

FERRAMENTAS ESTATÍSTICAS PARA MONITORAMENTO DE MODELOS
INFERENCIAIS

Kese Pontes Freitas

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO
LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA
(COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE
DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE
EM CIÊNCIAS EM ENGENHARIA QUÍMICA.

Aprovada por:

Prof. Enrique Luís Lima, DSc.

Prof. José Carlos Costa da Silva Pinto, DSc.

Prof. Tito Lívio Moitinho Alves, DSc.

Prof. Marcio Schwaab, DSc.

RIO DE JANEIRO, RJ - BRASIL

MAIO DE 2009



COPPE/UFRJ

FERRAMENTAS ESTATÍSTICAS PARA MONITORAMENTO DE MODELOS
INFERENCIAIS

Kese Pontes Freitas

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Química, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Química.

Orientadores: Enrique Luís Lima

José Carlos Costa da Silva Pinto

Rio de Janeiro

Maio de 2009

Freitas, Kese Pontes

Ferramentas Estatísticas para Monitoramento de Modelos Inferenciais/ Kese Pontes Freitas. – Rio de Janeiro: UFRJ/COPPE, 2009.

XIV, 103 p.: il.; 29,7 cm.

Orientador: Enrique Luís Lima

José Carlos Costa da Silva Pinto

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia Química, 2009.

Referencias Bibliográficas: p. 96-103.

1. Análise estatística. 2. Dados operacionais. 3. Modelos inferenciais. I. Lima, Enrique Luís *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Química. III. Título.

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

FERRAMENTAS ESTATÍSTICAS PARA O MONITORAMENTO DE MODELOS INFERENCIAIS

Kese Pontes Freitas

Maio/2009

Orientadores: Enrique Luís Lima

José Carlos Costa da Silva Pinto

Programa: Engenharia Química

Nesta dissertação é proposta uma abordagem estatística para monitoramento e avaliação da qualidade de inferências, com foco nos modelos inferenciais. Como ilustração é apresentado o caso da inferência da temperatura a 85% do diesel em um processo de refino. Propõe-se aqui que ferramentas estatísticas clássicas sejam realizadas tanto nos dados operacionais quanto no modelo inferencial empregado, a fim de investigar a consistência estatística da inferência. Ao longo do estudo, foi proposta uma adequação do modelo inferencial por meio do método estatístico de estimação de parâmetros. Todavia, o modelo inferencial em estudo não se apresenta estatisticamente adequado para fornecer a inferência confiável da temperatura a 85% do diesel, apresentando fortes indícios de super-parametrização. A metodologia de abordagem estatística proposta se mostra bastante eficaz na determinação da qualidade estatística de inferências, sendo um procedimento que apresenta grande potencial de aplicação em tempo real.

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

STATISTICAL TOOLS FOR MONITORING OF INFERENTIAL MODELS

Kese Pontes Freitas

May/2009

Advisors: Enrique Luís Lima

José Carlos Costa da Silva Pinto

Department: Chemical Engineering

In this dissertation a statistical approach is proposed for monitoring and evaluation of the quality of inferences and inferential models. As an illustration, the inference of the 85% temperature of diesel in a refining plant is presented. It is proposed that statistical analyses be performed both with the operational data and the inferential data in order to investigate the statistical consistency of the inferences. The adequacy of the studied inferential model is analyzed through parameter estimation. However, it is shown that the studied inferential model does not seem appropriate to provide reliable inferences for the 85% temperature of diesel, because there are strong indications of super-parameterization. The proposed statistical approach is very effective for determination of the quality of statistical inferences and presents great potential for applications in real time.

NODIAEMQUEEU SAÍ DE CASA

JOEL MARQUES

No dia em que eu saí de casa minha mãe me disse:

- Filho, vem cá!

*Passou a mão em meus cabelos olhou em meus olhos começou falar:
- Por onde você for eu sigo com meu pensamento sempre onde estiver
Em minhas orações eu vou pedir a Deus que ilumine os passos seus...*

*Eu sei que ela nunca compreendeu
Os meus motivos de sair de lá
Mas ela sabe que depois que cresce
O filho vira passarinho e quer voar...*

*Eu bem queria continuar ali
Mas o destino quis me contrariar
E o olhar de minha mãe na porta
Eu deixei chorando a me abençoar...*

*A minha mãe naquele dia me falou do mundo como ele é
Parece que ela conhecia cada pedra que eu iria por o pé
E sempre ao lado do meu pai da pequena cidade ela jamais saiu
Ela me disse assim:
- Meu filho vá com Deus que este mundo inteiro é seu...*

*Eu sei que ela nunca compreendeu
Os meus motivos de sair de lá
Mas ela sabe que depois que cresce
O filho vira passarinho e quer voar...*

*Eu bem queria continuar ali
Mas o destino quis me contrariar
E o olhar de minha mãe na porta
Eu deixei chorando a me abençoar...*

*E o olhar de minha mãe na porta
Eu deixei chorando a me abençoar...*

*E o olhar de minha mãe na porta
Eu deixei chorando a me abençoar...*

À Sandra Aparecida Duarte Pontes, senhora de minha vida.

À Ketly Pontes Soares, minha brilhante estrela.

À Alberto Assis Araújo, meu maravilhoso pai.

DEDICO.

Ao meu anjo, amigo e cúmplice.

À minha adorável amiga Beatriz.

OFEREÇO.

AGRADECIMENTOS

“Agradeço todas as dificuldades que enfrentei; não fosse por elas, eu não teria saído do lugar. As facilidades nos impedem de caminhar. Mesmo as críticas nos auxiliam muito.”

FRANCISCO CÂNDIDO XAVIER

Neste momento especial, o fechamento e o início de etapas importantes, agradeço e louvo ao Pai Celestial a vitória glorificante.

Amor sincero à tradicional família mineira (TFM), Duarte Pontes, presente e verdadeira.

À Minha amada mãe guerreira, ao meu pai tão generoso.

À minha irmã tão companheira e ao meu irmão amoroso.

Carinho à amiga família Prata, a qual serei sempre muito grata.

Dedicação aos Mafra que muito por mim fizeram, é de coração meu lar carioca eterno.

Quero eu aqui lembrar os amigos companheiros, os íntimos confidentes, os resistentes e os fuleiros.

André, paciência e bondade.

Diego, poesia e amizade.

Edu Lemos, disposição e fraternidade.

Edu Lima, concentração e sinceridade.

Elisa, carinho e simplicidade.

Fabrcio, dedicaçao e serenidade.

Heloisa, ternura e cumplicidade.

Isaias, benevolencia e religiosidade.

Joao Batista, lealdade e irmandade.

Jose da Paixao, polidez e tranqüilidade.

Matheus, confidencia e fidelidade.

Paulo Andre, bom humor e boa vontade.

Schwaab, inteligencia e generosidade.

E à minha querida turma (mestrado 2007), só tenho a bendizer: vocês são os melhores, marcaram o meu viver!

Admiração eterna aos meus orientadores; poderosos em conhecimento, ricos em humanidade.

Enrique Luis Lima, meu referencial.

Jose Carlos Costa da Silva Pinto, meu doutrinador.

Gratidão pela exímia orientação, pela humildade em compartilhar conhecimento na mais completa capacidade de se doar.

Carinho ao meu orientador acadêmico, Tito Lívio Moitinho.

Acolheste-me de braços abertos, me ensinastes os primeiros passos, companheiro desta caminhada.

Aos professores do PEQ não há o que acrescentar, são na íntegra completos e fazem deste um grande lar.

Aos funcionários do PEQ o meu sincero agradecer, foram em todos os momentos companheiros pra valer!

Amizade saudosa ao casal maravilhoso Andréia e Esly Costa Jr., presentes preciosos.

Gratidão à CAPES, pela oportunidade de execução do trabalho.

Agradecimentos à PETROBRÁS, em especial à Antônio Carlos Zanin, à Euclides Almeida e à equipe da RPBC (Refinaria Presidente Bernardes de Cubatão), pela preciosa contribuição ao disponibilizar os dados operacionais.

À cidade maravilhosa; não passas de promessas...

ÍNDICE

1. CAPÍTULO I - INTRODUÇÃO.....	01
2. CAPÍTULO 2 – REVISÃO BIBLIOGRÁFICA.....	04
2.1 ANÁLISE ESTATÍSTICA DE INFERÊNCIAS.....	04
2.1.1 Análise Estatística de Erros Grosseiros.....	06
2.2 SENSORES VIRTUAIS: MECANISMOS DE INFERÊNCIAS.....	08
3. CAPÍTULO 3 – MONITORAMENTO ESTATÍSTICO DE INFERÊNCIAS..	16
3.1 MONITORAMENTO ESTATÍSTICO DE INFERÊNCIAS.....	16
3.2 ESTUDO DE CASO: INFERÊNCIA DA T85% ASTM DO DIESEL PESADO..	18
3.2.1 O Modelo Inferencial.....	19
3.2.2 O Problema do <i>BIAS</i>	21
3.3 METODOLOGIA DE ABORDAGEM.....	21
3.3.1 Inferências.....	23
3.3.2 Resíduos de inferências.....	24
3.3.3 Caracterização Básica.....	24
3.3.4 Caracterização Dinâmica.....	26
3.3.5 Caracterização da Predição.....	33
3.3.6 Adequação do Modelo Inferencial.....	43
4. CAPÍTULO 4 – RESULTADOS.....	44
4.1 DADOS OPERACIONAIS.....	44
4.2 ANÁLISE DOS DADOS.....	45
4.3 ANÁLISE DO MODELO INFERENCIAL.....	61
4.4 ADEQUAÇÃO DO MODELO INFERENCIAL.....	76
4.5 ANÁLISE DO MODELO INFERENCIAL PROPOSTO.....	82
5. CAPÍTULO 5 – CONCLUSÃO E SUGESTÕES.....	94
6. REFERÊNCIAS.....	96

CAPÍTULO I

INTRODUÇÃO

“Sei que meu trabalho é só uma gota no oceano, mas sem ele o oceano seria menor.”

MADRE TERESA DE CALCUTÁ

Nos processos industriais um problema muito comum é a difícil medição de algumas variáveis cruciais para avaliação contínua do seu desempenho e qualidade. A origem desta dificuldade pode estar associada à falta de instrumentos que meçam essas variáveis ou à necessidade de amostragem e análises de laboratório demoradas. Neste último caso, a demora pode inviabilizar uma operação satisfatória do processo.

Uma abordagem bastante utilizada para contornar esta situação é usar um modelo matemático para inferir variáveis não medidas, ou medidas com baixa frequência, por meio de outras variáveis (secundárias ou auxiliares) que podem ser medidas com maior frequência.

Em geral, o conhecimento dos valores das variáveis, combinado com o uso de modelos matemáticos apropriados (sensor virtual), permite inferir e controlar a qualidade de um produto sem a necessidade de análises laboratoriais muito frequentes. No entanto, esta solução requer dados e modelo matemático de alta confiabilidade para seu êxito. Se esta combinação não é, ou deixa de ser, confiável, os valores inferidos podem trazer prejuízos maiores que os resultantes de análises de laboratório demoradas.

Desta forma é evidente a necessidade de monitorar a qualidade das inferências obtidas por meio de variáveis auxiliares e modelos matemáticos (inferenciais), antes de considerá-las satisfatórias para outros usos (como, por exemplo, controle de processos).

Como os dados medidos estão sujeitos a erros de medição, o monitoramento em tempo real da qualidade de inferências obtidas deve ser realizado sob a luz de ferramentas estatísticas apropriadas.

Nesta pesquisa é desenvolvida uma metodologia de monitoramento de qualidade de inferências, sendo aplicada a um processo de produção de diesel da indústria de refino. O problema analisado consiste na inferência da qualidade do diesel produto, que é composto pela mistura da corrente da torre atmosférica de diesel de destilação direta e de querosene e da corrente de fundo da unidade de solventes.

A especificação do diesel é feita através do parâmetro de qualidade T85%, que corresponde ao ponto 85% da curva de destilação ASTM D-86 do diesel produto (temperatura quando 85% do volume de uma amostra do produto são evaporados). No entanto, na prática, a amostragem para monitorar esta especificação do diesel é realizada a cada 12 horas, sendo que os resultados das análises laboratoriais são repassados à operação 3 horas após a coleta, indicando assim a necessidade de valores inferidos com maior frequência por meio de um sensor virtual. O sensor virtual estudado neste caso depende de informações oriundas da inferência de outras propriedades do processo, calculadas a partir de modelos de destilação multicomponente, que podem conter, além de possíveis erros de modelagem e parametrização, erros embutidos nas variáveis medidas.

Assim, o objetivo principal desta dissertação é a elaboração de uma metodologia, fundamentada em técnicas estatísticas, para monitoração e avaliação da qualidade dos dados produzidos por modelos de inferência. O principal resultado perseguido é a definição e proposição de uma metodologia de avaliação de fidelidade de modelos inferenciais, que permita definir de forma inequívoca se um modelo inferencial precisa ser reajustado aos dados disponíveis de processo. Neste sentido, o estudo foi elaborado ao longo de três linhas principais de ação:

- i. Monitoração da qualidade de inferências.

- ii. Comparação do desempenho de diferentes modelos de inferência.
- iii. Determinação de um conjunto básico de ferramentas estatísticas para avaliação e julgamento de dados.

Esta dissertação está organizada em cinco capítulos; incluindo esta introdução. O próximo capítulo, intitulado Revisão Bibliográfica, trata de uma breve revisão sobre o tema abordado, em termos das principais ferramentas de análise estatística utilizadas e o uso generalizado de sensores virtuais. No Capítulo III, intitulado Monitoramento Estatístico de Inferências, discute-se a relevância de monitorar inferências estatisticamente e apresentam-se a descrição do sistema estudado e a metodologia de abordagem aplicada ao problema. Já o quarto capítulo, intitulado Resultados, apresenta os resultados mais relevantes, acompanhados por uma análise e discussão. Finalmente, o Capítulo V, intitulado Conclusão e Sugestões, apresenta os principais ganhos, vantagens e desvantagens da abordagem empregada neste estudo, e ainda outras possíveis propostas a explorar no tema.

CAPÍTULO II

REVISÃO BIBLIOGRÁFICA

“Não foram os fatos que vieram a posteriori confirmar a teoria. A teoria é que veio subseqüentemente explicar e resumir os fatos.”

ALLAN KARDEC

Neste capítulo são apresentados os principais conceitos empregados neste estudo em uma breve revisão contextual sobre o tema. Esta parte do estudo foi dividida nos temas Análise Estatística de Processo e Sensores Virtuais.

2.1 ANÁLISE ESTATÍSTICA DE INFERÊNCIAS

Estratégias de controle são desenvolvidas para manter a operação do processo dentro de uma faixa pré-estabelecida ou ótima, geralmente o estado estacionário. Todavia, os processos reais são continuamente sujeitos as mudanças de condições operacionais de um estado estacionário para outro (MORAD *et al.*, 1999). Portanto, o

monitoramento efetivo das características da qualidade de um processo de produção depende frequentemente de ferramentas estatísticas para a detecção, identificação e análise das causas responsáveis por variações que afetam o comportamento do processo de maneira imprevisível.

De acordo com MONTGOMERY (2004), que apresenta uma breve história dos procedimentos usados para controle e melhoria da qualidade, a introdução da ciência estatística no meio industrial foi feita por Shewhart em 1924. Shewhart utilizava métodos estatísticos para alcançar o estado de controle de um sistema e para julgar quando este estado foi alcançado, ficando famoso por desenvolver o conceito estatístico de carta de controle (CC).

Ao longo do tempo, com a complexidade dos processos e o enorme volume de dados disponíveis em plantas industriais modernas automatizadas, abordagens empregando técnicas e ferramentas estatísticas foram continuamente difundidas, devido à necessidade de garantir as condições operacionais do processo e as especificações do produto. Desta forma, análise, monitoramento e controle de processos baseados em técnicas estatísticas são práticas comuns em várias indústrias há mais de 40 anos (KOURTI, 2006).

De acordo com MILETIC *et al.* (2004), a literatura tecno-científica reporta uma enorme variedade de ferramentas estatísticas disponíveis para análises de processo, que vêm ganhando popularidade no meio industrial, tais como: análises dos componentes principais, mínimos quadrados parciais, entre outras. Todavia, são as ferramentas estatísticas clássicas (como por exemplo, médias e variâncias) as responsáveis pela detecção e definição das condições de operabilidade dos processos. Sua importância é tal, que estratégias desenvolvidas para controle e otimização de processos as empregam como referenciais.

Pode-se dizer que as ferramentas estatísticas clássicas estão presentes em todos os métodos estatísticos, compondo a base destes métodos, atuando de forma secundária e muitas vezes exercendo o papel principal (como alguns exemplos destes métodos podem-se citar: o controle estatístico da qualidade (CEQ), o controle estatístico de processo (CEP), o planejamento de experimentos (PE), a análises de capacidade de

processos, a teoria da amostragem, dentre outros). Essas ferramentas clássicas também estão presentes em estudos que envolvem a regressão de modelos, como os procedimentos de estimação de parâmetros e reconciliação de dados (RD). Portanto, é inegável a aplicabilidade dessas ferramentas estatísticas, sobretudo no ambiente industrial, onde o nível de exigências a respeito da qualidade do produto, da segurança operacional e dos padrões da legislação ambiental é cada vez mais rigoroso.

Neste contexto, pode-se afirmar que o uso de ferramentas estatísticas são práticas estabelecidas e de grande conceito. É possível perceber, dada a popularidade dessas ferramentas, que as variáveis operacionais importantes já possuem uma estatística estabelecida na maioria das unidades industriais. Todavia, mesmo com o avanço da tecnologia digital, que possibilita a rápida evolução desta inteligente reunião entre a engenharia e a estatística, estudos estatísticos envolvendo qualidade de inferências parecem não ter sido reportados.

Apesar do conceito de inferências ser relativamente novo na literatura tecnocientífica, a importância destes elementos para os processos industriais é elevada, sendo empregadas como parâmetros da qualidade do produto e como medida no sistema de controle avançado. Portanto, esta dissertação é dedicada à análise estatística da qualidade deste importante e decisivo elemento, que se encontra presente na maioria dos processos atuais.

2.1.3 Análise Estatística de Erros Grosseiros

Outra importante aplicação da estatística na análise de processos está relacionada à detecção de erros grosseiros (DEG). A inferência representa um importante papel nos processos modernos de engenharia, sendo geralmente obtidas a partir de medidas do processo e de modelos construídos, com base fenomenológica ou empírica, usando dados históricos que provêm de plantas industriais. Contudo, dados amostrados inerentemente contêm erros, fazendo-se imprescindível uma avaliação da qualidade destes dados operacionais.

Embora seja natural admitir que os erros de medidas sejam aleatórios, normalmente distribuídos, não correlacionados, com média nula e variância conhecida, estas colocações nem sempre estão de acordo com a realidade de operação do processo. Erros grosseiros (associados à falhas eventuais ou a desvios sistemáticos dos sistemas de medições) estão muitas vezes presentes nos dados. De forma geral, não seguem o comportamento estatístico clássico.

De acordo com SCHWAAB e PINTO (2007), a expressão “erros grosseiros” é usada genericamente para designar pontos experimentais que parecem não se adequar a uma distribuição particular de probabilidade definida pela grande maioria dos demais pontos experimentais. Todavia, Gumbel em 1960 cautelosamente alerta para o fato de que a rejeição de erros grosseiros, mesmo em bases estatísticas, é e continua a ser um procedimento perigoso, pois a existência destes pontos pode ser a prova de que a população estudada não é, na realidade, o que se admitiu que fosse (KRUSKAL *et al.*, 1960). Apesar disso, o fato é que a presença de erros grosseiros pode invalidar a base estatística admitida para as variáveis envolvidas, levando a obtenção de estimativas errôneas para as variáveis e parâmetros (FREITAS *et al.*, 2008).

Segundo CHEN e ROMAGNOLI (1998) os erros grosseiros podem ser divididos em duas classes: os desvios sistemáticos (também chamados permanentes) e os valores espúrios. Desvios sistemáticos são valores consistentemente mais altos ou mais baixos das medidas em relação aos valores reais. O valor médio do erro neste caso não é nulo. Estes erros podem derivar da instalação ou calibração incorreta dos instrumentos de medida (LIEBMAN *et al.*, 1992, MCBRAYER e EDGAR, 1995). Já os valores espúrios são medidas obtidas a partir de algum comportamento anormal, resultante de “picos” de processo ou distúrbios não medidos. Um valor espúrio é uma medida que não segue qualquer distribuição estatística representativa do conjunto dos dados (ALBUQUERQUE e BIEGLER, 1996; CHEN e ROMAGNOLI, 1998).

Um fator limitante para o amplo uso de sensores virtuais em processos industriais é a sua incapacidade de lidar com dados ruidosos. Assim, para obter inferências satisfatórias, a influência dos menos frequentes erros grosseiros deve ser eliminada. Isto pode ser realizado por procedimentos de detecção de erros grosseiros, que trata de uma subdivisão da etapa completa de retificação de dados.

Algumas abordagens para detecção de erros grosseiros baseadas em técnicas estatísticas têm sido reportadas na literatura, dentre as quais podem ser citadas: a utilização das técnicas de análise de agrupamentos, do termo em inglês “*cluster*” (CHEN e ROMAGNOLI, 1998; PRATA *et al.*, 2008; FREITAS *et al.*, 2008), e a análise dos componentes principais (TONG e CROWE 1995, 1996; JIANG *et al.*, 1999).

2.2 SENSORES VIRTUAIS: MECANISMOS DE INFERÊNCIAS

A quantidade e a qualidade de informações obtidas de um dado processo (ou mais genericamente, de um sistema) são fatores fundamentais para operá-lo satisfatoriamente. Qualquer sistema de monitoramento e/ou controle de processos está vinculado ao emprego de elementos físicos, conhecidos como sensores, que permitam a coleta dessas informações. A importância dos sensores para processos é cada vez maior, envolvendo grande número de áreas de aplicação. Existem sensores de muitos tipos, que são especificados de acordo com a natureza da variável a ser medida e com as condições de aplicação.

PARK e HAN (2000) afirmam que a disponibilidade de sensores que superem as limitações de custo, confiabilidade e grandes atrasos de resposta favorece a implantação de sistemas de monitoramento e controle mais eficientes, possibilitando a redução nas perdas de produto e energia, a diminuição de problemas de segurança e maior controle sobre os produtos tóxicos gerados durante o processo.

Apesar da constante evolução tecnológica dos sensores, estes elementos ainda possuem algumas sérias limitações em vários processos. Alguns autores (THAM *et al.*, 1991; SOROUSH 1998; GONZÁLES, 1999; ALBERTOS e GODWIN, 2002; ZANATA, 2005; FORTUNA *et al.*, 2007) descrevem os seguintes problemas como os que mais afetam sua aplicação:

- i. Indisponibilidade no mercado.

- ii. Altos custos de aquisição e de manutenção.
- iii. Necessidade periódica de manutenção do equipamento.
- iv. Precisão grosseira.
- v. Falhas.
- vi. Elevado tempo de resposta.
- vii. Distância considerável do ponto de medição.
- viii. Ambiente de medição hostil.

Neste contexto, como tentativa de solucionar ou amenizar os problemas descritos acima e aproveitando os avanços da tecnologia digital, surge um novo conceito de sensores, capazes de manipular computacionalmente as informações medidas e transmiti-las de forma mais eficiente para o uso do processo. Estes sensores são chamados sensores virtuais e são responsáveis pela inferência de propriedades cruciais do processo.

A literatura técnico-científica reporta uma variedade de nomes empregada para denominar a técnica que, ao utilizar sensores parcial ou completamente programáveis, justifica a popularidade do termo sensor virtual. De acordo com ZANATA (2005) a origem da denominação sensor virtual está relacionada à sua principal aplicação: substituição dos sensores físicos.

Recentemente, o popular termo *soft sensor* (*software sensor*) teve patente requerida pela *Pavilion Technologies Inc.* (ZANATA, 2005). Esta empresa foi uma pioneira no uso da técnica para inferências das propriedades de qualidade da folha de papel durante a fabricação.

Portanto, os sensores virtuais representam um importante marco na evolução dos processos modernos de engenharia. Podem ser definidos como um algoritmo para tomada de medidas, cujo elemento principal é o modelo inferencial. O modelo inferencial corresponde a um modelo matemático que, por meio de um conjunto de variáveis medidas, torna possível a predição de variáveis que não podem ser medidas na planta e/ou em tempo real.

A idéia fundamental deste “instrumento inteligente” é a associação entre o sensor ou o conjunto de sensores e o modelo. Desta forma, o sensor virtual é composto por duas partes distintas:

- i. *Hardware*. Consiste no sensor, ou conjunto de sensores que medem variáveis relacionadas com a variável que se deseja monitorar.
- ii. *Software*. Consiste no algoritmo com o modelo implementado (modelo inferencial), capaz de fornecer por meio de simulações a variável desejada.

Diferentes estratégias têm sido propostas para desenvolver sensores virtuais. De acordo com FORTUNA *et al.* (2007), são três as principais abordagens utilizadas para o desenvolvimento destes sensores: modelagem fenomenológica, estatística multivariada e modelagem por inteligência artificial (Redes Neurais, Lógica *Fuzzy* e métodos híbridos).

A primeira destas abordagens, também conhecida como modelagem caixa branca, requer um vasto conhecimento das leis físicas. Como as plantas usualmente envolvem fenômenos que se identificam melhor com modelos não lineares e de maior complexidade, o desenvolvimento do modelo pode requer muito tempo e esforço de programação e implementação (FREITAS *et al.*, 2006). Além disso, com os atuais avanços nos sistemas de automação industrial, um grande número de variáveis de processo podem ser simultaneamente medidas e armazenadas, podendo ser facilmente acessadas, o que encoraja a construção de modelos a partir de dados históricos, já que o registro histórico acumulado representa uma fonte muito útil de informação, que permite identificar características relevantes do processo (FORTUNA *et al.*, 2007). Todavia, é importante enfatizar que estas abordagens são usadas para executar uma mesma função: transformar medidas disponíveis na planta em outras medidas não disponíveis, com auxílio de uma transformação matemática.

Devido a sua simplicidade, os sensores virtuais são empregados para solução de inúmeros problemas, tais como (FORTUNA *et al.*, 2007):

- i. Cópia do sistema de medição.
- ii. Simulações.
- iii. Redução da necessidade de sensores físicos (hardware).
- iv. Predição em tempo real para monitoramento e controle da planta.

- v. Validação de sensores.
- vi. Estratégias de diagnóstico e detecção de falhas.

Um problema mais específico, mas de grande importância que pode ser encontrado nos mais variados processos, em escala industrial ou laboratorial, são os chamados parâmetros intangíveis. Parâmetros intangíveis podem ser propriedades muito complexas de uma substância, tais como cor, sabor ou gosto, cremosidade, maciez ou suavidade, que podem ser somente definidos subjetivamente. De acordo com SOROUSH (1998), a relação entre a planta e o índice de qualidade de produtos que dependem dos parâmetros intangíveis geralmente é complicada e não bem entendida, baseando-se apenas na experiência sem um padrão quantitativo. Sensores virtuais podem ser muito úteis nesses casos, por permitirem acúmulo de experiência passada em modelos empíricos de monitoramento.

Os sensores virtuais e suas diversas aplicações têm sido amplamente estudados no ambiente industrial (FORTUNA *et al.*, 2007). Segundo GONZÁLES (1999), “os sensores são os olhos pelos qual o comportamento de uma planta é visto e a informação sobre seu desempenho é obtida”. Daí a grande importância dos sensores virtuais, capazes de “enxergar mais além”.

Um estudo de revisão realizado por GONZÁLES (1999) propõe uma abordagem integrada de vários métodos de desenvolvimento de sensores virtuais, com destaque para os modelos obtidos através de técnicas de modelagem empírica e modelagem híbrida. O autor também faz considerações relacionadas a estratégias de controle baseadas em modelos inferenciais.

Alguns dos aspectos mencionados por GONZÁLES (1999) também podem ser encontrados em ALBERTOS e GOODWIN (2002), que analisaram as propriedades básicas de sensores, com o foco nas operações de controle em tempo real. Estes autores revisaram diferentes abordagens para o desenvolvimento de sensores virtuais, além de algumas estratégias para aplicar estes projetos de acordo com as suas várias restrições na indústria.

Dentre as aplicações industriais para quais os sensores virtuais têm tido grande emprego, destaca-se o caso onde se faz necessária a obtenção de medidas em um intervalo pequeno de tempo. THAM *et al.* (1991) aborda este problema em vários processos industriais diferentes, apresentando diversas formulações para aplicações independentes. De acordo com YANG e CHAI (1997), estes sensores constituem uma poderosa ferramenta para aumento da frequência de amostragem.

De acordo com SOROUGH (1998), os sensores convencionais tornaram-se inadequados para algumas necessidades dos modernos sistemas de monitoramento e de controle de processos que requerem respostas rápidas durante a operação. Nesses casos, o emprego de sensores virtuais é a alternativa mais viável na crescente e agressiva competição global para produção de produtos de alta qualidade a baixos custos.

Uma abordagem inovadora na literatura técnica-científica dedicada a desenvolvimento de sensores virtuais é proposta por LIN *et al.* (2007). Estes autores apresentam uma sistemática baseada em técnicas de estatística robusta para desenvolvimento de sensores virtuais, construídos a partir de dados operacionais. Propõe-se que os sensores sejam utilizados como observadores de estados de variáveis relacionadas à qualidade dos produtos e à emissão de poluentes atmosféricos na indústria.

Devido a sua importância, o cenário de aplicações dos sensores virtuais é cada vez mais amplo. A seguir serão apresentados exemplos que ilustram esse cenário.

Uma aplicação bastante interessante de sensores virtuais foi relatada por CHOI e PARK (2001). Estes autores apontam o uso de medidores inferenciais como uma alternativa bastante viável nos processos de tratamento de efluentes, pois apresentam menor custo, satisfatória precisão e maior confiabilidade em relação aos sensores convencionais empregados para medir a composição de fósforo e nitrogênio dissolvidos. Vários outros trabalhos semelhantes vêm sendo desenvolvidos na área ambiental. Alguns exemplos interessantes são apresentados por BERTHOUEX e BOX (1996), MILLER *et al.* (1997), MASSON *et al.* (1999), SOTOMAYOR *et al.* (2002), MATSUMURA *et al.* (1998) e LIN *et al.* (1997).

Há também uma grande aplicabilidade destes sensores nas áreas de bioprocessos e de biotecnologia, bem como na engenharia genética, onde os sensores convencionais são caros ou inexistentes. Alguns trabalhos envolvendo sensores virtuais nestas áreas de pesquisa são apresentados por LINKO *et al.* (1999), JAMES *et al.* (2000), ASSIS e MACIEL FILHO (2002), BOGAERTS e WOUWER (2003), KOMIVES e PARKER (2003), BOGAERTS e WOUWER (2004), ARAÚZO-BRAVO *et al.* (2004), FEITKENHAUER e MEYER (2004), DESAI *et al.* (2006) e DAI *et al.* (2006). JAMES *et al.* (2000) fazem uma extensa revisão sobre o assunto, focando o desenvolvimento de medidores inferenciais para processos bioquímicos. Estes autores apresentam ainda um histórico organizado, que permite avaliar onde estão ocorrendo os avanços mais significativos e quais técnicas tem maior aplicação para estes processos.

Na indústria de celulose e papel sensores virtuais são empregados como inferidores para predição do número *Kappa* (índice de qualidade da celulose) e como medidores inferenciais para determinar parâmetros de qualidade considerados críticos de folhas de papel (FREITAS *et al.*, 2005; DUFOUR *et al.*, 2005).

Sensores virtuais também encontram grande aplicabilidade na indústria de polímeros, onde a qualidade do produto é especificada principalmente em termo de propriedades difíceis de medir e geralmente indisponíveis em tempo real (PRATA, 2005). Algumas aplicações de sensores virtuais para processos de polimerização são apresentadas por ZHANG *et al.* (1997), TALAIE *et al.* (1998), CHEN *et al.* (2004), SHARMIN *et al.* (2006), LIU (2007) e FACCO *et al.* (2008).

Entretanto, a indústria petrolífera pode ser considerada a pioneira no desenvolvimento e implementação de estratégias envolvendo sensores virtuais. Tal fato se deve à necessidade do conhecimento em tempo real de algumas propriedades para o bom funcionamento dos sistemas de monitoramento e controle avançado do processo e do produto. Em um artigo de discussão, FRIEDMAN (2005) afirma que foi um dos pioneiros em desenvolver modelos inferenciais fenomenológicos para unidades de destilação cruas, cujas atividades foram iniciadas em 1982. Este autor propôs um modelo inferencial, cujo nome comercial é GCC (*Generalized Cut-point Calculation*), que foi implementado para aplicações em malha fechada de controle avançado em aproximadamente 40 refinarias.

A Tabela (2.1) apresenta alguns trabalhos recentes desenvolvidos para várias aplicações de sensores virtuais no processo de refino.

Tabela 2.1: Aplicações de sensores virtuais no processo de refino.

Autor	Ano	Cenário de aplicação
Yang <i>et al.</i>	1998	Craqueamento Catalítico
Park e Han	2000	Coluna de Destilação
Zhong e Yu	2000	Coluna de Destilação
Zamprogna <i>et al.</i>	2001	Coluna de Destilação
Dave <i>et al.</i>	2003	Coluna de Destilação
Fortuna <i>et al.</i>	2003	Coluna de Destilação
Kano <i>et al.</i>	2003	Coluna de Destilação
Yan <i>et al.</i>	2003	Coluna de Destilação
Chatterjee e Saraf	2004	Coluna de Destilação
Yan <i>et al.</i>	2004	Coluna de Destilação
Fortuna <i>et al.</i>	2005	Coluna de Destilação
Frezza <i>et al.</i>	2005	Coluna de Destilação
Jana <i>et al.</i>	2005	Coluna de Destilação
Zamprogna <i>et al.</i>	2005	Coluna de Destilação
Skeie <i>et al.</i>	2006	Unidade de Separação Água/óleo
Fortuna <i>et al.</i>	2007	Vários*
Jain <i>et al.</i>	2007	Coluna de Destilação
Kano <i>et al.</i>	2008	Coluna de Destilação
Yan	2008	Coluna de Destilação

* Livro intitulado “*Soft Sensors for Monitoring and Control*”, apresenta também alguns trabalhos envolvendo sensores virtuais para monitoramento e controle deste processo.

Apesar de todo avanço e desenvolvimento relacionado à construção e implementação de sensores virtuais, tanto na academia quanto na indústria, surpreendentemente nada tem sido reportado a respeito de caracterização estatística da qualidade dos dados fornecidos ou do desempenho dos modelos implementados. De forma geral, a literatura considera que um modelo inferencial é bom quando a correlação obtida entre um certo conjunto de dados inferidos na planta e as respectivas

medidas obtidas fora de linha é superior a um certo valor limite pré-estabelecido (por exemplo, a correlação de 0,95). Ainda mais surpreendente é o fato de que o monitoramento estatístico das medidas de processo é reconhecidamente importante e, portanto, usualmente realizada nas plantas industriais.

Como as inferências podem constituir um elemento de grande importância nos processos modernos, responsáveis muitas vezes pela aceitação da qualidade do produto e decisivas durante a tomada de decisão em que sistemas de controle avançado, é altamente recomendável que o monitoramento rigoroso da qualidade dos modelos inferenciais seja realizado, com base em ferramentas estatísticas.

CAPÍTULO III

MONITORAMENTO ESTATÍSTICO DE INFERÊNCIAS

“Os métodos são as verdadeiras riquezas.”

FRIEDRICH NIETZSCHE

Este capítulo trata da apresentação e da interpretação dos resultados relacionados ao monitoramento estatístico de inferências. A apresentação é feita em três etapas: primeiramente é feita uma discussão sobre o monitoramento estatístico de inferências; em seguida é feito um estudo de caso, relacionado à inferência da T85% ASTM do diesel produto; finalmente, é apresentada uma proposta para abordagem metodológica.

3.1 MONITORAMENTO ESTATÍSTICO DE INFERÊNCIAS

O desenvolvimento de métodos para o monitoramento estatístico de processos foca principalmente na incorporação de técnicas que priorizam tomadas de ações e correções a partir da análise dados de medidas. A justificativa para tal consiste no fato de que em condições normais de operação, as unidades industriais já possuem uma

estatística conhecida para as variáveis operacionais importantes. Todavia, algumas propriedades cruciais para análise do desempenho do processo não estão disponíveis na planta e/ou em tempo real, sendo necessário o emprego de inferências.

Como as inferências são geralmente obtidas por meio de variáveis medidas disponíveis com o auxílio de um modelo, a elaboração de estatísticas para monitoramento das mesmas ainda é pouco comum, o que torna sua confiabilidade questionável. São dois os fatores que podem comprometer a confiabilidade das inferências:

- i. A acurácia das medidas das variáveis operacionais.
- ii. A fidedigna representação do processo.

O primeiro fator se refere ao fato de que os dados amostrados inerentemente contêm erros. Qualquer medida experimental apresenta algum desvio do valor verdadeiro, devido à precisão limitada dos instrumentos e à presença de efeitos ou fenômenos que foram desconsiderados ou que não são conhecidos (BOX *et al.*, 1978); como já foi discutido.

O segundo fator responsável por contaminar as inferências é o fato de que a descrição completa dos fenômenos envolvidos em um processo é matematicamente impossível. Desta forma, todos os modelos podem ser considerados incorretos, embora alguns sejam úteis (BOX, 1979). De acordo com SCHWAAB (2007), os modelos são representações idealizadas de um processo, cuja tentativa de descrição detalhada pode levar a estruturas altamente complexas, inviabilizando a sua aplicação prática.

Um outro fator a ser considerado diz respeito às distribuições de probabilidade das diferentes variáveis medidas empregadas nas inferências. Como os comportamentos estatísticos das diferentes medidas podem ser muitos distintos, o comportamento estatístico da inferência pode ser bastante complexo. Neste contexto, em que as inferências podem ser contaminadas por fatores extremamente complexos, faz-se imprescindível o monitoramento em tempo real dos modelos inferenciais e dos valores das variáveis medidas. Essa tarefa deve ser realizada sob a luz de ferramentas estatísticas, garantindo que ações de correção e de controle sejam tomadas somente

quando realmente forem necessárias. Nada parece ter sido proposto na literatura ainda para realizar essa atividade de forma estatisticamente rigorosa.

3.2 ESTUDO DE CASO: INFERÊNCIA DA T85% ASTM DO DIESEL PRODUTO

A qualidade do diesel e de outros produtos destilados da indústria de petróleo é analisada de acordo com a norma desenvolvida pela ASTM (*American Society for Testing and Materials*), intitulada Métodos de Ensaio para Destilação de Produtos do Petróleo (*Test Method for Distillation of Petroleum Products*), mais conhecida como ASTM D 86. De acordo com esta norma, a T85% é a principal especificação do diesel produto.

A T85% corresponde à temperatura em que 85% em volume são recolhidos na destilação a pressão atmosférica de uma amostra do produto. Sabe-se que esta temperatura é bastante influenciada pelo teor de componentes pesados contidos na amostra. Desta forma, supõe-se que o diesel produto influencia bastante esta especificação, por ser o componente mais denso do processo.

A Figura (3.1) apresenta o fluxograma simplificado da rota de destilação do diesel produto (MIRANDA e LUSA, 2003).

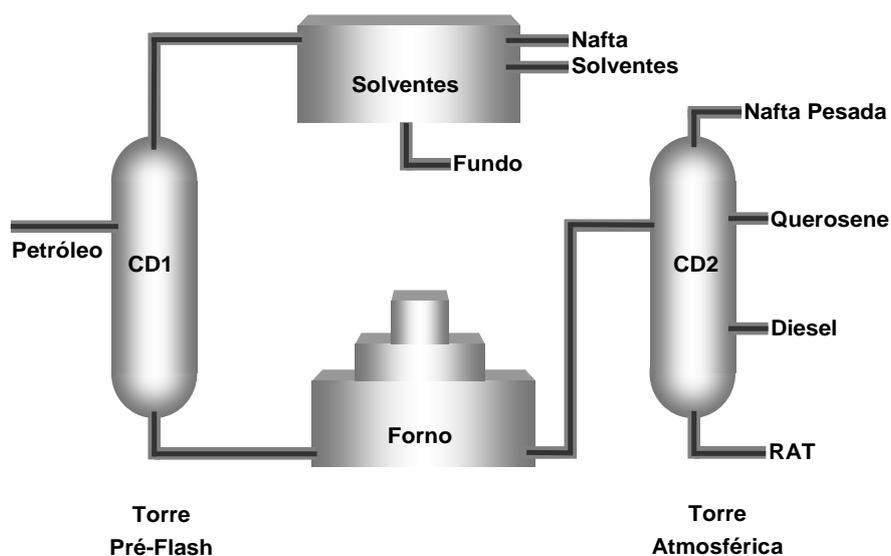


Figura 3.1: Fluxograma simplificado do processo de destilação do diesel produto.

O diesel produto é composto pela mistura das correntes de diesel e de querosene, produzidos na torre atmosférica CD2, e da corrente de fundo da unidade de solventes.

Embora a T85% seja uma propriedade-chave na definição do ponto ótimo operacional da coluna atmosférica, não há um analisador em linha que realize a sua medição. A especificação desta propriedade é restrita aos ensaios de laboratório, com períodos longos de amostragem (12 horas) e de análise (3 horas). Desta forma, as informações, além de serem infreqüentes, são obtidas com considerável defasagem, restringindo estratégias de controle e de otimização do processo, podendo acarretar graves prejuízos ao controle de qualidade do produto.

Para contornar esta situação, ZANIN (1993) desenvolveu um analisador virtual, capaz de acompanhar a qualidade do diesel que é produzido, mesmo na ausência da análise de laboratório.

3.2.1 O Modelo Inferencial

O modelo inferencial aqui descrito é utilizado para estimar o ponto 85% da curva de destilação ASTM D-86 do diesel produto (T85%) em linha. As inferências são calculadas por meio de uma estrutura empírica, que consiste em uma equação simplificada derivada da curva típica de destilação de uma coluna de destilação atmosférica. A estrutura utiliza como variáveis de entrada as temperaturas e vazões das bandejas de retiradas de produto da torre atmosférica.

A inferência implementada é obtida por meio de um modelo empírico, de acordo com a seguinte seqüência de cálculos apresentados a seguir (ZANIN, 1993).

a) Cálculo da temperatura 30% do diesel, a partir da temperatura da bandeja de retirada do diesel:

$$T_{30d} = (i_d \cdot T_d) + b_d \quad (3.1)$$

onde T_d é a temperatura 30% do diesel, T_d é a temperatura da bandeja de retirada do diesel e i_d e b_d são parâmetros ajustados com dados de laboratório.

b) Cálculo da temperatura 30% do RAT (resíduo atmosférico) a partir da temperatura da bandeja de retirada do RAT:

$$T_{30RAT} = (i_{RAT} \cdot T_{RAT}) + b_{RAT} \quad (3.2)$$

onde T_{30RAT} é a temperatura 30% do RAT, T_{RAT} é a temperatura da bandeja de retirada do RAT e i_{RAT} e b_{RAT} são parâmetros ajustados com dados de laboratório.

c) Cálculo da temperatura 85% do diesel a partir da temperatura 30% do RAT, da temperatura 30% do diesel, da vazão de RAT e da vazão de diesel:

$$V_d = (0,85 - 0,30) \cdot f_d / (0,7 \cdot f_d + 0,3 \cdot f_{RAT}) \quad (3.3)$$

$$T_{85d} = (T_{30RAT} - T_{30d}) \cdot V_d + T_{30d} + b \quad (3.4)$$

$$T_{85dF} = (T_{30RAT} - T_{30d}) \cdot V_d + T_{30d} + b \quad (3.5)$$

onde T_{85d} é a temperatura 85% do diesel, f_{RAT} é a vazão volumétrica de retirada do RAT, f_d é a vazão volumétrica de retirada do diesel, V_d é o volume correspondente do diesel, T_{85dF} é a temperatura 85% final calculada do diesel e b é o *bias*.

O *bias* é utilizado para corrigir a inferência com um resultado de laboratório e é calculado iterativamente na forma:

$$b = b + cor_b \quad (3.6)$$

$$cor_b = T_{85lab} - T_{85d} \quad (3.7)$$

onde Cor_b é o valor de correção do *bias* e T_{85lab} é a temperatura 85% do diesel produto medida em laboratório.

A inferência é recalculada a cada minuto com a Equação (3.5). Porém, a cada 12 horas, quando um novo resultado de laboratório é disponibilizado, um novo valor de *bias* pode ser automaticamente recalculado com as Equações (3.6) e (3.7). O modelo inferencial da T85% empregado em linha utiliza 19 variáveis de entrada medidas na planta e 5 parâmetros ($i_d, q_d, i_{RAT}, q_{RAT}, b$), sendo que um deles (o *bias*) é atualizado em tempo real.

3.2.2 O Problema do *BIAS*

Durante a operação do processo, a cada inferência realizada é feita uma comparação com a última medida de laboratório, que pode ter sido obtida 12 horas antes, por exemplo. Se o resultado inferido em linha parecer confiável, a correção (*bias*) da inferência deve ser descartada; caso contrário, o *bias* deve ser adicionado ao valor inferido. É importante ressaltar que esta decisão a respeito da confiabilidade da inferência é feita apenas qualitativamente, considerando a experiência dos profissionais da planta (neste caso, os operadores), sem o suporte de quaisquer ferramentas matemáticas e/ou estatísticas.

O acréscimo automático do *bias* ao modelo pode ser muito questionado. Vários argumentos atestam que existe a possibilidade de que este acréscimo aumenta a variabilidade do processo. QUELHAS e PINTO (2009) mostram de forma inequívoca que o uso do *bias* pode afetar a estatística do procedimento de inferência, muito significativamente, prejudicando o desempenho do inferidor. Em contrapartida, há argumentos que defendem o emprego do *bias* como um elemento calibrador do modelo e, portanto, como um meio de obter previsões mais confiáveis.

3.3 METODOLOGIA

As variáveis de processo são medidas para monitoramento de desempenho e implementação de procedimentos de otimização e controle. Com os atuais avanços nos

sistemas de automação industrial, um grande número de variáveis de processo é simultaneamente medida e armazenada, podendo as medidas ser facilmente acessadas, facilitando o uso de eficientes estratégias de operação em tempo real.

É importante ressaltar que os cálculos de grandezas estatísticas estão sempre vinculados à distribuição de probabilidade admitida para descrever o comportamento dos dados. A escolha mais comum consiste em admitir que os dados experimentais seguem uma distribuição normal de probabilidades (SCHWAAB, 2007). Apesar de ser difícil mostrar que um conjunto de dados é ou não é normalmente distribuído (RATKOWSKY, 1990), esta hipótese é geralmente razoável por causa das características intrínsecas da distribuição normal (BARD, 1974):

- i. Geralmente apresenta um comportamento próximo ao de muitas medidas experimentais.
- ii. Com o aumento do número de medidas, muitas distribuições aproximam-se da distribuição normal.
- iii. Definidas a média e a variância de um conjunto de medidas e usando os conceitos do cálculo variacional, é possível mostrar que a distribuição normal é a que insere a menor quantidade de informações extras ao problema.
- iv. O tratamento matemático é fácil, permitindo a definição das distribuições t de Student, χ^2 e F de Fisher, de fundamental importância para o estabelecimento de intervalos de confiança dos parâmetros e para os testes da qualidade do ajuste do modelo aos dados.

Deste modo, pode-se justificar o uso comum da distribuição normal para representação de dados industriais, já que estes são obtidos por meio dos registros históricos que contêm centenas de variáveis operacionais em grandes conjuntos amostrais. Portanto, a maioria das abordagens estatísticas empregadas atualmente na indústria baseia-se na hipótese de que as medidas seguem a distribuição Normal. De acordo com CRAMER (1946), *“Todos acreditam na lei de distribuição Normal de erros; os experimentalistas porque pensam que isto é um teorema matemático, e os matemáticos porque pensam que isto é um fato experimental”*. Nesta dissertação, admite-se que os dados apresentam uma distribuição Normal de probabilidades.

A realização das análises estatística do processo, em particular da inferência, foi dividida em três etapas: caracterização estatística básica, caracterização dinâmica e caracterização da predição. Na etapa de caracterização estatística básica são identificados os referenciais operacionais do processo. Para tal são empregadas médias, variâncias, covariâncias e correlações.

Já a etapa de caracterização dinâmica do processo busca a averiguação da presença de dinâmica no processo. Essa etapa é importante porque modelos inferenciais são usualmente projetados para trabalharem em condições estacionárias, embora estas condições sejam raramente encontradas em plantas. Aqui são empregadas médias e variâncias móveis, com seus respectivos intervalos de confiança, averiguação de conformidade da média e da variância, espectros de variância e espectros de auto-correlação.

Finalmente, na etapa de caracterização da predição são realizadas as análises empregando a representação matemática do processo (ou mais propriamente dito o modelo inferencial), a fim de avaliar a sua confiabilidade.

A seguir será descrita a seqüência lógica de etapas adotada para a realização com êxito da análise estatística de uma inferência. Assim, supõe-se que esteja disponível certo conjunto de dados, contendo as variáveis independentes e dependentes, sobre as quais será aplicado um modelo para a inferência das variáveis e posterior análise destas inferências.

3.3.1 Inferências

As inferências são feitas por meio de avaliações do modelo, para o qual são fornecidos os valores dos parâmetros $\underline{\theta}$ e das variáveis de entrada \underline{x} (independentes) em um dado instante, sendo então calculados os valores das variáveis de saída \underline{y} (dependentes),

$$\underline{\hat{y}} = f(\underline{x}, \underline{\theta}) \quad (3.8)$$

onde \hat{y} é o vetor de predição da variável dependente, \underline{x} é o vetor de medidas da variável independente, $\underline{\theta}$ é o vetor de valores dos parâmetros e $f(\underline{x}, \underline{\theta})$ é a função que os relaciona. Do ponto de vista de aplicação, $f(\underline{x}, \underline{\theta})$ é o que caracteriza o processo particular que está sendo analisado. Do ponto de vista de programação, $f(\underline{x}, \underline{\theta})$ é um procedimento numérico que deve ser fornecido pelo usuário.

3.3.2 Resíduos de inferências

Os resíduos da inferência são os desvios observados entre os valores medidos e os valores calculados pelo modelo para as variáveis de saída y (dependentes), definido como:

$$e_{ik} = y_{ik}^m - y_{ik}^c \quad (3.9)$$

onde e_{ik} é o resíduo da variável dependente i , y_{ik}^m e y_{ik}^c são, respectivamente, o valor medido e o valor calculado da variável dependente i e k o instante de amostragem.

3.3.3 Caracterização Básica

Esta etapa é dedicada à definição estatística do processo. Para tal, são realizados os cálculos das médias e das matrizes de covariância e de correlação. O objetivo destas análises é determinar os referenciais estatísticos da operação do processo.

A média de cada variável z_i (z pode ser tanto uma variável independente x como uma variável dependente y) é calculada pela seguinte equação (SCHWAAB e PINTO, 2007):

$$\bar{z}_i = \frac{\sum_{k=1}^N z_{ik}}{N} \quad (3.10)$$

onde z_{ik} é a variável i no instante k , \bar{z}_i é o valor médio da variável i e N é o número de amostragens disponíveis para a variável z_i no intervalo considerado.

Para caracterizar a variabilidade do conjunto de dados em torno do valor médio, são calculadas as matrizes de covariância e de correlação, conforme definido abaixo:

$$\mathbf{V} = \begin{bmatrix} v_{11}^2 & v_{12}^2 & \cdots & v_{1M}^2 \\ v_{21}^2 & v_{22}^2 & \cdots & v_{2M}^2 \\ \vdots & \vdots & \ddots & \vdots \\ v_{M1}^2 & v_{M2}^2 & \cdots & v_{MM}^2 \end{bmatrix} \quad (3.11)$$

onde \mathbf{V} é a matriz de covariâncias e cada termo v_{ij}^2 é calculado por meio da seguinte equação:

$$v_{ij}^2 = \frac{\sum_{k=1}^N (z_{ik} - \bar{z}_i)(z_{jk} - \bar{z}_j)}{N-1} \quad (3.12)$$

Então, o desvio-padrão pode ser calculado pela seguinte equação:

$$\sigma_{ii} = \sqrt{\frac{\sum_{k=1}^N (z_{ik} - \bar{z}_i)^2}{N-1}} \quad (3.13)$$

A matriz de correlações é definida como:

$$\mathbf{R} = \begin{bmatrix} r_{11}^2 & r_{12}^2 & \cdots & r_{1M}^2 \\ r_{21}^2 & r_{22}^2 & \cdots & r_{2M}^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1}^2 & r_{M2}^2 & \cdots & r_{MM}^2 \end{bmatrix} \quad (3.14)$$

onde \mathbf{R} é a matriz de correlação e cada termo r_{ij}^2 é calculado por meio da seguinte equação:

$$r_{ij}^2 = \frac{v_{ij}^2}{\sqrt{v_{ii}^2 v_{jj}^2}} \quad (3.15)$$

As médias e as matrizes de covariância e de correlação dos resíduos podem ser calculadas por meio das Equações (3.8) a (3.15), substituindo-se as variáveis z pelos resíduos e .

As médias e as variâncias são utilizadas para caracterização de desempenho dos modelos de inferência em vários momentos, a depender dos procedimentos particulares considerados. Particularmente, estas grandezas são fundamentais para caracterização de natureza estatística dos dados utilizados para fins de monitoramento, fornecendo importantes informações de entrada de vários procedimentos numéricos.

3.3.4 Caracterização Dinâmica

Esta etapa de caracterização tem como objetivo principal a investigação da presença de dinâmica no processo. Desta forma, são calculadas médias e variâncias móveis, e seus respectivos intervalos de confiança. Calculam-se ainda os espectros de covariância e de correlação.

As médias móveis podem ser usadas para acompanhar os dados do processo como flutuações filtradas devido ao fato destas não oscilarem tanto como os dados reais. A seguinte equação apresenta o cálculo das médias móveis,

$$\overline{z}m_{i,j} = \frac{1}{TJ} \sum_{k=i-TJ+1}^i z_{k,j} \quad (3.16)$$

onde j indica a variável, i indica o instante de amostragem superior, TJ é o tamanho da janela e $\overline{z}m_{i,j}$ representa a média móvel da variável z_i na janela móvel i .

É importante lembrar que o tamanho da janela (TJ) tem uma relação proporcional aos valores das flutuações; ou seja, quanto maior o valor de TJ usado para o cálculo das médias móveis, mais amenizadas ficam as flutuações do processo, uma vez que, a variância da média móvel é TJ vezes menor que a variância da medida de um dado isolado.

O cálculo das variâncias móveis é uma ferramenta importante para identificação de estados estacionários e transientes de processo. Calculadas as variâncias móveis, o conjunto de dados que apresenta a menor variância é em geral o melhor candidato do período analisado para identificação da operação estacionária do processo.

As variâncias móveis são calculadas por meio da equação

$$vm_{i,j}^2 = \frac{1}{TJ-1} \sum_{k=i-TJ+1}^i (z_{k,j} - \overline{zm}_{i,j})^2 \quad (3.17)$$

onde j indica a variável, i indica instantes de amostragem, TJ é o tamanho da janela, $\overline{zm}_{i,j}$ representa a média móvel calculada para a variável j na janela móvel i , z são as variáveis e $vm_{i,j}^2$ representa a variância móvel da variável j na janela de amostragem.

Para o cálculo dessas grandezas estatísticas, média e variância móveis, as janelas móveis são deslocadas de uma em uma durante a medida, de forma que, para um número total de amostras igual a Npt , o número de janelas NJ para um tamanho de janela de TJ é igual a $Npt-TJ+1$.

É importante salientar que médias e variâncias móveis não devem ser utilizadas de forma indiscriminada, sem que se faça uma análise detalhada dos dados e das condições de operação, já que as médias e variâncias móveis eliminam artificialmente efeitos dinâmicos dos conjuntos de dados amostrados. Portanto, o uso das médias e variâncias móveis é recomendado principalmente quando em que o usuário pode garantir que o processo operava de forma estável, em condições próximas da condição estacionária.

O intervalo de confiança das médias e das variâncias móveis fornece uma informação muito importante a respeito da confiabilidade destas medidas. O intervalo de confiança diz respeito ao quão próximo estas grandezas estatísticas estão dos seus valores amostrados. Quando o intervalo de confiança obtido é largo, muitos outros valores também distintos dos valores medidos podem ser considerados aceitáveis; logo, sua confiabilidade é baixa. Porém, quando o intervalo de confiança obtido é estreito, os valores admissíveis são próximos dos valores medidos; logo, sua confiabilidade é alta.

A definição do intervalo de confiança das médias móveis faz uso da distribuição estatística *t* de *Student* (SCHWAAB e PINTO, 2007). O cálculo do intervalo de confiança da média consiste em definir quais são os valores máximo e mínimo possíveis para a média de cada janela móvel empregando a distribuição *t*, de acordo com o número de amostras e o nível de confiança desejado.

As seguintes equações são empregadas para o cálculo dos limites inferior e superior da média,

$$\overline{zm}_{i,j,\text{mín}} = \overline{zm}_{i,j} - \sqrt{\frac{\overline{vm}_{i,j}^2}{TJ}} t_{TJ-1,1+\alpha/2} \quad (3.18)$$

$$\overline{zm}_{i,j,\text{máx}} = \overline{zm}_{i,j} + \sqrt{\frac{\overline{vm}_{i,j}^2}{TJ}} t_{TJ-1,1+\alpha/2} \quad (3.19)$$

onde $t_{n,p}$ representa a distribuição *t* de Student com *n* graus de liberdade e probabilidade *p*. As demais variáveis foram definidas anteriormente.

Já a definição do intervalo de confiança das variâncias móveis emprega a distribuição estatística χ^2 (SCHWAAB e PINTO, 2007). O cálculo do intervalo de confiança da variância consiste em definir quais são os valores máximo e mínimo possíveis para a variância de cada janela móvel empregando a distribuição χ^2 , dado um conjunto de pontos, de acordo com o número de amostras e o nível de confiança esperado.

As equações a seguir são empregadas para o cálculo dos limites inferior e superior da variância,

$$\overline{vm}_{i,j,\text{mín}}^2 = (TJ - 1) \cdot \frac{\overline{vm}_{i,j}^2}{\chi_{TJ-1,1+\alpha/2}^2} \quad (3.20)$$

$$\overline{vm}_{i,j,\text{máx}}^2 = (TJ - 1) \cdot \frac{\overline{vm}_{i,j}^2}{\chi_{TJ-1,\alpha/2}^2} \quad (3.21)$$

onde $\chi^2_{n,p}$ representa a distribuição χ^2 com n graus de liberdade e probabilidade p . As demais variáveis foram definidas anteriormente.

É importante lembrar que as comparações de intervalos de confiança, tanto para a média quanto para a variância, requerem dados (e, portanto, janelas móveis) independentes. Neste caso, as janelas móveis devem ser deslocadas de uma em uma janela durante a medida, de forma que, para um número total de amostras igual a N_{pt} , o número de janelas NJ móveis independentes para um tamanho de janela de TJ é igual a N_{pt}/TJ .

Referentes à análise de média e de variâncias móveis há ainda a análise conformidade destas. A análise de conformidade (ou análise comparativa) consiste em realizar comparações entre grandezas estatísticas de um mesmo tipo (média com médias e variância com variâncias) janela a janela, ao longo do período de amostragem, para verificar se estas podem ser consideradas similares entre si ou não.

Quando a grandeza estatística a ser “julgada” é a média, utiliza-se para o cálculo da conformidade desta a distribuição t de Student. Quando se trata da variância emprega-se a distribuição F de Fisher. Se o processo opera em estado estacionário, estas grandezas devem apresentar uma similaridade entre si; ou seja, a média da primeira janela móvel deve ser considerada igual à média das demais janelas e a variância da primeira janela deve ser considerada igual à variância das demais janelas, ao longo de todo o período amostral. É importante ressaltar que é possível obter resultados de conformidade positivos para uma grandeza e negativos para outra, dentro de um mesmo período amostral de operação. Este resultado é possível porque média e variância avaliam característica distintas da distribuição de dados: o ponto central da oscilação e a magnitude de oscilação.

A seguinte equação pode ser usada para calcular a conformidade da média,

$$\bar{z}m_{i,j} - \bar{z}m_{i,j+1} > \sqrt{\frac{S_{i,j}^2}{TJ}} t_{TJ-1,1+\alpha/2} + \sqrt{\frac{S_{i,j+1}^2}{TJ}} t_{TJ-1,1+\alpha/2} \quad (3.22)$$

onde as variáveis já forma definidas anteriormente. Se a desigualdade for satisfeita, diz-se que as médias são distintas com grau de confiança α . Caso contrário, as médias são ditas estatisticamente equivalentes.

A seguinte equação pode ser usada para cálculo da conformidade da variância,

$$F_{TJ-1, TJ-1, \alpha/2} > \frac{vm_{i,j}^2}{vm_{i,j+1}^2} > F_{TJ-1, TJ-1, 1+\alpha/2} \quad (3.23)$$

onde $F_{n,m,p}$ representa a distribuição F de Fisher, com graus de liberdade n e m para os dois conjuntos de dados e probabilidade p . Se as desigualdades acima forem satisfeitas, diz-se que as variâncias são estatisticamente equivalentes; caso contrário, diz-se que as variâncias são distintas com grau de confiança p .

No tocante a análise da variabilidade do processo, uma outra ferramenta importante é o cálculo do espectro de variância. Os espectros de variância consistem no cálculo de variâncias ao longo de todo um horizonte de amostragens, considerando intervalos independentes dos dados usados para o cálculo das variâncias. Portanto, semelhante à análise anterior (conformidade de médias e de variâncias), esta análise também utiliza a estratégia de janelas móveis independentes.

Para cada janela i a variância é calculada por meio das equações,

$$v_{i,j}^2 = \frac{1}{TJ-1} \sum_{k=1+(i-1)TJ}^{iTJ} (z_{k,j} - \bar{z}_{i,j})^2 \quad (3.24)$$

$$\bar{z}_{i,j} = \frac{1}{TJ} \sum_{k=1+(i-1)TJ}^{iTJ} z_{k,j} \quad (3.25)$$

onde as variáveis são como definidas anteriormente. Uma vez calculadas as variâncias nas NJ janelas independentes, pode ser calculada a média das variâncias calculadas em cada janela i com tamanho TJ , de acordo com a seguinte equação:

$$\bar{v}_{TJ,j}^2 = \frac{1}{NJ} \sum_{i=1}^{NJ} v_{i,j}^2 \quad (3.26)$$

A grandeza $\bar{v}_{TJ,j}^2$ representa a variância dos dados ao longo de todo o período considerado com $Npt - NJ$ graus de liberdade. Obtém-se o espectro de variâncias quando TJ é variado, sendo as variâncias obtidas como função do tamanho da janela.

Os espectros de variância são úteis para dizer se os dados são uniformes no tempo (variâncias essencialmente constantes) ou não. Quando as variâncias são uniformes, conclui-se que os dados analisados não experimentaram qualquer dinâmica; caso contrário, as variâncias tendem a aumentar com o tamanho da janela móvel (por incorporarem a dinâmica do processo). Neste caso, a variância mínima pode ser associada à variância típica da medição, enquanto a variância máxima pode ser associada à variância da operação do processo como um todo. Desta forma, o espectro de variâncias pode constituir uma importante ferramenta de análise de dados.

Os espectros de correlação consistem no cálculo de coeficientes de correlação quando uma das variáveis é deslocada no tempo, mantendo-se a outra variável fixa. Esta ferramenta estatística possibilita o acompanhamento de como uma variável influencia a outra com o passar do tempo, permitindo a identificação das constantes de tempo típicas do processo (quando ainda há correlação significativa, apesar dos deslocamentos) e de constantes de atraso (quando a máxima correlação entre duas variáveis é obtida para deslocamentos diferentes de zero). Quando o espectro de correlação é calculado para uma variável deslocada no tempo com relação a ela mesma, então chamado espectro de auto-correlação, o resultado pode ser usado para identificação do tempo normalmente requerido para que seja atingido um estado estacionário; ou seja, o tempo requerido para que as mudanças tenham caráter aleatório e não caráter dinâmico. Diz-se que há dinâmica, quando a correlação de dados deslocados no tempo para uma mesma variável é significativa.

Do ponto de vista do algoritmo, deve-se primeiramente ser calculado o número de deslocamentos que podem ser realizados. Este valor é definido como o menor valor entre 15 (definido de forma heurística como o valor mínimo para que a análise possa ser efetuada com segurança) e $2/3$ de Npt (definido de forma heurística para evitar um

deslocamento excessivo, quando o número de dados é pequeno). O menor destes dois valores é chamado de $Nmax$. Em seguida, varia-se k de 0 até $Nmax$, para um certo par de variáveis i e j , obtendo-se o espectro de correlações ou auto-correlações. Em seguida, são calculadas as médias para a variável i e para a variável j nos intervalos considerados, sendo que apenas a variável j é deslocada no tempo,

$$\bar{z}_{k0,i} = \frac{1}{Npt - k} \sum_{n=1}^{Npt-k} z_{n,i} \quad (3.27)$$

$$\bar{z}_{k+,j} = \frac{1}{Npt - k} \sum_{n=1+k}^{Npt} z_{n,j} \quad (3.28)$$

onde as variáveis foram definidas anteriormente. Deve-se perceber que é necessário reduzir o conjunto de dados de Npt para $Npt-k$ pontos, para manter a coerência da análise.

As seguintes equações podem ser usadas para calcular para as variâncias e as covariâncias, nos intervalos considerados.

$$v_{k0,ii}^2 = \frac{1}{Npt - k - 1} \sum_{n=1}^{Npt-k} (z_{n,i} - \bar{z}_{k0,i})^2 \quad (3.29)$$

$$v_{k+,jj}^2 = \frac{1}{Npt - k - 1} \sum_{n=1+k}^{Npt} (z_{n,j} - \bar{z}_{k+,j})^2 \quad (3.30)$$

$$cov_{k,i,j}^2 = \frac{1}{Npt - k - 1} \sum_{n=1}^{Npt-k} (z_{n,i} - \bar{z}_{k0,i})(z_{n+k,j} - \bar{z}_{k+,j}) \quad (3.31)$$

A correlação com um deslocamento k para as variáveis i e j é então calculada pela seguinte equação:

$$\rho_{k,i,j} = \frac{cov_{k,i,j}^2}{\sqrt{v_{k0,ii}^2 v_{k+,jj}^2}} \quad (3.32)$$

Todas estas análises pertinentes à etapa de caracterização estatística dinâmica também são calculadas para os resíduos, substituindo as variáveis z pelos resíduos e . Desta forma, é possível identificar se o desempenho do modelo é uniforme ou se varia ao longo do tempo.

3.3.5 Caracterização da Predição

A caracterização da predição é realizada empregando exclusivamente os resíduos da inferência. As ferramentas estatísticas utilizadas nesta etapa avaliam diretamente a qualidade da inferência com relação ao modelo inferencial. A princípio, são realizados testes estatísticos para avaliação do modelo inferencial. Os testes empregados são: o teste-t, o teste-F e o teste- χ^2 .

Foi realizada uma padronização dos resultados dos testes apresentados nesta etapa, empregando-se índices e atribuindo a estes significados estatísticos específicos para cada teste. Desta forma, convencionou-se que os índices empregados podem assumir os valores -1, 0 e +1. O objetivo desses índices é apresentar os resultados de forma mais simples e clara, de modo que decisões e ações de correção possam ser tomadas em tempo real por profissionais que não necessariamente entendam extensivamente de estatísticas de processos.

O teste-t é usado para análise da média dos desvios dos resíduos, verificando se existem desvios significativos em relação ao valor zero. O valor 0 indica que não há desvios significativos, o valor -1 indica que há um desvio negativo significativo e o valor +1 indica que há um desvio positivo significativo. Quando é detectado um desvio significativo, a ação que pode vir a ser tomada é a aplicação de um *bias* para deslocar toda a inferência, de modo a eliminar o efeito deste desvio significativo.

Este teste consiste em verificar se a seguinte desigualdade é satisfeita,

$$|e_i| \leq \sqrt{\frac{S_{ii}^2}{N}} t_{N-1, 1-\alpha/2} \quad (3.33)$$

onde as variáveis forma definidas anteriormente.

O teste F é usado para a comparação de duas variâncias; isto é, mostrar se duas variâncias independentes são significativamente diferentes entre si. Para este teste é utilizada a função de distribuição de probabilidade F de *Fisher*. Neste caso, o teste é usado para comparar a variância dos resíduos com a variância das medidas. Este teste avalia se os desvios entre o modelo e os dados amostrados são significativamente maiores que os desvios experimentais, indicando que o modelo pode não ser adequado. Também é avaliado se os desvios observados entre o modelo e os dados amostrados são significativamente menores que os desvios experimentais, indicando que o modelo pode estar super parametrizado. Como este teste é realizado nas variâncias dos resíduos de cada variável, é possível observar o desempenho do modelo com relação a cada uma das respostas do modelo. Se os desvios do modelo apresentam variância superior à dos dados experimentais, o modelo pode e deve ser aprimorado. Se os desvios do modelo apresentam variância inferior à dos dados experimentais, o modelo pode e deve ser simplificado. Nesse teste, o índice 0 indica que as variâncias dos resíduos e das medidas são similares. O índice -1 indica que a variância dos resíduos é significativamente menor que as variâncias da amostragem, indicando uma super parametrização do modelo ou que as variâncias da amostragem estão sendo influenciadas pela dinâmica do sistema. Já o índice +1 indica que a variância do resíduo é significativamente maior que as variâncias da amostragem, indicando que a inferência oscila mais que os dados amostrado, mostrando uma incapacidade do modelo em representar os dados.

O teste F consiste em verificar se a seguinte desigualdade é satisfeita,

$$F_{N-1, N-1}^{\alpha/2} \leq \frac{S_{ii}^2}{v_{ii}^2} \leq F_{N-1, N-1}^{1-\alpha/2} \quad (3.34)$$

onde S_{ii}^2 e v_{ii}^2 são respectivamente as variâncias dos resíduos e dos dados medidos da variável i , e as demais variáveis já foram definidas. O limite inferior é usado quando $S_{ii}^2 < v_{ii}^2$ e o limite superior é usado quando $S_{ii}^2 > v_{ii}^2$.

O teste- χ^2 é empregado para verificar o desempenho global do modelo, considerando todas as respostas do modelo no conjunto. Novamente, é possível verificar

se o modelo não é adequado ou se existe alguma super parametrização do mesmo, de forma similar ao teste anterior. Neste teste, todas as variáveis da inferência são consideradas e, assim, o modelo como um todo é avaliado. Quando a resposta deste teste é igual a 0, o desempenho do modelo está adequado aos dados operacionais. O índice -1 indica que a qualidade do modelo é muito alta; ou seja, pode indicar modelo super parametrizado ou variâncias amostrais muito grandes e mal avaliadas. Mas se o resultado do teste assume índice +1, a qualidade da inferência está provavelmente muito baixa, sendo necessária a reestimação dos parâmetros do modelo ou um aprimoramento do modelo visado.

É definida a soma ponderada dos resíduos,

$$SR = \sum_{k=1}^N \mathbf{e}_k^T \mathbf{V}_y^{-1} \mathbf{e}_k \quad (3.35)$$

onde \mathbf{e} é um vetor com dimensão NY contendo os resíduos no tempo atual de medição e \mathbf{V}_y é a matriz de covariância dos experimentos. Admitindo-se que os resíduos sejam variáveis com distribuição normal, a variável SR pode ser interpretada como uma variável com distribuição χ^2 com NY graus de liberdade. O seguinte teste pode então ser formulado para verificar a significância do valor de SR :

$$\chi_{NY, \alpha/2}^2 \leq SR \leq \chi_{NY, 1-\alpha/2}^2 \quad (3.36)$$

O ideal é que o valor de SR esteja compreendido no intervalo dado pela Equação (3.36). Se o limite superior é ultrapassado, o valor de SR é alto demais e indica uma predição ruim do modelo. Se o limite inferior é ultrapassado, o valor de SR é pequeno demais; ou seja, o ajuste está bom demais e indica que o modelo pode estar super parametrizado.

Também são calculadas as matrizes de sensibilidade do modelo, que correspondem às derivadas primeiras do modelo, em relação aos parâmetros e às variáveis independentes. A matriz de sensibilidade com relação aos parâmetros é definida como,

$$\mathbf{B}_{\theta,k} = \begin{bmatrix} \frac{\partial y_{1,k}}{\partial \theta_1} & \frac{\partial y_{1,k}}{\partial \theta_2} & \dots & \frac{\partial y_{1,k}}{\partial \theta_p} \\ \frac{\partial y_{2,k}}{\partial \theta_1} & \frac{\partial y_{2,k}}{\partial \theta_2} & \dots & \frac{\partial y_{2,k}}{\partial \theta_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{NY,k}}{\partial \theta_1} & \frac{\partial y_{NY,k}}{\partial \theta_2} & \dots & \frac{\partial y_{NY,k}}{\partial \theta_p} \end{bmatrix} \quad (3.37)$$

onde $\mathbf{B}_{\theta,k}$ é a matriz de sensibilidade com relação aos parâmetros θ no instante k .

Já a matriz de sensibilidade com relação às variáveis independentes é definida como,

$$\mathbf{B}_{x,k} = \begin{bmatrix} \frac{\partial y_{1,k}}{\partial x_{1,k}} & \frac{\partial y_{1,k}}{\partial x_{2,k}} & \dots & \frac{\partial y_{1,k}}{\partial x_{NX,k}} \\ \frac{\partial y_{2,k}}{\partial x_{1,k}} & \frac{\partial y_{2,k}}{\partial x_{2,k}} & \dots & \frac{\partial y_{2,k}}{\partial x_{NX,k}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_{NY,k}}{\partial x_{1,k}} & \frac{\partial y_{NY,k}}{\partial x_{2,k}} & \dots & \frac{\partial y_{NY,k}}{\partial x_{NX,k}} \end{bmatrix} \quad (3.38)$$

onde $\mathbf{B}_{x,k}$ é a matriz de sensibilidade com relação às variáveis independente x no instante k .

O cálculo de cada uma das derivadas acima é feito por diferenças finitas centrais, como definido nas equações abaixo,

$$\frac{\partial y_{i,k}}{\partial \theta_j} = \frac{y_{i,k}(\theta_j + \delta\theta_j) - y_{i,k}(\theta_j - \delta\theta_j)}{2\delta\theta_j} \quad (3.39)$$

$$\frac{\partial y_{i,k}}{\partial x_{j,k}} = \frac{y_{i,k}(x_{j,k} + \delta x_{j,k}) - y_{i,k}(x_{j,k} - \delta x_{j,k})}{2\delta x_{j,k}} \quad (3.40)$$

onde δ é um número pequeno, e corresponde ao valor da perturbação relativa que é feita para o cálculo das derivadas.

Para caracterizar a qualidade dos valores preditos, é calculada a matriz de covariância de predição $\hat{\mathbf{V}}_{y,k}$ (BARD, 1974). Esta matriz pode ser calculada a cada instante k , levando-se em consideração a matriz de covariância dos parâmetros $\hat{\mathbf{V}}_{\theta}$, a matriz de covariância experimental das variáveis independentes $\mathbf{V}_{x,k}$ e a matriz de covariância experimental das variáveis dependentes $\mathbf{V}_{y,k}$, como mostra a equação abaixo,

$$\hat{\mathbf{V}}_{y,k} = \mathbf{B}_{\theta,k} \hat{\mathbf{V}}_{\theta} \mathbf{B}_{\theta,k}^T + \mathbf{B}_{x,k} \mathbf{V}_{x,k} \mathbf{B}_{x,k}^T + \mathbf{V}_{y,k} \quad (3.41)$$

onde $\mathbf{B}_{\theta,k}$ e $\mathbf{B}_{x,k}$ são as matrizes de sensibilidade definidas nas Equações (3.37) e (3.38), respectivamente. A matriz de covariância dos parâmetros $\hat{\mathbf{V}}_{\theta}$ é obtida durante o procedimento de regressão do modelo (estimação de parâmetros e/ou reconciliação de dados) e deve ser considerada como um dado de entrada do problema. Eventualmente, esse dado pode ser gerado em linha e em tempo real, se uma estratégia adaptativa for utilizada.

Esta análise também é realizada para identificação dos elementos que concentram os erros e para identificação das medidas que mais influenciam os erros de predição. Para tal, a variância de predição é calculada sem o efeito de uma determinada variável, fazendo com que na matriz \mathbf{V}_x a linha j e a coluna j correspondente a uma variável x_j sejam zeradas. Assim, a componente da variância de predição correspondente à variável x_j é anulada. As variâncias de predição com e sem uma determinada variável podem ser comparadas através do teste F. Se as variâncias forem significativamente diferentes, pode-se dizer que a variável x_j tem um efeito dominante no erro de predição. Essa informação pode ser preciosa para a interpretação dos resultados obtidos com um modelo, sugerindo que medidas devam ser melhoradas na unidade real de produção. O mesmo procedimento é realizado para os parâmetros, a fim de verificar a influência destes elementos no erro de predição.

Para esta análise de averiguação de influência das variáveis no erro de predição, os resultados também são expressos em índices. Se for obtido índice +1, a variável e/ou

parâmetro cuja variância é anulada exerce um efeito dominante no erro de predição; se o índice for 0, a variância não influencia significativamente o erro de predição. O índice - 1 não deve ser obtido, já que quando se retira o efeito de um certo experimento, espera-se que a variância diminua. No entanto, quando as correlações são fortes, a remoção de um parâmetro pode levar ao aumento dos erros de predição, como demonstrado a seguir.

Seja o modelo,

$$y = a x + b = [a \quad b] \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (3.42)$$

onde a e b são parâmetros que devem ser estimados com o conjunto de dados experimentais.

$$x = \begin{bmatrix} x_1^e \\ x_2^e \\ \vdots \\ x_n^e \end{bmatrix} \quad (3.43)$$

$$y = \begin{bmatrix} y_1^e \\ y_2^e \\ \vdots \\ y_n^e \end{bmatrix} \quad (3.44)$$

Se a técnica de mínimos quadrados é usada para estimar os parâmetros, então,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (x_i^e)^2 & \sum_{i=1}^n (x_i^e) \\ \sum_{i=1}^n (x_i^e) & n \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n (y_i^e x_i^e) \\ \sum_{i=1}^n (y_i^e) \end{bmatrix} \quad (3.45)$$

$$\mathbf{V}_a = \begin{bmatrix} \sigma_a^2 & \sigma_{ab}^2 \\ \sigma_{ab}^2 & \sigma_b^2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n (x_i^e)^2 & \sum_{i=1}^n (x_i^e) \\ \sum_{i=1}^n (x_i^e) & n \end{bmatrix}^{-1} \quad (3.46)$$

onde \mathbf{V}_a é a matriz de covariâncias dos parâmetros. Portanto, a variância de predição (negligenciados os possíveis erros experimentais de medida) da variável calculada a partir da Equação (3.42) pode ser escrita na forma,

$$\hat{\sigma}_y^2 = [x \ 1] \mathbf{V}_a \begin{bmatrix} x \\ 1 \end{bmatrix} = [x \ 1] \begin{bmatrix} \sum_{i=1}^n (x_i^e)^2 & \sum_{i=1}^n (x_i^e) \\ \sum_{i=1}^n (x_i^e) & n \end{bmatrix}^{-1} \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (3.47)$$

$$\hat{\sigma}_y^2 = [x \ 1] \mathbf{V}_a \begin{bmatrix} x \\ 1 \end{bmatrix} = [x \ 1] \begin{bmatrix} \sigma_a^2 & \sigma_{ab}^2 \\ \sigma_{ab}^2 & \sigma_b^2 \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} =$$

$$[x \ 1] \begin{bmatrix} \frac{n}{n \sum_{i=1}^n (x_i^e)^2 - \left[\sum_{i=1}^n (x_i^e) \right]^2} & \frac{-\sum_{i=1}^n (x_i^e)}{n \sum_{i=1}^n (x_i^e)^2 - \left[\sum_{i=1}^n (x_i^e) \right]^2} \\ \frac{-\sum_{i=1}^n (x_i^e)}{n \sum_{i=1}^n (x_i^e)^2 - \left[\sum_{i=1}^n (x_i^e) \right]^2} & \frac{\sum_{i=1}^n (x_i^e)^2}{n \sum_{i=1}^n (x_i^e)^2 - \left[\sum_{i=1}^n (x_i^e) \right]^2} \end{bmatrix} \begin{bmatrix} x \\ 1 \end{bmatrix} \quad (3.48)$$

ou ainda,

$$\hat{\sigma}_y^2 = x^2 \sigma_a^2 + 2 x \sigma_{ab}^2 + \sigma_b^2 \quad (3.49)$$

Deseja-se saber se a remoção de uma fonte de variabilidade (por exemplo, um dos parâmetros a ou b) pode resultar em aumento da variabilidade. Em outras palavras, pergunta-se se:

$$\begin{aligned}
\hat{\sigma}_y^2 &= x^2 \sigma_a^2 + 2 x \sigma_{ab}^2 + \sigma_b^2 \leq x^2 \sigma_a^2 \\
2 x \sigma_{ab}^2 + \sigma_b^2 &\leq 0 \\
x &\leq -\frac{\sigma_b^2}{2 \sigma_{ab}^2}
\end{aligned} \tag{3.50}$$

ou ainda se:

$$\begin{aligned}
\hat{\sigma}_y^2 &= x^2 \sigma_a^2 + 2 x \sigma_{ab}^2 + \sigma_b^2 \leq \sigma_b^2 \\
x^2 \sigma_a^2 + 2 x \sigma_{ab}^2 &\leq 0 \\
-\frac{2 \sigma_{ab}^2}{\sigma_a^2} &\leq x \leq 0
\end{aligned} \tag{3.51}$$

As Equações (3.50) e (3.51) mostram de forma inequívoca que a remoção de fontes de variabilidade pode de fato resultar em aumento dos erros de predição. Para isso, é necessário que as covariâncias entre as fontes de variabilidade do problema sejam não nulas. Como mostrado nas Equações (3.45) a (3.51), isso depende fundamentalmente das condições experimentais usadas para estimar os parâmetros e das condições experimentais usadas para fazer as previsões com o modelo do processo.

Também são realizadas análises que permitem a avaliação do desempenho e a comparação quando vários modelos inferenciais para realização de uma mesma inferência são disponíveis. Este procedimento é geralmente conhecido como discriminação de modelos.

As probabilidades absolutas de cada modelo são obtidas com base na distribuição χ^2 , utilizando a soma dos desvios entre os dados medidos e os dados inferidos, ponderados pela variância de medição de cada variável. Admitindo-se que estes desvios são normalmente distribuídos, esta soma ponderada tem distribuição χ^2 para um certo número de graus de liberdade, que tem valor médio igual ao número de graus de liberdade do modelo inferencial. Deve ser observado que, quando são estimados NP parâmetros a partir do conjunto de dados, o número de graus de liberdade

equivale ao número de variáveis independentes multiplicado ao número de experimentos subtraindo o número de parâmetros do modelo ($NY.NE - NP$).

É definida a soma ponderada dos desvios entre os dados medidos e os dados inferidos como dado pela Equação (3.35). O cálculo da probabilidade absoluta do modelo é feito admitindo-se que a probabilidade de que um modelo represente adequadamente um conjunto de dados é igual à probabilidade de encontrar uma soma ponderada (representada na distribuição chi-quadrado) maior do que a obtida com o modelo (SCHWAAB *et al.*, 2008). Assim, por meio da definição da probabilidade da soma ponderada SR (Equação (3.35)), é possível escrever,

$$P_{abs,m} \equiv 1 - p[\chi^2_{\nu} \leq SR] \quad (3.52)$$

onde χ^2 representa a distribuição chi-quadrado com $\nu = NY.NE$ graus de liberdade.

Com o valor calculado da probabilidade absoluta do modelo, é possível comparar esta probabilidade absoluta com o nível de confiança α (por exemplo, 0,95) exigido para as inferências. Se a probabilidade é menor que o nível de confiança exigido (por exemplo, $0,02 < 1 - \alpha = 1 - 0,95 = 0,05$), existe uma indicação de que o modelo é incapaz de apresentar o desempenho exigido pelo analista, indicando problemas na sua formulação ou a existência de variâncias de processo subestimadas. Caso a probabilidade seja maior que o nível de confiança exigido (por exemplo, $0,97 > \alpha = 0,95$), o desempenho do modelo pode ser considerado exageradamente bom, de maneira que é possível que o modelo esteja super-parametrizado ou que as variâncias estejam superestimadas. Deve ser enfatizado que os cálculos e as conclusões obtidas dependem da definição adequada das variâncias de medição; ou seja, da variância experimental.

Finalmente é realizado o teste para a detecção de erros grosseiros. A análise de erros grosseiros é utilizada para identificar dados experimentais ruins, que podem ter sido corrompidos por má medição ou por falha de instrumentação. Também pode ser utilizada para identificar problemas no procedimento de análise, como a definição de

variâncias muito pequenas, onde um número muito grande de erros grosseiros pode ser artificialmente encontrado.

Este teste admite que os resíduos tenham distribuição normal e que, assim, seja possível determinar um intervalo de confiança para avaliar se um determinado resíduo corresponde ou não a um erro grosseiro. Novamente, foi definido que os resultados deste teste podem assumir índices iguais a 0 se o ponto não é erro grosseiro, -1 se o ponto é um erro grosseiro negativo e +1 se o ponto é um erro grosseiro positivo. Calcula-se também o número de erros grosseiros observados no conjunto de dados analisados.

O teste é descrito pela seguinte desigualdade,

$$|e_{k,i}| \leq \sqrt{\hat{v}_{ii}^2} N_{1-\alpha/2} \quad (3.53)$$

onde $N_{1-\alpha/2}$ é o valor da distribuição normal com média zero e variância 1, definindo um intervalo com um nível de confiança de α .

Quando a desigualdade expressa pela Equação (3.53) não é satisfeita, existe um forte indício de que o ponto experimental seja um erro grosseiro. Entretanto, a própria hipótese de distribuição normal dos resíduos já contempla a possibilidade da existência de erros grosseiros; isto é, como o intervalo de confiança corresponde a uma porcentagem de $(1-\alpha)$ 100% de todos os dados, existe a probabilidade média de uma fração α dos dados não estarem no intervalo definido pela Equação (3.53). Desta forma, é preciso identificar qual é o número máximo de erros grosseiros que podem ocorrer. Para tal procedimento estatístico, emprega-se a distribuição Binomial com uma probabilidade de $\alpha \cdot 100\%$ do ponto ser um erro grosseiro (SCHWAAB e PINTO, 2007).

A distribuição Binomial é considerada uma poderosa ferramenta para fins de controle de qualidade e análise de dados. Por se tratar de uma distribuição discreta, univariada e bi paramétrica, sendo possível apenas 2 resultados, esta distribuição vem sendo usada para identificar padrões de formação entre pontos experimentais obtidos diretamente do processo.

O teste consiste em calcular a probabilidade acumulada para que o número de erros grosseiros encontrados seja superior a $(1-p)$, sendo a probabilidade Binomial calculada de acordo com a seguinte equação:

$$P_{AC}(NO; N, \alpha) = \sum_{n=0}^{NO} \frac{N!}{n! (N-n)!} (1-\alpha)^n \alpha^{N-n} \quad (3.54)$$

Observe que α é a probabilidade do ponto amostral ser considerado estatisticamente um erro grosseiro, em concordância com a Equação (3.53), e P_{AC} é a probabilidade acumulada que define a probabilidade de que até NO erros grosseiros sejam encontrados no conjunto de pontos N . Se $\frac{\alpha}{2} < P_{AC} < 1 - \frac{\alpha}{2}$ diz-se que o número de erros grosseiros encontrados é compatível com o valor esperado, de maneira que não parece ser apropriada a remoção destes pontos do conjunto de dados. Se $P_{AC} > 1 - \frac{\alpha}{2}$, pode-se dizer que o modelo apresenta desempenho incompatível como as estatísticas de processo, porque muitos erros grosseiros estarão presentes.

3.4 IMPLEMENTAÇÃO

Todas estas análises e ferramentas estatísticas empregadas para avaliação de inferências constituem um bloco de rotinas, programado em FORTRAN 90 e implementadas em um computador pessoal cujas especificações principais são: processador AMD Athlon X2 Dual Core 6000+, 2,0 GB de memória RAM, placa mãe ASUS M2A-MVP e placa de vídeo ATI Radeon HD 3600. O tempo de cálculo destes procedimentos estatísticos é pequeno e não limita a implementação em tempo real do aplicativo.

CAPÍTULO IV

RESULTADOS

“Uma resposta aproximada da questão certa é mais valiosa do que uma resposta certa de um problema aproximado.”

JOHN TUKEY

Este capítulo trata a apresentação e a interpretação dos resultados relacionados ao monitoramento estatístico de inferências. Faz-se inicialmente a apresentação dos dados operacionais. Em seguida os dados são caracterizados e é feita a análise do desempenho do modelo inferencial.

4.1 DADOS OPERACIONAIS

Nos modernos processos industriais, fortemente suportados em tecnologias de informação e automação industrial, um grande número de variáveis de processo pode ser simultaneamente medidas, armazenadas e facilmente acessadas, facilitando o uso de eficientes estratégias de operação em tempo real. Sabe-se que a disponibilidade de

informações é um requisito essencial à operação satisfatória do processo, sendo o uso dessas tecnologias um diferencial de competitividade para qualquer setor industrial (FELDMAN, 2007). Todavia, os benefícios e as vantagens deste moderno cenário industrial estão diretamente vinculados à utilização da informação. Portanto, além da informação disponível, deve haver o bom uso desta informação (AMY, 1992; BOYER, 1999; FELDMAN, 2007).

Os estudos envolvendo unidades industriais são convenientes, não só pela disponibilidade de dados operacionais, mas pela relevância dos resultados obtidos; isto porque, neste caso a validação pode ter uma aplicabilidade imediata e efetiva. Portanto, pode-se dizer que os sistemas reais consolidam, em última escala, a validação de um estudo.

Neste estudo, a estrutura para monitoramento e avaliação de inferências foi aplicada ao processo de produção de diesel, com o objetivo de validar a abordagem proposta. Assim, foram empregados dados operacionais de uma indústria brasileira de refino do período de janeiro do ano 2008. Cada ponto de amostragem foi obtido diretamente do SDCD (Sistema Digital de Controle Distribuído), com intervalo de amostragem de 5 minutos. O conjunto de dados operacionais selecionados está em torno de 4032 medidas, que equivalem aproximadamente à metade de um mês de produção. Todavia, como os resultados das análises de laboratório para T85% são obtidas a cada 12 horas, o conjunto de pontos analíticos é composto por aproximadamente 31 medidas.

É importante ressaltar que os dados operacionais analisados no decorrer deste estudo foram coletados de um período de operação da planta considerado estacionário.

4.2 ANÁLISE DOS DADOS

A etapa inicial da análise consiste na seleção e tratamento dos dados operacionais. A grande disponibilidade de dados em sistemas industriais não é

necessariamente sinônimo de informações fidedignas do processo. Portanto, a seleção e o tratamento dos dados são imprescindíveis.

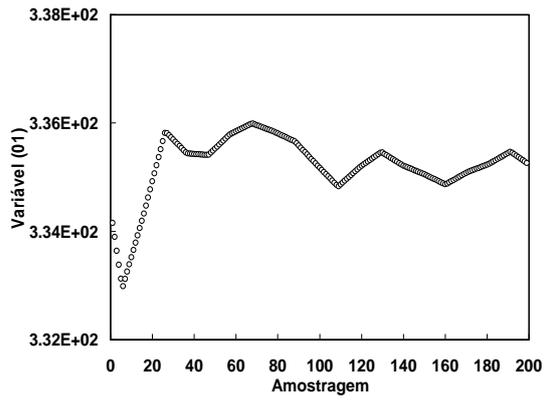
A Tabela (4.1) lista propriedades empregadas como variáveis de entrada do modelo inferencial. Como explicado posteriormente, as variâncias foram calculadas como as variâncias obtidas do espectro de variâncias para janelas de amostragem de menor tamanho ($TJ = 2$).

Tabela 4.2: Variáveis de entrada do modelo inferencial.

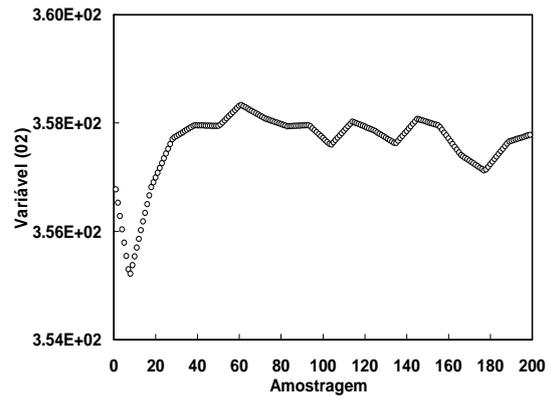
Propriedade	Código	Variância
Temperatura de topo do diesel produto (°C)	Variável (01)	2,07E-03
Temperatura de fundo da torre (°C)	Variável (02)	2,14E-03
Vazão de topo do diesel produto (m ³ /d)	Variável (03)	9,11E+01
Vazão de fundo da torre (m ³ /d)	Variável (04)	4,83E+02
Vazão de nafta pesada (m ³ /d)	Variável (05)	7,41E+01
Vazão de querosene (m ³ /d)	Variável (06)	1,32E+03
Vazão de diesel (m ³ /d)	Variável (07)	1,68E+02
Vapor de retificação (ton/h)	Variável (08)	8,11E-04
Pressão da zona de flash (kgf/cm ² g)	Variável (09)	8,61E-07
T85% (°C)	Variável (10)	1,36E-01

Estas propriedades foram selecionadas por ZANIN (1993) como as variáveis que mais afetam a qualidade do diesel.

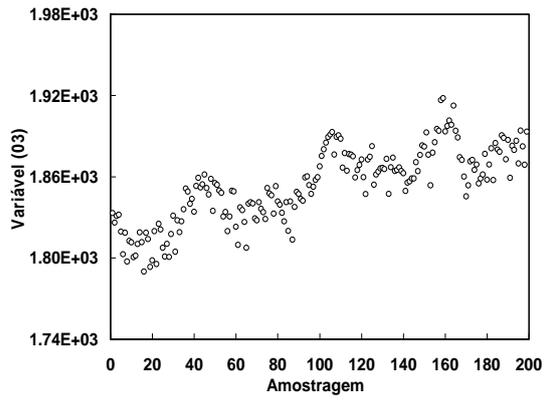
A Figura (4.1) ilustra o comportamento das variáveis operacionais, empregadas como variáveis independentes do modelo inferencial. Analisando o comportamento das variáveis durante o período considerado, nota-se que os gráficos (a), (b), (e), (g), (h) e (i) da Figura (4.1) apresentam trajetórias definidas (pouco aleatórias) durante a operação do processo. Observando estes gráficos, é possível também notar que as variáveis que apresentam maior aleatoriedade de comportamento são as variáveis representadas pelos gráficos (d) e (f). Essa informação é de grande relevância, pois indica a presença de dinâmica nos dados embora o modelo inferencial seja estacionário.



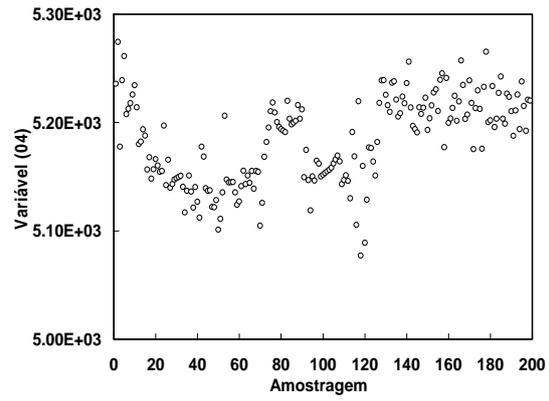
(a)



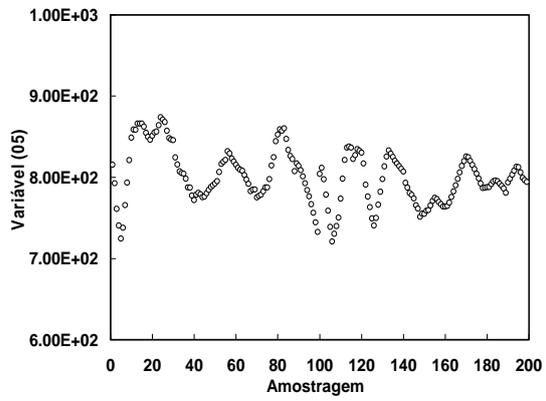
(b)



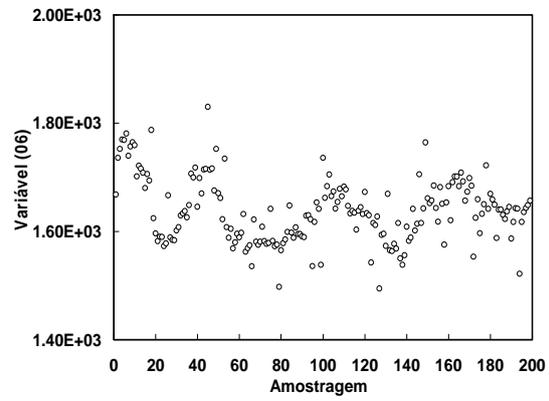
(c)



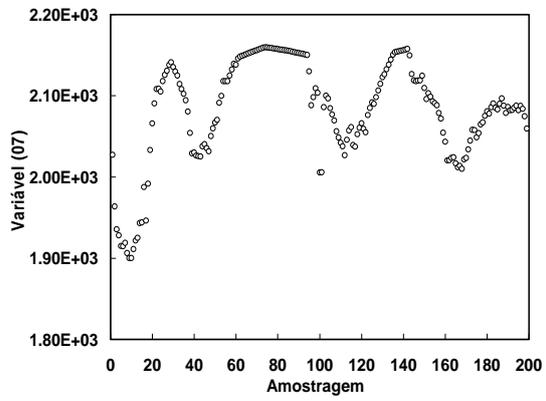
(d)



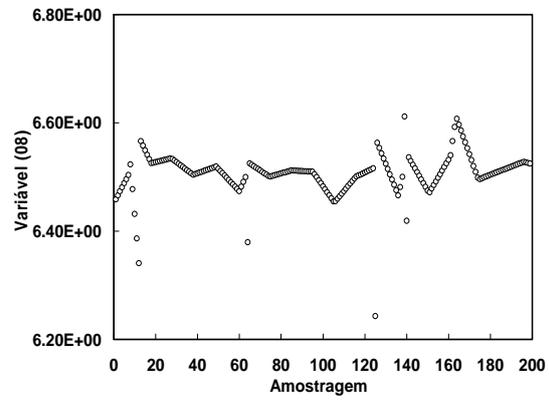
(e)



(f)



(g)



(h)

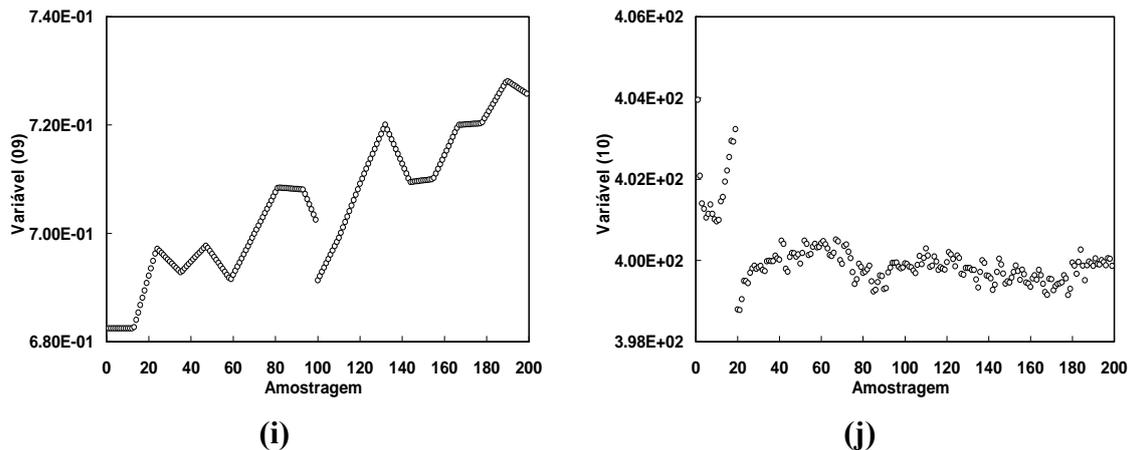


Figura 4.1: Comportamento das variáveis operacionais.

Uma outra observação possível é o fato de que as variáveis são medidas por instrumentos de diferentes precisões, acarretando em níveis de ruído diferentes nas diferentes variáveis. É importante que esse fato seja considerado, já que o desempenho dos diferentes instrumentos não é uniforme.

Os dados operacionais apresentados na Figura (4.1) foram filtrados empregando um filtro de ordem zero que consiste basicamente em substituir toda medida considerada inválida (contraditória ou nula) pela medida anterior.

Realizada a seleção e o pré-tratamento dos dados, passou-se caracterização estatística básica dos dados. Foram então realizados os cálculos de médias, de desvio-padrão, da variância e da correlação dos dados obtidos. Nestas análises, o objetivo principal é definir os valores de referências para o processo. As Tabelas (4.2) à (4.4), respectivamente, apresentam o valor médio, o desvio-padrão e a matriz de correlação de cada variável operacional analisada.

Tabela 4.2: Valor médio das variáveis operacionais.

Média									
Var(01)	Var(02)	Var(03)	Var(04)	Var(05)	Var(06)	Var(07)	Var(08)	Var(09)	Var(10)
335,22	357,64	1853,50	5183,45	800,68	1638,35	2082,25	6,51	0,71	400,01

Tabela 4.3: Desvios-padrão das variáveis operacionais.

Desvios-padrão									
Var(01)	Var(02)	Var(03)	Var(04)	Var(05)	Var(06)	Var(07)	Var(08)	Var(09)	Var(10)
0,57	0,60	27,20	39,65	32,82	58,81	63,43	0,04	0,01	0,73

Tabela 4.4: Matriz de correlação das variáveis operacionais.

Matriz de Correlação									
1	<u>0,865</u>	0,071	-0,327	0,124	-0,626	<u>0,852</u>	0,138	0,315	-0,491
<u>0,865</u>	1	0,345	-0,333	-0,156	-0,505	<u>0,78</u>	0,052	0,31	-0,502
0,071	0,345	1	0,256	-0,51	-0,002	0,065	0,041	0,659	-0,432
-0,327	-0,333	0,256	1	-0,097	0,077	-0,114	0,106	0,47	-0,089
0,124	-0,156	-0,51	-0,097	1	-0,205	0,04	0,13	-0,151	0,137
-0,626	-0,505	-0,002	0,077	-0,205	1	-0,733	-0,055	-0,314	0,369
<u>0,852</u>	<u>0,78</u>	0,065	-0,114	0,04	-0,733	1	0,021	0,32	-0,507
0,138	0,052	0,041	0,106	0,13	-0,055	0,021	1	0,259	-0,15
0,315	0,31	0,659	0,47	-0,151	-0,314	0,32	0,259	1	-0,557
-0,491	-0,502	-0,432	-0,089	0,137	0,369	-0,507	-0,15	-0,557	1

Os resultados obtidos para as correlações indicam a existência de uma associação significativa entre algumas variáveis. Quando este resultado é positivo, como por exemplo, no caso da associação entre a variável (01) e a variável (02) em que este valor é 0,865, indica que estas variáveis se influenciam positivamente; ou seja, são diretamente proporcionais, de tal forma que à medida que uma variável “cresce” a outra também “cresce”. Porém, quando este resultado é negativo, como no caso da associação entre a variável (09) e a variável (10) em que este valor é -0,557, indica que estas variáveis se influenciam negativamente; ou seja, são inversamente proporcionais, de tal forma que à medida que uma variável “cresce” a outra “decrece”.

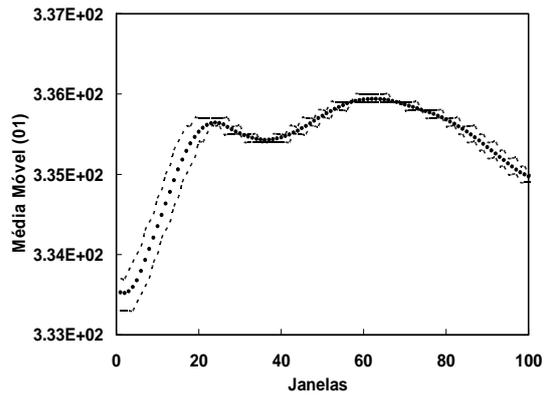
Neste caso, observa-se a presença de uma correlação significativa entre algumas variáveis, o que pode levar à interferências na qualidade da inferência durante a operação do processo, por perturbações aplicadas pelo sistema de controle e/ou pela própria estrutura do modelo que não prevê este tipo de relacionamento entre as variáveis. Observa-se em particular que os pares formados pelas variáveis (01), (02) e

(06), (07) apresentam forte correlação entre si. Do ponto de vista do processo, é justificável que as temperaturas de retiradas, tanto de topo quanto de fundo da torre estejam fortemente correlacionadas; o mesmo acontece para as vazões de retirada de diesel e querosene que são produtos que possuem propriedades físicas (como por exemplo, densidade e ponto de ebulição) muito similares entre si.

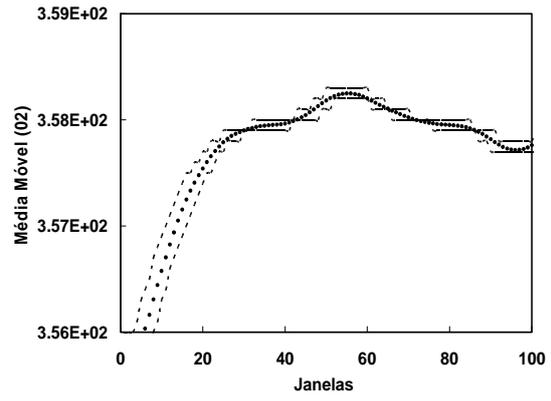
Embora a caracterização básica seja importante para o conhecimento e a análise operacional, sendo seus resultados muitas vezes utilizados como parâmetros de calibração de instrumentos e de sintonia de controladores, estes apresentam interpretação limitada no aspecto do monitoramento da inferência. Portanto, faz-se necessário o emprego de ferramentas capazes de investigar a presença de dinâmica do comportamento do processo.

Neste sentido, são empregadas as ferramentas estatísticas para caracterização dinâmica dos dados operacionais. O objetivo é a identificação de períodos de operação estacionária ou transiente. Para tal são realizados cálculos de médias e de variâncias móveis e seus respectivos intervalos de confiança e índices de conformidade. Também são calculados os espectros de variância e os espectros de correlação das variáveis operacionais.

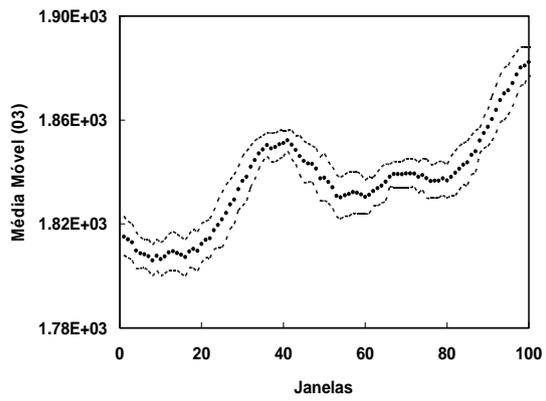
As Figuras (4.2) e (4.3) apresentam as médias e as variâncias móveis para as variáveis operacionais estudadas, e seus respectivos intervalos de confiança. Para a aplicação destas ferramentas estatísticas utilizaram-se janelas móveis com tamanho de 13 amostragens, representando 1 hora de operação do processo. Com o suporte da visualização do comportamento dos dados operacionais (Figura 4.1), pode-se observar claramente o efeito suavizador do cálculo das médias móveis (Figura 4.2), removendo flutuações das variáveis estudadas.



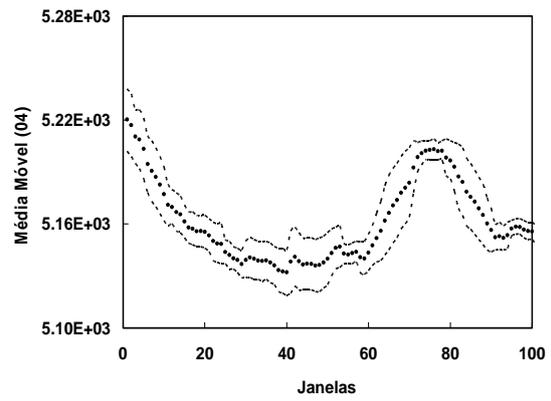
(a)



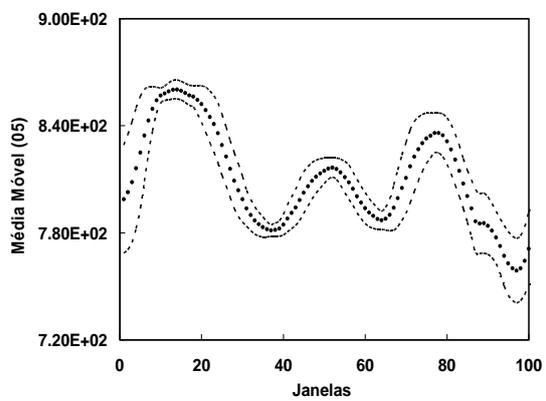
(b)



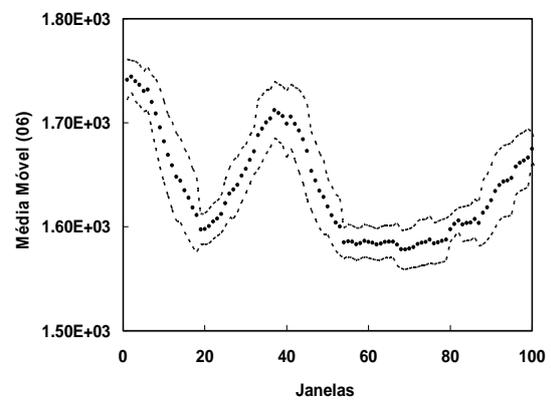
(c)



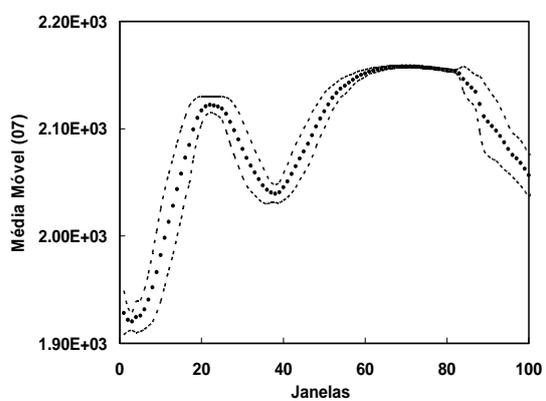
(d)



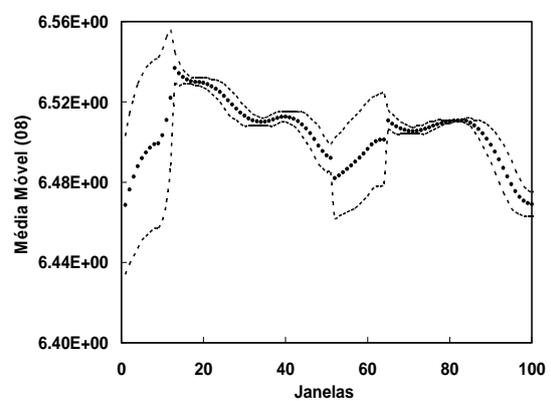
(e)



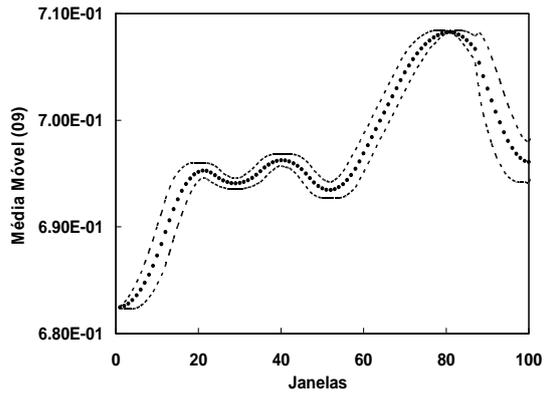
(f)



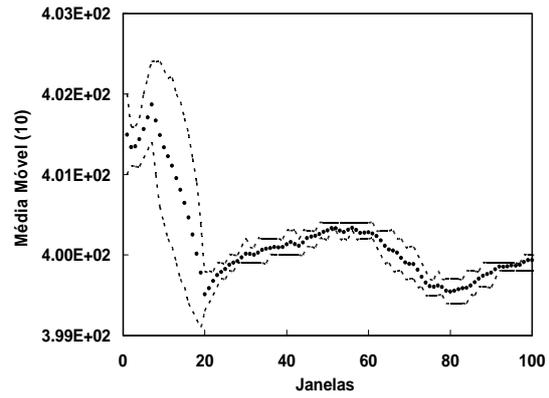
(g)



(h)

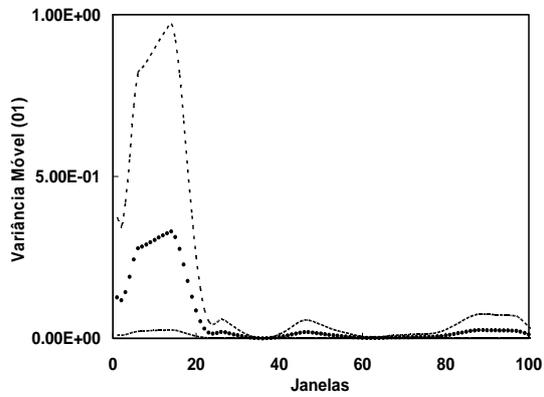


(i)

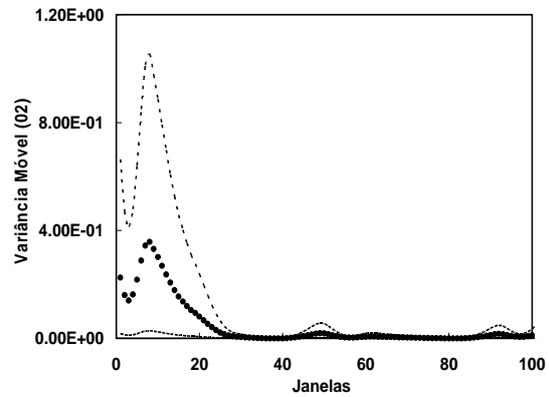


(j)

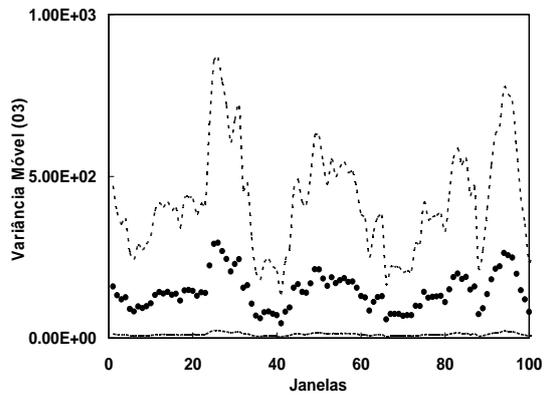
Figura 4.2: Intervalo de confiança (---) e média móvel (●) das variáveis operacionais.



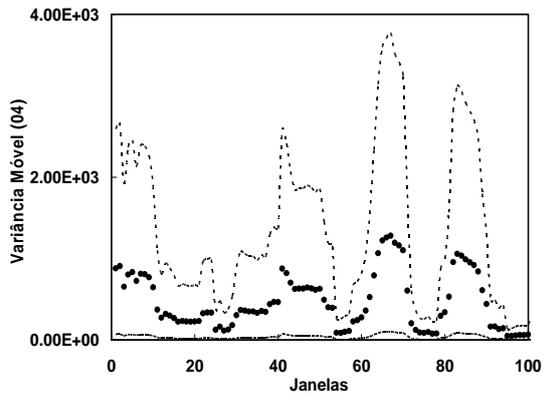
(a)



(b)



(c)



(d)

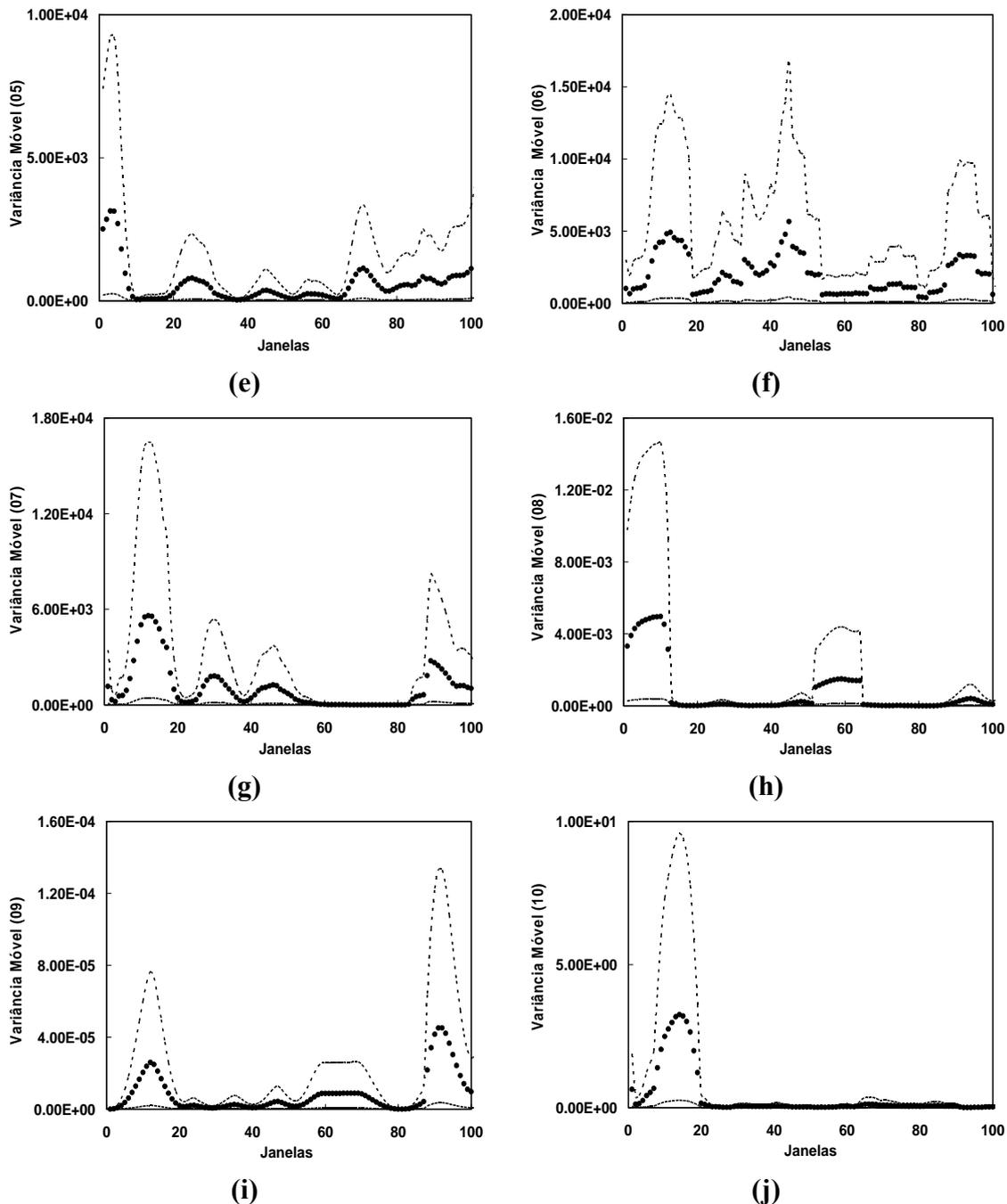


Figura 4.3: Intervalo de confiança (---) e variância móvel (●) das variáveis operacionais.

Na Figura (4.3) gráficos (c), (d), (f) e (g), as variáveis (3), (4), (6) e (7) apresentam largos intervalos de confiança para variância. A justificativa para tal pode estar no comportamento mais difuso (aleatório) destas variáveis, como apresentado na Figura (4.1) gráficos (c), (d), (f) e (g). Como os intervalos de confiança expressam uma relação de significância entre a grandeza estatística e o seu valor real, quanto o maior o

ruído ou aleatoriedade da medida, maior é possibilidade da variável assumir outros valores, ampliando a região em que possa ser encontrada.

A Figura (4.4) apresenta respectivamente, a análise de conformidade da média e da variância móvel das variáveis operacionais. O índice de conformidade foi definido como a fração de vezes em que a comparação entre duas médias ou variâncias resulta em medidas conformes, quando comparações são feitas entre medidas obtidas com uma janela de mesmo tamanho TJ .

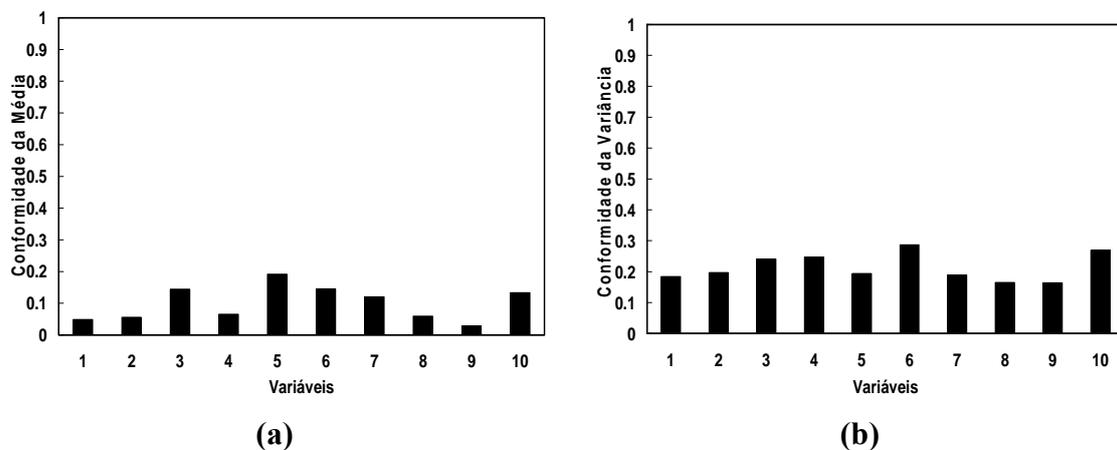


Figura 4.4: Conformidade da média e da variância para as variáveis operacionais.

Nesse caso, tanto a conformidade da média quanto a conformidade da variância móvel foram calculados com tamanho de janela móvel equivalente 1 hora de operação do processo. A conformidade foi expressa em forma de índice, que representa o resultado desta análise para cada variável ao longo da operação. Observa-se que, de um modo geral, os índices de conformidade da variância (Figura 4.4, gráfico (b)) são maiores que os índices obtidos para a conformidade da média (Figura 4.4, gráfico (a)). Provavelmente a justificativa está no sistema de controle da planta, que impõe limites rígidos para as oscilações do processo. Valores para conformidade da média móveis muito diferentes da variância móvel também são possíveis, já que os cálculos destas grandezas são independentes.

Os índices de conformidade da média para todas variáveis operacionais (exceto a Variável (06)) estão abaixo de 0,2, extremamente baixos. Estes índices demonstram que as médias são em geral diferentes, caracterizando o comportamento dinâmico do

processo. Acredita-se que há uma alteração constante do valor desejado (*set point*) durante a operação da planta, e que esta alteração pode estar sendo causada pelo *bias* acrescentado ao modelo inferencial a cada nova inferência tomada na planta. A variabilidade ao longo do processo parece apresentar amplitude e frequência de alteração relativamente alta, com índices de conformidade inferiores a 0,3. Isto significa que o nível de ruído observado também flutua no tempo, o que pode caracterizar o nível variável de ruído da operação ou o mau funcionamento dos sistemas de controle.

A fim de avaliar a influência do tamanho da janela móvel *TJ* sobre a análise de conformidade, foram realizadas comparações dos percentuais de conformidade da média e da variância móvel para diferentes tamanhos de janela móvel. As Tabelas (4.5) e (4.6) apresentam respectivamente os percentuais de conformidade da média (PCM) e da variância (PCV), calculados para diferentes tamanhos de janela móvel (*TJ*). O percentual de conformidade da média representa o resultado desta análise para cada variável ao longo da operação em percentual.

Tabela 4.5: Percentual de conformidade da média para diferentes TJ.

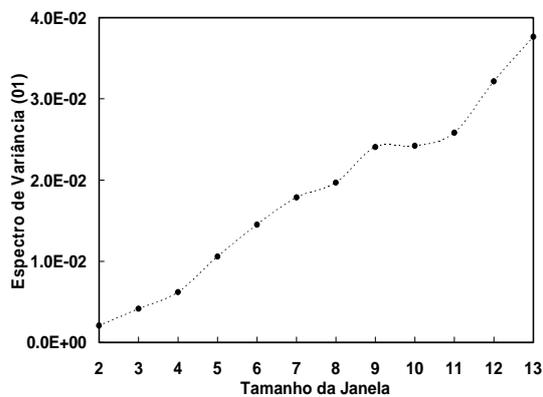
Percentual de Conformidade da Média – PCM (%)										
TJ (h)	Var(01)	Var(02)	Var(03)	Var(04)	Var(05)	Var(06)	Var(07)	Var(08)	Var(09)	Var(10)
0,5	4,16	4,88	17,12	8,24	19,61	17,63	12,28	6,15	2,69	16,47
1,0	4,88	5,59	14,45	6,55	19,16	14,55	12,05	5,89	2,85	13,35
1,5	4,68	6,11	13,28	6,64	21,73	13,21	11,34	6,34	3,04	12,46
2,0	5,53	6,01	13,39	6,17	18,76	12,88	12,03	6,93	3,18	11,65
2,5	5,55	6,17	13,06	6,30	19,48	12,62	12,08	6,69	2,81	11,16

Tabela 4.6: Percentual de conformidade da variância para diferentes TJ.

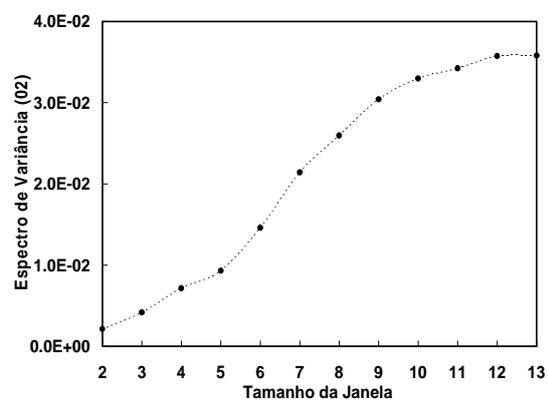
Percentual de Conformidade da Variância – PCV (%)										
TJ (h)	Var(01)	Var(02)	Var(03)	Var(04)	Var(05)	Var(06)	Var(07)	Var(08)	Var(09)	Var(10)
0,5	24,68	25,88	30,16	30,96	26,14	35,99	27,61	20,17	22,58	35,05
1,0	18,44	19,70	24,10	24,74	19,40	28,70	18,91	16,45	16,38	26,90
1,5	16,32	16,72	20,81	20,88	17,17	24,12	15,20	14,05	13,50	21,57
2,0	16,05	16,87	17,38	16,69	15,68	20,45	13,46	15,19	11,32	18,47
2,5	13,86	15,25	14,67	14,06	14,62	18,82	12,86	14,91	10,89	16,18

Enquanto o percentual de conformidade da média tem comportamento diferenciado para cada variável em relação ao tamanho da janela, o percentual de conformidade da variância decresce. A redução do PCV significa que, com o passar do tempo, flutuações dinâmicas estão sendo incorporadas, modificando o comportamento desta grandeza. O aumento do PCM, por sua vez, é um mero reflexo do efeito de filtragem da operação. De qualquer forma parece bastante claro que há comportamento dinâmico nos dados de processo utilizados e que esta conclusão independe do tamanho de janelas considerado.

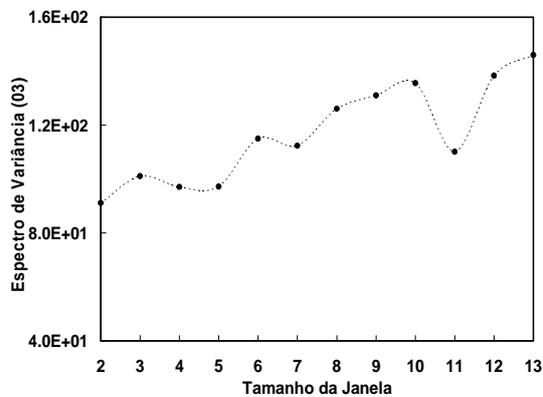
Para caracterização da variabilidade do processo, são usados também os espectros de variância das variáveis operacionais. Como já foi dito, esta ferramenta é muito útil para caracterizar uniformidade ou não dos dados ao longo do tempo. A Figura (4.5) apresenta os espectros de variância das variáveis operacionais, para janelas móveis de 10 minutos (2 pontos amostrais) até 1 hora (13 pontos amostrais) de operação.



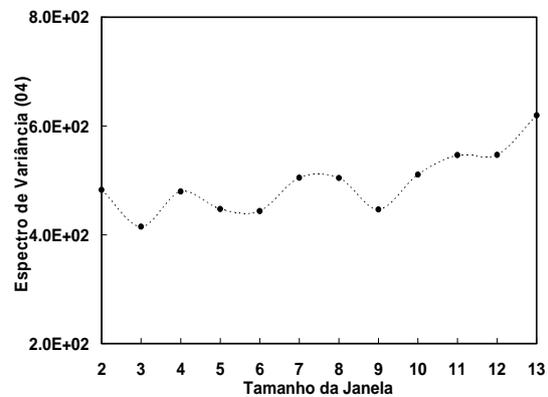
(a)



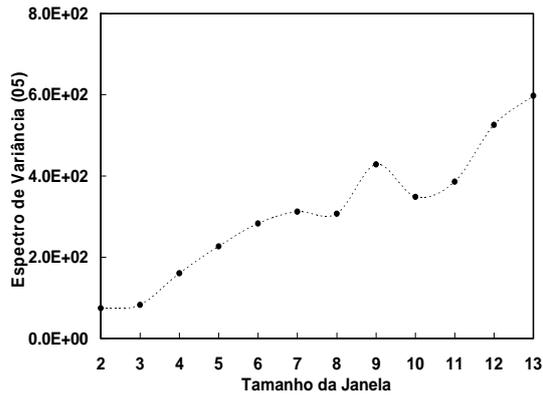
(b)



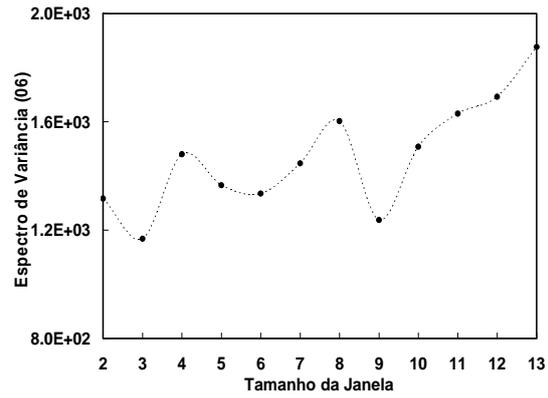
(c)



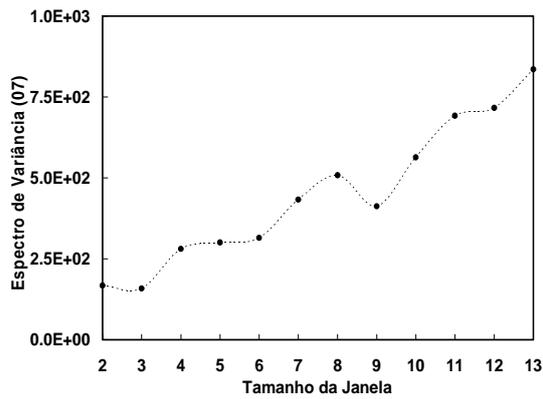
(d)



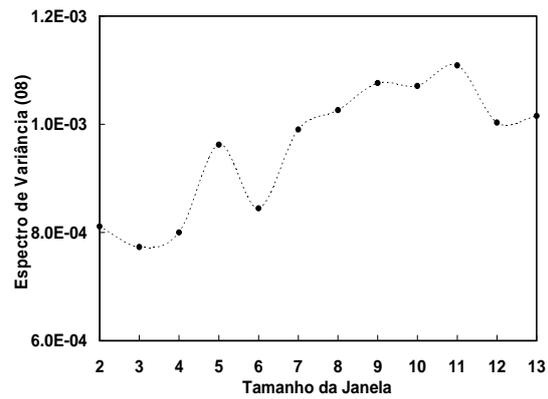
(e)



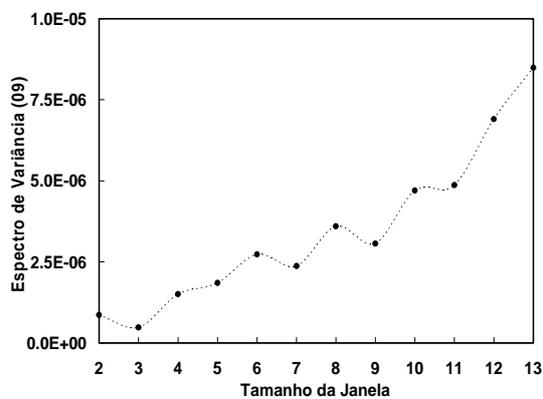
(f)



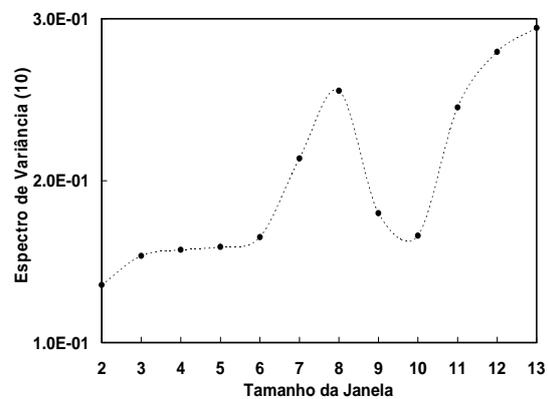
(g)



(h)



(i)



(j)

Figura 4.5: Espectro de variância das variáveis operacionais.

Observa-se que nos espectros de variância apresentados na Figura (4.5), as variâncias tendem a aumentar com o tamanho da janela móvel, o que indica que dados estatisticamente diferentes são incorporados ao longo do processo, caracterizando novamente a existência de dinâmica.

Para a Variável (04), Figura (4.5) gráfico (d), a variabilidade instrumental está bem próxima da variabilidade incorporada ao longo do processo representado, o que

demonstra comportamento estacionário típico. Possivelmente este fato está associado ao comportamento estocástico apresentado por esta variável, Figura (4.1) gráfico (d), condição que pode ser reafirmada pelos largos intervalos de confiança apresentados por suas média e variância móveis, gráficos (d) das Figuras (4.2) e (4.3). Esta também é a variável que apresenta a menor associação entre suas medidas ao longo da operação do processo.

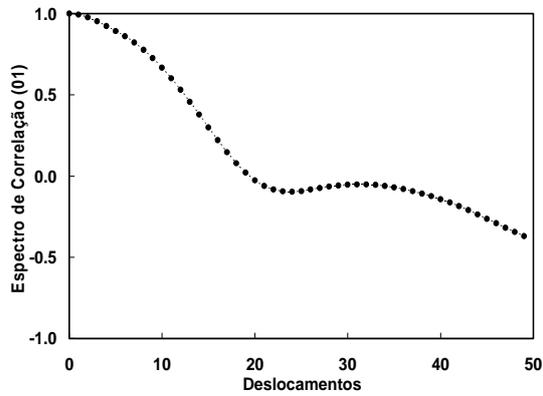
O comportamento que melhor caracteriza o espectro de variância ideal é a curva apresentada pela Variável (02), Figura (4.5) no gráfico (b). Nesta curva, cujo ponto inicial representa a variância instrumental, notam-se as flutuações do processo sendo incorporadas (crescentemente), à medida que se aumenta o tamanho da janela, até atingir a variabilidade final da variável de processo. Para as demais variáveis, pode-se dizer que há dinâmicas que se manifestam em tempo superior a 13 amostragens.

A oscilação presente nos espectros de variância das Variáveis (03), (06), (08) e (10) representados pela Figura (4.5) gráficos (c) (f) (h) (j) respectivamente, podem informar sobre períodos de estabelecimento de condições de estacionariedade, mudança estatística da variável, perturbações acrescentadas ao longo da operação do processo, sendo que também podem fundamentar questionamentos sobre a qualidade dos dados (a presença de erros grosseiros) e a possibilidade de falhas. A oscilação das variâncias está muito associada à existência de padrões dinâmicos de variação. Por exemplo, para a Variável (08), o máximo está intimamente relacionado às oscilações iniciais das medidas.

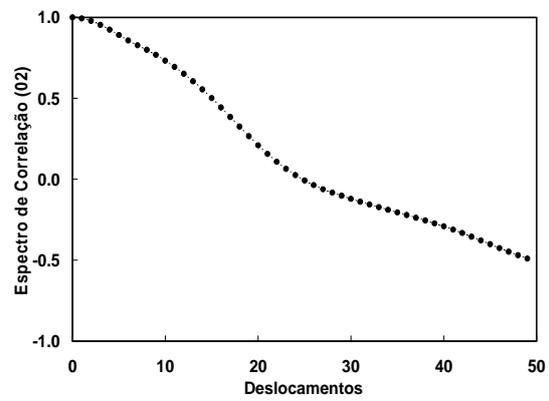
As variâncias mínimas, obtidas para janelas de tamanho TJ igual a 2, são mostradas na Tabela (4.1). Essas são as variâncias naturais de flutuação, consideradas para fins de caracterização do modelo.

Finalmente, a caracterização dinâmica das variáveis operacionais apresenta os espectros de correlação. Neste caso será empregado o espectro de auto-correlação, pois esta ferramenta estatística possibilita a identificação do tempo normalmente requerido para que seja atingido um estado estacionário; ou seja, o tempo requerido para que as mudanças tenham caráter aleatório e não caráter dinâmico.

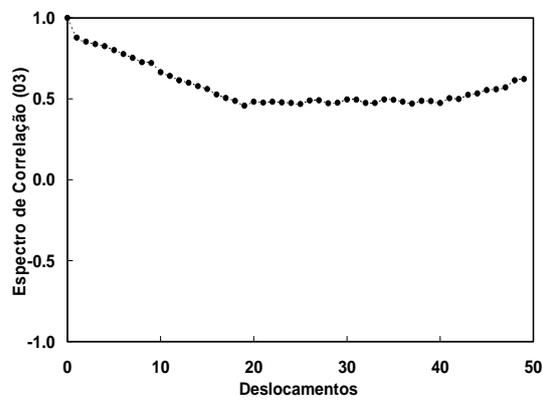
A Figura (4.6) apresenta os espectros de auto-correlação das variáveis operacionais, para aproximadamente 4 horas de operação do processo.



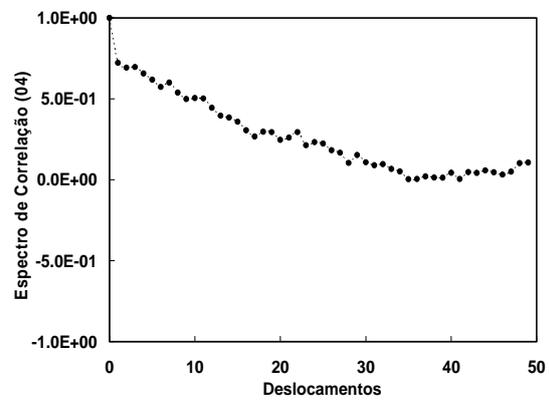
(a)



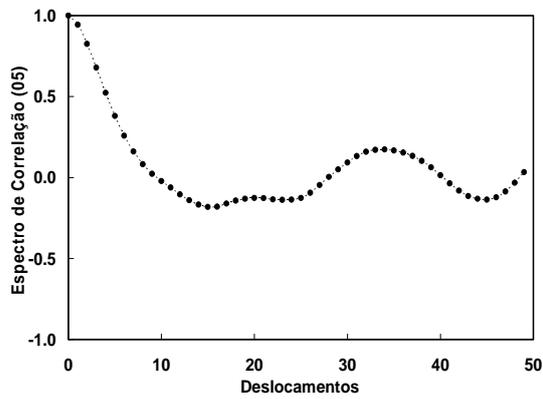
(b)



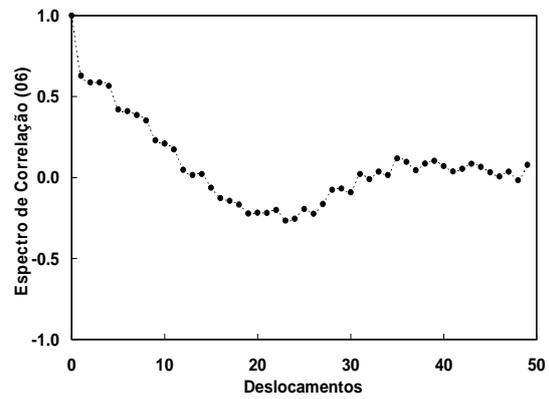
(c)



(d)



(e)



(f)

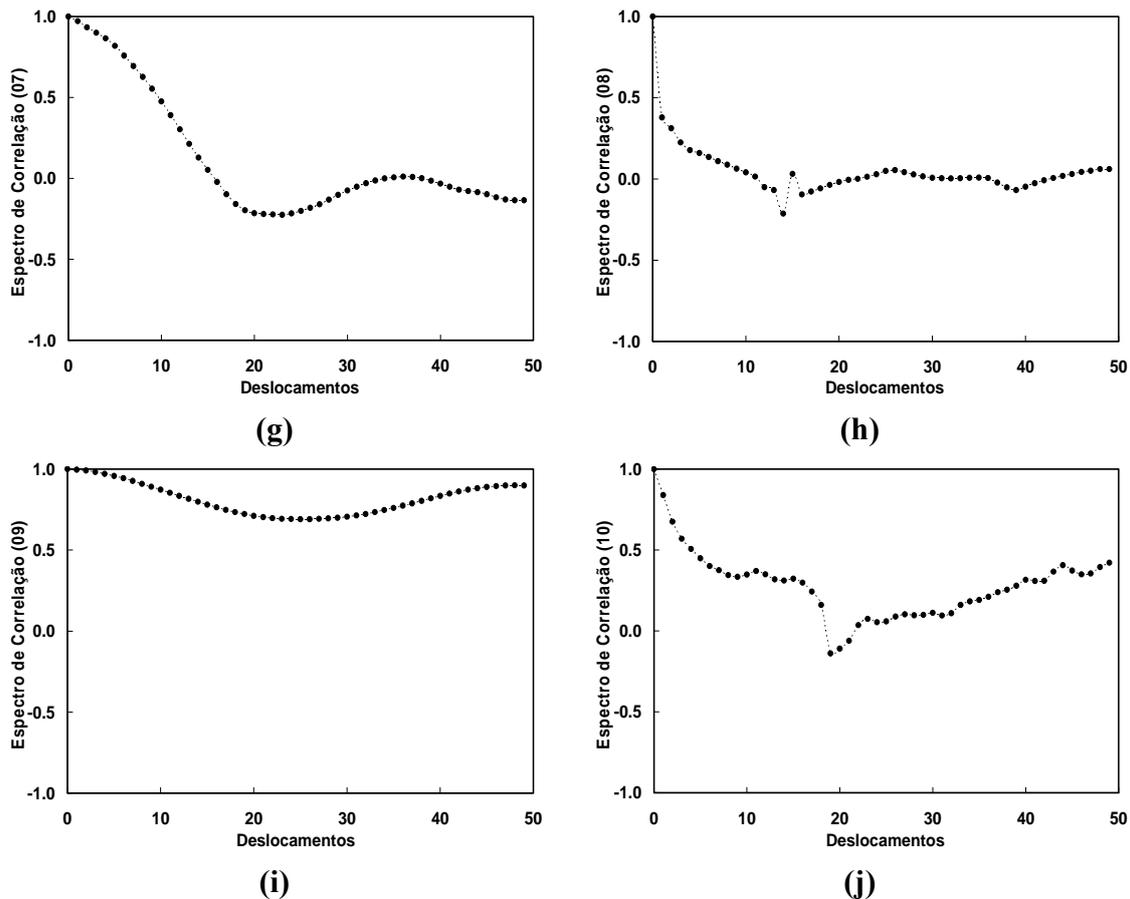


Figura 4.6: Espectros de auto-correlação das variáveis operacionais.

De acordo com os resultados, com exceção de algumas correlações, a maioria delas (60% das variáveis) fica próxima a 0,5 após 10 deslocamentos e reduzem para 0,0 após 20 deslocamentos. Este resultado está coerente com a operação real do processo, já que 10 deslocamentos correspondem a 50 minutos de operação e 20 deslocamentos correspondem a 10 dias de operação, e é esperado que não haja correlação significativa entre os sinais após este intervalo de tempo, pela natureza dinâmica do processo.

O comportamento oscilatório de algumas variáveis (05, 06, 07 e 10), pode significar mudanças ao longo da operação do processo.

Observa-se também que as correlações das variáveis (05), (06), (08) e (10) reduzem rapidamente para o valor de 0,5, após 5 deslocamentos, e equivale a uma constante de tempo de operação do processo em torno de 2,5 dias para estas variáveis, indicando que os dados amostrados experimentam dinâmica embora tenham sido coletados em um período amostral da operação considerado estacionário.

Os espectros de correlação das variáveis (03) e (09) mantêm-se constantes acima de 50%, sendo que no caso desta última o espectro não atinge em ponto algum de deslocamento um valor inferior a 0,5 de correlação. Também usualmente chamadas como variáveis de “lembrança”, estas variáveis apresentam o comportamento futuro fortemente dependente do passado.

Uma informação importante a partir dos espectros de correlação é a identificação da presença de dinâmica nos dados amostrados, com constante de tempo do processo bastante superiores a 1 dia.

A análise dos dados é importante para um maior entendimento e uma melhor descrição da operação do processo, bem como para a elaboração mais seguras de estatísticas do processo acerca das variáveis operacionais. Uma análise importante relacionada a este tema, não abordada diretamente neste estudo, é a detecção de erros grosseiros. Para tratar este problema, ferramentas estatísticas foram empregadas com base no modelo inferencial.

4.3 ANÁLISE DO MODELO INFERENCIAL

O modelo inferencial é um modelo que por meio de variáveis medidas prediz variáveis de interesses que não podem ser medidas (na planta e/ou em tempo real). O acompanhamento de modelos geralmente é realizado por meio dos resíduos. Os resíduos consistem nas diferenças entre os valores medidos (neste caso, as análises de laboratório) e os valores obtidos através do modelo (neste caso, o modelo inferencial). Portanto, os resíduos são considerados os resultados filtrados pelo modelo. A análise estatística deste elemento é de fundamental importância, já que este demonstra como o modelo se comporta em relação aos dados. Assim, o resíduo é o vínculo entre o modelo inferencial e os dados operacionais!

A mesma seqüência realizada para a análise e caracterizações estatística dos dados é feita para a inferência fornecida pelo modelo inferencial. Portanto, nesta parte

do trabalho serão apresentados resultados concernentes às etapas de caracterização básica e de caracterização dinâmica, com o acréscimo da etapa de caracterização da predição. Esta última é dedicada à análise da inferência exclusivamente por meio de modelos inferenciais.

Os dados operacionais empregados pelo modelo inferencial são os mesmos reportados na Seção 4.2 deste capítulo.

A caracterização básica da inferência é apresentada pela Tabela (4.7); a inferência e a sua variância são respectivamente apresentadas pelas Figuras (4.7) e (4.8).

Tabela 4.7: Média e desvio-padrão para a inferência.

Inferência	
Média	Desvio-padrão
400,023	0,742

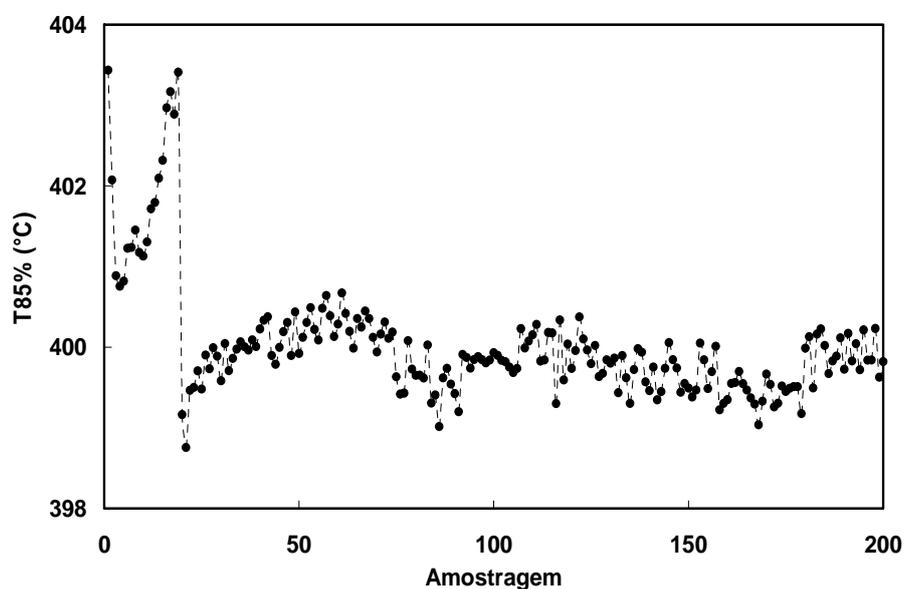


Figura 4.7: Inferência da T85% (°C).

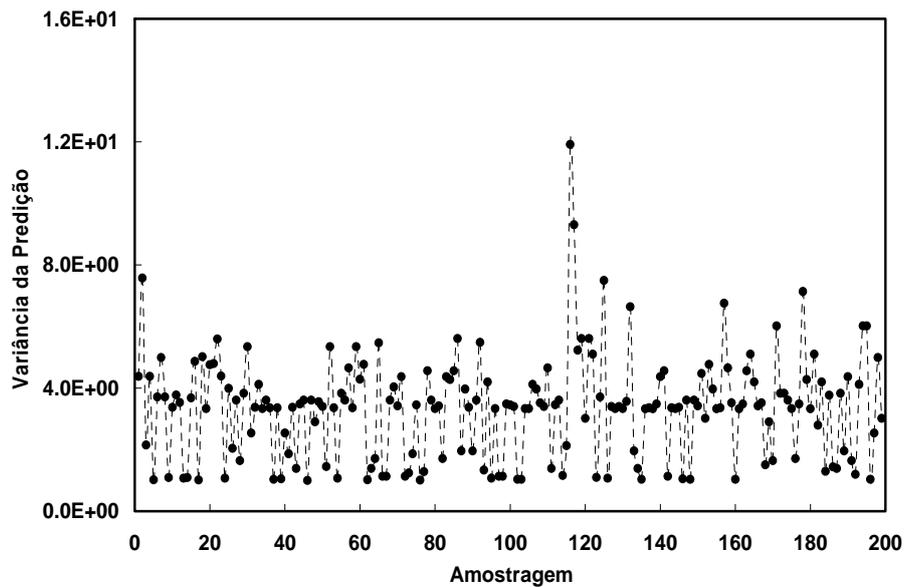


Figura 4.8: Variância da predição da T85% (°C).

A inferência da T85% oscila entre 398 a 404 °C, mantendo-se a maior parte da amostragem em torno de 400 °C. Do ponto de vista prático, deseja-se que a T85% se mantenha em torno de 398 °C e que sejam inferiores a 400 °C. Neste gráfico pode-se perceber que o comportamento inicial (até o ponto 19 da amostragem) da inferência difere significativamente do comportamento a longo do tempo amostral (os outros 179 pontos consecutivos), o que pode indicar um comportamento dinâmico da T85% e também levantar hipóteses sobre a presença de erros grosseiros nas medidas empregadas pelo modelo inferencial. De qualquer forma é importante observar que o comportamento da inferência não é uniforme, quer seja por mudanças no processo, quer seja por desempenho do modelo inferidor.

Os valores obtidos para a variância da predição da inferência são relativamente altos, conforme a Figura (4.8). A variância da predição é uma ferramenta importante para avaliação do erro de predição considerando todos os elementos envolvidos neste processo, pois é composta pela variância das variáveis (independentes e dependentes) e dos parâmetros, e ainda a variância experimental. Portanto, dentre as justificativas possíveis para uma variância de perdições altas estão: parâmetros correlacionados, altos valores de correlação experimental das variáveis independentes e/ou das variáveis dependentes, e erros grosseiros presentes nos dados operacionais.

Observa-se que há necessidade de se caracterizar a fonte de erro que eleva a variância da predição. Para tal, ao longo deste estudo serão reestimados os parâmetros do modelo inferencial, analisada a influência das variáveis (dependentes e independentes) sobre a inferência e realizado um procedimento de detecção de erros grosseiros a partir do modelo.

Seguindo a mesma sistemática adotada para avaliação dos dados, são calculados os resíduos e sua caracterização estatística (cálculo de média e de desvio-padrão).

A Figura (4.9) apresenta os resíduos da variável de saída, para 25 dias de operação do processo (amostragem de 12 em 12 horas).

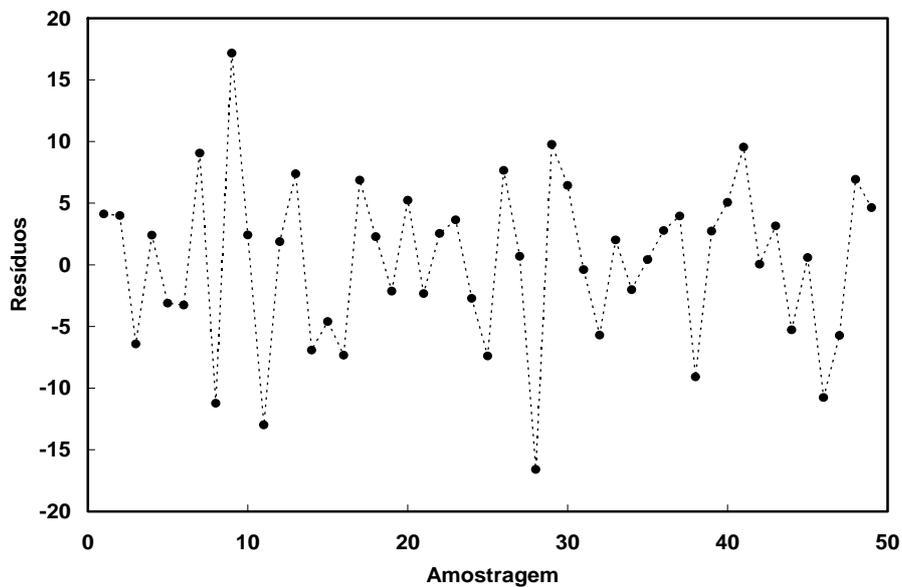


Figura 4.9: Comportamento dos resíduos da T85%.

Observa-se que os resíduos possuem valores altos, estão compreendidos em um intervalo de 20 a -20. Os altos valores obtidos para os resíduos demonstram que as inferências não são muito próximas do valor medido em laboratório, e indicam que as predições não são satisfatórias.

A Tabela (4.8) apresenta os resultados da média e do desvio-padrão para os resíduos.

Tabela 4.8: Média e desvio-padrão calculados para os resíduos.

Resíduos	
Média	Desvio-padrão
0,000	24,255

Embora a flutuação seja considerável, o que significa que as previsões oscilam bastante em torno do valor médio, os resíduos apresentam uma média cujo resultado é favorável ao modelo. Resíduos com média zero significam que o modelo não necessita ajuste (*bias*); ou seja, suas previsões não estão tendenciosas. Este resultado questiona a validade do *bias*.

Assim como para os dados foram realizados cálculos de médias e de variâncias móveis, de espectros de correlação e de variância, o mesmo também foi feito para os resíduos. Porém neste caso o resultado destas análises apresenta a informação sobre o ajuste do modelo inferencial aos dados.

As Figuras (4.10) e (4.11) apresentam média e variância móveis dos resíduos, calculadas com janelas móveis de 1 hora, para aproximadamente 28 horas de operação do processo.

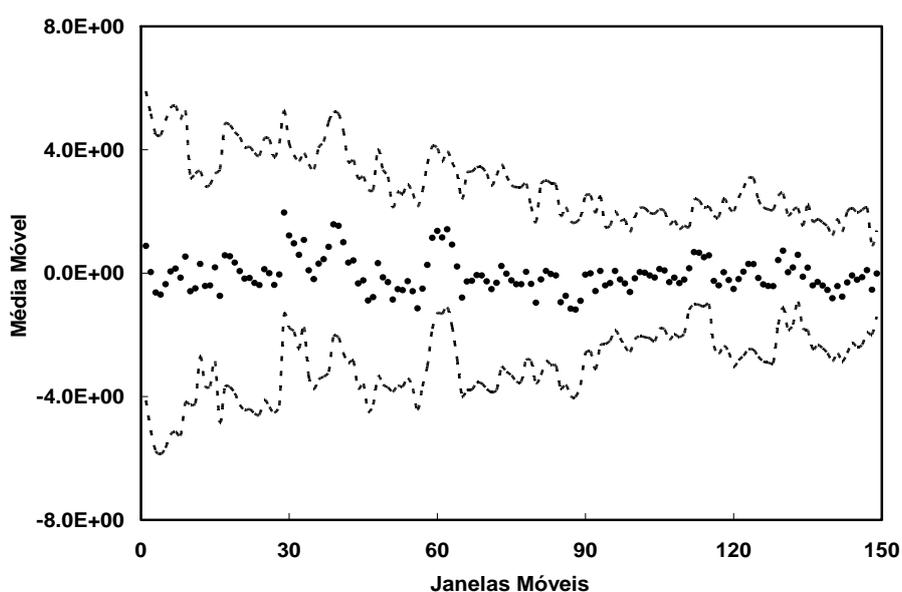


Figura 4.10: Intervalo de confiança (---) e média móvel (●) dos resíduos.

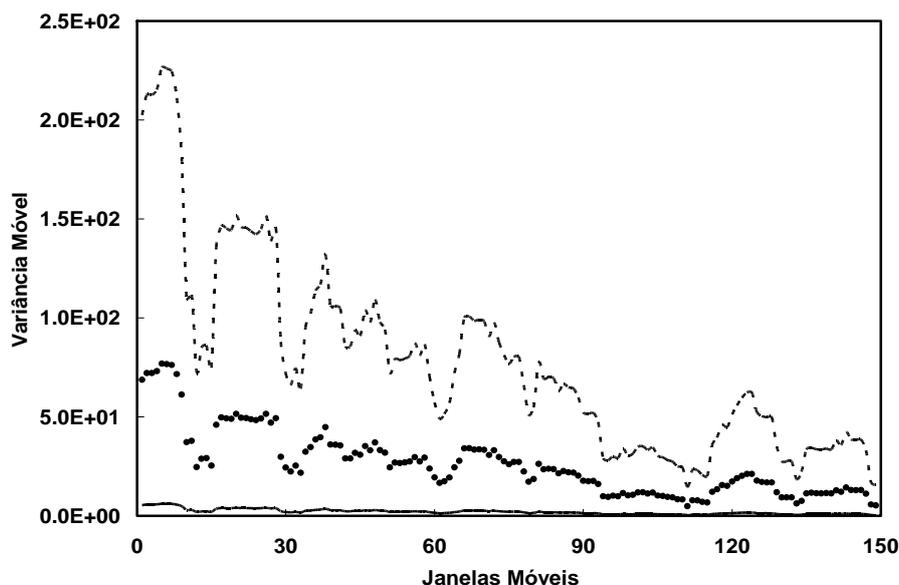


Figura 4.11: Intervalo de confiança (---) e variância móvel (●) dos resíduos.

Foi definido um intervalo amostral maior para análises das médias e variâncias móveis dos resíduos com o objetivo de analisar as regularidades destas ao longo da operação do processo, neste caso 1 dia de operação. O intervalo de amostral composto por 150 janelas móveis, possibilitou a percepção clara do afinamento dos intervalos de confiança destas duas entidades estatísticas, que vão estreitando progressivamente ao longo da amostragem. Esta observação implica em um aumento progressivo confiabilidade das médias e variâncias móveis ao longo da amostragem. Este fato pode estar relacionado a qualidade dos dados operacionais, mas provavelmente indicam que ao longo da amostragem a estabilidade é maior que no início, e que esta estabilidade aumenta com o tempo.

Quanto as médias dos resíduos, no geral, apresentam um bom comportamento, oscilando com valores positivos e negativos entre o valor zero, novamente descartando a idéia do *bias* do modelo, confirmando o resultado apresentado pela Tabela (4.8).

As elevadas estimativas das variâncias reportam a flutuação experimental e demonstram haver flutuações dinâmicas significantes no processo. Este resultado indica que a variância do modelo não pode ser considerada compatível com a flutuação característica do processo.

É importante lembrar que tanto a média móvel quanto a variância móvel apresentam largos intervalos de confiança. Estes resultados indicam que para estas entidades há muitos outros resultados possíveis; portanto, a confiabilidade destes valores não é elevada.

Posteriormente foram calculados os percentuais de conformidade da média e da variância, que como já foi visto as análises de médias e variâncias móveis, expressando em índices a suas validades.

A Figura (4.12) apresenta a conformidade da média e da variância, respectivamente, para janelas móveis de 1 hora.

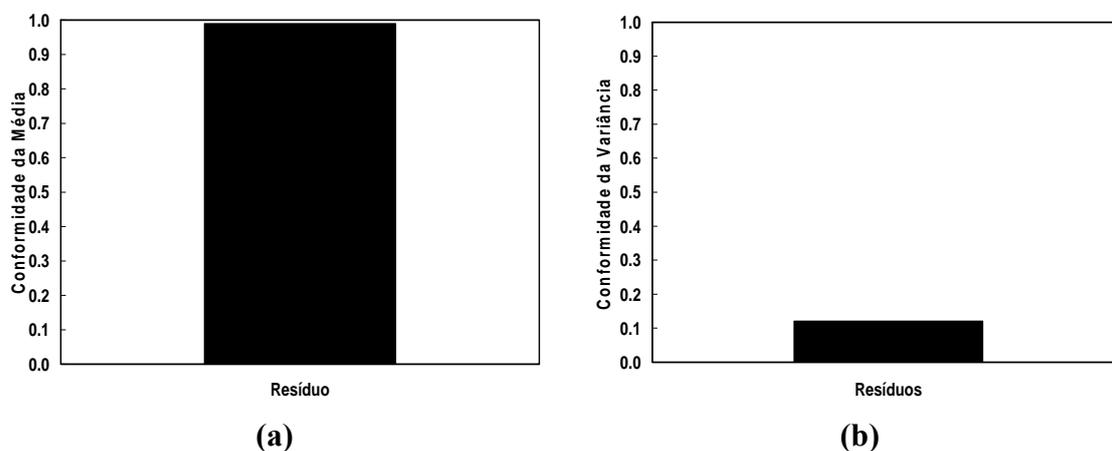


Figura 4.12: Conformidade da média e da variância para os resíduos.

A análise de conformidade da média, Figura (4.12) gráfico (a) apresenta o índice de conformidade próximo a 1,0, o que significa que as médias de cada janela móvel são estatisticamente similares entre si definindo claramente a não necessidade do acréscimo do *bias* ao modelo inferencial em linha.

Com relação à conformidade da variância, o baixo índice confirma que a flutuação dos dados operacionais não está conforme a flutuação das inferências. Este resultado permite especulações sobre a validade do modelo inferencial, que pode estar com problemas de parametrização (sub ou super parametrizado) ou ainda não representar adequadamente a relação funcional entre os dados operacionais e a variável desejada.

Um outro problema possível é o uso de dados mal amostrados contaminados por erros grosseiros no cálculo da inferência, pois estes erros usualmente inflam a média e variância causando estimativas desviadas de seus valores reais; porém, esta hipótese parece não parecer ter muita relevância, pois não só afetaria a variância móvel como também a média móvel.

Lidando com os dados operacionais (Figura 4.1), a flutuação do modelo inferencial é muito menor do que a apresentada por estes. Neste caso, é possível que o modelo esteja super parametrizado, de modo a garantir predições satisfatórias em linha por ajuste destes parâmetros.

Em seqüência, a Figura (4.13) apresenta o espectro de variância dos resíduos, para um tamanho da janela móvel de 1 hora.

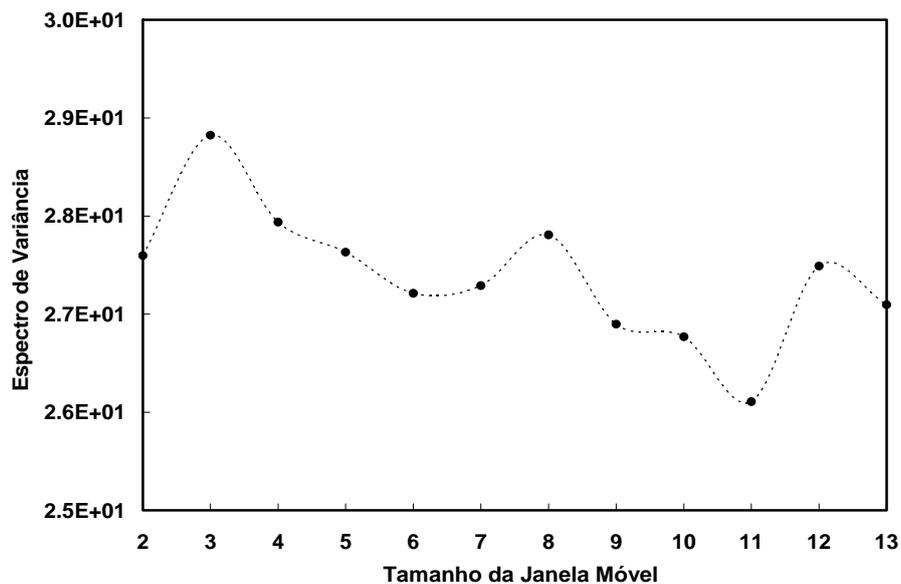


Figura 4.13: Espectro de variância dos resíduos.

O espectro de variância apresenta uma baixa oscilação e uma variação muito pequena (2°C). Estas características são típicas de estabilidade, demonstrando que não incorporação de variabilidade ao longo da operação do processo.

Uma observação significativa é o fato da oscilação do espectro de variâncias apresentar valor maior para janelas de menor tamanho (como por exemplo, janelas contendo 3 pontos amostrais). Isto significa que a variância de uma medida é mais afetada pela medida mais próxima a ela, e que ao longo do processo vai se estabelecendo uma condição de estabilidade.

O espectro de correlação dos resíduos, para um tamanho da janela móvel de 1 hora, é apresentado pela Figura (4.14).

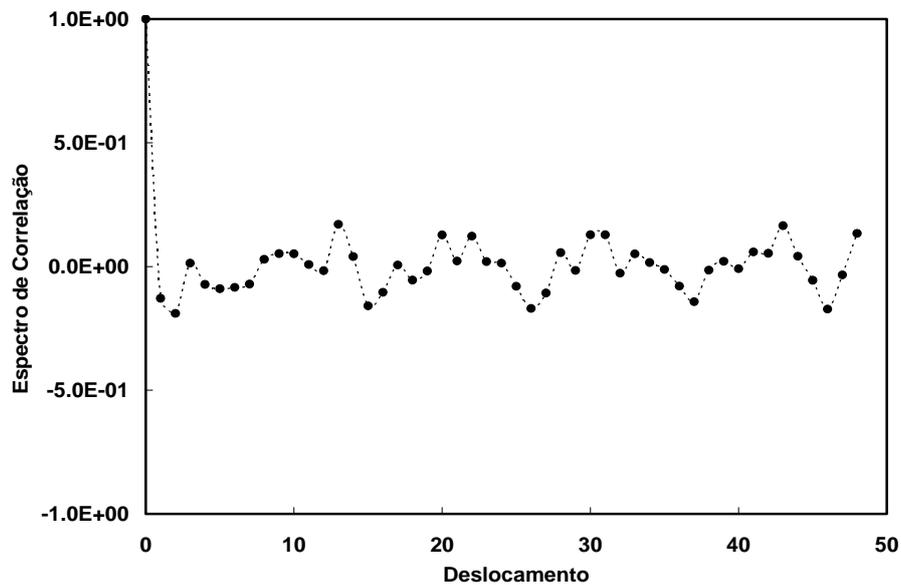


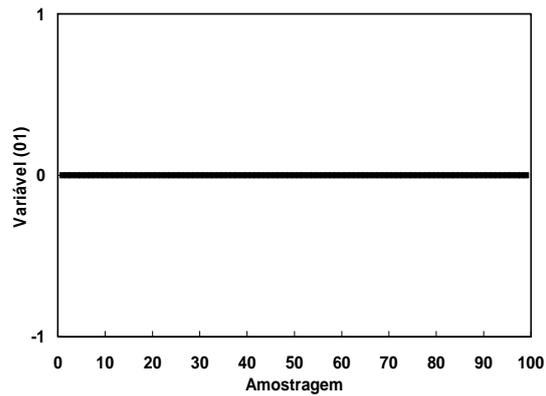
Figura 4.14: Espectro de auto-correlação dos resíduos.

Nota-se que a auto-correlação cai bruscamente logo após o primeiro deslocamento, oscilando em torno do valor zero, à medida que os deslocamentos aumentam. Este comportamento demonstra que não há dinâmica nos resíduos, que oscilam em torno do valor zero ao longo de todo o processo. Portanto, embora o modelo seja desenvolvido para operação em estado estacionário parece ser capaz de filtrar a dinâmica do processo.

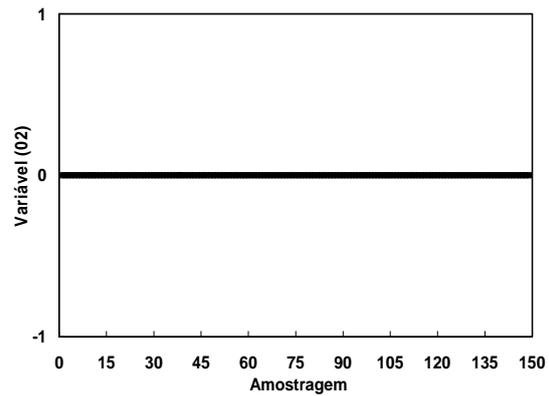
Uma importante informação obtida desta análise de correlação é o fato de não haver defasagem entre as medidas de processo e o modelo inferencial.

As próximas análises são realizadas especificamente para os modelos inferenciais e, portanto não constam na parte do estudo relacionada diretamente a análises dos dados (item 4.2).

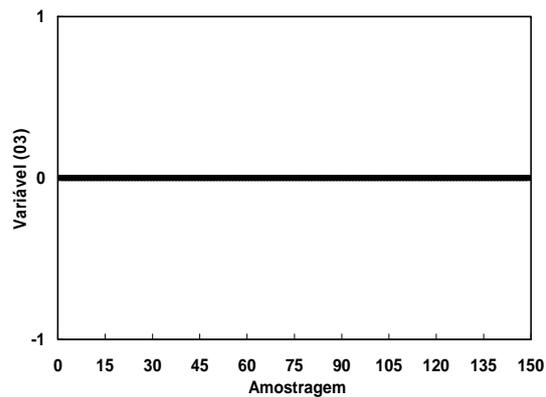
A Figura (4.15) apresenta do efeito de uma determinada variável independente sobre os erros de predição, por meio de índices.



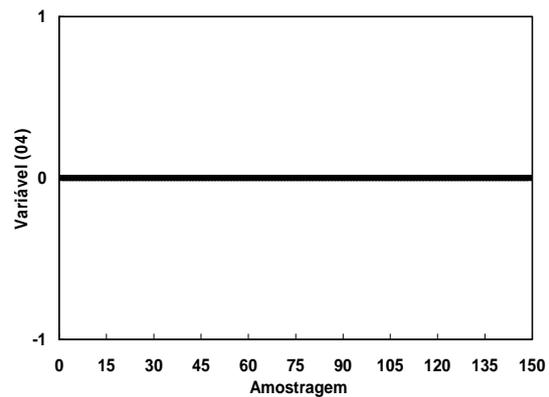
(a)



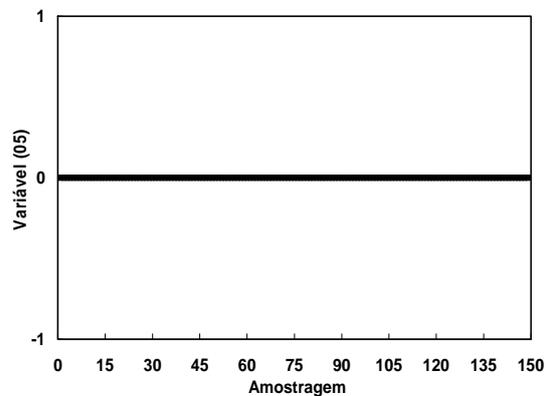
(b)



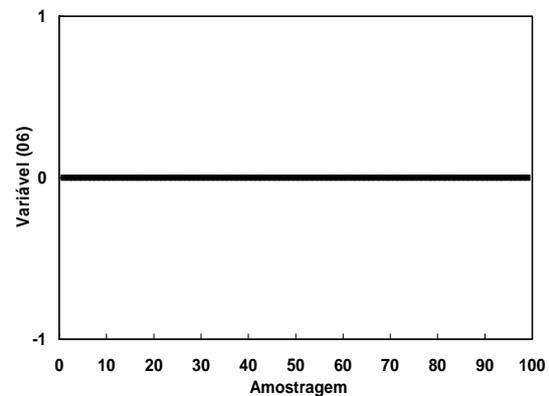
(c)



(d)



(e)



(f)

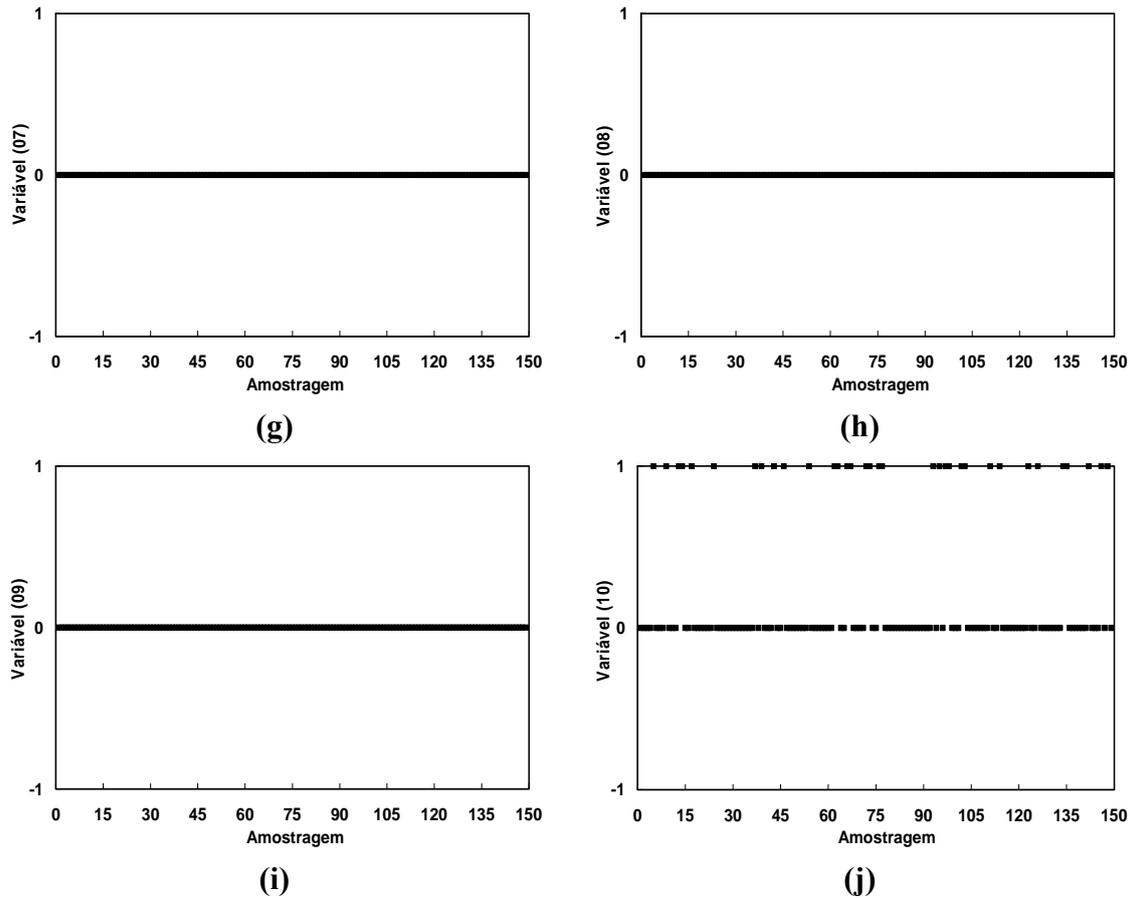


Figura 4.15: Efeito das variáveis independentes sobre a inferência.

Nos resultados acima, Figura (4.15) o erro de predição (que não é senão a variância de predição) é calculado alternando a cada gráfico o efeito (que não senão a variância) de cada variável independente presente no modelo inferencial. Por não estar disponível a matriz de covariância experimental, foi utilizada uma matriz composta dos primeiros valores obtidos para os espectros de variâncias das variáveis operacionais.

Observa-se que as variáveis (exceto a Variável (10)) não afetam de forma significativa os erros de predição, pois apresentam índice 0 de significância estatística para este caso. Entretanto, a retirada do efeito da Variável (10) afetou alguns experimentos, demonstrando que esta variável interfere no erro de predição.

A Figura (4.16) apresenta os resultados para identificação do efeito de um determinado parâmetro, e ainda a variável dependente, sobre os erros de predição.

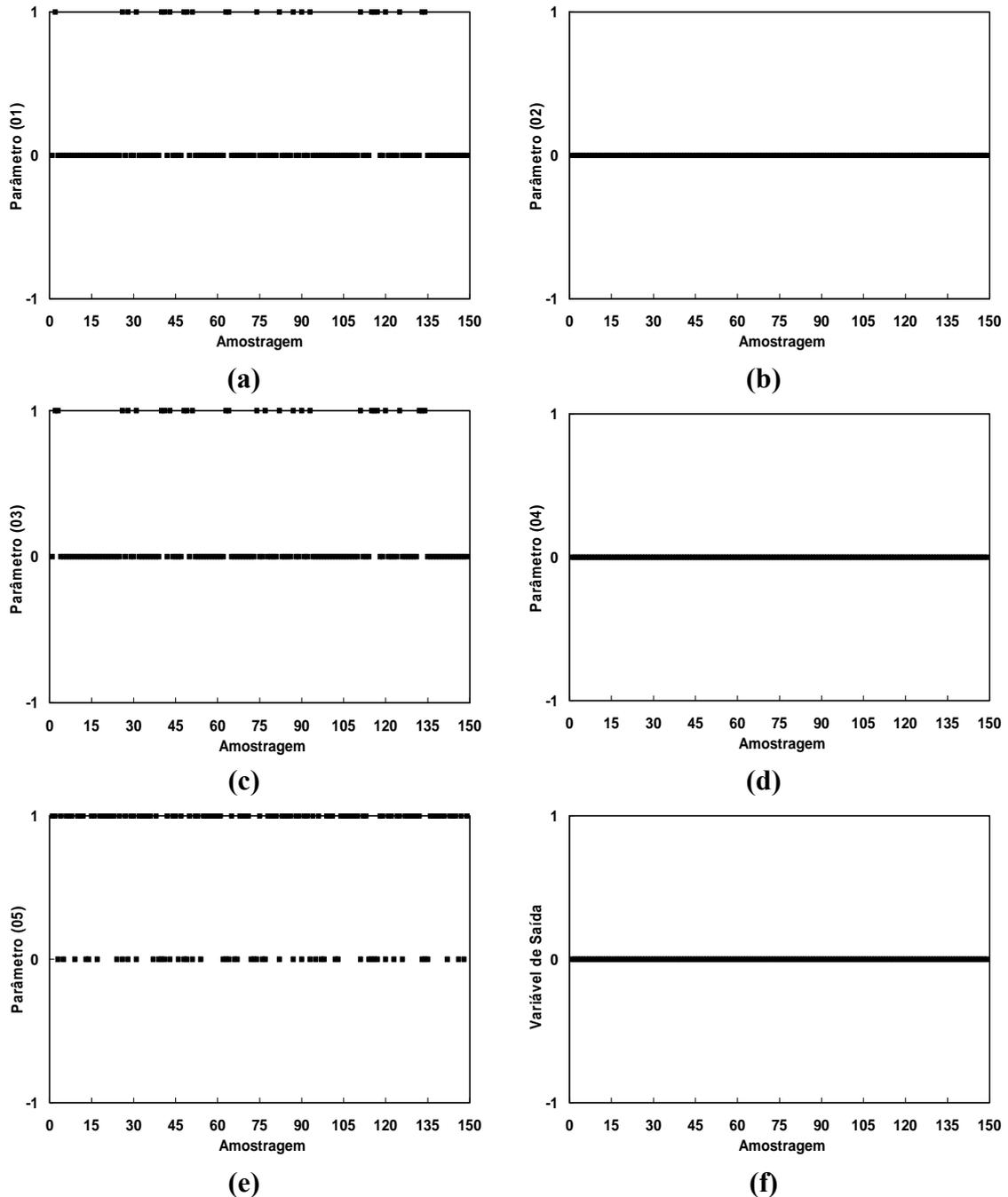


Figura 4.16: Efeito dos parâmetros e da variável de saída sobre a inferência.

Nos resultados acima, Figura (4.16) o erro de predição é calculado alternando o efeito de um determinado parâmetro do modelo a cada gráfico. Por não estar disponível a matriz de covariância dos parâmetros, foi utilizada uma matriz identidade para esta solução. A matriz de covariância dos parâmetros é geralmente obtida durante a regressão do modelo (reconciliação de dados ou estimação de parâmetros), procedimento não realizado neste trabalho. O gráfico (f) da Figura (4.16) apresenta os

resultados para a variável dependente, que assim como as variáveis independentes, não possui disponível sua variância experimental.

Dessa forma, os resultados apresentados acima não podem ser encarados com rigor de forma a reprovar estes elementos do modelo (parâmetros e variável de saída), mas como uma validação do procedimento numérico responsável por estas análises que permite a visualização da influência dos parâmetros e da variável de saída na inferência.

Mesmo neste contexto, parece claro que os parâmetros que não possuem uma influência estatisticamente significativa são os 2 e 4 (Figura (4.16), gráficos (b) e (d)), e da mesma forma atua a Variável de Saída no erro de predição.

Estas análises são de grande importância, pois ao identificar as os elementos do modelo inferencial que contribuem para o erro de predição (“as causas”) torna-se mais fácil avaliar este modelo e obter inferências de melhor qualidade.

Outras análises importantes de investigação do modelo inferencial estão relacionados à sua possibilidade de representar os dados do processo, quando vários modelos inferenciais são disponíveis para a realização da mesma inferência. Portanto, estas análises são empregadas para avaliação do desempenho do modelo inferencial e para a discriminação de modelos, e consiste no cálculo das probabilidades absolutas e relativas de cada um dos modelos propostos. A Figura (4.17) apresenta os resultados para as análises relativas à avaliação do desempenho do modelo.

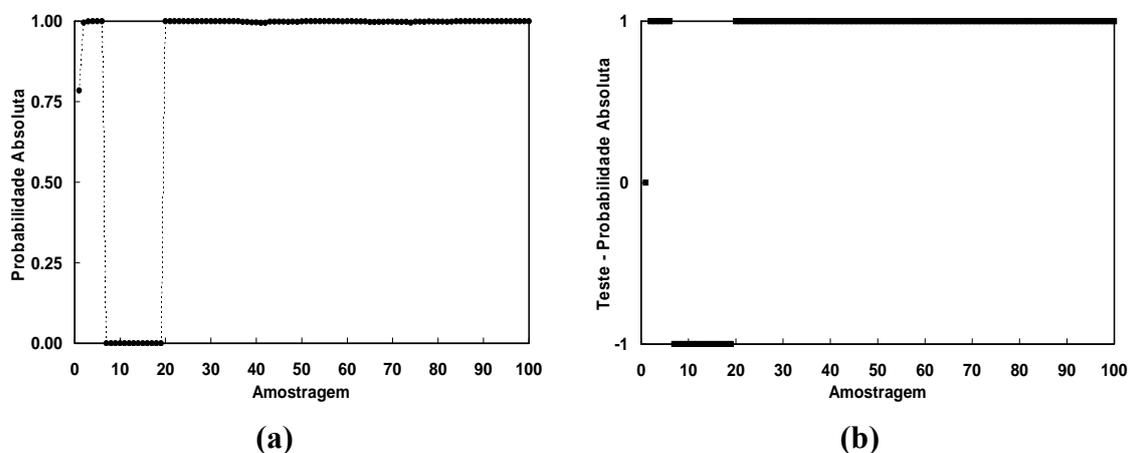


Figura 4.17: Probabilidades do modelo.

De acordo com análise de probabilidade absoluta do modelo inferencial apenas para o primeiro ponto amostral o modelo apresenta-se adequado. Neste caso o resultado é apresentado pelo índice 0 no teste de probabilidade absoluta.

Observa-se um trecho crítico (intervalo de amostragem do ponto 6 ao 20), representado pelos gráficos (a) e (c) da Figura (4.17), em que os valores obtidos para a probabilidade absoluta é uma constância de 0, apresentado mais claramente pelo teste de probabilidade pelo índice -1. O índice -1 implica em um modelo não adequado para representação dos dados, pois a probabilidade é menor que o nível de confiança exigido.

Ainda para a probabilidade absoluta, nota-se que a maior parte dos resultados obtidos assume valores de 1, representados no teste de probabilidade absoluta pelo índice 1. Este resultado implica em um modelo com desempenho exageradamente bom, pois a probabilidade é maior que o nível de confiança exigido. Neste caso, os resultados indicam que o modelo pode estar super parametrizado.

Analisando os valores obtidos para probabilidade relativa, observa-se que o modelo é considerado adequado para todos os experimentos. No entanto, estes resultados (incluindo o teste de probabilidade relativa) não expressam significado estatístico, visto que não foi empregado outro modelo inferencial para obtenção de resultados de comparação. Logo, não há dúvida que na ausência de outros modelos, um modelo inferencial comparado a ele mesmo é um modelo adequado. O objetivo destes resultados é apresentar a ferramenta estatística, que foi implementada neste estudo e está disponível para discriminação quando houver outros modelo inferenciais (da mesma propriedade) a serem analisados.

O objetivo dos testes é apresentar os resultados de forma mais simples e clara de modo que decisões e ações de correção possam ser tomadas em tempo real por profissionais que não necessariamente entendam extensivamente de estatísticas de processos (como por exemplo, os operadores).

Finalmente, são realizadas análises a partir dos resíduos capazes de determinar se um dado experimental pode ou não ser considerado um erro grosseiro. Além disso,

também são apresentados a análises que identificam o número de erros grosseiros observado num determinado conjunto de dados, e a probabilidade acumulada destes estarem presentes.

A Tabela (4.9) e a Figura (4.18) apresentam os resultados obtidos para a detecção de erros grosseiros nos resíduos.

Tabela 4.9: Detecção de erros grosseiros (EG) nos resíduos.

Resíduos		
Número de EG	Número máximo de EG	Probabilidade Acumulada de EG
1	8	0,00405

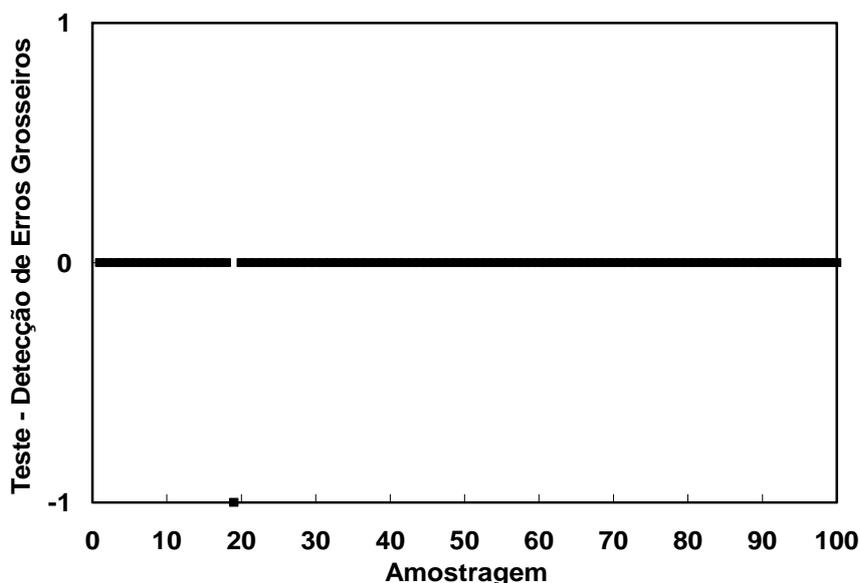


Figura 4.18: Detecção de erros grosseiros nos resíduos.

O número de erros grosseiros detectado pode expressar a boa qualidade dos dados amostrados ou o alto valor da variância de predição. Como mostrado na Tabela (4.9), para 100 dados eram esperados até 8 erros grosseiros com 95% de confiança estatística. No teste de detecção de erros grosseiros observa-se que os valores obtidos são dados por índices 0 e -1; o índice zero significa que o ponto não é considerado erro grosseiro, mas o índice 1 indica que o ponto é estatisticamente um erro grosseiro e este índice é positivo quando se trata de um erro grosseiro positivo e negativo quando considerado um erros grosseiro negativo (como neste caso apresentado). A detecção de

apenas 1 erro grosseiro resulta das altas variâncias de predição e sugerem que estes valores estão superestimados. Não parece haver dúvidas de que os parâmetros do modelo precisam ser reavaliados, assim como os desempenhos das medidas experimentais.

Cabe lembrar que a presença de erros grosseiros não necessariamente significa inconsistência estatística dos dados analisados; afinal o próprio fato de assumir uma hipótese de distribuição, ao definir um nível de confiança, contempla a possibilidade de existência de erros grosseiros.

4.4 ADEQUAÇÃO DO MODELO INFERENCIAL

Os modelos em geral são compostos basicamente por três elementos: as variáveis independentes x , as variáveis dependentes y e os parâmetros θ . A relação matemática entre as variáveis independentes (geralmente obtidas via sensores) e as variáveis dependentes (geralmente obtidas por meio de análises de laboratório) compõe a estrutura do modelo. Já os parâmetros são certos valores ou constantes que não podem ou não são facilmente medidos (como por exemplo, a energia de ativação de uma reação química), mas que podem ser inferidos a partir de observações experimentais (SCHWAAB, 2007).

O procedimento de inferência dos parâmetros de um modelo é chamado de estimação dos parâmetros e consiste em ajustar os valores dos parâmetros de tal forma que as previsões do modelo sejam as mais próximas possíveis dos valores medidos experimentalmente. SCHWAAB (2007) apresenta um estudo abrangente sobre este tema, incluindo um histórico envolvendo trabalhos remotos desenvolvidos nesta área comum à engenharia e à estatística.

Dentre os objetivos definidos para este estudo, a (re) modelagem não está inclusa. Frente à necessidade de uma melhor predição, requerendo a adequação do modelo, a abordagem empregada será a re-estimação dos parâmetros a fim de avaliar se

estes estão estatisticamente consistentes e, ainda, se o modelo está super ou sub parametrizado. Desta forma, é possível também definir se os problemas de predição do modelo estão ligados aos parâmetros ou se é um problema de estruturação/relação entre as variáveis.

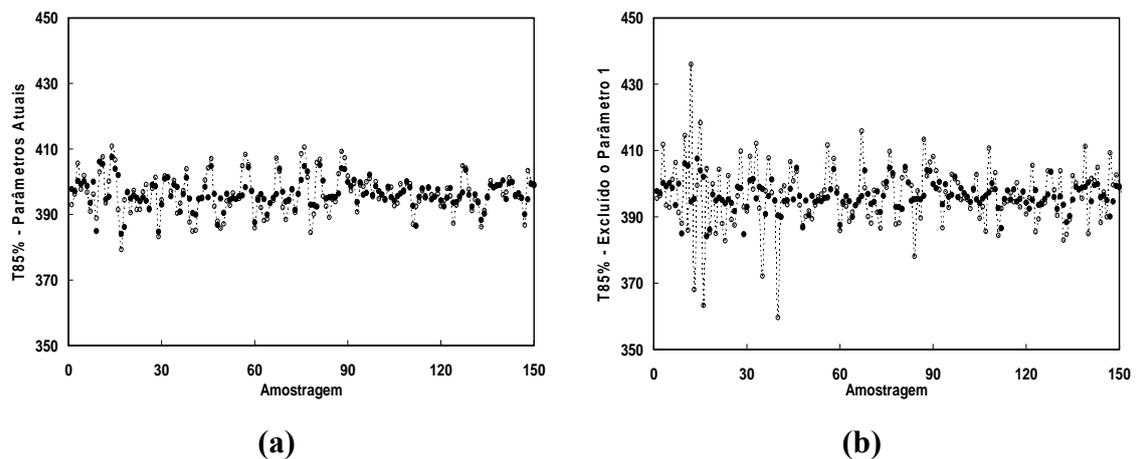
Os parâmetros originais empregados pelo modelo inferencial são apresentados pela Tabela (4.10).

Tabela 4.10: Parâmetros originais do modelo inferencial.

Parâmetro 1	Parâmetro 2	Parâmetro 3	Parâmetro 4	Parâmetro 5
1,090	-15,000	1,650	-129,700	0,000

Inicialmente, buscou-se averiguar de forma qualitativa a significância dos parâmetros. Esta verificação consiste na exclusão do parâmetro a ser averiguado do modelo inferencial, observando a influencia deste parâmetro no resultado após a simulação.

A Figura (4.19) apresenta a significância qualitativa dos parâmetros do modelo inferencial.



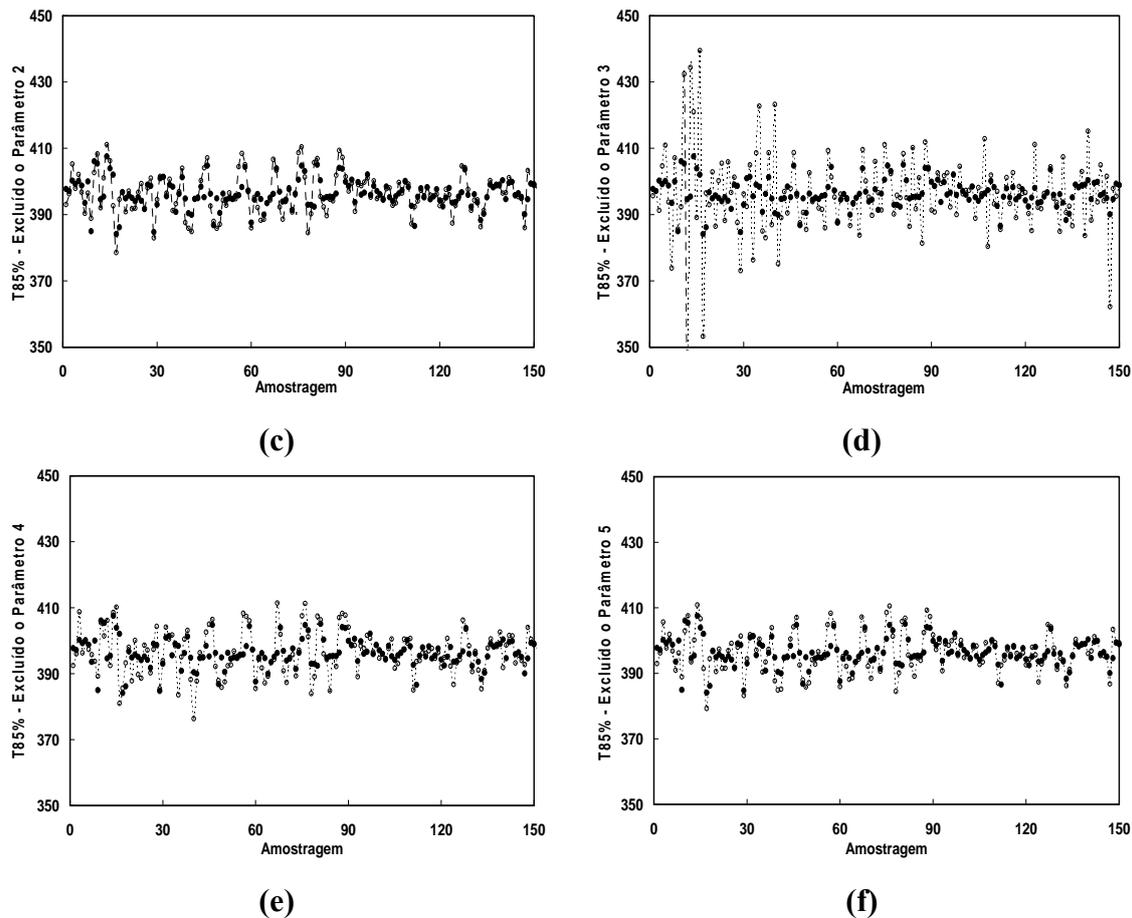


Figura 4.19: Inferência da T85% (-o-) sem um determinado parâmetro do modelo inferencial comparada aos dados medidos de laboratório (●)

A Figura (4.19) gráfico (a) representa a comparação da inferência com os dados medidos de laboratórios. Nesta figura os parâmetros empregados são os atualmente utilizados pelo modelo inferencial (em linha). Na Figura (4.19) gráfico (b) apresenta a comparação da inferência com os dados de laboratório, excluído o primeiro parâmetro; e assim, por diante.

Observa-se que a ausência dos parâmetros 2 e 4 (Figura 4.19, gráficos (b) e (d)) não alteram muito o valor da inferência, não apresentando influência significativa no modelo inferencial. Análises de variâncias demonstraram que esta observação está correta, pois a variabilidade neste caso não chega a 2 °C, enquanto para os demais parâmetros os valores para variabilidade ultrapassam a 8 °C. Em contrapartida, a retirada dos parâmetros 1 e 3 pode ser considerada significativa, pois afeta fortemente a qualidade da inferência, sendo que no caso da ausência do parâmetro 3 produz uma inferência extremamente ruidosa com elevada variância. A inferência obtida quando é

excluído o parâmetro 5 apresenta o mesmo resultado para a inferência realizada com os parâmetros atuais. A justificativa está no fato de que o modelo empregado em linha utiliza o quinto parâmetro com valor igual a zero. Portanto, a análise quantitativa dos dados confirma a análise de sensibilidades realizada anteriormente, já que ficou caracterizada a influência dos parâmetros no modelo.

Após esta avaliação, passou-se à etapa de estimação dos parâmetros. Para tal, empregou-se o pacote computacional ESTIMA (NORONHA *et al.*, 1993). O pacote resolve problema de máxima verossimilhança com o método Gaus-Newton. Os dados experimentais foram considerados independentes e as variâncias usadas foram aquelas apresentadas na Tabela (4.1).

As Tabelas (4.11) a (4.13) apresentam os resultados estatísticos obtidos para os novos parâmetros do modelo inferencial.

Tabela 4.11: Resultados da estimação de parâmetros do modelo inferencial.

	Parâmetro	Limite inferior	Limite superior	Desvio-padrão
Parâmetro 1	6,53E-01	5,69E-01	7,37E-01	4,25E-02
Parâmetro 2	-1,82E+01	-6,10E+01	2,45E+01	2,16E+01
Parâmetro 3	6,86E-01	4,77E-01	8,95E-01	1,06E-01
Parâmetro 4	-4,45E+01	-4,63E+01	1,35E+02	4,60E+01
Parâmetro 5	1,02E-01	-3,39E-02	2,39E-01	6,90E-02

Tabela 4.12: Resultados da matriz de covariância dos novos parâmetros.

Matriz de Covariância				
1,81E-03	-2,47E-01	-2,34E-03	1,18E+00	2,97E-04
-2,47E-01	4,68E+02	4,58E-01	2,29E+02	-2,55E-03
-2,34E-03	4,58E-01	1,12E-02	-4,23E+00	-9,69E-04
1,18E+00	2,29E+02	-4,23E+00	2,12E+03	4,33E-01
2,97E-04	-2,55E-03	-9,69E-04	4,33E-01	4,76E-03

Tabela 4.13: Resultados da matriz de correlação dos novos parâmetros.

Matriz de Correlação				
1,00E+00	-2,68E-01	-5,20E-01	6,03E-01	1,01E-01
-2,68E-01	1,00E+00	2,00E-01	2,30E-01	-1,71E-03
-5,20E-01	2,00E-01	1,00E+00	-8,68E-01	-1,33E-01
6,03E-01	2,30E-01	-8,68E-01	1,00E+00	1,36E-01
1,01E-01	-1,71E-03	-1,33E-01	1,36E-01	1,00E+00

Os novos parâmetros (4.11) apresentam valores em escala próximo aos parâmetros originais (exceto o parâmetro 4); o que implica em um ajuste do modelo inferencial às medidas de processo.

A significância estatística dos parâmetros 2, 4 e 5 é questionável, pois estes apresentam limites que passam pelo valor zero; o que indica possibilidade destes parâmetros assumirem este valor, não sendo significativos. Este fato também afetaria o modelo inferencial, pois o parâmetro oscila entre valores negativos e positivos, variando a tendência da inferência.

Esta análise estatística é concordante com a averiguação qualitativa realizada a respeito da significância dos parâmetros, onde nota-se a baixa (quase nula) influência de alguns parâmetros na qualidade da inferência (Figura (4.19)). Neste contexto, a eliminação de parâmetros do modelo fica associada essencialmente à qualidade da predição, já que a remoção de parâmetros pode prejudicar o desempenho do modelo, mesmo quando estes parâmetros não são significativos. Deve-se se observar, porém, que um modelo super parametrizado quando fora da região para qual foi realizada a estimação pode fornecer predições de qualidade insatisfatória (SCHWAAB, 2007).

Na Figura (4.20) é feita uma comparação entre os valores inferidos com os novos parâmetros e os dados de laboratório.

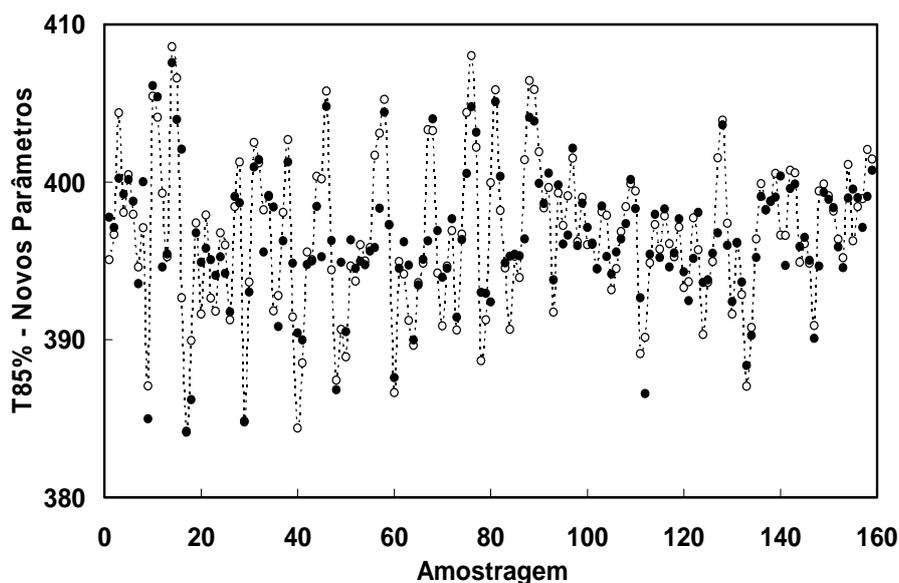


Figura 4.20: Inferência da T85% (-○-) comparada aos dados medidos de laboratório (●)

Com a re-estimação dos parâmetros que a variabilidade da inferência parece estar qualitativamente mais coerente com a flutuação dos dados. Portanto, pode-se dizer que a re-estimação dos parâmetros do modelo é capaz de melhorar a qualidade estatística da inferência; logo, esta inferência se apresenta mais confiável.

A Figura (4.21) compara os valores inferidos com os parâmetros antigos e com os novos parâmetros do modelo inferencial, em ambas as situações está sendo empregado o *bias*.

As inferências foram somadas ao *bias*, resultadas em nova inferências. Percebe-se que a re-estimação dos parâmetros não afetou muito a qualidade da inferência, se o *bias* continua a ser empregado. Este resultado significa que alguns parâmetros podem realmente não ter influencia na predição, indicando que o modelo pode estar super parametrizado. Neste caso, a má qualidade da inferência não deve ser atribuída aos parâmetros e levanta questionamentos sobre a validade da estrutura do modelo.

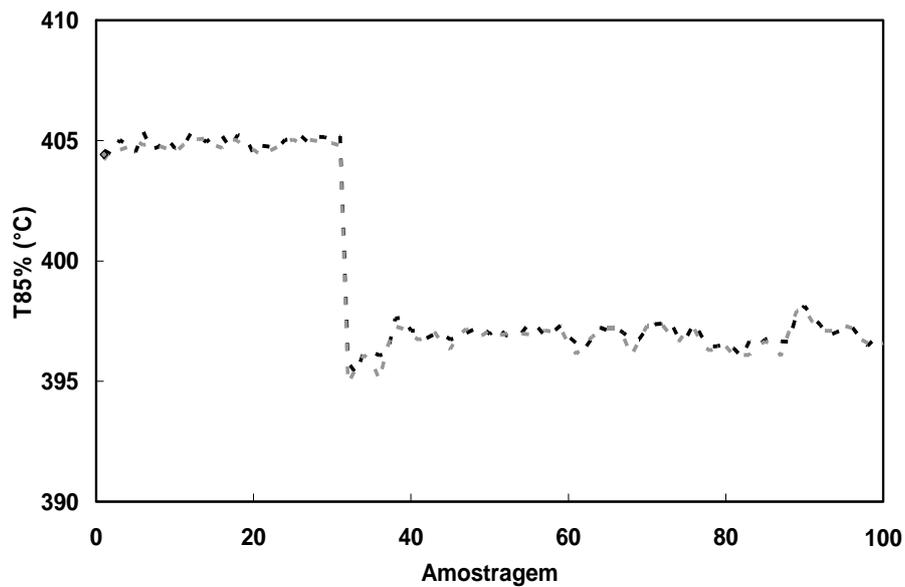


Figura 4.21: Comparação entre os valores inferidos empregando o *bias*, com os parâmetros antigos (---) e com os novos parâmetros (- -).

4.5 ANÁLISE DO MODELO INFERENCIAL PROPOSTO

Após a adequação do modelo, neste caso a re-estimação dos parâmetros, o modelo proposto é avaliado. Desta vez, o objetivo é avaliar melhoria na qualidade da inferência obtida pelo modelo com novos parâmetros.

As Figuras (4.22) e (4.23), respectivamente, apresentam a nova inferência e a variância de predição desta; enquanto a Tabela (4.14) apresenta a análise de caracterização básica (média e desvio-padrão).

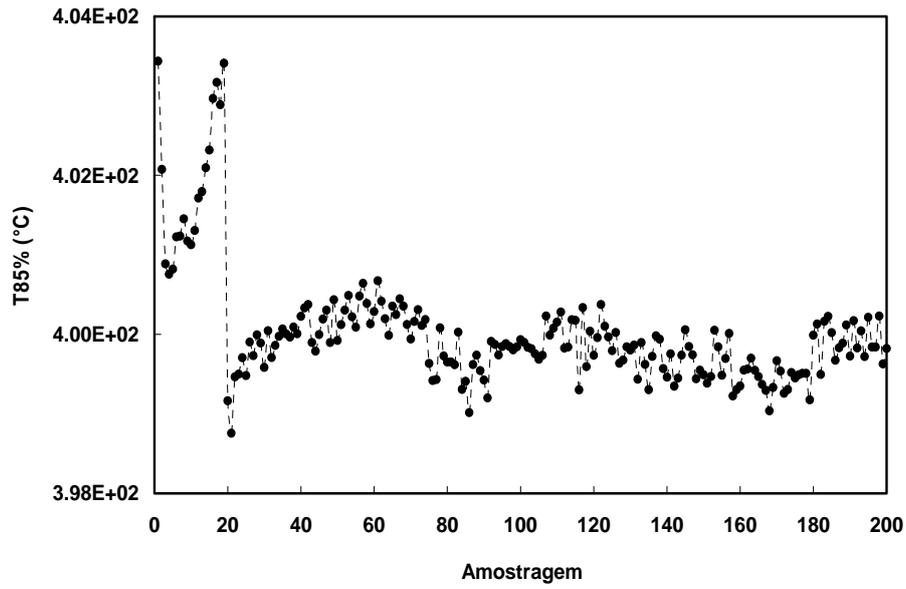


Figura 4.22: Inferência da T85%.

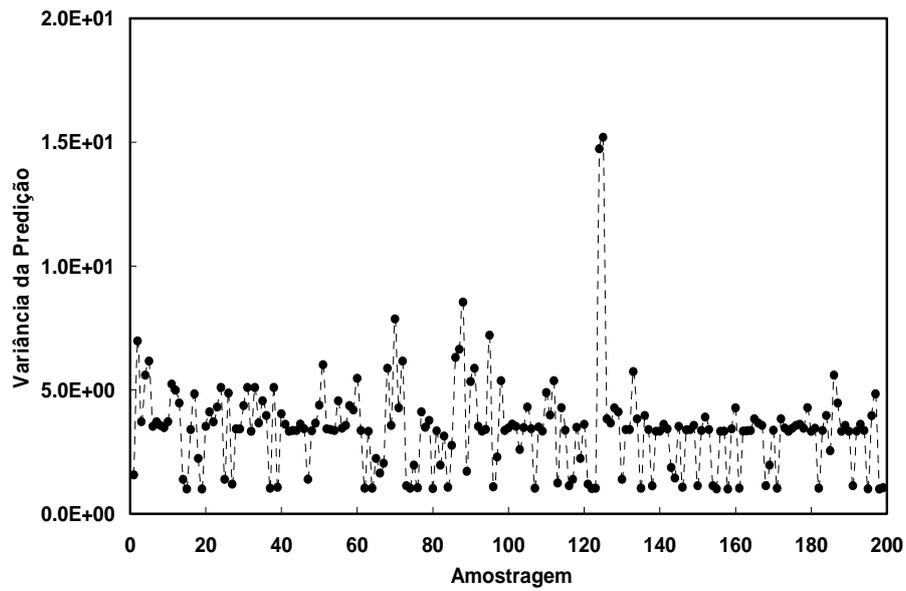


Figura 4.23: Variância da predição da T85%.

Tabela 4.14: Média e desvio-padrão para a inferência, após a adequação do modelo.

Inferência	
Média	Desvio-padrão
400,023	0,740

Constata-se uma melhoria na qualidade da predição (Figura 4.22), embora os valores obtidos para os novos parâmetros (Tabela 4.11) tenham se mantido próximos aos dos parâmetros empregados pelo modelo empregado neste estudo e pelo modelo inferencial em linha (como foi visto a diferença entre estes dois modelos é única e exclusivamente o *bias*). Esta pequena diferença entre os parâmetros pode indicar o ajuste do modelo aos dados operacionais deste determinado período proposto para re-estimação dos parâmetros. Este fato leva a interpretar que a estimação de parâmetros em linha pode ser uma boa alternativa para obtenção de inferências mais confiáveis.

Embora a re-estimação dos parâmetros do modelo inferencial tenha apresentado uma melhora na qualidade da inferência, a variância desta predição ainda se encontra alta. Contudo, não se pode negar que houve uma redução significativa do erro de predição, quando compara-se a Figura (4.8) com a Figura (4.23), o que demonstra que os parâmetros afetavam a qualidade da inferência, embora não sejam os únicos responsáveis por este problema. Os resultados da análise básica obtidos para a inferência após a adequação do modelo mostram que a variabilidade da inferência o modelo após a adequação do modelo inferencial é bem pequena. É importante lembrar, que a variância da predição calculada antes de ser realizada a adequação do modelo empregava como matriz de covariância dos parâmetros uma matriz identidade, sendo difícil avaliar essa melhoria do modelo inferencial só por esta análise

Mas são as análises dos resíduos que apresentam mais claramente a melhoria obtida a partir da re-estimação dos parâmetros do modelo inferencial. A Figura (4.24) apresenta os resíduos obtidos a partir do novo modelo inferencial proposto e a Tabela (4.15) caracterização estatística básica. Comparando a Figura (4.24) com a Figura (4.9), observa-se a grande melhoria do modelo, após a re-estimação dos parâmetros.

Como se pode observar, os resíduos possuem valores menores, compreendidos em um intervalo de amplitude muito menor, -5 a 2. Pode-se dizer que os valores estão compreendidos entre -1 e 1. Isto indica que as inferências obtidas a partir do novo modelo proposto estão mais próximas do valor medido em laboratório, do que com o modelo inferencial atualmente empregado em linha (Figura (4.9)), pois oscilam com menores amplitude e frequência ao longo da amostragem. É clara a existência de um

ponto discrepante dos demais. Se este ponto é desconsiderado pode-se dizer que o desempenho do modelo é excelente.

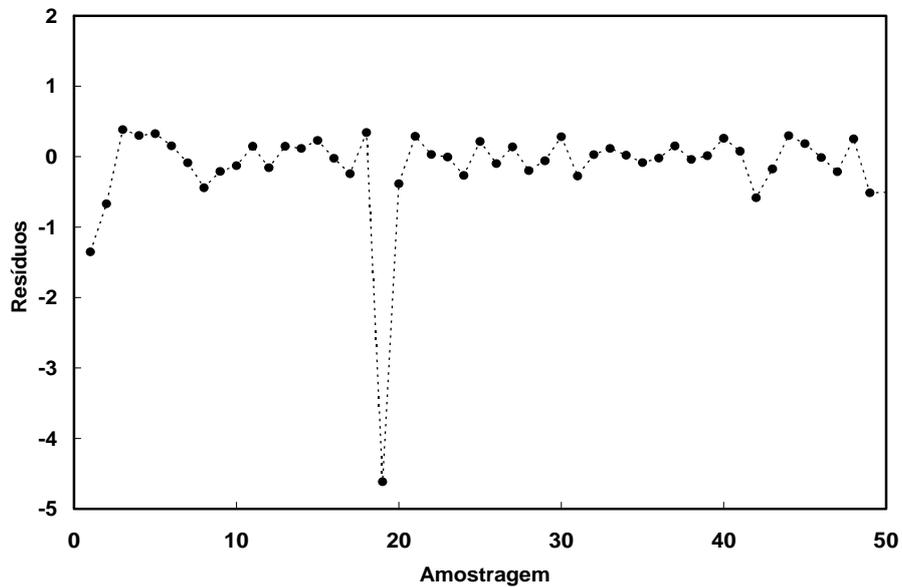


Figura 4.24: Comportamento dos resíduos.

As análises da caracterização básica apresentam a influência do ponto 19, cujo valor é -4,614, que torna a média negativa e o desvio-padrão maior. Se este ponto é considerado um erro grosseiro e é removido/substituído, o valor obtido para a média passa então a ser 0,000 e o valor do desvio-padrão é reduzido á metade. Como pode ser visto, erros grosseiros modificam a média e o desvio-padrão, fornecendo resultados equivocados e não representativos.

Tabela 4.15: Média e desvio-padrão para o resíduo.

Resíduos	
Média	Desvio-padrão
-0,032	0,426

Os resultados das médias e variâncias móveis também indicam melhora na qualidade da inferência após a adequação do modelo inferencial. As Figuras (4.25) e (4.26) apresentam respectivamente os resultados da média móvel e da variância móvel dos resíduos.

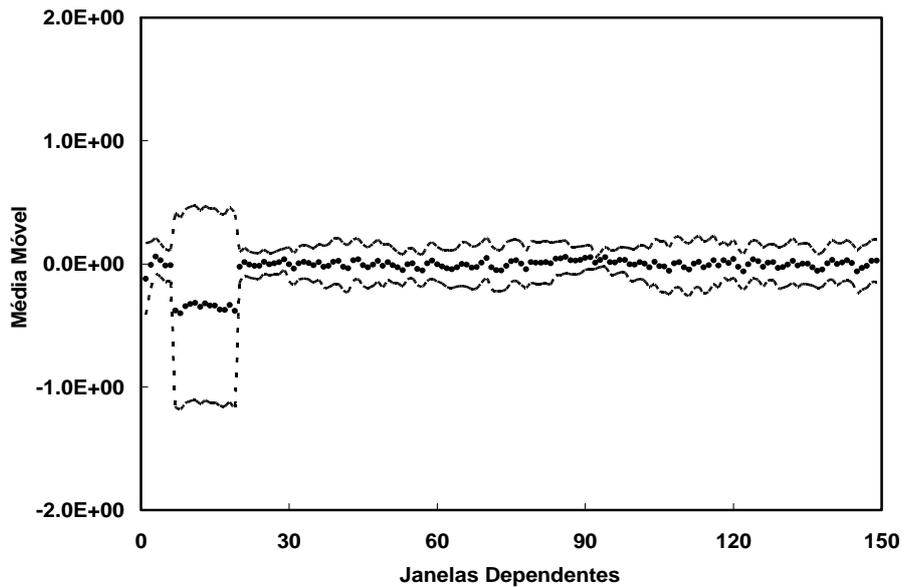


Figura 4.25: Intervalo de confiança (---) e média móvel (-●-).

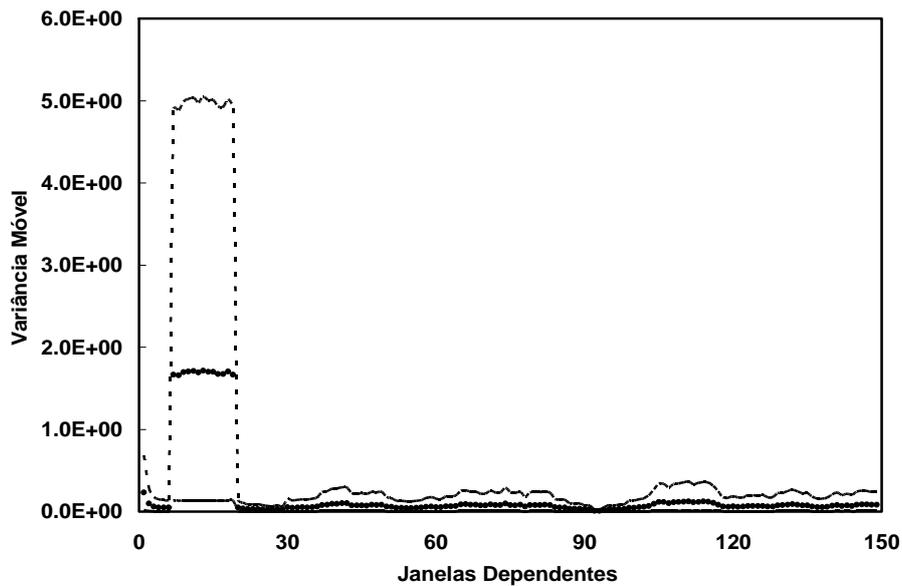


Figura 4.26: Intervalo de confiança (---) e variância móvel (-●-).

Em ambas as análises os intervalos de confiança aparecem bem estreitos, o que significa que os valores estão bem mais próximos do valor real. Observa-se também que os resultados obtidos no intervalo de amostragem de 5 a 20, são influenciados por um erro grosseiro, obviamente o resíduo do ponto de amostragem 19. É interessante relembrar o quão prejudicial pode ser um erro grosseiro em um conjunto de dados, dando margens a interpretações errôneas. Também é importante perceber o quanto é

importante o intervalo de confiança destas grandezas, visto que estes apontam claramente a dificuldade de estar representando um valor real ou próximo disto.

Uma informação relevante é o fato dos erros de predição do modelo inferencial com os parâmetros originais apresentar valores tão altos que encobriam a presença de erro grosseiro. Na ausência do erro grosseiro, o desempenho do modelo é homogêneo ao longo do intervalo de análise, mostrando a melhoria da inferência proposta.

Os resultados confirmam não só a melhoria da qualidade da inferência, mas também a não necessidade do *bias* aplicado ao modelo inferencial em linha. Os resultados da média móvel apontam que o modelo não necessita *bias*, afinal os resíduos oscilam em torno do zero de forma uniforme e compatível com os erros experimentais. Esta observação é confirmada pelos resultados obtidos para a variância móvel, cujos valores se apresentam bem próximos ao zero.

A Figura (4.27) apresenta os resultados para o cálculo da conformidade da média e da variância.

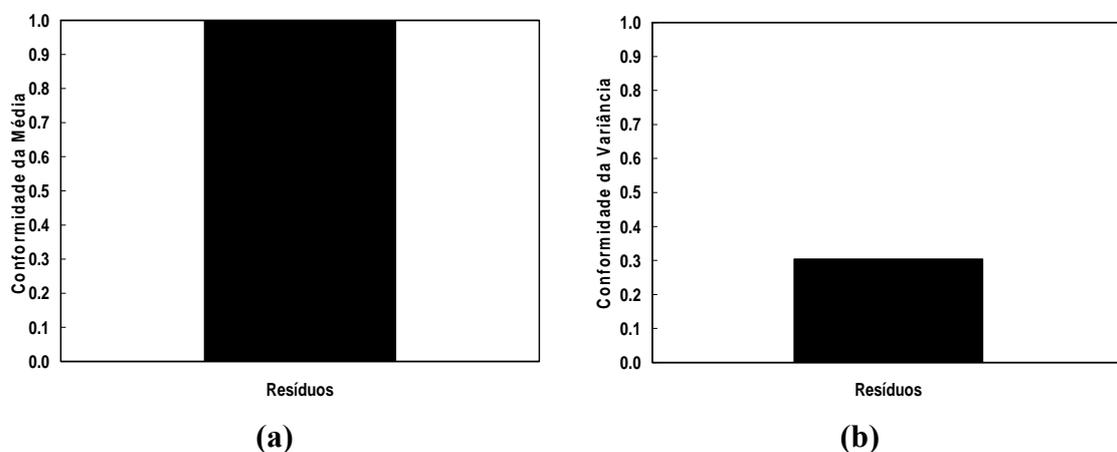


Figura 4.27: Conformidade da média e da variância para o resíduo.

O resultado de conformidade da média, Figura (4.27) gráfico (a) confirma claramente a não necessidade do acréscimo do *bias* ao modelo inferencial em linha. Este resultado indica que as médias entre os dados operacionais e a inferência apresentam total conformidade entre si, eliminando de todas as formas a hipótese do *bias*. O *bias* aparentemente se auto-degrada durante as inferências, não havendo motivo plausível

para empregá-lo, já que não mantém consistência estatística. Portanto, ao adicionar o *bias* à predição, está sendo adicionado ruído ao modelo.

Com relação à conformidade da variância, pode se observar que a re-estimação dos parâmetros pode elevar este índice, assegurando uma melhora da predição. No entanto, a flutuação dos resíduos não apresenta similaridade estatística ao longo do tempo.

Os espectros de variância e de correlação são apresentados respectivamente pelas Figuras (4.28) e (4.29).

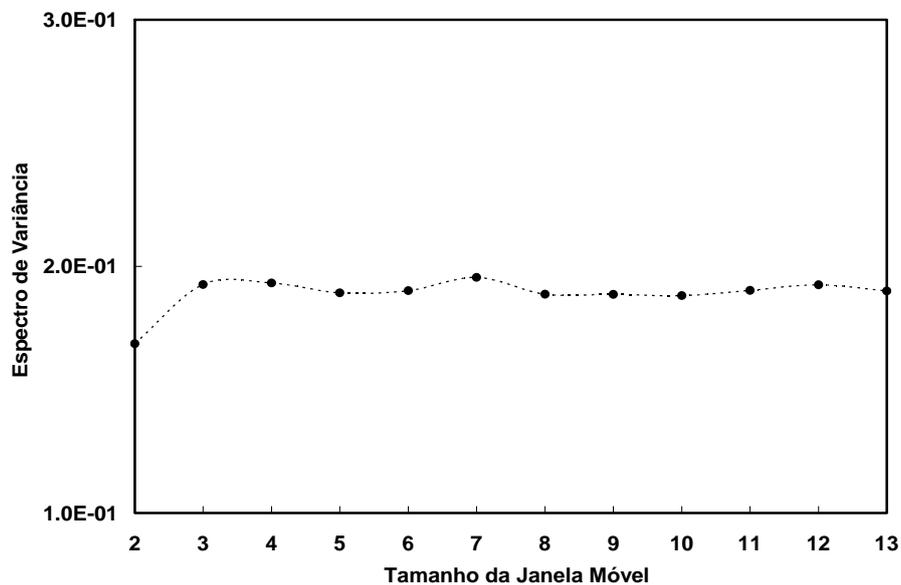


Figura 4.28: Espectro de variância dos resíduos.

O espectro de variância apresentado na Figura (4.28) expressa o comportamento típico estacionário. Há uma constância dos valores para os diferentes tamanhos de janela móvel, não havendo dinâmica na evolução das médias dos resíduos. O mesmo pode ser observado pelo espectro de auto-correlação, que cai bruscamente para o valor zero logo no primeiro deslocamento. Portanto, os espectros de variância e de correlação apresentados indicam que os resíduos estão se comportando de forma adequada para um processo operando em uma região de estabilidade. Então, apesar de haver uma dinâmica de fundo no processo, o modelo parece filtrá-la, mesmo tendo sido desenvolvido para representar o estado estacionário.

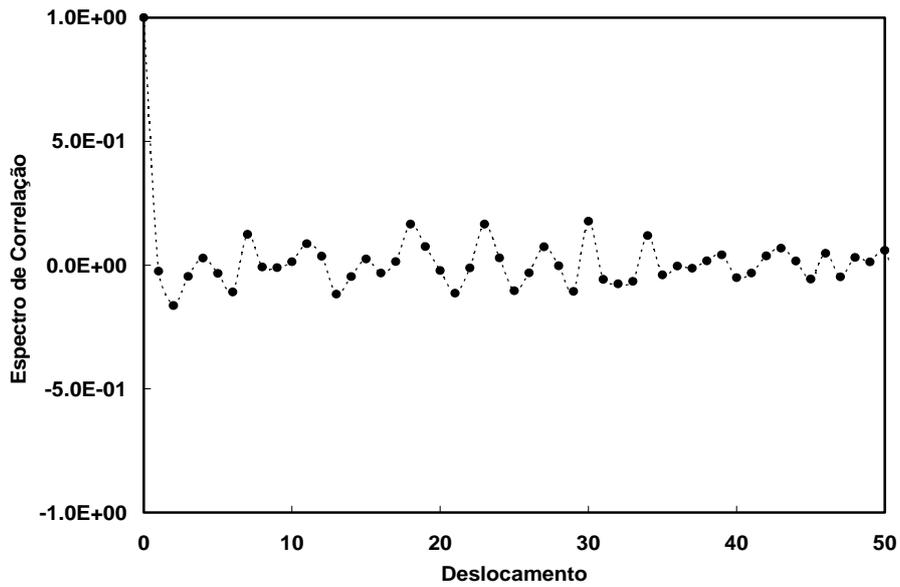
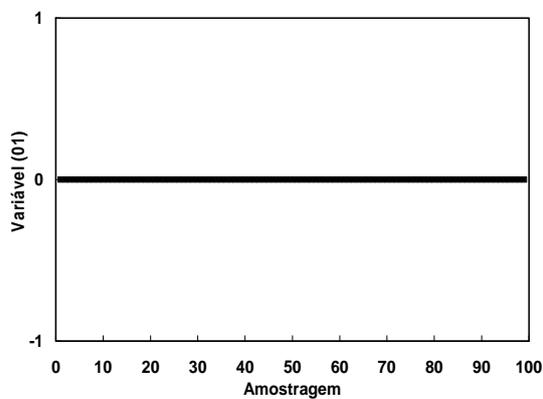


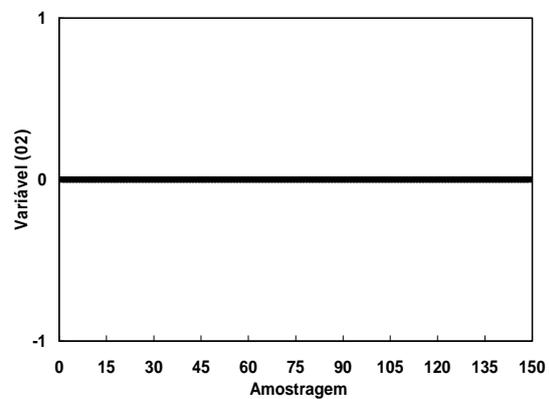
Figura 4.29: Espectro de auto-correlação dos resíduos.

Novamente são realizadas as análises de caracterização da influência das variáveis sobre o desempenho do modelo. Os resultados são apresentados nas Figuras (4.30) à (4.33) e na Tabela (4.16).

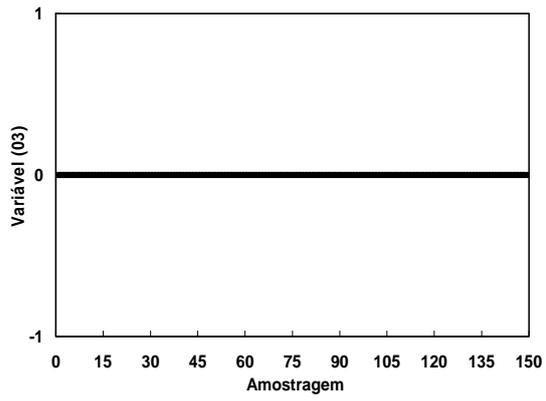
Na Figura (4.30) analisando a influência das variáveis no erro de predição observa-se que a Variável (10) interfere de modo significativo, enquanto as demais variáveis não apresentam qualquer influência. A justificativa para tal pode estar no fato desta variável ser representada pela inferência tomada na amostragem anterior da variável de saída atual.



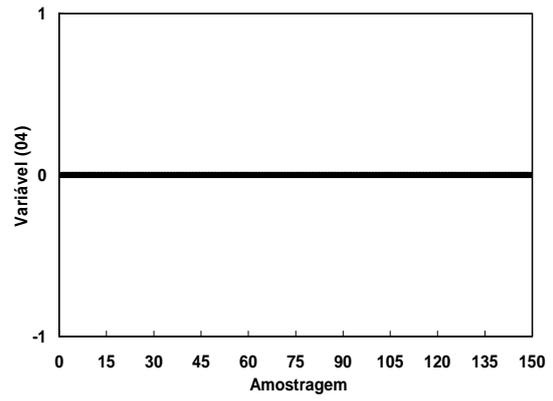
(a)



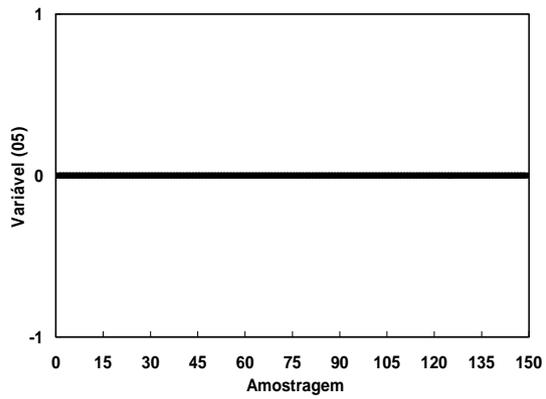
(b)



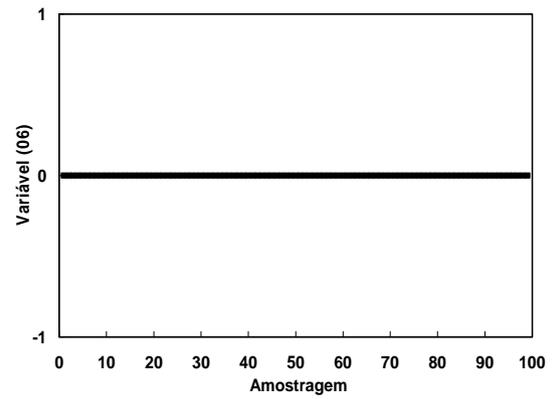
(c)



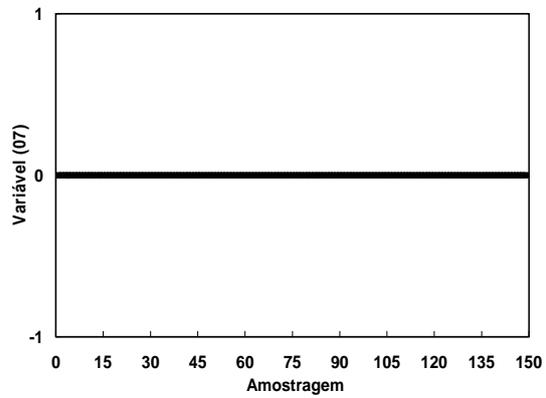
(d)



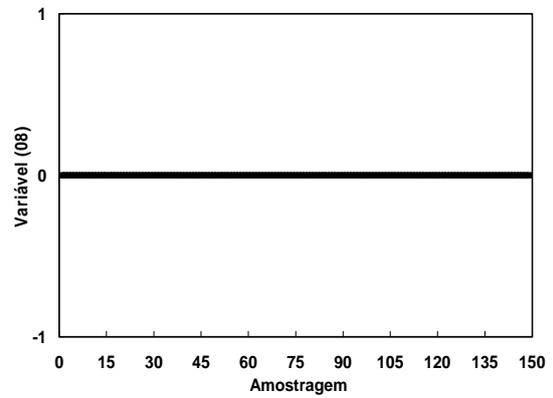
(e)



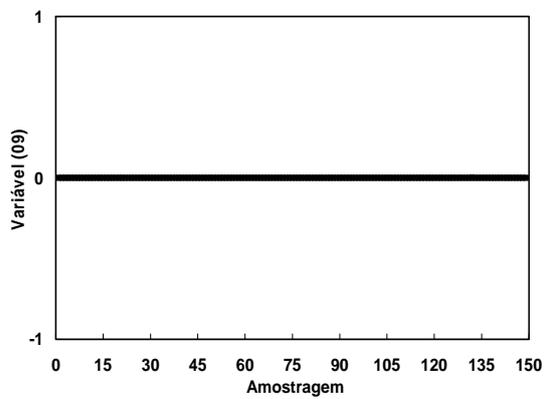
(f)



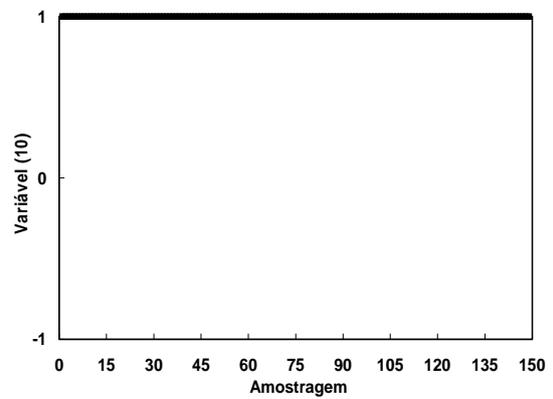
(g)



(h)



(i)



(j)

Figura 4.30: Efeito dominante das variáveis independentes sobre a inferência.

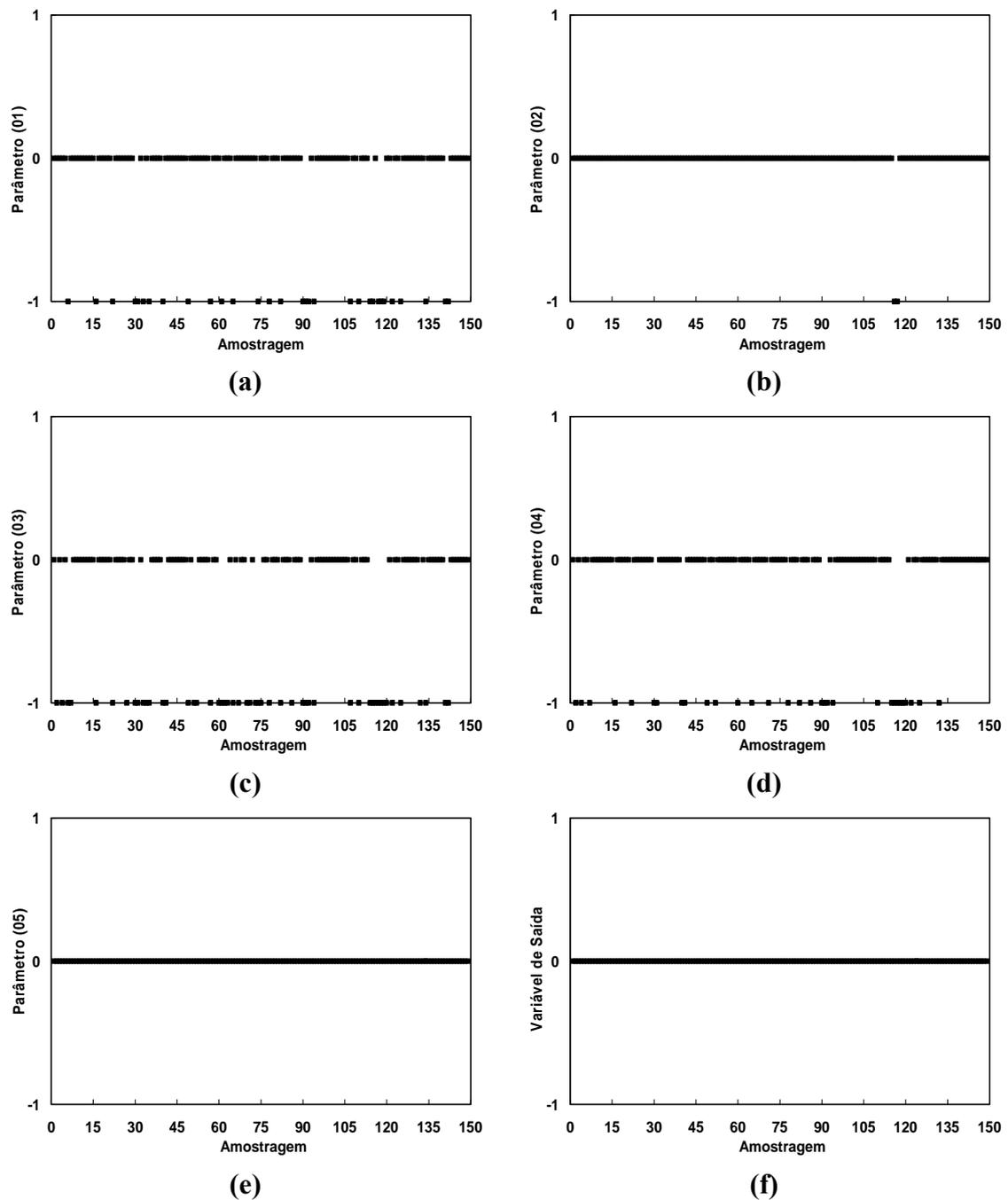


Figura 4.31: Efeito dos parâmetros e da variável de saída sobre a inferência.

Com relação aos parâmetros o mesmo pode ser observado na Figura (4.31), que com exceção do quinto parâmetro todos os demais afetam de forma significativa o erro de predição, sendo esta influência negativa. Estes resultados indicam o aumento do erro de predição quando a componente paramétrica é anulada. Apesar de parecer

matematicamente incoerente, estes resultados demonstram que os parâmetros do modelo inferencial podem estar fortemente correlacionados, como mostrado na seção 3.35. Portanto, parece óbvio que a correta caracterização das correlações é fundamental para a compreensão apropriada dos erros de predição do modelo.

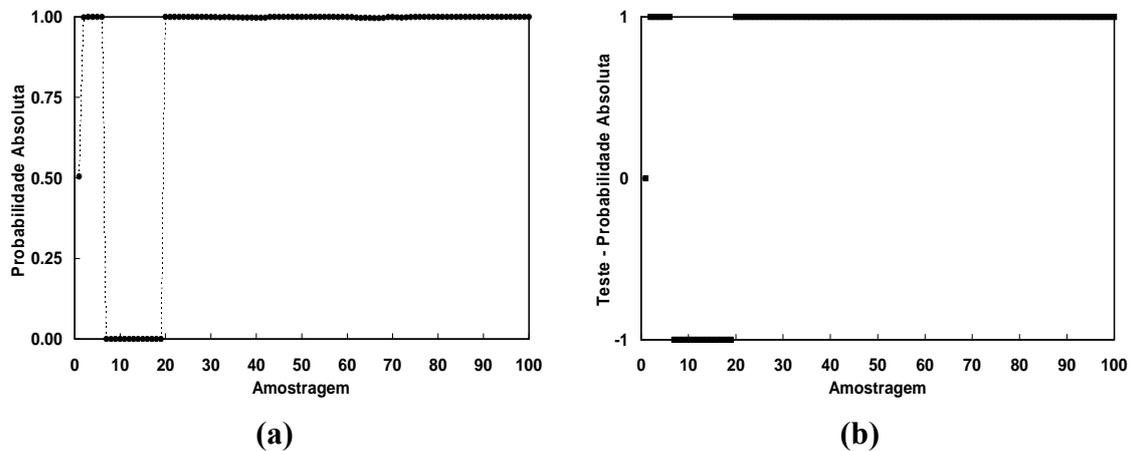


Figura 4.32: Probabilidades do modelo inferencial.

De acordo com análise de probabilidade absoluta do modelo inferencial apenas para o primeiro ponto amostral o modelo apresenta-se adequado, neste caso representado pelo índice 0 no teste de probabilidade absoluta. Observa-se que o trecho crítico particular (intervalo de amostragem do ponto 6 ao 20) em que o modelo não se apresenta adequado para representação dos dados, pois a probabilidade é menor que o nível de confiança exigido. Ao analisar o gráfico da inferência (Figura (4.7)) este trecho da amostragem realmente foi considerado bastante discrepante ao longo do período de operação; este resultado também foi observado nas médias e variâncias móveis dos resíduos (Figuras (4.25) e (4.26)) que são estatisticamente mais confiáveis à medida que se aumenta o número de amostragem.

Para os demais pontos do intervalo amostral a probabilidade absoluta continua a assumir os valores 1, representados no teste de probabilidade absoluta pelo índice 1. Este resultado implica em um modelo com desempenho exageradamente bom, pois a probabilidade é maior que o nível de confiança exigido. Neste caso, os resultados confirmam a possibilidade do modelo estar super parametrizado.

Tabela 4.16: Detecção de erros grosseiros (EG) nos resíduos.

Resíduos		
Número de EG	Número máximo de EG	Probabilidade Acumulada de EG
2	8	0,0182

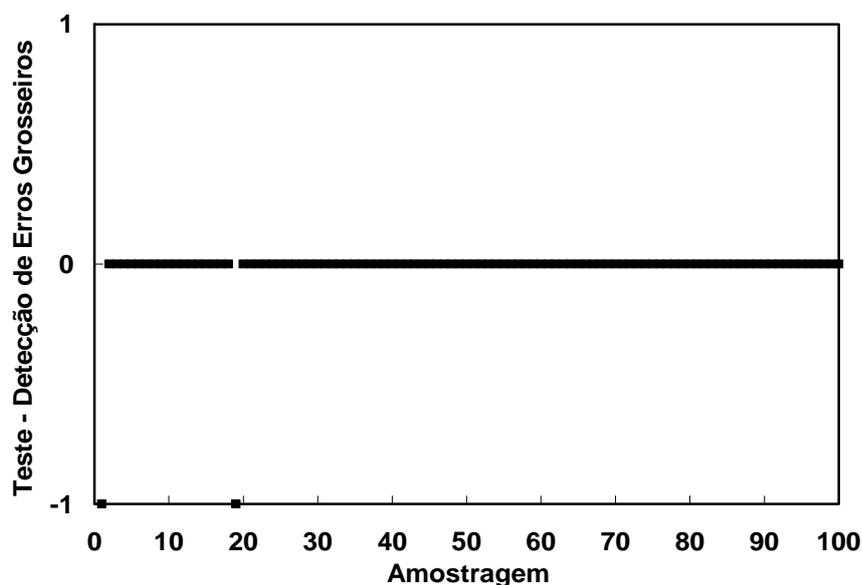


Figura 4.33: Detecção de erros grosseiros nos resíduos.

O número de erros grosseiros detectados continua a expressar a boa qualidade estatística dos dados amostrados, pois de acordo com o baixo valor obtido para a análise da probabilidade acumulada de erros grosseiros (Tabela 4.16). No teste de detecção de erros grosseiros apresentado pela Figura (4.34), observam-se 2 pontos que foram considerados erros grosseiros negativos, sendo que o ponto 19 pode ser reconhecido como tal ao observar os resíduos (Figura (4.24)). Neste caso, é possível que os altos valores obtidos para os resíduos calculados com o modelo empregando os parâmetros originais mascarassem esses pontos estatisticamente discrepantes do conjunto de dados.

CAPÍTULO V

CONCLUSÕES E SUGESTÕES

“Apesar de parecer um paradoxo, toda ciência exata é baseada na idéia de aproximação.”

BERTRAND RUSSEL

Neste estudo foi apresentada uma abordagem estatística para monitoramento e avaliação de inferências, com foco nos modelos inferenciais. Como ilustração, foi apresentada o caso da inferência da temperatura a 85% do diesel produto do processo de refino. Foram realizadas análises dos dados operacionais e do modelo inferencial empregado, a fim de investigar a consistência estatística da inferência. Posteriormente, foi proposta a adequação do modelo.

A análise estatística do modelo inferencial demonstra que embora, o modelo seja utilizado em linha, não se apresenta estatisticamente adequado para realização da inferência em questão. Esta análise também indica que o *bias* usado não é necessário do ponto de vista estatístico. Esse parâmetro pode estar acrescentando variância adicional ao processo e induzindo alterações dos valores desejados (*set points*) usados nos esquemas de controle, além aparentemente degradar ao longo da amostragem.

Embora a adequação do modelo por meio da reestimação dos parâmetros tenha tornado possível obter predições estatisticamente mais consistentes com a operação do

processo, há fortes indícios de que o modelo está super parametrizado. Essas conclusões podem ser relacionadas ao fato de que há parâmetros que pouca influência exerce sobre o desempenho de predição e ao fato de que as probabilidades do modelo são quase sempre muito próximas de um.

Assim, neste trabalho foram apresentadas ferramentas estatísticas para a análise de inferências, que avaliam a qualidade das informações do processo e que investigam a confiabilidade do modelo inferencial, permitindo ainda a discriminação entre modelos inferenciais quando outros modelos estão disponíveis. Os resultados obtidos em um problema modelo validam a utilização das ferramentas desenvolvidas. O baixo tempo de simulação encoraja o uso dessas ferramentas em linha.

Finalmente, são oferecidas algumas sugestões para trabalhos futuros. Para este estudo de caso, sugere-se trabalhar a estrutura do modelo inferencial empregado para a predição da temperatura a 85% do diesel produto. Se possível, deve-se adotar uma nova abordagem para modelagem do problema de inferências do diesel produto, empregando procedimentos estatísticos de redução de modelos. Não deixa de ser aconselhável um estudo aprofundado a respeito do processo, a fim de se obter um modelo mais preciso (fenomenológico), capaz de fornecer inferências mais precisas em tempo real. É sugerido ainda o uso de redes neuronais, caso um estudo mais aprofundado do processo não seja de interesse. É importante também avaliar cuidadosamente a estratégia de adição do *bias* ao modelo inferencial. Sugestões mais abrangentes compreendem:

- i. Adoção de uma estratégia de detecção de erros grosseiros que seja independente ao uso do modelo inferencial.
- ii. Adoção do procedimento de (re) estimação dos parâmetros em tempo real.
- iii. Adoção do procedimento de reconciliação de dados das variáveis operacionais em tempo real.
- iv. Inserção de testes de normalidade, para os dados operacionais.
- v. Definição das matrizes experimentais, para uma melhor avaliação dos erros experimentais e da influência destas variáveis nestes erros.
- vi. O estudo e implementação de ferramentas estatísticas para redução de modelos.

REFERÊNCIAS

- ALBERTOS, P., GOODWIN, G. C., 2002, “Virtual Sensors for Control Applications”, *Annual Reviews in Control*, v. 26, pp. 101-112.
- ALBUQUERQUE, J. S., BIEGLER, L. T., 1996, “Data reconciliation and gross-error detection for dynamic systems”, *AIChE Journal*, v. 42, pp. 2841-2856.
- AMY, L. T., 1992, *Automation Systems for Control and Data Acquisition*. 2 ed., North Carolina, ISA.
- ARAÚZO-BRAVO, M. J., CANO-IZQUIERDO, J. M., GÓMEZ-SÁNCHEZ, E., *et al.*, 2004, “Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems”, *Control Engineering Practice*, v. 12, pp. 1073-1090.
- ASSIS, A. J., MACIEL FILHO, R., 2000, “Soft sensors development for on-line bioreactor state estimation”, *Computers and Chemical Engineering*, v. 24, pp. 1099-1103.
- BARD, Y., 1974, *Nonlinear Parameter Estimation*. 1 ed., New York, Academic Press.
- BERTHOUEX, P. M., BOX, G. E., 1996, “Time Series Models for Forecasting Wastewater Treatment Plant Performance”, *Water Research*, v. 30, pp. 1865-1875.
- BOGAERTS, P., WOUWER, A. V. A. V., 2004, “Software sensors for bioprocesses”, *ISA Transactions*, v. 42, pp. 547-558.
- BOGAERTS, P., WOUWER, A. V. A. V., 2004, “Parameter identification for state estimation—application to bioprocess software sensors”, *Chemical Engineering Science*, v. 59, pp. 2465-2476.
- BOX, G. E.P., 1979, *Robustness in the strategy of scientific model building*, in *Robustness in Statistics*. New York, Academic Press: R.L. Launer and G.N. Wilkinson, Editors.

- BOX, G.E.P., HUNTER, W.G., HUNTER, J. S., 1978, *Statistics for experimenters*. 2 ed., New York, John Wiley & Sons.
- BOYER, S. A., 1999, *SCADA – Supervisory Control and Data Acquisition*. 2 ed., North Carolina, ISA – Instrumentation Systems and Automation Society.
- CHATTERJEE, T., SARAF, D. N., 2004, “On-line estimation of product properties for crude distillation units”, *Journal of Process Control*, v. 14, pp. 61–77.
- CHEN, X., GAO, F., CHEN, G., 2004, “A soft-sensor development for melt-flow-length measurement during injection mold filling”, *Materials Science and Engineering A*, v. 384, pp. 245–254.
- CHEN, J., ROMAGNOLI, J. A., 1998, “A strategy for simultaneous dynamic data reconciliation and outlier detection”, *Computers Chemical Engineering*, v. 22, pp. 559-562.
- CHOI, D., PARK, H., 2001, “A hybrid artificial neural network as a software sensor for optimal control of a wastewater treatment process”, *Water Research*, v. 35, pp. 3959–3967.
- CRAMER, H., 1946, *Mathematical Methods of Statistics*. New Jersey, Princeton University Press.
- DAI, X., WANG, W., DING, Y., SUN, Z., 2006, ““Assumed inherent sensor” inversion based ANN dynamic soft-sensing method and its application in erythromycin fermentation process”, *Computers and Chemical Engineering*, v. 30, pp. 1203–1225.
- DAVE, D. J., DABHIYA, M.Z.; SATYADEV, S.V.K.; GANGULY, S.; SARAF, D.N., 2003, “Online tuning of a steady state crude distillation model for real time applications”, *Journal Process Control*, v. 13, pp. 267–282.
- DESAI, K., BADHE, Y, TAMBE, S. S., KULKARNI, B. D., 2006, “Soft-sensor development for fed-batch bioreactors using support vector regression”, *Biochemical Engineering Journal*, v. 27, pp. 225–239.
- DUFOUR, P., BHARTIYA, S., DHURJATI, P. S., DOYLE, F. J., 2005, “Neural network-based software sensor: training set design and application to a continuous pulp digester”, *Control Engineering Practice*, v. 13, pp. 135–143.

- FACCO, P., DOPLICHER, F., BEZZO, F., BAROLO, M., 2008, “Moving average PLS soft sensor for online product quality estimation in an industrial batch polymerization process”, *Journal of Process Control*, v. 19, pp. 520–529.
- FEITKENHAUER, H., MEYER, U., 2004, “Software sensors based on titrimetric techniques for the monitoring and control of aerobic and anaerobic bioreactors”, *Biochemical Engineering Journal*, v. 17, pp. 147–151.
- FELDMAN, R. N., 2005. Reconciliação de Dados em Tempo Real para Monitoração e Detecção de Falhas em Terminal de Transporte e Armazenamento de Derivados de Petróleo. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- FORTUNA, L., GRAZIANIA, S., XIBILIA, M. G., 2005, “Soft sensors for product quality monitoring in debutanizer distillation columns”, *Control Engineering Practice*, v. 13, pp. 499-508.
- FORTUNA, L., RIZZO A., SINATRA, M., XIBILIA, M. G., 2003, “Soft analyzers for a sulfur recovery unit”, *Control Engineering Practice*, v. 11, pp. 1491-1500.
- FORTUNA, L., GRAZIANI, S., RIZZO, A., XIBILIA, M. G., 2007, *Soft Sensors for Monitoring and Control of Industrial Processes*. 1 ed. New York, Springer-Verlag.
- FREITAS, K. P., COSTA, A. O. S., COSTA Jr., E. F., 2006, “Descrição da deslignificação com oxigênio da celulose empregando redes neurais”, *In Proceedings: 9º Encontro de Modelagem Computacional*, Belo Horizonte, Novembro.
- FREITAS, K. P., PRATA, D. M., LIMA, E., PINTO, J. C., 2008, “Monitoramento em tempo real no processo de refino com base em soft sensors e análise de cluster”, *In Proceedings: 11º Encontro de Modelagem Computacional*, Volta Redonda, Novembro.
- FROZZA, E., LUZ Jr., L. F. L., 2005, “Elaboração de Modelo de Inferência para Controle de Unidade de Destilação Atmosférica de Petróleo”, *In: Proceedings of the 3º Congresso Brasileiro de P&D em Petróleo e Gás*, Salvador, Outubro.
- FRIEDMAN, Y. Z., 2005, “Comments on ‘‘On line estimation of product properties for crude distillation unit’’ by Tirtha Chatterjee and Deoki N. Saraf [Journal of Process Control 14 (2004) 61–77]”, *Journal of Process Control*, v. 15, pp. 847.

- GONZALEZ, G. D., 1999, "Soft Sensors for processing plants", *In: Proceedings 2nd International Conference on Intelligent Processing and Manufacturing of Materials – IPMM 99*, Hawaii, Julho.
- HIMMELBLAU, D. M., 1970, *Process analysis by statistical methods*. New York, John Wiley & Sons.
- JAIN, P., RAHMAN, I., KULKARNI, B. D., 2007, "Development of a soft sensor for a batch distillation column using support vector regression techniques", *Chemical Engineering Research and Design*, v. 85, pp. 283-287.
- JAMES, S. C., LEGGE, R. L., BUDMAN, H., 2008, "On-line estimation in bioreactors: A review", *Reviews in Chemical Engineering*, v. 16, pp. 311-340.
- JANA, A. K., SAMANTA, A. N., GANGULY, S., 2005, "Nonlinear model-based control algorithm for a distillation column using software sensor", *ISA Transactions*, v. 44, pp. 259-271.
- KANO, M., SHOWCHAIYA, N., HASEBE, S., HASHIMOTO, I., 2003, "Inferential Control of Distillation Compositions: Selection of Model and Control Configuration", *Control Engineering Practice*, v. 11, pp. 927-933.
- KOMIVES, C., PARKERY, R. S., 2003, "Bioreactor state estimation and control", *Current Opinion in Biotechnology*, v. 14, pp. 468-474.
- KOURTI, T., 2006, "The Process Analytical Technology initiative and multivariate process analysis, monitoring and control", *Analytical and Bioanalytical Chemistry*, v. 384, pp. 1043-1048.
- KRUSKAL, W., FERGUSON, T. S., TUKEY, J. W., GUMBEL, E. J., ANSCOMBE, F. J., 1960, "Discussion of the paper of Messrs. Anscombe and Daniel", *Technometrics*, v. 2, pp. 157-165.
- LIN, B., RECKE, B., KNUDSEN, J. K. H., JØRGENSEN, S. B., 2007, "A Systematic Approach for Soft Sensor Development", *Computers and Chemical Engineering*, v. 31, pp. 419-425.
- LIEBMAN, M. J., EDGAR, T. F. AND LASDON, L. S., 1992, "Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques", *Computers Chemical Engineering*, v. 16, pp. 963-986.

- LINKO, S., ZHU, Y., LINKO, P., 1999, “Applying neural networks as software sensors for enzyme engineering”, *Trends in Biotechnology*, v. 17b, pp. 155-162.
- LIU, J., 2007, “On-line soft sensor for polyethylene process with multiple production grades”, *Control Engineering Practice*, v. 15, pp. 769–778.
- MASSON, M. H., CANU, S., GRANDVALET, Y., LYNGGAARD-JENSEN, A., 1999, “Software sensor design based on empirical data”, *Ecological Modelling*, v. 12, pp. 131-139.
- MATSUMURA, S., IWAHARA, T., OGATA, K., FUJII, S., SUZUKI, M., 1998, “Improvement of de-NO_x device control performance using a software sensor”, *Control Engineering Practice*, v. 6, pp. 1267-1276.
- MCBRAYER, K. F. AND EDGAR, T. F., 1995, “Bias detection and estimation in dynamic data reconciliation”, *Journal Process of Control*, v. 5, pp. 285-289.
- MILETIC, I., QUINN, S., DUDZIC, M., VACULIK, V., CHAMPAGNE, M., 2004, “An Industrial Perspective on Implementing On-Line Applications of Multivariate Statistics”, *Journal of Process Control*, v. 14, pp. 821-836.
- MILLER, W.J., MCDONALD, J.F., PINSKER, W., 1997, “Molecular domestication of mobile elements”, *Genetica*, v. 100, pp. 261-270.
- MONTGOMERY, D. C., 2004, *Introdução ao controle estatístico da qualidade*. 4 ed. Rio de Janeiro, LTC.
- MORAD, K, SVRCEK, W. Y., MCKAY, I., 1999, “A robust direct approach for calculating measurement error covariance matrix”, *Computers and Chemical Engineering*, v. 23, pp. 889–897.
- NORONHA, F. B., PINTO, J. C., MONTEIRO, J. L., LOBÃO, M. W., SANTOS, T. J., 1993, “Um Pacote Computacional para Estimção de Parâmetros e Projeto de Experimentos”. *Technical Report*, PEQ/COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- PARK, S., HAN, C., 2000, “A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns”, *Computers and Chemical Engineering*, v. 24, pp. 871-877.
- PRATA, D. M., 2005. *Reconciliação de Dados em um Reator de Polimerização*. Tese

Dissertação de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

PRATA, D. M., FREITAS, K. P., LIMA, E. L., PINTO, J. C., 2008, “Análise de Clusters para Detecção de Erros Grosseiros e Reconciliação de Dados: Aplicação Industrial”, *In: Proceedings of the XVII Congresso Brasileiro de Engenharia Química– COBEQ 2008*, Recife, Setembro.

QUELHAS, A. D., PINTO, J. C., 2009, “Soft sensor models: Bias updating revisited”, *In: Proceedings of the International Symposium on Advanced Control of Chemical Processes 2009*, Istambul, Turquia.

RATKOWSKY, A. D., 1990, *Handbook of Nonlinear Regression Models*. New York, Marcel Dekker.

SCHLADT, M., HU, B., 2007, “Soft Sensors based on Nonlinear Steady-state Data Reconciliation in the Process Industry”, *Chemical Engineering and Processing*, v. 46, pp. 1107 – 1115.

SCHWAAB, M., 2007. *Desenvolvimento e implementação de novas técnicas de estimação de parâmetros e planejamento*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

SCHWAAB, M., SILVA, F. M., QUEIPO, C. A., BARRETO JR., A. G., NELE, M., PINTO, J. C., 2007, “A new approach for sequential experimental design for model discrimination”, *Chemical Engineering and Processing*, v. 61, pp. 5791 – 5806.

SCHWAAB, M., PINTO, J. C., 2007, *Análises de Dados Experimentais I - Fundamentos de Estatística e Estimação de Parâmetros*. 1 ed. Rio de Janeiro, e-papers.

SHARMIN, R., SUNDARARAJ, U., SHAH, S., GRIEND, L. V., SUN, Y., 2006, “Inferential Sensors for Estimation of Polymer Quality Parameters: Industrial Application of a PLS-based Soft Sensor for a LDPE Plant”, *Chemical Engineering Science*, v. 61, pp. 6372 – 6384.

SKEIE, N., MYLVAGANAM, S., LIE, B., 2006, “Using multi sensor data fusion for level estimation in a separator”, *In: Proceedings of the 16th European Symposium on Computer Aided Process Engineering and 9th International Symposium on Process Systems Engineering*, Garmisch-Partenkirchen, Julho.

- SOTOMAYOR, O. A. Z., PARK, S. W., GARCIA, C., 2002, "Software sensor for on line estimation of the microbial activity in activated sludge systems", *ISA Transactions*, v. 41, pp. 127-143.
- SOROUSH, M., 1998, "State and parameter estimations and their applications in process control", *Computers and Chemical Engineering*, v. 23, pp. 229-245.
- TALAIE, A., MIRHOSSEINI, A. R., ROMAGNOLI, J., TAGUCHI, T., 1998, "A new application of soft-computing theory to conducting polymer sensors: a pH estimation study", *Computational and Theoretical Polymer Science*, v. 18, pp. 291-295.
- THAM, M. T., MONTAGUE, G. A., MORRIS, A. J., LANT, P. A., 1991, "Soft-sensors for process estimation and inferential control", *Journal Process Control*, v. 1, pp. 3-14.
- TONG, H., CROWE, C.M., 1995, "Detection of Gross Errors in Data Reconciliation by Principal Component Analysis", *AIChE Journal*, v. 41, pp. 1712-1722.
- TONG, H., CROWE, C. M., 1996, "Detecting persistent gross errors by sequential analysis of principal components", *Computers & Chemical Engineering*, v. 20, pp. 733-738.
- YAN, W., SHAO, H., WANG, X., 2004, "Soft sensing modeling based on support vector machine and Bayesian model selection", *Computers and Chemical Engineering*, v. 28, pp. 1489-1498.
- YAN, X., 2008, "Modified nonlinear generalized ridge regression and its application to develop naphtha cut point soft sensor", *Computers and Chemical Engineering*, v. 32, pp. 608-621.
- YANG, Y., CHAI, T., 1997, "Application of statistical theory of adjustment to material balances", *In Proceedings of the American Control Conference*, Albuquerque, New México, Junho.
- ZAMPROGNA, E., BAROLO, M., SEBORG, D. E., 2001, "Composition estimations in a middle-vessel batch distillation column using artificial neural networks", *Chemical Engineering Research and Design*, v. 79, pp. 689-696.
- ZAMPROGNA, E., BAROLO, M., SEBORG, D. E., 2005, "Optimal Selection of Soft

Sensor Inputs for Batch Distillation Columns Using Principal Component Analysis”, *Journal of Process Control*, v. 15, pp. 39-52.

ZANATA, D. F. P., 2005. *Desenvolvimento de Sensor Virtual empregando Redes Neurais para Medição da Composição em uma Coluna de Destilação*. Tese de M.Sc., USP, São Paulo, SP, Brasil.

ZANIN, A. C., 1993, “Inferência da Destilação ASTM e do Ponto de Névoa do Diesel Linha da UC da RPBC”, *In Proceedings: IX Encontro Técnico de Destilação da PETROBRAS*, Curitiba , Novembro .

ZHANG, J., MARTIN, E. B., MORRIS, A. J., KIPARISSIDES, C., 1997, “Inferential Estimation of Polymer Quality Using Stacked Neural Networks”, *Computers Chemical Engineering*, v. 21, pp. 1025-1030.

ZHONG, W., YU, J., 2000, “MIMO Soft Sensors for Estimating Product Quality With On-Line Correction”, *Trans IChemE*, v. 78a, pp. 91-103.