



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
DEPARTAMENTO DE PROCESSOS QUÍMICOS

DESENVOLVIMENTO DE UM *SOFTWARE* PARA
RECONCILIAÇÃO DE DADOS DE PROCESSOS QUÍMICOS
E PETROQUÍMICOS

Dissertação de Mestrado apresentada à
Faculdade de Engenharia Química como parte
dos requisitos exigidos para a obtenção do título
de Mestre em Engenharia Química.

Autor: Agremis Guinho Barbosa

Orientador: Prof. Dr. Rubens Maciel Filho

Campinas – São Paulo

Novembro de 2003



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA QUÍMICA
DEPARTAMENTO DE PROCESSOS QUÍMICOS

DESENVOLVIMENTO DE UM *SOFTWARE* PARA
RECONCILIAÇÃO DE DADOS DE PROCESSOS QUÍMICOS
E PETROQUÍMICOS

Autor: Agremis Guinho Barbosa

Orientador: Prof. Dr. Rubens Maciel Filho

Campinas – São Paulo

Novembro de 2003

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

B234d Barbosa, Agremis Guinho
 Desenvolvimento de um software para reconciliação de
 dados de processos químicos e petroquímicos / Agremis
 Guinho Barbosa.--Campinas, SP: [s.n.], 2003.

 Orientador: Rubens Maciel Filho
 Dissertação (mestrado) - Universidade Estadual de
 Campinas, Faculdade de Engenharia Química.

 1. Estimativa de parâmetro. 2. Controle de processos
 químicos. 3. Redundância (Engenharia). 4. Localização de
 falhas (Engenharia). 5. Craqueamento catalítico. I. Maciel
 Filho, Rubens. II. Universidade Estadual de Campinas.
 Faculdade de Engenharia Química. III. Título.

Título em Inglês: Development of software for data reconciliation of chemical and
 petrochemical processes

Palavras-chave em Inglês: Data reconciliation, Gross error detection, Catalytic cracking

Área de concentração: Desenvolvimento de processos Químicos

Titulação: Mestre em Engenharia Química

Banca examinadora: Gustavo Paim Valença, Denis José Schiozer

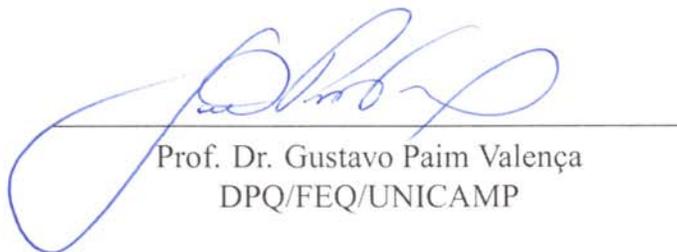
Data da defesa: 06/11/2003

Programa de Pós-Graduação: Engenharia Química

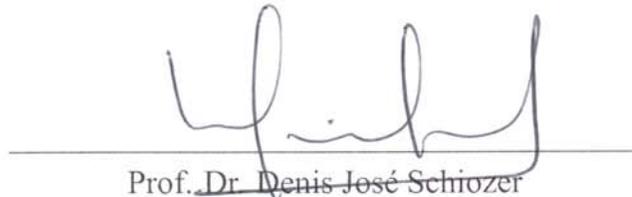
Dissertação de Mestrado defendida por Agremis Guinho Barbosa e aprovada em 6 de novembro de 2003 pela banca examinadora constituída pelos doutores:



Prof. Dr. Rubens Maciel Filho
Orientador



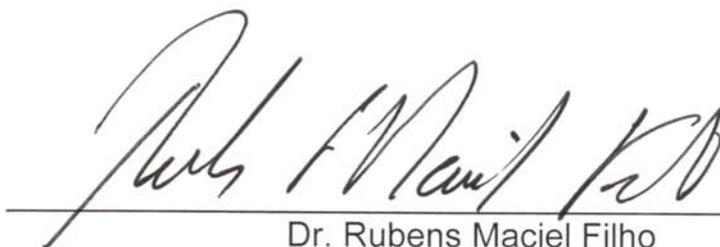
Prof. Dr. Gustavo Paim Valença
DPQ/FEQ/UNICAMP



Prof. Dr. Denis José Schiozer
DEP/FEM/UNICAMP

Esta página foi deixada intencionalmente em branco

Este exemplar corresponde à versão final da Dissertação de Mestrado em
Engenharia Química.

A handwritten signature in black ink, reading "Rubens Maciel Filho". The signature is written in a cursive style with a large initial 'R' and 'M'. Below the signature is a horizontal line.

Dr. Rubens Maciel Filho

Orientador

Esta página foi deixada intencionalmente em branco

Agradecimentos

À minha família: Lourdes, Ana e Luiza pelo apoio e por sua simples existência;

Ao Prof. Dr. Rubens Maciel pela orientação e pelo ambiente estimulante que proporciona;

À minha irmã de coração Pleycienne pela amizade e por toda ajuda que me dá;

À comunidade Shadu: Jefferson & Edvaldo pelas gargalhadas e fritura de neurônios;

Ao caro amigo Sebastião Coutinho pela companhia na caminhada da CIÊNCIA;

Ao caro amigo Mardonny pelas conversas sempre cultas e/ou divertidas;

Ao caro amigo URSO (também conhecido como Eduardo Coselli) outra fonte de conversas sempre agradáveis;

Aos queridos amigos do LOPCA pelo companheirismo;

Aos meus saudosos amigos da graduação pelo trabalho de equipe e por compartilhar esperanças;

A Donald Knuth e Leslie Lamport por criarem o $\text{T}_{\text{E}}\text{X}$ e o $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$, o que me permitiu manter o foco no trabalho (e não na editoração do texto);

Ao meu falecido pai ... É uma pena que não possa ver onde estou indo ...

À FAPESP pelo apoio financeiro e pelo trabalho de fomento à PESQUISA;

Resumo

O objetivo deste trabalho é o desenvolvimento de rotinas computacionais para o condicionamento de dados provenientes de um processo químico, de modo que estes sejam consistentes para a representação do comportamento do processo.

A descrição adequada do comportamento de um processo é a base fundamental de qualquer sistema de controle e/ou otimização, uma vez que será em resposta às medições deste processo (sua descrição) que os referidos sistemas atuarão. Desta forma o tratamento e correção dos erros de medição, especificamente, e a estimativa de parâmetros, de um modo mais geral, constituem uma etapa que não deve ser negligenciada no controle e otimização de processos.

O condicionamento de dados estudado neste trabalho é a **reconciliação de dados**, que tem como característica principal o uso de um modelo de restrições para condicionar a informação. Geralmente os modelos de restrição são balanços de massa e energia e os somatórios das frações mássicas e molares, mas outros modelos também podem ser usados.

Matematicamente, a reconciliação de dados é um problema de otimização sujeito a restrições. Neste trabalho, a formulação do problema de reconciliação é a dos mínimos quadrados ponderados sujeito a restrições e a abordagem para a sua solução é a fatoração QR.

Objetiva-se também reunir as rotinas desenvolvidas em uma única ferramenta computacional para a descrição, resolução e análise dos resultados do problema de reconciliação de dados, constituindo-se em um *software* de fácil utilização e que tenha ainda um mecanismo de comunicação com banco de dados, conferindo-lhe interatividade em tempo real com sistemas de aquisição de dados de processo.

Abstract

The purpose of this work is the development of computational routines for conditioning chemical process data in order to represent the process behavior as reliable as possible.

Reliable process description is fundamental for any control or optimization system development, since they respond to the process measurements (its description). Thus, data conditioning and correction of process measurement errors, and parameter estimation are a step that should not be neglected in process control and optimization.

The data conditioning considered in this work is **data reconciliation** which has as the main characteristic the use of a constraint model. In general constraint models are mass and energy balances and mass and molar fraction summation, but other models may be used.

Under a mathematical point of view, data reconciliation is an optimization subject to constraints. In this work, it is used the formulation of weighed least squares subject to constraints and QR factorization approach to solve the problem.

The additional objective of this work is to accommodate the developed routines in such a way to build up an integrated computational tool characterized by its easy to use structure, capability to solve and perform data reconciliation. Its structure takes into account the interaction with data bank, giving it real time interactiveness with process data acquisition systems.

Sumário

Lista de Figuras	p. xvii
Lista de Tabelas	p. xix
Nomenclatura	p. xxi
1 Introdução e Objetivos	p. 1
1.1 Introdução	p. 1
1.2 Organização da Dissertação	p. 2
2 Revisão Bibliográfica e Estado da Arte	p. 5
2.1 Introdução	p. 5
2.2 Conceitos Fundamentais	p. 5
2.2.1 Erros Aleatórios e Grosseiros	p. 5
2.2.2 Exatidão vs. Precisão	p. 7
2.2.3 Retificação, Reconciliação e Coaptação de Dados	p. 7
2.2.4 Benefícios da Reconciliação de Dados e da Detecção de Erros Grosseiros	p. 9
2.3 Reconciliação de Dados	p. 12
2.4 Detecção de Erros Grosseiros	p. 15

3	Reconciliação de Dados Linear em Estado Estacionário	p. 23
3.1	Introdução	p. 23
3.2	Conceitos Básicos	p. 24
3.3	Base Estatística da Reconciliação de Dados	p. 29
3.4	Formulação do Problema de Reconciliação de Dados	p. 31
3.5	Decomposição do Problema Geral de Estimativa	p. 35
3.6	Classificação das Variáveis de Processo	p. 38
3.6.1	Definições	p. 41
3.6.2	Análise da Topologia do Processo	p. 42
3.6.3	Abordagens para Solução do Problema de Classificação	p. 44
3.6.3.1	Técnicas Orientadas a Grafos	p. 45
3.6.3.2	Técnicas Orientadas a Equação	p. 46
3.7	Decomposição Usando Transformações Ortogonais	p. 49
3.7.1	Abordagem da Projeção de Matrizes	p. 49
3.7.2	Abordagem da Fatoração QR	p. 50
3.8	Reconciliação de Dados Linear com Todas as Variáveis Medidas	p. 53
3.8.1	Método dos Multiplicadores de Lagrange	p. 53
3.8.2	Método da Fatoração QR	p. 54
3.9	Reconciliação de Dados com Variáveis não Medidas	p. 56
3.10	Conclusão	p. 58
4	Detecção de Erros Grosseiros	p. 59
4.1	Introdução	p. 59

4.2	Conceitos Básicos	p. 59
4.3	Testes Estatísticos Básicos para a Detecção de Erros Grosseiros	p. 61
4.3.1	O Teste Global (<i>Global Test - GT</i>)	p. 64
4.3.2	Teste Nodal ou Teste da Restrição (<i>Nodal Test - NT</i>)	p. 66
4.3.3	Teste da Medida (<i>Measurement Test - MT</i>)	p. 68
4.3.4	Teste da Razão de Verossimilhança Generalizado (<i>Generalized Likelihood Ratio Test - GLR</i>)	p. 70
4.4	Testes Estatísticos para Modelos em Estado Estacionário	p. 76
4.5	Técnicas para Identificação de um Erro Grosseiro	p. 78
4.5.1	Estratégia de Eliminação Serial para Identificação de Erros Grosseiros Singulares	p. 79
4.6	Conclusão	p. 81
5	Programas Desenvolvidos	p. 83
5.1	Introdução	p. 83
5.2	Descrição do Problema	p. 84
5.2.1	Geração de Ruído	p. 88
5.3	Ciclo de Reconciliação	p. 88
5.3.1	Pré-condicionamento	p. 88
5.3.2	Reconciliação	p. 89
5.3.3	Detecção de Erros Grosseiros	p. 91
5.4	Apresentação e Análise dos Resultados	p. 92
5.5	Sistema de Testes	p. 92

5.5.1	Aplicação: Controle de um Reator CSTR não Isotérmico	p. 97
5.6	Conclusão	p. 100
6	Estudo de Caso: O Processo de Craqueamento Catalítico Fluido (FCC)	p. 101
6.1	Introdução	p. 101
6.2	Descrição	p. 102
6.3	Aplicação do Programa de Reconciliação ao FCC	p. 106
7	Conclusões	p. 109
7.1	Conclusões	p. 109
7.2	Sugestões para Trabalhos Futuros	p. 110
	Referências	p. 113
	Apêndice A – Ferramentas Utilizadas	p. 117
A.1	Introdução	p. 117
A.2	MATLAB	p. 117
A.3	Linguagens de Programação	p. 118
A.4	Linguagens de Programação Combinadas	p. 121

Lista de Figuras

2.1	Conjunto preciso e exato	p. 8
2.2	Conjunto preciso e inexato	p. 8
2.3	Conjunto impreciso e exato	p. 8
2.4	Conjunto impreciso e inexato	p. 8
3.1	Classificação das variáveis medidas	p. 41
3.2	Diagrama de fluxo para um sistema simples em série (ROMAGNOLI; STEPHANOPOULOS, 1980)	p. 43
3.3	Gráfico de fluxo de informação para o sistema descrito na Figura 3.2 (ROMAGNOLI; STEPHANOPOULOS, 1980)	p. 44
4.1	Sistema de troca de calor com <i>bypass</i> (NARASIMHAN; JORDACHE, 2000)	p. 64
5.1	Etapas do tratamento de dados	p. 84
5.2	Geração/disponibilização de dados simulados/medidos	p. 85
5.3	Seqüência de definição do problema de reconciliação	p. 86
5.4	Entrada de dados da corrente (nome da corrente, valor real e variância)	p. 87
5.5	Algoritmo do procedimento de reconciliação de dados	p. 89
5.6	Fluxograma da rotina de reconciliação de dados	p. 90
5.7	Saída de dados numéricos	p. 93
5.8	Abordagem de troca direta de dados	p. 94

5.9	Abordagem via banco de dados	p. 95
5.10	Esquematização do processo FCC sob o esquema de testes	p. 97
5.11	Interface gráfica para o simulador transiente de um CSTR não isotérmico	p. 99
5.12	Interface gráfica para controle e entrada de dados de um CSTR não isotérmico	p. 99
6.1	Esquema da seção reator/regenerador de uma unidade craqueamento catalítico fluido	p. 104
6.2	Esquema simplificado da seção reator/regenerador do FCC	p. 106
6.3	Representação do FCC (reator/regenerador)	p. 107
6.4	Corrente de ar	p. 108
6.5	Corrente de gasóleo	p. 108
6.6	Corrente de gases de combustão	p. 108
6.7	Corrente de produtos	p. 108
A.1	A interface de desenvolvimento integrado do Borland Delphi 6	p. 120
A.2	Object inspector	p. 121
A.3	<i>Form designer</i>	p. 122

Lista de Tabelas

- 4.1 Reconciliação de dados com a presença de erros grosseiros para o processo da Figura 4.1 (NARASIMHAN; JORDACHE, 2000) p. 65
- 4.2 Redução na estatística do teste global com a eliminação da medida i (NARASIMHAN; JORDACHE, 2000) p. 80
- 5.1 Parâmetros para simulação dinâmica de um CSTR não isotérmico (LUYBEN, 1989) p. 98
- 6.1 Propriedades das correntes de entrada e saída do FCC (reator/regenerador) p. 105

Nomenclatura

Letras Latinas

- A** Jacobiana de φ
- C** Jacobiana de ϕ
- g** Quantidade de valores em **x**
- H₀** Hipótese nula
- H₁** Hipótese alternativa
- I** Matriz identidade
- J** Função objetivo do problema de minimização de erros
- l** Quantidade de medidas em **y**
- L** Matriz de incidência
- P** Matriz de projeção de Crowe
- r** Grau de redundância
- r** Vetor dos resíduos
- \Re** Conjunto dos números reais
- w** Erro aditivo ao modelo de restrições adicionais
- x** Vetor dos valores reais das medidas
- \hat{x}** Vetor das estimativas dos valores reais
- y** Vetor das medidas com ruído
- \mathbf{z}_a** Matriz das estatísticas do teste da medida
- \mathbf{z}_r** Matriz das estatísticas do teste nodal

Letras Gregas

- α Nível de significância
- β Nível de significância modificado
- γ Estatística do teste global
- ε Erro aleatório
- σ Desvio padrão
- ϕ Modelo funcional da medição
- φ Modelo de restrições adicionais
- χ^2 Distribuição de probabilidade “chi-quadrado”
- Ψ Matriz de variância-covariância

Siglas

- CSTR *Continuous Stirred Tank Reactor* - Reator Tanque de Mistura Contínua
- DCS *Distributed Control System* - Sistema de Controle Distribuído
- DLL *Dynamic Link Library* - Biblioteca de Vínculo Dinâmico
- FCC *Fluid Catalytic Cracking* - Craqueamento Catalítico Fluido
- GLP Gás Liquefeito de Petróleo
- GLR *Generalized Likelihood Ratio* - Razão de Verossimilhança Generalizada
- GUI *Graphical User Interface* - Interface Gráfica com o Usuário
- MT *Measurement Test* - Teste da Medida
- NT *Nodal Test* - Teste Nodal
- RAD *Rapid Application Development* - Desenvolvimento Rápido de Aplicativos
- SGBDR Sistema Gerenciador de Banco de Dados Relacional
- SQP *Successive Quadratic Programming* - Programação Quadrática Sucessiva

1 Introdução e Objetivos

A good scientist is a person with original ideas. A good engineer is a person who makes a design that works with as few original ideas as possible. There are no prima donnas in engineering.

Freeman Dyson

1.1 Introdução

Nas últimas décadas, refinarias e outras indústrias de processos reconheceram a necessidade da melhoria da informação e do acesso mais rápido a ela. Para atingir esta meta, um grande investimento foi feito em DCS's (*Distributed Control System* – Sistema de Controle Distribuído), Bancos de Dados e infraestrutura de coleta de dados. O resultado de todo este investimento é que uma quantidade cada vez maior de dados têm sido disponibilizada com o acesso cada vez mais facilitado.

Infelizmente, quantidade de dados não é sinônimo de qualidade de informação. Devido aos erros inerentes à qualquer medida e a erros de origens diversas (falha humana, vazamentos, isolamento térmico ineficaz) uma grande massa de dados acaba se transformando em pouca informação ou em informação corrompida.

A grande quantidade de informação vinda de várias fontes resultou em inconsistências entre os dados, as quais resultam inevitavelmente em perda de confiança no sistema de medição. O mesmo conjunto de dados é, então, manipulado por diferentes grupos,

de diferentes maneiras, o que faz com que as mesmas medidas tenham uma variedade de ajustes, todos diferentes entre si.

Desta forma, se faz necessária uma metodologia padrão a ser aplicada sobre as fontes de dados, para que estes se convertam em informação. Médias, filtros simples e ajustes baseados na experiência são alguns dos possíveis métodos de condicionamento de dados, mas nenhum deles considera os balanços de massa e energia e podem levar o conjunto dos dados a uma inconsistência global.

A reconciliação de dados consiste em um tratamento matemático objetivando a melhoria da qualidade dos dados que representam processos experimentais ou industriais (PLÁCIDO, 1995) e é uma ferramenta relativamente nova e ainda pouco abordada e conhecida, mesmo no meio acadêmico. Os resultados do seu emprego podem ser bastante expressivos, principalmente na otimização de processos, mas também no diagnóstico de falhas em uma planta.

As técnicas de reconciliação de dados e **detecção de erros grosseiros** são eminentemente estatísticas mas fazem uso dos princípios de conservação de massa e energia como modelos de restrição. Neste trabalho é abordada a **reconciliação de dados linear em estado estacionário** que tem o seu principal emprego na melhoria da qualidade de informação sobre vazões mássicas de um processo.

1.2 Organização da Dissertação

No Capítulo 2 são introduzidos os conceitos fundamentais e a terminologia básica para logo em seguida se apresentar a revisão bibliográfica separada entre os dois temas principais: a reconciliação de dados e a detecção de erros grosseiros.

O Capítulo 3 descreve a fundamentação matemática do problema de reconciliação de dados, bem como discorre sobre abordagens para a sua solução.

O Capítulo 4 é dedicado à detecção de erros grosseiros, descrevendo os principais testes estatísticos usados na detecção deste tipo de erro.

No Capítulo 5 está a implementação prática de toda a teoria vista nos capítulos anteriores, sendo o Capítulo 6 a sua aplicação ao caso estudo, a Unidade de Craqueamento Catalítico Fluido (FCC - *Fluid Catalytic Cracking*).

No Capítulo 7 as conclusões deste trabalho são discutidas e são dadas sugestões para trabalhos futuros e no Apêndice A são mostradas e discutidas as principais ferramentas computacionais que foram usadas no desenvolvimento deste trabalho.

2 Revisão Bibliográfica e Estado da Arte

There is no substitute for data

John Tukey

2.1 Introdução

Neste capítulo faz-se uma exposição dos termos mais freqüentes na reconciliação de dados e na detecção de erros grosseiros, apresentando-se alguns conceitos importantes para a compreensão deste trabalho. Na seqüência é feita uma revisão bibliográfica sobre reconciliação de dados e detecção de erros grosseiros.

2.2 Conceitos Fundamentais

2.2.1 Erros Aleatórios e Grosseiros

Quando se descreve um processo, têm-se dois problemas principais: fazer com que as flutuações inerentes à dinâmica e os erros de diferentes naturezas não afastem o conjunto dos dados das restrições (balanços de massa, balanços de energia, somatório de frações molares) e, ainda, acessar dados que não são diretamente medidos. A eliminação dos erros aleatórios e o ajuste às restrições (reconciliação de dados) são feitos sobre

um conjunto redundante de dados e são normalmente precedidos pela detecção de erros grosseiros (Retificação de Dados).

Antes de se iniciar o detalhamento dos dois tipos de erro, é importante ressaltar que, segundo Crowe (1996), qualquer conjunto de variáveis de um processo que não obedece às leis de conservação (balanços de massa, de energia e somatórios de frações mássicas e molares) é considerado como contendo erros.

O que não é tão evidente é que o erro (aleatório ou grosseiro) pode advir de uma leitura errada, ou seja, de um problema no equipamento de medição ou pode ser fruto de flutuações naturais nos dados do processo, oriundas de diversas causas, que, mesmo sendo os dados corretamente medidos, levam à fuga dos referidos balanços. Esta análise primária é útil, pois a partir dela pode-se utilizar os procedimentos de reconciliação de dados para identificar problemas nos equipamentos de medição.

No tocante aos aspectos estocásticos dos erros, sabe-se que os erros aleatórios (também chamados de randômicos) são comuns a todo processo real de obtenção de dados, constituindo pequenos desvios em relação a um valor central. A concepção de regime permanente, rigorosamente falando, é extremamente útil na compreensão e modelagem matemática de diversos fenômenos e equipamentos na indústria de processos, mas em geral não corresponde à realidade por mais que um átimo de tempo, haja vista as perturbações naturais ou provocadas de diversas origens e tipos, como as ações de controle, impurezas, vibrações nas bombas, etc. Assim, uma avaliação das medidas disponíveis de um processo com base na precisão nominal dos medidores é incompleta quando não contabiliza também a “vibração” do processo. (PLÁCIDO, 1995).

Ao contrário dos erros aleatórios, normalmente distribuídos, com média zero, uma variância conhecida e de pequena magnitude, os erros grosseiros são associados a eventos não aleatórios como por exemplo:

- descalibração ou calibração deficiente dos instrumentos de medição que provocam desvios;
- vazamento insuspeito nas unidades do processo;

- mal-funcionamento ou falha nos instrumentos;
- regime transiente ou flutuações naturais do processo;
- flutuações nas condições ambientais;
- erros de amostragem e de análise;
- leituras feitas por diferentes observadores;
- dados levantados em tempos não-simultâneos, pertencentes a regimes permanentes distintos, etc. (PLÁCIDO, 1995).

2.2.2 Exatidão vs. Precisão

Um conjunto de medidas precisas não é necessariamente um conjunto de medidas exatas. A precisão está associada à densidade em torno de um valor médio com que as medidas se distribuem. Por outro lado, uma medida será tão mais exata, ou acurada, quanto mais se aproximar do valor real. Estes conceitos são ilustrados com o clássico exemplo do tiro ao alvo nas Figuras 2.1 até 2.4. Neste trabalho, usa-se as palavras exatidão e acurácia de forma indistinta.

2.2.3 Retificação, Reconciliação e Cooptação de Dados

A grande maioria dos procedimentos de reconciliação de dados está baseada na suposição de que as medidas estão sujeitas somente a erros aleatórios. Esta suposição não é normalmente confirmada pela realidade e as medidas dos processos estão também sujeitas a *biases* ou viéses¹ para fora da distribuição normal.

Existem ainda enfoques alternativos nos quais não seria necessário remover as variáveis portadoras de erros grosseiros para o procedimento de reconciliação, como o apresentado por Romagnoli (1983). Neste caso, o próprio modelo é modificado de

¹O termo *bias* é amplamente usado na literatura sobre reconciliação e retificação de dados e significa “tendência”. Neste trabalho, no lugar de se usar *bias*, optou-se pela palavra “viés”

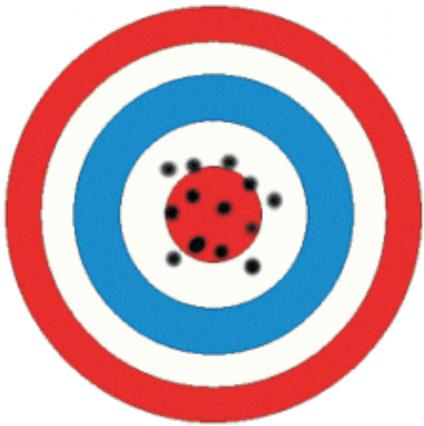


Figura 2.1: Conjunto preciso e exato

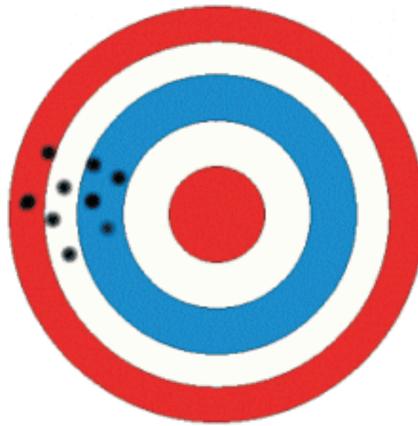


Figura 2.2: Conjunto preciso e inexato



Figura 2.3: Conjunto impreciso e exato

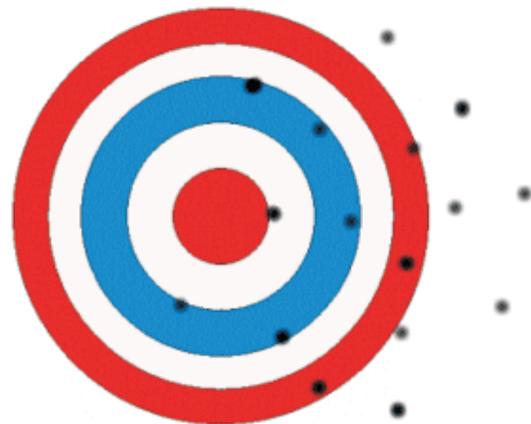


Figura 2.4: Conjunto impreciso e inexato

modo a considerar a presença de erros sistemáticos, acrescentando-se as magnitudes dos desvios em relação às médias, desconhecidas, como incógnitas a serem calculadas. Ambos os enfoques seriam equivalentes, já que levam a sistemas de equações que têm o mesmo grau de liberdade.

Em geral, a possibilidade de falha na detecção de erros grosseiros aumenta com a quantidade desses erros presentes.

O ajuste das medidas para compensar os erros aleatórios envolve a resolução de um problema de minimização sujeito a restrições, comumente do tipo de mínimos quadra-

dos sujeito a restrições. As equações de balanço são incluídas nas restrições e estas podem ser lineares mas são geralmente não lineares. A função objetivo é geralmente quadrática com relação ao ajuste das medidas e tem a **matriz da variância-covariância** dos erros das medidas como peso. Assim, essa matriz é essencial na obtenção de conhecimento confiável do processo.

A estimativa de parâmetros é também uma atividade importante no projeto, avaliação e controle de processos. Devido aos dados não satisfazerem às restrições do processo, o **método dos erros nas variáveis** provê tanto estimativas de parâmetros quanto estimativas de dados reconciliados que são consistentes com respeito ao modelo. Estes problemas representam uma classe especial de problemas de otimização porque a estrutura dos mínimos quadrados pode ser explorada no desenvolvimento de métodos de otimização. Uma revisão deste assunto pode ser encontrada no trabalho de Biegler *et al.* (1986).

As restrições podem ser usadas para estimar correntes ausentes ou que tenham sido descartadas: este uso das restrições é definido como **cooptação de dados** na literatura. Contudo, este procedimento propaga erros e deve ser realizado com cuidado.

Finalmente, algumas abordagens estão emergindo de dentro do problema de reconciliação de dados, tais como abordagens Bayesianas e técnicas robustas de estimativa, bem como estratégias que usam Análise de Componentes Principais. Elas oferecem alternativas viáveis aos métodos tradicionais e provêem novos terrenos para posterior aperfeiçoamento (CONSUL, 2002).

2.2.4 Benefícios da Reconciliação de Dados e da Detecção de Erros Grosseiros

Segundo Narasimhan e Jordache (2000), o desenvolvimento de um programa para reconciliação de dados e detecção de erros grosseiros para um sistema e sua implementação prática é uma tarefa complexa. A justificativa para o uso destas técnicas pode vir de várias aplicações importantes para melhora de desempenho, que requerem dados

acurados para alcançarem os benefícios esperados.

1. Uma aplicação direta da reconciliação de dados está na avaliação de produtos do processo ou na inferência do consumo de utilidades em diferentes unidades do processo. Valores reconciliados provêm estimativas mais acuradas quando comparadas com o uso de medições não tratadas. Por exemplo, a reconciliação de balanço material na escala completa de uma refinaria ajuda numa melhor estimativa da produção global da mesma. De uma maneira similar, uma auditoria dos balanços energéticos usando fluxos e temperaturas reconciliados ajuda numa melhor identificação de processos e equipamentos energeticamente ineficientes.
2. Aplicações como simulação e otimização de um equipamento de processo existente necessitam do modelo do equipamento com detalhes suficientes para sua adequada representação. Estes modelos comumente contêm parâmetros, os quais têm que ser estimados a partir de dados da planta. Isto é conhecido como *sintonia do modelo* ou *identificação do modelo*, para o qual dados acurados são essenciais. O uso de medidas errôneas na sintonia do modelo pode dar lugar a parâmetros incorretos que, eventualmente, anulem os benefícios alcançáveis através da otimização. Existem duas possibilidades de utilização da reconciliação de dados em tais aplicações e estas são ilustradas a seguir, tomando-se como exemplo uma coluna de destilação:
 - Considerando-se o problema de otimização do desempenho de uma coluna de destilação existente, pode-se obter dados operacionais, medidas de fluxo, temperatura e composição de todas as correntes de saída e entrada. Um caminho possível é reconciliar estas medidas usando somente balanços globais de massa e energia em torno da coluna. Estes dados reconciliados podem ser usados em conjunto com um modelo prato-a-prato detalhado da coluna, no sentido de se estimar parâmetros tais como eficiência de prato. O modelo sintonizado pode então ser usado para otimizar o desempenho da coluna.
 - Em uma outra abordagem, pode ser feita uma reconciliação de dados si-

- multânea com uma estimativa de parâmetros usando-se o modelo detalhado prato-a-prato da coluna. Neste caso, se as medidas de temperatura do prato e/ou composições estão disponíveis, estas também podem ser usadas e reconciliadas como parte do problema. Obviamente, a segunda abordagem leva a um significativo aumento de esforço e de tempo computacional. Esta última abordagem é também referida como modelagem rigorosa *on line* e vem sendo incorporada em vários simuladores em estado estacionário comerciais.
3. A reconciliação de dados pode ser muito útil no planejamento de manutenção de equipamento de processo. Dados reconciliados podem ser usados para estimar acuradamente o desempenho de parâmetros-chave dos equipamentos. Por exemplo, o coeficiente de troca térmica dos trocadores de calor ou o nível de atividade do catalisador nos reatores pode ser estimado e usado para determinar se o trocador deve ser limpo ou se o catalisador deve ser repostado/regenerado, respectivamente.
 4. Várias estratégias de controle avançado tais como controle baseado em modelo ou controle inferencial requerem estimativas acuradas das variáveis controladas. As técnicas de reconciliação de dados dinâmica podem ser usadas para derivar estas estimativas para um melhor controle de processos.
 5. A detecção de erros grosseiros não somente melhora a acurácia das estimativas dos procedimentos de reconciliação, como também é útil na identificação de problemas de instrumentação, os quais requerem manutenção especial e correção. Uma incipiente detecção de erros grosseiros pode reduzir os custos de manutenção e promover a operação da planta de uma maneira mais suave. Estes métodos também podem ser estendidos para detectar equipamentos defeituosos.

2.3 Reconciliação de Dados

Knepper e Gorman (1980) apresentaram as equações para o ajuste de dados e estimativa de parâmetros desenvolvidas a partir da teoria das inversas generalizadas. Elas se destinam à análise de conjuntos de dados sujeitos a restrições e ao planejamento de medidas experimentais, aplicando-se a casos lineares e não lineares. Examinaram a convergência de problemas não-lineares, para os quais sugerem o uso de um algoritmo computacional. Desenvolveram a matriz variância-covariância dos parâmetros calculados, a matriz dos dados ajustados e a covariância entre elas, que são usadas na decisão de validade de medidas experimentais adicionais para aumentar a precisão. Descrevem um teste χ^2 multivariado que permite verificar discrepâncias nas restrições inconsistentes com a matriz de variância-covariância dos dados medidos. Descrevem ainda outros testes estatísticos para que as estimativas sejam livres de erros grosseiros.

Romagnoli e Stephanopoulos (1980) estudaram a retificação de dados medidos em plantas químicas complexas que operam em estado estacionário e a obtenção de estimativas de variáveis não-medidas. Propõem o uso de um algoritmo para classificar as correntes medidas e as não-medidas, baseado na **matriz de ocorrência**², a fim de reduzir o tamanho do problema. Apresentam, também, diferentes algoritmos referentes à solução de problemas práticos especiais, como seleção de parâmetros corrigíveis, seleção de medidas necessárias para que um processo e também certos parâmetros sejam determináveis. Propõem o uso de testes estocásticos para verificar a consistência de dados do processo e detectar erros grosseiros. Este estudo se aplica tanto ao caso de restrições lineares quanto para não-lineares, considerando-se a situação na qual a composição de certas correntes é medida para todos os componentes.

Romagnoli e Stephanopoulos (1981) estudaram a retificação de dados de processo medidos na presença de erros grosseiros. Apresentaram um método para identificar a fonte de erros através da eliminação seriada de uma ou mais medidas de cada vez. Utilizaram o método dos mínimos quadrados para a obtenção dos ajustes, tendo como res-

²A matriz de ocorrência ou matriz de incidência é a matriz que carrega a informação da topologia do processo, descrevendo matematicamente quais correntes têm contato com quais unidades.

trições os balanços de massa nodais. No método desenvolvido, faz-se o processamento sequencial das equações de balanço, argumentando que isto simplifica a identificação de erros grosseiros durante o processo de reconciliação e facilita o tratamento de dados de processo de grandes dimensões, pois fornece um critério simples individualizado para cada balanço, ao contrário do teste global. O critério é baseado na distribuição χ^2 dos resíduos de balanço. Utilizam análise estrutural para identificar um conjunto reduzido de medidas suspeitas de erro. Este estudo é restrito ao caso de balanços lineares, que, na prática, são encontrados somente em balanços globais mássicos, mas que podem ser aplicados a balanços de componentes e de energia linearizáveis.

Mah (1981) apresenta uma revisão sobre estudos em diversas áreas relacionadas ao ajuste de dados de processo, envolvendo a reconciliação de dados, a classificação de variáveis, a detecção e identificação de erros grosseiros e a definição de pontos de tomada de medidas que formam uma base para o projeto e análise de sistemas de monitoramento. Observa que, inicialmente, dava-se ênfase aos procedimentos de cálculo para a reconciliação, que deu lugar à questão de classificação de variáveis medidas e à detecção e identificação de erros grosseiros. Argumenta que um trabalho posterior é necessário para avaliar e generalizar os algoritmos desenvolvidos e para comparar os ganhos específicos em diferentes aspectos de sistemas de monitoramento do desempenho para que tenham uso industrial direto.

Stanley e Mah (1981a) apresentam dois algoritmos para determinação da observabilidade local e global, redundância de variáveis individuais e medições de processos baseadas em técnicas de análise teórico-gráfica. O tratamento considera somente restrições de conservação de massa e de energia, porém observam que pode ser estendido a processos com outras restrições adicionais, como frações de partição especificadas.

Stanley e Mah (1981b) desenvolveram uma teoria sobre observabilidade e redundância aplicável a sistemas no estado estacionário, através de analogias com sistemas dinâmicos. Demonstram a importância desses conceitos na predição do desempenho qualitativo de estimadores, como os mínimos quadrados condicionados e outros. Mostram como a observabilidade local está diretamente ligada à unicidade local da solução

da equação de medida e como a não observabilidade local provoca a falha do estimador tanto no caso do estado estacionário quanto no quase-estacionário. Para restrições e medições lineares, as condições para observabilidade local também valem para a observabilidade global e o sistema é decomponível em subsistemas (redundante, estritamente observável e não observável).

Keller (1992) propõe um algoritmo analítico para estimar a matriz de covariância dos erros de medida, que é empregada na reconciliação de dados de processo. Ele se baseia nos resíduos de restrições lineares, calculados a partir dos dados disponíveis. Observa que uma exigência indispensável para a sua aplicação é a redundância espacial suficiente e que este algoritmo pode ser aplicado a medidas correlacionadas se todos os sensores sujeitos a tais medidas forem conhecidos.

Bagajewicz (1996) aborda o problema da distribuição de probabilidade dos erros na reconciliação de dados. Na quase totalidade dos trabalhos encontrados na literatura disponível, a formulação do problema de reconciliação recai na suposição da distribuição normal dos erros. O autor aponta que esta suposição é verdadeira para dispositivos de leitura que efetuem somente transformações lineares sobre as medidas. No caso de leituras como as de vazão que implicam em transformações não lineares sobre uma medida primária, mesmo que esta tenha uma distribuição normal, a leitura final a distorcerá e a formulação “clássica” torna-se ineficaz para captar a natureza dos erros nos procedimentos de reconciliação de dados.

Dempf e List (1998) procederam à reconciliação de dados com objetivos de otimização de processos em uma planta de acetato de vinila e outra de acetona. Os autores afirmam que o “ótimo” empírico está na maioria dos casos à 10% do ótimo global e que a abordagem do problema de reconciliação deve sempre partir de uma modelagem simples do balanço global de massa de toda a planta, com a gradual introdução de complexidade acompanhada de redução de escopo. Relatam, ainda, os seguintes tópicos entre os resultados alcançados em sua aplicação prática:

- A RD compõe uma metodologia bastante sensível para detectar partidas do estado

estacionário;

- Suporte à manutenção preventiva, detectando vazamentos e/ou disfunção dos sensores ou das operações unitárias;
- Alta credibilidade nas provas de conformidade com regras ambientais, de segurança e restrições econômicas (padrões de certificação);
- Várias análises laboratoriais puderam ser substituídas por medidas de temperatura e a taxa de amostragem foi reduzida em 50%;
- A eficiência da unidade de craqueamento da acetona pôde ser calculada.

Bagajewicz e Jiang (2000) compararam o desempenho entre a abordagem integral para a reconciliação de dados dinâmica e a reconciliação de dados estacionária. Foi mostrado que na ausência de vieses e vazamentos, o desempenho em ambas as abordagens é similar. Foi provado que uma vez que a variância apropriada seja escolhida, ambos os métodos são idênticos na ausência de termos de acúmulo. Finalmente, foi feita uma análise das discrepâncias como função do acúmulo.

2.4 Detecção de Erros Grosseiros

Vários autores relatam formas de tratar erros grosseiros que envolvem a sua detecção e identificação. A maior parte dos critérios propostos se baseia em testes estatísticos que se aplicam ao conjunto de dados como um todo ou a cada nó (testando o desbalanço nodal correspondente) ou ainda ao ajuste de cada medida. Às vezes são usados em conjunto, utilizando-se mais de uma estatística em um mesmo problema de retificação de dados. Alguns testes são aplicados após a reconciliação de dados, que é seguida da remoção da medida suspeita, refazendo-se o procedimento de ajuste de dados. Outros procedimentos são efetuados antes de realizar a reconciliação de dados, o que é possível de acordo com a estatística empregada.

Romagnoli e Stephanopoulos (1981) utilizaram uma função-teste dada pelo quadrado do vetor dos resíduos do balanço ponderado pela variância desses resíduos, para verificar a presença de erros grosseiros. Partem do pressuposto de que os erros de medida são normalmente distribuídos, com valor esperado igual a zero. Então, os resíduos das equações de balanço também são distribuídos da mesma forma e têm valor esperado igual a zero. Acrescentam que os erros sistemáticos causam o deslocamento da medida numa direção e as observações são distribuídas ao redor de um certo valor, diferente do valor verdadeiro do parâmetro. A diferença entre esses valores é o erro sistemático. A quantidade medida, neste caso, é o vetor dos resíduos de balanço, cujo valor teórico é conhecido e igual a zero. Considera-se que a incerteza do erro de medida é conhecida e então verifica-se se a média dos resíduos calculada difere significativamente de zero, de modo que a hipótese de média zero não se verifique. Como o vetor dos resíduos tem n elementos (n equações), a função-teste terá uma distribuição χ^2 com n graus de liberdade, que deverá ser testada contra o valor tabelado correspondente, uma vez que seja pré-estabelecida a probabilidade de erro admissível, sendo aceitável em muitos casos o valor de 0,10. Sugerem que a identificação de erros grosseiros seja feita através de análise estrutural, cujo procedimento é descrito. Propõem que, alternativamente, seja adotado o processamento seqüencial das equações de balanço. Ele inicia com a aplicação do teste a uma equação de balanço, seguida da remoção (e teste) de uma medida de cada vez, quando o teste estatístico revelar a presença de erros grosseiros, com exceção daquelas medidas que foram consideradas sem erros grosseiros nas etapas anteriores. A seguir, vão sendo adicionadas as equações de balanço, uma a uma, à função-teste, aumentando da mesma forma o número de graus de liberdade da distribuição χ^2 tabelada.

Tamhane (1982), estudando o uso de resíduos (ajustes individuais das medidas) para detectar valores extremos em regressões lineares, mostra que um teste baseado no vetor transformado dos resíduos tem, na classe das transformações lineares do vetor dos resíduos, certas propriedades ótimas para revelar a presença de um único valor extremo, quando o observador não está ciente de que há exatamente um desses valores. Observa que é mais poderoso do que outros testes baseados no uso de resíduos, neste caso específico.

Mah e Tamhane (1982) tratam o problema de detecção e identificação da presença de um ou mais erros grosseiros em dados de processo, aplicado ao caso em que a reconciliação de dados é sujeita a restrições lineares. Consideram a situação em que todas as variáveis são medidas e que os erros grosseiros estão antes associados às medições do que a erros no modelo do processo. Observam que um teste simples, baseado nos ajustes ou resíduos (diferença entre o valor medido e o valor estimado) de cada variável, pode ser usado para detectar e identificar diretamente a fonte de erros grosseiros. Descrevem ainda um outro teste de medida baseado no vetor dos resíduos transformado (obtido pelo produto do inverso da matriz variância-covariância pelo vetor resíduo), que conforme Tamhane (1982) possui máxima potência para detectar a presença de uma única medida com erro grosseiro. Este teste pode ser estendido ao caso de erros múltiplos, sem contudo ter assegurada suas propriedades de máxima potência.

Crowe *et al.* (1983) tratam da verificação e identificação de erros grosseiros na reconciliação de vazões, em balanços multicomponentes, no estado estacionário. Consideram que todas as medidas de concentração e vazão são independentes e normalmente distribuídas, com média desconhecida e variância conhecida. Observam que embora o produto de duas variáveis normais independentes não seja normalmente distribuído, este produto tem uma distribuição aproximadamente normal se as duas variáveis tiverem somente valores positivos e se seus coeficientes de variação (desvios padrão relativos) forem suficientemente pequenos, isto é, menores que 5%. Então, consideram que o vetor das medidas é uma amostra de uma distribuição normal multivariada com média desconhecida e matriz variância-covariância conhecida. Observam que medidas de diferentes correntes são, usualmente, tomadas como estatisticamente independentes, assim como medidas de concentração e de vazão total numa mesma corrente. Entretanto, as vazões dos componentes de uma mesma corrente não são independentes, pois a vazão total é comum. Utilizam três estatísticas-teste, sendo uma delas baseada na distribuição χ^2 , a qual é usada para detectar erros grosseiros, coletivamente, em um conjunto de combinações lineares dos erros nas equações do balanço reduzido. Observam que qualquer combinação linear desses erros pode ser testada contra uma distribuição normal e utilizam uma que testa o desbalanço particular em cada nó e outra que testa o ajuste

particular de cada medida. Aconselham que estas três estatísticas sejam empregadas em conjunto, para obter tantas evidências quanto possível e observam que, em um dos casos estudados, um dos testes não revelou a presença de erros grosseiros, enquanto os outros dois revelaram.

Madron (1985), estudando o problema de identificação de erros grosseiros em medidas, propõe uma abordagem que, ao complementar outras existentes, permite reduzir o número de elementos do conjunto de medidas suspeitas de portar erros, que deve ser testada na fase final da eliminação desse tipo de erro. Para isto, faz uso de informações disponíveis, a priori, sobre a credibilidade da medida, que representa o valor máximo possível do erro grosseiro de cada medida. Sugere o uso do teste global χ^2 para detectar erros grosseiros e de três métodos de identificação da fonte de erros descritos na literatura, os quais, restringindo-se ao caso de desvios de um modelo de erros pressuposto e não contemplando erros devidos à representação inadequada do modelo matemático (como vazamentos, por exemplo), considerando também a presença de um único erro grosseiro.

Jordache *et al.* (1985) tratam da avaliação do desempenho do teste de medida para detecção de erros grosseiros em dados de processo, proposto por Mah e Tamhane (1982), através da avaliação de sua potência, que é a probabilidade de detectar e identificar, corretamente a presença de erros grosseiros. Consideram que somente um erro grosseiro está presente nas medidas e exploram a influência de diferentes parâmetros na potência do teste. Observam que o efeito das restrições, da configuração do sistema e da posição da medida com erro grosseiro podem ser adequadamente considerados através da matriz transformada de restrições e que aqueles fatores que tornam as colunas (correntes) dessa matriz mais proporcionais tendem a reduzir as potências associadas a estas colunas e torná-las umas iguais às outras, enquanto fatores que as tornam menos proporcionais tendem a aumentar suas potências. Utilizam duas definições para a potência e observam que ela é fortemente afetada pela razão entre magnitude do erro grosseiro e o desvio-padrão. Além disso, as potências também dependem da faixa e da distribuição dos desvios-padrão, quando eles são diferentes. A razão entre os desvios máximo e

mínimo é uma medida do espalhamento das potências e quanto maior esta razão, maior a faixa das potências. Para uma mesma relação magnitude do erro grosseiro/desvio-padrão, quanto menor a magnitude do erro, menor a potência associada àquela medida.

Narasimham e Mah (1987) descrevem um método para detectar e identificar erros grosseiros devidos a desvios instrumentais e vazamentos em processos químicos no estado estacionário. Ele pode ser estendido a qualquer tipo de erro que possa ser modelado matematicamente. Como sub-produto, ele permite a obtenção de estimativa da magnitude do erro grosseiro. Este método é chamado de método da **razão de verossimilhança generalizada** (GLR - *Generalized Likelihood Ratio*) e se baseia no teste estatístico da razão da probabilidade condicional. Permite a identificação de vários erros grosseiros, para o quê utiliza uma estratégia baseada na compensação seriada de erros grosseiros. Esta forma de compensação necessita menor tempo de cálculo do que outras existentes, baseadas na eliminação seriada. Observam que o método GLR apresenta vantagens sobre outros métodos para identificação de erros grosseiros em processos químicos no estado estacionário, pois somente ele possibilita diferenciar os vários tipos de erros. Comentam que isto é importante, pois nem sempre se pode estar certo de que o processo opera estritamente no estado estacionário. Para avaliar o desempenho do teste relatam experiências de simulação, utilizando duas medidas de desempenho. Observam que quando somente erros grosseiros devidos a desvios instrumentais estão presentes, o teste GLR se reduz ao teste da medida e que ambos têm o mesmo desempenho, em condições idênticas.

Narasimham e Mah (1989) observam que a maioria dos testes estatísticos para identificar erros grosseiros, propostos até então, trata de modelos simples de processos, no estado estacionário. Neles, as variáveis são consideradas diretamente medidas ou não-medidas. Propõem técnicas para transformar adequadamente modelos gerais, no estado estacionário, em modelos simples nas mesmas condições de operação. Consideram os casos em que as variáveis de processo são medidas indiretamente e onde variáveis não medidas estão presentes nas restrições. Numa situação em que ambos os casos estejam presentes, o modelo geral do processo é transformado no modelo simples em duas

etapas. Inicialmente, eliminam-se as variáveis não medidas e logo após são tratadas as variáveis indiretamente medidas. Assim é possível aplicar qualquer um dos quatro testes estatísticos seguintes, previamente desenvolvidos: o teste global, o teste dos resíduos das restrições, o teste de medida e o teste da razão de probabilidade condicional generalizada.

Kao *et al.* (1990) observam que uma das maiores desvantagens da maioria dos métodos propostos na literatura para detecção de erros grosseiros é que eles partem do pressuposto de que as observações são não correlacionadas seriadamente (já que a independência implica em ausência de correlação), o que muitas vezes não é confirmado pelas evidências experimentais. Relatam que observações correlacionadas podem ser causadas por um número de fatores físicos associados com o tempo morto do processo, com sua dinâmica e controle, bem como devido a fatores relacionados a instrumentos de medição. Nesta investigação, tratam de erros de medida correlacionados seriadamente, em processos químicos no estado estacionário. Entretanto, consideram que variáveis diferentes medidas num mesmo tempo ou em diferentes tempos não são correlacionadas, isto é, que não há correlação cruzada. Usam como referência, o teste da medida e mostram que é bastante significativo o efeito de negligenciar a existência de correlação. Sugerem duas soluções para este problema, que são avaliadas analiticamente e através de simulações em computador. Uma delas se baseia na obtenção da fórmula correta para a variância, considera correlações seriadas dos dados do processo a outra envolve a filtragem para remoção de correlações seriadas destes dados, obtendo resíduos não correlacionados, aos quais podem ser aplicadas as técnicas desenvolvidas para a detecção de erros grosseiros em dados de processo independentes.

Yang *et al.* (1995) revisaram brevemente os métodos de detecção de erros grosseiros, quais sejam:

Baseados na suposição de distribuição normal

- Teste Global (1965);
- Teste da Máxima Potência (1975);

- Teste Nodal ou de Restrição (1976);
- Teste da Medida (1982);
- Método do Critério de Informação de Akaike (1986)
- Razões de Verossimilhança Generalizada (1989);
- Análise de Componentes Principais (1994).

Baseados em distribuições não normais

- Distribuições de Verossimilhança Bivariada (1991);
- Função de Densidade de Probabilidade Não-central (1993);

e **Baseados em redes neurais**. Observaram que nenhum deles apresenta garantias consistentes de encontrar todos os erros grosseiros nos dados de processo. Um método combinado é então proposto baseado no teste da medição (*measurement test* ou **MT**) e no teste nodal (*nodal test* ou **NT**) e este método tem aplicação prática e factibilidade, com óbvias vantagens sobre as abordagens similares prévias.

A vantagem do MT está no fato de que a localização dos erros grosseiros pode ser apontada nos dados do processo diretamente mas com a desvantagem de tender a gerar outros erros grosseiros (erro tipo I - apontar um erro quando este não ocorre realmente). O método NT não espalha os erros grosseiros por todos os dados mas tem como desvantagem a possibilidade de não detectar um erro grosseiro (erro tipo II - não detectar a presença de um erro quando este de fato ocorre) no caso de haver dois erros grosseiros de mesma magnitude e sinais opostos na entrada ou saída do sistema ou dois erros grosseiros de mesma magnitude e sinal, um na entrada e outro na saída do sistema, cancelando-se mutuamente. Outra desvantagem é que este método não aponta qual corrente contém o erro grosseiro diretamente, sendo necessário o teste de diversas combinações de nós diferentes para rastrear o erro, o que torna o processo lento.

A abordagem combinada do MT e NT baseia-se em primeiramente encontrar os dois nós ligados à corrente detectada pelo MT, então checa-se estes dois nós pelo NT. Se nenhum deles apresenta o erro, passa-se à próxima corrente candidata a portar o erro grosseiro que tenha sido apontada pelo MT e assim sucessivamente.

Zhang *et al.* (2001) mostram que por vezes a eliminação de medidas detectadas como portadoras de erros grosseiros provoca perda de precisão na solução do problema de reconciliação de dados subsequente, de modo que propõem um método de análise de redundância, baseado na medida da precisão da reconciliação, que permite a eliminação criteriosa e seqüencial de medidas portadoras de erros grosseiros com o objetivo de preservar a solvabilidade do problema de reconciliação. Os autores aplicam o método proposto a uma unidade de síntese de metanol e na Fujian Petrochemical Ltd. Co., da República Popular da China.

3 *Reconciliação de Dados Linear em Estado Estacionário*

All models are false, but some are useful

John Tukey

3.1 Introdução

Quando um estimador é comparado com a observação, algumas questões podem surgir como, por exemplo, qual será o efeito do posicionamento da medida (observação) no desempenho do estimador e qual é o efeito de medidas enviesadas sobre esse estimador. Estas e outras questões são levadas em conta no problema de estimativa de parâmetros e na seleção da estrutura das medidas para o monitoramento ou controle de um dado processo. Está claro que os conceitos de redundância e a alocação das medidas têm um importante papel no problema da estimativa. Além disso a redundância é útil como segurança quando há vieses nas medidas ou imperfeições no modelo da situação física sob consideração (ROMAGNOLI; SÁNCHEZ, 2000).

Neste capítulo, a formulação matemática do problema de reconciliação de dados é colocada e são mostradas algumas abordagens para a sua solução.

3.2 Conceitos Básicos

Para se ter um problema de estimativa de parâmetros, qualquer que seja ele, em primeiro lugar deve haver um sistema com várias medidas disponíveis.

O sistema é algum objeto físico e o seu comportamento pode ser descrito por equações. O sistema pode ser dinâmico (discreto ou contínuo) ou estático. Daqui por diante será discutido um processo em estado estacionário.

Seja uma quantidade (ou vetor quantidade), associado com a operação de um sistema cujo valor se deseja a cada instante de tempo. Se esta quantidade não for diretamente mensurável ou só puder ser medida com erro, deve-se assumir que \mathbf{y} medidas com ruído¹ estão disponíveis e que um experimento tenha sido projetado para medir ou estimar a referida quantidade associada do sistema. O conjunto dos valores reais das variáveis pode ser escrito como o vetor:

$$\mathbf{x}^T = [x_1, x_2, \dots, x_g] \quad (3.1)$$

A situação mais geral é aquela na qual as variáveis desejadas não podem ser observadas (medidas) diretamente e devem, portanto, ser indiretamente medidas como uma função de observações diretas. Assim, assume-se também que um conjunto de l medidas \mathbf{y} possa ser expresso como uma função de g elementos de um vetor constante \mathbf{x} mais um erro aleatório aditivo ε . Desta forma, as medidas do processo podem ser modeladas como:

$$\mathbf{y} = \phi(\mathbf{x}) + \varepsilon, \quad \mathbf{y} \in \mathfrak{R}^l, \quad \mathbf{x} \in \mathfrak{R}^g \quad (3.2)$$

onde ϕ representa o modelo funcional da medição.

Se $\varepsilon = \mathbf{0}$, então $\mathbf{y} = \phi(\mathbf{x})$ e diz-se que as medidas são perfeitas (não contêm erros). Se $\varepsilon \neq \mathbf{0}$, então elas contêm ruído. Nos casos os quais ϕ é diferenciável no ponto \mathbf{x}^0 ,

¹O ruído pode ser definido genericamente como tudo o que aparece sobreposto a um sinal e que não faz parte do sinal.

pode-se definir a matriz \mathbf{C} :

$$\mathbf{C}(\mathbf{x}^0) = \left. \frac{\partial \phi}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}^0} \quad (3.3)$$

onde \mathbf{C} é a versão linearizada das equações não lineares das medidas. Para sistemas lineares, a matriz \mathbf{C} é constante e independente de \mathbf{x} . Em geral, o sistema linear ou linearizado é expresso por

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \varepsilon, \quad \mathbf{C} \in \mathfrak{R}^{l \times g}, \quad (3.4)$$

onde \mathbf{C} é a Jacobiana ($l \times g$) de ϕ . Assim, quando as observações forem planejadas, deve ser especificado um modelo funcional geral sobre o sistema a ser avaliado (matriz \mathbf{C}). Tal modelo funcional, que se refere a um sistema finito fechado, é determinado por um certo número de variáveis e pelas relações entre elas.

Há sempre um número mínimo de variáveis independentes que determinam unicamente um modelo. Denota-se este número por g . A menos que as observações sejam suficientes para determinar as g variáveis, a situação será obviamente deficiente. Estas observações devem ser linearmente independentes, isto é, nenhuma das l observações pode ser derivada de nenhuma outra das $(l - 1)$ restantes.

Define-se um sistema como redundante quando a quantidade de dados disponíveis (informação) excede a quantidade mínima necessária para uma determinação singular das variáveis independentes do modelo.

Para o sistema na Equação 3.2, quando l é maior que g , é dito haver redundância. Esta redundância, denotada por r , é dada por:

$$r = l - g \quad (3.5)$$

é igual ao conceito estatístico de graus de liberdade.

Visto que os dados são comumente obtidos de observações que estão sujeitas a flutuações probabilísticas, os dados redundantes são em geral inconsistentes no sentido de que cada subconjunto com um número suficiente de variáveis para uma determinação

singular do sistema provê um resultado diferente dos demais. Para se obter uma única solução, um critério adicional se faz necessário. Se o princípio dos mínimos quadrados for aplicado, entre todas as soluções que são consistentes com o modelo de medição, as estimativas que são as mais próximas possível das medidas são consideradas como sendo a solução do problema de estimativa. Define-se o problema de estimativa pelos mínimos quadrados como:

$$\text{Min } J = (\mathbf{y} - \mathbf{C}\mathbf{x})^T(\mathbf{y} - \mathbf{C}\mathbf{x}) \quad (3.6)$$

A solução dos mínimos quadrados é a minimização da soma dos quadrados do $J = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}$ residual. A equação em \mathbf{x} ,

$$\mathbf{C}^T \mathbf{C} \mathbf{x} = \mathbf{C}^T \mathbf{y} \quad (3.7)$$

obtida pela diferenciação de J , é chamada de equação normal. Pode-se agora definir a propriedade da estimabilidade:

DEFINIÇÃO 3.1 Diz-se que um sistema é estimável se a equação normal admite uma única solução e, naturalmente, \mathbf{x} é único.

Assim, as condições necessárias para a estimabilidade podem ser definidas. Para que as variáveis do processo, \mathbf{x} , sejam estimáveis, o seguinte deve ser verdadeiro (RAO, 1973).

TEOREMA 3.1 O sistema descrito pela Equação 3.7 é globalmente estimável se e somente se

$$\text{posto } \mathbf{C} = g \quad (3.8)$$

onde g é a dimensão do sistema.

Reciprocamente, se

$$\text{posto } \mathbf{C} < g \quad (3.9)$$

o sistema é globalmente não-estimável.

Quando o sistema é não-estimável, um valor estimado de \mathbf{x} , $\hat{\mathbf{x}}$, não é uma solução única do problema de mínimos quadrados. Neste caso, a solução é possível somente

com a adição de informação. Esta informação é introduzida através de equações do modelo do processo (equações de restrição). Elas ocorrem na prática quando algumas ou todas variáveis do sistema devem obedecer às relações que surgem das restrições físicas do processo.

Em alguns casos, a introdução de equações adicionais do modelo do processo pode aumentar o número de variáveis a serem estimadas, não diminuindo assim a deficiência de estimabilidade.

Com a introdução de restrições adicionais tem-se:

$$\begin{aligned}\mathbf{0} &= \varphi(\mathbf{x}) & \mathbf{x} &\in \mathfrak{R}^g \\ \mathbf{y} &= \phi(\mathbf{x}) + \varepsilon & \mathbf{y} &\in \mathfrak{R}^l\end{aligned}\tag{3.10}$$

onde $\varphi \in \mathfrak{R}^m$, com m sendo o número de equações de restrição adicionais.

Estas relações funcionais que caracterizam o comportamento de processos reais não são nunca conhecidas exatamente. Uma maneira convencional de contabilizar a falta de acurácia gerada pelas aproximações é introduzir um ruído aditivo, o qual, de certa forma, reflete o grau de erro na modelagem, isto é,

$$\begin{aligned}\mathbf{0} &= \varphi(\mathbf{x}) + \mathbf{w} & \mathbf{x} &\in \mathfrak{R}^g \\ \mathbf{y} &= \phi(\mathbf{x}) + \varepsilon & \mathbf{y} &\in \mathfrak{R}^l\end{aligned}\tag{3.11}$$

Assumindo que $\varphi(\mathbf{x})$ e $\phi(\mathbf{x})$ sejam diferenciáveis em \mathbf{x}^0 , e aplicando-se uma expansão em série de Taylor, usando somente os termos zero e de primeira ordem (descartando os termos de segunda ordem em diante), chega-se ao sistema linear ou linearizado descrito por:

$$\begin{aligned}\mathbf{0} &= \mathbf{A}\mathbf{x} + \mathbf{w} \\ \mathbf{y} &= \mathbf{C}\mathbf{x} + \varepsilon\end{aligned}\tag{3.12}$$

onde \mathbf{A} e \mathbf{C} são as matrizes $(m \times g)$ e $(l \times g)$ da Jacobiana de φ e ϕ . Neste caso, a condição de redundância será satisfeita quando $(m + l) > g$. Pode-se agora definir o

problema dos mínimos quadrados da seguinte forma:

$$\text{Min } J = (\mathbf{z} - \mathbf{M}\mathbf{x})^T(\mathbf{z} - \mathbf{M}\mathbf{x}), \quad (3.13)$$

onde

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}, \mathbf{z} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} \quad (3.14)$$

A equação normal é dada por

$$\mathbf{M}^T\mathbf{M}\mathbf{x} = \mathbf{M}^T\mathbf{z} \quad (3.15)$$

De modo similar ao caso anterior, as condições gerais para a estimabilidade podem ser colocadas como se segue

TEOREMA 3.2 *O sistema descrito pelas equações 3.14 e 3.15 são globalmente estimáveis se, e somente se,*

$$\text{posto } \mathbf{M} = \text{posto} \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = g \quad (3.16)$$

Reciprocamente, se

$$\text{posto } \mathbf{M} = \text{posto} \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} < g \quad (3.17)$$

o sistema é globalmente não-estimável.

Define-se agora uma forma mais geral da função objetivo quadrática, a qual permite atribuir pesos pré-determinados aos componentes. Qual seja:

$$J = \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\varepsilon} \end{bmatrix}^T \mathbf{W} \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\varepsilon} \end{bmatrix} \quad (3.18)$$

onde \mathbf{W} é matriz de ponderação (matriz de variância-covariância), restrita a ser tanto simétrica quanto definida positivamente, ou seja, $\mathbf{W} = \mathbf{W}^T > \mathbf{0}$. A introdução da matriz de ponderação define o problema como mínimos quadrados ponderados e as mesmas

condições estabelecidas pelos teoremas 3.1 e 3.2 também se aplicam nesta situação. (ROMAGNOLI; SÁNCHEZ, 2000).

3.3 Base Estatística da Reconciliação de Dados

Descrita a formulação do problema de reconciliação de dados a partir de um ponto de vista quase que puramente intuitivo, especialmente com respeito aos pesos da função objetivo a serem usados em diferentes medidas, o problema de reconciliação de dados pode ser também explicado usando-se uma base teórica estatística que não somente ajuda na compreensão deste assunto, bem como dispõe informação quantitativa sobre a melhora da acurácia nos dados obtidos através da reconciliação e das propriedades estatísticas das estimativas resultantes. Estas propriedades podem ser usadas para identificar dados grosseiramente incorretos ou para projetar redes de sensores.

A base estatística para a reconciliação de dados vem das propriedades que são pressupostas para erros aleatórios nas medidas. Geralmente se assume que os erros aleatórios seguem uma distribuição normal multivariada com média zero e uma matriz de variância-covariância, Ψ , conhecida. Contudo, deve ser levado em conta que algumas vezes o sinal primário mensurado é transformado na variável final de interesse. Se a transformação é não-linear como na Equação 3.19

$$F = k\sqrt{\frac{p_0\Delta p}{T}} \quad (3.19)$$

onde k é a constante do orifício da placa, Δp é a diferença de pressão no orifício, p_0 é a pressão de entrada no orifício e T é a temperatura do fluido. Então o erro na variável indicada não tem necessariamente uma distribuição normal.

Somente a forma linearizada pode ser aproximada por uma distribuição normal. Assim, se possível, as variáveis \mathbf{x} no modelo de medida da Equação 3.4 devem representar as variáveis primárias mensuradas sendo que as relações entre a variável primária mensurada e as variáveis de interesse devem ser incluídas como restrições. No caso de restrições não lineares, uma técnica de reconciliação de dados não linear deve ser usada.

A matriz Ψ contém informação sobre a acurácia das medidas e as correlações entre elas. Os elementos na diagonal de Ψ , σ_i^2 , são a variância na i -ésima variável mensurada e os elementos fora da diagonal, σ_{ij}^2 são a covariância dos erros nas variáveis i dada uma variável j . Se os valores medidos são dados pelo vetor \mathbf{y} , então as estimativas mais prováveis para \mathbf{x} são obtidas pela maximização da função de verossimilhança da distribuição normal multivariada:

$$\max_x \frac{1}{(2\pi)^{\frac{n}{2}} |\Psi|^{\frac{n}{2}}} e^{-0,5(\mathbf{y}-\mathbf{x})^T \Psi^{-1} (\mathbf{y}-\mathbf{x})} \quad (3.20)$$

onde $|\Psi|$ é o determinante de Ψ . O problema de máxima verossimilhança acima é equivalente à minimização da função

$$\min_x (\mathbf{y} - \mathbf{x})^T \Psi^{-1} (\mathbf{y} - \mathbf{x}) \quad (3.21)$$

Como σ_i é o desvio padrão do erro na i -ésima medida, o fator de ponderação Ψ na Equação 3.21 é inversamente proporcional ao desvio padrão deste erro. Desta forma, um valor mais alto de desvio padrão implica em uma medida menos acurada e assim tem-se pesos maiores para medidas mais acuradas.

É possível também derivar propriedades estatísticas das estimativas obtidas através de reconciliação de dados. Considerando-se o caso quando todas as variáveis são mensuradas, as estimativas são dadas por:

$$\hat{\mathbf{x}} = \mathbf{y} - \Psi \mathbf{A}^T (\mathbf{A} \Psi \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{y} = [\mathbf{I} - \Psi \mathbf{A}^T (\mathbf{A} \Psi \mathbf{A}^T)^{-1} \mathbf{A}] \mathbf{y} = \mathbf{B} \mathbf{y} \quad (3.22)$$

A Equação 3.22 mostra que as estimativas são obtidas usando-se transformações lineares das medidas. As estimativas, portanto, são normalmente distribuídas, com valor esperado e matriz de covariância dados por:

$$E[\hat{\mathbf{x}}] = \mathbf{B} E(\mathbf{y}) = \mathbf{B} \mathbf{x} = \mathbf{x} \quad (3.23)$$

$$\text{Cov}[\hat{\mathbf{x}}] = E\{(\mathbf{B}\mathbf{y})(\mathbf{B}\mathbf{y})^T\} = \mathbf{B}\Psi\mathbf{B}^T \quad (3.24)$$

A Equação 3.23 implica em estimativas não enviesadas o que é uma propriedade de um estimador de máxima verossimilhança para sistemas lineares. A Equação 3.24 dá uma medida da acurácia das estimativas.

No caso em que algumas das variáveis não são medidas, também é possível se derivar propriedades estatísticas semelhantes. Estas propriedades são exploradas para a identificação de medidas portadoras de erros grosseiros bem como para projetar redes de sensores. (NARASIMHAN; JORDACHE, 2000).

3.4 Formulação do Problema de Reconciliação de Dados

Como foi colocado nas seções anteriores, a reconciliação de dados melhora a acurácia dos dados através do ajuste dos valores mensurados de modo que estes satisfaçam restrições do processo. A quantidade do ajuste feito sobre as medidas é minimizada posto que os erros nas medidas são tidos como pequenos. No caso geral, não são todas as variáveis que são medidas devido à limitações de ordem técnica ou econômica.

As estimativas das variáveis não-medidas bem como os parâmetros do modelo são também obtidos como parte do problema de reconciliação. A estimativa de valores baseada em medidas reconciliadas é também conhecida como coaptação de dados. De um modo geral, a reconciliação de dados pode ser formulada pelo seguinte problema de otimização por mínimos quadrados sujeitos a restrições:

$$\min_{x_i, u_j} \sum_{i=1}^n \omega_i (y_i - x_i)^2 \quad (3.25)$$

sujeito a

$$g_k(x_i, u_i) = 0, \quad k = 1, \dots, m \quad (3.26)$$

A função objetivo 3.25 define a soma total dos quadrados ponderados dos ajustes feitos sobre as medidas, onde os ω_i são os pesos, y_i é a medida e x_i é a estimativa reconciliada para a variável i e os u_j são as estimativas das variáveis não medidas. A Equação 3.26 define o conjunto das restrições do modelo. Os pesos ω_i são escolhidos dependendo da acurácia das diferentes medidas.

As restrições do modelo são geralmente balanços de massa e energia, mas também podem incluir relações de desigualdade impostas pela factibilidade das operações no processo. As leis determinísticas naturais da conservação de massa e energia são tipicamente usadas como restrições para a reconciliação de dados pois elas são conhecidas e válidas no universo dentro do qual foram feitas as hipóteses que as originaram. Equações empíricas e de outros tipos envolvendo vários parâmetros não medidos não são recomendadas, posto que elas são, na melhor das hipóteses, conhecidas somente de modo aproximado. Forçar as variáveis medidas a obedecer relações inexatas pode levar a uma reconciliação de dados inexata e ao diagnóstico incorreto de erros grosseiros.

Qualquer lei de conservação de massa ou energia pode ser expressa pela seguinte fórmula geral:

$$\text{entrada} - \text{saída} + \text{geração} - \text{consumo} - \text{acúmulo} = 0$$

A quantidade para qual a equação acima é escrita pode ser o fluxo material global, o fluxo de componentes individuais ou o fluxo de energia. Se não há acúmulo de qualquer destas quantidades, então estas restrições são algébricas e definem uma operação em estado estacionário.

Para um processo dinâmico, contudo, o termo do acúmulo não pode ser desprezado e as restrições são equações diferenciais. Para a maioria das unidades de processo, não há geração ou depleção de material. No caso dos reatores, porém, a geração ou consumo de componentes individuais devido à reação deve ser levada em consideração.

Para algumas unidades mais simples como *splitters* (divisores de corrente), não há nenhuma variação nem na composição nem na temperatura das correntes. Para tais

unidades, os balanços de componente e de energia se reduzem à forma simples:

$$x_i = x_j \quad (3.27)$$

onde a variável x_i representa tanto a temperatura quanto a composição da corrente i . A equação acima é útil também quando dois ou mais sensores são usados para medir a mesma variável – uma taxa de fluxo ou a temperatura de uma corrente, por exemplo.

O tipo das restrições que são impostas na reconciliação de dados dependem do escopo do problema e do tipo da unidade. Além disso, a complexidade das técnicas de solução usadas depende fortemente das restrições impostas. Por exemplo, se o interesse é reconciliar somente taxas de fluxos de todas as correntes, então as restrições dos balanços materiais são lineares nas variáveis de fluxo e daí resulta um problema de reconciliação de dados linear. Por outro lado, se o interesse é reconciliar composições, temperaturas ou pressão em conjunto com os fluxos, então tem-se um problema de reconciliação de dados não linear.

Um quesito importante é o tipo de restrições que podem ser legitimamente impostas na aplicação da reconciliação de dados. Haja vista que a reconciliação de dados força as estimativas de todas as variáveis a satisfazerem as restrições impostas, esta questão tem grande importância. Normalmente, as restrições de balanço de massa e energia são incluídas porque estas são leis físicas válidas. Deve ser notado, contudo, que estas equações são geralmente escritas assumindo que não há perda de material ou de energia do processo para o meio ambiente. Enquanto isto pode ser válido para o fluxo material, uma perda significativa de energia pode ocorrer fruto, por exemplo, do isolamento impróprio de trocadores de calor. Em tais casos, é melhor não impor os balanços de energia ou, alternativamente, incluir um termo não conhecido de perda na equação do balanço e este pode ser estimado como parte do problema de reconciliação.

Outra restrição além da conservação de massa e energia pode ser um modelo de uma unidade de processo que contenha equações envolvendo parâmetros da unidade. Por exemplo, o modelo de um trocador de calor pode incluir a equação relacionado a carga térmica ao coeficiente global de troca térmica, a área de troca e os fluxos e

temperaturas das correntes. A Equação 3.28 descreve esta relação.

$$Q - UA\Delta T_{ln} = 0 \quad (3.28)$$

onde Q é a carga térmica, U é o coeficiente global de troca térmica e ΔT_{ln} é a diferença média logarítmica de temperatura.

Geralmente, uma equação como esta pode ser incluída como restrição em uma reconciliação sobre um trocador de calor se o coeficiente de troca for desconhecido e tenha que ser estimado a partir de dados medidos. Se não houver nenhuma informação prévia sobre U e nenhuma restrição de factibilidade, então a inclusão desta restrição não fornece nenhuma informação adicional e as estimativas de todas as outras variáveis serão as mesmas sendo esta restrição incluída ou não. Desta forma, o problema de reconciliação de dados pode ser resolvido sem esta restrição e U pode ser subsequentemente estimado pela equação anterior, usando-se os valores reconciliados de fluxos e temperaturas.

Por outro lado, se U tem que estar dentro de limites específicos ou se há uma boa estimativa de U vinda de um ciclo anterior de reconciliação, então a restrição deve ser incluída junto com a informação adicional sobre U como parte do problema de reconciliação. O coeficiente global de troca térmica pode também ser relacionado às propriedades físicas das correntes, seus fluxos, temperaturas e as características do trocador de calor usando-se correlações. Não é aconselhável se usar uma equação destas no modelo de reconciliação, pois as correlações podem conter muitos erros nelas mesmas e se estas são usadas, força-se os fluxos e temperaturas a se ajustarem a esta equação, o que pode aumentar a inacurácia das estimativas.

Outra questão importante é se se deve proceder à reconciliação de dados usando-se um modelo estacionário ou dinâmico do processo. Na prática, um processo nunca está verdadeiramente em estado estacionário. Apesar disso, uma planta é operada por várias horas ou dias em uma região em torno de um ponto de operação nominalmente em estado estacionário.

Para aplicações como otimização *on line*, nas quais a reconciliação é realizada uma vez em algumas horas, é apropriado empregar-se a reconciliação de dados em estado estacionário sobre a média das medidas em um intervalo de tempo de interesse.

Durante um regime transiente, quando a saída do estado estacionário é significativa, a reconciliação de dados estacionária não deve ser aplicada porque resultará em grandes ajustes aos valores médios. Medidas tomadas durante tais períodos transientes podem ser reconciliadas, se necessário, usando-se um modelo dinâmico do processo. De um modo similar, para aplicações de controle de processo, nas quais a reconciliação precisa ser realizada a cada poucos minutos, a reconciliação de dados dinâmica é mais apropriada.

A reconciliação de dados é baseada no pressuposto que somente erros aleatórios estão presentes nas medidas, os quais seguem uma distribuição normal (Gaussiana). Se um erro grosseiro devido a um viés na medição está presente em alguma medida ou se há um vazamento significativo no processo que não tenha sido contabilizado no modelo das restrições, então os dados reconciliados podem ser bastante inaccurados. É, portanto, necessário identificar e remover tais erros grosseiros. Isto é conhecido como o **problema de detecção de erros grosseiros**.

Erros grosseiros podem ser detectados baseando-se na extensão na qual as medidas violam as restrições ou na magnitude dos ajustes feitos às medidas em uma reconciliação preliminar. Ainda que as técnicas de detecção de erros grosseiros tenham sido desenvolvidas primariamente para melhorar a acurácia de estimativas reconciliadas, elas também são úteis na identificação de instrumentos de medição que precisam ser trocados ou recalibrados. (NARASIMHAN; JORDACHE, 2000).

3.5 Decomposição do Problema Geral de Estimativa

As seções anteriores discutiram a formulação matemática do problema de estimativa de parâmetros e as condições gerais de redundância e estimabilidade. Será analisada agora a decomposição do problema geral de estimativa. A divisão de sistemas lineares

em suas partes observáveis e não observáveis foi primeiro sugerida por Kalman (1960). O mesmo tipo de argumento pode ser usado aqui para decompor um sistema considerado em estado estacionário.

Quando os resultados da teoria de matrizes são aplicados ao problema geral de estimativa de parâmetros, o seguinte pode ser definido:

TEOREMA 3.3 *Para o sistema descrito pela equações 3.14 e 3.15, se*

$$\text{posto } \mathbf{M} = \text{posto} \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = j < g, \quad (3.29)$$

então existe uma matriz não singular \mathbf{T} *tal que*

$$\mathbf{MT} = \begin{bmatrix} \mathbf{A}_U & \mathbf{0} \\ \mathbf{C}_U & \mathbf{0} \end{bmatrix}, \quad (3.30)$$

onde \mathbf{A}_U *e* \mathbf{C}_U *têm* j *colunas e*

$$\text{posto} \begin{bmatrix} \mathbf{A}_U \\ \mathbf{C}_U \end{bmatrix} = j. \quad (3.31)$$

O sistema de equações 3.12 pode ser escrito usando a forma escalonada em colunas da matriz \mathbf{M} tal como segue (ROMAGNOLI; SÁNCHEZ, 2000):

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} = \mathbf{MTT}^{-1}\mathbf{x} + \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_U & \mathbf{0} \\ \mathbf{C}_U & \mathbf{0} \end{bmatrix} \mathbf{T}^{-1}\mathbf{x} + \begin{bmatrix} \mathbf{w} \\ \boldsymbol{\varepsilon} \end{bmatrix} \quad (3.32)$$

Dependendo da estrutura de \mathbf{T}^{-1} , duas situações podem surgir:

1. Se cada linha de \mathbf{T}^{-1} tem somente um elemento não-zero, isto significa fisicamente que nas novas coordenadas $\mathbf{x}_c = [x_r, x_{g-r}]$, onde \mathbf{x}_r é um vetor j -dimensional, o subsistema

$$\begin{aligned} \mathbf{0} &= \mathbf{A}_U \mathbf{x}_r + \mathbf{w} \\ \mathbf{y} &= \mathbf{C}_U \mathbf{x}_r + \boldsymbol{\varepsilon} \end{aligned} \quad (3.33)$$

é estimável. O sistema como um todo admite uma decomposição em dois sub-sistemas menores: um estimável, de dimensão j e outro não estimável, de dimensão $(g - j)$. O primeiro inclui as variáveis \mathbf{x}_r e o último contém as variáveis em \mathbf{x}_{g-r} .

2. Se algumas linhas de \mathbf{T}^{-1} têm mais de um elemento não zero, existem combinações lineares entre as variáveis em \mathbf{x}_r e as variáveis em \mathbf{x}_{g-r} . Assim, a porção estimável do sistema é de dimensão ob tal que $(ob < j)$ e a porção não estimável é de dimensão $(g - ob)$.

Uma medida é considerada redundante se a sua remoção não causa perda de estimabilidade. Se o posto de $\mathbf{M} = g$ e $(m + l) > g$, isto é, há mais informação disponível que o necessário para uma determinação singular do sistema, então o seguinte teorema pode ser enunciado:

TEOREMA 3.4 *Se o sistema de equações (3.14) e (3.15) é estimável e redundante, isto é, $(m + l > g)$, com $(l - i)$ medidas redundantes e se as linhas de \mathbf{C} são permutadas de modo que as primeiras $(l - i)$ linhas correspondam às medidas redundantes (\mathbf{y}_1) , ou seja,*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix} \text{ e } i > \mathbf{0}, \quad (3.34)$$

então existe uma matriz \mathbf{F} ($g \times g$), não-singular tal que

$$\mathbf{MF} = \begin{bmatrix} \mathbf{A}_U & \mathbf{0} \\ \mathbf{C}_{1U} & \mathbf{0} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \quad (3.35)$$

e

$$\text{posto } \mathbf{C}_{22} = i, \quad \text{posto } [\mathbf{A}_{1U}] = \text{posto} \begin{bmatrix} \mathbf{A}_U \\ \mathbf{C}_{1U} \end{bmatrix} = g - i \quad (3.36)$$

com cada uma das medidas do sistema $\mathbf{A}_{1U} = \begin{bmatrix} \mathbf{A}_U \\ \mathbf{C}_{1U} \end{bmatrix}$ sendo redundante.

A partir dos resultados do teorema anterior, conclui-se que qualquer sistema que seja estimável e redundante, ($r > 0$), admite a decomposição em suas partes redundante (\mathbf{x}_1) e não redundante (\mathbf{x}_2). Essa decomposição permite uma nova formulação em duas partes do problema geral dos mínimos quadrados.

PROBLEMA 1

Problema dos mínimos quadrados:

$$\min_{\mathbf{x}_1} (\mathbf{z}_1 - \mathbf{A}_{1U}\mathbf{x}_1)^T \mathbf{W}_1 (\mathbf{z}_1 - \mathbf{A}_{1U}\mathbf{x}_1) \quad (3.37)$$

onde

$$\mathbf{z} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y}_1 \end{bmatrix} \quad (3.38)$$

PROBLEMA 2

Como a decomposição permite que \mathbf{x}_1 seja determinado primeiro, o passo seguinte é calcular \mathbf{x}_2 usando os valores já conhecidos de \mathbf{x}_1 e \mathbf{y}_2 .

Esta formulação dividida em dois problemas distintos resulta numa significativa redução de dimensionalidade em relação ao problema original.

3.6 Classificação das Variáveis de Processo

Uma planta de processos químicos é um sistema físico contendo uma grande quantidade de unidades e correntes. Por exemplo, a contagem dos equipamentos no setor de processo e de utilidades (considerando os misturadores e os divisores de corrente) de uma planta petroquímica pode revelar a existência de aproximadamente 1000 unidades interconectadas e cerca de 2500 correntes. Em cada corrente as variáveis de interesse podem ser vazão, composição, temperatura, pressão e entalpia. Assim o tratamento dos dados de uma planta típica envolve a solução de um problema em grande escala (ROMAGNOLI; SÁNCHEZ, 2000).

A idéia original de reduzir os sistemas de equações usados no problema da reconci-

liação é devida a Václavek (1969), o qual propôs um procedimento de correção baseado somente em um subconjunto reduzido de equações e medidas. A idéia consiste na exploração da topologia para classificar as variáveis do processo e eliminar do problema original as que não são medidas, resultando em um subconjunto de equações envolvendo somente variáveis medidas. Várias estratégias foram desenvolvidas desde então para alcançar o mesmo objetivo, qual seja, a decomposição do processo para reduzir a dimensionalidade do problema. Algumas destas estratégias são baseadas na teoria dos grafos (Mah *et al.* (1976); Kretsovalis e Mah (1988a, 1988b); Meyer *et al.* (1993)), e outras abordagens orientadas a equação (Crowe *et al.* (1983); Crowe (1986, 1989a); Romagnoli e Stephanopoulos (1980); Joris e Kalitventzeff (1987)).

Do que foi apresentado, torna-se evidente que a aplicação de técnicas de reconciliação de dados em grandes plantas, representadas por modelos não lineares complexos é um problema desafiador. A decomposição através da classificação das variáveis do processo é uma importante ferramenta no tratamento da dimensionalidade do problema. O que é mais importante é que a compreensão da estrutura topológica da planta não somente permite decompô-la, mas também pode ser muito útil no projeto ou na análise de um sistema completo de monitoramento.

Seja um processo contendo K unidades denotadas por $k = 1, \dots, K$, e J correntes orientadas, $j = 1, \dots, J$, com C componentes, $c = 1, \dots, C$. A topologia da planta pode ser representada pela Matriz de Incidência, \mathbf{L} , com as linhas correspondendo às **unidades** e colunas correspondendo às **correntes**. Assim

$$\begin{aligned} L_{jk} &= 1 && \text{se a corrente } j \text{ entra no nó } k \\ L_{jk} &= -1 && \text{se a corrente } j \text{ sai do nó } k \\ L_{jk} &= 0 && \text{se a corrente } j \text{ não tem contato com o nó } k \end{aligned}$$

As restrições de balanço para uma unidade de processo sem reações químicas e transferência de calor podem ser expressas tal como segue (ROMAGNOLI; SÁNCHEZ, 2000):

Balanços de Massa:

$$\sum_j \mathbf{L}_{j,k} f_j = 0 \quad (3.39)$$

Balanços de Massa para os Componentes:

$$\sum_j \mathbf{L}_{j,k} f_j \mathbf{M}_{c,j} = 0 \quad (3.40)$$

Balanços de Entalpia:

$$\sum_j \mathbf{L}_{j,k} f_j h_j = 0 \quad (3.41)$$

Equações de Normalização:

$$\sum_c f_j \mathbf{M}_{c,j} - f_j = 0 \quad (3.42)$$

onde f_j é o fluxo total na corrente j , $\mathbf{M}_{c,j}$ é a fração do componente c na corrente j e h_j é a entalpia específica da corrente j .

De um modo geral, o modelo de uma planta operando em estado estacionário é constituído de um sistema de equações algébricas não lineares da forma

$$\varphi(\mathbf{x}, \mathbf{u}) = \mathbf{0}, \quad \varphi \in \mathfrak{R}^m, \quad (3.43)$$

onde φ é uma função não linear tendo como variáveis \mathbf{x} e \mathbf{u} , os vetores das variáveis do processo, medidas e não medidas, respectivamente. Para balanços lineares de massa, a eq.(3.43) se torna:

$$\mathbf{A}_1 \mathbf{x} + \mathbf{A}_2 \mathbf{u} = \mathbf{0}, \quad \mathbf{x} \in \mathfrak{R}^g, \quad \mathbf{u} \in \mathfrak{R}^n \quad (3.44)$$

onde \mathbf{A}_1 , e \mathbf{A}_2 são matrizes compatíveis de dimensão $(m \times g)$ e $(m \times n)$, respectivamente.

Se o estado do sistema for diretamente medido, então o modelo da medição é representado por:

$$\mathbf{y} = \mathbf{x} + \varepsilon \quad (3.45)$$

Neste caso, a Jacobiana das funções de medição \mathbf{C} é igual à matriz identidade e o vetor

do erro aleatório das medidas é

$$\varepsilon = \mathbf{y} - \mathbf{x} \quad (3.46)$$

Fica disposto, assim, que a formulação da reconciliação de dados é apenas um caso especial do problema geral de estimativa de parâmetros.

3.6.1 Definições

As variáveis não medidas do processo podem ser classificadas em **determináveis** e **não determináveis**.

Uma variável não medida é dada como determinável se esta pode ser avaliada a partir das medidas disponíveis com auxílio das equações de balanço. Por outro lado, se esta variável não puder ser avaliada da forma mencionada, ela é dita “não determinável”.

As variáveis medidas também se dividem em duas categorias: **redundantes** e **não redundantes** (Figura 3.1).

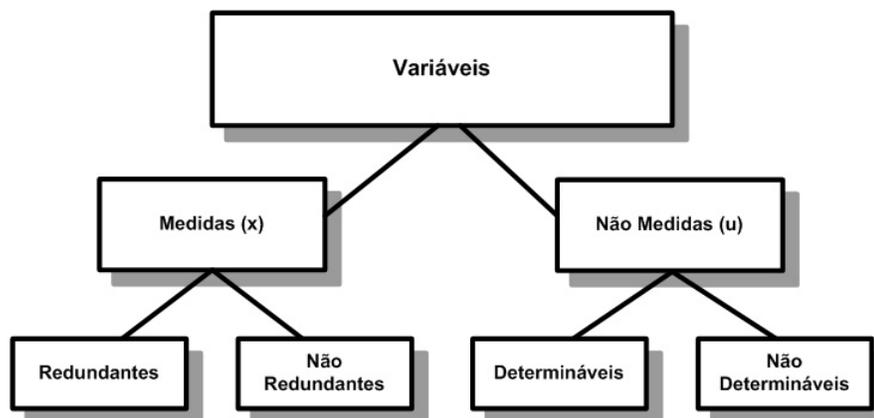


Figura 3.1: Classificação das variáveis medidas

Uma medida é dita redundante (ou sobre-determinada) se esta puder ser encontrada através das equações de balanço aliadas às outras medições disponíveis. De modo contrário, se esta medida só pode ser alcançada através da própria medição, esta é dita não redundante.

Baseado na formulação anterior, os seguintes problemas podem ser definidos:

1. Classificar as variáveis não medidas;
2. Definir o subconjunto de equações redundantes a serem usadas no ajuste das medidas;
3. Classificar as variáveis medidas.

As ferramentas básicas para a avaliação estrutural das equações do processo serão brevemente discutidas. Estas ferramentas permitem analisar sistematicamente a estrutura topológica das equações de balanço e resolver os três problemas que foram colocados.

3.6.2 Análise da Topologia do Processo

Seguindo a divisão das variáveis do processo nos vetores \mathbf{x} e \mathbf{u} , medidas e não medidas, respectivamente, os sistemas de equações lineares ou linearizadas podem ser divididos nas matrizes compatíveis \mathbf{A}_1 e \mathbf{A}_2 através da Equação 3.44.

Esta divisão sugere uma representação estrutural do sistema onde as matrizes \mathbf{A}_1 e \mathbf{A}_2 consistem de alguns elementos que são geralmente não-nulos e outros que são sempre nulos.

O sistema de matrizes \mathbf{A}_1 e \mathbf{A}_2 descrevem a topologia estrutural das correntes e unidades em termos de variáveis e equações as quais podem ser associadas a um gráfico mostrando as suas influências mútuas.

Sejam os nós do gráfico as variáveis de processo e suas fronteiras as relações (equações de balanço) entre elas. Há uma fronteira em comum entre o nó a e o nó i , se a pertence ao intervalo de i , ou seja, se a é necessário para avaliar i .

Os conceitos expostos acima podem ser ilustrados na Figura 3.2, onde três unidades são conectadas a sete correntes, sendo que destas, apenas as correntes 1, 2 e 4 são

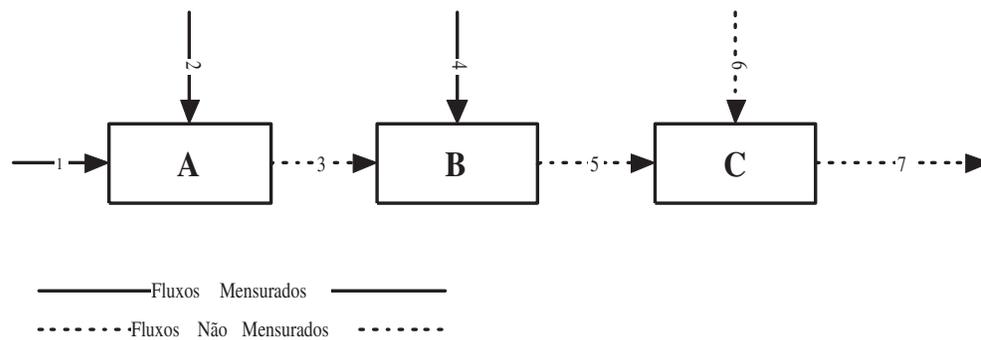


Figura 3.2: Diagrama de fluxo para um sistema simples em série (ROMAGNOLI; STEPHANOPOULOS, 1980)

medidas. Procedendo a um balanço de massa total sobre cada unidade de acordo com a Equação 3.39, tem-se:

$$f_1 + f_2 - f_3 = 0$$

$$f_3 + f_4 - f_5 = 0$$

$$f_5 + f_6 - f_7 = 0$$

A resolução destas três equações para as variáveis f_3 , f_5 e f_6 proporciona o gráfico de fluxo de informação mostrado na Figura 3.3. São introduzidos agora alguns conceitos em conexão com os sistemas estruturais e os seus gráficos associados

DEFINIÇÃO 3.2 (Inacessibilidade) *Define-se um nó i como inacessível a partir do nó a se não houver possibilidade de alcançar i partindo de a (o qual corresponde a uma variável mensurada) e indo para o nó i na direção das setas ao longo de um caminho no gráfico de fluxo de informação.*

DEFINIÇÃO 3.3 (Determinabilidade) *Define-se um nó i como determinável se qualquer caminho para o nó i começar em um nó mensurado.*

Na Figura 3.3, aplicando a definição de acessibilidade:

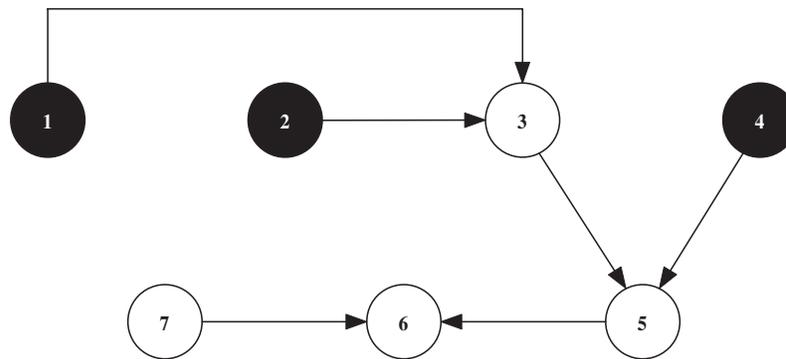


Figura 3.3: Gráfico de fluxo de informação para o sistema descrito na Figura 3.2 (ROMAGNOLI; STEPHANOPOULOS, 1980)

- Os nós 3, 5 e 6 são acessíveis;
- O nó 7 é inacessível.

Na mesma Figura, aplicando agora o conceito de determinabilidade:

- Os nós 3 e 5 são determináveis;
- Os nós 6 e 7 são não determináveis.

Conseqüentemente, f_3 e f_5 são variáveis do processo não mensuradas mas são determináveis, ao passo que f_6 e f_7 são indetermináveis.

3.6.3 Abordagens para Solução do Problema de Classificação

Durante as três últimas décadas, várias estratégias foram formuladas para realizar a classificação de variáveis. Estas estratégias podem ser divididas em dois grandes grupos. Um dos grupos aplica os conceitos da **teoria dos grafos**² e o outro faz uso de técnicas

²A Teoria dos Grafos é o ramo da matemática que estuda as propriedades de grafos. Um grafo é um conjunto de pontos, chamados vértices (ou nós), conectados por linhas, chamadas de arestas (ou arcos). Dependendo da aplicação, arestas podem ou não ter direção, pode ser permitido ou não arestas ligarem um vértice a ele próprio e vértices e/ou arestas podem ter um peso (numérico) associado. Se as arestas têm uma direção associada (indicada por uma seta na representação gráfica) tem-se um grafo direcionado, ou digrafo.

de ordenação de matrizes e procedimentos computacionais. Está disposta a seguir uma breve revisão sobre o assunto que pode ser encontrada em Romagnoli e Sánchez (2000).

3.6.3.1 Técnicas Orientadas a Grafos

Dada a topologia de um processo, um grafo não orientado é construído onde os nós correspondem às unidades e os arcos às correntes do processo. O grafo do processo contém um nó para o meio-ambiente do qual o processo recebe as correntes de alimentação e para o qual vão as correntes de produtos finais.

As principais contribuições às técnicas orientadas a gráficos são devidas aos seguintes autores.

Václavek

Václavek (1969) foi quem primeiro definiu os conceitos de observabilidade e redundância. Ele formulou duas regras para se categorizar as variáveis em modelos lineares de plantas:

1. Agregar dois nós conectados com uma corrente não mensurada. O Esquema de Balanço Reduzido resultante contém somente medidas redundantes;
2. Retirar todas as correntes medidas e procure por ciclos no grafo reduzido. Os ciclos no grafo resultante representam fluxos não determináveis.

Václavek e Loucka (1976) estenderam a abordagem a processos multicomponentes com a suposição que, em qualquer corrente, ou todas as frações mássicas são medidas ou nenhuma é. As reações químicas são contabilizadas pela adição de correntes fictícias no grafo. Os *splitters* não são considerados em sua formulação.

Mah e Colaboradores

Apresentaram uma abrangente teoria e algoritmos para a classificação de variáveis medidas e não medidas. Para processos com um único componente (somente balanços

de massa), Mah *et al.* (1976) derivaram um procedimento simples de classificação baseado na teoria dos grafos. Em um trabalho posterior, Kretsovalis e Mah (1987) descreveram a categorização das variáveis para fluxos de multicomponentes sem pressuposições sobre a localização dos sensores. Não foram levadas em conta reações químicas nem *splitters*. Kretsovalis e Mah (1988a, 1988b) estenderam o seu tratamento para incluir reatores, *splitters* e unidades onde ocorrem fluxos de energia. As seguintes variáveis das correntes foram consideradas em sua análise: fluxo de massa, frações mássicas, fluxos de componentes e de energia e temperaturas. O conjunto de medidas é restrito a fluxos mássicos, frações mássicas e temperaturas. Foi pressuposto que existe uma correspondência unívoca entre temperatura e entalpia por unidade de massa.

A técnica requer uma análise extensa do grafo do processo e dos seus sub-grafos derivados (16 + número de componentes). Eles são testados por um conjunto de 19 teoremas de observabilidade e redundância. Estes subgrafos são atualizados durante a execução do procedimento. A classificação das variáveis não medidas é alcançada usando as regras derivadas somente da teoria dos grafos e da álgebra matricial.

Meyer et al.

Os autores (MEYER *et al.*, 1993) introduziram um método variante derivado do Kretsovalis e Mah (1987) que permite o tratamento de reações químicas e *splitters*. Ele leva à diminuição do tamanho do problema de reconciliação de dados bem como um particionamento das equações para classificação das variáveis não medidas.

3.6.3.2 Técnicas Orientadas a Equação

Dada a topologia do processo e um conjunto de medidas, estas estratégias geram primeiramente o sistema de equações do modelo para a planta, procedendo posteriormente diferentes tipos de rearranjos e cálculos envolvendo matrizes e equações não lineares de modo a classificar as variáveis do processo. As principais contribuições para essa abordagem são dispostas a seguir:

Romagnoli e Stephanopoulos

Romagnoli e Stephanopoulos (1980) propuseram um procedimento de classificação de variáveis baseado na aplicação de um algoritmo de atribuições de saída (*output set assignment*) sobre a sub-matriz das variáveis não medidas, associado com as equações lineares ou não lineares do modelo. Uma variável designada como não medida é classificada como determinável depois de se checar se seu cálculo pode ser feito através da resolução da equação correspondente ou sub-conjunto de equações.

Um conjunto de equações redundantes é construído a partir das equações não designadas sem variáveis indetermináveis e de equações de balanço específicas em torno de sistemas disjuntos de unidades. As medições envolvidas neste conjunto são classificadas como redundantes.

O procedimento foi originalmente aplicado à classificação de variáveis para sistemas bilineares de equações. No caso de balanços multicomponentes, foi considerado que a composição de uma corrente ou é completamente conhecida ou nenhuma composição o é.

Crowe

Para modelos lineares, Crowe *et al.* (1983) usaram a projeção de matrizes para obter um conjunto reduzido de equações que permitem a classificação das variáveis medidas. Eles identificaram que as variáveis não medidas através da redução de colunas da sub-matriz correspondente a estas variáveis.

Crowe (1986) estendeu esta metodologia para a classificação de variáveis envolvidas em balanços bilineares de componentes. O modelo é modificado para uma forma linear usando o conhecimento da topologia do processo, da localização dos instrumentos e um conjunto de medidas que devem ser consistentes com as restrições do processo. Crowe (1989a) propôs um algoritmo de classificação de variáveis baseado numa série de lemas. Nesta formulação, balanços bilineares de energia são incluídos nas equações do modelo, assumindo uma correspondência unívoca entre a temperatura e a entalpia

por unidade de massa.

O procedimento permite a inclusão de medições em lugares arbitrários, reações químicas, fluxos em *splitters* e fluxos de energia.

Joris e Kalitventzeff

O procedimento desenvolvido por Joris e Kalitventzeff (1987) objetiva classificar as variáveis e medições envolvidas em qualquer tipo de modelo de planta. O sistema de equações que representa a operação da planta envolve variáveis de estado (temperatura, pressão, taxas de fluxos parciais molares de componentes e extensão de reação), medidas e variáveis de ligação (as que relacionam certas medidas com as variáveis de estado). Este sistema é composto de balanços de massa e energia, relações de equilíbrio líquido-vapor e etc. A classificação de variáveis não medidas e das medições é alcançada pela permuta de linhas e colunas da matriz de ocorrência correspondente à matriz Jacobiana do modelo.

Na maioria dos casos, o procedimento estrutural é capaz de determinar se as medições podem ser corrigidas e quando elas possibilitam o cômputo de todas as variáveis de estado do processo. Em algumas configurações esta técnica, usada sozinha, falha na detecção de variáveis indetermináveis. Esta situação surge quando a Jacobiana usada na resolução é não inversível.

Madron

O procedimento de classificação desenvolvido por Madron é baseado na conversão da matriz com as equações do modelo da planta lineares ou linearizadas para a forma canônica. Inicialmente é formada uma matriz composta, contendo variáveis medidas e não medidas e um vetor de constantes. A seguir é realizada uma eliminação Gauss-Jordan, usada para pivotar as colunas pertencentes às quantidades não mensuradas. Na fase seguinte, o procedimento aplica a eliminação sobre uma sub-matriz resultante que contém variáveis mensuradas. A forma canônica final é obtida pelo rearranjo de linhas e colunas da macro-matriz, a qual permite a classificação de ambos os tipos de variáveis.

O usuário deste procedimento deve prover estimativas iniciais para todas as variáveis. Esta estratégia é extensamente descrita na monografia de Madron (1993).

3.7 Decomposição Usando Transformações Ortogonais

Crowe *et al.* (1983) propuseram uma importante estratégia para o desacoplamento de variáveis mensuradas a partir das equações lineares das restrições. Este procedimento permite tanto a redução do problema de reconciliação de dados quanto a classificação das variáveis do processo. Ele é baseado no uso de projeção de matrizes para eliminar as variáveis não medidas. Posteriormente Crowe estendeu essa metodologia (CROWE, 1986, 1989a).

Uma decomposição equivalente pode ser realizada usando as transformações ortogonais QR (em Sánchez *et al.* (1992)). As fatorações ortogonais foram primeiro usadas para por Swartz (1989) no contexto das técnicas de sucessivas linearizações para eliminar as variáveis não mensuradas das equações de restrição (ROMAGNOLI; SÁNCHEZ, 2000).

3.7.1 Abordagem da Projeção de Matrizes

Representando um processo em estado estacionário através de:

$$\mathbf{A}_1\mathbf{x} + \mathbf{A}_2\mathbf{u} = \mathbf{0}, \quad \mathbf{x} \in \mathfrak{R}^g; \mathbf{u} \in \mathfrak{R}^n, \quad (3.47)$$

onde \mathbf{x} é o vetor ($g \times 1$) das variáveis medidas e \mathbf{u} é o vetor ($n \times 1$) das variáveis não medidas. \mathbf{A}_1 e \mathbf{A}_2 são matrizes compatíveis de dimensão ($m \times g$) e ($m \times n$).

Uma matriz de projeção \mathbf{P} foi definida por Crowe tal que \mathbf{P} pré-multiplicando a matriz Jacobiana \mathbf{A}_2 resulta em:

$$\mathbf{P}\mathbf{A}_2 = \mathbf{0} \quad (3.48)$$

As colunas de \mathbf{P} abarcam o espaço nulo de \mathbf{A}_2 , e assim as variáveis não medidas são eliminadas. Para obter a matriz de projeção \mathbf{P} , Crowe propôs o seguinte procedimento:

1. Reduzir as colunas de \mathbf{A}_2 para obter a matriz \mathbf{X} com colunas linearmente independentes

$$\mathbf{A}_2\mathbf{A}_3 = \begin{bmatrix} \mathbf{X} & \mathbf{0} \end{bmatrix} \quad (3.49)$$

onde \mathbf{A}_3 representa a matriz inversível que realiza as operações necessárias sobre \mathbf{A}_2 .

2. Particionar \mathbf{X} tal que

$$\mathbf{A}_4\mathbf{A}_2\mathbf{A}_3 = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{0} \end{bmatrix} \quad (3.50)$$

com \mathbf{X}_1 quadrada e inversível. Então \mathbf{P} é calculado pela seguinte expressão:

$$\mathbf{P} = \begin{bmatrix} -\mathbf{X}_2\mathbf{X}_1^{-1} & \mathbf{I} \end{bmatrix} \mathbf{A}_4 \quad (3.51)$$

Finalmente o problema reduzido é formulado como:

$$\mathbf{G}\mathbf{x} = \mathbf{0} \quad (3.52)$$

onde

$$\mathbf{G} = \mathbf{P}\mathbf{A}_1 \quad (3.53)$$

A reconciliação de dados pode agora ser realizada sobre um subsistema reduzido contendo somente variáveis mensuradas.

3.7.2 Abordagem da Fatoração QR

Uma decomposição alternativa pode ser realizada usando a fatoração **QR** da matriz \mathbf{A}_2 para desacoplar as variáveis não medidas das medidas (SÁNCHEZ; ROMAGNOLI, 1996).

TEOREMA 3.5 (Teorema da Fatoração QR (DAHLQUIST; BJORK, 1974))

Seja \mathbf{A} uma matriz $(m \times n)$ com $m \geq n$ e n colunas linearmente independentes. Então existe uma única matriz \mathbf{Q} $(m \times m)$,

$$\mathbf{Q}^T \mathbf{Q} = \mathbf{D}_i, \quad \mathbf{D}_i = \begin{bmatrix} d_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & d_n \end{bmatrix}; \quad d_k > 0, \quad k = 1, \dots, n \quad (3.54)$$

e uma única matriz triangular superior \mathbf{R} $(m \times n)$, com $R_{kk} = 1, k = 1, \dots, n$ tal que

$$\mathbf{A} = \mathbf{Q}\mathbf{R} \quad (3.55)$$

Se \mathbf{A} tem o seu posto deficiente a fatoração QR pode ser modificada de um modo simples para que fique da seguinte forma:

$$\mathbf{A}\Pi = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.56)$$

onde $r = \text{posto}(\mathbf{A})$, \mathbf{Q} é ortogonal, \mathbf{R}_{11} é triangular superior e Π é uma permutação. Se $\mathbf{A}\Pi = [a_{c1}, \dots, a_{cn}]$ e $\mathbf{Q} = [\mathbf{q}_1, \dots, \mathbf{q}_m]$, então para $k = 1, \dots, n$, tem-se que

$$\mathbf{a}_{ck} = \sum_{i=1}^{\min\{r,k\}} r_{ik} \mathbf{q}_i \in \text{span}\{\mathbf{q}_1, \dots, \mathbf{q}_r\} \quad (3.57)$$

Segue que para qualquer vetor que satisfaça $\mathbf{A}\mathbf{x} = \mathbf{b}$,

$$\Pi^T \mathbf{x} = \begin{bmatrix} \mathbf{s} \\ \mathbf{z} \end{bmatrix} \quad \text{e} \quad \mathbf{Q}^T \mathbf{b} = \begin{bmatrix} \mathbf{i} \\ \mathbf{l} \end{bmatrix} \quad (3.58)$$

onde \mathbf{s} e \mathbf{i} são vetores de dimensão r , \mathbf{z} é um vetor de dimensão $(n - r)$ e \mathbf{l} é um vetor de dimensão $(m - r)$.

De volta ao problema da reconciliação, a fatoração QR da matriz \mathbf{A}_2 permite obter as matrizes \mathbf{Q}_u e \mathbf{R}_u e a matriz de permuta Π_u tais que:

$$\mathbf{A}_2 \boldsymbol{\Pi}_u = \mathbf{Q}_u \mathbf{R}_u \quad (3.59)$$

onde \mathbf{Q}_u e \mathbf{R}_u podem ser divididas em:

$$\mathbf{Q}_u = \begin{bmatrix} \mathbf{Q}_{u1} & \mathbf{Q}_{u2} \end{bmatrix}, \mathbf{R}_u = \begin{bmatrix} \mathbf{R}_{u1} & \mathbf{R}_{u2} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.60)$$

com $r_u = \text{posto}(\mathbf{A}_2) = \text{posto}(\mathbf{R}_{u1})$. Tem-se que \mathbf{Q}_u é uma matriz ortogonal e \mathbf{R}_{u1} é uma matriz inversível triangular superior de dimensão r_u . Do mesmo modo as variáveis não mensuradas do processo podem ser particionadas em dois subconjuntos:

$$\boldsymbol{\Pi}_u^T \mathbf{u} = \begin{bmatrix} \mathbf{u}_{r_u} \\ \mathbf{u}_{n-r_u} \end{bmatrix} \quad (3.61)$$

Pré-multiplicando as restrições linearizadas por $\mathbf{Q}_u^T = \mathbf{Q}_u^{-1}$, obtém-se

$$\begin{bmatrix} \mathbf{Q}_{u1}^T \mathbf{A}_1 & \mathbf{R}_{u1} & \mathbf{R}_{u2} \\ \mathbf{Q}_{u2}^T \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{u}_{r_u} \\ \mathbf{u}_{n-r_u} \end{bmatrix} = \mathbf{0} \quad (3.62)$$

as primeiras r_u equações para \mathbf{u}_{r_u} podem ser escritas em termo das outras variáveis:

$$\mathbf{u}_{r_u} = \mathbf{R}_{u1}^{-1} \mathbf{Q}_{u1}^T \mathbf{A}_1 \mathbf{x} - \mathbf{R}_{u1}^{-1} \mathbf{R}_{u2} \mathbf{u}_{n-r_u} \quad (3.63)$$

como as variáveis não medidas não aparecem nas equações restantes, o primeiro sub-problema reduzido se torna:

PROBLEMA 1

$$\min_x (\mathbf{y} - \mathbf{x})^T \boldsymbol{\Psi}_x^{-1} (\mathbf{y} - \mathbf{x}) \quad (3.64)$$

$$\mathbf{G}_x \mathbf{x} = \mathbf{0} \quad (3.65)$$

onde

$$\mathbf{G}_x = \mathbf{Q}_{u2}^T \mathbf{A}_1$$

PROBLEMA 2

Estimar as variáveis não medidas, \mathbf{u} , resolvendo a Eq. 3.63 onde os componentes \mathbf{u}_{n-r_u} são arbitrariamente colocados. A unicidade de \mathbf{u} é relacionada a estimabilidade do sistema.

3.8 Reconciliação de Dados Linear com Todas as Variáveis Medidas

Este problema pode ser formulado como:

$$\begin{aligned} \min_x J(\mathbf{y} - \mathbf{x})^T \Psi^{-1} (\mathbf{y} - \mathbf{x}) \\ \mathbf{A}_1 \mathbf{x} = \mathbf{0} \end{aligned} \quad (3.66)$$

onde \mathbf{A}_1 é uma matriz ($m \times g$) com constantes conhecidas. Neste caso todas as variáveis são redundantes.

3.8.1 Método dos Multiplicadores de Lagrange

Introduzindo o erro da medida nas restrições do processo tem-se que

$$\mathbf{A}_1 (\mathbf{y} - \boldsymbol{\varepsilon}) = \mathbf{0} \quad (3.67)$$

Conseqüentemente, o problema de otimização se torna agora

$$\begin{aligned} \min_{\boldsymbol{\varepsilon}} \boldsymbol{\varepsilon}^T \Psi^{-1} \boldsymbol{\varepsilon} \\ \mathbf{A}_1 \mathbf{x} = \mathbf{0} \end{aligned} \quad (3.68)$$

A solução é obtida pelo **Método dos Multiplicadores de Lagrange**. O Lagrangiano para este problema é

$$L = \boldsymbol{\varepsilon}^T \Psi^{-1} \boldsymbol{\varepsilon} - 2\boldsymbol{\lambda}^T (\mathbf{A}_1 \mathbf{y} - \mathbf{A}_1 \boldsymbol{\varepsilon}) \quad (3.69)$$

Como Ψ é positivamente definida e as restrições são lineares, as condições necessárias e suficientes para a minimização são:

$$\begin{aligned}\frac{\partial L}{\partial \varepsilon} &= 2\Psi^{-1}\varepsilon + 2\mathbf{A}_1^T\lambda = \mathbf{0} \\ \frac{\partial L}{\partial \lambda} &= \mathbf{A}_1(\mathbf{y} - \varepsilon) = \mathbf{0}\end{aligned}\quad (3.70)$$

com

$$\varepsilon = -\Psi\mathbf{A}_1^T\lambda \quad (3.71)$$

$$\lambda = -(\mathbf{A}_1\Psi\mathbf{A}_1^T)^{-1}\mathbf{A}_1\mathbf{y} \quad (3.72)$$

E finalmente, o estimador para as variáveis do processo, $\hat{\mathbf{x}}$, pode ser obtido como

$$\hat{\mathbf{x}} = \mathbf{y} - \Psi\mathbf{A}_1^T(\mathbf{A}_1\Psi\mathbf{A}_1^T)^{-1}\mathbf{A}_1\mathbf{y} \quad (3.73)$$

3.8.2 Método da Fatoração QR

Usando-se o método da fatoração QR, o problema de estimativa por mínimos quadrados ponderados sujeito a restrições é transformado em um problema não sujeito a restrições. Para tanto, os seguintes passos devem ser dados:

Passo1: Computar a solução geral do sistema indeterminado ($\mathbf{A}_1\mathbf{x} = \mathbf{0}$). Usando o procedimento indicado na Seção 3.7.2. A Fatoração ortogonal QR de \mathbf{A}_1 produz as matrizes \mathbf{Q}_x , \mathbf{R}_x e Π_x que permitem o cálculo de \mathbf{Q}_{x1} , \mathbf{Q}_{x2} , \mathbf{R}_{x1} , \mathbf{R}_{x2} , \mathbf{x}_{r_x} e \mathbf{x}_{g-r_x} , tais que:

$$\mathbf{A}_1\Pi_x = \mathbf{Q}_x\mathbf{R}_x \quad (3.74)$$

$$\mathbf{Q}_x = \begin{bmatrix} \mathbf{Q}_{x1} & \mathbf{Q}_{x2} \end{bmatrix}, \quad \mathbf{R}_x = \begin{bmatrix} \mathbf{R}_{x1} & \mathbf{R}_{x2} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.75)$$

$$\Pi_x^T \mathbf{x} = \begin{bmatrix} \mathbf{x}_{r_x} \\ \mathbf{x}_{g-r_x} \end{bmatrix} \quad (3.76)$$

onde $r_x = \text{posto}(\mathbf{R}_{x1}) = \text{posto}(\mathbf{A}_1)$. A solução geral do problema é:

$$\mathbf{x}_{r_x} = -\mathbf{R}_{x1}^{-1} \mathbf{R}_{x2} \mathbf{x}_{g-r_x} \quad (3.77)$$

onde \mathbf{x}_{g-r_x} é um vetor arbitrário

Passo 2: Formulação do problema não restrito. Aplicando-se os resultados anteriores, o vetor $(\mathbf{y} - \mathbf{x})$ da função objetivo é modificado tal como segue:

$$\begin{aligned} (\mathbf{y} - \mathbf{x}) &= \mathbf{y} - \begin{bmatrix} \mathbf{I}_{x1} & \mathbf{I}_{x2} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{r_x} \\ \mathbf{x}_{g-r_x} \end{bmatrix} = \\ &= \mathbf{y} + \mathbf{I}_{x1} \mathbf{R}_{x1}^{-1} \mathbf{R}_{x2} \mathbf{x}_{g-r_x} - \mathbf{I}_{x2} \mathbf{x}_{g-r_x} = \\ &= \mathbf{y} + (\mathbf{I}_{x1} \mathbf{R}_{x1}^{-1} \mathbf{R}_{x2} - \mathbf{I}_{x2}) \mathbf{x}_{g-r_x} \end{aligned} \quad (3.78)$$

onde

$$\mathbf{\Pi}_x = \begin{bmatrix} \mathbf{I}_{x1} & \mathbf{I}_{x2} \end{bmatrix}, \quad \tilde{\mathbf{I}} = \mathbf{I}_{x1} \mathbf{R}_{x1}^{-1} \mathbf{R}_{x2} - \mathbf{I}_{x2} \quad (3.79)$$

\mathbf{I} representa uma matriz identidade $(g \times g)$ e $\tilde{\mathbf{I}}$ é uma matriz $[g \times (g - r_x)]$ com colunas independentes.

A minimização não restrita pode agora ser declarada como:

$$\min(\mathbf{y} + \tilde{\mathbf{I}} \mathbf{x}_{g-r_x})^T \Psi^{-1} (\mathbf{y} + \tilde{\mathbf{I}} \mathbf{x}_{g-r_x}) \quad (3.80)$$

Passo 3: Estimativa de \mathbf{x} . A solução do problema acima é:

$$\hat{\mathbf{x}}_{g-r_x} = -(\tilde{\mathbf{I}}^T \Psi^{-1} \tilde{\mathbf{I}})^{-1} \tilde{\mathbf{I}}^T \Psi^{-1} \mathbf{y} \quad (3.81)$$

usando-se o valor de \mathbf{x}_{g-r_x} , a equação 3.77 é resolvida para se calcular $\hat{\mathbf{x}}_{r_x}$.

Nota-se também que a dimensão do problema de otimização não sujeito a restrições é menor que a do problema original.

3.9 Reconciliação de Dados com Variáveis não Medidas

Segundo Romagnoli e Sánchez (2000), o pressuposto que todas as variáveis são medidas é geralmente falso, pois na prática, algumas variáveis não são medidas e precisam ser estimadas. Na Seção anterior, a decomposição de um problema de reconciliação de dados linear envolvendo somente variáveis mensuradas foi discutida, levando a um problema de mínimos quadrados reduzido. Nesta seção, são usados estes conceitos para prover uma solução geral do problema de reconciliação de dados linear quando algumas das variáveis não são medidas. A solução é baseada no desacoplamento entre as variáveis não medidas e as variáveis medidas, usando-se fatoração ortogonal QR. Desta forma, o problema global de estimativa é dividido em dois sub-problemas.

Considerando-se as equações de restrição

$$\mathbf{A}_1 \mathbf{x} + \mathbf{A}_2 \mathbf{u} = \mathbf{0} \quad (3.82)$$

Realizando-se uma decomposição QR sobre a matriz \mathbf{A}_2 , as matrizes \mathbf{Q}_u , \mathbf{R}_u e Π_u são obtidas tais que:

$$\mathbf{A}_2 \Pi_u = \mathbf{Q}_u \mathbf{R}_u \quad (3.83)$$

$$\mathbf{Q}_u = \begin{bmatrix} \mathbf{Q}_{u1} & \mathbf{Q}_{u2} \end{bmatrix}, \quad \mathbf{R}_u = \begin{bmatrix} \mathbf{R}_{u1} & \mathbf{R}_{u2} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (3.84)$$

onde $r_u = \text{posto}(\mathbf{A}_2) = \text{posto}(\mathbf{R}_{u1})$. o vetor das variáveis não medidas é particionado em dois subconjuntos

$$\Pi_u^T \mathbf{u} = \begin{bmatrix} \mathbf{u}_{r_u} \\ \mathbf{u}_{n-r_u} \end{bmatrix} \quad (3.85)$$

Pré-multiplicando as restrições lineares por \mathbf{Q}^T , obtém-se

$$\begin{aligned} \mathbf{Q}_{u1}^T \mathbf{A}_1 \mathbf{x} + \mathbf{R}_{u1} \mathbf{u}_{r_u} + \mathbf{R}_{u2} \mathbf{u}_{n-r_u} &= \mathbf{0} \\ \mathbf{Q}^T \mathbf{A}_1 \mathbf{x} &= \mathbf{0} \end{aligned} \quad (3.86)$$

Realizando a reconciliação no subsistema desacoplado representado pelas variáveis mensuradas \mathbf{x} e as restrições

$$\mathbf{Q}_{u2}^T \mathbf{A}_1 \mathbf{x} = \mathbf{G}_x \mathbf{x} = \mathbf{0} \quad (3.87)$$

a solução do sistema é:

$$\hat{\mathbf{x}} = \mathbf{y} - \Psi \mathbf{G}_x^T (\mathbf{G}_x \Psi \mathbf{G}_x^T)^{-1} \mathbf{G}_x \mathbf{y} \quad (3.88)$$

Contudo, este problema pode ser reduzido ainda mais, usando-se os conceitos desenvolvidos anteriormente, para o caso no qual todas as variáveis são medidas, levando à solução de uma seqüência de sub-problemas menores.

Para as variáveis não medidas, tem-se em geral:

$$\mathbf{u}_{r_u} = -\mathbf{R}_{u1}^T \mathbf{Q}_{u1}^T \mathbf{A}_1 \hat{\mathbf{x}} - \mathbf{R}_{u1}^{-1} \mathbf{R}_{u2} \mathbf{u}_{n-r_u} \quad (3.89)$$

onde os componentes \mathbf{u}_{n-r_u} são dispostos arbitrariamente. Pode-se ter, assim, dois casos:

1. $\text{Posto}(\mathbf{R}_{u1}) = n$
2. $\text{Posto}(\mathbf{R}_{u1}) < n$

Caso(1) Todos os parâmetros não medidos são estimáveis (observáveis) e uma solução única para as variáveis não medidas é possível usando-se os valores mensurados ajustados e as equações de balanço;

Caso(2) Algumas variáveis não mensuradas do processo não são estimáveis e é possível um infinito número de soluções. Assim, a solução básica é:

$$\mathbf{u}_{r_u} = -\mathbf{R}_{u1}^{-1} \mathbf{Q}_{u1}^T \mathbf{A}_1 \hat{\mathbf{x}}; \quad \mathbf{u}_{n-r_u} = 0 \quad (3.90)$$

3.10 Conclusão

A abordagem para solução do problema de minimização utilizada neste trabalho é a da fatoração QR, que foi escolhida no lugar da matriz de projeção de Crowe por sua implementação mais simples a partir do pacote de métodos numéricos utilizado, o IMSL.

Os procedimentos vistos neste Capítulo, em conjunto com as técnicas exploradas no próximo, permitem uma implementação prática para solução do problema de reconciliação, retificação e cooptação de dados. Esta implementação, para reconciliação linear de vazão mássica e em estado estacionário, é descrita no Capítulo 5.

4 *Detecção de Erros Grosseiros*

It is only possible to assess uncertainty, not to prove hypotheses

John Tukey

4.1 Introdução

A técnica de reconciliação de dados depende crucialmente do pressuposto da presença exclusiva de erros aleatórios nos dados e da ausência de erros sistemáticos tanto nas medidas quanto nas equações do modelo. Se esse pressuposto for inválido, a reconciliação pode levar a grandes ajustes sobre os valores medidos e as estimativas resultantes podem ser bastante inexatas e por vezes infactíveis. Assim, é importante identificar tais erros sistemáticos ou grosseiros antes das estimativas reconciliadas finais serem obtidas. Os fundamentos das técnicas de detecção são abordados neste capítulo.

4.2 Conceitos Básicos

No Capítulo 3, foi dito que a reconciliação pode ser realizada somente se restrições estão presentes. O mesmo pode ser dito com respeito à detecção de erros grosseiros. Sem a disponibilidade de restrições para a verificação das medidas, a detecção de erros grosseiros não pode ser realizada. Portanto, a reconciliação de dados e a detecção de erros grosseiros exploram a mesma informação disponível a partir das medidas e das

restrições. Estas técnicas estão portanto relacionadas e interagem no processamento de dados.

Existem dois tipos principais de erros grosseiros, como indicado no Capítulo 2. Um está relacionado ao desempenho do instrumento e inclui viéses de medida, tendências, descalibração e falha total de instrumentação. O outro é relacionado ao modelo de restrições e inclui perdas não contabilizadas de material e energia resultantes de vazamentos em equipamentos de processo ou inaccurácias de modelagem devido a parâmetros inexatos. Várias técnicas foram propostas e desenvolvidas para a detecção e eliminação destes dois tipos de erros grosseiros. Antes de descrever estas técnicas, é melhor inicialmente dispor de forma clara os requisitos de uma estratégia de detecção de erros grosseiros. Isto leva a um melhor entendimento sobre a variedade de técnicas que são propostas, suas interrelações e os resultados possíveis de cada uma.

Qualquer estratégia abrangente de detecção de erros grosseiros deve possuir as seguintes capacidades:

- Habilidade de detectar a presença de um ou mais erros grosseiros entre os dados (problema de detecção);
- Habilidade de identificar o tipo e a localização do erro grosseiro (problema de identificação);
- Habilidade de localizar e identificar múltiplos erros grosseiros que possam estar presentes simultaneamente entre os dados (problema da identificação de múltiplos erros grosseiros);
- Habilidade de estimar a magnitude dos erros grosseiros (problema de estimativa).

Nem todas as técnicas de detecção de erros grosseiros podem preencher os requisitos propostos acima. O último dos requisitos, ainda que útil, não é absolutamente necessário. Uma estratégia de detecção de erros grosseiros pode ser analisada em termos dos métodos que esta usa para atacar os três principais problemas: detecção, identificação de um único erro grosseiro e identificação de múltiplos erros grosseiros, sendo que

o desempenho de uma estratégia é função direta destes métodos que lhe compõe. No presente trabalho não será abordada a identificação de múltiplos erros grosseiros.

4.3 Testes Estatísticos Básicos para a Detecção de Erros Grosseiros

Este componente da estratégia de detecção de erros grosseiros tenta responder à questão de estarem presentes erros grosseiros ou não entre os dados. Como todos os métodos de detecção se utilizam direta ou indiretamente do fato que a presença de erros grosseiros entre as medidas causam a violação das restrições do modelo, então se as medidas não contêm quaisquer erros aleatórios, a violação de qualquer restrição do modelo por um valor medido pode ser imediatamente interpretada como devida à presença de erros grosseiros. Este é um método puramente determinístico. Já foi pressuposto, contudo, que todas as medidas contêm erros aleatórios, devido aos quais não se pode esperar que as medidas satisfaçam estritamente qualquer restrição de modelo, mesmo se erros grosseiros estão ausentes. Assim, há que ser feita uma “permissão” de violação das restrições devido a erros aleatórios.

Sob uma distribuição de probabilidades pré-definida para os erros aleatórios, uma abordagem probabilística é usada para resolver este problema.

O princípio básico na detecção de erros grosseiros é derivado da detecção de deslocamento de dados em aplicações estatísticas. O erro aleatório inerente a qualquer medida é pressuposto seguir uma distribuição normal com média zero e variância conhecida. O *erro normalizado*, que é a diferença entre o valor medido e o valor médio esperado, dividido pelo seu desvio padrão, segue uma distribuição normal padrão. A maioria dos erros normalizados está dentro de um intervalo de confiança $(1 - \alpha)$ a um nível de significância α escolhido. Qualquer valor (um erro normalizado) que caia fora deste intervalo de confiança é declarado como um *deslocamento* ou um erro grosseiro.

Um grande número de testes estatísticos é derivado deste princípio básico e estes

são capazes de detectar erros grosseiros, mas nem todos testes são capazes de identificar os diferentes tipos e a localização destes erros. Alguns testes básicos são capazes de detectar somente erros de medidas (viéses). Outros testes só podem detectar erros no modelo do processo ou vazamentos. Por outro lado, o teste generalizado da razão de verossimilhança, o qual é derivado do princípio da estimativa por máxima verossimilhança em estatística, pode ser usado tanto para detectar problemas de instrumentação como vazamentos no processo.

As técnicas estatísticas mais comumente usadas para detecção de erros grosseiros são baseadas no teste de hipótese. Em um caso de detecção de erro grosseiro, a hipótese nula, H_0 , é a de que não há erros grosseiros presentes e a hipótese alternativa, H_1 , é a de que há um ou mais erros grosseiros no sistema. Todas as técnicas estatísticas para a escolha entre estas duas hipóteses fazem uso de uma estatística de teste a qual é uma função das medidas e do modelo de restrições. A estatística de teste é comparada com um valor pré-determinado de referência e a hipótese nula é rejeitada ou aceita dependendo se esta estatística excede o valor de referência ou não. O valor de referência é também conhecido como critério de teste, valor crítico ou valor crítico de teste.

O resultado de um teste de hipótese não é perfeito. Um teste estatístico pode declarar a presença de erros grosseiros quando, de fato, não há nenhum erro grosseiro presente (H_0 é verdadeiro). Neste caso, o teste comete um **Erro do Tipo I**, ou levanta um *alarme falso*. Por outro lado, o teste pode detectar a medida como livre de erros, quando de fato esta contém um ou mais erros. Este é um **Erro do Tipo II**. O poder de um teste estatístico, que corresponde à probabilidade da correta detecção, é igual à probabilidade de $1 - \text{Erro do Tipo II}$. O poder e a probabilidade do Erro do Tipo I em qualquer teste estatístico estão intimamente relacionados.

O poder, ou a potência, de um teste estatístico pode ser aumentado através da permissão para uma probabilidade maior para o Erro do Tipo I. Portanto, quando se projeta um teste estatístico, a potência do teste deve ser balanceada face à probabilidade de falsa detecção. Se a distribuição de probabilidade da estatística de teste pode ser obtida sob a consideração da hipótese nula, então o critério de teste pode ser selecionado de modo

que a probabilidade do Erro do Tipo I seja menor ou igual a um valor especificado α . O parâmetro α é também chamado de nível de significância para o teste estatístico.

Serão, a seguir, considerados testes estatísticos para a detecção de erros grosseiros considerando-se modelos lineares e operação em estado estacionário.

Assume-se um modelo de restrições lineares dado por

$$\mathbf{Ax} = \mathbf{c} \quad (4.1)$$

onde \mathbf{A} é a matriz das restrições lineares e o vetor \mathbf{c} contém coeficientes conhecidos. Normalmente, para fluxos lineares, \mathbf{c} é o vetor zero, a menos que algumas das variáveis sejam conhecidas exatamente. Como nos capítulos anteriores, os erros nas medidas têm uma distribuição pressupostamente normal e têm uma matriz de variância-covariância, Ψ , conhecida.

Quatro testes estatísticos básicos foram desenvolvidos e amplamente aplicados para a detecção de erros grosseiros. Para simplificar a descrição destes testes, um modelo linear com todas as variáveis medidas será inicialmente proposto.

Os dois primeiros testes são baseados no vetor dos resíduos de balanço, \mathbf{r} , o qual é descrito por:

$$\mathbf{r} = \mathbf{Ay} - \mathbf{c} \quad (4.2)$$

Na ausência de erros grosseiros, o vetor \mathbf{r} segue uma distribuição normal multivariada com média zero e uma matriz de variância-covariância, \mathbf{V} , dada por

$$\mathbf{V} = \text{cov}(\mathbf{r}) = \mathbf{A}\Psi\mathbf{A}^T \quad (4.3)$$

Portanto, sob H_0 , $\mathbf{r}(\mathbf{0}, \mathbf{V})$. Na presença de erros grosseiros, os elementos do vetor residual \mathbf{r} refletem o grau de violação das restrições do processo (leis de conservação

de massa e energia). Por outro lado, a matriz \mathbf{V} contém informação sobre a estrutura do processo (matriz \mathbf{A}) e sobre a matriz de variância-covariância. As duas quantidades, \mathbf{r} e \mathbf{V} , podem ser usadas para construir estatísticas para detectar a presença de erros grosseiros.

4.3.1 O Teste Global (*Global Test - GT*)

O Teste Global usa o teste descrito por

$$\gamma = \mathbf{r}^T \mathbf{V}^{-1} \mathbf{r} \quad (4.4)$$

Sob H_0 , a estatística acima segue uma distribuição χ^2 com v graus de liberdade, com v igual ao posto da matriz \mathbf{A} . Se o critério de teste for escolhido como $\chi^2_{1-\alpha, v}$, onde $\chi^2_{1-\alpha, v}$ é o valor crítico da distribuição χ^2 a um dado nível de significância α , então H_0 é rejeitado e um erro grosseiro é detectado se $\gamma \geq \chi^2_{1-\alpha, v}$. Esta escolha do critério de teste assegura que a probabilidade de Erro do Tipo I para este teste é menor ou igual a α . O teste global combina todos os resíduos das restrições na obtenção da estatística de teste e assim dá lugar a um teste multivariado ou coletivo.

Para esclarecer a implementação deste teste é considerado um sistema de troca de calor com *bypass* (NARASIMHAN; JORDACHE, 2000), mostrado na Figura 4.1.

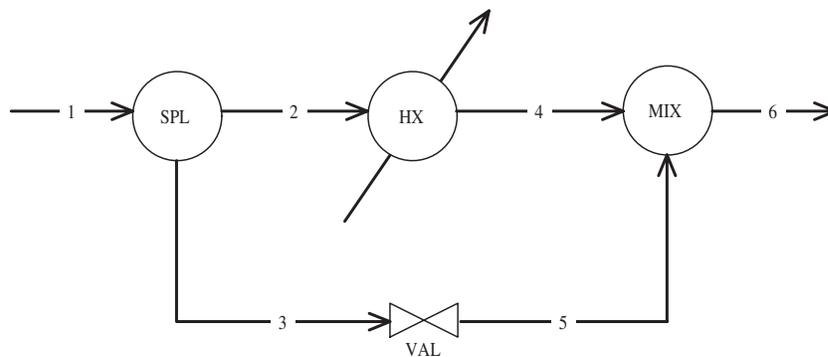


Figura 4.1: Sistema de troca de calor com *bypass* (NARASIMHAN; JORDACHE, 2000)

Tabela 4.1: Reconciliação de dados com a presença de erros grosseiros para o processo da Figura 4.1 (NARASIMHAN; JORDACHE, 2000)

Número da Corrente	Valores Verdadeiros dos Fluxos (kg/min)	Valores Mensurados dos Fluxos (kg/min)	Valores Reconciliados (kg/min)
1	100	100,91	100,89
2	64	68,45	65,83
3	36	34,65	35,05
4	64	64,20	65,83
5	36	36,44	35,05
6	100	98,88	100,89

APLICAÇÃO 4.1 Considerando-se a reconciliação sobre os fluxos do trocador de calor ilustrado na Figura 4.1, assume-se que todos os fluxos sejam medidos e os valores verdadeiros, os medidos e os reconciliados (desconsiderando-se a presença de erros grosseiros) estão dispostos na Tabela 4.1. A matriz das restrições para este processo é dada por

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

onde as linhas correspondem aos balanços materiais (fluxos) para o splitter, o trocador de calor, a válvula de bypass e o mixer e as colunas correspondem às seis correntes. Os resíduos das restrições, dado pela Equação 4.2 para as dadas medidas são iguais a

$$\mathbf{r} = \begin{bmatrix} -2,19 & 4,25 & -1,79 & 1,76 \end{bmatrix}$$

e a matriz variância-covariância dos resíduos das restrições, através de 4.3, é igual a

$$\mathbf{V} = \begin{bmatrix} 3 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 3 \end{bmatrix}$$

Usando-se a Equação 4.4, a estatística do teste global é computada como 16,674. Isto pode ser verificado como sendo igual à soma dos quadrados das diferenças entre os valores medidos e reconciliados. O critério de teste, em um nível de significância de 5%, para uma distribuição χ^2 com 4 graus de liberdade é igual a 9,488. Desta forma o teste global rejeita a hipótese nula, H_0 , e um erro grosseiro é detectado.

4.3.2 Teste Nodal ou Teste da Restrição (*Nodal Test - NT*)

O vetor \mathbf{r} também pode ser usado para derivar estatísticas de teste, uma para cada restrição i , na forma

$$\mathbf{z}_{r,i} = \frac{|r_i|}{\sqrt{V_{ii}}}, \quad i = 1, 2, \dots, m \quad (4.5)$$

ou escrevendo na forma vetorial

$$\mathbf{z}_r = [\text{diag}(\mathbf{V})]^{1/2} \mathbf{r} \quad (4.6)$$

onde $\text{diag}(\mathbf{V})$ é a matriz diagonal cujos elementos são V_{ii} . O teste nodal ou teste da restrição usa as estatísticas de teste $z_{r,i}$ para a detecção de erros grosseiros. $z_{r,i}$ segue uma distribuição normal padrão $N(0,1)$ sob H_0 . Se qualquer uma das estatísticas $z_{r,i}$ excede o critério de teste $z_{1-\alpha/2}$, onde $z_{1-\alpha/2}$ é o valor crítico da distribuição normal padrão para um nível α de significância, então um erro grosseiro é detectado.

Ao contrário do teste global, o teste nodal processa cada resíduo de restrição separadamente e dispõe m testes univariados. Como múltiplos testes são realizados usando o mesmo valor crítico, isto aumenta a probabilidade que um destes testes possa ser rejeitado mesmo se nenhum erro grosseiro estiver presente. Em outras palavras, a probabilidade do Erro do Tipo I será maior que o valor especificado de α . Se é desejável controlar a probabilidade de Erro do Tipo I, o nível de significância modificado β , proposto por Mah e Tamhane (1982) pode ser usado:

$$\beta = 1 - (1 - \alpha)^{1/m} \quad (4.7)$$

Para qualquer valor especificado de α , o valor modificado de β pode ser computado a partir da Equação 4.7 e o critério de teste para todos os testes nodais pode ser escolhido como $z_{1-\beta/2}$. Isto assegura que a probabilidade de qualquer um dos testes nodais ser rejeitado sob H_0 seja menor ou igual a α . Deve-se notar que α é apenas o limite superior da probabilidade de Erro do Tipo I e no sentido de se assegurar que a probabilidade de

Erro do Tipo I seja exatamente igual a α , pode-se adotar um critério de teste por tentativa e erro através de simulação. Como alternativa, Rollins e Davis (1992) propuseram o uso de um valor crítico baseado no intervalo de confiança de Bonferroni, o qual é dado por

$$\beta = \alpha/m \quad (4.8)$$

Para valores grandes de m , a Equação 4.7 se reduz à Equação 4.8.

É possível se obter outras formas do teste nodal através do uso de transformações lineares dos resíduos das restrições. Contudo, nem todas estas formas possuem a mesma potência de detecção de erros grosseiros. Crowe (1989b) obteve uma forma em particular do teste nodal que detém a propriedade da *máxima potência*. As estatísticas do teste nodal de máxima potência são dadas por:

$$z_{r,i}^* = \frac{|[V^{-1}]_i|}{\sqrt{[V]_{ii}}} \quad (4.9)$$

ou, se escrevendo na forma vetorial:

$$\mathbf{z}_r^* = [\text{diag}(\mathbf{V}^{-1})]^{1/2} \mathbf{V}^{-1} \mathbf{r} \quad (4.10)$$

O critério de teste é escolhido como sendo o mesmo do teste nodal padrão. Se existir um erro grosseiro no processo, então o valor esperado do máximo entre as estatísticas do teste dadas pela Equação 4.10 é maior que o valor esperado do máximo entre as estatísticas do teste dadas pela Equação 4.5. Isto implica que se houver um erro grosseiro, então o teste nodal baseado na estatística da Equação 4.10 tem uma maior probabilidade de detectá-lo do que o teste baseado nas estatísticas da Equação 4.5. Se as estatísticas do teste nodal são derivadas usando qualquer outra transformação linear sobre os resíduos, estas não possuem esta propriedade. Assim, o teste nodal baseado nas estatísticas da Equação 4.10 tem a propriedade da *máxima potência* - *MP*.

APLICAÇÃO 4.2 Para o processo considerado no exemplo 4.1, os resíduos das restrições e sua matriz de covariância foram calculados. A partir destes dados, as estatísticas do teste nodal são obtidas: $\left[0,687 \quad 3,0052 \quad 1,2657 \quad 1,0161 \right]$. O critério de teste em uma distribuição normal padrão, a um nível de significância de 5% é de 1,96, assim, somente o teste para o resíduo da restrição 2 é rejeitado.

4.3.3 Teste da Medida (*Measurement Test - MT*)

O terceiro teste é baseado no vetor dos ajustes às medidas

$$\mathbf{a} = \mathbf{y} - \hat{\mathbf{x}} \quad (4.11)$$

onde $\hat{\mathbf{x}}$ é o vetor das estimativas reconciliadas obtidas a partir da Equação 3.73, por exemplo. Usando esta solução, os ajustes também podem ser escritos como

$$\mathbf{a} = \Psi \mathbf{A}^T \mathbf{V}^{-1} \mathbf{r} \quad (4.12)$$

o qual segue, sob H_0 , uma distribuição normal multivariada: $N(0, \bar{W})$, onde

$$\bar{W} = cov(\mathbf{a}) = \Psi \mathbf{A}^T \mathbf{V}^{-1} \Psi \quad (4.13)$$

A seguinte estatística,

$$z_{a,j} = \frac{|a_j|}{\sqrt{\bar{W}_{jj}}}, \quad j = 1, 2, \dots, n \quad (4.14)$$

conhecida como a estatística do teste da medida, segue uma distribuição normal padrão, $N(0,1)$ sob H_0 . Tamhane (1982) mostrou que para uma matriz de covariância Ψ não diagonal, um vetor de estatísticas de teste com máxima potência para a detecção de um único erro grosseiro é obtido pela pré-multiplicação de \mathbf{a} por Ψ^{-1} , o que resulta em

$$\mathbf{d} = \Psi^{-1}\mathbf{a} \quad (4.15)$$

Sob H_0 , \mathbf{d} é também normalmente distribuído com média zero e uma matriz de covariância

$$W = cov(\mathbf{d}) = \mathbf{A}^T(\mathbf{A}\Psi\mathbf{A}^T)^{-1}\mathbf{A}^T \quad (4.16)$$

Mah e Tamhane (1982) propuseram a seguinte estatística

$$z_{d,j} = \frac{|d_j|}{\sqrt{W_{jj}}}, \quad j = 1, 2, \dots, n \quad (4.17)$$

conhecida como teste da medida de máxima potência, o qual segue uma distribuição normal padrão, $N(0,1)$ sob H_0 . De um modo similar ao teste nodal, o teste da medida também envolve múltiplos testes univariados.

APLICAÇÃO 4.3 A partir dos valores medidos e reconciliados listados na Tabela 4.1, os ajustes são computados: $\left[1,0233 \ 2,6167 \ -0,4033 \ -1,6333 \ 1,3867 \ -2,0067 \right]$. A matriz de covariância dos ajustes às medidas é dada por

$$\bar{\mathbf{W}} = \begin{bmatrix} 0,6667 & -0,1667 & -0,1667 & -0,1667 & -0,1667 & -0,3333 \\ -0,1667 & 0,6667 & 0,1667 & -0,3333 & 0,1667 & -0,1667 \\ -0,1667 & 0,1667 & 0,6667 & 0,1667 & -0,3333 & -0,1667 \\ -0,1667 & -0,3333 & 0,1667 & 0,6667 & 0,1667 & -0,1667 \\ -0,1667 & 0,1667 & -0,3333 & 0,1667 & 0,6667 & -0,1667 \\ -0,3333 & -0,1667 & -0,1667 & -0,1667 & -0,1667 & 0,6667 \end{bmatrix}$$

As estatísticas do teste da medida são então obtidas: $\left[1,2533 \ 3,2047 \ 0,494 \ 2,0004 \ 1,6983 \ 2,4577 \right]$, devido a matriz de covariância dos erros nas medidas ser diagonal neste exemplo, as estatísticas de máxima potência para o teste da medida são também as mesmas. Para um nível de significância de 5%, o critério de teste normal padrão é 1,96. A partir destas informações, os testes para as medidas 2, 4 e 6 são rejeitados. O nível de significância dado pela desigualdade de Sidak (Eq. 4.7) é igual a 0,0085, enquanto o baseado no intervalo de confiança de Bonferroni (Eq. 4.8) é igual a 0,0083. Correspondendo a estes níveis de significância modificados, os critérios de teste são, respectivamente, 2,6315 e 2,6396. Assim, se forem usados os níveis de significância modificados, somente a medida 2 é rejeitada.

4.3.4 Teste da Razão de Verossimilhança Generalizado (*Generalized Likelihood Ratio Test - GLR*)

O quarto teste para a detecção de erros grosseiros em estado estacionário é o teste da razão de verossimilhança generalizado, baseado no princípio da máxima verossimilhança. Em contraste com outros testes, a formulação deste teste requer um modelo do processo na presença de erros grosseiros, também conhecido como **modelo de erros grosseiros**. Este teste pode identificar diferentes tipos de erros grosseiros para os quais sejam fornecidos modelos.

O modelo para erros grosseiros para um viés de magnitude desconhecida b na medida j é dado por:

$$\mathbf{y} = \mathbf{x} + \varepsilon + b\mathbf{e}_j \quad (4.18)$$

onde \mathbf{e}_j é o vetor unidade com valor 1 na j -ésima posição e zero nas demais.

Por outro lado, vazamentos de material devem ser modelados como parte das restrições, posto que podem ser consideradas “correntes”. Um vazamento de fluxo mássico em um nó i do processo com uma magnitude b desconhecida, pode ser modelado como:

$$\mathbf{A}\mathbf{y} - b\mathbf{m}_i = \mathbf{c} \quad (4.19)$$

Os elementos do vetor \mathbf{m}_i são relativamente fáceis de definir quando somente balanços com fluxos totais estão envolvidos. Se o vazamento vem de uma unidade i , então somente a restrição de fluxo para este vetor da unidade é afetado e assim \mathbf{m}_i é idêntico a \mathbf{e}_i . Contudo, se as restrições também incluem balanços de componentes e balanços de energia (com valores conhecidos de composição e temperatura), então o vetor \mathbf{m}_i só pode ser definido aproximadamente a partir de uma decisão arbitrária. Uma recomendação de Narasimham e Mah (1987) é escolher os elementos de \mathbf{m}_i da seguinte forma:

1. Correspondendo à restrição de fluxo de massa total da unidade i , \mathbf{m}_i tem o valor de 1 na i -ésima posição;
2. Correspondendo à restrição de fluxo de energia associada ao nó i , o valor do i -ésimo elemento de \mathbf{m}_i pode ser escolhido como a entalpia específica média das correntes que incidem no nó i . O mesmo pode ser aplicado à restrição do fluxo de um componente para o nó i , substituindo a entalpia específica pela concentração;
3. Os elementos em \mathbf{m}_i não associados com restrições sobre o nó i , são escolhidos como iguais a zero.

Perdas de energia e de fluxos de componentes no nó i também podem ser modeladas pela Equação 4.19, escolhendo-se o elemento correspondente em \mathbf{m}_i igual a 1 e todos os outros elementos iguais a zero.

É possível se derivar a distribuição estatística dos resíduos das restrições sob H_1 , utilizando-se modelos de erros grosseiros quando um erro deste tipo está presente entre as medidas ou nas restrições. Os resíduos das restrições seguem uma distribuição normal sob H_0 , com média zero e matriz de covariância dada pela Equação 4.3. Sob H_1 , os resíduos das restrições continuam seguindo uma distribuição normal com matriz de covariância dada pela eq. 4.3, mas o valor esperado depende do tipo de erro grosseiro presente. Se um erro grosseiro devido a um viés de magnitude b está presente na medida j , então:

$$E[\mathbf{r}] = b\mathbf{Ae}_j \quad (4.20)$$

Por outro lado, se um erro grosseiro devido a um vazamento no processo está presente nas medidas no nó i , então:

$$E[\mathbf{r}] = b\mathbf{m}_i \quad (4.21)$$

Portanto, quando um erro grosseiro devido a um viés ou vazamento no processo estiver presente, pode-se definir:

$$E[\mathbf{r}] = b\mathbf{f}_k \quad (4.22)$$

onde

$$\mathbf{f}_k \begin{cases} \mathbf{Ae}_j & \text{para um viés na medida } i \\ \mathbf{m}_i & \text{para um vazamento no nó } i \end{cases} \quad (4.23)$$

Os vetores \mathbf{f}_k são também conhecidos como vetores de assinatura de erro grosseiro.

Se define-se μ como o valor esperado desconhecido de \mathbf{r} , pode-se formular a hipótese de detecção de erros grosseiros como:

$$\begin{aligned} H_0 : \mu &= 0 \\ H_1 : \mu &= b\mathbf{f}_k \end{aligned} \quad (4.24)$$

onde H_0 é a hipótese nula que nenhum erro grosseiro esteja presente e H_1 é a hipótese alternativa de que esteja presente ou um viés de medida ou um vazamento no processo. A hipótese alternativa tem duas incógnitas, b e \mathbf{f}_k . O parâmetro b pode ser qualquer número real e \mathbf{f}_k pode ser qualquer vetor do conjunto F dado por:

$$F = \{\mathbf{A}\mathbf{e}_j, \mathbf{m}_i : i = 1, \dots, m : j = 1, \dots, n\} \quad (4.25)$$

onde m é o número de nós ou unidades do processo e n é o número de variáveis mensuradas.

No sentido de testar as duas hipóteses dadas pela Eq. 4.24, pode-se usar o teste da razão da verossimilhança. Este teste, para este caso é dado por:

$$\lambda = \sup \frac{P\{\mathbf{r}|H_1\}}{P\{\mathbf{r}|H_0\}} \quad (4.26)$$

onde $P\{\mathbf{r}|H_0\}$ e $P\{\mathbf{r}|H_1\}$ são as probabilidades de se obter o vetor residual \mathbf{r} sob as hipóteses H_0 e H_1 . O supremo (sup na Equação 4.26) é computado sobre todos os valores possíveis dos parâmetros presentes nas hipóteses. Usando-se uma função de densidade de probabilidade normal para \mathbf{r} , pode-se reescrever a Equação 4.26 como:

$$\lambda = \sup_{b, \mathbf{f}_k} \frac{e^{-0,5(\mathbf{r}-b\mathbf{f}_k)^T \mathbf{V}^{-1}(\mathbf{r}-b\mathbf{f}_k)}}{e^{-0,5\mathbf{r}^T \mathbf{V}^{-1}\mathbf{r}}} \quad (4.27)$$

como o lado direito da Equação 4.27 é sempre positivo, pode-se simplificar o cálculo através da escolha da estatística:

$$T = 2 \ln \lambda = \sup_{b, \mathbf{f}_k} [\mathbf{r}^T \mathbf{V}^{-1} \mathbf{r} - (\mathbf{r} - b \mathbf{f}_k)^T \mathbf{V}^{-1} (\mathbf{r} - b \mathbf{f}_k)] \quad (4.28)$$

A contabilidade de T é realizada da seguinte forma: para qualquer vetor \mathbf{f}_k , calcula-se a estimativa b^* de b , que dá o supremo na Equação 4.28. Assim, obtém-se uma estimativa de máxima verossimilhança:

$$b^* = (\mathbf{f}_k^T \mathbf{V}^{-1} \mathbf{f}_k)^{-1} (\mathbf{f}_k^T \mathbf{V}^{-1} \mathbf{r}) \quad (4.29)$$

substituindo b^* na Equação 4.28 e denotando o valor correspondente de T por T_k , tem-se

$$T_k = d_k^2 / C_k \quad (4.30)$$

onde

$$d_k = \mathbf{f}_k^T \mathbf{V}^{-1} \mathbf{r} \quad (4.31)$$

$$C_k = \mathbf{f}_k^T \mathbf{V}^{-1} \mathbf{f}_k \quad (4.32)$$

Este procedimento é realizado para cada vetor \mathbf{f}_k no conjunto F e a estatística T é portanto obtida como:

$$T = \sup_k T_k \quad k = 1, \dots, m+n \quad (4.33)$$

Seja \mathbf{f}^* o vetor que leva ao supremo na Eq. 4.33. A estatística T é comparada com um valor de referência T_{cr} pré-determinado e um erro grosseiro é detectado se T excede T_{cr} . Pode-se interpretar T_k como uma estatística para testar a presença do erro grosseiro k . Como T é o máximo entre os T_k , o teste *GLR* detecta um erro grosseiro se qualquer das estatísticas T_k exceder o valor crítico. Assim o teste *GLR*, como o teste da medida

e o teste nodal, realiza múltiplos testes univariados para detectar um erro grosseiro. A distribuição de T_k , sob H_0 , é uma distribuição χ^2 central com um grau de liberdade. Desta forma, com o objetivo de manter a probabilidade do Erro do Tipo I do teste GLR menor ou igual a um dado valor α , pode-se escolher um critério de teste como $\chi^2_{1-\beta,1}$, o quantil $1 - \beta$ da distribuição χ^2 com um grau de liberdade, onde β é dado por:

$$\beta = 1 - (1 - \alpha)^{\frac{1}{m+n}} \quad (4.34)$$

APLICAÇÃO 4.4 *Considerando-se erros grosseiros causados por viéses nas medidas para o processo considerado nas aplicações anteriores, o vetor de assinatura para os erros grosseiros para um viés numa medida i é a i -ésima coluna da matriz de restrições mostrada na aplicação 4.1. As estatísticas do teste GLR computadas a partir dos resíduos das restrições e sua matriz de covariância dada na aplicação 4.2 são $\left[1,5708 \quad 10,2704 \quad 0,244 \quad 4,0017 \quad 2,8843 \quad 6,0401 \right]$. As estatísticas do teste GLR são o quadrado dos testes da medida de máxima potência, computados na aplicação 4.3. O critério de teste a um nível de significância de 5% e a dos níveis modificados de significância - Sidak e Bonferroni, são simplesmente o quadrado do critério de teste normal padrão $\left[3,8415 \quad 6,925 \quad 6,9676 \right]$, respectivamente. Assim os testes GLR para as medidas 2, 4 e 6 são rejeitadas a um nível de significância de 5% enquanto que somente o teste para a medida 2 é rejeitado sob os níveis de significância modificados.*

Se se deseja também testar a presença de vazamentos em todos os quatro nós, então os vetores de assinatura para estes quatro erros grosseiros são simplesmente os vetores unidade. As estatísticas do teste GLR para estes quatro erros grosseiros são dadas por $\left[1,5708 \quad 13,2496 \quad 0,3844 \quad 6,0401 \right]$. Os testes GLR para vazamentos nos nós 2 e 4 são rejeitados a um nível de significância de 5% enquanto o teste para vazamento sob os níveis de significância modificados rejeitam somente o nó 2.

4.4 Testes Estatísticos para Modelos em Estado Estacionário

Nas seções anteriores, diferentes testes estatísticos para detecção de erros grosseiros foram descritos e implementados para o caso mais simples no qual todas as variáveis são medidas diretamente. De um modo geral, como já foi explanado no Capítulo 3, podem estar presentes variáveis não medidas e/ou as medidas podem ser indiretamente relacionadas às variáveis. Narasimham e Mah (1989) descreveram transformações simples através das quais os modelos gerais em estado estacionário podem ser convertidos no modelo simples acima mencionado.

Usando-se estas transformações, todos os testes estatísticos podem ser derivados como descrito a seguir.

Se variáveis não medidas estão presentes, o modelo de restrições é descrito por:

$$\mathbf{A}_1 \mathbf{x} + \mathbf{A}_2 \mathbf{u} = \mathbf{c} \quad (4.35)$$

onde $\mathbf{x} : n \times 1$ é o vetor das variáveis mensuradas, $\mathbf{u} : p \times 1$ é o vetor das variáveis não medidas e \mathbf{A}_2 tem o posto completo, p .

Como mostrado no Capítulo 3, as variáveis não medidas podem ser eliminadas pela pré-multiplicação das restrições por uma matriz de projeção $\mathbf{P} : (m - p) \times m$, de posto $m - p$, onde m é o número de restrições, de modo a gerar as restrições reduzidas

$$\mathbf{P} \mathbf{A}_1 \mathbf{x} = \mathbf{P} \mathbf{c} \quad (4.36)$$

Os resíduos das restrições para o conjunto reduzido de restrições é definido em uma forma exatamente análoga à Equação 4.2:

$$\rho = \mathbf{P}(\mathbf{A}_1 \mathbf{y} - \mathbf{c}) \quad (4.37)$$

A matriz de variância-covariância do vetor ρ é:

$$\mathbf{V}_\rho = \text{cov}(\rho) = \mathbf{P}\mathbf{A}_1\Psi(\mathbf{P}\mathbf{A}_1)^T \quad (4.38)$$

As estatísticas para os testes global, nodal e da medida podem ser obtidas usando $\mathbf{P}\mathbf{A}_1$, ρ e \mathbf{V}_ρ no lugar de \mathbf{A} , \mathbf{r} e \mathbf{V} , respectivamente, nas equações apropriadas. Para se derivar estatísticas para os testes GLR, os vetores de assinatura de erros grosseiros para viéses e vazamentos são também transformados pela matriz de projeção.

Este vetor de assinatura transformado é dado por:

$$\mathbf{f}_{\rho k} = \mathbf{P}\mathbf{f}_k \quad (4.39)$$

onde \mathbf{f}_k é dado pela Equação 4.23. As estatísticas do teste GLR são então obtidas usando as equações 4.30 até a 4.33, substituindo-se \mathbf{f}_k , \mathbf{V} e \mathbf{r} por $\mathbf{f}_{\rho k}$, \mathbf{V}_ρ e ρ , respectivamente.

Em alguns casos, as medidas não podem ser diretamente relacionadas às variáveis como na Equação 3.2. Por exemplo, na Equação 3.19, onde a queda de pressão é relacionada ao quadrado da variável de taxa de fluxo. Outro exemplo é a relação entre uma medida de pH e a concentração de íons de hidrogênio e talvez a temperatura do processo. Estas relações são tipicamente não lineares, mas por simplicidade serão representadas aqui por equações lineares:

$$\mathbf{y} = \mathbf{D}\mathbf{x} + \boldsymbol{\varepsilon} \quad (4.40)$$

assumindo-se as restrições sejam dadas por

$$\mathbf{A}\mathbf{x} = \mathbf{c} \quad (4.41)$$

define-se as variáveis artificiais \mathbf{x}_a como

$$\mathbf{x}_a = \mathbf{D}\mathbf{x} \quad (4.42)$$

então a Equação 4.40 se torna

$$\mathbf{y} = \mathbf{x}_a + \boldsymbol{\varepsilon} \quad (4.43)$$

As equações 4.41 e 4.42 podem ser escritas em conjunto como

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \mathbf{x}_a + \begin{bmatrix} \mathbf{A} \\ -\mathbf{D} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{c} \\ \mathbf{0} \end{bmatrix} \quad (4.44)$$

As equações 4.43 e 4.44 representam um modelo alternativo equivalente do processo no qual as variáveis \mathbf{x} seriam “variáveis não medidas” e as variáveis \mathbf{x}_a seriam as “variáveis diretamente medidas”. Portanto, o método descrito para o tratamento de variáveis não medidas pode ser usado para derivar estatísticas para todos os testes.

A técnica descrita acima pode ser aplicada mesmo quando as medidas são relacionadas às variáveis através de equações não-lineares. Contudo, as equações de restrição modificadas serão não lineares e assim técnicas de reconciliação e de detecção de erros grosseiros não lineares terão que ser usadas para resolução destes processos.

4.5 Técnicas para Identificação de um Erro Grosseiro

O segundo componente de uma estratégia de detecção de erros grosseiros trata do problema de identificar corretamente o tipo e a localização de um erro grosseiro que seja detectado por um teste. É importante notar que o problema de identificação somente surge se os testes de detecção rejeitam a hipótese nula. Nem todos os testes de detecção descritos nas seções anteriores são projetados para distinguir os diferentes tipos de erros grosseiros. Somente o teste GLR é adequado para tanto porque este usa informação concernente ao efeito de cada tipo de erro grosseiro sobre o modelo do processo.

No sentido de se comparar as diferentes técnicas desenvolvidas em conjunto com os diferentes testes para identificação de erros grosseiros e se obter uma visão geral das interrelações, esta seção se restringirá a considerações sobre erros causados por vieses nas medidas e somente a um único erro grosseiro presente entre as medidas. Neste caso, o problema de identificação se reduz a simplesmente identificar corretamente a medida que contém o erro grosseiro.

As técnicas para identificação da medida contendo o erro vão de uma regra simples até estratégias complexas, dependendo do teste que é usado. O teste da medida e o teste GLR, em virtude do modo como são derivadas suas estatísticas, usam uma regra simples para identificar o erro grosseiro. O teste GLR e o teste da medida de máxima potência são idênticos considerando-se erros grosseiros causados somente por vieses. Neste caso, há uma estatística correspondente a cada medida. A regra de identificação usada nestes testes pode ser formulada como “*Identificar o erro grosseiro na medida que corresponda à máxima estatística que exceda ao critério de teste*”.

Devido à simplicidade da regra acima, diz-se que os testes GLR e da medida não necessitam de uma estratégia propriamente dita para identificação de erros grosseiros.

4.5.1 Estratégia de Eliminação Serial para Identificação de Erros Grosseiros Singulares

Quando se usa o teste global para a detecção de erros grosseiros, uma estratégia comparativamente mais complexa tem que ser usada.

Ripps (1965) foi quem primeiro delineou um procedimento que foi posteriormente estudado e refinado por Serth e Heenan (1986) e Rosenberg *et al.* (1987). Este procedimento é conhecido como *procedimento de eliminação serial*. Nele, cada medida é deletada por vez e a estatística do teste global é recomputada. Quando se elimina uma medida, faz-se a variável correspondente passar por não medida, assim o teste global tem que ser recomputado usando o conjunto deduzido das restrições.

Devido ao aumento no número de variáveis não medidas, o valor da função objetivo

e, por conseguinte, a estatística do teste global, vai diminuir. Ripps (1965) sugeriu que o erro grosseiro pode ser identificado naquela medida cuja eliminação leva à maior redução na função objetivo.

Ao invés de resolver o problema de reconciliação de dados repetidamente ou computar a projeção de matrizes para a eliminação de cada medida por vez, Crowe (1988) derivou expressões simplificadas para a redução do valor da função objetivo da reconciliação de dados com a eliminação da medida i . Esta expressão é dada por:

$$\Delta J_i = J - J_i = \frac{(\mathbf{e}_i \mathbf{A}^T \mathbf{V}^{-1} \mathbf{r})^2}{\mathbf{e}_i \mathbf{A}^T \mathbf{V}^{-1} \mathbf{e}_i} \quad (4.45)$$

Pode ser prontamente verificado que a redução no valor da função objetivo através da eliminação da medida i é igual à estatística do teste GLR (ou o quadrado da estatística do teste da medida) para a variável i . Isto implica que se a regra usada em conjunto com o teste global for “*Identificar o erro grosseiro na medida que dá o máximo ΔJ_i* ”, então isto é precisamente a mesma regra usada no teste GLR (ou no teste da medida de máxima potência) para identificação do erro grosseiro na medida correspondente à máxima estatística. Em outras palavras, o teste global em combinação com a estratégia de eliminação serial é equivalente ao teste GLR.

Tabela 4.2: Redução na estatística do teste global com a eliminação da medida i (NARASIMHAN; JORDACHE, 2000)

Medida Eliminada	Redução na Estatística do Teste Global
1	1,571
2	10,27
3	0,244
4	4,002
5	2,884
6	6,040

APLICAÇÃO 4.5 *A partir dos resultados das aplicações 4.3 e 4.4, observa-se que quando se escolhe identificar um erro grosseiro na medida correspondente à máxima estatística do teste, então o erro grosseiro na medida 2 é corretamente identificada tanto pelo teste da medida de máxima potência e pelo teste GLR. De fato mesmo considerando-se erros grosseiros devidos a vazamentos, a maior estatística do teste GLR corresponde a um viés na medida 2. O teste global também rejeita a hipótese nula e então pode se utilizar a estratégia de eliminação de variáveis para identificar a localização do erro grosseiro. A Tabela 4.2 mostra a redução na estatística do teste global quando diferentes medidas são eliminadas.*

Como a maior redução na estatística do teste global é obtida quando a medida 2 é eliminada, o procedimento de eliminação em conjunto com o teste global identificou o erro corretamente.

4.6 Conclusão

Com a teoria vista neste e no capítulo anterior, está exposta a fundamentação teórica necessária para o desenvolvimento das rotinas de tratamento de dados, exploradas no próximo capítulo.

A implementação prática da detecção de erros grosseiros se dará através de um método combinado, proposto por Yang *et al.* (1995) e também está descrito no próximo capítulo.

5 *Programas Desenvolvidos*

As soluções, eu já as possuo há muito tempo, mas ainda não sei como cheguei a elas.

Johann C. F. Gauss

5.1 Introdução

As aplicações das teorias apresentadas nos Capítulos 3 e 4 são relatadas neste Capítulo, detalhando-se o fluxo de informação desde o dado cru, que pode ser proveniente de uma medição verdadeira, como pode ser gerado com adição de ruído a um “valor real” arbitrado, até a disponibilização final das estimativas.

Nem sempre se dispõe de dados reais de operação (*on* ou *off line*) para se testar rotinas de reconciliação e retificação de dados. Narasimhan e Jordache (2000) sugerem um procedimento para contornar esta dificuldade e, assim, se poder testar a efetividade das rotinas e a redução de erros que pode ser alcançada. O procedimento, que é usado neste trabalho, precisa dos seguintes dados de entrada:

1. O *flowsheet* do processo, o qual indica o número de unidades, as correntes e suas conexões. A natureza da unidade do processo não precisa ser especificado se a simulação é sobre fluxos globais;
2. Os valores “reais” ou nominais de fluxo no estado estacionário de todas as correntes. Estes valores “verdadeiros” são úteis no julgamento sobre o aumento de

acurácia alcançado através da reconciliação de dados;

3. Um conjunto de fluxos mensurados do processo e o desvio padrão do erro em cada medida.

Em primeiro lugar, são gerados erros aleatórios os quais seguem uma distribuição normal com média zero e um dado desvio padrão. Estes erros são adicionados aos valores “verdadeiros” no sentido de se obter “medidas” simuladas. A matriz de restrições, \mathbf{A} , é obtida baseada na informação sobre a conectividade do processo e as submatrizes \mathbf{A}_x e \mathbf{A}_u , são também obtidas, correspondendo às variáveis mensuradas e não mensuradas, respectivamente.

A Figura 5.1 descreve as etapas gerais do tratamento de dados desenvolvido neste trabalho: definição do problema (com acesso opcional a uma fonte de dados), seguido do ciclo de reconciliação para finalmente a exposição dos resultados

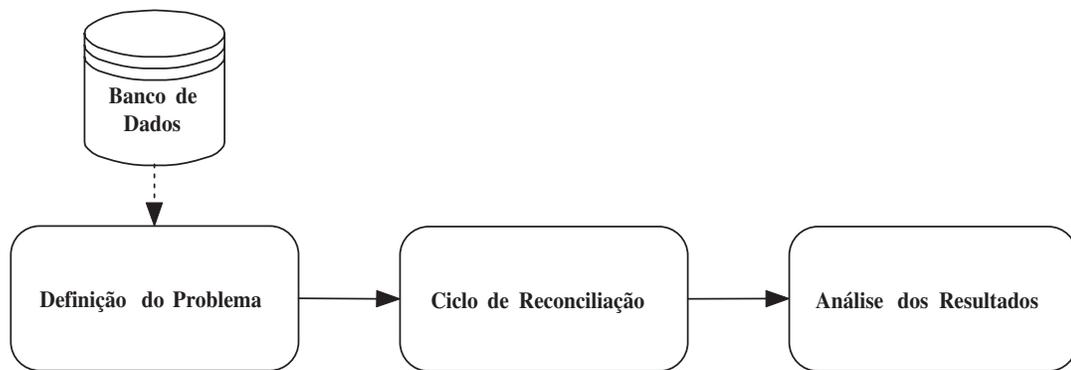


Figura 5.1: Etapas do tratamento de dados

5.2 Descrição do Problema

A primeira etapa no diagrama da Figura 5.1 é a descrição do problema de reconciliação de dados. O procedimento usado neste trabalho para esta descrição é ilustrado na Figura 5.2.

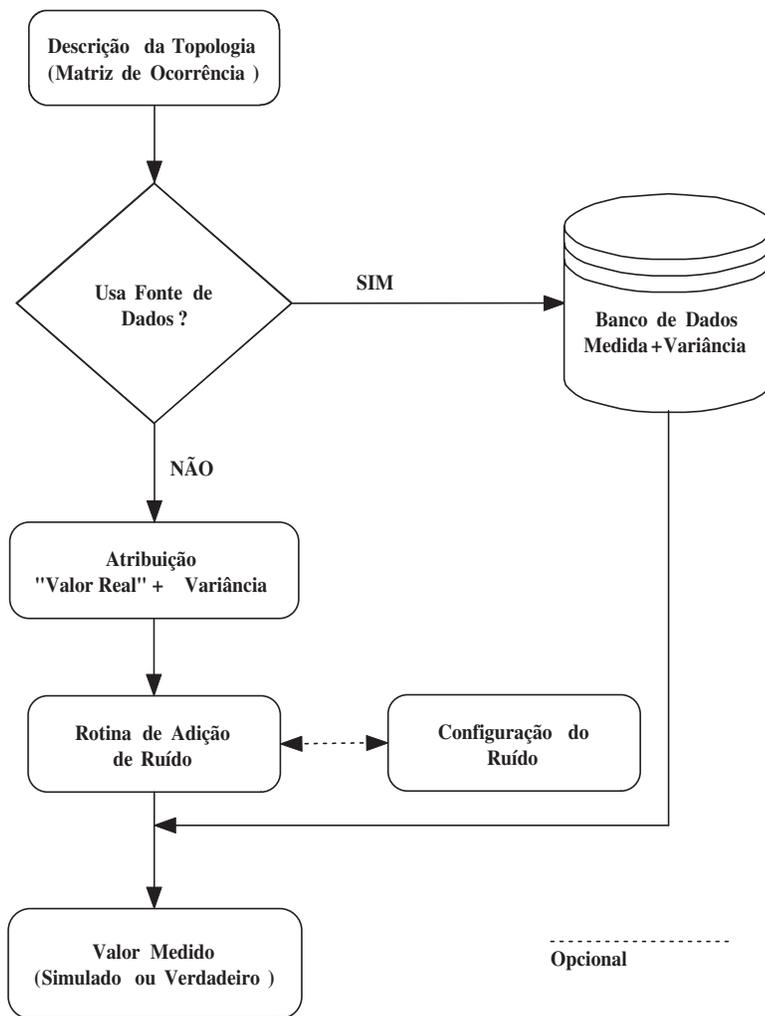


Figura 5.2: Geração/disponibilização de dados simulados/medidos

Um dos objetivos deste trabalho é a criação de um ambiente visual para facilitar a definição do problema de reconciliação, ou seja, a representação da topologia do processo, unidade ou equipamento o qual se deseja reconciliar os dados.

Um ambiente com interface gráfica com o usuário (**GUI** - *Graphical User Interface*)¹ guia o processo de definição, diminuindo e, até certo ponto, impedindo os erros na representação de uma topologia. Uma topologia (matriz de ocorrência) errada leva

¹Ver Apêndice A

a uma solução sem conexão com o problema original ou, até mesmo, à impossibilidade de solução.

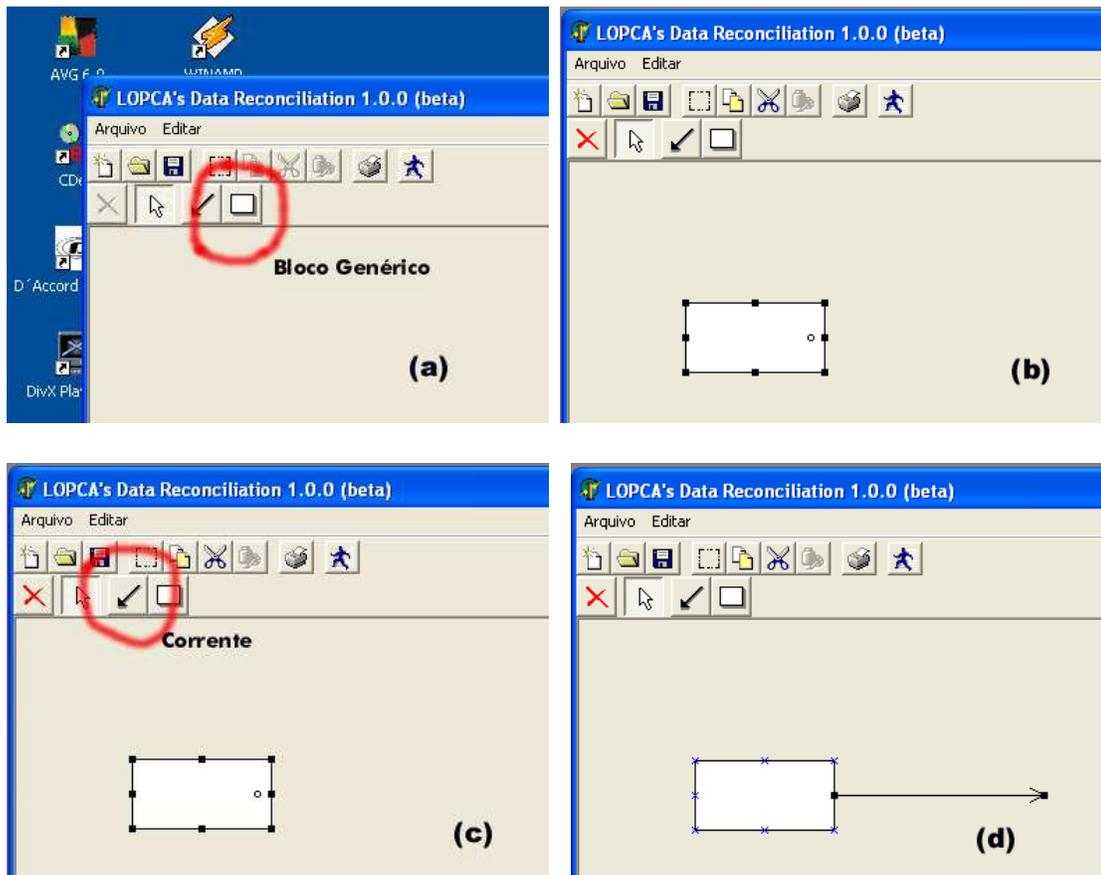


Figura 5.3: Seqüência de definição do problema de reconciliação

Por se tratar de reconciliação de dados linear em estado estacionário, a representação de qualquer equipamento, ou conjunto de equipamentos, é feita através de blocos genéricos para onde afluem e de onde saem correntes carregando a informação de vazões mássicas. A seqüência de definição do problema está ilustrada na Figura 5.3. Inicialmente seleciona-se o ícone correspondente a um bloco genérico de processo na barra de ferramentas e depois clica-se na área de desenho para colocá-lo lá. As correntes são adicionadas selecionando-se o ícone correspondente à corrente na barra de ferramentas e, do mesmo modo que o bloco, são adicionadas clicando-se na área de desenho. As correntes devem ser conectadas aos blocos em pontos especiais que estão em evidência

no seu contorno (*link points*).

Uma vez que o processo tenha sido reproduzido, as propriedades de cada corrente devem ser acessadas através de um clique duplo na mesma. A janela para entrada de dados está ilustrada na Figura 5.4.

Uma outra facilidade que foi implementada no *software* foi a possibilidade de se salvar e recuperar em arquivo tanto a topologia quanto os dados associados às correntes, bem como opções de execução do programa. Isto permite grande praticidade na comparação de problemas de reconciliação.

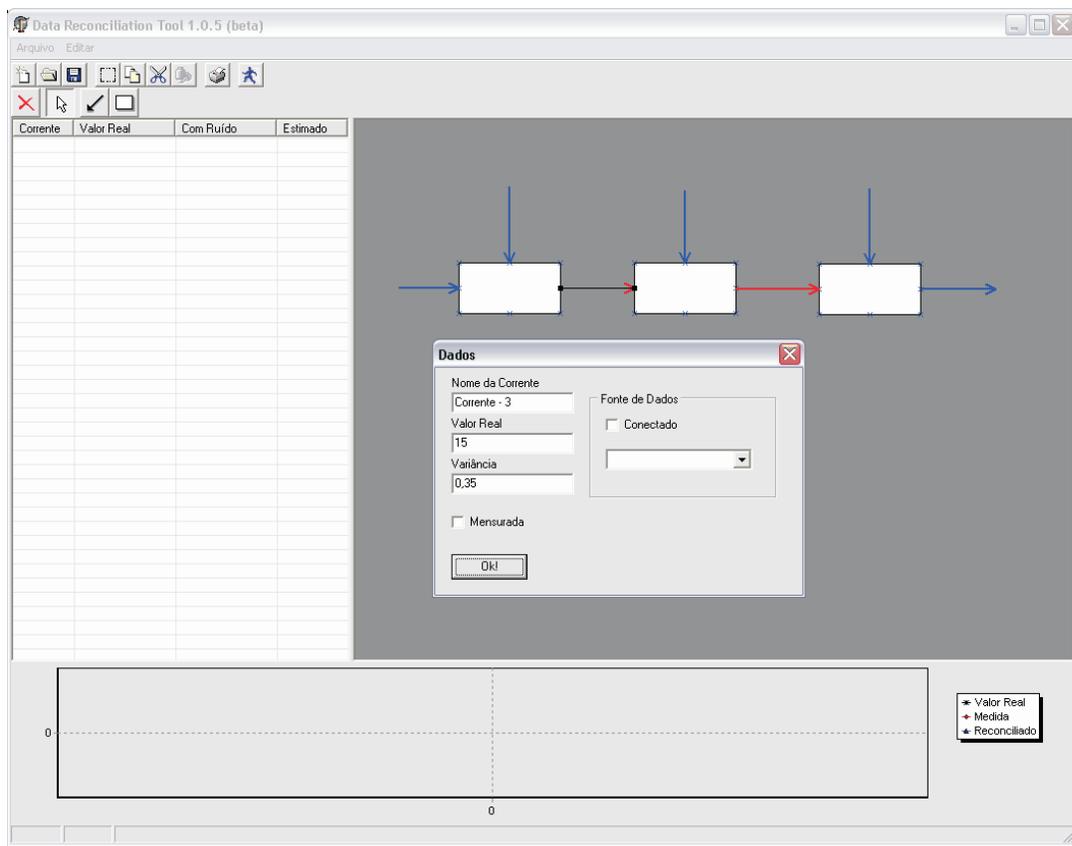


Figura 5.4: Entrada de dados da corrente (nome da corrente, valor real e variância)

5.2.1 Geração de Ruído

A rotina para geração de dados com ruído provê a adição de um erro aleatório a partir da seguinte fórmula:

$$\mathbf{y} = \mathbf{x} + (\text{RAN} \times 2 - 1.0) \times \text{Amp} \times \sigma \quad (5.1)$$

onde \mathbf{y} é o valor “medido” (simulando uma leitura), \mathbf{x} é o valor verdadeiro (a partir do modelo), RAN é uma função de geração de números aleatórios na faixa $0 \leq \text{RAN} \leq 1$, σ é o desvio padrão da variável e Amp é a amplitude dentro da qual se deseja gerar ruído. Na maioria dos testes realizados, usou-se uma amplitude de 2 a 3 desvios padrão.

5.3 Ciclo de Reconciliação

Nesta etapa do fluxo de informações, são aplicadas as rotinas de reconciliação de dados e detecção de erros grosseiros, conforme é mostrado no fluxograma da Figura 5.5.

Esta etapa se inicia com o pré-condicionamento dos dados, seguido da primeira fase de reconciliação. Os dados passam então por uma detecção de erros grosseiros e se esta apresentar medidas suspeitas dentre os dados, estas são retiradas e passam a integrar o rol das variáveis não medidas, passando-se, então, por uma segunda fase de reconciliação e assim por diante enquanto houver condições necessárias para a resolução do problema de reconciliação de dados: redundância entre um conjunto de medidas. Se um determinado problema não apresenta tais condições é mostrada uma mensagem ao usuário, informando-o desta situação.

5.3.1 Pré-condicionamento

Segundo o que é sugerido em Narasimhan e Jordache (2000), os dados devem ser pré-condicionados antes dos procedimentos de reconciliação. No caso específico abordado neste trabalho, a reconciliação linear em estado estacionário, e levando-se em

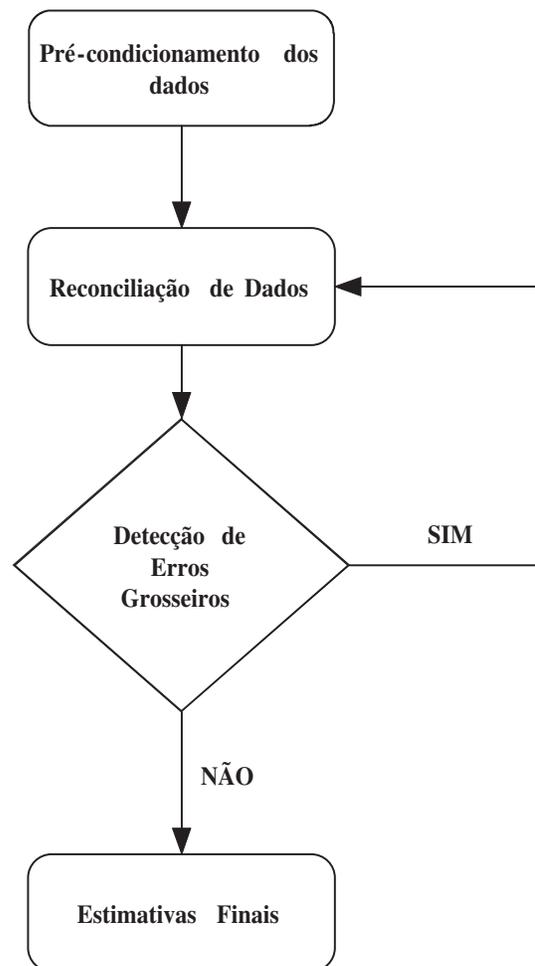


Figura 5.5: Algoritmo do procedimento de reconciliação de dados

conta a finalidade de reconciliação para otimização, optou-se por pré-condicionar os dados corrompidos através de uma média simples de 50 valores de cada variável “medida”. Outra possibilidade de pré-condicionamento de dados é o uso de filtros.

5.3.2 Reconciliação

O cerne deste *software* é a rotina de reconciliação de dados. Foi codificada uma subrotina de reconciliação com duas funções básicas: reconciliação de um conjunto de variáveis todas medidas e reconciliação com coaptação de dados sobre um conjunto de

variáveis medidas e não medidas. A abordagem para decomposição do problema de reconciliação e resolução do problema de mínimos quadrados resultante foi a fatoração QR.

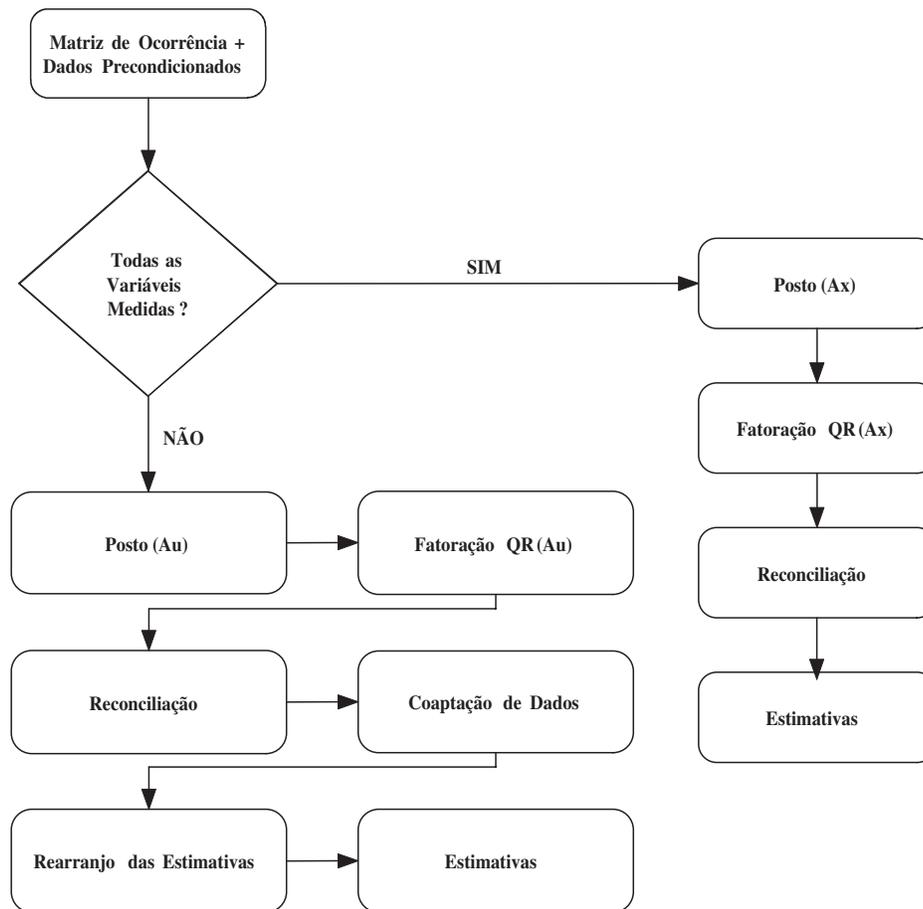


Figura 5.6: Fluxograma da rotina de reconciliação de dados

Conforme pode ser visto no fluxograma da Figura 5.6, um conjunto básico de ações é realizado similarmente para os dois casos: retorno do posto da matriz pertinente, fatoração QR, procedimentos de reconciliação (solução da minimização) e disponibilização das estimativas.

Quando há variáveis não medidas, estas são analisadas quanto à observabilidade e redundância e, dependendo da configuração, o problema de coaptação de dados pode ser resolvido na íntegra (estimativa de todas as variáveis não medidas), parcialmente

resolvido ou não ser resolvido por completo.

Os ajustes sobre as “medições” é armazenado com a finalidade de ser analisado posteriormente pelas rotinas de detecção de erros grosseiros.

5.3.3 Detecção de Erros Grosseiros

A estratégia de detecção de erros grosseiros usada neste trabalho baseia-se em um procedimento proposto por Yang *et al.* (1995), de se combinar o Teste da Medida (*Measurement Test*) com o Teste Nodal (*Nodal Test*). A vantagem do MT é que este aponta diretamente a localização dos erros grosseiros, mas tem como desvantagem a tendência de apontar erros grosseiros extra (erro do tipo I). O NT não espalha os erros grosseiros pelos dados, mas tem como desvantagem a tendência de incorrer no erro do tipo II². A combinação de ambos os métodos cancela mutuamente as desvantagens e potencializa as respectivas qualidades.

O procedimento pode ser assim descrito:

1. Encontra-se os nós³ que ligam a corrente que tem o maior Z_a (MT);
2. Verifica-se estes nós através do NT, comparando os Z_r com valor crítico, Z_{rc} ;
3. Se o Z_r for menor que Z_{rc} , isto significa que a corrente está descartada como portadora de erro grosseiro. Retorna-se ao MT e se escolhe o segundo maior Z_a ;
4. Se um ou dois Z_r excedem Z_{rc} , isto significa que pelo menos uma das correntes ligando este nó contém um erro grosseiro e a primeira suspeita é a corrente que recebeu o maior ajuste relativo (a_i/y_i). Verifica-se outro nó em contato com esta corrente Z_r . Se for menor que Z_{rc} , passa-se ao segundo maior ajuste relativo e se este resultar em um valor maior que o valor crítico Z_{rc} , então um erro grosseiro foi encontrado;

²Ver Seção 4.3

³Unidade do Processo

5. Move-se o erro grosseiro fora do conjunto das medidas (para o conjunto das não medidas);
6. Passa-se ao segundo maior Z_a que exceda $Z_{a,c}$ e assim por diante.

5.4 Apresentação e Análise dos Resultados

No *software* desenvolvido, a apresentação dos resultados consiste em dados numéricos e gráficos, expostos na mesma interface com o usuário. A disponibilização dos resultados na mesma ferramenta de descrição e resolução do problema de reconciliação faz com que o aplicativo desenvolvido seja quase que auto-suficiente, não sendo necessário recorrer a outras ferramentas para a análise do desempenho das rotinas desenvolvidas.

Os dados numéricos estão nas colunas à esquerda do aplicativo, ordenados por **nome da corrente, valor “real”, valor com ruído e valor estimado** (Figura 5.7).

Na parte inferior do aplicativo fica a área de gráficos, onde são plotados os gráficos para o valor “medido”, o valor “real” e a estimativa da reconciliação. O gráfico pertinente à corrente é exposto quando se seleciona a corrente desejada na lista da esquerda do aplicativo.

Há também as funcionalidades de exportar os gráficos e a área de descrição do problema (*flowsheet*) para arquivos de imagens.

5.5 Sistema de Teste de Processos Distribuído em Rede: Aplicação para Integração de Processos em Tempo Real

Neste trabalho, foi proposta a criação de rotinas genéricas para a resolução de problemas de reconciliação de dados, detecção e identificação de erros grosseiros e, como

Uma dos conceitos centrais deste sistema de testes é possibilitar ao pesquisador que este se concentre em seu objetivo principal, seja modelagem, controle ou otimização, para citar os mais comuns, evitando retrabalho quando for realizar algum tipo de teste usando partes desenvolvidas por terceiros. Outra vantagem é que, usando-se o computador para simular cada parte em modo dedicado, tem-se um maior desempenho.

Apesar de esta estrutura ter sido projetada visando funcionamento em rede, pode-se operar todos os executáveis em um mesmo computador, ou ainda distribuir os programas entre os computadores da maneira que se desejar, o que evidencia sua flexibilidade.

Inicialmente, este sistema de testes tem duas possíveis abordagens. Numa delas, diferentes computadores conectados a uma rede, o que possibilita inclusive o uso da *Internet* com tais finalidades, trocam dados diretamente entre si como ilustrado na figura 5.8.

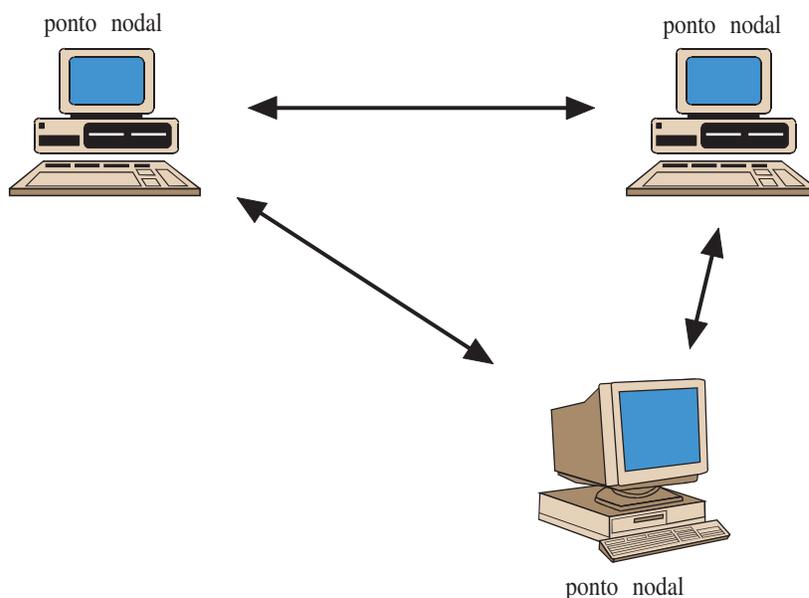


Figura 5.8: Abordagem de troca direta de dados

Esta abordagem tem como vantagem a rapidez na troca de dados entre as unidades e como desvantagem o esforço para o projeto de cada unidade, haja vista a necessidade de se sincronizar eventos de modo que uma informação não seja processada estando

incompleta e que o envio de um bloco de dados dispare o processamento na outra parte. A esta abordagem foi dado o nome de **Abordagem em um Único Plano**, pois o processamento e o controle do fluxo de informações se dão em um único plano.

A abordagem em um único plano é toda montada sobre protocolo de rede TCP/IP e cada nó de processamento (modelo, controle, otimização, reconciliação, corrupção de dados, etc.) é responsável por enviar e capturar blocos de dados no momento correto e com o já citado sincronismo.

A outra abordagem foi batizada de **Abordagem em dois Planos** ou **Abordagem via Banco de Dados**. Nesta, o plano do processamento é separado do plano onde o fluxo de informação é controlado, ficando esta última tarefa para um servidor de Banco de Dados Relacional Cliente-Servidor. Esta abordagem é ilustrada pela figura 5.9.

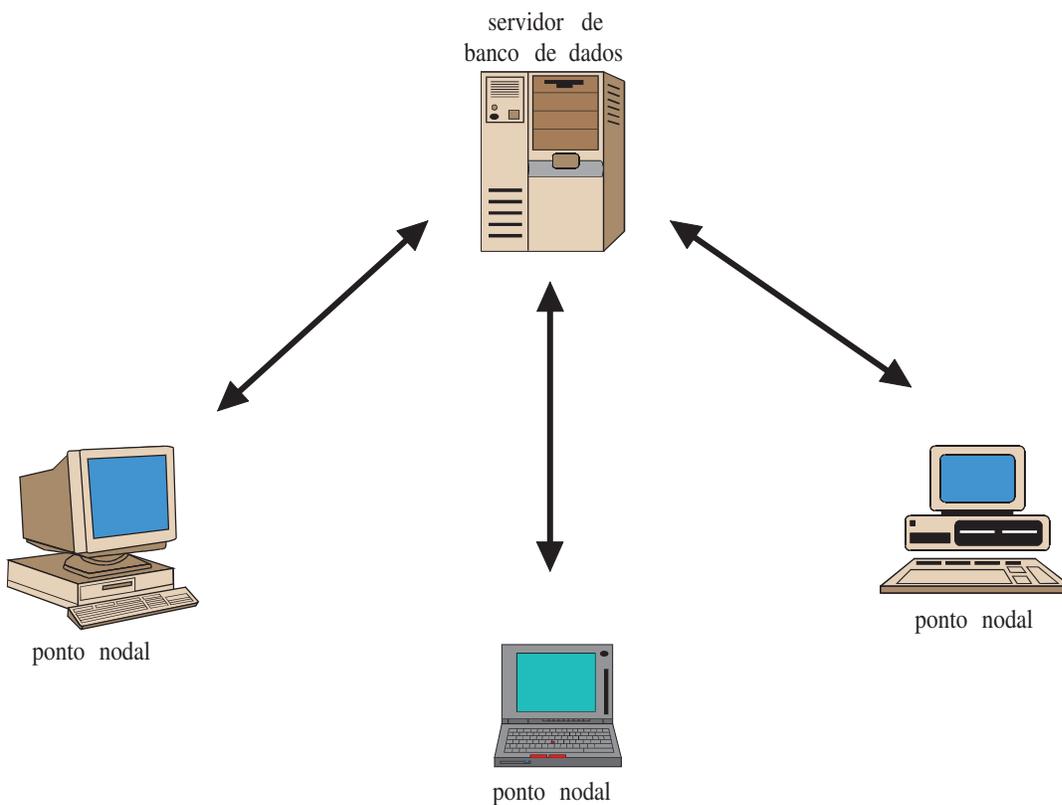


Figura 5.9: Abordagem via banco de dados

Esta abordagem facilita muito o projeto de integração pois pode-se abstrair a complexidade de cada nó de processamento de modo a enxergá-los como “caixas pretas” com uma interface pública através da qual entram e saem informações.

Com este esquema, também é fácil se armazenar um histórico do processamento e se retornar informações a respeito do estado de cada nó. Isto é útil para, por exemplo, monitorar se a temperatura de um reator ultrapassou o limite máximo, não havendo mais sentido em se prosseguir com os cálculos.

Dada a praticidade com a qual os dados são armazenados e retornados, fica fácil a exposição em tempo de execução dos resultados em gráficos, o que favorece muito a sensibilidade sobre os casos em estudo, posto que se pode observar instantaneamente efeitos de alterações realizadas pelo usuário.

O Banco de Dados escolhido para o desenvolvimento desta abordagem é o *MySQL* por diversos motivos que começam com a sua gratuidade para fins não comerciais (mesmo para fins comerciais o valor de sua licença é centenas de vezes menor que uma licença de um Sistema Gerenciador de Banco de Dados Relacional (SGBDR) comercial a exemplo do *Oracle*), passam pela sua robustez e vão até principalmente sua rapidez, sendo considerado um dos SGBDR mais rápidos, o que é um requisito fundamental quando se fala em cálculos que simulam operações em tempo real.

No atual esquema de testes, o servidor de banco de dados é executado no sistema operacional Linux e os diversos clientes são executados em máquinas Windows, com programas provistos com interfaces gráficas para facilitar sua utilização.

Foi constituído um esquema de testes para a Unidade FCC, caso estudo principal deste trabalho, ilustrado na figura 5.10, de modo que a unidade FCC é executada em *loop* infinito, gerando dados de vazão. Estes dados são em seguida corrompidos em um outro módulo que permite alterar parâmetros como desvio padrão de cada “medida” e o número de desvios padrão dentro dos quais será adicionado erro, além da possibilidade de se provocar um erro grosseiro. Finalizando o esquema de testes, está o acoplamento do programa para reconciliação de dados desenvolvido, mostrando a evolução das va-

riáveis em sua interface de modo a se comparar o valor “real” vindo do modelo antes da corrupção com os valores corrompidos e reconciliados, dispondo estas informações em forma numérica e gráfica.

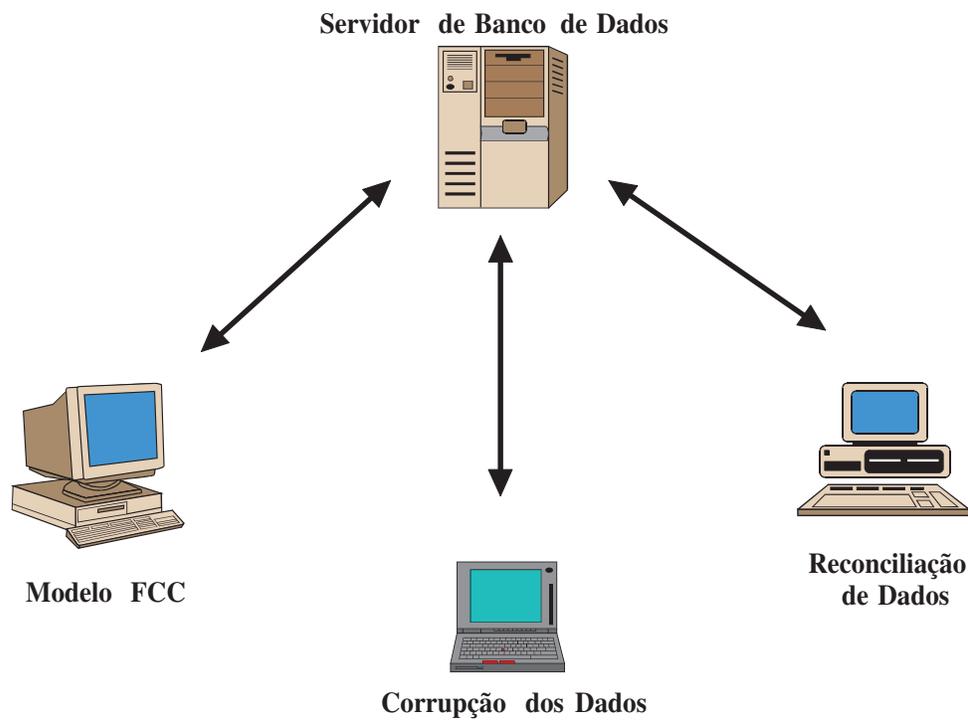


Figura 5.10: Esquematização do processo FCC sob o esquema de testes

5.5.1 Aplicação: Controle de um Reator CSTR não Isotérmico

Uma aplicação para o sistema de testes em rede é o controle de um reator CSTR (*Continuous Stirred Tank Reactor* – Reator Tanque de Mistura Perfeita), descrito em Luyben (1989), onde a modelagem dinâmica de um reator CSTR não isotérmico é submetida a um controlador do tipo P (proporcional). Os parâmetros do problema estão na Tabela 5.1.

Nesta aplicação o modelo transiente do reator é executado em *loop* infinito recebendo e enviando dados para um outro programa que pode ser diretamente a aplicação

executando o controlador (abordagem em um único plano) ou pode ser o Servidor de Banco de Dados (abordagem em dois planos).

Na interface do controlador são inseridos também os parâmetros do reator e os valores iniciais de operação, além dos parâmetros do próprio controlador como o ganho e os *set points*. Pode ser ainda controlada a frequência de atuação do controlador bem como desligá-lo completamente, operando-se em malha aberta. Este tipo de funcionalidade seria difícil de implementar se o controle e o modelo estivessem acoplados em uma única aplicação e os programas fossem em modo texto (*Console Application*). Estas características são interessantes do ponto de vista didático e podem ser usadas como uma ferramenta para este fim.

As Figuras 5.11 e 5.12 ilustram as interfaces gráficas do modelo e do controlador.

Tabela 5.1: Parâmetros para simulação dinâmica de um CSTR não isotérmico (LUYBEN, 1989)

Valores do Estacionário	
Vazão no reator (\bar{F})	$40 \text{ ft}^3/\text{h}$
Concentração de alimentação (\bar{C}_{A0})	$0,50 \text{ lb} \cdot \text{mol}_A/\text{ft}^3$
Temperatura no reator (\bar{T})	600°R
Vazão na jaqueta (\bar{F}_J)	$49,9 \text{ ft}^3/\text{h}$
Volume do reator (\bar{V})	48 ft^3
Concentração na saída (\bar{C}_A)	$0,245 \text{ lb} \cdot \text{mol}_A/\text{ft}^3$
Temperatura da jaqueta (\bar{T}_J)	$594,6^\circ\text{R}$
Temperatura de entrada no reator (\bar{T}_0)	530°R
Parâmetros	
Volume da jaqueta (V_J)	$3,85 \text{ ft}^3$
Energia de ativação (E)	$30000 \text{ Btu}/\text{lb} \cdot \text{mol}$
Coefficiente global de troca térmica (U)	$150 \text{ Btu}/\text{h ft}^2 \cdot ^\circ\text{R}$
Temperatura de entrada na jaqueta (T_{J0})	530°R
Capacidade calorífica reacional (C_p)	$0,75 \text{ Btu}/\text{lb}_m \cdot ^\circ\text{R}$
Densidade da massa reacional (ρ)	$50 \text{ lb}_m/\text{ft}^3$
Ganho (Kc)	$4 \text{ ft}^3/\text{h}/^\circ\text{R}$
Fator pré-exponencial (α)	$7,08 \times 10^{10} \text{ h}^{-1}$
Constante universal dos gases (R)	$1,99 \text{ Btu}/\text{lb} \cdot \text{mol} \cdot ^\circ\text{R}$
Área de troca térmica (A_H)	250 ft^2
Calor de reação (λ)	$-30000 \text{ Btu}/\text{lb} \cdot \text{mol}$
Capacidade calorífica do refrigerante (C_J)	$1,0 \text{ Btu}/\text{lb}_m \cdot ^\circ\text{R}$
Densidade do refrigerante (ρ_J)	$62,3 \text{ lb}_m/\text{ft}^3$
Temperatura <i>setpoint</i> (T^{set})	600°R

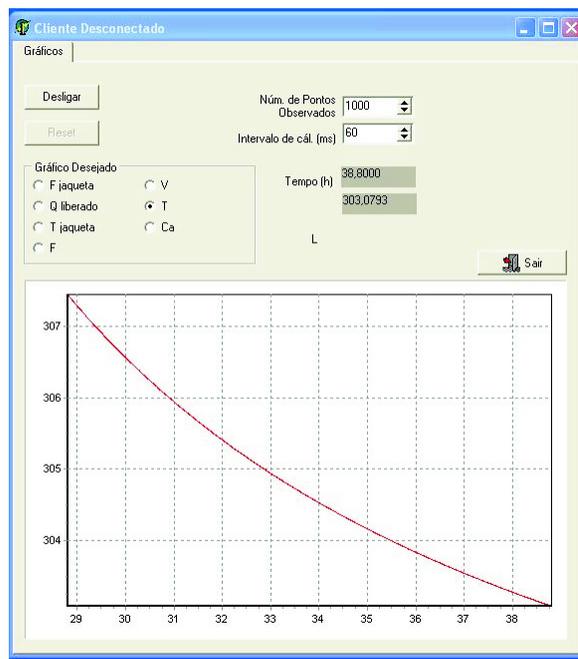


Figura 5.11: Interface gráfica para o simulador transiente de um CSTR não isotérmico

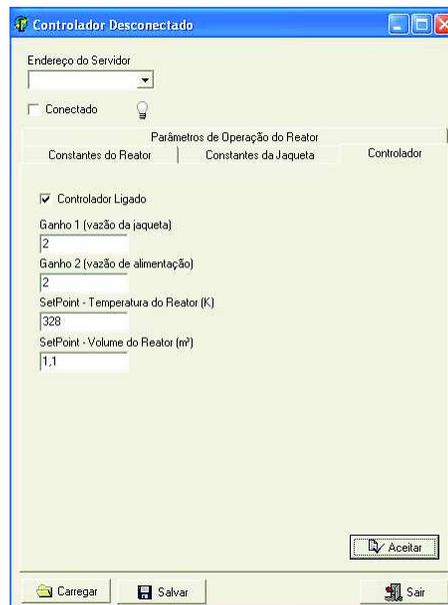


Figura 5.12: Interface gráfica para controle e entrada de dados de um CSTR não isotérmico

5.6 Conclusão

Neste Capítulo foi descrita toda a metodologia de construção do aplicativo de reconciliação de dados e detecção de erros grosseiros. Foram implementadas as técnicas expostas nos Capítulos 3 e 4 bem como mostrado o ambiente de simulação de processos: Sistema de Teste de Processos Distribuído em Rede, no qual os aplicativos desenvolvidos neste trabalho se inserem. A aplicação ao caso estudo, o sistema reator/regenerador FCC (Unidade de Craqueamento Fluido Catalítico), será mostrada no próximo Capítulo.

6 *Estudo de Caso: O Processo de Craqueamento Catalítico Fluido (FCC)*

An appropriate answer to the right problem is worth a good deal more than an exact answer to an approximate problem

John Tukey

6.1 Introdução

A unidade FCC tem como função a conversão de gasóleo oriundo do petróleo em hidrocarbonetos de cadeias menores, como por exemplo gasolina e gás liquefeito de petróleo (GLP).

O craqueamento no conversor FCC é o maior processo catalítico do mundo pois envolve uma grande quantidade de processamento, grande dimensão dos equipamentos envolvidos e grande quantidade de catalisador. Além disso, tem um grande impacto financeiro, pois apresenta um lucro líquido de aproximadamente US\$ 30/ton processada o que perfaz, em uma unidade típica, um total de US\$ 87,6 milhões por ano.

A unidade FCC produz alguns dos produtos de maior valor agregado nas refinarias, como por exemplo a gasolina e o GLP.

O sistema FCC é um equipamento de operação complexa pois envolve um processo

explosivo, multivariável, altamente não-linear, com fortes interações entre as variáveis e apresenta restrições nas variáveis de entrada e saída.

6.2 Descrição

A corrente de alimentação é introduzida no conversor FCC (*riser*), onde ocorre a reação de craqueamento catalítico. Os produtos formados contendo as frações mais leves de hidrocarbonetos são levados para a etapa de recuperação, constituída de colunas de destilação ou “fracionadoras”. O catalisador desativado é enviado para o regenerador que, depois de regenerado, é enviado continuamente para o riser para catalisar novamente a reação.

O conversor pode ser dividido em três partes:

Riser Local onde ocorre a reação de craqueamento em leito de arraste na presença de catalisador sob condições de controle de temperatura. O *riser* é assumido como sendo um reator tubular. A vazão de catalisador introduzida no *riser* é controlada pela abertura de uma válvula. A vazão e a temperatura da corrente de alimentação são variáveis também usadas para controlar a operação do conversor. A temperatura do *riser* deve ser controlada para permitir a obtenção de uma conversão adequada da reação de craqueamento, a qual pode ser caracterizada pela severidade da reação.

Reator Local onde os produtos gasosos da reação são separados do catalisador pela ação da gravidade. Os produtos (vapores) provenientes da reação são levados para uma coluna de destilação (fracionadora) principal e separados de acordo com o seu ponto de ebulição, enquanto o catalisador coqueado (desativado) passa por um *stripper* (onde o vapor d’água é usado para retirar o excesso de hidrocarbonetos do sólido) e entra na 3^a parte da unidade FCC, o regenerador.

Regenerador Local onde ocorre a regeneração do catalisador. O coque (depósito de carbono que fica aderido à superfície do catalisador impedindo que a reação cata-

lítica ocorra) é queimado na presença de ar, cuja vazão é manipulada. Usualmente as temperaturas do regenerador são elevadas e, devido ao risco de danos metalúrgicos no conversor, elas devem ser bem controladas. O regenerador é ainda dividido em cinco partes, descritas a seguir:

- Fase densa do primeiro estágio - onde o catalisador que vem do reator é depositado e sofre a primeira combustão;
- Fase densa do segundo estágio - onde o catalisador que transborda do estágio anterior sofre uma segunda queima do coque e segue retornando ao *riser*.

Acima dos compartimentos de depósito de catalisador, encontram-se os gases de combustão, os quais apresentam características distintas em diferentes partes do regenerador, sendo que três zonas principais podem ser identificadas:

- Fase diluída do primeiro estágio - localizado acima da fase densa e é constituído pelos gases de combustão proveniente da queima do coque desta fase;
- Fase diluída do segunda estágio - localizado acima do segundo estágio da fase densa;
- Fase geral diluída - formada pela mistura dos gases das outras fases, dando origem a uma região ainda mais diluída cujos gases são levados para uma fornalha.

Cada uma das fases do regenerador é modelada como um reator de mistura perfeita.

Uma representação esquemática do reator/regenerador *Orthoflow F Kellog* é mostrada na Figura 6.1. As reações de craqueamento do FCC levam a produtos com ponto de ebulição mais baixos tais como gasolina, óleo de reciclo leve, butanos e gás. A alimentação no caso estudado é gasóleo na faixa de temperatura de ebulição de 400 a 570 °C. Para garantir a conversão e a seletividade adequadas, a temperatura de reação deve ser mantida na faixa de 520 a 550 °C. Reações secundárias geram coque que é depositado sobre a superfície do catalisador. Na seção de craqueamento, é consumida energia

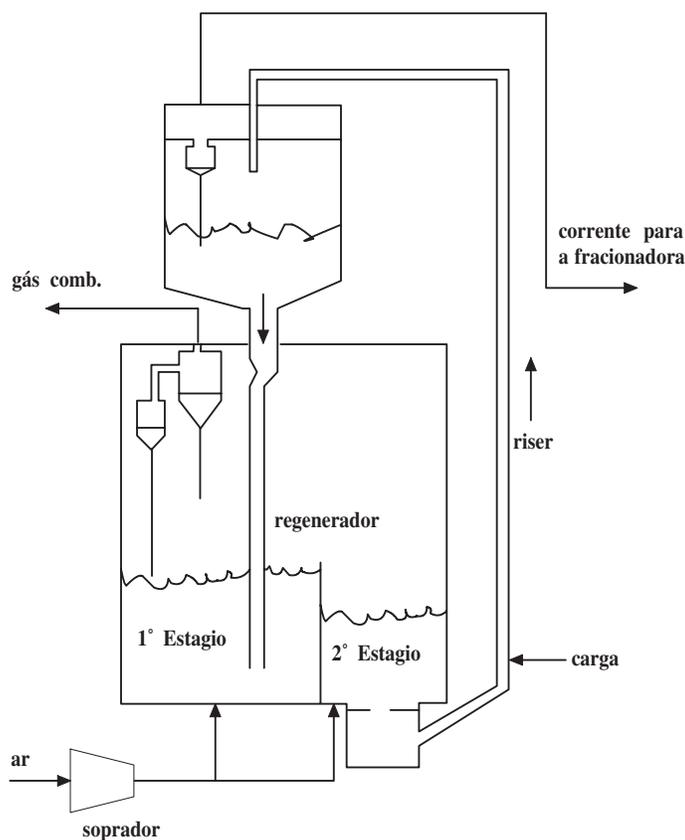


Figura 6.1: Esquema da seção reator/regenerador de uma unidade craqueamento catalítico fluido

pelas reações endotérmicas e o catalisador quente regenerado provê a energia necessária para fechar o balanço energético (MORO; ODLOAK, 1995).

Neste trabalho foi utilizada a modelagem dinâmica da seção reator/regenerador descrita em Moro e Odloak (1995) e programada em Fortran por Odloak *et al.* (SIMULADOR SIMFCCOF, versão 4.0 de janeiro de 1997).

Eles apresentam uma extensão do modelo de Kurihara (*apud* Moro e Odloak, 1995) e basicamente direcionam o trabalho no sentido da predição das seguintes variáveis:

- Conteúdo de coque do catalisador gasto e regenerado do primeiro e segundo estágios do regenerador;

- Temperatura do *riser*;
- Temperatura da fase densa do primeiro e segundo estágios;
- Inventário de catalisador no reator e no primeiro e segundo estágios do regenerador;
- Temperatura das fases diluídas do regenerador e dos ciclones;
- Conteúdo de oxigênio nos gases de combustão que saem do regenerador;
- Pressão no regenerador e no reator.

Como neste trabalho o interesse é voltado para as vazões mássicas que entram e saem da seção reator/regenerador do FCC, o programa foi utilizado para inferir estes dados de modo a disponibilizá-los como valores “reais”. Foi desconsiderado o *makeup* de catalisador neste balanço de massa.

Alguns dados provenientes do programa para simulação do FCC foram completados e/ou confrontados com dados disponíveis em Zanin (2001). Estes dados estão dispostos na Tabela 6.1.

A Figura 6.2 mostra um esquema simplificado da seção reator/regenerador do FCC que será usada para os testes de reconciliação de dados apresentados na seção 6.3.

Tabela 6.1: Propriedades das correntes de entrada e saída do FCC (reator/regenerador)

	Densidade (kg/m^3)	Vazão Volumétrica (m^3/h)	Vazão Mássica (kg/h)	Vazão Mássica (ton/min)
Gasóleo	$0,93 \times 10^3$	283	$26,319 \times 10^4$	4,39
Ar	1,293	221×10^3	$28,57 \times 10^4$	4,76
Gases Comb.	1,1	$253,64 \times 10^3$	$27,90 \times 10^4$	4,65
Produtos	$0,96 \times 10^3$	281,25	$27,00 \times 10^4$	4,5

6.3 Aplicação do Programa de Reconciliação ao FCC

A partir dos dados da Tabela 6.1, a Unidade FCC (reator/regenerador) foi representada na interface gráfica do programa de reconciliação de dados como pode ser visto na Figura 6.3. Com um desvio padrão correspondente a 2,5% de precisão em cada medida, foram estimados valores para as correntes e os gráficos correspondentes se encontram nas Figuras 6.4 a 6.7. Em todos os gráficos é perceptível a redução dos erros e a grande aproximação das estimativas com o “valor real”.

Para a corrente de gasóleo (Figura 6.5) e para a corrente de produtos (Figura 6.7), a estimativa é quase que coincidente com o “valor real”. Esta simulação da seção reator/regenerador também foi feita usando o esquema descrito na Seção 5.5, resultando

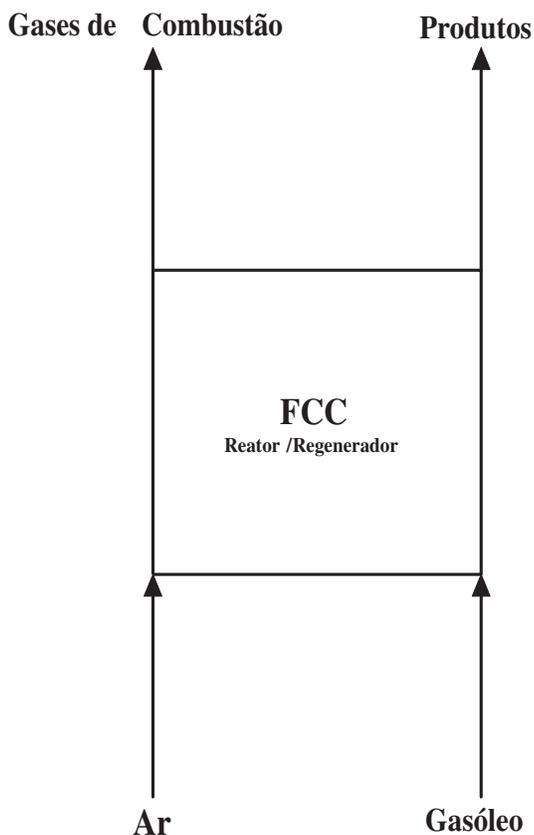


Figura 6.2: Esquema simplificado da seção reator/regenerador do FCC

no mesmo comportamento visto nas Figuras de 6.4 a 6.7.

Uma aplicação como esta é importante no controle da contabilidade da produção que é o que define o preço final do produto, bem como permite que seja feita uma otimização efetiva da operação.

Esta aplicação também torna evidente a flexibilidade do aplicativo na representação de um processo, unidade ou equipamento. Uma unidade, tal como descrita na Figura 6.1, foi abstraída para um único bloco virtual. O mesmo pode ser feito com, por exemplo, um conjunto de tanques. O sentido inverso também é possível: pode-se fragmentar um equipamento e representar as partes através dos blocos genéricos, como por exemplo, uma coluna de destilação pode ser representada por N blocos (um para cada estágio) com suas respectivas “correntes” de transferência de massa.

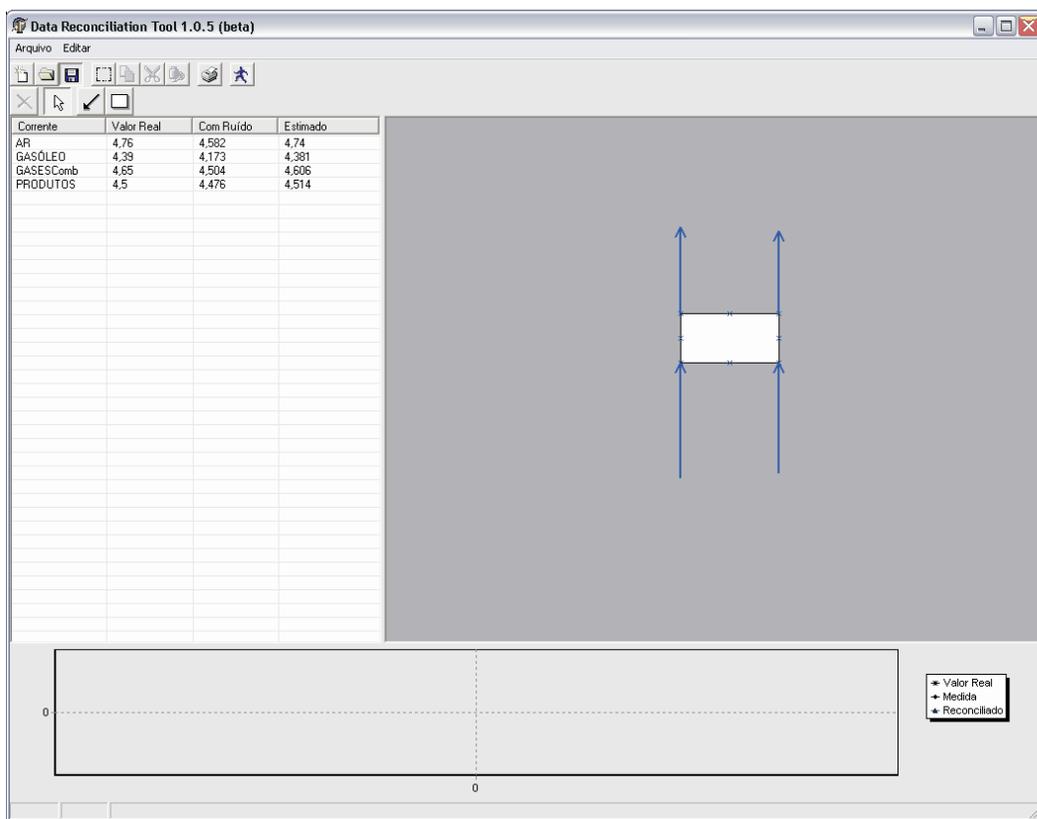


Figura 6.3: Representação do FCC (reator/regenerador)

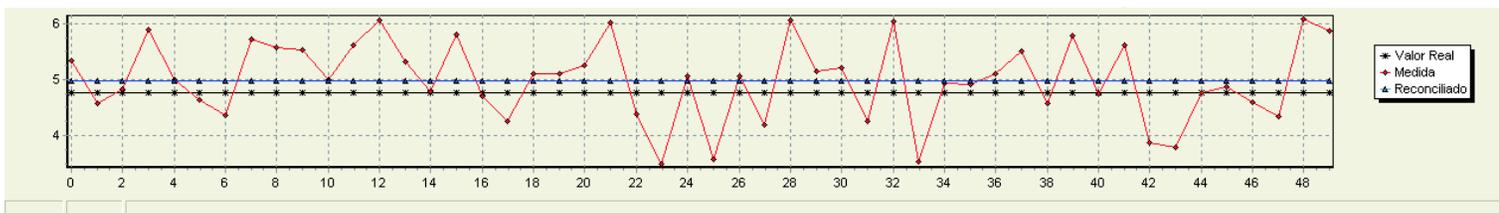


Figura 6.4: Corrente de ar

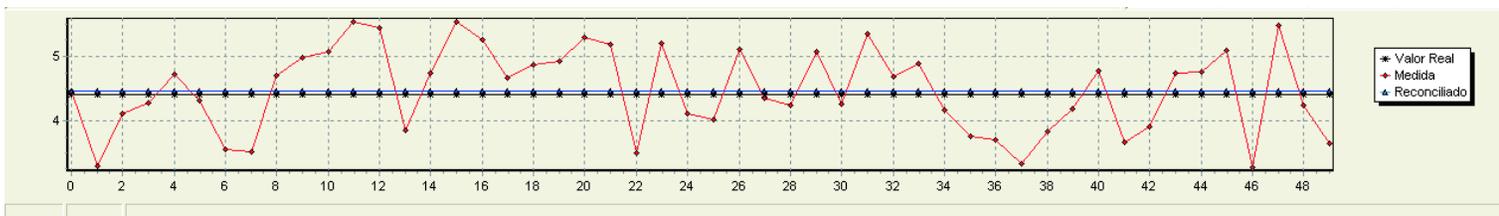


Figura 6.5: Corrente de gásóleo

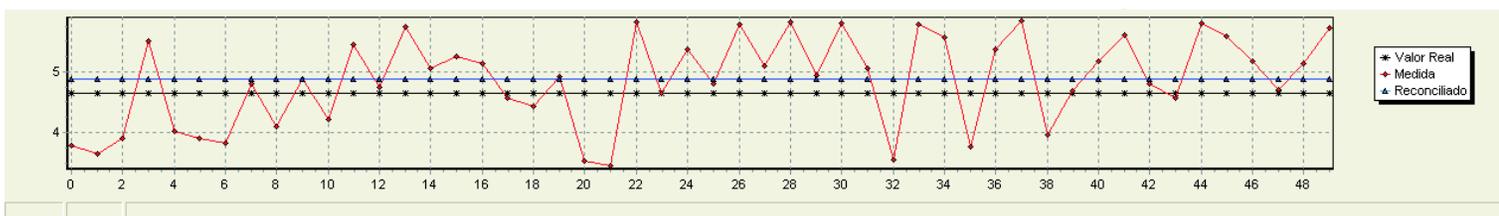


Figura 6.6: Corrente de gases de combustão

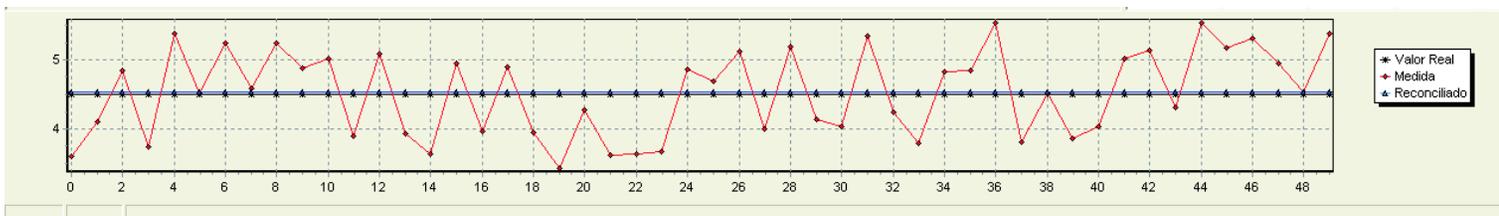


Figura 6.7: Corrente de produtos

7 *Conclusões*

Without assumptions, there can be no conclusions

John Tukey

7.1 **Conclusões**

A reconciliação de dados linear em estado estacionário, estudada neste trabalho, permite fazer ajustes de vazões mássicas em processos, unidades ou mesmo equipamentos com medidas redundantes. Os procedimentos estudados e implementados levam, dentro de certas condições, à melhora na qualidade da informação proveniente de um conjunto de medidas.

Foi adotado um seqüenciamento simples, mas eficiente, no condicionamento da informação e que vem sendo usado em aplicações comerciais na área, qual seja: pré-condicionamento das informações através da média simples de dados coletados em uma janela de operação, realização de uma reconciliação de dados preliminar, detecção de erros grosseiros a partir dos dados reconciliados, reestruturação do problema de reconciliação com a eventual retirada de medidas portadoras de erros grosseiros e finalmente uma nova reconciliação de dados, no caso da presença de erros grosseiros ter sido detectada, e por fim as estimativas das medidas redundantes e das variáveis observáveis. Para desvios padrão não superiores a 5% do valor real, as técnicas abordadas aqui se mostraram eficientes.

O uso da reconciliação se justifica quando comparado com outros métodos de condicionamento de dados que são mais simples, como no caso de uma média dentro de um intervalo de tempo, mas que não garantem a consistência dos dados para o fechamento de balanços de massa e energia. Desta forma pode-se concluir que a reconciliação de Dados Linear em estado estacionário é de fato uma ferramenta importante para a otimização de processos químicos e ainda que a abordagem que foi dada neste trabalho permite o uso facilitado desta ferramenta.

O sistema de testes de processos distribuído em rede se mostrou bastante útil nos testes de reconciliação de dados e é uma solução versátil para simular integração de processos em tempo real.

7.2 Sugestões para Trabalhos Futuros

No decorrer deste trabalho foram levantadas algumas questões que merecem aprofundamento, mas que fugiam ao escopo do Projeto de Pesquisa. Dentre outras questões, as seguintes podem ser destacadas e ficam como sugestão para trabalhos futuros:

- Estudo de distribuições não normais de erros em instrumentos de medida com modelos não lineares (como medidas de vazão a partir de queda de pressão);
- Avaliação de técnicas mais elaboradas de otimização para a solução do problema de minimização de erros envolvido na reconciliação de dados, como, por exemplo, a associação de métodos baseados em Algoritmos Genéticos com métodos formais de otimização do tipo SQP (*successive quadratic programming*);
- Estudo e avaliação do pré-condicionamento de dados através de filtros;
- Avaliação do efeito da elaboração matemática/metodológica sobre o tempo de cálculo com vistas a contemplar o pré-requisito da disponibilidade da informação em tempo real para as rotinas de otimização e controle;

- Desenvolvimento de conexão com fontes de dados padrão na indústria para testes com dados reais.

Referências

- BAGAJEWICZ, M. J. On the probability distribution and reconciliation of process plant data. *Computers Chem. Engng.*, v. 20, n. 6/7, p. 813–19, 1996.
- BAGAJEWICZ, M. J.; JIANG, Q. Comparison of steady state and integral dynamic data reconciliation. *Comp. and Chem. Engng.*, v. 24, p. 2367–83, 2000.
- BARBOSA, A. G. *Desenvolvimento de Software Educacional Voltado para a Engenharia Química - Relatório de Estágio Integrado*. Campina Grande, Maio 2001.
- BIEGLER, L. T.; DAMIANO, J. J.; BLAU, G. E. Non-linear parameter estimation: A case study comparison. *AIChE Journal*, n. 32, p. 29–43, 1986.
- CHAPMAN, S. J. *Fortran 90/95 for Scientists and Engineers*. 1st. ed. [S.l.]: WBC McGraw-Hill, 1998.
- CONSUL, C. M. D. *Técnicas Estatísticas Multivariadas para o Monitoramento de Processos Industriais Contínuos*. Tese (Tese de Doutorado) — Faculdade de Engenharia Química/UNICAMP, 2002.
- CROWE, C. M. Reconciliation of process flow rates by matrix projection. part ii: The non-linear case. *AIChE Journal*, n. 32, p. 616–23, 1986.
- CROWE, C. M. Recursive identification of gross errors in linear data reconciliation. *AIChE Journal*, n. 34, p. 541–50, 1988.
- CROWE, C. M. Observability and redundancy of process data for steady state reconciliation. *Chem. Eng. Sci.*, n. 44, p. 2909–17, 1989.
- CROWE, C. M. Test of maximum power for detection of gross errors in process constraints. *AIChE Journal*, n. 35, p. 869–72, 1989.
- CROWE, C. M. Data reconciliation - progress and challenges. *J. Proc. Cont.*, v. 6, n. 2/3, p. 89–98, 1996.
- CROWE, C. M.; CAMPOS, Y. A. G.; HRYMAK, A. Reconciliation of process flow rates by matrix projection. part i: Linear case. *AIChE Journal*, n. 29, p. 881–88, 1983.

- DAHLQUIST, G.; BJORK, A. *Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1974.
- DEMPF, D.; LIST, T. On-line data reconciliation in chemical plants - industrial application of known methods. *Computers Chem. Engng.*, v. 22, 1998.
- JORDACHE, C. I.; MAH, R. S. H.; TAMHANE, A. C. Performance studies of the measurements test for detection of gross errors in process data. *AIChE Journal*, v. 31, n. 7, p. 1187–201, 1985.
- JORIS, P.; KALITVENTZEFF, B. Process measurements analysis and validation. *Proc. CEF'87: Use Comput. Chem. Eng.*, p. 41–6, 1987.
- KALMAN, R. E. Contributions to the theory of optimal control. *Boletin de la Sociedad Matematica Mexicana*, n. 5, p. 102–19, 1960.
- KAO, C. S.; TAMHANE, A. C.; MAH, R. S. H. Gross error detection in serially correlated process data. *Ind. Eng. Chem. Res.*, v. 29, n. 6, p. 1004–12, 1990.
- KELLER, J. Y. Analytical estimator of measurement error variances in data reconciliation. *Computers Chem. Engng.*, v. 16, n. 3, p. 185–8, mar. 1992.
- KNEPPER, J. C.; GORMAN, J. M. Statistical analysis of constrained data sets. *AIChE Journal*, v. 26, n. 2, p. 260–4, mar. 1980.
- KRETISOVALIS, A.; MAH, R. S. H. Observability and redundancy classification in multicomponent process networks. *AIChE Journal*, n. 33, p. 70–82, 1987.
- KRETISOVALIS, A.; MAH, R. S. H. Observability and redundancy classification in generalised process networks. i:theorems. *Comput. Chem. Eng.*, n. 12, p. 671–87, 1988a.
- KRETISOVALIS, A.; MAH, R. S. H. Observability and redundancy classification in generalised process networks. ii:algorithms. *Comput. Chem. Eng.*, n. 12, p. 689–703, 1988b.
- LUYBEN, W. L. *Process Modeling, Simulation, and Control for Chemical Engineers*. [S.l.]: McGraw-Hill, 1989.
- MADRON, F. A new approach to the identification of gross errors in chemical engineering measurements. *Chem. Eng. Sci.*, v. 40, n. 10, p. 1855–60, 1985.
- MADRON, F. *Process Plant Performance: Measurement and Data Processing for Optimization and Retrofits*. [S.l.]: Ellis Horwood Ltd., 1993. (Ellis Horwood Series in Chemical Engineering).

- MAH, R. S. Design and analysis of process performance monitoring systems. *Engineering Foundation Conf. on Chemical Process Control II*, p. 525–40, 1981.
- MAH, R. S. H.; STANLEY, G.; DOWNING, D. Reconciliation and rectification of process flow and inventory data. *Ind. Eng. Chem. Process Des. Dev.*, n. 15, p. 175–83, 1976.
- MAH, R. S. H.; TAMHANE, A. C. Detection of gross erros in process data. *AIChE Journal*, v. 28, n. 5, p. 828–30, 1982.
- MEYER, M.; KOEHRET, B.; ENJALBERT, M. Data reconciliation on multicomponent network process. *Comput. Chem. Eng.*, n. 17, p. 807–17, 1993.
- MORO, L. F. L.; ODLOAK, D. Constrained multivariable control of fluid catalytic cracking converters. *J. Proc. Cont.*, v. 5, n. 1, p. 29–39, 1995.
- NARASIMHAM, S.; MAH, R. S. H. Generalized likelihood ratio method for gross error indentification. *AIChE Journal*, v. 33, n. 8, p. 1514–21, 1987.
- NARASIMHAM, S.; MAH, R. S. H. Treatment of general steady state process models in gross error detection. *Comput. Chem. Engng.*, v. 13, n. 7, p. 851–3, 1989.
- NARASIMHAN, S.; JORDACHE, C. *Data Reconciliation and Gross Error Detection - an intelligent use of process data*. [S.l.]: Gulf Publishing Company, 2000.
- PLÁCIDO, J. *Desenvolvimento e Aplicação Industrial da Técnica de Reconciliação de Dados*. Dissertação (Tese de Mestrado) — POLI/USP, 1995.
- RAO, C. R. *Linear Statistical Inference and its Applications*. [S.l.]: Wiley, 1973.
- RIPPS, D. L. Adjustment of experimental data. *Chem. Eng. Progress - Symp. Series*, n. 61, p. 8–13, 1965.
- ROLLINS, D. K.; DAVIS, J. F. Unbiased estimation of gross errors in process measurements. *AIChE Journal*, v. 38, p. 563–72, 1992.
- ROMAGNOLI, J. A. On data reconciliation: Constraints processing and treatment of bias. *Chem. Eng. Sci.*, v. 38, n. 7, p. 1107–17, 1983.
- ROMAGNOLI, J. A.; SÁNCHEZ, M. C. *Data Processing and Reconciliation for Chemical Process Operations*. [S.l.]: Academic Press, 2000.
- ROMAGNOLI, J. A.; STEPHANOPOULOS, G. On the rectification of measurements errors for complex chemical plants. *Chem. Eng. Sci.*, v. 35, p. 1067–81, 1980.

- ROMAGNOLI, J. A.; STEPHANOPOULOS, G. Rectification of process measurements data in the presence of gross errors. *Chem. Eng. Sci.*, v. 36, n. 11, p. 1849–63, 1981.
- ROSENBERG, J.; MAH, R. S. H.; JORDACHE, C. Evaluation of schemes for detecting and identification of gross errors in process data. *Ind. & Eng. Chem. Proc. Des. Dev.*, n. 26, p. 555–64, 1987.
- SÁNCHEZ, M. C.; BANDONI, A.; ROMAGNOLI, J. A. Pladat - a package for process variable classification and plant data reconciliation. *Comput. Chem. Engn.*, n. S16, p. 499–506, 1992.
- SÁNCHEZ, M. C.; ROMAGNOLI, J. A. Use of orthogonal transformations in data classification - reconciliation. *Comput. Chem. Engng.*, n. 20, p. 483–93, 1996.
- SERTH, R. W.; HEENAN, W. A. Gross errors detection and data reconciliation in steam-metering systems. *AIChE Journal*, n. 30, p. 243–7, 1986.
- STANLEY, G.; MAH, R. S. H. Observability and redundancy in process data estimation. *Chem. Engng. Sci.*, v. 36, p. 259–83, 1981a.
- STANLEY, G.; MAH, R. S. H. Observability and redundancy classification in process networks. *Chem. Engng. Sci.*, v. 36, p. 1941–54, 1981b.
- SWARTZ, C. L. Data reconciliation for generalised flowsheet applications. *197th Natl. Meet. Am. Chem. Soc.*, 1989.
- TAMHANE, A. C. A note on use of residuals for detecting an outlier in linear regression. *Biometrika*, v. 69, n. 2, p. 448–9, 1982.
- VÁCLAVEK, V. Studies on system engineering. iii. optimal choice of the balance measurements in complicated chemical engineering systems. *Chem. Eng. Sci.*, v. 24, p. 947–55, 1969.
- VÁCLAVEK, V.; LOUCKA, M. Selection of measurements necessary to achieve multicomponent mass balances in chemical plants. *Chem. Engn. Sci.*, n. 31, p. 1199–205, 1976.
- YANG, Y.; TEN, R.; JAO, L. A study of gross errors detection and data reconciliation in process industries. *Computers Chem. Engng.*, v. 19, n. suppl., p. s217–22, 1995. European Symposium on Computer Aided Process Engineering.
- ZANIN, A. C. *Implementação Industrial de um Otimizador em Tempo Real*. Tese (Tese de Doutorado) — Escola Politécnica - USP, 2001.
- ZHANG, P.; RONG, G.; WANG, Y. A new method of redundancy analysis in data reconciliation and its application. *Comp. and Chem. Engng.*, v. 25, p. 941–9, 2001.

APÊNDICE A – Ferramentas Utilizadas

A.1 Introdução

Foram utilizados neste trabalho o ambiente de computação matemática **MATLAB 6** e as linguagens de programação **Fortran 90/95** e o **Borland Delphi 6**. Serão apresentadas neste capítulo as justificativas para o uso de tais ferramentas bem como alguns detalhes pertinentes à técnica de combinar linguagens de programação diferentes para obtenção de resultados com características profissionais (*mixed-language programming*).

A.2 MATLAB

O MATLAB é o ambiente ideal para o desenvolvimento e teste de algoritmos devido à facilidade com que as subrotinas de cálculo são chamadas e os resultados são processados e expostos (inclusive na forma gráfica). Um dos inconvenientes da utilização do MATLAB está associado à sua lentidão se comparado com um *software* dedicado a resolver uma única tarefa. No escopo deste projeto, o uso do MATLAB é voltado para a criação, verificação e análise inicial das subrotinas de reconciliação e retificação de dados.

A.3 Linguagens de Programação

Segundo Chapman (1998) o computador ao executar um programa segue uma cadeia de operações muito simples, tais como carregar, estocar, adicionar, subtrair e multiplicar. Cada operação tem um único padrão binário chamado de *operation code* (código de operação) ou *op code* para especificá-lo. O programa que o computador executa é tão somente uma cadeia de *op codes* (e os dados associados aos *op codes*) em uma ordem necessária para se alcançar um objetivo. Os *op codes* são coletivamente chamados de linguagem de máquina, pois eles são a verdadeira linguagem que o computador reconhece e executa (BARBOSA, 2001).

A linguagem de máquina é muito difícil de se trabalhar, sendo chamada de “linguagem de baixo nível”. É muito mais produtivo trabalhar com linguagens que exprimam as operações a serem realizadas em orações e expressões algébricas tais como as encontradas no uso cotidiano. Tais linguagens de programação são chamadas de “linguagens de alto nível”.

Segundo Barbosa (2001), quando se trabalha com linguagens de alto nível, os compiladores e “*linkadores*” convertem para linguagem que o computador compreende: a linguagem de máquina. Existem várias linguagens de alto nível com características diversas. Algumas são apropriadas para uso científico, outras para uso de banco de dados e outras ainda, são de uso geral. O importante é utilizar a ferramenta (ou a combinação de ferramentas) apropriada para o problema em questão

No presente trabalho optou-se pelo uso do Fortran para a codificação dos cálculos necessários na resolução dos problemas matemáticos pelo seu legado de código científico e pela qualidade das subrotinas de cálculo numérico disponíveis. Por outro lado, O compilador Fortran do Laboratório de Otimização de Processos e Controle Avançado - LOPCA, onde este trabalho foi desenvolvido, não dispõe de um ambiente de desenvolvimento rápido de *front-ends* (interfaces gráficas), sendo assim necessário o uso de linguagens de programação combinadas com o fito de se obter um resultado de características profissionais. A linguagem escolhida para a geração do *front-end* e trata-

mento parcial das informações foi o Borland Delphi 6. A justificativa para o uso desta ferramenta e pormenores da técnica de combinar linguagens diferentes em um único aplicativo são expostas a seguir.

O Borland Delphi não é especificamente uma linguagem de programação, ele é uma ferramenta **RAD** - *Rapid Application Development* ou Desenvolvimento Rápido de Aplicativos. De fato, a linguagem utilizada é o **Object Pascal** que é a versão orientada a objetos do Pascal.

O Delphi é característico pela facilidade com que proporciona a construção de um aplicativo fazendo com que a fase de prototipagem (construção da interface gráfica) seja quase que totalmente dissociada da criação do código funcional. O próprio Delphi gera todo código necessário para declaração de métodos (sub-rotinas e funções) e de classes e objetos que são criados no protótipo através do paradigma *drag'n'drop*¹. Cabe ao programador escrever o código essencial do aplicativo e que responderá às ações do usuário².

A área de trabalho do Delphi (figura A.1) é composta basicamente de barras de ferramentas, uma palheta de componentes (que podem ser estendidos por terceiros programando no próprio Object Pascal), o **Form Designer** (para o *Design* das Janelas), Object Inspector (figura A.2 - para inspeção e edição das propriedades dos objetos) e uma janela para visualização do código.

Na palheta de componentes se encontram agrupados por afinidade diversos tipos de componentes visuais e não visuais que são arrastados e soltos no *Form Designer* (figura A.3) para compor, de uma forma totalmente visual, a aplicação, janela por janela.

Os componentes são simplesmente soltos no *Form Designer* e nele tem-se ferramentas para gerenciar todas as posições e tamanhos dos componentes visuais relacionando-os entre si de uma forma estética e funcional.

Ao passo que são adicionados componentes no *Form Designer*, são adicionadas au-

¹arraste e solte

²Orientação a eventos

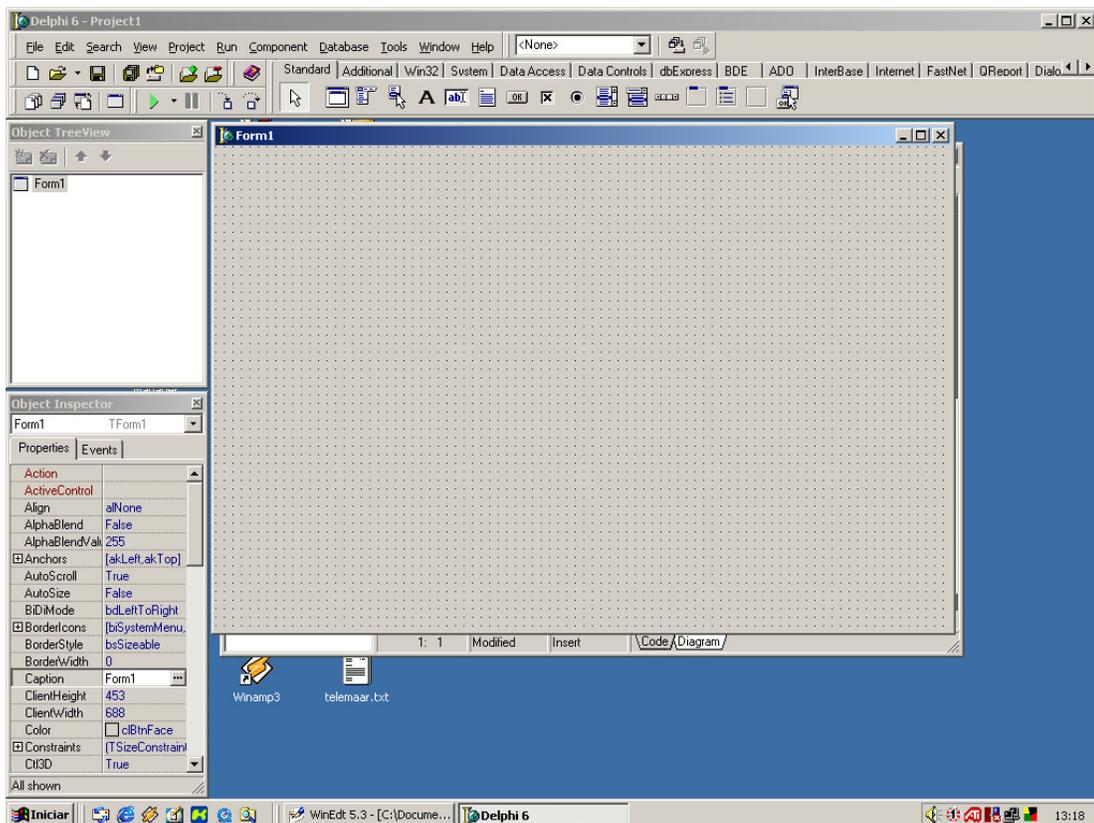


Figura A.1: A interface de desenvolvimento integrado do Borland Delphi 6

automaticamente as linhas de código necessárias para a declaração das classes que correspondem a estes componentes. Este é um dos aspectos mais interessantes do Delphi: ele escreve uma boa parte do código automaticamente e geralmente é a parte mais tediosa, deixando para o programador o foco no objetivo (BARBOSA, 2001).

Finalmente, no Object Inspector, são visualizadas as propriedades e os eventos associados com um componente visual ou não visual. Por exemplo, quando se seleciona um botão no *Form Designer*, no Object Inspector tem-se como propriedades a **caption** que é o texto que aparece no botão (ok, cancela, etc.), a propriedade **name** que é o nome do botão acessado via código, **Height** que é a altura do botão e várias outras propriedades. Na parte de eventos, para o mesmo botão temos o **OnMouseMove** que é disparado toda vez que o ponteiro do mouse passar pelo botão. Quando se seleciona um evento, o

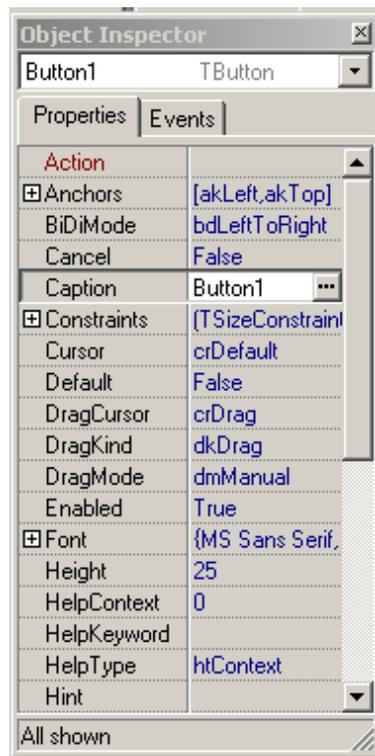


Figura A.2: Object inspector

Delphi automaticamente escreve no código um *handler* de evento, que é na verdade um bloco de código que só é executado quando o evento ocorre.

Além de toda a facilidade de desenvolvimento de aplicativos que foi exposta, tem-se ainda como forte justificativa para o uso do Borland Delphi a vasta disponibilidade de literatura nos mais diversos níveis.

A.4 Linguagens de Programação Combinadas

A técnica de *mixed-language programming* utilizada neste trabalho para combinar código escrito nas linguagens Fortran e Delphi é a das DLL's (*Dynamic Link Libraries* - Bibliotecas de Vínculo Dinâmico). Estas bibliotecas de vínculo dinâmico são, grosso modo, executáveis chamados a partir de outros executáveis.

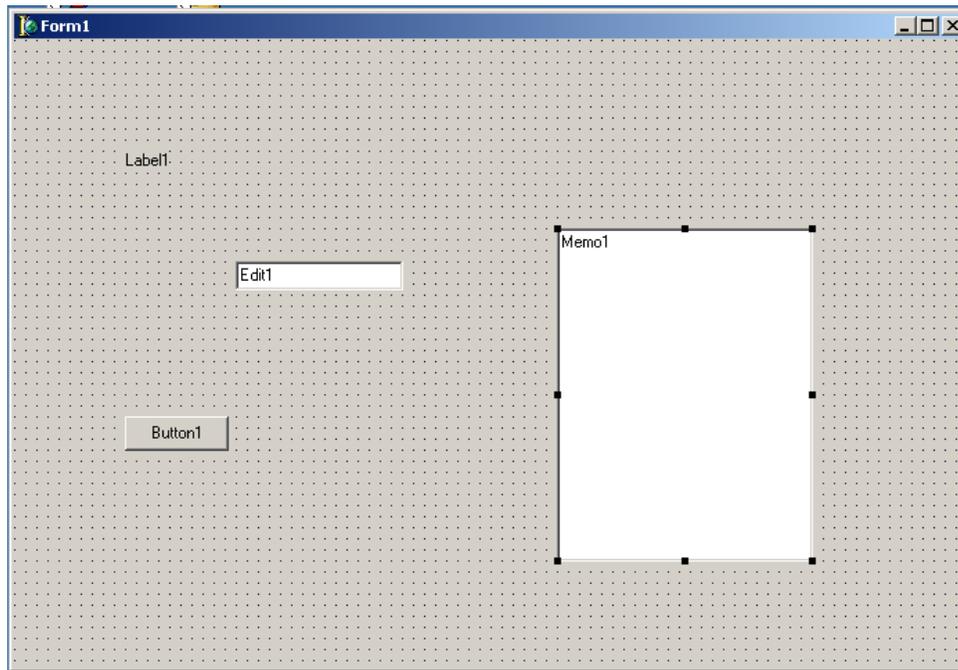


Figura A.3: *Form designer*

Geralmente as DLL's do sistema Windows contêm funções que realizam uma tarefa que pode ser requisitada por diferentes aplicativos e assim a DLL, que é carregada uma só vez na memória pode ser compartilhada entre aplicativos de uma maneira transparente para o usuário.

A abordagem deste trabalho é armazenar nestes arquivos separados do executável (programa principal), as sub-rotinas escritas em Fortran que efetuam os cálculos que exigem robustez e velocidade.

Para construir uma DLL no ambiente de desenvolvimento **Fortran COMPAQ 6.1**, por exemplo, basta criar um novo *workspace*, defini-lo como uma DLL, adicionar uma unidade de código e usar exclusivamente *subroutines* e *functions* dentro do projeto.

Depois de se definir a sub-rotina e seus parâmetros a serem passados, é usada uma diretiva de compilação, após a declaração da subrotina, para tornar a DLL uma DLL de “exportação”, ou seja, torná-la visível para chamadas externas.

Esta diretiva de compilação define os atributos da DLL, tais como a convenção de chamada (leitura de parâmetros da esquerda para direita ou o contrário), o nome de “exportação” da sub-rotina (o nome que será chamado externamente) e outros detalhes do funcionamento da sub-rotina.

Feitas estas etapas, parte-se então para a codificação em si, que, excetuando-se alguns cuidados, é exatamente como em um programa comum, podendo conter inclusive outras sub-rotinas. Depois de codificada, basta compilar e gerar o arquivo de extensão .dll que será chamado por outro executável.

Para se chamar uma sub-rotina contida em uma DLL a partir de outro executável tem-se duas maneiras: o carregamento dinâmico e o carregamento estático. O primeiro é mais sofisticado e evita erros de carregamento mapeando a sub-rotina somente quando esta é necessária, descarregando-a em seguida. O segundo carrega a sub-rotina no início do programa e, se esta não estiver presente ou apresentar algum problema, simplesmente o programa principal não é iniciado.

Baseado no que foi exposto, é plenamente justificável o uso combinado de duas linguagens de programação pois cada uma delas apresenta uma especialidade.