modelos empíricos..... | 24 |
| 2.3.1 Regressão linear multivariável..... | 25 |
| 2.3.2 Modelos baseados em PCA..... | 26 |
| 2.3.3 Modelos baseados em PLS | 28 |
| 2.4 Seleção de variáveis | 30 |
| 2.4.2 Avaliação de todos os modelos possíveis | 32 |
| 2.4.3 Procedimentos para seleção de variáveis..... | 33 |
| 2.4.4 Outros métodos..... | 34 |
| 2.5 Validação dos modelos..... | 34 |
| 2.5.1 K-fold..... | 34 |
| 2.5.2 y-rank..... | 34 |
| 2.5.3 Validação cruzada para modelos dinâmicos | 35 |
| Metodologia proposta | 37 |
| 3.1 Estudo da unidade e pré-seleção das variáveis | 37 |
| 3.2 Seleção dos dados e pré-processamento..... | 38 |
| 3.3 Detecção dos estados estacionários..... | 39 |
| 3.4 Seleção de variáveis auxiliares | 40 |
| 3.5 Obtenção do modelo estacionário | 40 |
| 3.6 Inclusão de dinâmicas nas entradas do modelo | 41 |
| Estudo de caso – UPGN | 43 |
| 4.1 Descrição da UPGN | 43 |

| | |
|---|-----------|
| 4.1.1 Sobre o gás natural e seu processamento..... | 43 |
| 4.1.2 Descrição da unidade trabalhada..... | 45 |
| 4.2 Obtenção dos dados de planta..... | 46 |
| 4.3 Pré-tratamento dos dados de planta..... | 49 |
| 4.3.1 Detecção de outliers..... | 49 |
| 4.3.2 Remoção do tempo morto..... | 50 |
| 4.3.3 Identificação dos estados estacionários..... | 51 |
| 4.3.4 Separação do conjunto de validação..... | 52 |
| 4.3.6 Tempo gasto na seleção de variáveis..... | 58 |
| 4.4 Escolha da estrutura do modelo e modelagem..... | 59 |
| 4.4.1 Não-linearidade..... | 59 |
| 4.4.2 Escolha do modelo..... | 59 |
| 4.5 Estudo dos componentes principais..... | 66 |
| Modelo dinâmico..... | 69 |
| 5.1 Modelo estático..... | 69 |
| 5.2 Modelo dinâmico ARX..... | 71 |
| 5.3 Obtenção dos valores ótimos de dinâmicas..... | 74 |
| 5.3.1 Caso 1 com busca exaustiva..... | 75 |
| 5.3.2 Caso 1 com método stepwise..... | 77 |
| 5.3.3 Caso 2 com busca exaustiva..... | 80 |
| 5.3.4 Caso 2 com o método stepwise..... | 82 |
| 5.4 Discussão..... | 83 |
| Conclusão..... | 85 |
| 6.1 Considerações finais..... | 85 |
| 6.2 Sugestão para trabalhos futuros..... | 86 |
| Referências..... | 87 |
| Apêndice A..... | 89 |

Lista de figuras

| | |
|---|----|
| Figura 2.1: Metodologia consolidada para o desenvolvimento de analisadores virtuais apresentada por Kadlec, Gabrys <i>et al.</i> (2009). | 22 |
| Figura 2.2: Exemplo de aplicação do PCA (Shlens, 2009). | 27 |
| Figura 2.3: Mudança de base realizada por PCA. | 27 |
| Figura 2.4: Decomposição PLS..... | 29 |
| Figura 2.5: Comparação entre o erro de ajuste e o erro no conjunto de predição a medida que se aumenta a dimensão do modelo. | 32 |
| Figura 3.1: Diagrama esquemático da metodologia proposta..... | 38 |
| Figura 4.1: Típico processo de tratamento de gás em terra simplificado (Mokhatab, Poe <i>et al.</i> , 2006). | 44 |
| Figura 4.2: Fluxograma simplificado do processo. | 48 |
| Figura 4.3: Dados coletados antes do tratamento..... | 49 |
| Figura 4.4: Método de remoção de outliers 3σ em janela..... | 50 |
| Figura 4.5: Exemplo do resultado da detecção de estados estacionários. | 51 |
| Figura 4.6: Distinção entre os pontos utilizados para validação e modelagem utilizando o método y-rank. | 52 |
| Figura 4.7: Determinação do número de variáveis principais ótimo para o Caso 1 por PLS com e sem validação cruzada. | 54 |
| Figura 4.8: Determinação do número de variáveis principais ótimo para o Caso 1 por PCA com e sem validação cruzada. | 55 |
| Figura 4.9: Gráfico de dispersão entre as variáveis selecionadas por busca exaustiva e a variável modelada. | 56 |
| Figura 4.10: Determinação do número de variáveis principais ótimo para o Caso 2 por PLS com e sem validação cruzada. | 57 |
| Figura 4.11: Determinação do número de variáveis principais ótimo para o Caso 2 por PCA com e sem validação cruzada. | 57 |
| Figura 4.12: Gráfico de dispersão entre as variáveis selecionadas por busca exaustiva e a variável modelada..... | 58 |
| Figura 4.13: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até seis (baixo dir.) no modelo PLS com busca exaustiva para o Caso 1. | 61 |
| Figura 4.14: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até seis (baixo dir.) no modelo PLS com busca exaustiva para o Caso 1. ... | 62 |
| Figura 4.15: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até seis (baixo dir.) no modelo QPLS com busca exaustiva para o Caso 1. | 62 |
| Figura 4.16: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até seis (baixo dir.) no modelo QPLS com busca exaustiva para o Caso 1..... | 63 |
| Figura 4.17: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo PLS com o método <i>stepwise</i> para o Caso 2..... | 64 |

| | |
|---|----|
| Figura 4.18: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo PLS com método <i>stepwise</i> para o Caso 2..... | 64 |
| Figura 4.19: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo QPLS com o método <i>stepwise</i> para o Caso 2..... | 65 |
| Figura 4.20: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo QPLS com método <i>stepwise</i> para o Caso 2..... | 65 |
| Figura 5.1: Validação do modelo estático para o Caso 1 com variáveis selecionadas por busca exaustiva..... | 70 |
| Figura 5.2: Validação do modelo estático para o Caso 1 com variáveis selecionadas pelo método <i>stepwise</i> | 70 |
| Figura 5.3: Validação do modelo estático para o Caso 2 com variáveis selecionadas por busca exaustiva..... | 70 |
| Figura 5.4: Validação do modelo estático para o Caso 2 com variáveis selecionadas por método <i>stepwise</i> | 71 |
| Figura 5.5: Validação do modelo dinâmico direto para o Caso 1 com variáveis selecionadas por busca exaustiva..... | 72 |
| Figura 5.6: Resposta ao degrau para o modelo de ordem 4 para o Caso 1 com busca exaustiva, no eixo das abcissas tempo em 10000 segundos..... | 72 |
| Figura 5.7: Validação do modelo dinâmico direto para o Caso 1 com variáveis selecionadas pelo método <i>stepwise</i> | 72 |
| Figura 5.8: Resposta ao degrau para o modelo de ordem 1 para o Caso 1 com <i>stepwise regression</i> , no eixo das abcissas tempo em 10000 segundos..... | 73 |
| Figura 5.9: Validação do modelo dinâmico direto para o Caso 2 com variáveis selecionadas por busca exaustiva..... | 73 |
| Figura 5.10: Resposta ao degrau para o modelo de ordem 4 para o Caso 1 com busca exaustiva, no eixo das abcissas tempo em 10000 segundos..... | 73 |
| Figura 5.11: Validação do modelo dinâmico direto para o Caso 2 com variáveis selecionadas pelo métodos <i>stepwise</i> | 74 |
| Figura 5.12: Resposta ao degrau para o modelo de ordem 2 para o Caso 2 com <i>stepwise regression</i> , no eixo das abcissas tempo em 10000 segundos..... | 74 |
| Figura 5.13: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 1 com busca exaustiva, onde Td é a constante de tempo..... | 77 |
| Figura 5.14: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 1 com busca exaustiva..... | 77 |
| Figura 5.15: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 1 com o método <i>stepwise</i> , onde Td é a constante de tempo..... | 79 |
| Figura 5.16: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 1 com método <i>stepwise</i> | 80 |
| Figura 5.17: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 2 com busca exaustiva, onde Td é a constante de tempo..... | 81 |
| Figura 5.18: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 2 com busca exaustiva..... | 82 |

| | | |
|---------------------|--|----|
| Figura 5.19: | Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 2 com o método <i>stepwise</i> , onde T_d é a constante de tempo. | 83 |
| Figura 5.20: | Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 2 com método <i>stepwise</i> | 83 |

Lista de tabelas

| | |
|---|----|
| Tabela 2.1: Representação do conjunto de dados para um modelo linear. | 25 |
| Tabela 4.1: Composição típica do gás natural (Mokhatab, Poe <i>et al.</i> , 2006)..... | 43 |
| Tabela 4.2: Variáveis auxiliares pré-selecionadas..... | 47 |
| Tabela 4.3: Determinação do tempo morto..... | 50 |
| Tabela 4.4: Resumo do conjunto de dados antes e após a identificação dos estados estacionários..... | 52 |
| Tabela 4.5: Determinação da dimensão do problema para o Caso 1..... | 54 |
| Tabela 4.6: Variáveis selecionadas para o Caso 1..... | 55 |
| Tabela 4.7: Determinação da dimensão do problema para o Caso 2..... | 56 |
| Tabela 4.8: Variáveis selecionadas para o Caso 2..... | 58 |
| Tabela 4.9: Comparação do tempo gasto para seleção de variáveis..... | 59 |
| Tabela 4.10: Resultado para variação do número de variáveis latentes para o Caso 1. .. | 60 |
| Tabela 4.11: Resultado para variação do número de variáveis latentes. | 63 |
| Tabela 4.12: Componentes principais para o Caso 1 por PLS, onde variáveis com asterisco foram selecionadas para modelo com busca exaustiva e valores sublinhados são o maior peso de cada componente principal. | 67 |
| Tabela 4.13: Comparação entre modelos com variáveis selecionadas por busca exaustiva e através dos pesos dos componentes principais. | 67 |
| Tabela 4.14: Componentes principais para o Caso 2 com PLS, onde variáveis com asterisco foram selecionadas para modelo com busca exaustiva e valores sublinhados são o maior peso de cada componente principal. | 68 |
| Tabela 5.1: Resultados para os modelos estáticos..... | 70 |
| Tabela 5.2: Resultado dos modelos dinâmicos obtidos diretamente com dados de planta. | 71 |
| Tabela 5.3: Definição da ordem do modelo para o Caso 1 com busca exaustiva. | 72 |
| Tabela 5.4: Definição da ordem do modelo para o Caso 1 com o método <i>stepwise</i> | 73 |
| Tabela 5.5: Definição da ordem do modelo para o Caso 2 com busca exaustiva. | 74 |
| Tabela 5.6: Definição da ordem do modelo para o Caso 2 com o método <i>stepwise</i> | 74 |
| Tabela 5.7: Resultados para o modelo dinâmico proposto comparado com o resultado para o modelo estacionário. | 75 |
| Tabela 5.8: Resultado da otimização das dinâmicas para o Caso 1 com busca exaustiva. | 76 |
| Tabela 5.9: Dinâmicas obtidas para o Caso 1 com busca exaustiva..... | 76 |
| Tabela 5.10: Resultado da otimização das dinâmicas para o Caso 1 com o método <i>stepwise</i> | 78 |
| Tabela 5.11: Dinâmicas obtidas para o Caso 1 com o método <i>stepwise</i> | 78 |
| Tabela 5.12: Resultado da otimização das dinâmicas para o Caso 2 com busca exaustiva. | 80 |
| Tabela 5.13: Dinâmicas obtidas para o Caso 2 com busca exaustiva..... | 81 |
| Tabela 5.14: Resultado da otimização das dinâmicas para o Caso 2 com o método <i>stepwise</i> | 82 |
| Tabela 5.15: Dinâmicas obtidas para o Caso 2 com o método <i>stepwise</i> | 82 |
| Tabela A.1: Valores usados na modelagem estacionária para as onze primeiras variáveis..... | 89 |

| | |
|---|----|
| Tabela A.2: Valores usados na modelagem estacionária para as dez últimas variáveis e a variável de saída AI-02.C2..... | 91 |
|---|----|

Capítulo 1

Introdução

1.1 Motivação

Inferências são modelos matemáticos preditivos construídos a partir de medidas de um determinado sistema, geralmente com o objetivo de gerar informações rápidas e frequentes de outra propriedade de difícil medição direta. As inferências têm diversas aplicações na indústria de processos, atuando como ferramentas para monitoramento, detecção de falhas, back-up para outros sensores e no controle de processos.

Essa última é a sua principal aplicação, sendo geralmente a mais encontrada na literatura (Kano, Miyazaki *et al.*, 2000; Fortuna, Graziani *et al.*, 2005; Bolf, Galinec *et al.*, 2009). A automação de processos químicos, além de ter um papel essencial na segurança e estabilidade da unidade industrial, também colabora diretamente na garantia da especificação dos produtos. Se há boas medidas dessas especificações, é possível controlá-las diretamente através de controle avançado ou até mesmo com um controle retroalimentado. Na maioria dos casos é necessário o desenvolvimento de uma inferência, já que essas medidas costumam envolver ensaios complexos de laboratório o que dificulta a realização de análises frequentes. Em algumas situações, é possível fazer a medida com um analisador em linha, porém os tempos de análise e amostragem ainda são altos para controle direto. Outro problema enfrentado por esse tipo de instrumento é a baixa confiabilidade, já que muitas vezes há falta de corpo técnico qualificado para sua manutenção. Isso aponta novamente às inferências como excelente forma de contornar esses problemas e garantir um melhor controle da especificação do produto.

As inferências podem auxiliar no monitoramento dos estados do processo, ajudando a identificar as causas quando o processo está fora da normalidade. Essa tarefa geralmente é feita de forma monovariável pelo operador baseado em sua experiência. Ao se utilizar ferramentas multivariadas, é possível se tomar ações mais rápidas e precisas na correção dos problemas. Outra aplicação semelhante é na detecção de falhas de sensores. Em caso de problemas de medição, o resultado do modelo fica discrepante do apresentado na realidade,

sendo possível identificar a causa da diferença. O modelo também pode ser usado para substituir esse sensor com problemas (Kadlec, Gabrys *et al.*, 2009).

Neste trabalho todas as etapas do processo de desenvolvimento de inferências são estudadas e uma nova metodologia sistemática é proposta. A abordagem utilizada é segmentada, onde a modelagem é feita em duas partes, sendo a primeira a obtenção de um modelo estacionário, e a segunda a determinação das dinâmicas de cada entrada do modelo.

1.2 Inferências e analisadores virtuais

Na literatura internacional, as inferências são geralmente referidas como *soft-sensors*, palavra gerada a partir da união de *software* com *sensor*, já que uma inferência nada mais é do que um instrumento de medida realizado somente com código de computador. Outros nomes utilizados são *inferential sensor*, *virtual on-line analyser* (Kadlec, Gabrys *et al.*, 2009).

Em português, dois nomes são mais utilizados para esse tipo de modelo: inferência e analisador virtual. Facchin (2005) sugeriu uma diferenciação entre esses dois nomes, chamando de inferência o modelo matemático em si e de analisador virtual o conjunto que engloba o modelo – ou seja, a inferência – e o esquema de correção baseado no erro entre a predição do modelo e a medição da variável inferida, que pode ser feita em laboratório ou através de um analisador em linha. Essa correção pode ser realizada de formas simples, como a soma dessa diferença no resultado da predição ou com esquemas mais complexos, utilizando estimadores de estado.

O modelo da inferência pode ser classificado de acordo com sua natureza, sendo chamado de caixa branca o derivado dos princípios da física, também conhecido como modelo de princípios básicos. Esse modelo tem um ótimo poder de extrapolação e pode capturar não linearidades do processo. Porém, o seu desenvolvimento é mais trabalhoso e costuma depender de parâmetros de difícil obtenção, o que o torna menos utilizado. Outro tipo de modelo é o chamado caixa-preta, também conhecido como empírico ou baseado em dados. Como o último nome sugere, esse tipo de modelo utiliza apenas as informações de operação do sistema, sem nenhum conhecimento teórico. A sua principal vantagem é a facilidade de criação, porém o poder de extrapolação é limitado sendo necessária grande quantidade de dados para se obter um modelo de boa qualidade. Um terceiro tipo de modelo também pode ser obtido, conhecido como modelo caixa-cinza. Neste caso, utiliza-se conhecimento fenomenológico do processo, porém a parte mais complexa do modelo é feita utilizando-se dados históricos de operação do sistema em estudo de forma análoga ao caixa-preta.

Mais uma forma de classificação é utilizada por Kano, Miyazaki *et al.* (2000) para modelos caixa-preta onde eles são separados em estáticos, estacionários e dinâmicos. Modelos estacionários são definidos como modelos derivados de dados em estado estacionário. Modelos estáticos são análogos aos estacionários, porém usam dados dinâmicos de processo. Já os modelos dinâmicos usam os mesmos dados dos estáticos, porém incluem dados amostrados em tempos diferentes como entradas.

1.3 Objetivo

Neste trabalho, será sugerida uma nova metodologia para o desenvolvimento de inferências baseadas em dados e será estudado o processo tradicional. A inclusão da etapa correção do modelo não será abordada, apesar de ser uma necessária para a maior confiabilidade do modelo. Serão discutidos apenas os modelos do tipo caixa-preta, porém no estudo de caso serão comparados modelos estáticos, estacionários e dinâmicos.

A metodologia aqui proposta tem como foco facilitar a manutenção da inferência baseada em dados. Para isso o modelo estacionário é obtido separadamente da dinâmica, que é inserida posteriormente. Cada uma das partes é independente e pode ser alterada sem prejuízo à outra. É possível se inserir com mais facilidade o conhecimento do processo, já que formas dinâmicas previamente conhecidas podem ser usadas como base no cálculo dos valores. Por exemplo, caso se saiba de uma resposta inversa existente na relação entre duas variáveis, é possível fixar esse tipo de dinâmica no algoritmo de cálculo, sendo necessária apenas a determinação da grandeza dos parâmetros. Essa forma de abordagem busca também a automatização da manutenção do modelo. Em trabalhos futuros, a inferência poderá ser acoplada a *softwares* de auditoria de processo, que podem fazer a detecção automática de períodos em estacionário e, simultaneamente, melhorar o modelo.

1.4 Estrutura da dissertação

Essa dissertação está dividida em seis capítulos, sendo o primeiro a introdução. No próximo capítulo será feita uma análise geral sobre desenvolvimento de inferências caixa-preta. Serão apresentadas técnicas de tratamento de dados, seleção de variáveis auxiliares, modelagem, entre outras. No terceiro capítulo, a metodologia proposta é esquematizada sendo aplicada em um estudo de caso nos dois próximos capítulos. O capítulo quatro aborda o tratamento de dados e o desenvolvimento do modelo estático. O capítulo cinco mostra a etapa de inserção de dinâmicas no modelo previamente trabalhado. O último capítulo conclui o trabalho e sugere temas para trabalhos futuros partindo do desenvolvimento aqui realizado.

Capítulo 2

Revisão bibliográfica

Neste capítulo será apresentada uma visão geral do desenvolvimento de inferências e analisadores virtuais. Serão apresentadas e discutidas diversas técnicas que abrangem cada uma das etapas do desenvolvimento, que vão desde a coleta de dados até a validação final do modelo. Por fim será feita uma análise dos trabalhos de aplicação industrial destas técnicas existentes na literatura. O estudo realizado se focou apenas em processos contínuos, apesar de a aplicação em processos em batelada ser extensa.

Uma metodologia consolidada existente para o desenvolvimento de analisadores virtuais está mostrada na Figura 2.1. A primeira etapa consiste em uma avaliação inicial dos dados para se identificar problemas, como variáveis congeladas, por exemplo, e também avaliar a riqueza dos dados, verificando se há necessidade de realizar perturbações na planta para melhorá-los. Na próxima etapa o conjunto de dados é selecionado com ajuda da avaliação feita anteriormente e os períodos em que a planta estava em estado estacionário são identificados. A terceira etapa é responsável por lapidar os dados de forma a eliminar problemas como *outliers* e prepará-los para a modelagem, realizando a sua normalização para média zero e desvio-padrão unitário. Na quarta etapa se define o tipo de modelo a ser usado para depois ajustá-lo e quantificar seu desempenho. Essa etapa é de grande importância e tem forte impacto no resultado final, em geral testam-se diversos modelos e compara-se o seu desempenho para realizar a escolha. Por fim, o modelo passa a ser usado *online* no processo. Com o passar do tempo, seu desempenho deve ser reavaliado para que sejam feitos ajustes. Maiores detalhes sobre e cada etapa podem ser vistos nos trabalhos de Fortuna, Graziani *et al.* (2007) e Kadlec, Gabrys *et al.* (2009), onde são feitas extensas análises do estado da arte do desenvolvimento de inferências.

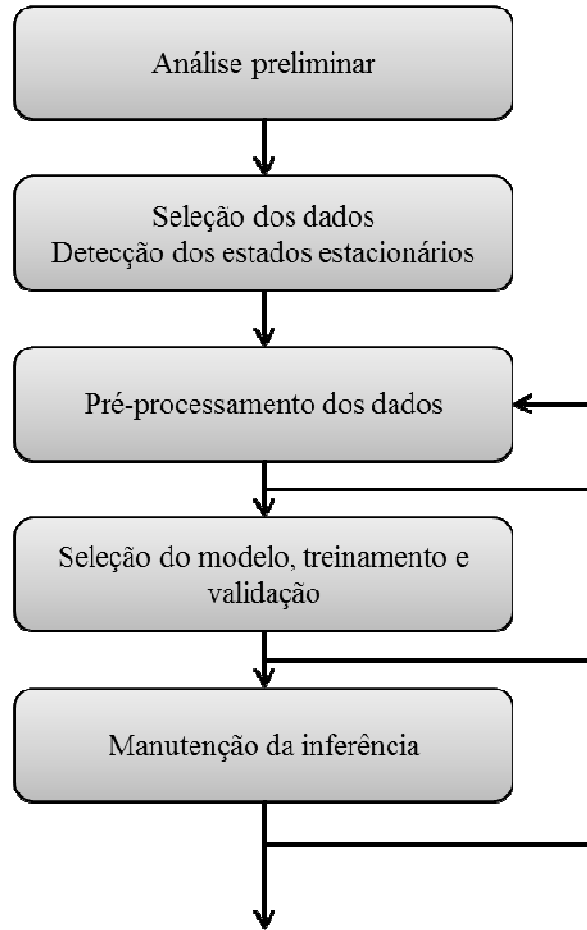


Figura 2.1: Metodologia consolidada para o desenvolvimento de analisadores virtuais apresentada por Kadlec, Gabrys *et al.* (2009).

2.1 Tratamento de dados

Grandes indústrias tem necessidade de coletar e armazenar grandes quantidades de dados das principais variáveis do seu processo, seja por motivos de segurança, monitoramento da qualidade, otimização do processo, ou qualquer outro. Isso abre caminho para o uso desses dados na identificação de um modelo. Porém, muitas vezes esses dados apresentam problemas, como tempo de amostragem inadequado e a presença de *outliers* que podem torná-los inadequados para esse tipo de tarefa (Fortuna, Graziani *et al.*, 2007). Alguns desses problemas e suas consequências são resumidos a seguir.

2.1.1 Compressão de dados

Um problema bastante conhecido é a compressão de dados utilizada pelos historiadores de processo. Thornhill, Choudhury *et al.* (2004) mostraram que isso interfere em diversos tipos de análises baseadas em dados e recomendam fortemente que se tenha precaução no uso de dados arquivados com compressão. Nesse mesmo trabalho, os autores apresentaram um método para quantificação da compressão baseado no cálculo da segunda derivada do sinal que pode ajudar na avaliação de dados comprimidos.

2.1.2 Outliers

A presença de *outliers* é outro problema comum e deve ser tratado e eliminado. *Outliers* são observações não consistentes com a maioria dos dados e incluem pontos inexistentes, falhas de leituras, problemas de transmissão etc. Um modelo de inferência baseado em dados tem seu resultado deteriorado caso o conjunto utilizado possua muitos *outliers*, como explica Lin, Recke *et al.* (2007).

Parte dos *outliers* pode ser facilmente removida ao se analisar os dados em conjunto com o processo em questão. Valores que violam uma limitação técnica ou física são facilmente identificados quando esses limites são conhecidos e devem ser eliminados através de um processamento lógico simples.

A outra parte dos outliers necessita de técnicas mais complexas. Uma forma comum de se tratar esse problema é remover os pontos que estão além de um limite de três desvios-padrão do conjunto de dados, conhecido como método 3σ . Uma extensa aplicação dessa técnica é feita por Fortuna, Rizzo *et al.* (2007), que compara o resultado com outras existentes na literatura.

O método 3σ , porém, apresenta problemas, pois no cálculo da estatística, os próprios *outliers* também são incluídos, o que distorce o valor da média e do desvio-padrão. Para contornar esse problema, Lin, Recke *et al.* (2007) utilizaram em seu trabalho um método que substitui esses dois parâmetro por outros baseados no cálculo do desvio absoluto da mediana (MAD). Esse método é conhecido como identificador de *Hampel* e, de acordo com Pearson (2001), é uma forma bastante efetiva de se lidar com *outliers*.

Esses dois métodos são baseados em estatística monovariável. Existem outras técnicas que utilizam estatística multivariável, sendo a maioria delas baseadas na análise dos componentes principais (PCA). Um dos métodos mais utilizados é o parâmetro de Jolliffe, que foi estudado por Fortuna, Rizzo *et al.* (2007) e aplicado por Warne, Prasad *et al.* (2004).

2.2 Identificação de estados estacionários

A identificação de estados estacionários tem diversas aplicações práticas na indústria. É importante determinar os períodos de operação em estacionário para a realização de uma modelagem fenomenológica em um simulador estático de processos, por exemplo. Além disso, os dados estacionários auxiliam no acompanhamento da qualidade da produção e ajudam a identificar problemas de operação. Neste trabalho ela será útil para a obtenção de um modelo estacionário baseado em dados, onde essa tarefa é essencial.

De acordo com Kadlec, Gabrys *et al.* (2009) geralmente usa-se método manual de seleção. Porém, essa tarefa pode ser bastante trabalhosa caso se disponha de um conjunto grande de dados. Na literatura há diversos métodos disponíveis para realização da detecção automática de estados estacionários e alguns deles estão descritos a seguir.

2.2.1 Método de Cao-Rhinehart

Um dos métodos mais conhecidos foi proposto por Cao e Rhinehart (1995) e se baseia na comparação da variância do ruído do sinal estimada por duas maneiras distintas. Caso essas variâncias sejam aproximadamente equivalentes, diz-se que a média no processo é constante e trata-se de um período em estado estacionário. Caso essas variâncias sejam consideradas distintas, diz-se que tem-se uma mudança de média e não se está em estado estacionário. Esse método, porém, requer o ajuste de quatro parâmetros que são difíceis de serem definidos, o que torna o método pouco confiável.

2.2.2 Método da derivada

Outro método recentemente proposto é baseado no cálculo da derivada do sinal (Duarte, Mejía *et al.*, 2010; Mejía, Farias *et al.*, 2010). O índice de estacionariedade desenvolvido baseia-se na estimativa das primeiras e segundas derivadas dos sinais suavizados, $\hat{\dot{x}}_l$ e $\hat{\ddot{x}}_l$ respectivamente. Com esses dois valores, calcula-se o índice de estacionariedade, $I_{DER,i}$, apresentado na equação 2.1.

$$I_{DER,i} = \left| \hat{\dot{x}}_l \right| + 10 \left| \hat{\ddot{x}}_l \right| \quad (2.1)$$

A estimação da derivada é feita com o auxílio de transformadas wavelets discretas. Essa técnica tem problemas pelo efeito de bordas. O sinal de entrada tem um comprimento finito, e assim um transiente aparente abrupto pode acontecer nos limites dos dados. Para evitar isto é usado o método da transformação de translação e rotação (TRT) eliminando este transiente e subtraindo uma componente linear do sinal de entrada para reduzir o efeito de bordas. Para implementar o transformada wavelet discreta (DWT) é usado o algoritmo *à trous*, por se tratar da implementação original e mais simples.

Outros métodos de detecção podem ser encontrados nos trabalhos de Duarte, Mejía *et al.* (2010) e Cao e Rhinehart (1995).

2.3 Desenvolvimento de modelos empíricos

Modelos empíricos são baseados somente em dados, não levando em consideração a origem dessas informações e qual o fenômeno está por trás da sua geração. Isso faz com que seu uso seja difundido em diversas áreas da ciência que vão desde a medicina até a economia e ciências sociais. Os modelos empíricos requerem grande quantidade de dados e com grande riqueza de informação. Esses modelos podem ser interpolados sem grandes problemas, porém a sua extrapolação pode implicar em resultados bastante diferentes da realidade (Cinar, Palazoglu *et al.*, 2007)

A seguir serão apresentadas algumas técnicas utilizadas para o desenvolvimento desses modelos. Os primeiros são modelos lineares que vão desde a regressão multivariável simples até modelos baseados em técnicas de redução de dimensionalidade, como a regressão em componentes principais (PCR) e a regressão baseada em mínimos quadrados parciais

(PLSR). Entre os tipos de modelos não-lineares está uma versão quadrática do mínimos quadrados parciais (PLS) e as redes neurais.

2.3.1 Regressão linear multivariável

A formulação de um modelo linear relacionando p variáveis auxiliares x_1, x_2, \dots, x_p com a variável resposta y está mostrado na equação a seguir.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon \quad (2.2)$$

Na Eq. 2.2, os índices $\beta_0, \beta_1, \dots, \beta_p$ são constantes que representam os coeficientes de regressão do modelo e ε representa o erro do modelo ou um distúrbio aleatório. Uma representação do conjunto de dados para n observações está mostrado na Tabela 2.1 (Chatterjee e Hadi, 2006).

Tabela 2.1: Representação do conjunto de dados para um modelo linear.

| Observações | Variável resposta | Variáveis auxiliares | | | |
|-------------|-------------------|----------------------|----------|----------|----------|
| | y | x_1 | x_2 | ... | x_3 |
| 1 | y_1 | x_{11} | x_{12} | ... | x_{1p} |
| 2 | y_2 | x_{21} | x_{22} | ... | x_{2p} |
| 3 | y_3 | x_{31} | x_{32} | ... | x_{3p} |
| \vdots | \vdots | \vdots | \vdots | \ddots | \vdots |
| N | y_n | x_{n1} | x_{n2} | ... | x_{np} |

O problema da regressão linear consiste em estimar o vetor de parâmetros β de forma a obter a melhor representação da variável de resposta y . Existem diversos métodos para realizar essa tarefa, sendo o mais difundido o método dos mínimos quadrados, que busca a minimização da soma do quadrado dos erros (Chatterjee e Hadi, 2006). A solução analítica é dada pela equação a seguir.

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (2.3)$$

A solução do problema determinará um conjunto de coeficientes de regressão estimados representados por $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p \quad (2.4)$$

Os estimadores baseados em mínimos quadrados têm diversas propriedades com as quais é possível se realizar inferências estatísticas como a estimação do intervalo de confiança e testes de hipótese (Chatterjee e Hadi, 2006). Porém, como ressaltou Facchin (2005), a solução do problema de mínimos quadrados requer o cálculo da inversa da matriz $\mathbf{X}^T \mathbf{X}$ relacionada com o conjunto de dados que pode levar a problemas numéricos. Dados colineares (variáveis auxiliares altamente correlacionadas) gera uma matriz posto-deficiente o

que impossibilita o cálculo da inversa e, desta forma, a solução do problema de mínimos quadrados.

2.3.2 Modelos baseados em PCA

A análise dos componentes principais é uma ferramenta padrão para a análise de dados moderna, pois é um método simples para a extração de informações relevantes em grandes conjuntos de dados. Com pouco esforço, o PCA fornece um caminho para a redução da dimensão de um conjunto complexo de dados e ajuda na visualização das estruturas por trás dele (Shlens, 2009).

Um exemplo está mostrado na Figura 2.2, onde um sistema massa-mola é monitorado por três câmeras. Esse sistema pode ser descrito em apenas uma variável, já que sua variação se dá apenas ao longo do eixo das abcissas. Porém, a medição das três câmeras nos dão seis variáveis para descrever o problema. Essas medições são todas redundantes e apenas uma é necessária. A análise PCA permite identificar o grau de correlação entre as variáveis e permite determinar quantas seriam necessárias para descrever o problema satisfatoriamente.

Esse tipo de conjunto de dados é chamado multicolinear, ou seja, algumas – ou todas, de suas variáveis são altamente correlacionadas (como é o caso do exemplo da Figura 2.2) (Chatterjee e Hadi, 2006). A análise dos componentes principais é uma forma de detectar essa multicolinearidade e se baseia no fato de que qualquer conjunto de variáveis pode ser transformado em outro conjunto de variáveis ortogonais da seguinte forma:

$$T = XP \quad (2.5)$$

Na Eq. 2.5 a matriz \mathbf{X} representa os dados originais, \mathbf{P} é a matriz dos vetores de projeção (*loading vectors*) e \mathbf{T} é matriz dos vetores de coordenada (*score vectors*). \mathbf{T} nada mais é do que os dados originais representados nessa nova base ortogonal. \mathbf{P} é a transformação linear que faz a mudança de base.

Essa transformação linear é escolhida de forma que a primeira das novas variáveis – t_1 – tenha a máxima variância possível, a próxima variável será escolhida da mesma forma, porém ela deve ser ortogonal a primeira escolhida. O resultado desse procedimento pode ser melhor entendido com a visualização da transformação em um conjunto de duas variáveis, como é mostrado na Figura 2.3. Para este caso, é fácil perceber que a nova variável t_1 exprime a grande maioria da variabilidade no problema, sendo a variabilidade presente em t_2 basicamente devido ao ruído de medição. Desta forma, seria necessária apenas a variável t_1 para representar com fidelidade o problema que antes precisava das duas variáveis x_1 e x_2 .

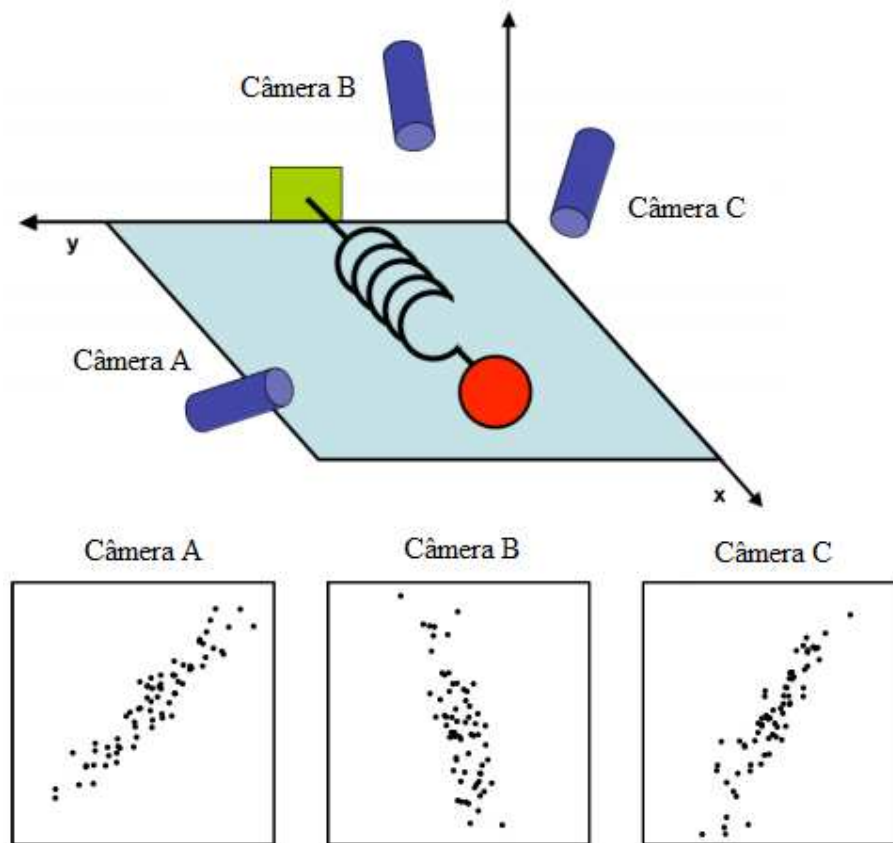


Figura 2.2: Exemplo de aplicação do PCA (Shlens, 2009).

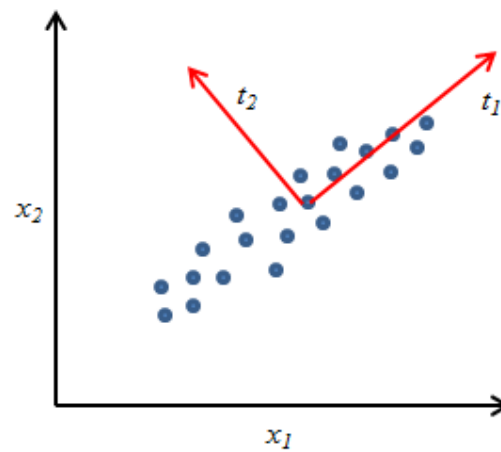


Figura 2.3: Mudança de base realizada por PCA.

Para a obtenção das matrizes \mathbf{P} utiliza-se normalmente o algoritmo NIPALS, que é o mais difundido método para tal propósito. Mais detalhes sobre o seu funcionamento podem ser obtidos em Facchin (2005).

Regressão de componentes principais – PCR

Segundo Cinar, Palazoglu *et al.* (2007), a regressão dos componentes principais é uma das técnicas disponíveis para se lidar com matrizes de dados mal condicionadas ao fazer a regressão das variáveis auxiliares em cima da matriz de vetores de coordenada \mathbf{T} . O problema de regressão linear mostrado anteriormente pode ser reescrito da seguinte forma:

$$y = \beta_0 + \beta_1 t_1 + \beta_2 t_2 + \dots + \beta_p t_p + \varepsilon \quad (2.6)$$

Onde a matriz de coeficientes de regressão ótima pode ser obtida analiticamente pelo método dos mínimos quadrados da seguinte forma:

$$\hat{\beta} = (T^T T)^{-1} T^T Y \quad (2.7)$$

Ao contrário da inversão de $\mathbf{X}^T \mathbf{X}$, onde alguns dos vetores x podiam ser colineares, a inversão de $\mathbf{T}^T \mathbf{T}$ não causa esse problema, pois os vetores são ortogonais (Cinar, Palazoglu *et al.*, 2007).

2.3.3 Modelos baseados em PLS

O método dos mínimos quadrados parciais pode ser visto como uma melhoria do método PCA. Enquanto o PCA trabalha apenas com a matriz das variáveis auxiliares \mathbf{X} , o PLS utiliza também as informações da matriz das variáveis a serem modeladas \mathbf{Y} no procedimento de determinação das variáveis principais no problema, aqui chamadas e variáveis latentes. Esse método aplica transformações lineares nas matrizes \mathbf{Y} e \mathbf{X} de forma a buscar novas direções que maximizam a relação entre essas duas matrizes. Pode-se pensar no problema de forma análoga ao PCA, porém desta vez se faz mudanças em \mathbf{Y} na busca da nova base ótima para o problema. As transformações lineares aplicadas estão mostradas nas equações a seguir.

$$X = TP^T + E \quad (2.8)$$

$$Y = UQ^T + F \quad (2.9)$$

Uma melhor visualização do problema pode ser obtida na Figura 2.4. Como se vê, o número de variáveis auxiliares é reduzido de p para a , sendo a o número de variáveis latentes escolhido para o problema. \mathbf{E} e \mathbf{F} representam as matrizes residuais, que ocorre devido à redução da dimensão do problema. Para a determinação da solução desse problema, o algoritmo mais utilizado também é o NIPALS. Mais detalhes sobre o algoritmo podem ser obtidos em Facchin (2005).

O modelo de regressão baseado em PLS nada mais é do que a aplicação de uma regressão linear multivariável entre a matriz de variáveis auxiliares transformada \mathbf{T} e a matriz de variáveis de interesse transformada \mathbf{U} . Mais informações sobre esse problema podem ser obtidos em Wold, Sjostrom *et al.* (2001) e Cinar, Palazoglu *et al.* (2007). Em seu trabalho,

Wold, Sjostrom *et al.* apresentam diversos exemplos de aplicações da regressão baseada em PLS para problema da quimiometria e mostram as vantagens de se utilizar esse método com alternativa à regressão multivariável quando os dados são altamente correlacionados.

$$\begin{array}{c}
 \begin{array}{c} p \\ \boxed{X} \\ n \end{array} = \begin{array}{c} a \\ \boxed{T} \\ n \end{array} \begin{array}{c} p \\ \boxed{P'} \\ a \end{array} + \begin{array}{c} p \\ \boxed{E} \\ n \end{array} \\
 \\
 \begin{array}{c} q \\ \boxed{Y} \\ n \end{array} = \begin{array}{c} a \\ \boxed{U} \\ n \end{array} \begin{array}{c} q \\ \boxed{Q'} \\ a \end{array} + \begin{array}{c} q \\ \boxed{F} \\ n \end{array}
 \end{array}$$

Figura 2.4: Decomposição PLS.

2.3.4 PLS quadrático

De acordo com Finkler (2003), ao se lidar com sistemas físicos e químicos complexos, o método PLS linear nem sempre é adequado para modelar a estrutura subjacente, que pode ser altamente não-linear. Ao se aplicar o método PLS linear em problemas não-lineares, muitas vezes se descarta informações importantes juntamente com as dimensões menos relevantes do problema. Isso ocorre porque a variabilidade dos blocos que não pode ser capturada pelo mapeamento linear acaba sendo confundida com os resíduos.

Para modelar relações não-lineares entre X e Y, as suas projeções devem ser relacionadas de forma não-lineares entre si (Wold, Kettanehwold *et al.*, 1989). O PLS quadrático (QPLS) utiliza uma função polinomial de segunda ordem para realizar tal tarefa, como mostra a equação a seguir.

$$u_i = c_{0i} + c_{1i}t_i + c_{2i}t_i^2 + \varepsilon_i \quad (2.10)$$

Na Eq. 2.10, c_{0i} , c_{1i} e c_{2i} são constantes e ε_i representa o vetor com os erros.

O algoritmo de cálculo deste método é uma modificação do NIPALS, utilizado para os métodos lineares. Essa modificação foi introduzida no trabalho de Wold, Kettanehwold *et al.* (1989) e é aplicável para qualquer função genérica relacionando T e U, sendo a relação quadrática um caso particular. Mais detalhes sobre o algoritmo podem ser obtidas também nos trabalhos de Finkler (2003) e Facchin (2005).

2.3.5 Outros métodos

Na literatura, inúmeros outros tipos de modelos foram usados no desenvolvimento de inferência de propriedade de processos. Dentre os modelos não-lineares, os mais comuns são

as redes neuronais, cuja inspiração é a estrutura biológica das células neuronais do cérebro. Salvatore (2007) utilizou redes neuronais para inferir a concentração de enxofre no óleo diesel produto de uma unidade de hidrotreatamento. Nesse trabalho são descritos os fundamentos dos principais tipos de redes neuronais utilizados no desenvolvimento de inferências. Outra aplicação pode ser encontrada em Fortuna, Graziani *et al.* (2005).

Recentemente, modelos utilizando ICA (*Independent Component Analysis*) foram relatados na literatura. O ICA é aplicado no processamento de sinais multivariáveis com o objetivo de encontrar fatores ocultos em um determinado conjunto de variáveis, de forma semelhante ao PCA. Um exemplo de aplicação do ICA no desenvolvimento de inferências pode ser visto no trabalho de Kaneko, Arakawa *et al.* (2009).

Também são encontrados modelos que utilizam mais de uma técnica na sua obtenção. Modelos que combinam PCA com redes neuronais são bastante difundidos. Nesse método, no lugar a regressão linear, utiliza-se uma rede neuronal para relacionar a matriz \mathbf{T} com a matriz de variáveis de interesse \mathbf{Y} .

2.4 Seleção de variáveis

De acordo com Facchin (2005), a etapa de seleção de variáveis é de fundamental importância na construção de qualquer modelo e é fortemente responsável pelo sucesso da modelagem. Deve-se ter cuidado para não incluir um número excessivo de variáveis, pois isso pode causar problemas ao modelo. A adição de uma variável a mais no modelo baseado em estimação por mínimos quadrados quase sempre aumenta a variabilidade da resposta final, o que torna o modelo resultante muito sensível a ruídos e pequenas mudanças no conjunto de dados, resultando em um modelo pouco robusto (Warne, Prasad *et al.*, 2004; Chatterjee e Hadi, 2006).

Técnicas simples podem ser usadas na avaliação preliminar das variáveis disponíveis, como gráficos de dispersão e coeficientes de correlação. A análise do gráfico de cada variável independente versus a dependente permite uma busca visual pelas relações existentes entre elas. Se os pontos se organizam na forma de uma linha reta, isso indica uma relação linear. Um conjunto de pontos curvado pode indicar uma relação não-linear entre as duas variáveis. Caso não haja nenhum padrão, é possível que não haja relação nenhuma entre as duas variáveis. Coeficientes de correlação também podem ser usados e dão uma indicação da relação linear entre duas variáveis e podem ajudar na seleção de uma variável auxiliar. Porém, o cálculo de simples coeficiente de correlação entre variáveis independentes e a variável dependente ignora o efeito das outras variáveis candidatas no conjunto. Desta forma, em um caso multivariável a verdadeira correlação pode ser ocultada pelo efeito de outras variáveis. Essa técnica sozinha não é satisfatória para selecionar em definitivo as variáveis auxiliares (Warne, Prasad *et al.*, 2004).

2.4.1 Critérios para avaliação dos modelos

Parte importante da seleção das variáveis é a forma com que a qualidade dos modelos será comparada. Existem diversos fatores disponíveis na literatura e a grande maioria deles se baseia na soma dos quadrados do erro, SS_{Erro} (Facchin, 2005), reproduzido a seguir.

$$SS_{\text{Erro}} = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2.11)$$

Um dos índices mais utilizados é o coeficiente de determinação (R^2), que compara a soma dos quadrados do erro de ajuste de modelo com a soma do quadrado total.

$$SS_{\text{Total}} = \sum_{i=1}^N (y_i - \bar{y}_i)^2 \quad (2.12)$$

$$R^2 = 1 - \frac{SS_{\text{Erro}}}{SS_{\text{Total}}} \quad (2.13)$$

Esse índice faz uma comparação entre o ajuste obtido e a simples média dos dados. Quanto melhor o modelo for em relação a esse modelo representado pela média simples dos dados, mais próximo de 1 será o índice R^2 .

Esse índice, porém, não é bom para avaliar o incremento de variáveis, pois seu valor sempre melhorará à medida que se adicionam variáveis e ele tenderá assintótica e monotonicamente para o melhor valor de R^2 . Desta forma, outros índices foram propostos penalizando o desempenho do modelo pelo número de variáveis utilizado. Vários deles estão mostrados em Facchin (2005) e Chatterjee e Hadi (2006), sendo um dos mais difundidos o critério de Akaike, mostrado na equação a seguir, onde n é o número de amostras testadas e p o número de variáveis contidas no modelo.

$$AIC = \log\left(\frac{SS_{\text{Erro}}}{n}\right) + \frac{2p}{n} \quad (2.14)$$

Os índices apresentados anteriormente não preveem a separação dos dados em dois conjuntos, sendo um para modelagem e outro para validação. Se essa divisão for feita, não é necessário utilizar índices penalizando o modelo pelo número de variáveis. Ao se avaliar o modelo em um outro conjunto de dados previamente separado, a soma do erro quadrático nesse conjunto irá começar a aumentar quando o número de variáveis utilizado ficar excessivo, como mostra a Figura 2.5.

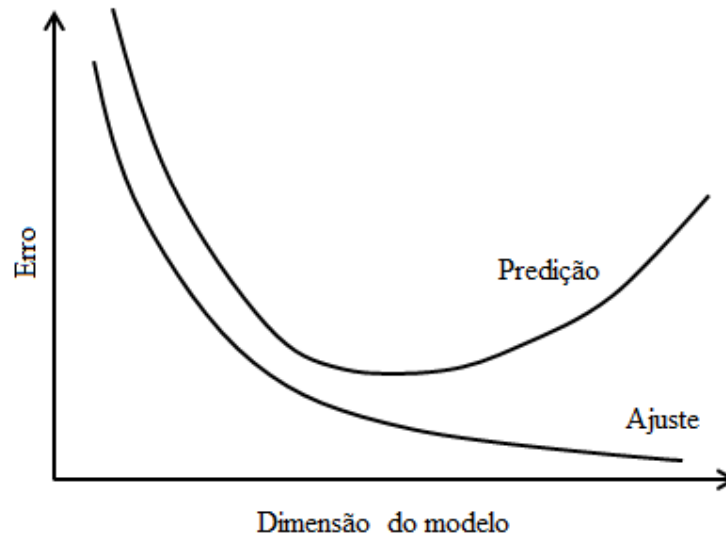


Figura 2.5: Comparação entre o erro de ajuste e o erro no conjunto de predição a medida que se aumenta a dimensão do modelo.

À soma do quadrado dos erros neste conjunto separado para avaliação do modelo dá-se o nome de PRESS (*predictive sum of squares*), como mostra a equação a seguir.

$$PRESS = \sum_{i=1}^N (y_i - \hat{y}_{pred,i})^2 \quad (2.15)$$

Esse índice é utilizado como principal forma de avaliação de modelos em diversos trabalhos (Kano, Miyazaki *et al.*, 2000; Facchin, 2005; Lin, Recke *et al.*, 2007). Um índice equivalente ao R^2 foi utilizado por Kaneko, Arakawa *et al.* (2009) e está apresentado na equação a seguir.

$$Q^2 = 1 - \frac{PRESS}{SS_{Total}} \quad (2.16)$$

2.4.2 Avaliação de todos os modelos possíveis

A forma mais simples de seleção de variáveis consiste na avaliação de todas as combinações possíveis para obter o conjunto que melhor modela o sistema. Esse método é também conhecido como busca exaustiva e tem como ponto negativo exigir grande carga computacional para ser usado nos casos em que o número de variáveis a serem escolhidas for elevado.

O número de modelos a serem avaliados por essa metodologia é proporcional ao número de variáveis auxiliares candidatas. Supondo que exista também um modelo para zero variáveis consistindo de apenas uma constante, o número total de modelos é 2^q , onde q é o número de variáveis auxiliares testadas. Desta forma, se esse número for 4, teremos 16 modelos para comparação. Caso o valor aumente para 5, serão 32 modelos. Subindo para 6

variáveis, teremos 64 e assim por diante (Facchin, 2005). Esse fato era impeditivo para seu uso antigamente, tanto que foram desenvolvidas alternativas menos custosas para seleção de variáveis, descritas a seguir. Atualmente o método de busca exaustiva é bastante viável e deve ser sempre avaliado como opção, já que seu resultado será o melhor possível, uma vez que o ótimo global será sempre atingido.

2.4.3 Procedimentos para seleção de variáveis

Para casos em que o número de variáveis é muito grande, pode-se utilizar procedimentos sistemáticos para avaliação dos modelos que não necessitem a avaliação de todos os modelos possíveis. Para q variáveis, esses métodos necessitam que avaliemos no máximo $(q + 1)$ modelos, ao contrário de 2^q que seriam todas as combinações possíveis. Esses procedimentos são classificados em duas categorias: o procedimento de *forward selection* e o de *backward elimination*. Há também uma modificação bastante popular do método FS, chamado de método *stepwise*. Esses três métodos estão descritos a seguir (Chatterjee e Hadi, 2006).

Forward Selection

Neste método se inicia com um modelo sem nenhuma variável auxiliar, apenas uma constante. A primeira variável a ser adicionada é a que tem a maior correlação com a variável principal y . Se o coeficiente de regressão β_1 é significativamente diferente de zero, a variável é retida no modelo e se busca a segunda variável. A próxima variável será a que tiver a maior correlação com o resíduo da regressão anterior, ou seja, $y - \hat{y}$. A significância da regressão com a segunda variável é então testada. Se o coeficiente de regressão é significativo, a busca pela terceira variável é iniciada. O procedimento se encerra quando a última variável entrando no modelo tem um coeficiente de correlação insignificante ou todas as variáveis estão incluídas no modelo. A significância do coeficiente de regressão de cada variável adicionada é avaliado por um teste- t calculado para o último modelo obtido.

Backward elimination

O método de *backward elimination* inicia com um modelo contendo todas as variáveis auxiliares que são removidas sucessivamente uma por uma. As variáveis são eliminadas baseando-se na sua contribuição para a redução da soma do quadrado dos erros. A primeira variável eliminada é a que apresenta menor contribuição. Isso é equivalente a eliminar a variáveis que tem o menor teste- t no modelo. Se todos os testes- t são significativos, todas as variáveis são retidas no modelo. Caso haja uma ou mais que não sejam significativas, a com menor significância será removida. Um novo modelo com as $(q - 1)$ variáveis que sobraram é então ajustado e os coeficientes da regressão obtidos são avaliados com testes- t . O procedimento termina quando todos os testes- t são relevantes. Caso o modelo termine sem nenhuma variável, ou seja, somente o termo constante, terão sido avaliados $(q + 1)$ modelos. Isso pode acontecer caso o critério de avaliação do teste- t seja muito rigoroso.

Método stepwise

O método *stepwise* é essencialmente um procedimento *forward selection* com uma etapa adicional em que há a possibilidade de se eliminar uma variável da mesma forma com que era feito no procedimento de *backward elimination*. Neste método a variável que entra nos primeiros estágios da seleção pode ser eliminada nas etapas finais. Os cálculos feitos são iguais aos já descritos para os métodos *forward selection* e *backward elimination*.

2.4.4 Outros métodos

Alguns outros métodos são utilizados na literatura, porém não serão tratados com mais detalhes nesse trabalho. Dentre eles está a técnica dos algoritmos genéticos, que foi comparada por Facchin (2005) com o método *stepwise* e a busca exaustiva sem obter bons resultados.

2.5 Validação dos modelos

Para a escolha do melhor modelo, é necessário um critério para compará-los. Como foi comentado na etapa de seleção de variáveis, é importante fazer a separação dos dados em dois conjuntos, um para modelagem e outro para validação. O modelo é ajustado com o conjunto de modelagem e os índices de avaliação são calculados com os valores preditos pelo modelo no conjunto de validação. Essa técnica é conhecida como validação cruzada e evita que o modelo se torne viciado em um determinado conjunto de dados, prejudicando seu desempenho quando utilizado em outro conjunto.

Como o conjunto de dados é finito e as vezes escasso, diversas técnicas foram desenvolvidas na busca da melhor maneira de se fazer essa separação. Deseja-se obter um bom número de pontos para a modelagem, para não prejudicar essa etapa, porém é necessário também um conjunto rico para fazer a melhor comparação possível. A seguir serão descritas algumas técnicas disponíveis.

2.5.1 K-fold

Nessa técnica primeiramente deve-se definir um valor k que será o número de conjuntos em que os dados originais serão divididos. Geralmente utiliza-se k igual a 10. Os dados são divididos aleatoriamente nesses conjuntos e cada um deles servirá de conjunto de validação para um modelo que será criado com união dos outros conjuntos. Ou seja, será criado um número total de modelos igual ao valor de k . Para cada modelo, um índice de desempenho será calculado para a sua predição no conjunto de validação. O índice final de validação será a soma do resultado para todos os conjuntos.

2.5.2 y-rank

Esse método primeiramente faz uma classificação dos dados de forma crescente da variável a ser modelada. Depois disso, os conjuntos são separados colocando-se um ponto no conjunto de validação e dois no conjunto de calibração. Pode-se utilizar relações diferentes entre os tamanhos do conjunto, ficando essa definição como um grau de liberdade dos

métodos. No trabalho de Facchin (2005), diversas técnicas de seleção de separação de conjuntos foram comparadas sendo a *y-rank* a que mostrou melhores resultados.

2.5.3 Validação cruzada para modelos dinâmicos

Os dois métodos apresentados anteriormente são válidos apenas para conjuntos de dados em estado estacionário. Geralmente, para modelos dinâmicos faz-se a seleção dos conjuntos manualmente, procurando garantir que nos dois as variáveis tenham as maiores variâncias possíveis, varrendo um grande número de variáveis.

Existem algumas técnicas que só podem ser utilizadas para modelos dinâmicos, como mostram Fortuna, Graziani *et al.*, (2005). Uma delas consiste na determinação da função de autocorrelação do resíduo do modelo. Quanto melhor o modelo, mais próximo o resíduo estará de um ruído branco e a função de autocorrelação será zero logo após o primeiro atraso. Essas técnicas podem ser usadas ao lado da validação cruzada para obter uma melhor avaliação do modelo dinâmico.

Capítulo 3

Metodologia proposta

Nesse capítulo será apresentada a metodologia sugerida para o desenvolvimento de inferências de processos. Um diagrama esquemático está apresentado na Figura 3.1 mostrando as principais etapas da metodologia. A seguir cada uma dessas etapas será detalhada mostrando as melhores práticas a serem adotadas.

Como já foi mencionado na introdução, o diferencial dessa metodologia em comparação com as apresentadas na literatura (Fortuna, Graziani *et al.*, 2007; Kadlec, Gabrys *et al.*, 2009) é a abordagem segmentada do problema. O modelo estacionário é desenvolvido numa primeira etapa e a dinâmica é inserida posteriormente, utilizando um algoritmo de otimização. Esse tipo de abordagem permite uma maior flexibilidade no desenvolvimento da inferência, sendo possível usar diferentes técnicas para cada etapa. Outro diferencial é a possibilidade de se inserir conhecimento prévio das dinâmicas do processo. Caso já se saiba de antemão o valor do tempo morto ou a existência de uma resposta inversa, por exemplo, é possível utilizar esses fatos para dar apoio à etapa de inserção de dinâmicas, melhorando o resultado. A manutenção do modelo também fica facilitada com essa abordagem, pois é possível fazê-la da mesma forma segmentada. Caso o modelo estacionário seja a causa do mau desempenho, pode-se refazê-lo apenas. Por outro lado, se estamos com problemas nas dinâmicas, basta utilizar um novo conjunto de dados para refazê-la, mantendo o modelo estacionário intacto.

3.1 Estudo da unidade e pré-seleção das variáveis

O primeiro passo a ser dado no desenvolvimento de uma inferência de processo, é a consolidação do conhecimento da unidade industrial em questão. É importante obter auxílio de operadores e engenheiros experientes para que se conheça os valores típicos de cada variável e as diferentes formas de operar da unidade.

Recomenda-se que um fluxograma simplificado da unidade seja desenvolvido nessa etapa, como forma de consolidar o entendimento do processo. Esse fluxograma deve ser

usado como principal referência na posterior seleção das variáveis auxiliares. Ele também será útil na etapa de avaliação do modelo dinâmico obtido para o processo.

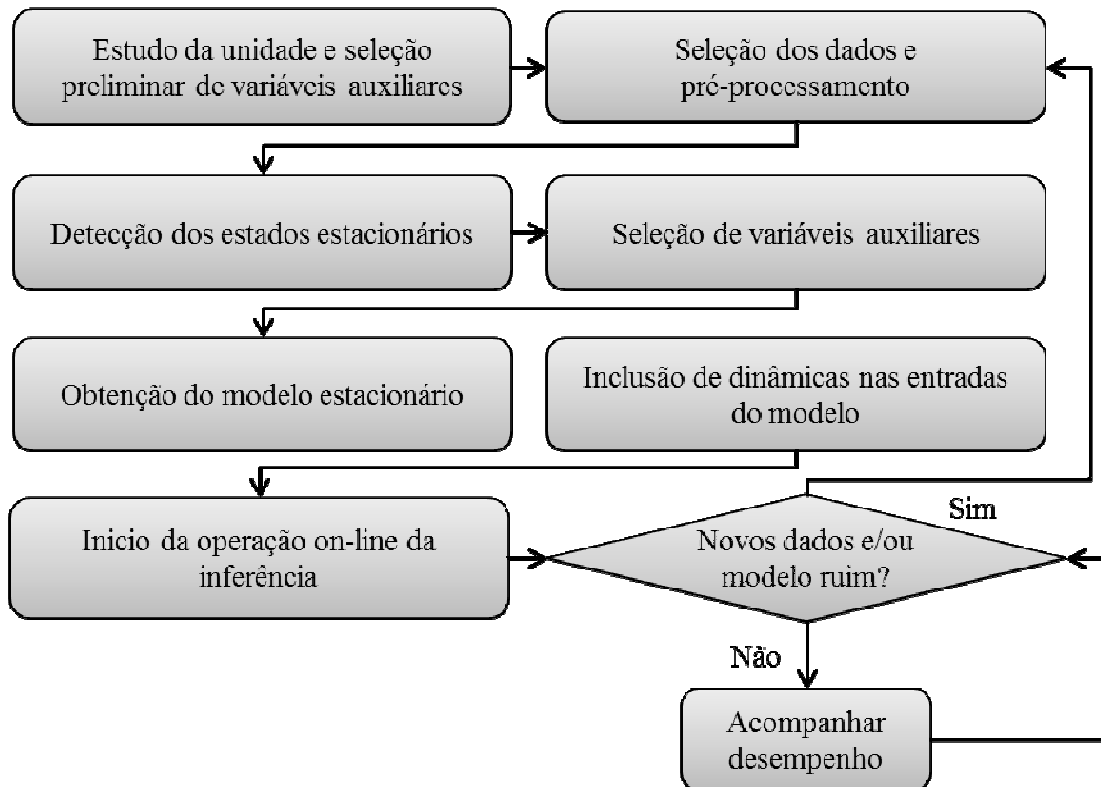


Figura 3.1: Diagrama esquemático da metodologia proposta.

Durante o estudo da unidade, uma listagem das candidatas a variáveis auxiliares deve ser feita. Esse trabalho geralmente é simples se houver auxílio de um operador experiente da unidade, pois esse tipo de profissional consegue distinguir rapidamente as variáveis importantes das irrelevantes. Em caso de dúvida, a variável deve ser incluída na lista, já que haverá uma etapa posterior de refinamento, onde apenas as variáveis estatisticamente significantes serão selecionadas, eliminando a redundância de informação.

3.2 Seleção dos dados e pré-processamento

Nessa etapa será feita a aquisição inicial dos dados e o seu pré-processamento. Inicialmente, é interessante coletar o maior período possível de dados existente, pois quanto maior a quantidade de informação, melhor tende a ser o resultado da inferência. Porém, é importante ter em mente que diversos equipamentos têm suas características alteradas com o tempo, como trocadores de calor e caldeiras. Neste caso a coleta de dados muito antigos pode acarretar numa perda de qualidade do modelo.

Os dados brutos obtidos devem ser trabalhados para que sejam removidas informações que não representam a operação normal da planta de processo. O conhecimento obtido previamente sobre as variáveis será valioso nessa etapa. Cada variável deve ser analisada para

que sejam removidos períodos em que seus valores estejam fora da faixa normal. Valores de vazões negativos, indicações congeladas em fim ou início de escala ou com variação além do que seria razoável ocorrer fisicamente no processo devem ser eliminados nesse momento.

A próxima etapa de tratamento é a eliminação dos *outliers* mais difíceis de serem detectados. Para isso é necessário o auxílio de um algoritmo, como os apresentados na revisão bibliográfica. Nesta metodologia, foi utilizado o método 3σ em janela. Esse algoritmo é semelhante ao 3σ original, bastante difundido na literatura (Fortuna, Graziani *et al.*, 2007). A diferença está em se utilizar uma janela móvel para o cálculo do desvio-padrão e média, em vez de se fazer esse cálculo para todos os dados. Têm-se então um grau de liberdade que é o tamanho dessa janela. Esse valor deve ser selecionado por tentativa e erro de forma a eliminar apenas uma fração das amostras correspondente aos dados espúrios. Geralmente também são eliminadas amostras corretas, porém se a quantidade total for baixa, não deve haver prejuízo para o resultado final. É importante fazer a interpolação dos dados por esse método, caso contrário as etapas de detecção de estacionários e seleção de variáveis podem ser prejudicadas. No próximo capítulo será mostrado em exemplo prático, com dados reais, o método apresentado.

Quando se trabalha com cromatógrafos, é importante fazer a remoção do tempo morto antes das etapas de seleção de variáveis auxiliares e modelagem. Cromatógrafos possuem um grande tempo morto inerente a sua medida, devido ao tempo necessário para coleta das amostras e análise na coluna cromatográfica. Esse atraso é prejudicial tanto para controle quanto para modelagem. O resultado do modelo é melhorado significativamente se eliminarmos esse tempo morto. O método utilizado para detecção do valor do tempo morto consiste na modelagem dos dados para diversos valores de deslocamento do vetor da variável de interesse. O que resultar no melhor modelo será o tempo morto estimado do processo (Ljung, 2002).

3.3 Detecção dos estados estacionários

Com os dados tratados, sem *outliers*, parte-se para a detecção dos estados estacionários. Essa etapa é necessária para a obtenção de um modelo estacionário, embora na literatura costuma-se omiti-la, pois a maioria dos trabalhos utilizam modelos estáticos ou dinâmicos com modelagem direta. O método a ser utilizado baseia-se na estimação da derivada do sinal após suavização. Mais informações sobre seu funcionamento estão dadas na revisão bibliográfica e em trabalhos apresentados nas referências (Duarte, Mejía *et al.*, 2010; Mejía, Farias *et al.*, 2010).

O algoritmo de detecção determina os períodos em que cada uma das variáveis está em estado estacionário. Para se chegar ao conjunto final de dados em estado estacionário é necessária mais uma etapa de tratamento, onde são escolhidos apenas os períodos em que todas as variáveis estão em estado estacionário e estes períodos, então, devem ser transformados em um ponto cada. É calculada a média de cada período e esse valor é usado como correspondente ao estado estacionário.

Durante essa etapa, é importante avaliar a quantidade de pontos resultantes do método. Se o valor for muito pequeno, poderá prejudicar a modelagem. Geralmente, para cada parâmetro do modelo é recomendado que se utilize pelo menos cinco pontos na obtenção do modelo. Além disso, serão necessários mais pontos para fazer a validação do modelo. Caso o número de pontos do conjunto final for pequeno, pode-se alterar os parâmetros do algoritmo de detecção de estacionários para se obter um número maior. Se essa ação for tomada, o resultado deve ser analisado visualmente com ajuda de um operador ou engenheiro com conhecimento do processo para garantir a qualidade do estacionário detectado.

3.4 Seleção de variáveis auxiliares

Em função do número de variáveis ser pequeno nos modelos de processos químicos, especialmente quando comparado a outras áreas como a quimiometria, por exemplo, a busca exaustiva se torna um método viável na seleção de variáveis auxiliares. Aqui ela será usada como principal técnica, embora no estudo de caso também serão realizados testes com o método *stepwise* para comparação.

Com o objetivo de acelerar o método de busca exaustiva, é sugerida nesse trabalho uma etapa preliminar para determinação da dimensão do problema a ser modelado. Essa etapa consiste na aplicação de técnicas de redução de dimensionalidade no conjunto de variáveis selecionadas previamente para se obter o número ótimo de variáveis para a modelagem do problema. Com a análise dos componentes principais em mãos, podemos fazer a busca exaustiva com o número de variáveis final fixado. Isso permite uma redução de esforço computacional sem que haja perda de qualidade da etapa de seleção de variáveis.

3.5 Obtenção do modelo estacionário

Com as variáveis auxiliares selecionadas e a matriz de dados estacionários, parte-se para a etapa de modelagem. Sugere-se inicialmente uma análise do gráfico de dispersão entre as variáveis auxiliares e a variável modelada para verificação de linearidade e correlação.

Depois disso são feitos dois modelos, um linear baseado em PLS e outro quadrático baseado em QPLS. Baseado na qualidade desses modelos será escolhido o que deve seguir para a próxima etapa. Caso a diferença de desempenho em relação entre o linear e o não-linear for pequena, deve-se preferir o linear, caso contrário, segue-se com o não-linear. Um gráfico de dispersão entre o valor modelado e o real também auxilia na identificação de não-linearidades, como será mostrado no estudo de caso no próximo capítulo.

Deve ser usada validação cruzada para fazer comparação entre os modelos. A técnica escolhida para separar o conjunto de dados em validação e modelagem foi o *y-rank*. O modelo é treinado com o conjunto de modelagem e o índice de desempenho calculado com o conjunto de validação. O índice a ser utilizado é o PRESS, apresentado no capítulo anterior, por ser o mais usada em artigos apresentados na literatura.

3.6 Inclusão de dinâmicas nas entradas do modelo

Tendo o modelo estático, deve-se determinar as dinâmicas de cada uma das variáveis auxiliares. Para isso, será assumida uma dinâmica de primeira ordem para cada uma, cuja constante de tempo será obtida por otimização. Cabe ressaltar que os tempos mortos entre as variáveis foram previamente determinados. A função objetivo a ser minimizada é mostrada na Equação 3.1. Para solução, foi utilizado o método dos poliedros flexíveis.

$$\text{Minimizar: } f(\tau_i) = \left(\sum_{i=1}^p \hat{\beta} \left(\mathcal{L}^{-1} \left(\frac{1}{\tau_i s + 1} \mathcal{L}(x_i) \right) \right) - y \right)^2 \quad (3.1)$$

$$\text{Sujeito a: } \tau_i > 0 \quad i = 1, 2, \dots, p$$

Onde τ é o vetor com o valor das dinâmicas, $\hat{\beta}$ é o vetor com os coeficientes estimados do modelo linear ajustado anteriormente, x_i é o vetor com a variável de entrada i do modelo, y é o vetor com a variável a ser modelada, p é o número de variáveis de entrada do modelo e \mathcal{L} representa a transformada de Laplace.

Nessa etapa, são utilizados dados dinâmicos de processo. O método *y-rank* utilizado anteriormente não se aplica a esse tipo de dados, sendo necessária outra forma de separação do conjunto de modelagem e validação. Nesse caso, os dados podem ser separados arbitrariamente, procurando-se obter dois conjuntos relevantes para cada um. Caso haja dados com períodos de operação distintos, pode-se utilizar um período para modelagem e outro para validação.

Por ser um problema de otimização não convexo, o valor inicial de dinâmica utilizado tem forte influência no resultado final. Para obtenção do melhor resultado, dez valores de dinâmica foram arbitrados inicialmente. O caso que resultou em menor resíduo no conjunto de validação foi escolhido como valor final da dinâmica. Para solução, escolheu-se o algoritmo *Simplex* (Edgar e Himmelblau, 1988), também conhecido como método dos poliedros flexíveis.

Capítulo 4

Estudo de caso – UPGN

Neste capítulo a metodologia desenvolvida e apresentada anteriormente será aplicada a uma unidade de processamento de gás natural (UPGN) cujo processo está descrito a seguir.

4.1 Descrição da UPGN

4.1.1 Sobre o gás natural e seu processamento

Gás natural é o nome dado aos hidrocarbonetos que são obtidos na forma gasosa durante a exploração de reservas geológicas naturais de petróleo. A composição desse gás é majoritariamente de metano, porém estão presentes em quantidades significativas hidrocarbonetos de até cinco carbonos junto com dióxido de carbono, nitrogênio e sulfeto de hidrogênio, como mostra a Tabela 4.1.

Tabela 4.1: Composição típica do gás natural (Mokhatab, Poe *et al.*, 2006).

| Nome | Formula | Volume (%) |
|-----------------------|--------------------------------|------------|
| Metano | CH ₄ | >85 |
| Etano | C ₂ H ₆ | 3-8 |
| Propano | C ₃ H ₈ | 1-2 |
| Butano | C ₄ H ₁₀ | <1 |
| Pentano | C ₅ H ₁₂ | <1 |
| Dióxido de carbono | CO ₂ | 1-2 |
| Sulfeto de hidrogênio | H ₂ S | <1 |
| Nitrogênio | N ₂ | 1-5 |
| Hélio | He | <0.5 |

No Brasil, o gás natural apresenta um teor maior de hidrocarbonetos mais pesados, sendo que propano e butano somados representam, em média, 5,8 % do volume total do gás (Vaz, Maia *et al.*, 2008).

Para ser comercializado, o gás natural precisa ser especificado de acordo com a exigência da ANP, sendo hoje o seu processamento fundamental e indispensável para aproveitamento em todos os setores da indústria. Nesse processamento, o metano é purificado através da remoção de água, H_2S e hidrocarbonetos mais pesados. Após essa etapa ele passa a ser chamado, em geral, de gás combustível ou gás residual. São obtidos também o GLP – produto de maior valor agregado – e a gasolina natural, ou fração C_5+ , que normalmente é misturada com correntes de petróleo por não ter especificação bem definida. Recentemente se tem produzido o etano petroquímico, que é usado como matéria-prima na fabricação de polietilenos (Vaz, Maia *et al.*, 2008).

Um exemplo típico do processo de tratamento de gás natural em terra está mostrado na Figura 4.1. Às vezes, pouco processamento é realizado, mas no mínimo é necessária uma planta para remoção de impurezas, água e excesso de hidrocarbonetos líquidos e para controlar a pressão do gás comercializado (Mokhatab, Poe *et al.*, 2006).

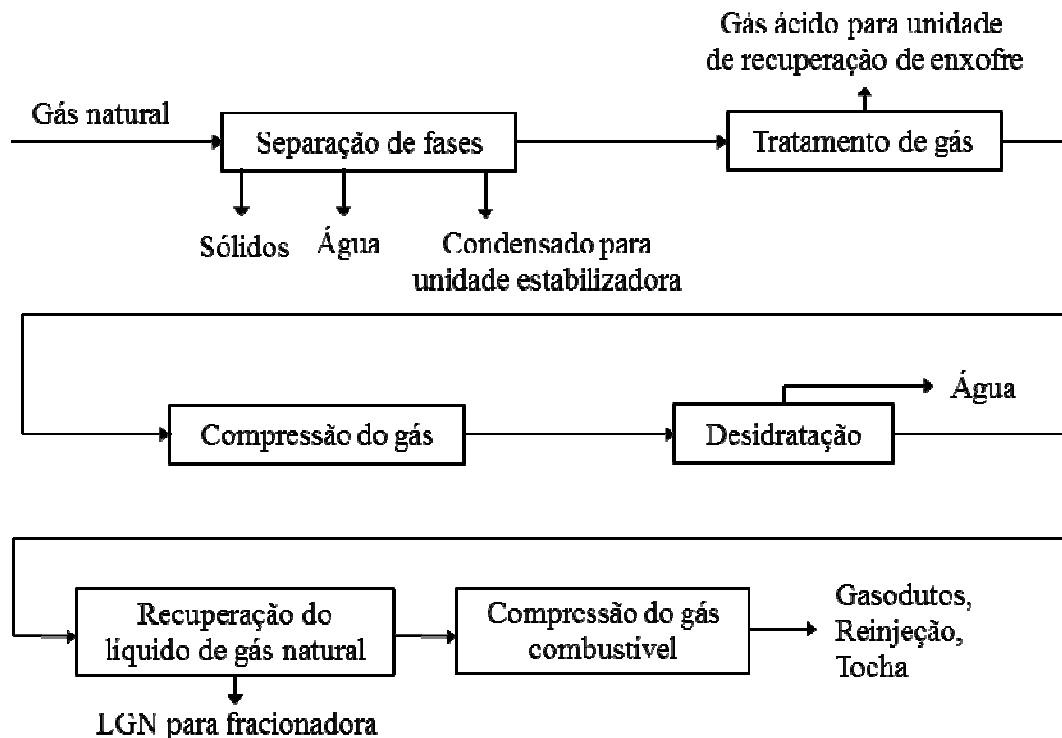


Figura 4.1: Típico processo de tratamento de gás em terra simplificado (Mokhatab, Poe *et al.*, 2006).

O ponto mais importante de uma unidade de processamento de gás natural é o sistema de refrigeração, responsável pela liquefação dos componentes pesados do gás natural. O processo termodinâmico escolhido nessa etapa define o tipo de unidade a ser utilizada (Vaz, Maia *et al.*). No Brasil, encontram-se principalmente unidades de turbo-expansão e absorção refrigerada. O processo de turbo-expansão é atualmente o mais eficiente e mais utilizado, consistindo basicamente na refrigeração e expansão do gás em uma turbina para liquefação

dos hidrocarbonetos mais pesados, podendo-se atingir temperaturas abaixo de $-95\text{ }^{\circ}\text{C}$. O processo de absorção refrigerada consiste primeiramente na condensação dos componentes mais pesados por refrigeração e, após essa etapa, na absorção do restante através da lavagem do gás natural com um solvente em contracorrente (Vaz, Maia *et al.*, 2008; Marcon, Campos *et al.*, 2010).

A escolha do tipo de processo a ser utilizado depende principalmente da composição do gás natural disponível, características do reservatório e do mercado consumidor. Devem ser observados, principalmente: a qualidade necessária para os produtos, tipo de produto desejado, duração das reservas de gás natural existentes, vazão de gás natural disponível e características do consumidor, como seu porte e proximidade (Vaz, Maia *et al.*, 2008).

4.1.2 Descrição da unidade trabalhada

A unidade de processamento de gás natural a ser estudada é do tipo turbo-expansão. Por tratar um gás natural com riqueza elevada (alto de teor de hidrocarbonetos com mais de dois carbonos), essa unidade utiliza um ciclo de refrigeração a propano para liquefazer as frações mais pesadas sem comprometer o desempenho da turbina (Vaz, Maia *et al.*, 2008). Ela consiste resumidamente de duas colunas de destilação, onde na primeira é separado o metano e o etano como produto de topo e na segunda é obtido no topo propano e butanos (GLP) e no fundo a gasolina natural.

O gás natural primeiramente é dessulfurizado e desidratado antes de ser refrigerado para ser enviado para a unidade de separação propriamente dita. Ao chegar à unidade, sua temperatura é diminuída ao passar por diversos permutadores de calor para condensar os hidrocarbonetos pesados. Inicialmente utiliza-se o propano refrigerante e em seguida os produtos efluentes do topo da primeira torre de fracionamento (T-01). O líquido que é condensado em cada uma dessas etapas de resfriamento é separado e alimentado nessa torre em pontos diferentes com o objetivo de otimizar a separação, economizando energia. O produto de topo é comprimido e comercializado ou utilizado nos fornos da própria unidade. O produto de fundo alimenta a segunda coluna de destilação.

A segunda torre tem por objetivo obter GLP como produto de topo, especificando-o para a comercialização e a gasolina natural no fundo, que será misturada a correntes de petróleo processado em outras unidades. Há um controle de temperatura em um prato sensível na seção de absorção da torre para aperfeiçoar a produção de GLP. O controle de pressão é feito através de uma válvula que atua no *by-pass* do condensador – conhecido como *hot by-pass*. A corrente de carga dessa coluna e do GLP produto são analisadas em um cromatógrafo gasoso em linha. Outros detalhes do processo podem ser vistos no fluxograma ilustrado na Figura 4.2.

O estudo de caso deste trabalho será feito na coluna fracionadora de GLP (T-02); dessa forma é importante entender quais são os requisitos para especificação deste produto. O GLP deve condensar facilmente sob pressão para transporte e armazenamento; e também deve evaporar com facilidade nas condições ambientes para queima. O primeiro é garantido pela

PVR (pressão de vapor Reid), que serve como indicação da pressão em que deve ser armazenado o produto e é uma medida indireta da presença de hidrocarbonetos leves. O segundo requisito é avaliado pelo teste de intemperismo, que mede a dificuldade de vaporização do GLP na pressão atmosférica e é uma medida indireta da presença de hidrocarbonetos pesados. Um terceiro requisito de qualidade é o teor de etano, que deve ser limitado para evitar o descolamento da chama durante seu uso no fogão.

4.2 Obtenção dos dados de planta

Seguindo a metodologia proposta, o projeto de uma inferência de composição começa com o estudo da unidade industrial em questão. Manuais, cópias das telas do sistema supervisório e P&Is foram analisados e utilizados para consulta durante o trabalho. Foi importante a colaboração da equipe de operação da planta e também dos engenheiros envolvidos para discussão. Um fluxograma do processo, apresentado na Figura 4.2, foi criado para aumentar o entendimento do processo e servir como uma melhor referência nas etapas posteriores do trabalho. A confecção desse fluxograma é bastante importante e facilita o trabalho, pois ele resume bem as partes importantes na elaboração das inferências, sem informações irrelevantes como acontece com os P&Is e telas do supervisório.

Na análise preliminar da unidade, são levantadas todas as variáveis que podem ter alguma relação com a composição que se deseja modelar. Para isso, discutiu-se com os operadores e engenheiros até se chegar às variáveis mostradas na Tabela 4.2.

Com a lista das variáveis em mãos, é feita a coleta de dados de planta para o maior período possível. Os dados geralmente estão disponíveis através de historiadores de processo em grandes bancos de dados, como ocorreu para este estudo de caso. Inicialmente, se quer ter o máximo de dados possíveis, embora um conjunto grande de dados não implique necessariamente um bom resultado. Os dados precisam ser representativos de todas as regiões de operação da unidade, representando o maior número de pontos de operação possíveis.

É importante que esses dados não tenham sofrido nenhuma alteração por algoritmos de compressão de dados, o que é muito comum na indústria. Dados comprimidos perdem seu valor estatístico e muitas das técnicas usadas no processo de obtenção de inferência não funcionam corretamente. É recomendável alterar a configuração do historiador para remover qualquer tipo de tratamento antes de obter os dados, caso haja tempo para esperar que um conjunto significativo seja gerado. Para este estudo de caso, a compressão da maioria das variáveis foi removida.

Em posse dos dados brutos, é necessário tratá-los de forma a manter apenas as informações significativas. Isso é feito na próxima etapa, o pré-tratamento de dados.

Tabela 4.2: Variáveis auxiliares pré-selecionadas.

| Variável | Descrição |
|-----------------|--|
| AI-01.C2 | Análise de C2 na carga da T-02 |
| FIC-03 | Vazão de gás para unidade |
| PIC-03 | Pressão do vaso de carga |
| TI-01 | Temperatura do gás no topo da torre T-01 |
| TI-02 | Temperatura do gás no P-07 para T-01 |
| PIC-01 | Indicador de pressão T-01 |
| FIC-01 | Vazão de refluxo T-01 |
| TI-03 | Temperatura no prato 36 da torre T-01 |
| TIC-01 | Indicador de temperatura - T-24001 |
| TI-04 | Temperatura no fundo da torre T-24001 |
| TI-05 | Produto de fundo da torre desetanizadora do refervedor P-24009 |
| TI-06 | Produto de fundo da torre desetanizadora T-24001 |
| TI-07 | Temperatura do produto de fundo da torre desetanizadora |
| TI-08 | Temperatura no prato 14 da torre T-24004 |
| PI-01 | Pressão da torre T-24002 debutanizadora |
| PDI-01 | Pressão diferencial na torre T-02 |
| TIC-02 | Indicador de temperatura - T-02 |
| TI-10 | Temperatura no fundo da T-24002 para P-24011 |
| TI-11 | Temperatura do produto no topo da torre desbutanizadora T-02 |
| PIC-02 | Indicador de pressão - V-24007 |
| FIC-02 | Indicador de fluxo - refluxo - T-02 |

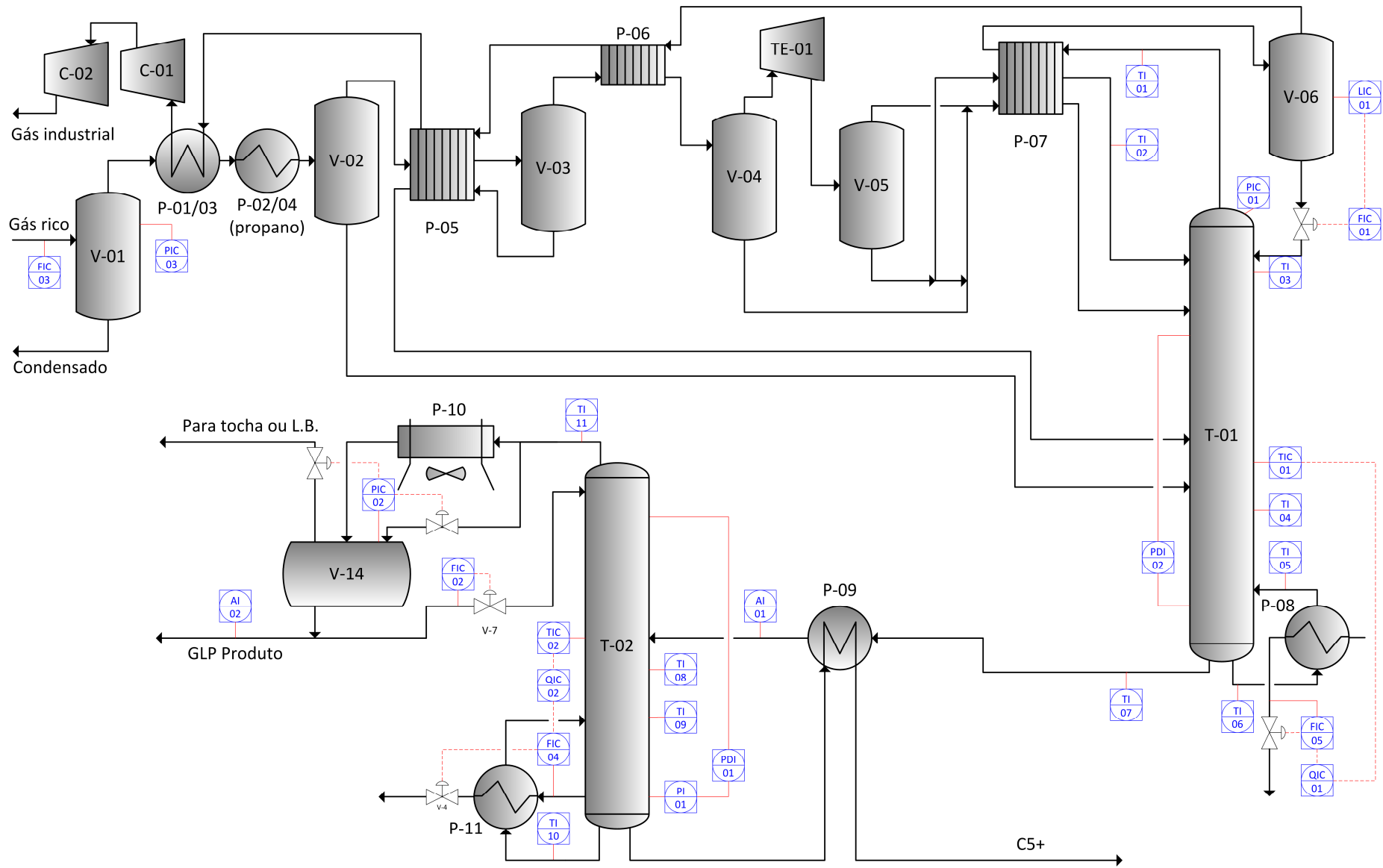


Figura 4.2: Fluxograma simplificado do processo.

4.3 Pré-tratamento dos dados de planta

Os dados obtidos devem ser tratados em etapas de forma a garantir sua qualidade e representatividade. Períodos em que a planta de processo encontra-se em parada, com algum sensor em falha, ou em alguma rotina diferente de operação, por exemplo, devem ser removidos. As variáveis pré-selecionadas devem ser avaliadas com mais rigor para que apenas aquelas que tenham importância real sejam mantidas. Para isso, utilizam-se métodos estatísticos. Na última etapa, é feita a detecção dos estados estacionários, para permitir a obtenção de um modelo estacionário a partir de dados de planta.

A Figura 4.3 mostra dados brutos típicos do analisador da UPGN em questão. Vê-se no gráfico que em vários momentos, a indicação congelou em algum valor e em outros apresentou variações muito além do esperado para um analisador baseado em cromatografia gasosa – ocasionada por uma falha no equipamento. Esses dados não podem ser usados diretamente, pois levariam a péssimos resultados.

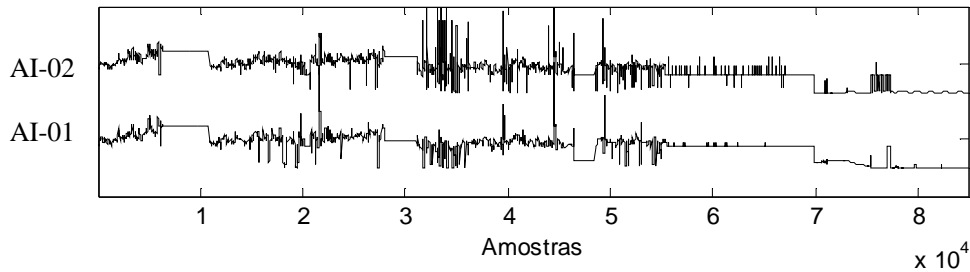


Figura 4.3: Dados coletados antes do tratamento.

4.3.1 Detecção de outliers

Inicialmente, os dados obtidos para cada variável apresentada na Tabela 4.2 foram avaliados para garantir que os valores estavam dentro de um intervalo válido, definido após o estudo da unidade em questão. Os dados que estavam fora desta faixa foram removidos. Desta forma é possível eliminar as paradas de planta – por exemplo, através do indicador de vazão da carga, que estará em zero – entre outros problemas de operação.

Depois da inspeção dos dados, aplicou-se o método de 3σ em janela para o tratamento de *outliers*. Utilizou-se a janela com 100 pontos, onde foram obtidos os melhores resultados. A Figura 4.4 mostra um exemplo gráfico do método utilizado. As duas linhas externas representam o desvio-padrão calculado em cada ponto e a linha interna é a série temporal da variável analisada. Valores que ficam fora das linhas externas são considerados *outliers* e são removidos e substituídos por novos pontos obtidos através de interpolação.

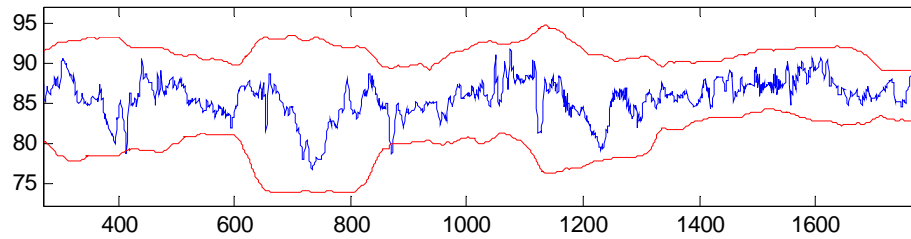


Figura 4.4: Método de remoção de outliers 3σ em janela.

4.3.2 Remoção do tempo morto

Como o vetor de dados a ser modelado é produto de análises de um cromatógrafo em linha, ele possui um grande atraso inerente ao processo de análise da corrente de gás produto da torre de destilação, além do atraso do processo em si. Para solucionar esse problema, os vetores provenientes de cromatógrafos foram deslocados em relação aos demais para que este atraso fosse removido.

O valor do atraso a ser usado foi definido a partir de inspeção. Um modelo linear foi criado para cada valor de atraso e o que apresentou melhor resultado foi usado. A Tabela 4.3 a seguir mostra o resultado destes testes.

Tabela 4.3: Determinação do tempo morto

| Avanço no tempo (em número de amostragens) | PRESS |
|--|-------------|
| 0 | 1498 |
| 1 | 1409 |
| 2 | 1325 |
| <u>3</u> | <u>1297</u> |
| 4 | 1322 |
| 5 | 1458 |
| 6 | 1519 |

Para esta modelagem foram usados os dados dinâmicos somente com os *outliers* removidos. A modelagem para cada caso seguiu a metodologia desenvolvida no trabalho, mas sem a detecção de estacionários, já que isso distorceria o resultado. Para este teste funcionar, os dados precisam ser contínuos e dinâmicos; a detecção de estacionários segmentaria as informações.

O melhor resultado obtido foi para um avanço de três amostragens. Como os dados utilizados estão amostrados a cada oito minutos, isso significa um tempo morto estimado para o cromatógrafo e o processo de aproximadamente vinte e quatro minutos.

4.3.3 Identificação dos estados estacionários

Com todos os dados tratados e os *outliers* removidos, foi utilizado o método de cálculo da derivada (Duarte, Mejía et al., 2010; Mejía, Farias et al., 2010) para a identificação dos períodos em que a planta se encontrava em estado estacionário.

A Figura 4.5 ilustra o resultado obtido pelo método de detecção para um pequeno período de dados e para apenas três variáveis da planta estudada. O gráfico mostra nas áreas com fundo hachurado o momento em que a variável está em estado estacionário e nas áreas com fundo branco, o transiente. O estacionário global é o momento em que todas as variáveis analisadas encontram-se em estacionário.

O trabalho de identificação foi realizado em todas as variáveis pré-selecionadas e em todo o período de dados obtidos. Para cada intervalo em estacionário, foi tomada a média a partir do momento do início do estacionário até o seu final e este vetor foi armazenado em uma tabela contendo todos os pontos estacionários. Em outras palavras, um período em que a unidade inteira ficou 20 minutos em estado estacionário, por exemplo, será traduzido em apenas um ponto na tabela, visto que todos os dados desse intervalo representam apenas um estado estacionário, ou ponto de operação.

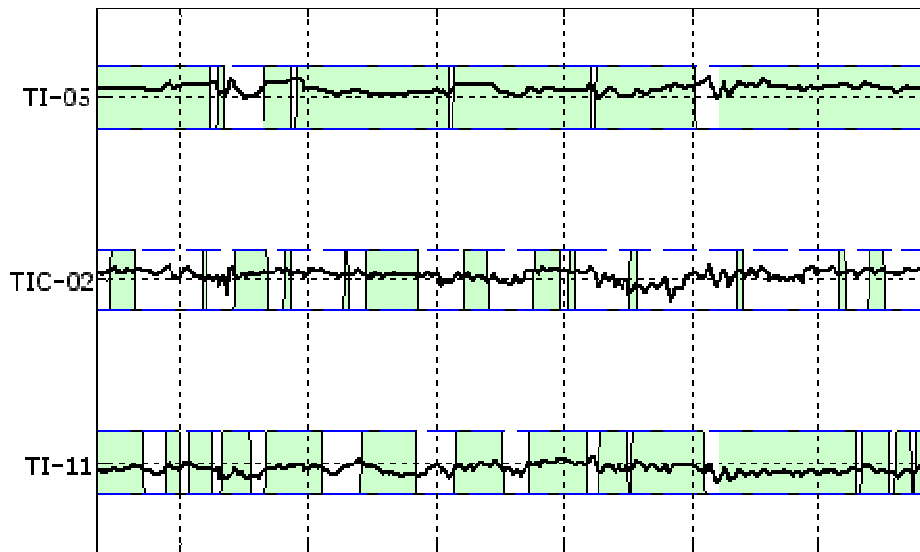


Figura 4.5: Exemplo do resultado da detecção de estados estacionários.

A Tabela 4.4 mostra uma comparação entre o número de pontos antes e depois desta etapa. Vê-se que há uma diminuição significativa no número de dados, mas, se a detecção de estados estacionários foi feita corretamente, a quantidade de informação relevante em cada um dos conjuntos é semelhante, já que redundância na informação foi reduzida.

Tabela 4.4: Resumo do conjunto de dados antes e após a identificação dos estados estacionários.

| | Conjunto inicial | Conjunto em E.E. |
|------------------|------------------|------------------|
| Número de pontos | 12541 | 124 |

4.3.4 Separação do conjunto de validação

Para validação cruzada do modelo é necessário um método de seleção dos conjuntos de teste e validação. Neste estudo de caso foi utilizado o y-rank, que abrange da melhor forma todas as regiões dos dados utilizados (Facchin, 2005).

Os dados estacionários obtidos são ordenados de forma crescente do valor do vetor y a ser modelado. Escolheu-se a relação 3 para 1 entre o tamanho do conjunto de modelagem e de validação por apresentar melhores resultados. Os dados são então varridos de uma ponta a outra e três pontos colocados no conjunto de modelagem, para, a seguir, um ser colocado no conjunto de validação. O resultado desse procedimento pode ser melhor entendido e visualizado através da Figura 4.6 que mostra os dois conjuntos em um mesmo gráfico.

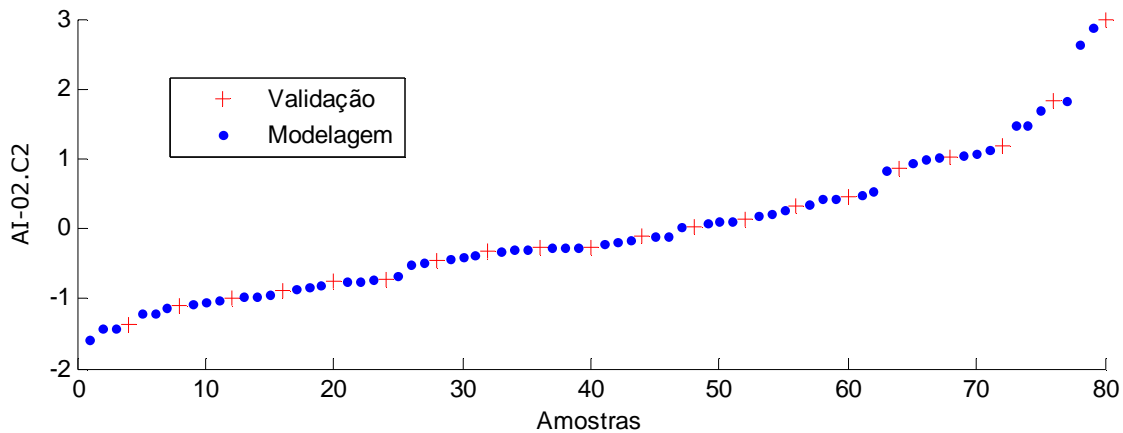


Figura 4.6: Distinção entre os pontos utilizados para validação e modelagem utilizando o método y-rank.

4.3.5 Seleção de variáveis auxiliares

Na Tabela 4.2 foram apresentadas as variáveis pré-selecionadas para o estudo de caso. Geralmente esse número é bastante grande, visto que por motivos de segurança as plantas de processos químicos possuem grande número de instrumentos de medição e malhas de controle. Isso é muito bom para quem desenvolve inferências do processo, porém a maior parte dessa informação é redundante e desnecessária. Para diminuir o número de variáveis usado na modelagem, uma segunda etapa de seleção de variáveis auxiliares é de grande importância. Quanto menor o número de variáveis, mais simples será o modelo e mais fácil a

sua manutenção e todas as etapas de desenvolvimento que se seguirão a partir daqui. Porém é preciso ter cuidado para não tornar o modelo simples demais.

Serão utilizados e comparados dois métodos de seleção de variáveis nessa etapa do trabalho. O primeiro deles é o método *stepwise*, que consiste num procedimento sistemático de adição e remoção de variáveis de acordo com sua relevância estatística para o modelo. O segundo será a busca exaustiva de variáveis, onde todas as combinações possíveis serão testadas e a melhor delas selecionada. Para diminuir a carga computacional requerida por este método, serão utilizadas três técnicas de redução de dimensionalidade para definir o número de variáveis ótimo para cada problema: PCA, PLS e QPLS. Ou seja, será estimada a dimensão – ou número de variáveis – que explicam de melhor forma a variabilidade da variável de saída. Com essa informação, a busca exaustiva terá fixo o número de variáveis a ser encontrado, sendo necessário apenas testar todas as combinações possíveis para esse número já determinado.

Além disso, o problema será avaliado em dois casos diferentes. Um deles utilizando a composição de carga da torre como variável auxiliar e o outro sem ela. Este teste é interessante, pois a composição de carga tem uma correlação intensa com a qualidade do produto de topo da torre, porém por ser uma variável medida através de um cromatógrafo em linha, possui um grande tempo morto e uma frequência de análise baixa. Além disso, tem baixa confiabilidade, pois apresenta problemas de funcionamento com elevada frequência, como foi verificado na prática ao se analisar os dados históricos da unidade em questão.

Na prática, não é interessante usá-la como variável auxiliar, visto que sua aplicação geralmente é em controle de processos, que requer rápido tempo de resposta e alta disponibilidade. Porém, a comparação entre os casos é mais uma forma de avaliar a qualidade da inferência que não utiliza a composição da carga no modelo.

Caso 1 (sem a composição da carga)

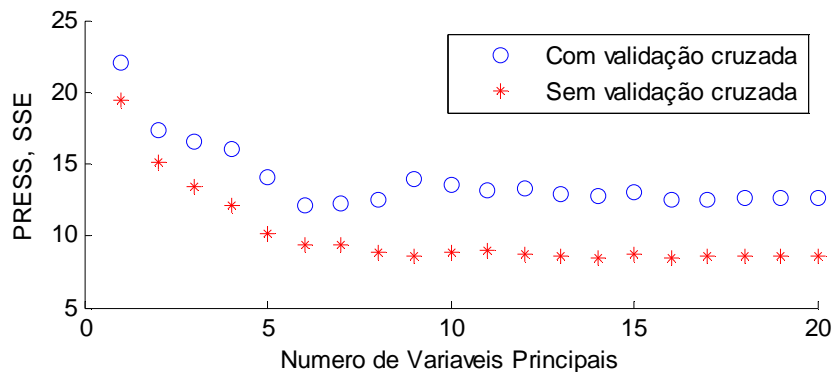
Os resultados das três técnicas para redução de dimensionalidade estão mostrados na Tabela 4.5. Como se vê, foram obtidos valores bastante diferentes para cada uma delas.

Para determinar esse valor, foram criados diversos modelos, variando-se o número de componentes principais (ou variáveis latentes, como geralmente são chamados no PLS) de um até o número total de variáveis. Cada um desses modelos foi desenvolvido com os dados estacionários previamente obtidos e validados com os conjuntos separados anteriormente. Ao se fazer a validação cruzada, é esperado que haja um valor ótimo intermediário que corresponda à verdadeira dimensão do problema. Para comparar os diversos modelos utilizou-se o resíduo no conjunto de validação do modelo (PRESS), tendo como melhor desempenho o modelo de dimensão seis baseado em PLS. O número obtido como resultado está coerente com os apresentados na literatura para problemas de inferências em colunas de destilação (Kano, Miyazaki *et al.*, 2000; Fortuna, Graziani *et al.*, 2005), o que reforça o resultado obtido.

Tabela 4.5: Determinação da dimensão do problema para o Caso 1.

| Método | Dimensão | PRESS mínimo |
|--------|----------|--------------|
| PLS | 6 | 12,19 |
| PCA | 20 | 12,65 |
| QPLS | 1 | 12,21 |

O resultado obtido para o PLS está ilustrado na Figura 4.7 e para o PCA na Figura 4.8. Na primeira se pode ver o valor obtido para cada um dos vinte modelos desenvolvidos com PLS, onde foi variada dimensão do problema, representado pelo número de variáveis principais. Nessa Figura se vê que a diferença é sutil entre a maioria dos valores acima de seis variáveis principais. Se a técnica de validação for alterada, por exemplo, possivelmente se obteria um resultado levemente diferente. Na segunda Figura, vê-se que o valor máximo de variáveis correspondeu ao valor ótimo, que não era esperado. Isso pode estar relacionado a problemas de escalonamento, que são possíveis no PCA, mas contornados no PLS por esse fazer utilizar a variável de saída y na redução de dimensão. Futuramente, com mais dados de planta, é interessante refazer esse estudo para verificar se esse resultado se mantém, ou é peculiar aos dados disponíveis atualmente.

**Figura 4.7:** Determinação do número de variáveis principais ótimo para o Caso 1 por PLS com e sem validação cruzada.

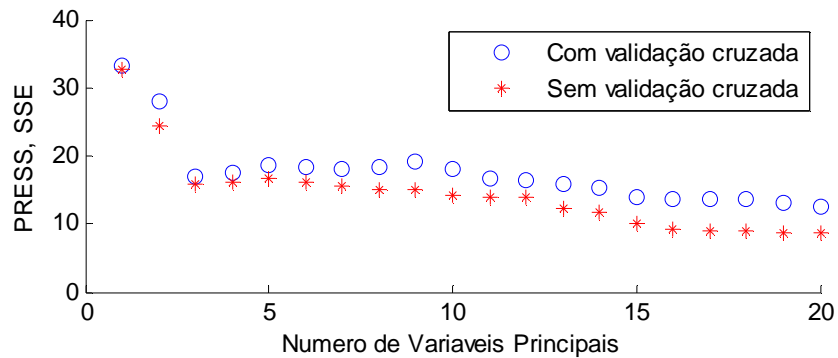


Figura 4.8: Determinação do número de variáveis principais ótimo para o Caso 1 por PCA com e sem validação cruzada.

Depois de determinada a dimensão, foi feita a seleção de variáveis pelos dois métodos apresentados anteriormente. Esses resultados estão mostrados na Tabela 4.6. As variáveis selecionadas estão dentro do esperado, já que todas elas têm sabidamente forte relação com a composição final do produto.

Foram selecionadas duas variáveis que pertencem à primeira torre de destilação, que separa uma corrente de C1 e C2 de outra de C3 e mais pesados, que é carga da segunda torre. É importante lembrar que nesta UPGN a primeira das duas torres tem a capacidade de remover o etano contido na carga, visto que na temperatura e pressão de operação da segunda torre, ele não se liquefaz e sai todo como produto de topo; isso explica a relevância dessas duas variáveis. As outras variáveis são tipicamente utilizadas para controle da pureza do produto, como é o caso da temperatura de topo, vazão de refluxo e pressão de operação da coluna.

Tabela 4.6: Variáveis selecionadas para o Caso 1.

| Método | Variáveis |
|---------------------|--|
| Busca exaustiva | FC-02, TI-11, TC-02, PI-01, TI-04 e TI-01 |
| Stepwise Regression | PC-02, TI-11, PDI-01, PI-01, TI-06, TI-05, FC-01, TI-02, TI-01 |

A Figura 4.9 mostra como estão correlacionadas as seis variáveis selecionadas pela busca exaustiva com a composição do GLP a ser modelada. Vê-se que algumas das variáveis aparentemente não têm correlação significativa com a composição de topo, porém por se tratar de um problema multivariável, a interação entre mais de uma variável pode mascarar informações relevantes. Por isso é importante a utilização de técnicas multivariáveis na seleção das variáveis. A variável TIC-02 representa o prato sensível da coluna de acordo com o projeto, ela apresente correlação quase nula, o levaria a descartá-la numa análise mais simplória.

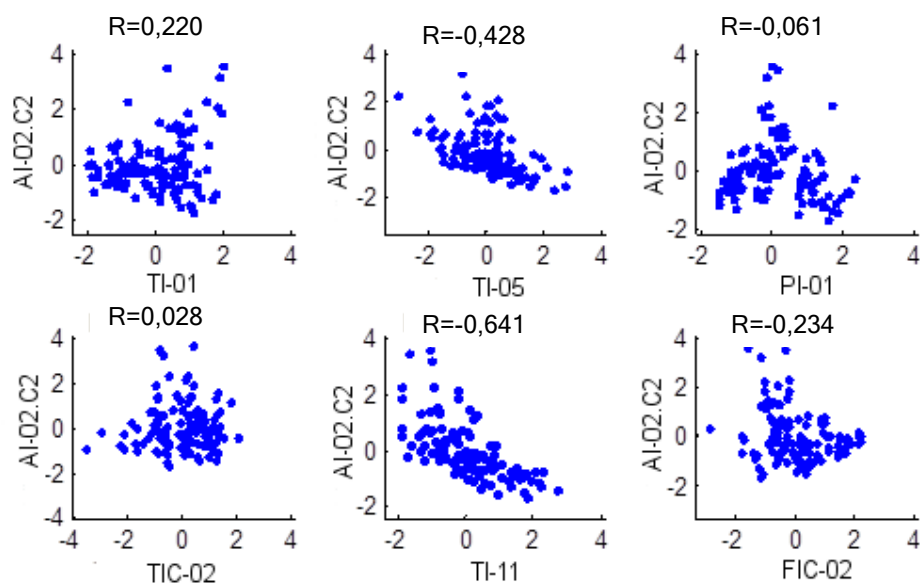


Figura 4.9: Gráfico de dispersão entre as variáveis selecionadas por busca exaustiva e a variável modelada.

Caso 2 (com a composição da carga)

Neste segundo caso, será analisado o mesmo problema com a adição da composição de carga como variável auxiliar. O procedimento realizado é análogo ao feito para o primeiro caso. Na Tabela 4.7 estão mostrados os resultados para a análise de dimensão do problema, como vemos, os valores obtidos são semelhantes aos do Caso 1.

Tabela 4.7: Determinação da dimensão do problema para o Caso 2.

| Método | Dimensão | PRESS mínimo |
|--------|----------|--------------|
| PCA | 16 | 1,72 |
| PLS | 7 | 1,58 |
| QPLS | 1 | 1,99 |

Na Figura 4.10 está mostrada a variação do resíduo no conjunto de validação com a adição de variáveis latentes para o modelo PLS. O gráfico obtido está coerente com o apresentado na literatura (Warne, Prasad *et al.*, 2004). Novamente se vê uma diferença pequena entre os valores do resíduo para cinco variáveis ou mais, o que mostra a sensibilidade existente na definição da dimensão do problema.

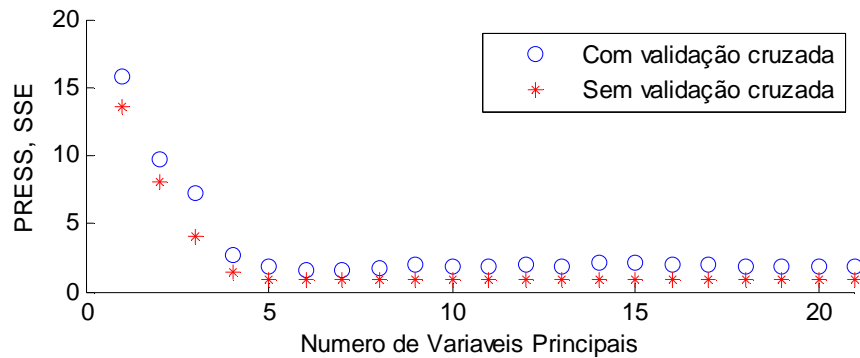


Figura 4.10: Determinação do número de variáveis principais ótimo para o Caso 2 por PLS com e sem validação cruzada.

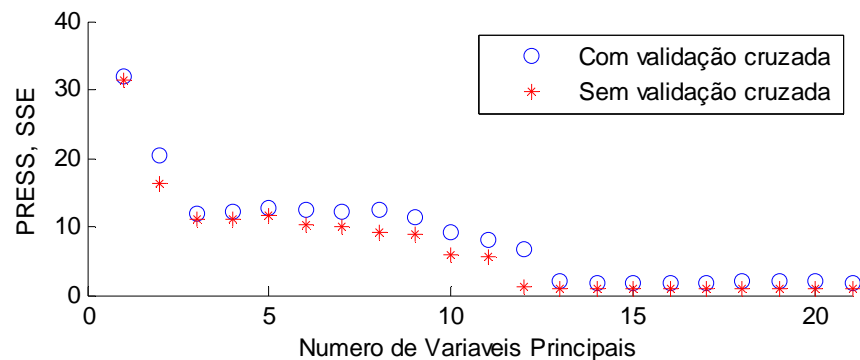


Figura 4.11: Determinação do número de variáveis principais ótimo para o Caso 2 por PCA com e sem validação cruzada.

As variáveis selecionadas para o Caso 2 estão mostradas na Tabela 4.8. Como já era esperado, a composição da carga foi selecionada pelos dois métodos. Ela apresenta uma forte correlação com a pureza do produto de topo e melhora significativamente o resultado do modelo. As outras variáveis selecionadas são semelhantes às escolhidas para o Caso 1 e estão dentro do esperado.

Ao se comparar as demais variáveis, vemos que o PI-01 foi selecionada apenas no Caso 1, indicando que a presença do AI-01 torna essa variável desnecessária. O TI-11 foi selecionado para todos os casos, indicando que a temperatura de topo é a temperatura melhor correlacionada com a composição do GLP.

Para o Caso 1 com *Stepwise Regression*, foram selecionadas duas medidas de pressão para a mesma torre, PC-02 e PI-01, além da diferença de pressão, PDI-01. Nesse caso, porém não foi selecionada a vazão de refluxo, FC-02. Esse resultado é incoerente e indica problemas na utilização desse método.

O Caso 1 com Busca Exaustiva foi o único a não selecionar os TI-05 e TI-06, porém foi o único a selecionar os TI-04 e TC-02. Como essas variáveis são próximas, isso mostra

que suas indicações são redundantes, podendo se utilizar um conjunto ou outro com bons resultados.

Tabela 4.8: Variáveis selecionadas para o Caso 2.

| Método | Variáveis |
|---------------------|--|
| Busca exaustiva | AI-01.C2, PC-01, TI-03, TI-05, TI-06, TI-11, FC-02 |
| Stepwise Regression | AI-01.C2, TI-05, TI-06, TI-11, FC-02 |

A Figura 4.12 mostra que a fração de etano na carga da coluna é a variável com maior correlação com a variável primária, e não se poderia esperar resultado diferente. Essa análise também permite concluir visualmente a relação fortemente linear entre essas duas variáveis. Já para as outras variáveis, a relação não é tão clara. As temperaturas selecionadas tem uma leve correlação não-linear comum de ser observada na relação temperatura de pratos com composição.

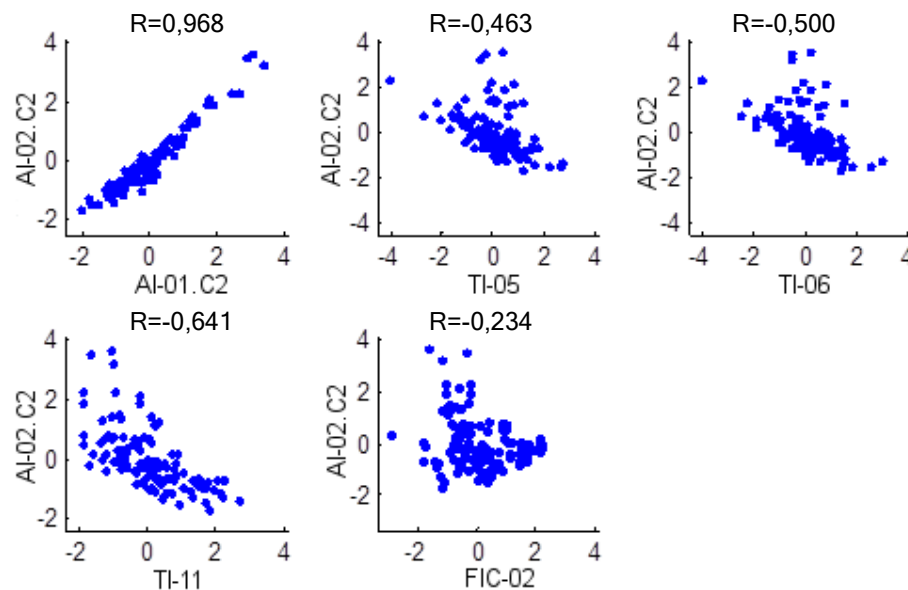


Figura 4.12: Gráfico de dispersão entre as variáveis selecionadas por busca exaustiva e a variável modelada.

4.3.6 Tempo gasto na seleção de variáveis

O tempo necessário para cada método de seleção de variáveis está mostrado na Tabela 4.9. Foram usadas vinte e uma variáveis para os testes e nelas aplicadas os métodos de seleção. O conjunto foi testado antes e depois da detecção de estados estacionários, ou seja, com 124 e 12538 pontos. O número de variáveis a ser encontrados pela busca exaustiva foi fixado em seis. O resultado mostra que mesmo para um conjunto grande de pontos, o tempo de execução da busca exaustiva foi razoável, permitindo sua utilização. Os testes foram feitos

utilizando um notebook com processador Intel® Core i5 2430M 2,4GHz e 4 GB de memória RAM.

Tabela 4.9: Comparação do tempo gasto para seleção de variáveis.

| Método | Número de pontos | Número de variáveis | Tempo (s) |
|----------------------------|------------------|---------------------|-----------|
| Busca exaustiva | 124 | 21 | 9 |
| <i>Stepwise regression</i> | 124 | 21 | 0,02 |
| Busca exaustiva | 12538 | 21 | 317 |
| <i>Stepwise regression</i> | 12538 | 21 | 0,8 |

4.4 Escolha da estrutura do modelo e modelagem

Nessa etapa será avaliada a necessidade de um modelo não-linear para o caso em questão, será apresentado e ilustrado o método utilizado na validação do modelo e comparados três tipos de modelos.

4.4.1 Não-linearidade

Para avaliar a não linearidade dos modelos, será feita uma comparação entre o resultado da modelagem com técnicas lineares e não-lineares (neste estudo de caso serão usados PLS e QPLS). Caso o desempenho dos dois seja semelhante, será dada preferência para o uso do modelo linear. Caso contrário, o modelo não-linear será utilizado nas etapas subsequentes.

4.4.2 Escolha do modelo

Para esta modelagem foram testados três tipos de modelos, dois lineares – baseados em PCA e PLS – e um não-linear – baseado em QPLS. Todos esses modelos utilizam redução de dimensionalidade para melhorar o seu resultado e evitar problemas numéricos resultantes de variáveis altamente correlacionadas (Wold, Sjostrom *et al.*, 2001).

O que é preciso definir nessa etapa é o número de variáveis principais usadas pelo algoritmo de cálculo de cada um dos métodos. Para isso, obtêm-se diversos modelos variando-se o número de variáveis latentes desde um até o número de variáveis auxiliares total selecionado – que é o máximo possível, o que representa um modelo sem redução de dimensionalidade, ou seja, uma regressão linear multivariável simples (MLR) para os modelos lineares. O trabalho é basicamente o mesmo feito para a determinação da dimensão do problema, mas desta vez não irá se usar todas as variáveis e sim somente as selecionadas anteriormente.

Com os conjuntos de dados de modelagem e validação em mãos, se utiliza o primeiro deles para calcular o modelo e o segundo para avaliar o resultado. Dessa forma se consegue garantir que não se está utilizando um modelo mais complexo do que o necessário,

modelando-se informações que são apenas ruídos do processo de medição. Como já se utilizou o PLS para determinação do número de variáveis necessárias para o problema, não são esperados ganhos significativos ao se usar modelos baseados em PCA e PLS. Pode ser interessante fazer somente uma regressão linear multivariável por simplicidade nesta etapa.

Caso 1 (sem a composição da carga)

Os resultados das modelagens e da determinação do número de variáveis principais estão mostrados na Tabela 4.10. Como era esperado, o método de seleção por busca exaustiva obteve um melhor resultado do que o método *stepwise*, mesmo com o primeiro utilizando menos variáveis auxiliares. O modelo quadrático também se mostrou levemente melhor do que os modelos lineares para este caso, mostrando que há algum grau de não-linearidade no problema.

O número de variáveis principais selecionadas foi um pouco menor do que o número de variáveis auxiliares utilizado, o que pode acontecer, visto que agora o número de variáveis testadas é diferente e as combinações possíveis são menores. Além disso, na determinação do número de variáveis a ser utilizado pela busca exaustiva, a diferença é sutil entre alguns números e pode mudar um pouco ao se utilizar conjuntos de dados diferentes, metodologias distintas ou quando se remove ou acrescenta alguma variável secundária.

Tabela 4.10: Resultado para variação do número de variáveis latentes para o Caso 1.

| | Número de variáveis | Número de componentes principais | | | PRESS | | |
|----------------------------|---------------------|----------------------------------|-----|------|-------|-------|-------|
| | | PCA | PLS | QPLS | PCA | PLS | QPLS |
| Busca exaustiva | 6 | 6 | 4 | 1 | 10,58 | 10,58 | 9,57 |
| Stepwise regression | 9 | 9 | 8 | 1 | 11,70 | 11,50 | 10,80 |

Uma forma visual de avaliar os modelos estáticos está nas Figuras 4.13 e 4.14. A primeira mostra os valores preditos e medidos juntos em um gráfico tendo no eixo das abcissas o número da amostragem. Uma melhora no resultado é percebida quando se aumenta o número de variáveis latentes (esquerda para direita, cima para baixo), porém é difícil chegar a uma conclusão concreta somente utilizando gráficos, por isso é necessária a utilização de índices mais conclusivos, como o PRESS.

A segunda Figura mostra um gráfico dos valores medidos *versus* preditos. Também é possível avaliar a qualidade dos modelos através desse gráfico, porém sua característica mais interessante é facilitar a visualização de não-linearidades presentes. Como se pode ver nos gráficos, os pontos apresentam o certo desvio da linha $f(x) = x$, o que é um indício de que para esse caso o uso de um modelo não-linear poderia melhorar o resultado.

Comparando-se os resultados da Tabela 4.10 com os obtidos para os modelos com todas as variáveis, mostrados na Tabela 4.5, vê-se uma diminuição do PRESS para o caso com menos variáveis. Para o Caso 1 com PLS usando vinte variáveis, o PRESS mínimo foi de 12,19, enquanto que se usando seis variáveis, seu valor foi de 10,58. O mesmo comportamento se observa para o Caso 2, que será apresentado a seguir. O PRESS mínimo com vinte e uma variáveis foi de 1,58, enquanto que para seis variáveis, o PRESS mínimo foi de 1,04, como mostram as Tabelas 4.7 e 4.11, respectivamente. Esses resultados mostram mais uma vantagem de se diminuir o número de variáveis totais do modelo.

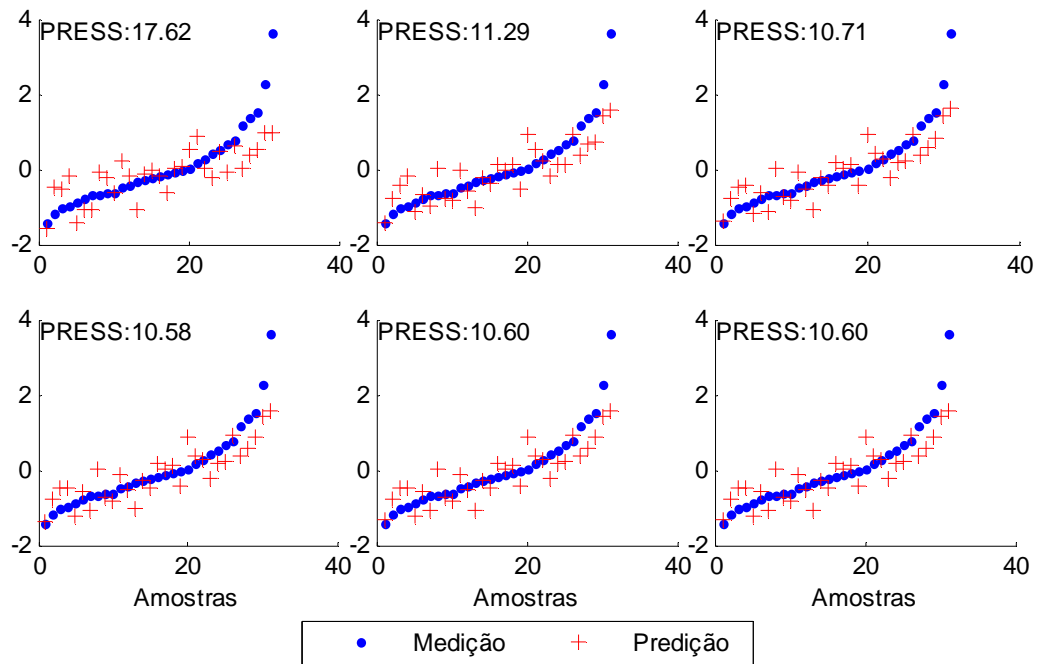


Figura 4.13: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até seis (baixo dir.) no modelo PLS com busca exaustiva para o Caso 1.

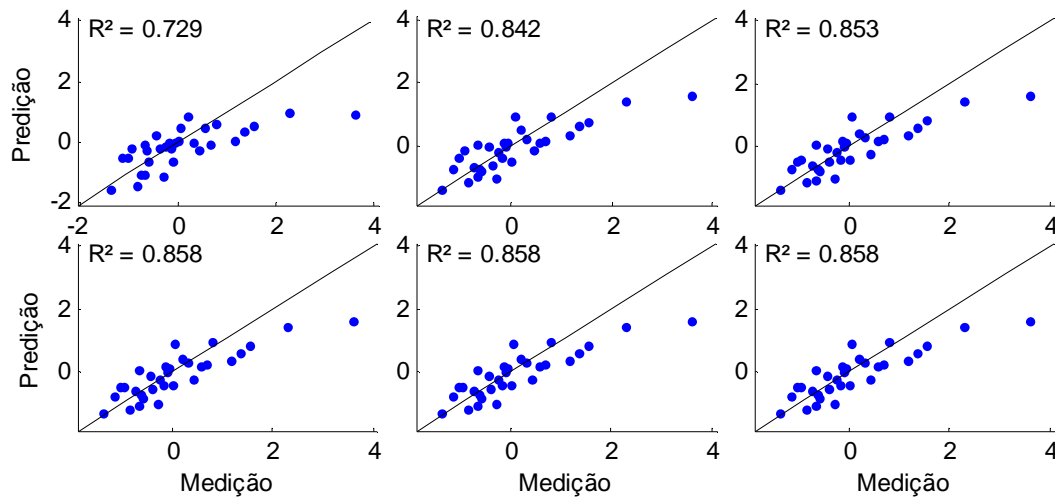


Figura 4.14: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até seis (baixo dir.) no modelo PLS com busca exaustiva para o Caso 1.

As Figuras 4.15 e 4.16 são análogas, mas para o modelo QPLS. Percebe-se que a resultado é levemente melhor, porém ainda assim não se conseguiu uma correlação muito boa na região de valores mais elevados. Isso indica que há outro tipo de não linearidade, provavelmente mais intensa, na forma exponencial.

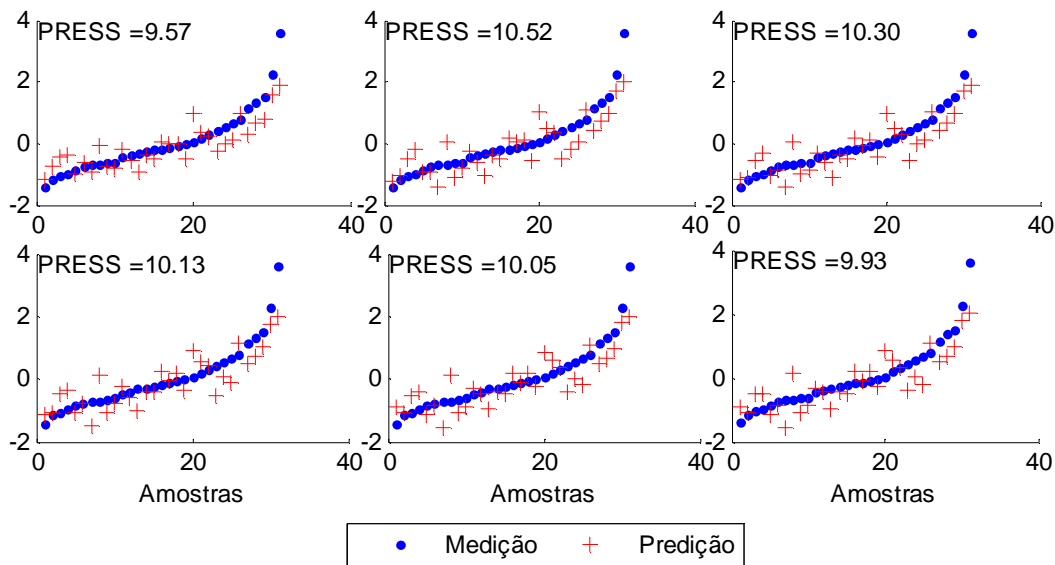


Figura 4.15: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até seis (baixo dir.) no modelo QPLS com busca exaustiva para o Caso 1.

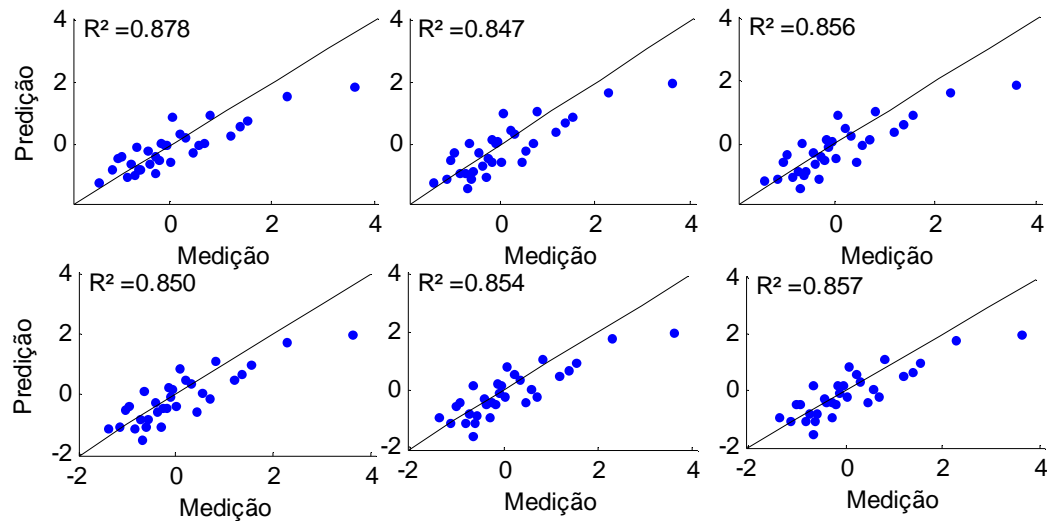


Figura 4.16: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até seis (baixo dir.) no modelo QPLS com busca exaustiva para o Caso 1.

Caso 2 (com a composição da carga)

O resultado obtido na determinação do número de variáveis principais para cada tipo de modelo, tanto para o conjunto selecionado por busca exaustiva como pelo método *stepwise* está mostrado na Tabela 4.11. Vemos que o resultado é bastante semelhante com o Caso 1 para o número de variáveis principais.

Tabela 4.11: Resultado para variação do número de variáveis latentes.

| | Número de variáveis | Número de componentes principais | | | PRESS | | |
|----------------------------|---------------------|----------------------------------|-----|------|-------|------|------|
| | | PCA | PLS | QPLS | PCA | PLS | QPLS |
| Busca exaustiva | 7 | 7 | 7 | 1 | 1,04 | 1,04 | 0,83 |
| Stepwise regression | 5 | 5 | 5 | 2 | 1,11 | 1,11 | 0,80 |

Comparando-se com o Caso 1, os valores do resíduo foram bem melhores, já que temos a inclusão da composição de carga como variável auxiliar. Vê-se pelas Figuras 4.17 e 4.18 a relação mais estreita entre os valores medidos e preditos pelo modelo. A Figura 4.18 também mostra que não há mais a aparente não-linearidade vista antes. As Figuras 4.19 e 4.20 mostram o mesmo resultado para o modelo feito com QPLS. O resultado é bastante parecido com o PLS, sendo percebida apenas uma leve diminuição do PRESS.

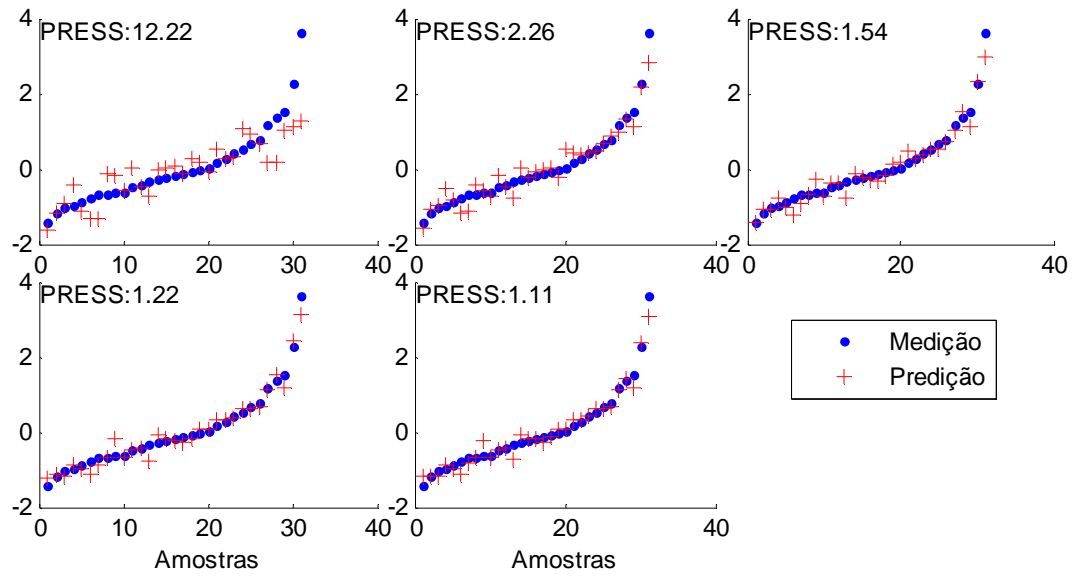


Figura 4.17: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo PLS com o método *stepwise* para o Caso 2.

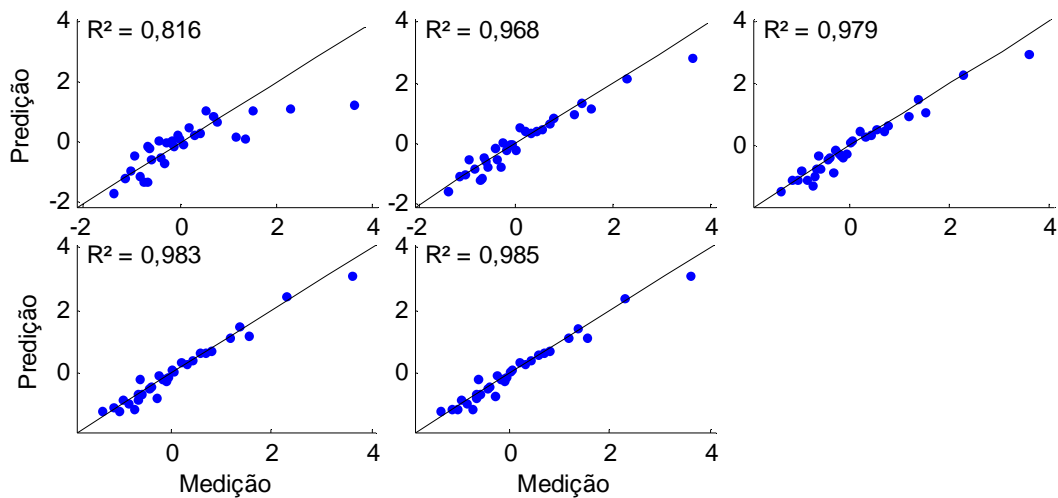


Figura 4.18: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo PLS com método *stepwise* para o Caso 2.

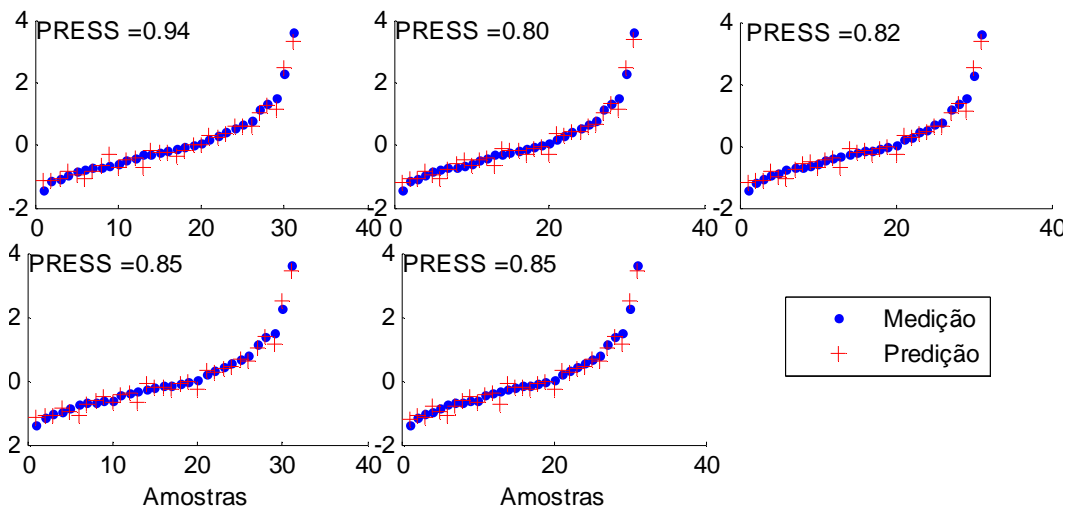


Figura 4.19: Valores preditos e medidos traçados juntos para o conjunto de dados de validação para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo QPLS com o método *stepwise* para o Caso 2.

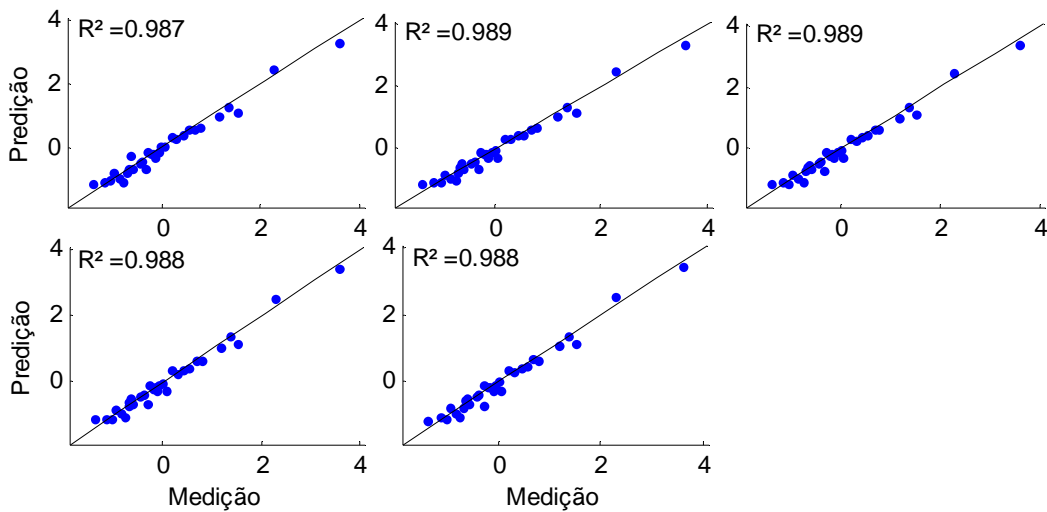


Figura 4.20: Gráfico de dispersão entre valores medidos traçados no eixo das abcissas e preditos no eixo das ordenadas para uma variável principal (cima esq.) até cinco (baixo dir.) no modelo QPLS com método *stepwise* para o Caso 2.

4.5 Estudo dos componentes principais

Para cada componente principal, é possível buscar uma interpretação física avaliando quais variáveis possuem maior peso. Na Tabela 4.12, fica claro que o primeiro componente principal representa algumas temperaturas altamente correlacionadas, que pertencem em sua maioria à primeira coluna. Já o segundo componente principal representa as temperaturas da carga da unidade. Com esse tipo de análise é possível ter uma ideia de quais variáveis são mais significativas para a modelagem. Variáveis que possuem pesos altos em um mesmo componente principal são fortemente correlacionadas entre si, como é o caso da TI-01 e da TI-02 no segundo componente principal da Tabela 4.12. Nessa situação, geralmente será necessário apenas uma das duas variáveis no modelo final, já que elas possuem informações redundantes.

Na Tabela 4.12 estão mostrados os pesos em cada componente principal para o Caso 1 obtidos com o método PLS para um total de seis componentes. A Tabela 4.14 é análoga para o Caso 2, onde foi obtido um total de sete componentes principais. O número total de componentes foi determinado pelo resultado da análise de dimensão do sistema. Em cada Tabela, estão destacadas com um asterisco as variáveis que haviam sido selecionadas previamente pelo método da busca exaustiva. Para cada componente principal foi sublinhado o peso de maior valor, que corresponde à variável que melhor representa aquele componente.

O objetivo desta análise é verificar se há relação entre as variáveis selecionadas por busca exaustiva e as variáveis com maior peso em cada componente principal. Analisando as Tabelas, vemos que para o Caso 1 apenas duas das seis variáveis possuíam o maior peso em um dos componentes principais. Para o Caso 2, foram duas em sete selecionadas. Esse resultado já sugere que o haverá diferenças nos resultados dos modelos criados com as variáveis selecionadas por busca exaustiva e pelas variáveis com maior peso nos componentes principais. Para uma comparação quantitativa, uma modelagem foi feita para cada caso e está mostrada na Tabela 4.13.

Os números apresentados na Tabela 4.13 mostram que os resultados dos modelos criados usando-se as variáveis selecionadas pelos maiores pesos de cada componente principal foram piores do que os modelos por busca exaustiva. Como a busca exaustiva atinge o ótimo global, não era esperado um modelo melhor, porém os valores de PRESS foram quase duas vezes maiores, o que mostra que não há boa relação entre as variáveis com maiores pesos nos componentes principais e as variáveis ótimos para a modelagem.

Tabela 4.12: Componentes principais para o Caso 1 por PLS, onde variáveis com asterisco foram selecionadas para modelo com busca exaustiva e valores sublinhados são o maior peso de cada componente principal.

| Variável | Componentes Principais | | | | | |
|----------------|------------------------|--------------|--------------|-------------|--------------|--------------|
| | 1° | 2° | 3° | 4° | 5° | 6° |
| FIC-03 | -3.48 | 1.46 | -5.23 | 3.91 | -2.36 | 2.27 |
| PIC-03 | -0.79 | 8.24 | -4.16 | 2.14 | 0.59 | 0.11 |
| TI-01* | -0.79 | 10.05 | -0.07 | -0.86 | 1.09 | -0.05 |
| TI-02 | -0.84 | <u>10.14</u> | -0.15 | -0.86 | 0.91 | -0.29 |
| PIC-01 | 3.68 | 4.09 | 3.88 | 0.07 | <u>-6.47</u> | 0.49 |
| FIC-01 | -3.73 | 7.10 | <u>-6.35</u> | 0.50 | 1.34 | 1.10 |
| TI-03 | -5.45 | -0.64 | 4.89 | <u>4.62</u> | -4.30 | 0.77 |
| TIC-01 | -8.01 | 5.96 | -0.30 | 1.80 | -0.07 | -1.15 |
| TI-04* | -8.64 | 5.66 | 0.34 | -0.37 | 1.59 | -0.38 |
| TI-05 | -8.35 | 2.09 | 5.03 | -3.76 | 0.82 | 1.57 |
| TI-06 | -8.30 | 1.16 | 5.31 | -4.13 | 0.58 | 1.40 |
| TI-07 | -8.66 | 1.89 | 4.69 | -3.67 | 0.44 | 1.71 |
| TI-08 | 3.65 | -5.49 | 3.48 | -2.99 | 4.94 | -3.82 |
| PI-01* | -3.88 | 6.58 | -4.79 | 4.13 | 0.93 | -1.39 |
| PDI-01 | -4.05 | -6.06 | -2.49 | 3.96 | -0.26 | -0.30 |
| TIC-02* | -0.12 | 0.10 | 2.19 | -2.34 | 5.97 | <u>-6.18</u> |
| TI-10 | -1.40 | 2.32 | 3.10 | 0.75 | 0.32 | -4.95 |
| TI-11* | <u>-8.69</u> | 1.38 | -3.41 | -2.38 | -0.58 | -1.80 |
| PIC-02 | -6.26 | 1.84 | -4.56 | 3.29 | -1.31 | -4.08 |
| FIC-02* | -0.79 | -8.09 | 0.93 | 2.44 | -0.25 | 0.10 |

Tabela 4.13: Comparação entre modelos com variáveis selecionadas por busca exaustiva e através dos pesos dos componentes principais.

| | | Número de variáveis | Número de componentes principais | PRESS |
|---------------|------------------------|---------------------|----------------------------------|-------|
| Caso 1 | Busca exaustiva | 6 | 4 | 10,58 |
| | Pesos dos CPs | 6 | 3 | 15,53 |
| Caso 2 | Busca exaustiva | 7 | 7 | 1,04 |
| | Pesos dos CPs | 6 | 6 | 1,82 |

Tabela 4.14: Componentes principais para o Caso 2 com PLS, onde variáveis com asterisco foram selecionadas para modelo com busca exaustiva e valores sublinhados são o maior peso de cada componente principal.

| Variável | Componentes principais | | | | | | |
|------------------|------------------------|--------------|--------------|-------------|--------------|--------------|-------------|
| | 1° | 2° | 3° | 4° | 5° | 6° | 7° |
| AI-01.C2* | <u>9.67</u> | 2.90 | 2.54 | 3.21 | 1.07 | 0.80 | 0.86 |
| FIC-03 | -3.56 | 1.65 | -4.27 | <u>4.76</u> | -2.08 | 2.02 | 1.93 |
| PIC-03 | -0.38 | 7.99 | -4.57 | 2.13 | 0.12 | -0.24 | -0.63 |
| TI-01 | -0.17 | 9.97 | -1.06 | -1.44 | 0.63 | -0.82 | -2.33 |
| TI-02 | -0.26 | <u>10.02</u> | -1.27 | -1.59 | 0.40 | -0.98 | -2.18 |
| PIC-01* | 3.91 | 3.75 | 2.30 | -2.52 | -6.08 | 2.45 | 3.03 |
| FIC-01 | -3.38 | 7.08 | <u>-6.01</u> | 2.19 | 2.01 | 0.68 | -0.10 |
| TI-03* | -5.09 | 0.15 | 4.94 | 2.42 | <u>-6.41</u> | 0.44 | -0.24 |
| TIC-01 | -7.53 | 6.52 | -0.36 | 1.49 | -1.02 | -1.35 | -0.86 |
| TI-04 | -8.23 | 6.27 | 0.30 | -0.16 | 1.25 | -0.78 | -1.03 |
| TI-05* | -8.14 | 2.88 | 4.85 | -3.74 | 1.65 | 1.66 | -0.39 |
| TI-06* | -8.17 | 1.93 | 5.09 | -4.15 | 1.52 | 1.66 | -0.54 |
| TI-07 | -8.48 | 2.67 | 4.50 | -3.63 | 1.32 | 1.85 | -0.37 |
| TI-08 | 3.28 | -5.56 | 3.53 | -3.01 | 4.18 | -4.84 | 0.83 |
| PI-01 | -3.28 | 6.80 | -4.24 | 4.70 | -0.05 | -1.81 | 0.76 |
| PDI-01 | -4.29 | -5.68 | -1.19 | 4.55 | -0.78 | -0.24 | -0.14 |
| TIC-02 | -0.09 | 0.24 | 2.16 | -2.27 | 4.87 | <u>-6.91</u> | <u>4.62</u> |
| TI-10 | -1.16 | 2.54 | 2.65 | -0.68 | -1.07 | -4.85 | 4.26 |
| TI-11* | -8.62 | 1.92 | -2.64 | -0.10 | 1.14 | -0.46 | 3.21 |
| PIC-02 | -6.13 | 2.11 | -4.04 | 3.86 | -1.73 | -2.91 | 2.13 |
| FIC-02* | -1.14 | -7.78 | 1.93 | 2.45 | -1.07 | -0.77 | -3.22 |

Capítulo 5

Modelo dinâmico

Neste capítulo serão mostrados os resultados da inclusão de dinâmicas nas entradas dos modelos estáticos obtidos previamente. O resultado de cada um dos casos será comparado com um modelo estático obtido através de modelagem direta dos dados dinâmicos e um modelo dinâmico autorregressivo.

5.1 Modelo estático

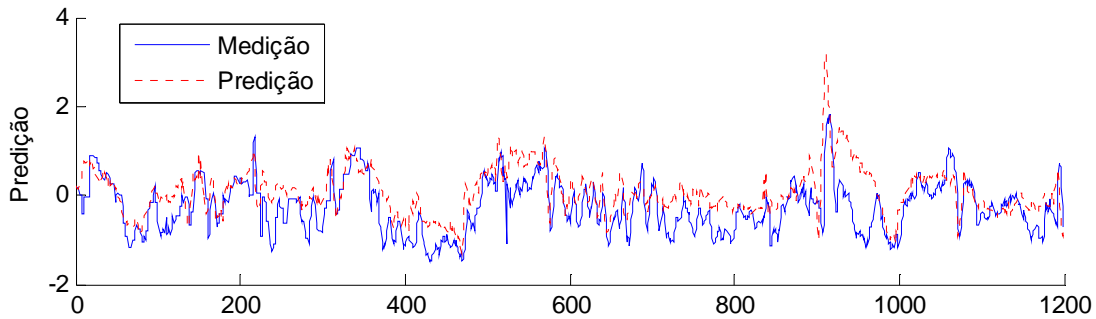
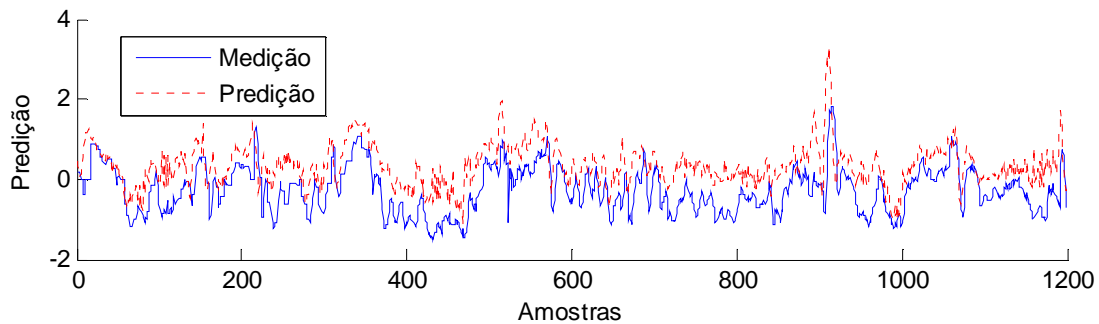
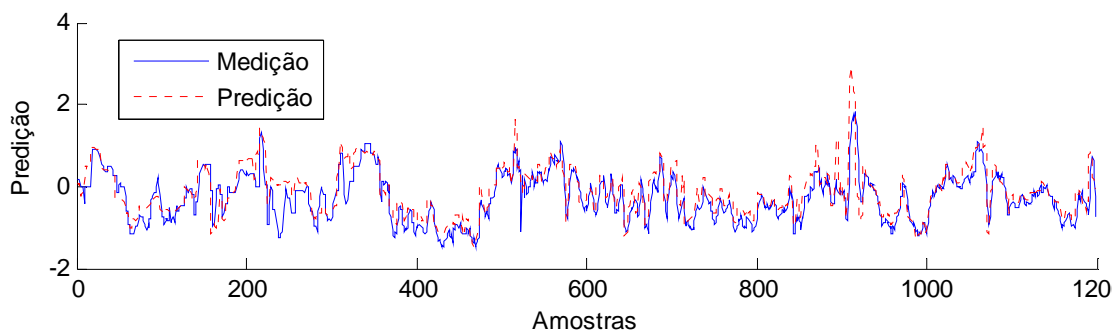
Como já mencionado anteriormente, um modelo é classificado como estático quando se utiliza dados dinâmicos para modelagem, mas a estrutura do modelo não inclui como entradas valores das variáveis com atraso no tempo. Caso isso fosse feito, o resultado obtido seria um modelo do tipo dinâmico.

Para a obtenção desse modelo, utilizou-se uma metodologia equivalente à usada para os modelos estacionários. A única alteração a ser feita é omitir a etapa de detecção dos estados estacionários, ou seja, foi usado o mesmo conjunto de dados. Para permitir uma melhor comparação entre os resultados, as variáveis auxiliares usadas foram as mesmas selecionadas anteriormente. Dessa forma, obteve-se quatro modelos estáticos a serem comparados com os modelos estacionários equivalentes com dinâmicas nas entradas.

Na Tabela 5.1 estão apresentados os resultados obtidos para esses modelos estáticos. As Figuras 5.1, 5.2, 5.3 e 5.4 mostram num mesmo gráfico a variável modelada medida e predita no conjunto de validação para Casos 1 e 2 utilizando os dois conjuntos de variáveis auxiliares selecionados

Tabela 5.1: Resultados para os modelos estáticos.

| | Número de variáveis principais | | | PRESS | | |
|--------------------|--------------------------------|-----|------|--------|--------|--------|
| | PCA | PLS | QPLS | PCA | PLS | QPLS |
| Caso 1 – BE | 6 | 6 | 6 | 388,51 | 388,51 | 381,81 |
| Caso 1 – SR | 9 | 9 | 7 | 683,32 | 683,32 | 683,37 |
| Caso 2 – BE | 7 | 7 | 6 | 151,61 | 151,61 | 149,43 |
| Caso 2 – SR | 5 | 5 | 3 | 151,00 | 151,00 | 149,27 |

**Figura 5.1:** Validação do modelo estático para o Caso 1 com variáveis selecionadas por busca exaustiva.**Figura 5.2:** Validação do modelo estático para o Caso 1 com variáveis selecionadas pelo método *stepwise*.**Figura 5.3:** Validação do modelo estático para o Caso 2 com variáveis selecionadas por busca exaustiva.

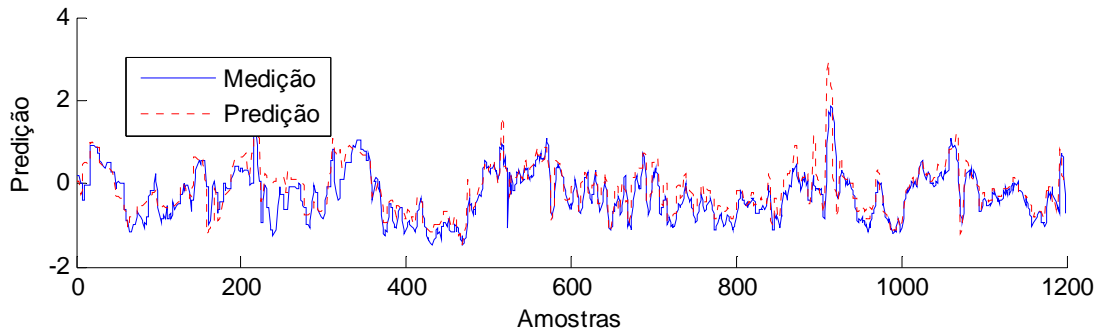


Figura 5.4: Validação do modelo estático para o Caso 2 com variáveis selecionadas por método *stepwise*.

5.2 Modelo dinâmico ARX

O modelo dinâmico direto usado para comparação foi baseado em um modelo autorregressivo do tipo ARX. De forma análoga ao modelo estático, foi usado o mesmo conjunto de dados e a mesma metodologia, porém sem a etapa de determinação dos estados estacionários.

O resultado obtido está na Tabela 5.2 para os quatro casos estudados. As Figuras 5.5, 5.7, 5.9 e 5.11 mostram num mesmo gráfico a variável modelada medida e predita no conjunto de validação para os casos estudados. Para definição da ordem do modelo, a ordem foi variada de 1 até 10 e o melhor modelo escolhido para representar o caso. As Tabelas 5.3, 5.4, 5.5 e 5.6 mostram os valores obtidos para a escolha da ordem do modelo. As Figuras 5.6, 5.8, 5.10 e 5.12 mostram as respostas para os melhores modelos de cada caso. Como é possível perceber, as dinâmicas determinadas foram semelhantes para as variáveis que aparecem em mais de um problema.

Tabela 5.2: Resultado dos modelos dinâmicos obtidos diretamente com dados de planta.

| | Ordem do modelo autorregressivo | PRESS |
|---|---------------------------------|-------|
| Caso 1 – Busca exaustiva (BE) | 4 | 410,4 |
| Caso 1 – <i>Stepwise Regression</i> (SR) | 1 | 486,2 |
| Caso 2 – BE | 2 | 148,1 |
| Caso 2 – SR | 2 | 151,0 |

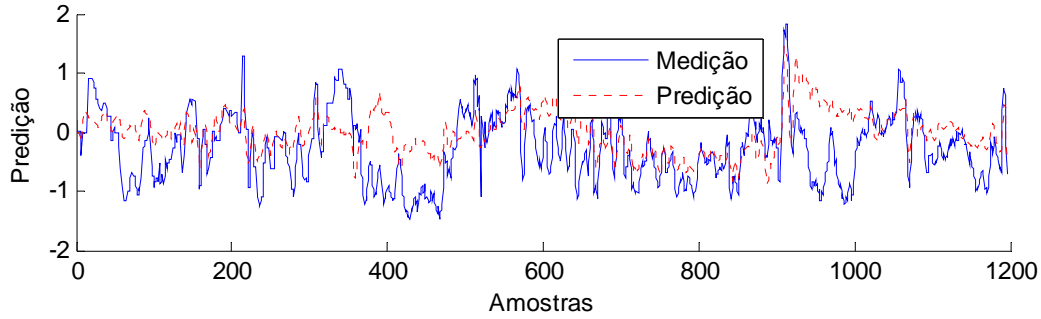


Figura 5.5: Validação do modelo dinâmico direto para o Caso 1 com variáveis seleccionadas por busca exaustiva.

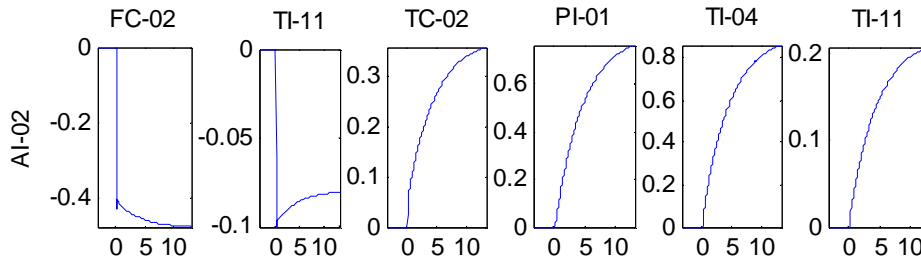


Figura 5.6: Resposta ao degrau para o modelo de ordem 4 para o Caso 1 com busca exaustiva, no eixo das abcissas tempo em 10000 segundos.

Tabela 5.3: Definição da ordem do modelo para o Caso 1 com busca exaustiva.

| Ordem | PRESS |
|----------|--------------|
| 1 | 505,5 |
| 2 | 529,7 |
| <u>4</u> | <u>410,4</u> |
| 6 | 590,6 |
| 8 | 774,0 |
| 10 | 696,7 |

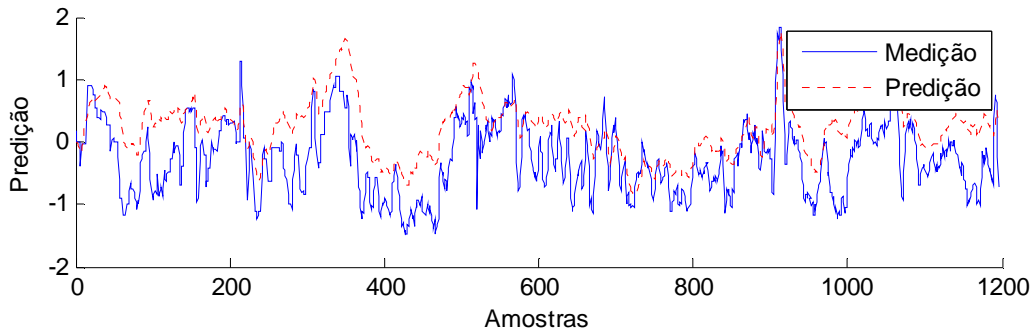


Figura 5.7: Validação do modelo dinâmico direto para o Caso 1 com variáveis seleccionadas pelo método *stepwise*.

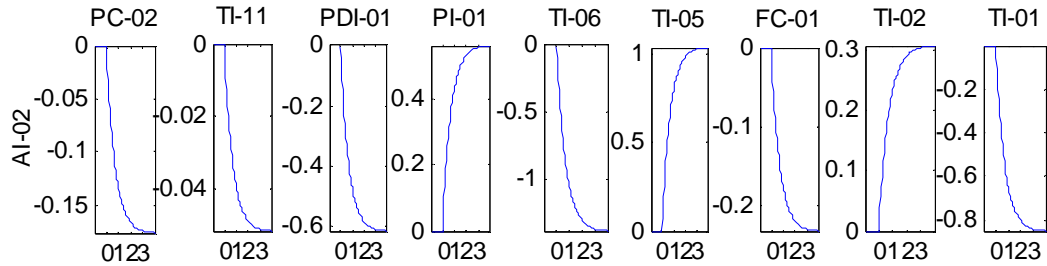


Figura 5.8: Resposta ao degrau para o modelo de ordem 1 para o Caso 1 com *stepwise regression*, no eixo das abcissas tempo em 10000 segundos.

Tabela 5.4: Definição da ordem do modelo para o Caso 1 com o método *stepwise*.

| Ordem PRESS | |
|-------------|--------|
| 1 | 486,2 |
| 2 | 621,7 |
| 4 | 586,5 |
| 6 | 846,3 |
| 8 | 1112,6 |
| 10 | 1082,9 |

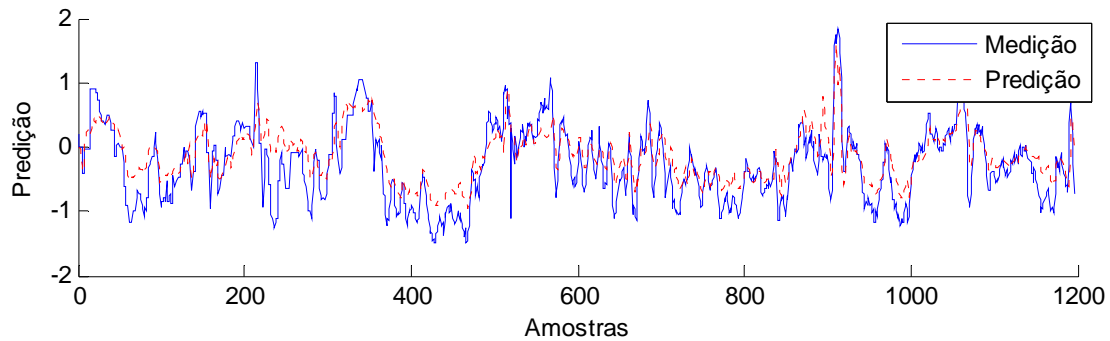


Figura 5.9: Validação do modelo dinâmico direto para o Caso 2 com variáveis selecionadas por busca exaustiva.

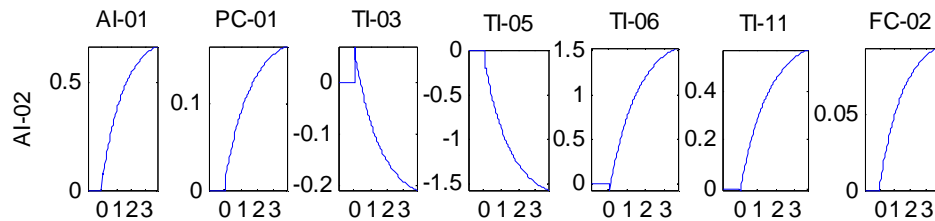
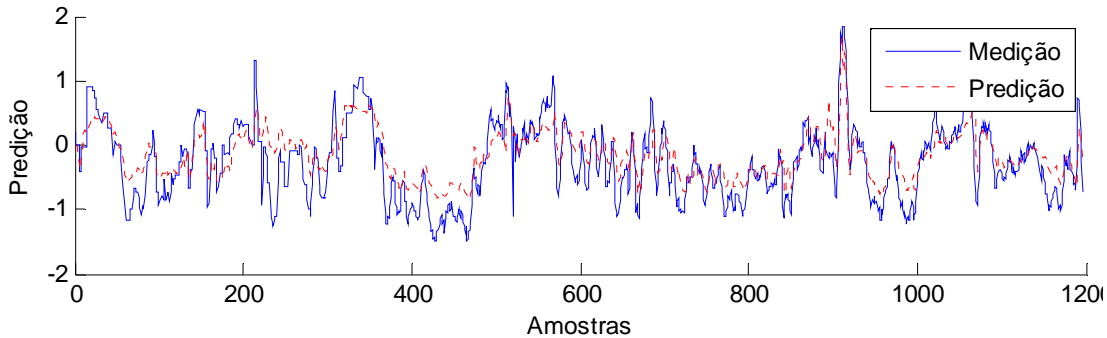
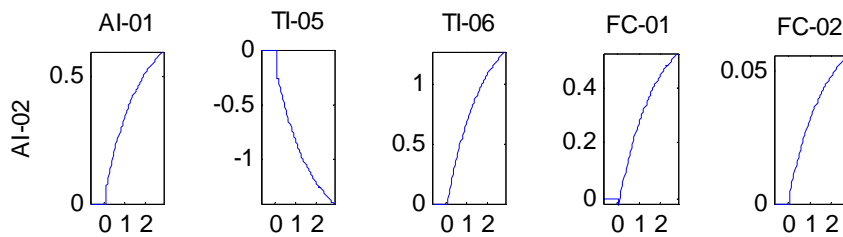


Figura 5.10: Resposta ao degrau para o modelo de ordem 4 para o Caso 1 com busca exaustiva, no eixo das abcissas tempo em 10000 segundos.

Tabela 5.5: Definição da ordem do modelo para o Caso 2 com busca exaustiva.

| Ordem | PRESS |
|-----------------|---------------------|
| 1 | 173,8 |
| <u>2</u> | <u>148,1</u> |
| 4 | 332,9 |
| 6 | 426,5 |
| 8 | 368,7 |
| 10 | 340,7 |

**Figura 5.11:** Validação do modelo dinâmico direto para o Caso 2 com variáveis selecionadas pelo métodos *stepwise*.**Figura 5.12:** Resposta ao degrau para o modelo de ordem 2 para o Caso 2 com *stepwise regression*, no eixo das abcissas tempo em 10000 segundos.**Tabela 5.6:** Definição da ordem do modelo para o Caso 2 com o método *stepwise*.

| Ordem | PRESS |
|-----------------|---------------------|
| 1 | 169.0 |
| <u>2</u> | <u>151,0</u> |
| 4 | 373,3 |
| 6 | 423,0 |
| 8 | 368,1 |
| 10 | 347,3 |

5.3 Obtenção dos valores ótimos de dinâmicas

Esta é a última etapa da aplicação da metodologia proposta. Os modelos estacionários desenvolvidos no capítulo anterior são usados como base para desenvolvimento de um modelo dinâmico. Foram incluídas dinâmicas de primeira ordem na entrada de cada uma das variáveis auxiliares utilizadas e seu valor ótimo foi buscado para o mesmo conjunto de dados

utilizado no desenvolvimento dos modelos estáticos e dinâmicos diretos apresentados anteriormente. Os resultados para esse modelo estão apresentados na Tabela 5.7. Nessa tabela são mostrados também os resultados obtidos para o modelo estacionário desenvolvido no capítulo anterior aplicado ao mesmo conjunto de dados dinâmicos.

Tabela 5.7: Resultados para o modelo dinâmico proposto comparado com o resultado para o modelo estacionário.

| | Modelo dinâmico proposto | | Modelo estacionário | | Modelo ARX |
|--------------------|--------------------------|--------|---------------------|--------|------------|
| | PRESS | | PRESS | | PRESS |
| | PLS | QPLS | PLS | QPLS | |
| Caso 1 - BE | 333,76 | 368,62 | 454,34 | 537,36 | 410,4 |
| Caso 1 - SR | 476,42 | 426,67 | 695,41 | 826,51 | 486,2 |
| Caso 2 - BE | 186,31 | 197,34 | 222,36 | 221,09 | 148,1 |
| Caso 2 - SR | 209,64 | 193,24 | 218,52 | 271,97 | 151,0 |

A seguir, estão mostrados com mais detalhes os resultados obtidos para cada um dos casos. Como esse problema de otimização apresenta muitos mínimos locais, especialmente devido a erros de medição, não-linearidades na dinâmica e mudanças de condição de processo, foram testados vários valores iniciais para o problema, resultando em valores distintos para as dinâmicas. Os valores iniciais foram variados desde 0 até 500. Esse intervalo foi escolhido de forma a abranger os valores esperados para essas dinâmicas. O modelo final selecionado foi o que obteve melhor resultado para os índices de avaliação.

5.3.1 Caso 1 com busca exaustiva

Para este caso, os resultados das diversas soluções para o problema de otimização de dinâmicas estão mostrados na Tabela 5.8. As dinâmicas obtidas para cada variável estão mostradas na Tabela 5.9. A Figura 5.13 mostra uma comparação entre as variáveis auxiliares com e sem a dinâmica obtida. A Figura 5.14 mostra o resultado do modelo aplicado ao conjunto de validação.

Como se vê na Tabela 5.9, alguns valores de dinâmica para o modelo QPLS foram bastante elevados, o que seria o equivalente a excluir essa variável auxiliar do modelo. Esse resultado é condizente com o que foi obtido durante a elaboração do modelo estático. Lá o valor da dimensão do problema obtido para o PLS quadrático foi 7, o que permite excluir duas variáveis sem perdas de informação.

Os valores obtidos para as dinâmicas no modelo PLS estão condizentes com as estimativas dos operadores. As dinâmicas entre os sensores de temperatura dos pratos da torre e a composição são rápidas. O mesmo pode ser dito da vazão de refluxo de topo. Para a TI-01, o valor foi um pouco abaixo do esperado, já que está temperatura pertence à primeira torre da unidade.

Tabela 5.8: Resultado da otimização das dinâmicas para o Caso 1 com busca exaustiva.

| Valor inicial (minutos) | PRESS | |
|--------------------------------|--------------|-------------|
| | PLS | QPLS |
| Sem dinâmica | 454,34 | 537,36 |
| 0 | 454,33 | 537,36 |
| 1 | 341,90 | 389,61 |
| 2 | 341,90 | 389,61 |
| 5 | 341,90 | 389,61 |
| 10 | 341,90 | 388,74 |
| 20 | 341,90 | 388,74 |
| 50 | 333,76 | 389,89 |
| 100 | 336,39 | 368,62 |
| 200 | 340,87 | 368,62 |
| 500 | 340,87 | 512,64 |

Tabela 5.9: Dinâmicas obtidas para o Caso 1 com busca exaustiva.

| Variável | Constante de tempo (min) | |
|-----------------|---------------------------------|-------------|
| | PLS | QPLS |
| TI-01 | 1,5 | 3,6 |
| TI-04 | 69,9 | 1000,1 |
| PI-01 | 28,5 | 26,8 |
| TIC-02 | 0,0 | 798,0 |
| TI-11 | 25,1 | 33,5 |
| FIC-02 | 20,1 | 340,7 |

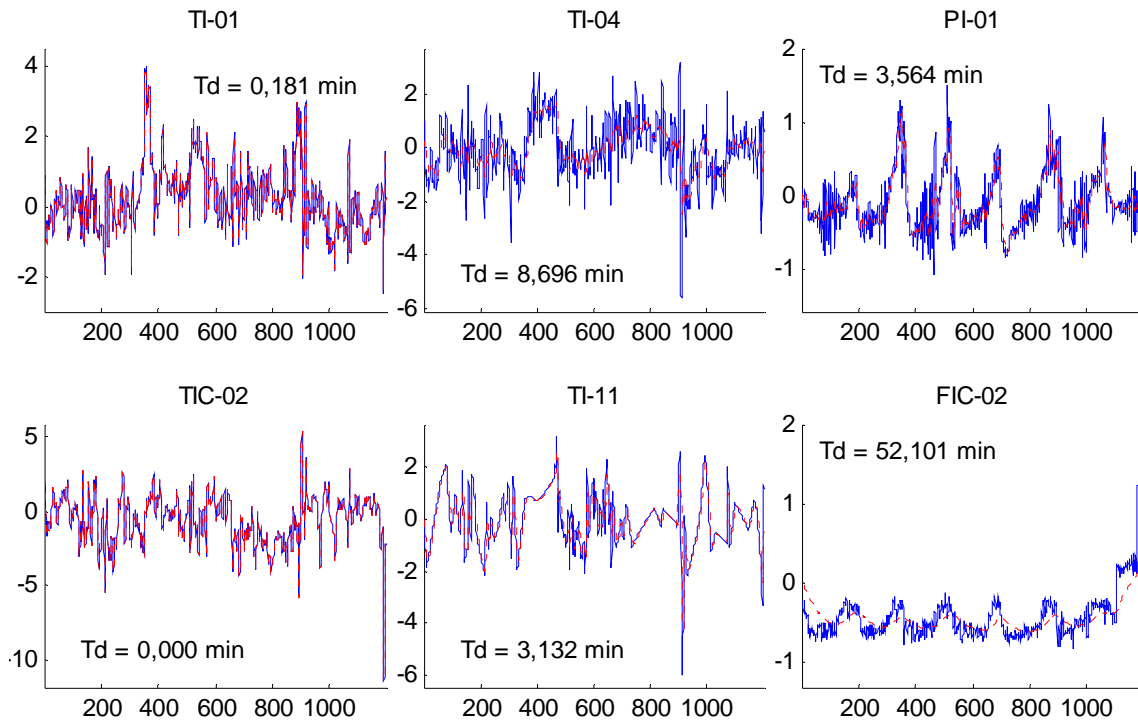


Figura 5.13: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 1 com busca exaustiva, onde T_d é a constante de tempo.

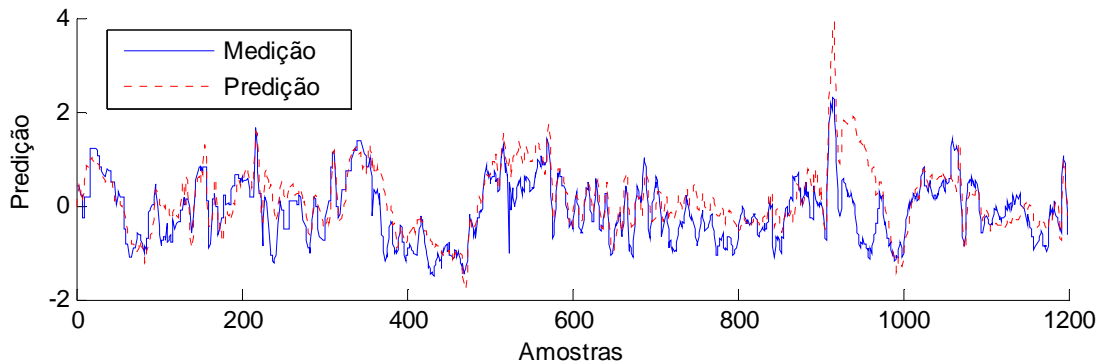


Figura 5.14: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 1 com busca exaustiva.

5.3.2 Caso 1 com método stepwise

Os resultados desse caso estão mostrados nas Tabelas 5.11 e 5.10. A Figura 5.15 mostra o gráfico das variáveis auxiliares com e sem as dinâmicas na entrada. A Figura 5.16 mostra os valores medidos e preditos da variável de interesse para o conjunto de validação.

As dinâmicas obtidas para esse caso são parecidas com as anteriores para o modelo PLS. O modelo QPLS, porém não apresentou nenhuma variável com valor de dinâmica muito elevada, fato que havia ocorrido antes. Os ganhos com a adição de dinâmica para este caso

foram expressivos, chegando a resultar em uma diminuição em quase 50% no PRESS para o modelo QPLS se comparado ao modelo sem dinâmicas.

Tabela 5.10: Resultado da otimização das dinâmicas para o Caso 1 com o método *stepwise*.

| Valor inicial (minutos) | PRESS | |
|----------------------------|---------|---------|
| | PLS | QPLS |
| Sem dinâmica | 695,41 | 826,51 |
| 0 | 528,10 | 766,88 |
| 1 | 476,42 | 426,67 |
| 2 | 476,42 | 426,67 |
| 5 | 550,83 | 1269,70 |
| 10 | 1030,49 | 1253,85 |
| 20 | 1030,49 | 1253,85 |
| 50 | 1030,49 | 1216,39 |
| 100 | 1030,49 | 1216,39 |
| 200 | 746,50 | 1374,26 |
| 500 | 741,41 | 513,65 |

Tabela 5.11: Dinâmicas obtidas para o Caso 1 com o método *stepwise*.

| Variável | Constante de tempo (min) | |
|---------------|--------------------------------|-------|
| | PLS | QPLS |
| TI-01 | 0,0 | 0,1 |
| TI-02 | 84,1 | 111,0 |
| FIC-01 | 1,5 | 2,5 |
| TI-05 | 0,0 | 0,0 |
| TI-06 | 12,9 | 4,0 |
| PI-01 | 38,1 | 40,0 |
| PDI-01 | 0,0 | 0,0 |
| TI-11 | 29,2 | 59,8 |
| PIC-01 | 16,2 | 14,5 |

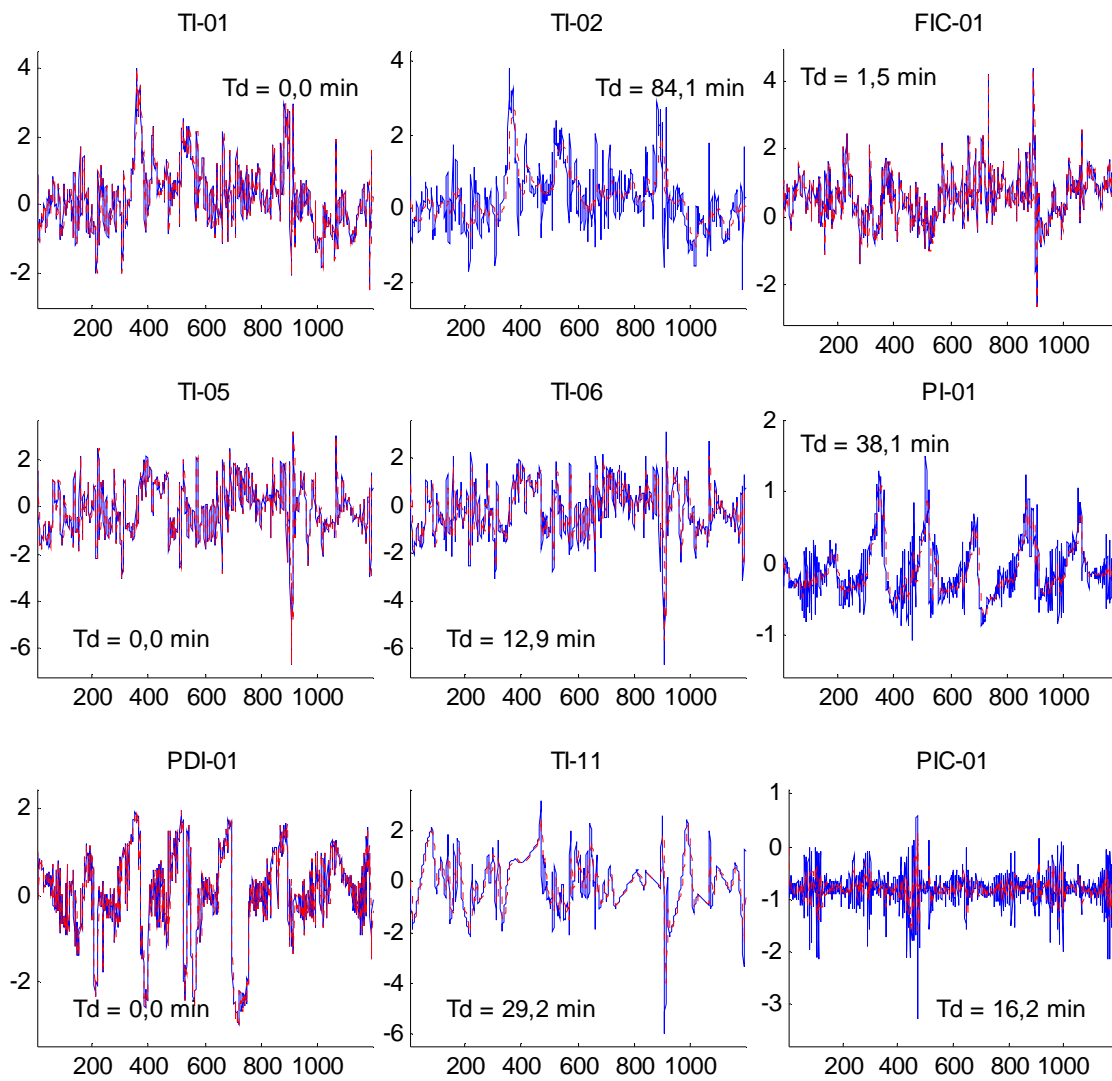


Figura 5.15: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 1 com o método *stepwise*, onde T_d é a constante de tempo.

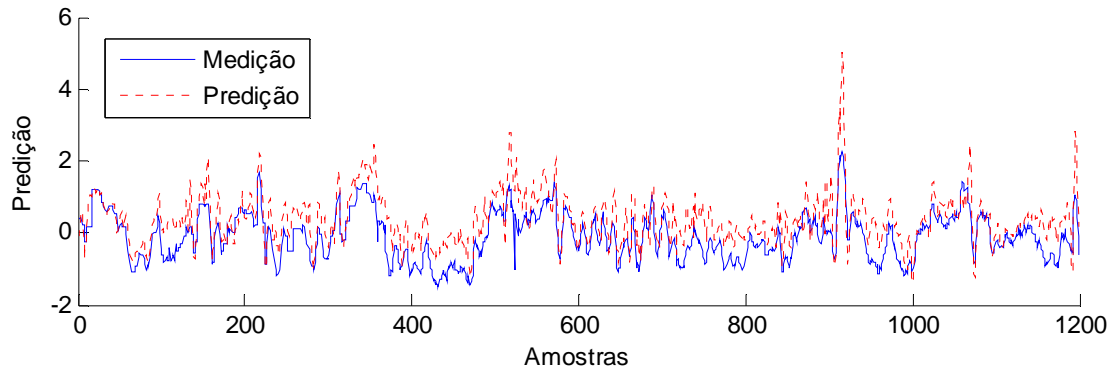


Figura 5.16: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 1 com método *stepwise*.

5.3.3 Caso 2 com busca exaustiva

O resultado para este caso está mostrado nas Tabelas 5.12 e 5.13. A Figura 5.17 mostra as variáveis de entrada com e sem as dinâmicas enquanto a Figura 5.18 ilustra o resultado do modelo para o conjunto de validação.

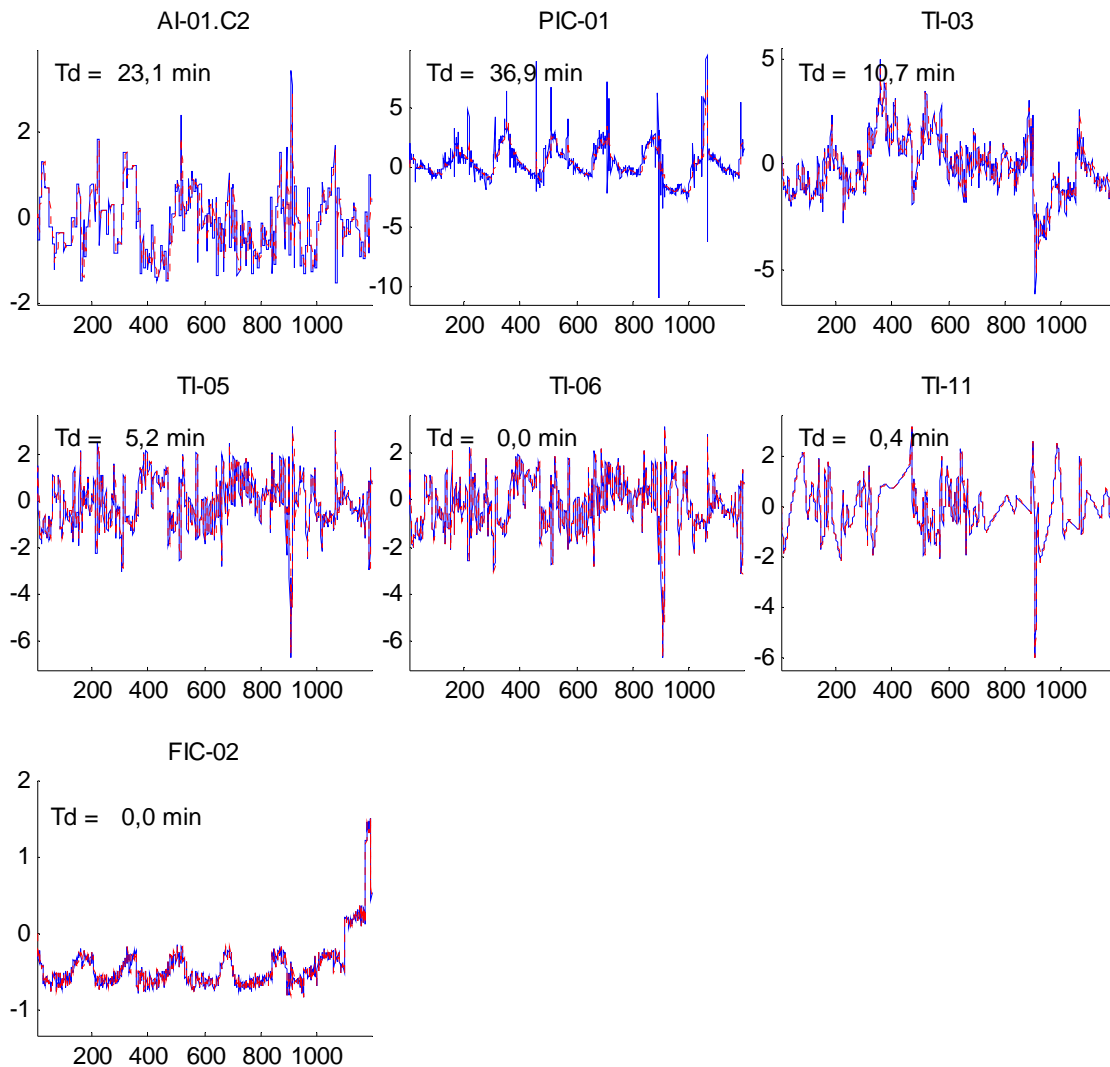
Verifica-se que para o QPLS foi obtida um valor infinito para uma das dinâmicas, ou seja, o resultado ficou melhor sem esta variável auxiliar. Novamente, o resultado de determinação da dimensão do problema realizada durante a obtenção do modelo estático mostrou que isso seria possível, já que o número de variáveis latentes ótimo foi de seis. Tirando essa diferença, os valores foram praticamente os mesmos para o modelo PLS e o QPLS.

Tabela 5.12: Resultado da otimização das dinâmicas para o Caso 2 com busca exaustiva.

| Valor inicial (minutos) | PRESS | |
|----------------------------|--------|--------|
| | PLS | QPLS |
| Sem dinâmica | 222,36 | 221,09 |
| 0 | 219,77 | 212,52 |
| 1 | 186,31 | 197,34 |
| 2 | 186,31 | 197,34 |
| 5 | 215,57 | 271,58 |
| 10 | 261,01 | 271,58 |
| 20 | 260,20 | 270,89 |
| 50 | 323,09 | 385,02 |
| 100 | 323,09 | 385,02 |
| 200 | 260,20 | 385,02 |
| 500 | 257,33 | 381,52 |

Tabela 5.13: Dinâmicas obtidas para o Caso 2 com busca exaustiva.

| Variável | Constante de tempo (min) | |
|-----------------|--------------------------|------|
| | PLS | QPLS |
| AI-01.C2 | 23,1 | 26,0 |
| PIC-01 | 36,9 | 57,8 |
| TI-03 | 10,7 | 6,7 |
| TI-05 | 5,2 | 3,7 |
| TI-06 | 0,0 | 0,0 |
| TI-11 | 0,4 | Inf |
| FIC-02 | 0,0 | 0,0 |

**Figura 5.17:** Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 2 com busca exaustiva, onde T_d é a constante de tempo.

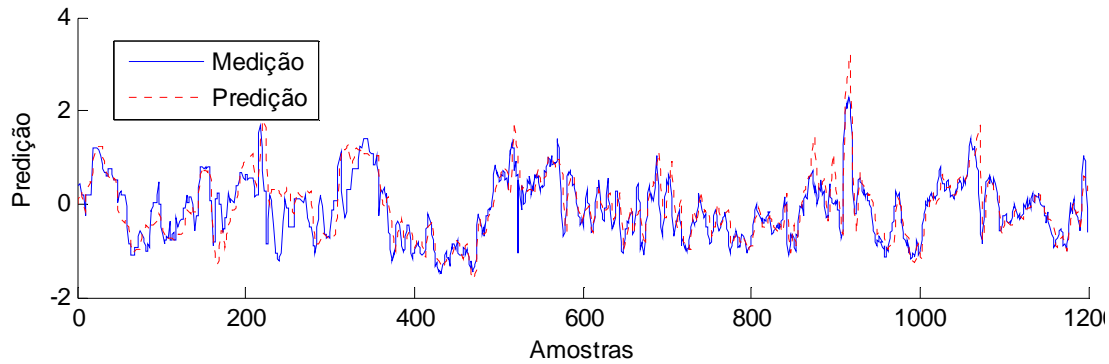


Figura 5.18: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 2 com busca exaustiva.

5.3.4 Caso 2 com o método *stepwise*

As Tabelas 5.15 e 5.14 trazem o resultado para este caso e as Figuras 5.19 e 5.20 ilustram de maneira análoga aos casos passados o resultado obtido para o modelo dinâmico. Novamente os resultados para os modelos PLS e QPLS foram bastante semelhantes. Desta vez, porém, uma das dinâmicas do modelo PLS foi muito elevada, o que não era esperado.

Tabela 5.14: Resultado da otimização das dinâmicas para o Caso 2 com o método *stepwise*.

| Valor inicial (minutos) | PRESS | |
|----------------------------|--------|--------|
| | PLS | QPLS |
| Sem dinâmica | 218,52 | 271,97 |
| 0 | 218,72 | 218,60 |
| 1 | 209,64 | 193,24 |
| 2 | 209,64 | 193,24 |
| 5 | 209,64 | 262,37 |
| 10 | 248,03 | 262,37 |
| 20 | 248,03 | 261,80 |
| 50 | 309,32 | 373,72 |
| 100 | 309,32 | 373,72 |
| 200 | 309,32 | 373,72 |
| 500 | 309,32 | 373,72 |

Tabela 5.15: Dinâmicas obtidas para o Caso 2 com o método *stepwise*.

| Variável | Constante de tempo (min) | |
|-----------------|--------------------------|--------|
| | PLS | QPLS |
| AI-01.C2 | 22,2 | 25,3 |
| TI-05 | 5,1 | 3,5 |
| TI-06 | 0,0 | 0,0 |
| TI-11 | 4069,0 | 4069,0 |
| FIC-02 | 0,0 | 0,0 |

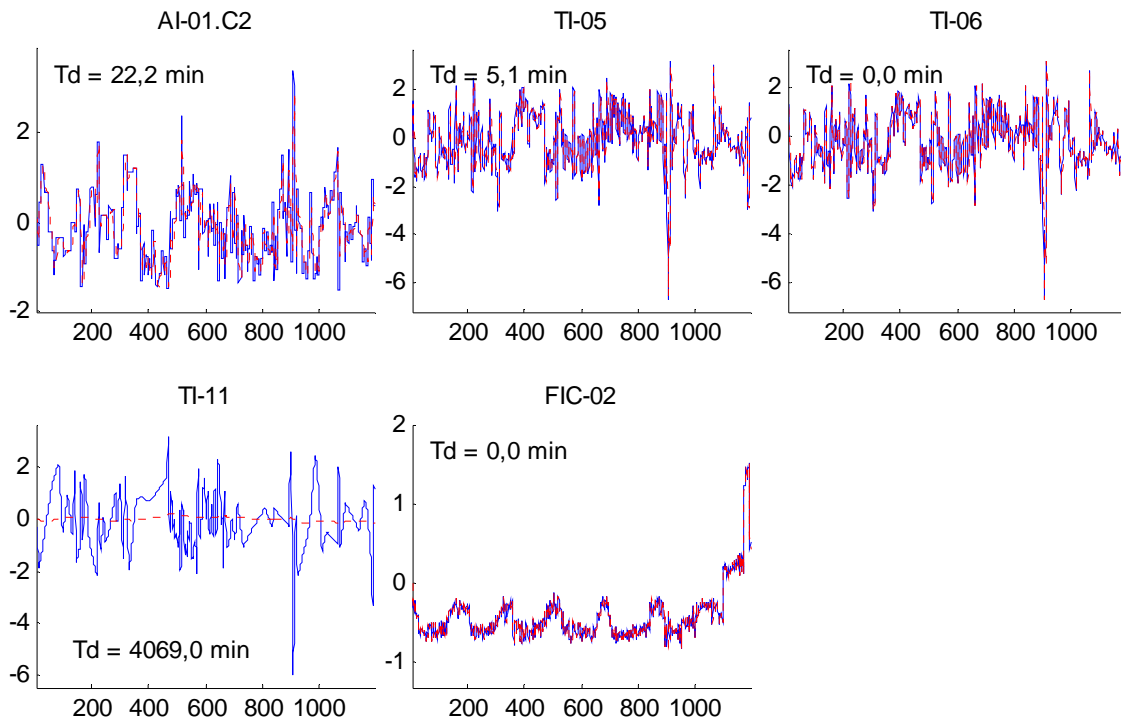


Figura 5.19: Comparação entre as variáveis auxiliares com e sem dinâmica para o Caso 2 com o método *stepwise*, onde T_d é a constante de tempo.

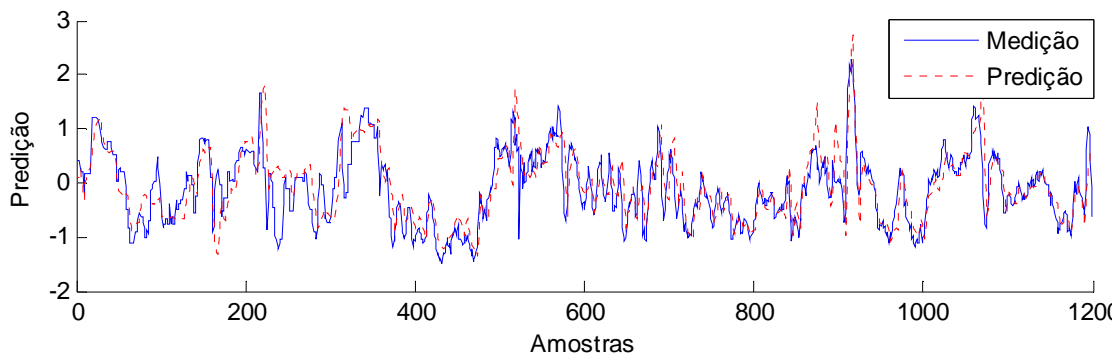


Figura 5.20: Comparação entre os valores preditos e medidos no conjunto de validação para o Caso 2 com método *stepwise*.

5.4 Discussão

É importante reforçar que todos os modelos apresentados tiveram como ponto de partida o trabalho realizado no capítulo anterior. Ou seja, as variáveis foram selecionadas usando dados estacionários, o que de certa forma beneficia o resultado final de todos. O método de separação dos conjuntos de modelagem e validação para tal tarefa, *y-rank*, permite um aproveitamento de todas as regiões dos dados, como mostrou Facchin (2005). Caso se utilizassem somente dados dinâmicos, esse método não poderia ser aplicado. Seria interessante fazer a comparação também com os métodos de seleção de variáveis sem a

determinação dos estados estacionários, porém essa tarefa não entrou no escopo deste trabalho.

Um fato a se destacar é a melhora significativa que o modelo estacionário teve após serem inseridas as dinâmicas nas variáveis auxiliares. Isso foi verdade especialmente para os modelos que não incluíam a composição da carga como entrada. Podemos dizer que pelo fato do modelo com a variável do cromatógrafo já ter apresentado um excelente resultado no primeiro modelo, havia pouco espaço para melhora com a inclusão da dinâmica.

Os resultados para o primeiro caso foram melhores quando se utilizou a metodologia proposta, como mostra a Tabela 5.7. Isso se justifica porque a inclusão da etapa de detecção de estacionários torna o método mais robusto a problemas nos dados. O tratamento para detecção de estacionários elimina qualquer *outlier* que pode ter passado pelas etapas anteriores. Também são eliminados automaticamente períodos onde há forte oscilação, que podem causar problemas especialmente na determinação do modelo estático.

Para este caso, o modelo estático teve um desempenho melhor do que o modelo dinâmico tradicional. Isso não era esperado, já que se pode considerar o modelo estático um subcaso do modelo dinâmico, onde a ordem é zero. Uma possível explicação dessa diferença é o período selecionado para validação. O modelo estático se adaptou melhor com o conjunto selecionado e espera-se que se outros conjuntos forem testados, o resultado será diferente.

O resultado da inserção de dinâmicas pode ser melhorado com o uso de um algoritmo de otimização global. Isso deverá ser percebido especialmente para o Caso 2, onde as dinâmicas apresentaram valores pequenos, abaixo do esperado. Além disso, o tipo de dinâmica pode ser mais complexo do que o utilizado aqui. O resultado pode ser melhorado com a inclusão de tempo morto e dinâmicas de segunda ordem. Porém esse tipo de problema não foi tratado nesse trabalho e exige um melhor estudo do problema de otimização envolvido.

Capítulo 6

Conclusão

6.1 Considerações finais

Inferências podem ser criadas através de diversos métodos, utilizando-se diferentes tipos de modelos, diferentes técnicas de pré-tratamento e validação. A metodologia aqui apresentada representa apenas mais uma alternativa, que busca ampliar as possibilidades nesse campo da ciência.

Como visto, os resultados alcançados com a metodologia proposta foram levemente melhores aos obtidos através de técnicas tradicionais, fato que torna o método competitivo. Somado a isso, deve-se lembrar que a abordagem segmentada, permite uma maior flexibilidade na manutenção do modelo, tornando viável a realização dessa tarefa de forma automatizada com auxílio de um aplicativo de auditoria de processos.

A comparação dos dois casos estudados na aplicação da metodologia mostrou a diferença na qualidade de uma inferência quando se usa a composição de carga da coluna como variável auxiliar. Essa variável auxiliar tem a maior correlação com as composições dos produtos, porém sua frequência de amostragem é baixa e deve ser usada com cuidado.

No desenvolvimento da metodologia, mostrou-se que na etapa seleção de variáveis, o método da busca exaustiva representa a melhor alternativa quando aplicado a modelos de processos industriais. Para o número de variáveis auxiliares presente nesse tipo de problema, o custo computacional já não é um obstáculo para a aplicação da técnica. Além disso, se fixarmos o número de variáveis auxiliares a serem buscadas – através da determinação da dimensão do problema – o tempo gasto se torna irrelevante, facilitando a automatização dessa etapa.

Os resultados mostraram que o PLS é a melhor técnica linear para estudar a dimensão de um modelo. Seu resultado esteve dentro do esperado e foi melhor do que o PCA em todas

as etapas. O método usado para avaliação da dimensão foi eficiente, mostrando bons resultados para o modelo com número reduzido de variáveis.

Outro resultado interessante foi o fato da qualidade dos modelos melhorarem à medida que se removiam variáveis auxiliares até se atingir um valor ótimo. Isso se mostrou válido até mesmo com o uso de técnicas de redução de dimensionalidade, como PLS, que são menos sensíveis a problemas numéricos frutos do mau condicionamento do sistema. Para uma regressão linear, é sabido que o acréscimo de variáveis implica o aumento da variabilidade da resposta do modelo (Chatterjee e Hadi, 2006). No estudo de caso apresentado, o mesmo raciocínio foi válido para regressões feitas com o auxílio de PCA e PLS. Portanto, deve sempre se buscar usar o mínimo de variáveis, mesmo quando se usa técnicas de redução de dimensionalidade.

6.2 Sugestão para trabalhos futuros

No desenvolvimento do modelo dinâmico, diferentes formas de validação poderiam ser utilizadas. Técnicas como as apresentadas em Fortuna (2005) podem auxiliar na obtenção de um melhor modelo dinâmico e podem ser testadas em trabalhos futuros nessa área. Outros tipos de modelos poderiam ser usados na comparação dos resultados, como as redes neurais

Na etapa de otimização das dinâmicas é possível realizar melhorias. O tipo de dinâmica ainda é fixado como sendo de primeira ordem, sendo interessante incluir um parâmetro de tempo morto nessas dinâmicas. Sistemas com ordens maiores também podem ser testados, permitindo um melhor resultado final.

A metodologia de inferência proposta ainda deve ser avaliada em tempo real, com o auxílio de um esquema de correção, formando um analisador virtual. Isso pode ser feito de várias formas, com o auxílio de técnicas de adaptação de modelos, como o filtro de Kalman.

Como forma de melhorar a metodologia, é sugerido fazer uma avaliação da razão ótima de separação dos conjuntos de validação e modelagem. Não existe uma regra definida para tal, ficando a cargo do desenvolvedor da inferência escolher o valor de que lhe for mais conveniente.

É interessante também avaliar se o resultado da busca exaustiva de variáveis com número de variáveis fixado por PLS é tão bom quanto o resultado de busca exaustiva global. Comparando-se com o método *stepwise* os resultados foram melhores, porém não foi avaliado se era possível aprimorá-los usando-se números de variáveis diferentes com o método de busca exaustiva.

Referências

BOLF, N.; GALINEC, G.; IVANDIC, M. Soft Sensors for Kerosene Properties Estimation and Control in Crude Distillation Unit. **Chemical and Biochemical Engineering Quarterly**, v. 23, n. 3, p. 277-286, Sep 2009. ISSN 0352-9568

CAO, S. L.; RHINEHART, R. R. An efficient method for online identification of steady-state. **Journal of Process Control**, v. 5, n. 6, p. 363-374, Dec 1995. ISSN 0959-1524

CHATTERJEE, S.; HADI, A. S. **Regression analysis by example**. 4th. Hoboken, N.J.: Wiley-Interscience, 2006. xv, 375 p. ISBN 0471746967

CINAR, A.; PALAZOGLU, A.; KAYIHAN, F. **Chemical process performance evaluation**. Chemical industries 118. Boca Raton, FL: CRC Press, 310 p. 2007

DUARTE, M. B.; MEJÍA, R. I. G.; TRIERWEILER, J. O. **Avaliação de desempenho e ajuste automático de métodos de identificação de estado estacionário**. 2010. (Trabalho de conclusão de curso). Departamento de Engenharia Química, UFRGS

EDGAR, T. F.; HIMMELBLAU, D. M. **Optimization of chemical processes**. New York: McGraw-Hill, 1988. xvii, 652 p. ISBN 0070189919

FACCHIN, S. **Técnicas de Análise Multivariável aplicadas ao Desenvolvimento de Analisadores Virtuais**. 2005. (Dissertação de Mestrado). PPGEQ, UFRGS

FINKLER, T. F. **Desenvolvimento de uma Ferramenta para Obtenção de Modelos Empíricos**. 2003. (Dissertação de Mestrado). PPGEQ, UFRGS

FORTUNA, L.; GRAZIANI, S.; XIBILIA, M. G. Soft sensors for product quality monitoring in debutanizer distillation columns. **Control Engineering Practice**, v. 13, n. 4, p. 499-508, Apr 2005. ISSN 0967-0661

FORTUNA, L. et al. **Soft sensors for monitoring and control of industrial processes**. **Advances in industrial control**. London: Springer, xviii, 270 p. 2007

KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven Soft Sensors in the process industry. **Computers & Chemical Engineering**, v. 33, n. 4, p. 795-814, Apr 2009. ISSN 0098-1354

KANO, M. et al. Inferential control system of distillation compositions using dynamic partial least squares regression. **Journal of Process Control**, v. 10, n. 2-3, p. 157-166, Apr-Jun 2000. ISSN 0959-1524

KANEKO, H.; ARAKAWA, M.; FUNATSU, K. Development of a New Soft Sensor Method Using Independent Component Analysis and Partial Least Squares. **AiChE Journal**, v. 55, n. 1, p. 87-98, Jan 2009. ISSN 0001-1541

LIN, B. et al. A systematic approach for soft sensor development. **Computers & Chemical Engineering**, v. 31, n. 5-6, p. 419-425, May-Jun 2007. ISSN 0098-1354

LJUNG, L. **Identification for Control: Simple Process Models**. 41st IEEE Conference on Decision and Control, 2002. Las Vegas, Nevada USA

MEJÍA, R. I. G.; FARIAS, R. C.; TRIERWEILER, J. O. **Derivatives Estimation Based on Smoothing Techniques 2010**.

MARCON, S. M. et al. Implantação de Sistemas de Controle Avançado em Unidades de Processamento de Gás Natural. **Brazil Automation**. São Paulo 2010.

MOKHATAB, S.; POE, W. A.; SPEIGHT, J. G. **Handbook of natural gas transmission and processing**. Burlington, MA: Gulf Professional Pub.: xxxi, 636 p. p. 2006.

PEARSON, R. K. Exploring process data. **Journal of Process Control**, v. 11, n. 2, p. 179-194, Apr 2001. ISSN 0959-1524.

SALVATORE, L. **Inferência do Teor de Enxofre em Unidades de Hidrotratamento de Diesel Baseada em Modelagem Heurística**. 2007. (Dissertação de Mestrado). Escola de Química, UFRJ

SHLENS, J. **A Tutorial on Principal Component Analysis**: New York University 2009.

THORNHILL, N. F.; CHOUDHURY, M.; SHAH, S. L. The impact of compression on data-driven process analyses. **Journal of Process Control**, v. 14, n. 4, p. 389-398, Jun 2004. ISSN 0959-1524.

VAZ, C. E. M.; MAIA, J. L. P. M.; SANTOS, W. G. D. S. **Tecnologia da Indústria do Gás Natural**. 1ª ed. São Paulo: 2008.

WARNE, K. et al. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. **Engineering Applications of Artificial Intelligence**, v. 17, n. 8, p. 871-885, Dec 2004. ISSN 0952-1976.

WOLD, S.; KETTANEHWOLD, N.; SKAGERBERG, B. Nonlinear PLS Modeling. **Chemometrics and Intelligent Laboratory Systems**, v. 7, n. 1-2, p. 53-65, Dec 1989. ISSN 0169-7439.

WOLD, S.; SJOSTROM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109-130, Oct 2001. ISSN 0169-7439.

Apêndice A

Conjunto de dados da modelagem estacionária

As Tabelas 1.1 e 1.2 mostram os dados usados na modelagem estacionária do estudo de caso.

Tabela A.1: Valores usados na modelagem estacionária para as onze primeiras variáveis

| AI-01.C2 | FIC-03 | PIC-03 | TI-01 | TI-02 | PIC-01 | FIC-01 | TI-03 | TIC-01 | TI-04 | TI-05 |
|----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| -0,8063 | -1,4940 | -1,0354 | -0,3125 | -0,1022 | 0,1927 | -0,5750 | -0,7268 | -1,0468 | -0,6606 | 0,7544 |
| -0,8867 | -1,4796 | -1,1612 | -0,3018 | -0,5295 | 0,3653 | -0,7318 | -0,7268 | -0,7126 | -0,3411 | 0,7023 |
| -1,1563 | -1,4381 | -1,0194 | -0,0712 | -0,1999 | 0,4647 | -0,6245 | -0,1212 | -0,5036 | 0,1530 | 1,0004 |
| -1,7936 | -3,1039 | -1,4243 | 0,2147 | 0,3597 | -0,0209 | -1,7245 | 0,0873 | -0,3081 | 1,1649 | 2,6979 |
| -0,2329 | -3,9968 | -3,4520 | 0,8795 | 0,8531 | -0,1291 | -2,4051 | -0,9844 | -1,3425 | -1,1676 | 0,6399 |
| 0,0241 | -0,7326 | -2,6375 | -1,7820 | -1,8368 | -0,8162 | -1,0761 | -0,2614 | -0,8417 | -0,6797 | 0,1409 |
| 0,1967 | -0,6169 | -2,2875 | -2,0085 | -2,0268 | -0,8167 | -0,9688 | -0,0622 | -0,6559 | -1,0749 | -0,3851 |
| 0,2228 | -0,2089 | -2,0715 | -1,2638 | -1,2934 | -0,7680 | -0,5939 | 0,3987 | -0,9692 | -1,4609 | -1,1055 |
| 0,0241 | -0,7870 | -2,4266 | -1,1750 | -1,1458 | -0,7840 | -1,0081 | 0,4325 | -1,0275 | -1,2241 | -0,2951 |
| -0,1694 | -0,2896 | -1,1864 | -1,6471 | -1,7114 | 0,5286 | -0,6908 | 0,1657 | -0,3527 | -0,8274 | -0,7483 |
| 0,0399 | 0,2279 | -0,7799 | -1,3612 | -1,4596 | -0,6812 | -0,6520 | -0,1896 | -0,7694 | -0,9899 | -0,0411 |
| -0,5191 | -0,0573 | -1,4157 | -1,6807 | -1,9224 | -0,5054 | -1,5795 | 1,0779 | -0,3492 | -0,1541 | 0,7145 |
| -0,4860 | -0,4898 | -1,1521 | -1,5184 | -1,5145 | -0,5072 | -1,3062 | -0,5251 | -1,6883 | -0,4643 | -0,1079 |
| -0,4238 | 1,1752 | -0,2648 | -2,0566 | -2,1979 | -0,1956 | -0,2592 | -0,9024 | -0,0298 | 0,1699 | 0,8374 |
| 0,3423 | -0,2906 | -0,9803 | -0,7479 | -0,7407 | 0,7515 | -1,7096 | 1,2032 | -1,2511 | -1,5452 | -0,5976 |
| -0,1344 | 0,0045 | -1,5192 | -1,1035 | -0,5986 | -0,5181 | -1,0510 | 0,8394 | -1,1317 | -0,7847 | -0,9441 |
| -0,5793 | -0,0630 | -0,5805 | -1,0890 | -1,0907 | -0,4601 | -0,9146 | 0,9626 | -0,4408 | -0,2022 | 0,0690 |
| -0,5520 | 0,3768 | -0,2889 | -1,0890 | -1,0780 | -0,1876 | -0,8811 | 0,9466 | -0,3672 | -0,2393 | -0,3950 |
| -1,2137 | -0,8114 | -1,4035 | -0,6351 | -0,6526 | -0,7773 | -1,3947 | 1,3681 | -0,0781 | 0,2018 | 1,2650 |
| -0,5122 | -0,0771 | -1,7018 | -0,8614 | -0,8878 | -0,1544 | -1,2340 | 1,4077 | 0,0621 | 0,3129 | 1,5055 |
| -0,5545 | -0,8424 | -1,5119 | -1,2487 | -1,2294 | -0,9742 | -1,3754 | 1,1867 | -0,3483 | -0,7055 | -0,0965 |
| -0,1904 | -0,7984 | -1,7172 | -0,7194 | -0,7423 | -0,6491 | -1,7672 | 0,6640 | -1,2004 | -0,9158 | 0,0657 |
| -0,4164 | -0,4140 | 0,0880 | -0,4444 | -0,3443 | -0,4861 | -1,1480 | 0,2904 | -0,7172 | -0,6487 | 0,2278 |
| -0,5237 | -0,4999 | -0,5418 | -0,8167 | -0,6283 | 0,5651 | -0,6787 | 0,2616 | -0,2283 | 0,0953 | 1,5614 |
| -0,2884 | 0,0448 | -0,3939 | -1,4064 | -1,3128 | -0,3857 | -0,9732 | 0,4869 | -0,4105 | -0,7399 | 0,5614 |
| -0,9081 | -0,7099 | -3,2453 | -0,7957 | -0,8258 | 0,6348 | -1,4660 | 0,9952 | -0,0304 | 0,4063 | 1,1647 |
| -1,0943 | 2,8957 | -0,1102 | -1,5566 | -1,7151 | 0,2593 | 0,0240 | 2,5399 | 0,2892 | 0,1730 | 0,9398 |
| -0,1334 | 1,7301 | -0,5093 | -1,1296 | -1,0805 | -0,3735 | -0,2707 | -0,0257 | -0,0967 | -0,5216 | -0,3223 |
| -0,0988 | -0,1032 | -0,4690 | -1,4168 | -1,6641 | -0,5075 | -1,2711 | 0,3247 | -0,8491 | -0,8665 | -0,1857 |
| -0,0994 | 1,1272 | -0,5850 | -2,1843 | -2,0354 | -0,2277 | -1,0767 | 0,4200 | -1,2550 | -1,5873 | -0,8562 |
| -0,6524 | 0,0499 | -0,6048 | -0,3664 | -0,3673 | -0,7438 | -1,1337 | 2,2909 | 0,7980 | 0,6609 | 0,3292 |
| -0,6676 | -0,4694 | -0,5812 | -1,3227 | -1,2203 | -0,9597 | -1,2443 | 1,4043 | 0,1714 | 0,0428 | 0,1669 |
| -0,7829 | -1,3139 | -0,5937 | -1,1418 | -1,1901 | -0,9302 | -1,7866 | 1,2991 | 0,1994 | 0,2386 | 0,4658 |
| -0,7291 | -0,8065 | -0,9844 | -1,1585 | -1,2199 | -1,0388 | -1,3069 | 2,1017 | -0,5109 | -0,3600 | 0,3849 |
| -0,0140 | 1,0757 | -0,3960 | -0,8365 | -0,7764 | -0,2083 | -0,2721 | 0,0507 | -0,5960 | -0,4044 | -0,2881 |
| -0,5810 | -0,7124 | -0,3728 | -1,2544 | -1,1260 | -0,0316 | -1,1966 | -0,6841 | -0,9894 | -0,6885 | 0,3095 |
| 0,7650 | -0,0047 | -0,6413 | -1,8680 | -1,7537 | 1,0357 | -1,3482 | 0,0022 | -0,8172 | -1,5453 | -1,0593 |
| 0,8215 | 1,5174 | -0,2528 | -1,2259 | -1,1821 | 0,2973 | -1,0988 | -0,4697 | -1,7409 | -1,9283 | -2,7102 |
| 0,9236 | 0,3862 | -0,5213 | -1,3061 | -1,3476 | -0,0051 | -1,0502 | -0,7074 | -1,4343 | -1,4141 | -1,5523 |

| | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0,9270 | 0,3454 | -0,1893 | -1,2256 | -1,4821 | -0,1388 | -1,1734 | -0,7775 | -1,7887 | -1,8371 | -2,0897 |
| 0,2549 | -0,1138 | -0,5230 | -1,0589 | -1,4498 | 0,5591 | -1,1458 | -0,4348 | -1,3528 | -1,0271 | -0,3915 |
| 0,9721 | -0,1277 | -0,7778 | -1,2324 | -1,2095 | 0,4643 | -1,3205 | -1,1597 | -1,4880 | -1,7558 | -1,6674 |
| -0,3074 | 0,0997 | 0,1512 | -0,5127 | -0,4287 | 2,3031 | -1,3011 | 0,2954 | -0,2298 | -0,0711 | 0,8590 |
| 0,8317 | -1,1545 | -0,6462 | -0,6087 | -0,7972 | 0,4444 | -1,9797 | 0,5028 | -1,7957 | -2,3910 | -1,3313 |
| 0,6185 | -0,6917 | 0,1657 | 0,1283 | 0,2703 | 1,9350 | -0,2052 | -1,4519 | -0,4491 | -0,4040 | 0,1839 |
| -0,1979 | -0,7005 | 0,2300 | 0,7762 | 1,1394 | 2,4444 | 0,0892 | -0,6090 | -0,1788 | 0,3579 | 0,6544 |
| -0,2455 | 0,4479 | 0,9428 | 1,7130 | 1,7890 | 1,9393 | 0,4747 | 1,8370 | 1,5912 | 0,9874 | 0,8164 |
| 1,4046 | -0,4092 | 0,4243 | 0,4088 | 0,3423 | 0,9081 | -0,1855 | 0,3180 | -0,9429 | -1,9099 | -2,2491 |
| 0,6847 | 0,2402 | 0,4481 | 0,7609 | 0,6974 | 1,0225 | 0,4410 | 0,3122 | -0,1381 | -1,0315 | -1,7215 |
| 3,3797 | -0,1179 | 0,8524 | 1,9381 | 1,8595 | 2,6995 | -0,0727 | -0,0211 | -0,1715 | -0,8493 | -0,5478 |
| 2,4270 | -0,9010 | -0,8052 | 1,8251 | 1,6620 | 0,9324 | -0,3004 | -1,1683 | -0,9136 | -0,7105 | -0,1056 |
| 1,7536 | -0,5714 | -0,5988 | 1,7910 | 1,6057 | 3,1315 | -0,0486 | 1,4535 | 0,6388 | 0,4170 | 0,7831 |
| 1,7038 | -0,8034 | 0,4185 | 0,9012 | 1,0730 | 1,0199 | -0,0051 | 0,4691 | 0,0678 | 0,1665 | 0,3953 |
| 1,1050 | 0,1105 | 0,4889 | 0,8363 | 0,7482 | 0,7749 | 0,5396 | -0,5646 | 0,0320 | 0,1394 | 0,5431 |
| 1,3331 | 0,3290 | 0,6212 | 0,9114 | 0,8230 | 1,0099 | 0,7722 | -0,2380 | -0,1640 | -0,0394 | 0,7826 |
| 1,2746 | -0,6447 | 0,4092 | 0,5982 | 0,5063 | 0,6238 | 0,0402 | 0,9564 | 0,2875 | 0,3483 | -0,0399 |
| 1,3538 | -0,8700 | 0,3577 | 0,7359 | 0,4821 | 0,8363 | -0,2247 | 0,5493 | 0,4241 | 0,3867 | 0,0889 |
| 1,2401 | -0,7089 | 0,3658 | 0,0462 | 0,2993 | 0,6068 | -0,0748 | -0,9747 | -0,8872 | -0,5745 | -1,1025 |
| 0,9881 | -0,2889 | 0,4836 | 0,7738 | 0,8398 | 0,9004 | -0,2107 | 0,4655 | -0,0964 | -0,3039 | -1,4482 |
| 0,6073 | -0,5854 | 0,8309 | 1,5416 | 1,4232 | 2,6753 | -0,4804 | 1,1740 | 0,3349 | 0,4950 | 0,5739 |
| 0,6607 | -0,4737 | 0,4917 | 0,6846 | 0,4971 | 0,6236 | 0,1442 | -0,4831 | -0,1278 | -0,3645 | -0,6577 |
| 1,1753 | -0,2061 | 0,6729 | 0,6082 | 0,7177 | -0,0252 | 0,1403 | -0,9700 | -0,1020 | 0,4662 | 0,0316 |
| 1,3596 | 0,0847 | 0,8598 | 0,6514 | 1,1632 | 1,4836 | 0,3356 | -0,3820 | -0,0778 | -0,0628 | -0,1560 |
| 0,5636 | 0,2249 | -0,0907 | 0,8032 | 0,4146 | -0,7017 | 0,5980 | -0,9497 | -0,1381 | 0,0607 | -0,4005 |
| 1,8722 | 0,1970 | 2,7257 | 1,9125 | 1,8669 | 0,6261 | -0,0541 | -0,4028 | -0,2620 | -0,0098 | -0,2143 |
| 0,3156 | -0,3844 | -1,4532 | 0,6039 | 0,4986 | -0,9482 | 0,3265 | -1,0503 | -0,0982 | 0,4119 | -0,2882 |
| -0,1742 | 0,4293 | 2,3423 | 0,8563 | 0,5757 | -0,5696 | 1,0283 | -1,0593 | -0,0878 | 0,3680 | -0,3987 |
| 0,3263 | 0,3600 | -0,0100 | 0,7063 | 0,4555 | 0,2554 | 0,3942 | -0,6486 | -0,1380 | 0,0292 | -0,1027 |
| 3,0808 | -0,5607 | 0,4914 | 1,8364 | 1,7571 | -0,9523 | 0,3582 | -0,3208 | -0,1424 | -0,1531 | 0,3180 |
| 2,9149 | -0,4342 | 0,5716 | 0,2969 | 0,1917 | -0,7976 | 0,0909 | -0,7166 | -0,8013 | -0,0584 | -0,3030 |
| 1,3795 | -3,0493 | 0,4015 | 1,0736 | 1,0061 | -0,0939 | -0,6547 | -0,2205 | 0,1720 | 0,5935 | 1,1073 |
| 0,9482 | -1,2226 | 0,1236 | 0,9628 | 0,6785 | -0,6741 | -0,0719 | -1,3490 | -0,9644 | -0,8861 | -0,4364 |
| -0,1583 | 0,6102 | 0,5217 | -0,4789 | -0,4289 | -0,1995 | 0,7965 | -0,9000 | 0,2542 | 0,0943 | -0,4513 |
| -0,0385 | 0,7552 | 0,5433 | -0,8128 | -0,7939 | -0,1424 | 0,8745 | -0,7961 | 0,5678 | 0,3697 | -0,1178 |
| -0,3418 | 1,6652 | 0,7371 | -0,2627 | -0,2108 | 0,5169 | 1,2193 | -0,1785 | 0,4564 | 0,1667 | 0,3175 |
| -0,4493 | -2,3375 | 0,3350 | -0,0110 | 0,0379 | -1,8627 | -0,4164 | -3,0323 | -2,0878 | -1,4623 | 0,4853 |
| -0,2859 | 0,2274 | 0,4439 | 0,0770 | 0,1302 | -0,3457 | 0,6593 | 0,4892 | 0,6726 | 0,1403 | -0,3512 |
| -0,9607 | 0,7239 | 0,4092 | 0,9578 | 0,9075 | -0,1063 | 1,1105 | 0,5407 | 0,8949 | 0,9988 | 0,1433 |
| -0,8721 | 0,3122 | 0,4716 | 0,5169 | 0,5403 | -0,2142 | 0,9454 | -0,7360 | 0,6324 | 0,8606 | 0,2954 |
| -0,0895 | 0,5322 | 0,5108 | 0,0449 | 0,0693 | -0,1441 | 0,7477 | -0,1785 | 0,3053 | -0,2538 | -0,8090 |
| -0,6477 | 0,2001 | 0,5481 | 0,0472 | 0,0453 | -0,4135 | 0,8571 | -1,7104 | -0,4055 | -0,1266 | -0,4641 |
| -0,7783 | 0,1722 | 0,4859 | -0,0071 | 0,0790 | -0,1474 | 0,7588 | -0,5613 | 0,7914 | 0,6310 | 0,3280 |
| -0,2319 | 0,2996 | 0,4327 | 0,1340 | 0,1428 | 0,0157 | 0,8552 | -1,5060 | -0,7741 | -0,7575 | -1,0566 |
| 0,4301 | 0,0384 | 0,6960 | -0,0381 | -0,0418 | 0,1122 | 0,5816 | -1,6005 | -1,1012 | -0,9636 | -1,1827 |
| -0,1254 | 0,5584 | 1,1649 | 0,1735 | 0,2931 | 1,2402 | 0,7363 | 0,0351 | 0,4497 | 0,2837 | 0,8843 |
| 0,1394 | 0,8529 | 0,6395 | 0,2084 | 0,2169 | 0,2287 | 1,2764 | -1,3329 | -0,3854 | -0,6233 | -0,4176 |
| 0,4469 | 0,4293 | 0,9622 | 0,0916 | 0,1935 | 1,8387 | 0,6961 | -0,9740 | -0,1375 | -0,3472 | -0,9345 |
| 0,1175 | -0,3219 | 0,4187 | -0,2609 | -0,1974 | -0,4789 | 0,3761 | -0,8900 | 0,4348 | 0,2660 | -0,8885 |
| 0,0658 | -0,1752 | 0,4043 | 0,5680 | 0,6043 | -0,1844 | 0,7154 | -0,8357 | 0,9218 | 0,8041 | -0,3380 |
| 0,4393 | 0,2378 | 0,5278 | -0,2858 | -0,2073 | -0,2741 | 0,5341 | -0,7949 | 0,2501 | 0,3077 | -1,2577 |
| -0,5235 | -0,5967 | 0,3645 | 0,2366 | 0,2500 | -0,5491 | 0,3840 | -1,5184 | -0,0618 | 0,0637 | 0,3487 |
| -0,1193 | 0,2220 | 0,5356 | 0,4282 | 0,4388 | -1,6496 | 1,0694 | -1,1350 | 0,3371 | 0,3875 | -0,0710 |
| -0,1728 | -0,1614 | 0,5521 | 0,6701 | 0,7105 | -1,3146 | 0,7609 | -1,3685 | 0,3771 | 0,3586 | -0,2697 |

| | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| -0,8314 | 1,0440 | 0,7352 | 1,4596 | 1,4047 | -1,2946 | 2,1548 | -2,0428 | 0,5742 | 1,3970 | 0,9901 |
| -0,1616 | 0,6087 | 0,5531 | 0,4040 | 0,4223 | -1,3184 | 1,3034 | -2,0164 | -0,1249 | 0,1182 | -0,5285 |
| -0,7166 | 0,8850 | 0,5512 | 0,3205 | 0,3903 | -1,0188 | 1,3559 | -1,2190 | 0,3151 | 0,6905 | 0,2724 |
| -1,1935 | 0,2583 | 0,9882 | 1,2343 | 1,2765 | 0,5630 | 0,6943 | 0,5911 | 1,2266 | 1,0859 | 1,4128 |
| 0,8026 | -0,0073 | 0,6441 | -0,5367 | -0,4497 | -1,4883 | 0,0983 | -0,4423 | 0,3291 | -0,0765 | -1,5413 |
| -0,8863 | 0,2496 | 0,7076 | 0,5780 | 0,5718 | -1,3313 | 0,7567 | -0,2004 | 0,5299 | 0,4631 | -0,0211 |
| -0,2334 | 0,1168 | 0,8258 | 1,0072 | 1,0415 | 0,1273 | 0,4238 | 0,9586 | 0,9832 | 0,6647 | 0,7373 |
| -1,1052 | 1,5552 | 1,6139 | 1,2702 | 1,3705 | 0,3320 | 1,4735 | 0,4161 | 0,4465 | 0,8595 | 1,5858 |
| -1,2787 | -0,0758 | 0,7294 | 1,7024 | 1,6617 | -1,3100 | 0,6419 | 1,4687 | 2,1080 | 1,6445 | 0,4509 |
| -1,7808 | 0,1782 | 0,5844 | 1,0493 | 1,0792 | 0,1914 | 0,8817 | 1,3588 | 3,0172 | 2,6989 | 2,6023 |
| -2,0206 | -0,0862 | 0,1805 | 1,1159 | 1,0778 | -1,1842 | 0,8284 | 0,7354 | 2,5633 | 2,3473 | 1,1756 |
| -1,0693 | 0,1782 | 0,5375 | 0,5917 | 0,5667 | -1,0461 | 0,6083 | 1,8533 | 3,1431 | 2,7580 | 1,2529 |
| -0,6448 | -0,0568 | 0,5554 | 0,5020 | 0,5245 | -1,3334 | 0,4754 | 0,9168 | 2,3575 | 2,0701 | 0,5285 |
| -0,2808 | 0,4315 | 1,0055 | 1,4156 | 1,3524 | -0,5943 | 0,6627 | 1,5762 | 2,1675 | 1,6347 | -0,1171 |
| 0,3480 | -0,2976 | 0,0406 | 0,5235 | 0,5356 | 1,4860 | -0,2040 | 0,0462 | -0,0187 | -0,3543 | -1,1657 |
| -1,1311 | 0,5839 | 0,9897 | 0,6287 | 0,6482 | -1,0774 | 1,2210 | -0,2873 | 1,3474 | 1,6781 | 1,0244 |
| -0,7690 | 0,4955 | 0,7114 | 0,0651 | 0,1490 | -0,9997 | 0,9190 | -0,1053 | 1,6357 | 1,8875 | 0,6236 |
| -0,3131 | 0,1110 | 0,7633 | 1,1677 | 1,2159 | -0,4632 | 0,5274 | 0,3122 | 0,0907 | -0,5968 | -0,9464 |
| -0,1794 | 0,0957 | 0,2638 | 0,0472 | 0,1213 | -1,2043 | 0,6333 | 0,1425 | 0,7290 | 0,4707 | -0,3605 |
| 0,3934 | 0,3614 | 0,0282 | -0,5818 | -0,5117 | -0,8428 | 0,4213 | 0,2008 | 0,7508 | 0,8515 | -0,2725 |
| 2,6552 | -0,5095 | -0,3866 | -0,7227 | -0,5945 | 1,2680 | 0,0333 | -1,3749 | -2,5292 | -2,9882 | -4,0180 |
| -1,0687 | 1,1461 | 0,7874 | 0,4124 | 0,4607 | 2,2911 | 1,2529 | 1,2967 | 1,3993 | 1,0084 | 0,9706 |
| -0,0614 | 0,6674 | 1,0015 | -0,1560 | -0,0989 | 1,4526 | 0,9001 | 0,2380 | 0,1016 | -0,3541 | -0,3353 |
| -1,3407 | 0,1733 | -0,0651 | 0,4847 | 0,5126 | -0,9030 | 0,8815 | 0,1191 | 1,2869 | 1,1884 | 0,9772 |
| -1,2918 | 1,6855 | 0,8621 | 1,2856 | 1,3462 | 0,5493 | 1,7654 | 0,8786 | 1,1505 | 0,8633 | 1,0838 |
| -0,9098 | 1,8225 | 0,7013 | 0,5787 | 0,6292 | 0,9441 | 1,4670 | 0,9969 | 1,4227 | 1,2394 | 1,7620 |
| -1,5404 | 3,6337 | 0,8085 | 0,7731 | 0,7396 | 0,4585 | 2,7565 | -0,0203 | 0,3795 | 1,5054 | 2,1494 |
| -1,1323 | 1,6987 | 0,7941 | 0,4327 | 0,4399 | -0,9332 | 1,6474 | 0,4705 | 0,6494 | 1,1152 | 0,1132 |
| -1,0159 | 1,0721 | 0,9091 | 0,1521 | 0,2297 | -1,0364 | 1,1957 | 1,1211 | 1,0346 | 1,1773 | 0,2489 |
| 0,0932 | 1,5431 | 1,2550 | 0,1473 | 0,1957 | -0,5440 | 1,5777 | 0,4417 | 0,5765 | 0,6061 | -0,0109 |
| 0,1290 | 0,3443 | 0,5750 | 0,4165 | 0,4330 | -1,2359 | 0,6643 | 1,3511 | 1,5235 | 0,8820 | -1,1372 |

Tabela A.2: Valores usados na modelagem estacionária para as dez últimas variáveis e a variável de saída AI-02.C2

| TI-06 | TI-07 | TI-08 | PI-01 | PDI-01 | TIC-02 | TI-10 | TI-11 | PIC-02 | FIC-02 | AI-02.C2 |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| 0,8773 | 0,5894 | 1,5192 | -1,1253 | 0,3100 | 0,0261 | -0,1284 | -0,1133 | -0,4876 | -0,0789 | -0,6826 |
| 0,9283 | 0,5691 | 0,9388 | -1,5328 | 0,0840 | 0,5533 | -0,5514 | -0,1482 | -0,3447 | 0,6845 | -0,7265 |
| 1,2112 | 0,9882 | 1,1553 | -1,5306 | -0,0711 | 1,1460 | -0,3761 | 0,0255 | -0,3998 | 0,8392 | -1,0024 |
| 2,9671 | 2,5894 | 2,4476 | -1,0176 | 0,1003 | 0,9773 | -1,4955 | 0,4031 | -0,8150 | 0,1874 | -1,2959 |
| 0,9664 | 0,4897 | 1,4222 | -1,2413 | -0,4622 | 0,1765 | -0,3735 | -1,2136 | -0,3522 | 0,4916 | -0,3478 |
| 0,1770 | 0,0022 | 1,4977 | -1,0425 | 0,5083 | 0,8366 | 0,0067 | -0,6457 | -0,3451 | 2,1816 | -0,1712 |
| -0,2328 | -0,4142 | 0,7789 | -1,0890 | 0,7995 | 0,8420 | -0,0516 | -0,6658 | -0,3954 | 2,1445 | -0,0058 |
| -0,9306 | -1,0687 | 1,5057 | -1,0530 | 0,5489 | -0,8359 | -1,3519 | 0,1677 | -0,3497 | 1,6832 | -0,4454 |
| -0,1003 | -0,2718 | 1,4731 | -1,4008 | 0,5162 | 0,3515 | -1,1681 | 0,5413 | -0,3666 | 1,6950 | -0,6457 |
| -0,5467 | -0,6495 | 0,4788 | -0,4419 | 1,6388 | -1,4424 | -0,0895 | 0,4332 | -0,3641 | 1,4386 | -0,6611 |
| -0,0237 | -0,4348 | 1,1259 | -0,8792 | 0,3896 | -0,6438 | 0,7003 | 0,8238 | -0,4009 | 1,1231 | -0,5627 |
| 0,6833 | 0,6278 | 1,1985 | -1,2065 | 0,3090 | 0,9318 | -0,6437 | -0,3879 | -0,3783 | 1,1039 | -0,4019 |
| 0,3098 | -0,0815 | 1,2336 | -1,2053 | 0,2855 | -0,5085 | -0,2250 | -0,1493 | -0,3467 | 1,6067 | -0,6924 |
| 0,7628 | 0,8610 | 1,0480 | -1,1915 | 0,4843 | 2,0392 | 0,8315 | 1,0392 | -0,4528 | 1,4556 | -0,4282 |
| -0,4796 | -0,6295 | -0,4406 | -0,8308 | 1,9724 | -1,0584 | -1,0234 | -1,2069 | -0,3521 | 2,0792 | 0,1819 |
| -0,7545 | -0,8191 | 1,3325 | -1,2138 | 0,8338 | -0,6317 | 0,3844 | -0,1275 | -0,3650 | 1,8278 | -0,2758 |

| | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 0,0325 | 0,0263 | -0,6958 | -1,1483 | 0,1826 | 0,0711 | -0,4861 | 0,1471 | -0,3753 | -0,3821 | -0,5336 |
| -0,3629 | -0,3445 | -0,7162 | -1,3132 | 0,5102 | -1,2790 | -0,3693 | -0,3655 | -0,3918 | -0,4521 | -0,4558 |
| 1,5198 | 1,2374 | 0,6661 | -1,5341 | 0,9109 | 0,0652 | -0,2261 | 0,7216 | -0,4154 | 0,2053 | -1,1590 |
| 1,4550 | 1,4390 | 0,5176 | -1,2949 | 0,7427 | 1,2535 | 0,3329 | -0,0232 | -0,3686 | 0,1961 | -0,5934 |
| 0,1488 | -0,0420 | 0,1836 | -1,4283 | -0,2861 | -0,7953 | -2,0072 | -0,7802 | -0,5194 | 0,6842 | -0,3174 |
| 0,0724 | 0,0552 | 1,0882 | -1,2126 | 0,3314 | 0,5275 | -1,0356 | -1,0158 | -0,3941 | 1,6995 | -0,0403 |
| 0,2696 | 0,2060 | 1,9896 | -1,1925 | 0,5995 | 1,3106 | 1,3389 | -0,7685 | -0,3446 | 1,7326 | -0,2906 |
| 1,3488 | 1,4629 | 0,9793 | -0,1390 | 1,3519 | -0,1870 | -0,4217 | -0,6238 | -0,3286 | 2,0935 | -0,2808 |
| 0,5020 | 0,9406 | 0,0400 | -0,6856 | -1,9138 | -1,6516 | 0,7947 | -1,3248 | -0,3523 | 2,1180 | -0,0007 |
| 1,4029 | 1,2636 | 1,1213 | -0,3967 | 1,6663 | -0,4794 | 0,3969 | -0,3462 | -0,3425 | 1,4181 | -0,7914 |
| 1,3314 | 1,3436 | -0,8051 | -1,0971 | 0,5128 | 0,0540 | 0,7276 | 1,5869 | -0,3416 | 0,6812 | -0,9643 |
| -0,3583 | -0,2341 | -0,3625 | -1,0636 | 0,2832 | -1,1159 | 0,4432 | 0,0654 | -0,3657 | 0,3616 | -0,2113 |
| -0,1430 | -0,3006 | 0,8980 | -1,0007 | 0,1802 | 0,5163 | 0,3370 | -0,8340 | -0,3715 | 1,1581 | 0,0362 |
| -0,9468 | -0,7343 | 0,9287 | -0,9486 | 0,2270 | -0,5501 | 0,2373 | -0,6640 | -0,3232 | 1,0126 | 0,0615 |
| 0,3812 | 0,4040 | 0,3621 | -0,9599 | -0,0001 | 0,1650 | -1,2223 | 0,6511 | -0,3202 | 0,3189 | -0,6033 |
| 0,3839 | 0,3209 | 0,8774 | -1,1471 | 0,2085 | -0,6938 | -0,5025 | 0,1840 | -0,3398 | -0,0895 | -0,6201 |
| 0,9255 | 0,5392 | 0,0459 | -1,1211 | -0,0555 | 0,8518 | 0,6064 | 0,4627 | -0,4297 | -1,4540 | -0,6766 |
| 0,5457 | 0,3979 | 0,1245 | -1,2752 | 0,0156 | 0,7361 | 0,5505 | 1,7597 | -0,3722 | -1,3109 | -0,7552 |
| -0,2242 | -0,2435 | -0,1773 | -0,9612 | 0,3617 | -0,0881 | -0,4175 | 0,5479 | -0,4291 | -0,9001 | -0,2831 |
| 0,6324 | 0,2689 | 0,1615 | -0,9201 | -0,0019 | 0,9834 | 0,6212 | 0,2918 | -0,2596 | -0,5589 | -0,6403 |
| -1,0311 | -1,0477 | -0,0611 | -0,0902 | 1,6519 | -0,8690 | 0,9810 | -1,9151 | -0,3758 | 0,9989 | 0,5286 |
| -2,5189 | -2,6180 | 0,1414 | -0,8153 | 0,2755 | -1,0592 | 0,8577 | -1,0081 | -0,2907 | 0,8639 | 0,7304 |
| -1,3767 | -1,8197 | -0,2034 | -0,8974 | 0,5367 | 0,7115 | 0,7040 | -1,2613 | -0,3588 | 0,8338 | 0,6686 |
| -1,9436 | -2,0499 | 0,0824 | -1,1909 | 0,7331 | 0,1721 | 0,5946 | -0,3115 | -0,3742 | 0,8219 | 0,5402 |
| -0,3590 | -0,4263 | -0,7899 | -0,7411 | 0,3672 | -1,9208 | -2,0316 | -1,5899 | -0,3788 | 0,9250 | 0,1938 |
| -1,4841 | -1,6043 | 1,5940 | -0,3921 | 0,3720 | 0,5027 | 0,9776 | -1,2000 | -0,3295 | 0,4012 | 0,8402 |
| 0,8197 | 0,8047 | 0,0660 | -0,5856 | -1,4781 | -0,6666 | 0,9696 | -0,8160 | -0,3920 | 1,4881 | -0,1060 |
| -1,1606 | -1,3400 | 1,4365 | -0,7902 | 0,6549 | -0,3367 | -0,5307 | -1,8796 | -0,3253 | 0,9896 | 0,7869 |
| 0,0650 | 0,1034 | -0,9331 | -0,5264 | -2,7218 | 0,6602 | 1,5495 | -0,9159 | -0,6180 | -2,8687 | 0,3252 |
| 0,5470 | 0,3645 | -0,4752 | -0,5406 | -2,7980 | 0,3867 | 1,1708 | -0,6529 | -0,6549 | -1,8423 | 0,0486 |
| 0,9276 | 1,0863 | -2,0230 | -0,2900 | -2,4746 | -1,6542 | 1,2529 | 0,1159 | -0,6831 | 0,0311 | -0,9951 |
| -2,2980 | -2,1883 | -0,1985 | -0,0489 | 0,4196 | -0,9272 | -1,3027 | -1,3645 | -0,5710 | -0,5606 | 1,3367 |
| -1,9245 | -1,6089 | 0,8159 | -0,1112 | 0,4093 | -0,9360 | -0,5754 | 0,6937 | -0,6756 | -0,6392 | 0,2028 |
| -0,5423 | -0,5905 | -0,6532 | -0,1678 | -2,4991 | -0,7317 | 0,3096 | -0,9822 | -0,8167 | -1,1622 | 3,2114 |
| -0,0888 | -0,1357 | 0,5067 | -0,1222 | -2,6278 | 0,2026 | -1,3180 | -0,9438 | -0,7920 | -1,0588 | 2,2704 |
| 0,7557 | 1,0030 | 0,1641 | -0,3455 | -2,5469 | 0,0917 | -0,1121 | -0,2123 | -0,7993 | -0,6147 | 2,1137 |
| 0,1845 | 0,3252 | 0,8215 | -0,2553 | -2,5487 | 0,4154 | 0,8616 | -0,2378 | -0,7551 | -1,0648 | 1,8859 |
| 0,4664 | 0,6676 | -0,7339 | 0,3256 | 0,0080 | 0,2073 | 0,2657 | 0,1919 | -0,9823 | -1,0875 | 1,1662 |
| 0,6814 | 0,7578 | -0,7449 | 0,3211 | -0,1787 | -0,0860 | 0,7276 | 0,3181 | -0,7268 | -1,1642 | 1,2446 |
| -0,2452 | 0,0995 | -0,0039 | 0,2062 | 0,1181 | -0,2280 | 0,3494 | 0,1150 | -0,7715 | -0,9154 | 1,4657 |
| 0,0200 | 0,1822 | 0,4329 | 0,0002 | 0,0931 | 0,3583 | 1,2030 | -1,0682 | -0,8105 | -0,9847 | 1,4203 |
| -1,0985 | -1,1944 | 1,1634 | 0,2711 | -0,0082 | 1,2812 | 0,3173 | -0,8497 | -0,7877 | -0,2763 | 1,5228 |
| -1,4004 | -1,3382 | 0,4469 | 0,1536 | 0,0816 | 1,7501 | 0,8075 | 0,2057 | -0,8043 | -1,0165 | 1,1418 |
| 0,5516 | 0,5817 | -0,4680 | 0,5383 | 1,1036 | 1,3807 | -0,4031 | -0,1360 | -0,7897 | -0,4572 | 0,7357 |
| -0,6550 | -0,5629 | 0,0151 | 0,1850 | -0,1291 | 0,0088 | -0,3378 | -0,7500 | -0,6963 | -0,8369 | 0,7775 |
| -0,2088 | -0,6961 | 0,6193 | 0,0202 | -2,2623 | 1,2587 | 0,4238 | -1,3137 | -0,8334 | -0,8556 | 1,3256 |
| -0,6622 | -0,7504 | -0,1926 | 0,0668 | -2,6898 | 0,2298 | 0,1202 | -0,7884 | -0,8571 | -0,7235 | 1,3825 |
| -0,4182 | -0,3624 | 0,0378 | -0,1573 | -0,2456 | 0,3951 | 0,2677 | -0,3958 | -0,7753 | -0,2240 | 0,6662 |
| -0,3885 | -0,6547 | -0,9092 | -0,0682 | -2,7287 | -0,9587 | 0,0571 | -1,8791 | -0,7783 | -0,2298 | 1,8902 |
| -0,3336 | -0,2127 | 0,4135 | 0,0095 | -0,1287 | 0,8088 | -1,1551 | -0,2308 | -0,7174 | -0,3117 | 0,2954 |
| -0,3762 | -0,3139 | -0,1988 | 0,0038 | -0,3099 | 0,2254 | -0,7165 | 0,4714 | -0,7659 | -0,3113 | -0,1312 |
| -0,2773 | -0,2091 | 0,5019 | -0,0103 | -2,0645 | 1,1610 | 1,5339 | -0,6076 | -0,8273 | -0,3517 | 0,3821 |
| 0,1940 | 0,2901 | 0,9466 | -0,0507 | -2,8613 | 0,4146 | 0,4187 | -1,0442 | -0,8232 | -1,6126 | 3,6021 |
| -0,5264 | -0,7990 | 0,3429 | 0,1296 | 0,1724 | -0,8623 | -0,1531 | -1,6748 | -2,5022 | -0,3725 | 3,4979 |

| | | | | | | | | | | |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1,4661 | 1,0694 | -1,2858 | 0,3589 | 0,9267 | -0,9360 | -0,9595 | -0,7801 | -0,7600 | -0,5967 | 1,3558 |
| -0,2864 | -0,2889 | 0,4891 | 0,4509 | 0,0016 | -1,1684 | 0,0207 | -1,0759 | -0,8217 | 0,0829 | 0,6931 |
| -0,5265 | -0,4211 | 0,2770 | -0,1442 | -0,1721 | 0,5210 | 0,8948 | -0,2390 | -0,7568 | 0,2316 | -0,1348 |
| -0,3019 | -0,0757 | 0,1332 | -0,1815 | 0,0585 | 0,8446 | 0,8050 | -0,0229 | -0,7645 | 0,1377 | -0,0139 |
| 0,0977 | 0,3833 | -0,5979 | -0,1603 | 0,3869 | 0,1813 | 0,4515 | 0,3350 | -0,8532 | -0,0899 | -0,2607 |
| 0,5816 | 0,4051 | 1,0132 | -0,3516 | -0,3005 | 0,4934 | 0,9894 | -1,7427 | -0,7014 | -0,6222 | -0,1815 |
| -0,3977 | -0,2767 | -2,3366 | -0,2271 | 0,0321 | -2,9758 | -1,9576 | 0,2987 | -0,8162 | -0,6456 | -0,2072 |
| 0,2019 | 0,2656 | -2,6836 | -0,2375 | 0,0645 | -3,5678 | -2,5161 | 0,1185 | -0,8224 | -0,6227 | -0,9666 |
| 0,3082 | 0,3879 | -1,6052 | -0,2854 | -0,0797 | -2,3084 | -1,2385 | -0,1338 | -0,8546 | -0,6230 | -0,7964 |
| -0,8658 | -0,7117 | -0,0288 | -0,3150 | 0,0857 | 0,1929 | -0,7247 | 0,2478 | -0,7859 | -0,6451 | -0,0756 |
| -0,4235 | -0,3470 | 0,7552 | -0,1477 | -0,3429 | 0,2240 | -0,8572 | 0,4387 | -0,9570 | -0,5714 | -0,6811 |
| 0,2673 | 0,4151 | 0,3941 | -0,4353 | 0,0299 | 1,1063 | 0,1001 | 1,6954 | -0,7529 | -0,5541 | -0,9150 |
| -1,0941 | -0,9603 | 0,0738 | -0,2660 | 0,2263 | -0,0534 | -0,5442 | 0,8731 | -0,8887 | -0,7105 | 0,2112 |
| -1,2487 | -1,1201 | 0,6170 | -0,3027 | 0,3978 | 0,5646 | 1,0480 | 0,1151 | -0,8331 | -0,6175 | 0,5697 |
| 0,6601 | 0,8928 | 0,3621 | -0,2236 | 0,5857 | 1,3364 | 1,1641 | -0,6521 | -0,8017 | -0,2897 | 0,0280 |
| -0,5244 | -0,3773 | 0,2760 | -0,4290 | -0,1146 | 0,2528 | -0,5561 | -0,5963 | -0,7600 | -0,7706 | 0,2541 |
| -1,0193 | -0,9563 | -0,7140 | 0,3141 | 1,1307 | -0,6826 | -1,2900 | -1,3780 | -0,7652 | -0,4748 | 0,5316 |
| -0,9151 | -0,9242 | 0,4200 | 0,2041 | -0,4001 | 0,6791 | -1,1782 | 0,1547 | 0,0881 | -0,8113 | -0,0514 |
| -0,3565 | -0,3452 | 1,0535 | 0,1449 | -0,3357 | 0,9686 | 0,0049 | 0,0461 | 0,0148 | -0,7322 | -0,1318 |
| -1,3480 | -1,3336 | 0,1545 | 0,9345 | 0,1137 | 0,5993 | -0,2963 | -0,4401 | 0,9030 | 0,3335 | 0,4711 |
| 0,3010 | 0,4049 | 1,1671 | 0,9081 | -0,0405 | 1,5127 | 0,2602 | 1,0539 | 0,8472 | 1,4971 | -0,4357 |
| -0,1322 | 0,0187 | 0,5506 | 0,7992 | 0,3090 | 0,8461 | -0,7853 | 0,0662 | 0,7289 | 1,3820 | -0,3613 |
| -0,3728 | -0,2919 | 0,5199 | 0,8820 | 0,3009 | 1,0897 | 0,2881 | -0,7810 | 0,8727 | 1,4842 | 0,0530 |
| 0,9574 | 1,0462 | 0,4495 | 0,8493 | -0,5545 | 1,3224 | 0,8600 | 2,1538 | 1,0365 | -0,6846 | -1,1293 |
| -0,6482 | -0,4277 | 0,0811 | 0,8275 | -0,1253 | 0,9340 | 1,1855 | 0,4431 | 0,8197 | -0,7300 | -0,1738 |
| 0,1226 | 0,3530 | 0,0384 | 0,7675 | -0,0433 | 1,2294 | 1,0665 | 1,4552 | 0,9264 | -0,7847 | -0,6049 |
| 1,3364 | 1,3662 | -0,5630 | 2,1334 | 0,9529 | 0,8744 | -0,8749 | 1,6004 | 2,0163 | -0,5316 | -0,7628 |
| -1,5856 | -1,6514 | 0,3833 | 1,2073 | -0,0148 | 1,2976 | 0,1254 | 0,0241 | 1,2523 | -0,7916 | 0,6689 |
| -0,0179 | 0,0746 | -0,9162 | 1,3714 | -0,2080 | -0,6617 | -0,3563 | 1,9011 | 1,6573 | -0,3255 | -1,0209 |
| 0,5907 | 0,7235 | -0,5269 | 2,1723 | 1,1987 | 0,9938 | 1,7639 | 2,2504 | 2,1875 | 0,0184 | -0,7083 |
| 1,4581 | 1,5938 | -0,4679 | 1,7014 | 0,9951 | 0,5920 | -1,3145 | 2,6851 | 1,4608 | 0,0344 | -1,4022 |
| 0,5555 | 0,5496 | -0,5423 | 1,4465 | 0,0145 | 0,5349 | 0,2114 | 2,1805 | 1,5977 | -1,2756 | -1,2402 |
| 2,4945 | 2,5847 | -0,5960 | 1,8164 | -0,0910 | 1,2510 | 1,6364 | 1,6948 | 2,2900 | -1,0942 | -1,4682 |
| 1,3281 | 1,2320 | -0,8426 | 1,5807 | -0,2685 | -0,5356 | -0,4845 | 1,8020 | 2,1784 | -1,1936 | -1,6808 |
| 1,1682 | 1,0752 | -1,0131 | 0,7095 | -0,1611 | 0,0173 | 2,2383 | 1,5055 | 1,0675 | -1,3723 | -0,8552 |
| 0,5128 | 0,4418 | -0,7493 | 1,1555 | -0,2981 | 0,1380 | -0,1401 | 2,0411 | 1,5083 | -1,8158 | -0,6923 |
| -0,1681 | -0,1491 | -0,4392 | 0,9931 | 0,3332 | 1,0858 | 2,5145 | 0,8003 | 1,2509 | -1,7370 | -0,1322 |
| -1,1509 | -1,1466 | 0,0074 | 1,3598 | -0,1815 | 0,5189 | 0,5319 | -0,8047 | 1,5941 | -0,6966 | 0,2999 |
| 1,0498 | 1,1022 | -1,0103 | 1,1733 | 0,0204 | -0,9225 | -1,6373 | 0,5620 | 1,3422 | 0,1592 | -1,1207 |
| 0,4577 | 0,4928 | -0,8643 | 1,2656 | 0,2247 | 0,0145 | -0,2562 | 0,7872 | 1,6335 | 0,0893 | -0,3181 |
| -0,8427 | -0,8938 | -1,6614 | 2,3037 | 0,9431 | -2,1492 | -2,1861 | -0,3136 | 2,2930 | 0,4684 | -0,2846 |
| -0,4040 | -0,2630 | -1,2874 | 1,4398 | 0,2442 | -0,9412 | -1,5646 | 0,0994 | 1,5788 | 0,1622 | -0,1449 |
| -0,4194 | -0,4490 | -1,3828 | 1,3126 | 0,4773 | -0,9128 | 0,1989 | -0,3684 | 1,6516 | -0,2493 | 0,4244 |
| -4,0420 | -3,8672 | -0,5403 | 1,6814 | -0,1446 | -0,5030 | -1,8563 | -1,8798 | 2,1055 | -0,2251 | 2,2787 |
| 0,8746 | 0,9879 | -0,7475 | 1,5741 | 0,6827 | -0,0614 | 0,3985 | 0,3828 | 1,7994 | -0,0525 | -0,8378 |
| -0,4931 | -0,2420 | -1,3105 | 1,4064 | 0,2079 | -0,6671 | -0,6440 | -0,1167 | 1,6678 | -0,3064 | -0,1917 |
| 0,8766 | 1,0088 | -1,4765 | 1,0713 | 0,4921 | -1,0010 | 1,0275 | 0,2336 | 1,3182 | 0,0784 | -1,0488 |
| 0,8993 | 1,0865 | -2,9804 | 1,9417 | 1,1990 | -2,1420 | 0,4863 | 1,0621 | 1,9379 | 0,4519 | -0,9435 |
| 1,4726 | 1,7486 | -2,7228 | 2,0990 | 1,4398 | -1,7248 | 1,3179 | 1,3112 | 1,9567 | 0,4369 | -0,6948 |
| 1,7721 | 2,1836 | -2,0667 | 0,7131 | 0,6875 | -0,5607 | 2,7707 | 0,9261 | 0,8253 | 0,3564 | -1,4854 |
| 0,1208 | 0,2556 | -0,8274 | 1,3888 | 0,1264 | -0,5967 | -1,1663 | 1,3253 | 1,6560 | 0,4700 | -1,2252 |
| 0,1748 | 0,3651 | -0,7912 | 1,8094 | 0,4930 | 0,2323 | 0,5227 | 1,2708 | 1,2420 | 0,2349 | -1,0803 |
| -0,0802 | 0,0613 | -0,8678 | 1,1118 | 0,4939 | -0,5403 | -0,5456 | 0,7831 | 1,4077 | 0,1480 | -0,6724 |
| -1,0676 | -1,1534 | -0,6723 | 1,1351 | 0,2303 | -0,6773 | -1,1809 | 0,0712 | 1,3221 | 0,1801 | -0,3918 |
