

Uma proposta para detecção de erros grosseiros em balanços de massa de processos industriais

A. M. Oliveira Júnior¹, E.J. Santos¹, E. L. Lima²; J. C. C. S. Pinto²

¹Programa de Pós-graduação em Engenharia Química, Universidade Federal de Sergipe, 49100-000, São Cristóvão-SE, Brasil

²Programa de Engenharia Química, COPPE/UFRJ, 21941-972, Rio de Janeiro-RJ, Brasil

antonio_martins@pq.cnpq.br

(Recebido em 7 de abril de 2011; aceito em 25 de abril de 2011)

As plantas químicas modernas são compostas por diversas unidades industriais como reatores, colunas de destilação etc. que estão conectados por uma complexa rede de tubulações. Variáveis de processo (vazões mássicas, temperaturas) são regularmente usadas para controle e avaliação de desempenho dos processos industriais. É esperado que essas medições satisfaçam os balanços de massa e energia associadas a restrições industriais quando o mesmo encontra-se no estado estacionário. Infelizmente, medições industriais não são perfeitas, e, de fato, sujeito a erros aleatórios, e, possivelmente, erros grosseiros. Os erros grosseiros podem ser causados devido ao mau funcionamento dos instrumentos, ruídos, ou, até mesmo, vazamentos no processo. A detecção de erros grosseiros é de grande importância, pois invalida a base estatística dos procedimentos de reconciliação de dados devido aos dados corrompidos, além de ser a base para o processo de diagnóstico de falha em uma unidade industrial.

Palavras-chave: reconciliação de dados, erros grosseiros, processos industriais.

A modern chemical plant consists of a large number of process units such as reaction vessels, distillation columns, and so forth, which are interconnected together by a complicated network of streams. Measurements of mass flow rates, temperature, and so forth are routinely made for purpose of process control and process performance evaluation. These measurements are expected to satisfy mass and energy balance constraints associated with the process network when the process is in a steady state. Unfortunately, the process measurements data are never perfectly corrected and are in fact subject to random and possibly gross errors. Malfunctioning instruments, measurement biases and process leaks, could cause the gross errors. Detection of the gross errors is of great of importance, because, firstly, the presence of gross errors will corrupt the data reconciliation calculation and spread the errors over all relatively corrected data; secondly, it is the basis of process fault diagnosis.

Keywords: Data reconciliation, Gross erros, Industrial Process.

1. INTRODUÇÃO

Nos projetos de indústrias químicas, instrumentos de medição, amostradores, etc, são colocados em várias correntes do processo, mas não em todas as correntes devido ao custo elevado. Associado a incertezas e imprecisões nas medidas, a técnica de reconciliação de dados tornou-se de fundamental importância na operação de plantas industriais.

A realidade é, claro, que todos os dados contêm erros ou incertezas em maior ou menor grau independentemente da técnica ou instrumentação utilizada. O que é requerido de um sistema de monitoramento é que ele seja suficientemente simples e preciso para uso pelos operadores da planta e engenheiros de operação, permitindo ajuste e correção das medidas, detecção de erros grosseiros e caracterização do grau de confiança desejado nas medições.

As medidas do processo são sujeitas a erros, sejam eles de natureza aleatória ou sistemática, fazendo com que as leis de conservação não sejam obedecidas. O procedimento conhecido como reconciliação de dados baseia-se na proposição de ajuste para que as medidas se adequem às leis de conservação, e a outras restrições impostas pelo processo.

A reconciliação de dados, normalmente, é baseada na hipótese de que a medida está corrompida por um erro, cujo modelo segue uma distribuição gaussiana. Desta forma, a possível existência de erros grosseiros deve ser identificada, e os erros grosseiros devem ser removidos

antes do ajuste das variáveis redundantes. Diferentes métodos para detecção destes erros podem ser encontrados em Mah (1990).

2. ERROS GROSSEIROS

Como medidas de variáveis de processo (vazões, concentrações, temperaturas) estão sujeitas a erros de medição e a perturbações relacionadas à diversidade e variabilidade dos processos, não deve esperar que um conjunto de medidas obedeça às leis de conservação. Espera-se que os erros das medidas sejam aleatórios. Contudo, podem ser encontrados também erros tendenciosos, causados pelo mau funcionamento dos instrumentos, má calibração ou amostragem indevida. Erros de transcrição de dados, como aqueles relacionados ao arquivamento equivocado de valores, podem ser considerados erros grosseiros.

De maneira geral, as medidas podem conter vários tipos de erros:

- a) Erros aleatórios pequenos
- b) Ruídos sistemáticos
- c) Erros grosseiros

Os pequenos erros aleatórios flutuam em torno da média zero e freqüentemente apresentam uma distribuição similar à distribuição gaussiana. A relação entre uma variável medida na planta e seu verdadeiro valor pode ser representada na forma:

$$x = \tilde{x} + \delta \quad (1)$$

onde \tilde{x} é o valor medido; x é o valor verdadeiro desconhecido; e δ é o erro aleatório. Os ruídos sistemáticos ocorrem quando as medidas fornecem valores errôneos consistentes. O terceiro tipo de medida errônea é o erro grosseiro, usualmente causado por uma seqüência de eventos não aleatórios.

A presença de erros grosseiros invalida a base estatística normalmente usada para construir os procedimentos de reconciliação de dados, sendo, por isso, essencial um tratamento apropriado desses dados. O tratamento de erros grosseiros pode ser geralmente dividido em três estágios. No primeiro estágio, técnicas de detecção de erros grosseiros são usadas para verificar se erros grosseiros estão presentes nas medições. Se a presença de erros grosseiros for detectada, o próximo estágio é a identificação apropriada das fontes desses erros. O estágio final é a eliminação das medidas corrompidas por erros grosseiros do conjunto de medidas experimentais.

Existem vários testes estatísticos construídos para detecção de erros grosseiros (McBreyer e Edgard, Bagajewicz e Jiang, 1997, Amand *et al*, 2001, Crowe, 1988, Chen *et al*, 1998, Abul-el-zeet *et al*, 2002). Narasimhan e Mah (1987) e Tamhane *et al* (1985), usaram testes seqüenciais, onde os dados são comparados usando um modelo de referência usando regressão de mínimos quadrados com os resíduos sendo estudados usando testes estatísticos tradicionais. Essa aproximação tem a desvantagem que os testes de distribuição são baseados nos resíduos de uma regressão que pode ter sido altamente afetada pela presença de *outliers*. Devido à natureza seqüencial do método, muitas regressões podem ter que ser feitas até que o critério de parada seja atingido.

Somente alguns métodos para identificação de erros grosseiros são capazes de discernir o local e o tamanho do erro (Bagajewicz, 2000). Depois da identificação dos erros grosseiros, uma das possíveis tomadas de decisão é normalmente executada: eliminar a medida que contém o erro grosseiro; ou corrigir a medida e efetuar a reconciliação novamente. Todos os testes estatísticos e eliminações seqüenciais ou técnicas de compensações seqüenciais propostas podem ser usadas para detecção e identificação de erros grosseiros em processos descritos por modelos não lineares. O procedimento comum é linearizar o modelo, seguido do método de identificação por equações lineares. Tipicamente, as variáveis medidas são reconciliadas sob a hipótese de que não há erros grosseiros e que as equações de restrição são linearizadas em torno das medidas reconciliadas. Essa estratégia, apesar de muito usual, pode não ser adequado para processos altamente não lineares com dados corrompidos por erros grosseiros significativos. Processos não lineares são muito mais complexos e requerem métodos especiais para detecção de erros grosseiros. A primeira alternativa é a que está implementada em softwares comerciais,

que somente considera os ruídos. Essa aproximação deixa o sistema com um pequeno grau de redundância, e, conseqüentemente, a precisão das variáveis reconciliadas se deteriorou.

3. ESTRATÉGIA UTILIZADA

Normalmente assumimos que erros presentes nas medidas são normalmente distribuídos com média zero e covariância conhecidas. Entretanto, os dados do processo não satisfazem as equações de conservação de massa e energia. O tamanho da inconsistência é relativamente pequeno caso não exista erros grosseiros nas medições. Na prática, os dados devem conter outros tipos de erro causados por eventos não aleatórios. Para evitarmos estas circunstâncias, precisamos conferir a viabilidade dos dados medidos do ponto de vista de erros grosseiros. Com esse propósito, usaremos o critério de teste estatístico de inconsistência de dados para um problema não linear geral (Romagnoli e Chen, 1998). Assim, o conjunto e as equações de balanço podem ser escritos da seguinte forma:

$$g(x) = 0 \quad (2)$$

Nas situações reais, a função g é não linear desde que tenhamos, por exemplo, o produto de duas leituras. Então consideraremos uma função g não linear, cuja expansão de Taylor em x fornece :

$$g(x) = g(x_0) + g'(x_0)(x - x_0) + \dots \quad (3)$$

onde x_0 é a solução inicial que pode ser indicada pela indicação de medida \tilde{x} . Então temos aproximadamente:

$$g(x) = g(\tilde{x}) + G(\tilde{x})(x - \tilde{x}) = 0 \quad (4)$$

ou,

$$-g(x) = G(\tilde{x})(x - \tilde{x}) \quad (5)$$

Como $g(\tilde{x}) = -e$, conseqüentemente

$$G(\tilde{x})(x - \tilde{x}) = e \quad (6)$$

Desde que $x - \tilde{x} = \delta$, a equação acima torna-se:

$$G(\tilde{x})\delta = e \quad (7)$$

Para o caso quadrático específico, temos que:

$$g_i(x) = a_{i0}x + x^T E_{i0}x + b = 0 \quad (8)$$

Esta equação pode ser escrita como:

$$G^T \delta = e \quad (9)$$

onde $G_i = (a_{i0} + E_{i0}^T \tilde{x})$ é a i -ésima coluna da matriz G . Desta forma obtemos um conjunto de equações lineares não quadráticas, com os erros das medidas e das equações sendo

consideradas variáveis. Aqui, considera-se conhecido, porém, na realidade, não podemos resolver este problema pois o número de linhas é geralmente menor que o número de colunas. Pode ser demonstrado que os erros são normalmente distribuídos com média zero e matriz de covariância C_{ee} :

$$E(e) = 0 \quad (10)$$

$$C_{ee} = D(e, e^T) = GC_{\delta\delta}G^T \quad (11)$$

onde $C_{\delta\delta}$ é a matriz de covariância dos erros da medida.

A seguir formularemos uma função teste para detecção de erros grosseiros:

$$h = e^T C_{ee}^{-1} e \quad (12)$$

A função h segue uma distribuição Qui-quadrado, χ^2 , com ng graus de liberdade. Então, para um determinado nível de significância α ,

$$P(h \geq \chi_{1-\alpha}^2 (ng)) = \alpha \quad (13)$$

onde ng é o número de equações do sistema.

Isto significa que a probabilidade particular de um evento h exceder o valor crítico χ^2 com um nível de significância $1 - \alpha$ com ng graus de liberdade é α . Isto nos provê um teste para determinar inconsistência de um conjunto de dados na presença de erros grosseiros. Faz-se necessário preestabelecer uma probabilidade de erro permissível que nos fornecerá um valor crítico de h . A escolha desta probabilidade dependerá das características do processo. Por exemplo, 0,10 é aceitável em diversos casos. Se um erro extremo for encontrado nas funções teste, precisa-se identificar qual medida contém o erro grosseiro para direcionar qualquer ação corretiva. Para identificação destes pontos extremos existem trabalhos consagrados na literatura como Nogita (1972).

Para casos práticos, propõe-se usar um modelo de detecção do erro com simples valores do vetor e_r , se a razão entre o erro extremo e a dispersão dos erros regulares não for tão pequena. De maneira a mostrar um teste para validação do modelo proposto, foi utilizado o exemplo proposto por Ripps(1965). O problema consiste de quatro fluxos mássicos sendo duas correntes de entrada e duas correntes de saída do reator. Todos os fluxos são medidos. As equações de balanço são:

$$\begin{aligned} 0,1X_1 + 0,6X_2 - 0,2X_3 - 0,7X_4 &= 0 \\ 0,8X_1 + 0,1X_2 - 0,2X_3 - 0,1X_4 &= 0 \\ 0,1X_1 + 0,3X_2 - 0,6X_3 - 0,2X_4 &= 0 \end{aligned} \quad (14)$$

As matrizes para o exemplo citado são:

$$G = \begin{bmatrix} 0,1 & 0,6 & -0,2 & -0,7 \\ 0,8 & 0,1 & -0,2 & -0,1 \\ 0,1 & 0,3 & -0,6 & -0,2 \end{bmatrix}$$

$$C_{\delta\delta} = \begin{bmatrix} 0,000289 & & & \\ & 0,0025 & & \\ & & 0,000576 & \\ & & & 0,04 \end{bmatrix}$$

A matriz de covariância do erro é dada pela Equação 11 e sua inversa pode ser encontrada como sendo :

$$C_{ee}^{-1} = \begin{bmatrix} 485 & -241 & -1340 \\ -241 & 5864 & -2061 \\ -1340 & -2061 & 5499 \end{bmatrix}$$

Com

$$e = \begin{bmatrix} -0,06694 \\ -0,00591 \\ -0,05707 \end{bmatrix}$$

Finalmente, com o auxílio da equação 12, obtém-se o valor de h:

$$|h| = e^T C_{ee}^{-1} e = 8,469$$

Se considerarmos um a probabilidade de erro igual a 0,10, então

$$8,469 > 6,25$$

e pode-se afirmar que a inconsistência é significativa com uma probabilidade de erro de 0,10.

A estrutura de classificação de variáveis desenvolvida por Oliveira Jr (2006) permite reduzir o problema combinatorial na seleção da medida que contém o erro grosseiro. As matrizes reduzidas G e E permitem a construção de um teste simples e efetivo para identificação de erros grosseiros persistentes.

4. CONCLUSÃO

O desenvolvimento da reconciliação de dados, detecção de erros grosseiros e sua implementação prática é trabalhosa e apresenta um alto custo, necessitando ser justificado através dos benefícios de uma determinada atividade industrial. A justificativa para a detecção de erros grosseiros durante o procedimento de reconciliação pode advir de alguns campos pouco explorados como a simulação e otimização de processos existentes cujos modelos contêm parâmetros que podem ser estimados através de dados industriais, onde os autores estão desenvolvendo uma tese de doutorado para processos de polimerização. O uso de medidas erradas no ajuste do modelo podem levar a parâmetros incorretos que podem anular os benefícios trazidos para operação da planta com os procedimentos de otimização.

A detecção de erros grosseiros não somente melhorará as precisões da estimativa dos procedimentos de reconciliação de dados, como também será útil em problemas de identificação na instrumentação que requeiram uma manutenção especial. A detecção incipiente de erros grosseiros de forma prática irá reduzir os custos de manutenção e prover a operação de unidades industriais com o mínimo de problemas. Os métodos podem ainda ser estendidos a detecção de falhas em equipamentos.

O presente trabalho centrado no desenvolvimento de estratégia de ramificação e contorno para visualização rápida e eficiente das medidas (em termos da função teste h) se mostrou eficiente para identificação de erro grosseiro em balanços de massa em estado estacionário.

1. ABUL-EL-ZEET, Z.H., BECERRA, V.M., ROBERTS, P.D., Combined Bias And Outlier Identification In Dynamic Data Reconciliation, *Comp. & Chem. Engin.* v.26, p. 921-935, 2002.
2. AMAND, Th., HEYEN, G., KALITVENTZEFF, B., Plant monitoring and fault Detection – Synergy between Data Reconciliation and Principal Component Analysis, *Comp. & Chem. Engin.* v.25, p. 501-507, 2001.
3. BAGAJEWICZ, M. J., A Brief Review of Recent Developments in data Reconciliation and gross Error Detection/Estimation, *Latin Amer. Apl. Res.*, v. 30, p. 335-342, 2000.
4. BAGAJEWICZ, M.J., JIANG, Q., Integral Approach To Plant Linear Dynamic Reconciliation, *AIChE Journal*, v. 43, p.2546-2558, 1997.
5. CHEN, J., BANDONI, A., ROMAGNOLI, J.A., Outlier Detection in Process Plant Data, *Comp. & Chem. Engin.*, v.22, p. 641-646,1998.
6. CHEN, J., ROMAGNOLI, J.A., A Strategy for Simultaneous Dynamic Data Reconciliation and Outlier Detection, *Comp. & Chem. Engineering.* v. 22, p. 559-562, 1998.
7. CROWE, C.M., Recursive Identification of Gross Errors in Linear Data Reconciliation, *AIChE Journal*, v. 34, p. 541-550, 1988.
8. MAH, R. S. H., *Chemical Process Structures and Information Flows*, 1 Ed. Stoneham. Butterworth,1990.
9. MCBRAYER, K.F., EDGAR, T.F., Bias Detection And Estimation In Dynamic Data Reconciliation”, *J. of Proc. Control.* v. 5, p. 285-289,1995.
10. NARASIMHAN, S., MAH, R. S. H., Generalized Likelihood ratio Method for Gross Error Detection, *AIChE Journal*, v. 33, p.1514-1521, 1987.
11. NOGITA, S., Statistical test and Adjustment of Process Data *Ind. Engin. and Chem. Proc. Dsg Devel.* 11, p.197-203, 1972.
12. OLIVEIRA JR, A . M., *Estimação de parâmetros em modelos de processo usando dados industriais e técnicas de reconciliação de dados.* Tese de Doutorado do Programa de Engenharia Química da Universidade Federal do Rio de Janeiro, 2006, 197 p.
13. RIPPS, D. L., 1965, *Chemical Engineering Processing Symposium Series*, n 61,p.8.
14. TAMHANE, A.C., MAH, R.S., Data Reconciliation and Gross error detection in Chemical Process Networks, *Technometrics*, v. 27, p. 409-422, 1985