



# PROPAGAÇÃO DE INCERTEZAS EM RECONCILIAÇÃO DE DADOS SEM REDUNDÂNCIA DE MEDIÇÃO.

*Leonardo Silva de Souza*<sup>1</sup>, *Ricardo Kalid*<sup>2</sup>, *Márcio A. F. Martins*<sup>3</sup>, *Asher Kiperstok*<sup>4</sup>, *Karla Esquerre*<sup>5</sup>, *Luciano Queiroz*<sup>6</sup>

<sup>1</sup> Rede de Tecnologias Limpas da UFBA (TECLIM), Salvador, Brasil, [leosilvasouza2001@gmail.com](mailto:leosilvasouza2001@gmail.com)

<sup>2</sup> Rede de Tecnologias Limpas da UFBA (TECLIM), Salvador, Brasil, [kalid@ufba.br](mailto:kalid@ufba.br)

<sup>3</sup> Programa de Pós-Graduação em Engenharia Industrial da UFBA, Salvador, Brasil, [marcio.engquimica@gmail.com](mailto:marcio.engquimica@gmail.com)

<sup>4</sup> Rede de Tecnologias Limpas da UFBA (TECLIM), Salvador, Brasil, [asher@ufba.br](mailto:asher@ufba.br)

<sup>5</sup> Rede de Tecnologias Limpas da UFBA (TECLIM), Salvador, Brasil, [karlaesquerre@ufba.br](mailto:karlaesquerre@ufba.br)

<sup>6</sup> Rede de Tecnologias Limpas da UFBA (TECLIM), Salvador, Brasil, [lmqueiroz@ufba.br](mailto:lmqueiroz@ufba.br)

**Resumo:** O presente trabalho visa estudar a avaliação da incerteza após reconciliação de dados em balanços hídricos, através do método de reconciliação de dados sem redundância (RECLIM). Os resultados mostram uma diminuição na incerteza da informação após a reconciliação de dados.

**Palavras chave:** reconciliação de dados, incerteza de medição, método de Monte Carlo, qualidade de informação.

## 1. INTRODUÇÃO

A inserção do computador na indústria possibilitou a aquisição de grande número de informações que podem ser usados para aperfeiçoar a precisão e a consistência dos dados de processo. Essa tecnologia quase que elimina o erro proveniente do registro manual, mas pode trazer dados corrompidos por erros (de medida, devido ao processamento ou transmissão do sinal). Existem diversas metodologias para eliminar esses erros inerentes as técnicas de medição, como o uso de filtros digitais, teste de controle de qualidade através de técnicas estatísticas e a reconciliação de dados.

As duas primeiras são muito importantes na análise individual dos dados, mas não garantem uma consistência inter-relacional entre as grandezas. A reconciliação de dados possibilita uma melhora na exatidão das medições, além de aumentar a possibilidade de que o resultado obtido seja fisicamente plausível; a expectativa dessa técnica é que os valores reconciliados sejam mais exatos do que os medidos. Baseado em [1], a formulação do problema de reconciliação de dados sujeito a restrições lineares passa por resolver a Eq. (1):

$$\min \Phi(\mathbf{V}_R, \mathbf{V}_M) = (\mathbf{V}_M - \mathbf{V}_R)^T (\mathbf{u}_M^2)^{-1} (\mathbf{V}_M - \mathbf{V}_R) \quad (1)$$

sujeito às seguintes restrições:

$$\mathbf{A}\mathbf{V}_R = \mathbf{0} \quad (2)$$

onde:

- $\mathbf{V}_R$  é o vetor das vazões reconciliadas;
- $\mathbf{V}_M$  é o vetor das vazões medidas;
- $\mathbf{u}_M^2$  é a matriz covariância das vazões medidas;

- $\mathbf{A}$  é a matriz incidência.

Através do método dos multiplicadores de Lagrange, a solução do problema definido pelas Eqs. (1) e (2) é dado pela Eq. (3) [2]

$$\mathbf{V}_R = \mathbf{S}_M^T \mathbf{V}_M = \left( \mathbf{I} - \mathbf{u}_M^2 \mathbf{A}^T (\mathbf{A} \mathbf{u}_M^2 \mathbf{A}^T)^{-1} \mathbf{A} \right) \mathbf{V}_M \quad (3)$$

onde  $\mathbf{I}$  é a matriz identidade.

Por definição [2,3], a matriz de covariância  $\mathbf{V}_R$  é dada por

$$\mathbf{u}_R^2 = \int (\mathbf{V}_R - \mathbf{E}[\mathbf{V}_R]) (\mathbf{V}_R - \mathbf{E}[\mathbf{V}_R])^T p(\mathbf{V}_R) d\mathbf{V}_R \quad (4)$$

sendo  $\mathbf{E}[\cdot]$ , o operador esperança, e  $p(\mathbf{V}_R)$ , a função de densidade de probabilidade (PDF) conjunta das variáveis  $\mathbf{V}_R$ . Conhecendo  $\mathbf{V}_R$  da Eq. (3), e aplicando a lei de propagação de incertezas multivariável [4], a matriz de covariância associada  $\mathbf{V}_R$  temos o resultado é dada por

$$\mathbf{u}_R^2 = \mathbf{S}_M^T \mathbf{u}_M^2 \mathbf{S}_M \quad (5)$$

Neste trabalho a incerteza é assumida como sendo a raiz quadrada positiva da variância (diagonal da matriz de covariância), ou seja, assume-se que as covariâncias são desprezíveis, i.e.,

$$\mathbf{u}_R = \sqrt{\text{diag}(\mathbf{S}_M^T \mathbf{u}_M^2 \mathbf{S}_M)} \quad (6)$$

É possível observar que a solução da reconciliação de dados passa por conhecer a matriz de covariância  $\mathbf{u}_M^2$ . Pelo método clássico (GUM), essa matriz é obtida a partir das medições existentes do sistema estudado. Devido à grande ausência de medidores em sistemas de transferência de água em indústrias petroquímicas, a avaliação da incerteza pelo método clássico não pode ser utilizada. Contudo, a metodologia da Rede TECLIM ([www.TECLIM.ufba.br](http://www.TECLIM.ufba.br)) para reconciliação de dados sem redundância de medição (RECLIM) [5] possibilita utilizar a vazão presumida ou mapeada de uma corrente, as quais devem ser qualificadas conforme a confiança depositada na fonte de informação, aqui classificada como Qualidade da Informação (QI). Isto é, um valor é atribuído para cada vazão mapeada de acordo com a qualidade dessas fontes de informação. Em outras palavras, a

$QI$  representa o nível de certeza que se tem sobre determinada informação. Portanto, a  $QI$  é inversamente proporcional a incerteza de uma variável.

## 2. OBJETIVO

Avaliar a incerteza das vazões industriais após a reconciliação de dados em sistemas sem redundância de medição.

## 3. MÉTODOS

A metodologia RECLIM permite caracterizar e gerir os fluxos de massa (e de energia) sem redundância de medição. As etapas são as seguintes:

- Identificação das fontes e consumidores de água (energia) e geradores de efluentes aquosos da unidade estudada;
- Atribuição da qualidade da informação ( $QI$ ) para cada vazão de água (fluxo de energia) do sistema;
- Especificação das restrições operacionais;
- Execução da reconciliação de dados aos balanços de massa (energia) dos sistemas de transferência de águas internas, fazendo uso das correntes mapeadas (vazões ou fluxos de energia) e da  $QI$ .

Chamamos a atenção para uma mudança de nomenclatura de vazão medida para vazão mapeada. Nessa metodologia, as vazões que não são medidas são estimadas a partir do conjunto de informações existente na empresa, tais como: relatórios, simulações, balanços de massa e informações oriundas das pessoas que operam o sistema. A cada um dessas estimativas fazemos uma análise da fonte de informação [6-7] quanto a sua confiabilidade e representatividade. Fazemos uma correlação temporal, se a fonte de informação nos fornece os dados recente ou não, e uma correlação tecnológica que depende da forma pela qual se obteve os dados. O resultado desta análise determina a  $QI$ .

A proposta de relação entre a  $QI$  e incerteza expandida é visto em [5]

$$QI_i = K_i \frac{V_{m_i}}{U_{m_i}} \quad (7)$$

onde  $V_{m_i}$  é a vazão mapeada da corrente  $i$ ,  $U_{m_i}$  é a incerteza expandida associada a vazão mapeada da corrente  $i$  antes da reconciliação e  $QI_i$  é a qualidade da informação de cada vazão mapeada. O parâmetro  $K_i$  é uma constante intrínseca do sistema estudado.

Para identificar as variáveis medidas é utilizado o subscrito  $M$ ; por outro lado, as variáveis que são conhecidas por outros meios, aqui denominada de variáveis mapeadas, recebem o subscrito  $m$ ; contudo essa distinção é apenas didática, pois ambas são estimadas, e a cada uma pode ser associada uma  $QI$  e uma incerteza expandida ( $U$ ) ou incerteza padrão ( $u$ ); além disso, podem-se converter essas propriedades umas nas outras, como será visto logo adiante.

Martins *et al.* [5] mostrou que o parâmetro  $K$  é praticamente constante para balanços hídricos, portanto combinando a Eq. (1) com a Eq. (7) e considerando  $K_i = K$  constante, obtém-se

$$\min \Phi(V_{R_i}) = \sum_{i=1}^N (V_{m_i} - V_{R_i})^2 \left( \frac{QI_i}{V_{m_i}} \right)^2 \quad (8)$$

A Eq. (8), agora, é a função objetivo sujeito às mesmas restrições da Eq. (2). Com a Eq. (8) é possível aplicar reconciliação de dados a sistemas com medição precária, utilizando informações qualitativas representadas pelo  $QI$ .

Para avaliar as  $QIs$  após a reconciliação de dados, é necessário avaliar as incertezas expandidas (as quais dependem das incertezas padrão e dos fatores de abrangência) das vazões reconciliadas, o que implica conhecer as incertezas expandidas das vazões mapeadas.

Na prática ou as vazões são medidas e pode-se obter a  $QI$  correspondente através da Eq. (7) ou é conhecido apenas valores máximo e mínimo de cada vazão, e neste caso é razoável assumir uma PDF uniforme para as vazões mapeadas, cuja incerteza expandida ( $U_{m_i}$ ) de cada vazão mapeada é dada por

$$U_{m_i} = k_{m_i}(\beta) \Delta_{m_i} \quad (9)$$

onde:  $k_{m_i}$  é o fator de abrangência, o qual depende da probabilidade de abrangência  $\beta$  escolhida;  $\Delta_{m_i}$  é a amplitude da PDF uniforme. Logo, da Eq. (7) podemos obter as incertezas expandidas (associadas às  $QIs$  atribuídas) das vazões mapeadas, enquanto que da Eq. (9) podemos determinar as amplitudes das PDFs uniformes, como consequência, as incertezas padrão de cada vazão mapeada podem ser obtidas pela Eq. (10):

$$u_{m_i} = \frac{\Delta_{m_i}/2}{\sqrt{3}} \quad (10)$$

Caso a PDF não seja uniforme, por exemplo,  $t$ -Student, o procedimento adotado envolvem as Eqs. (7), (11) e (12).

$$k_{m_i} = t^{-1}(\nu_i, \beta_i) \quad (11)$$

$$u_{m_i} = \frac{U_{m_i}}{k_{m_i}} \quad (12)$$

onde:  $k_{m_i}$  é o fator de abrangência que assume valores diferentes a depender dos graus de liberdade ( $\nu_i$ ) e da cada probabilidade de abrangência  $\beta_i$  de cada variável. A  $QI$  e os  $\nu_i$  tem uma relação conceitual entre si: quanto maior os  $\nu_i$  de um sistema, possuímos mais e melhores informações sobre o mesmo; portanto pode-se sugerir uma tabela de correspondência entre o  $QI$  e o  $\nu_i$ , a ser ajustado a cada problema. Por exemplo, se é considerado que a informação é totalmente confiável, pelo Princípio da Entropia Máxima (PEM) [8], o  $\nu_i$  é infinito e o fator de abrangência é definido por uma PDF normal.

Neste trabalho adotamos a probabilidade de abrangência ( $\beta$ ) igual a 90%, pois para dados de campo, a variabilidade das vazões é bastante significativa, logo valores maiores de  $\beta$

(95 ou 99%) conduzem a resultados fisicamente inexistentes (incertezas expandidas muito grandes).

Adotamos uma métrica para avaliar o quanto a reconciliação influenciou de forma global nos valores reconciliados: qualidade de informação global  $QI_{GR}$  após a reconciliação.

$$QI_{GR} = \frac{\sum V_{R_i} QI_{R_i}}{\sum V_{R_i}} \quad (13)$$

e, por conveniência, podemos definir a qualidade de informação global antes da reconciliação como

$$QI_{G_m} = \frac{\sum V_{m_i} QI_{m_i}}{\sum V_{m_i}} \quad (14)$$

É também interessante avaliar o impacto da reconciliação de dados sobre a incerteza do sistema. Neste trabalho adotamos a norma euclidiana para representar a incerteza global antes e depois da reconciliação, conforme representado pelas Eqs. (15) e (16), respectivamente.

$$u_{G_m} = \|\mathbf{u}_M^2\| = \sqrt{\sum \text{diag}(\mathbf{u}_M^2)} \quad (15)$$

$$u_{G_R} = \|\mathbf{u}_R^2\| = \sqrt{\sum \text{diag}(\mathbf{u}_R^2)} \quad (16)$$

#### 4. RESULTADOS

Um estudo de caso será mostrado aqui, reportado em [5]. Tem-se um sistema industrial com 8 nós e 13 correntes, como mostrado na Figura 1:

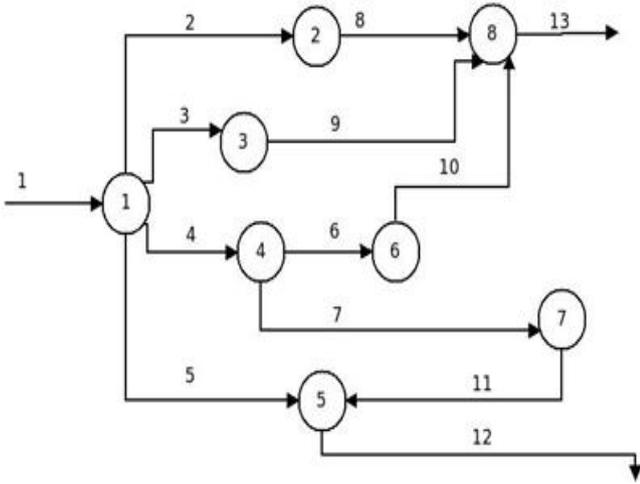


Fig. 1: Topologia de um sistema de distribuição de água.

Temos as informações de vazão mapeada e suas respectivas  $QI$ s. Aplicamos a função objetivo definida pela Eq. (8) e todos os passos da metodologia RECLIM, citados na seção 2.

Conforme é recomendado no trabalho de Martins *et al.* [5], a constante  $K_i$  utilizada no estudo de caso é igual a 0,1. Os resultados dos cálculos são apresentados na Tabelas 1, 2 e 3.

Tabela 1: Resultado da reconciliação de dados usando a metodologia RECLIM.

Correntes	$V_{m_i}$ (kg/s)	$V_{R_i}$ (kg/s)	$\left(\frac{V_{m_i}-V_{R_i}}{V_{m_i}}\right)$ (%)
1	28,1	28,2	0
2	3,1	2,9	-5
3	5,3	5,8	-10
4	8,9	8,5	4
5	11,4	11	4
6	3,4	4,5	-34
7	4,2	4,0	5
8	2,8	2,9	-5
9	5,8	5,8	0
10	4,4	4,5	-2
11	4,0	4,0	-2
12	15,0	14,9	0
13	13,3	13,3	0

É possível identificar que houve uma mudança nos valores de vazão e que agora zera-se os balanços no nosso sistema. Com os dados da Tabela 1, fez-se a avaliação da qualidade da informação após a reconciliação ( $QI_R$ ) através da metodologia de propagação de incerteza, usando as Eqs. (3) e (6). Os resultados estão nas Tabelas 2 e 3.

Tabela 2: Comparação entre as incertezas padrão antes e após a reconciliação de dados.

Corrente	$u_{m_i}$ (kg/s)	$u_{R_i}$ (kg/s)	$\left(\frac{u_{m_i}-u_{R_i}}{u_{m_i}}\right)$ (%)
1	0,09	0,05	44
2	0,02	0,01	50
3	0,42	0,03	93
4	0,05	0,03	40
5	0,91	0,05	94
6	0,27	0,02	92
7	0,33	0,02	93
8	0,01	0,01	0
9	0,03	0,03	0
10	0,02	0,02	0
11	0,02	0,02	0
12	0,05	0,04	20
13	0,04	0,03	25
$u_G$	1,10	0,1	90

Tabela 3: Comparação entre as  $QI$ s antes e após a reconciliação de dados.

Corrente	$QI_{m_i}$	$QI_{R_i}$	$\left(\frac{QI_{m_i}-QI_{R_i}}{QI_{m_i}}\right)$ (%)
1	10,0	19,0	90
2	6,0	8,6	44
3	0,4	7,3	1714
4	6,0	10,8	80
5	0,4	7,5	1783
6	0,4	7,3	1732
7	0,4	6,6	1549
8	6,0	8,6	44
9	6,0	7,3	21
10	6,0	7,3	22
11	6,0	6,6	10
12	10,0	11,2	12
13	10,0	15,3	53
$QI_G$	6,8	12,1	77

Nas Tabelas 2 e 3 observa-se que as incertezas (ou  $QI$ s) individuais tenderam a diminuir. Isso é possível, pois agregações de informações (topologia dos fluxos hídricos ou balanço de massa do sistema) ao sistema influenciam positivamente no resultado final, ou seja, diminuem a incerteza e aumentam a  $QI$ . A incerteza global ( $QI$  global) do sistema aumentou depois da reconciliação de dados. A incerteza diminuiu, ou a  $QI$  aumentou, significativamente, levando-se a concluir que a técnica melhora a informação sobre o modelo usado.

Na metodologia RECLIM, é possível que certos valores de  $QI$  fiquem acima do limite máximo inicial, veja a Tabela 3. A escala de  $QI$  mudou, após a reconciliação, o maior valor 19,0 e o menor é 6,6. O re-escalamento do problema levou a uma faixa maior de valores de  $QI$ . Antes a diferença entre o pior valor e o melhor valor era de 10, após a reconciliação passou para 12.

### 3. CONCLUSÃO

Na metodologia RECLIM de reconciliação de dados, observa-se que as incertezas ( $QI$ s) das vazões melhoraram, ou seja, diminuem (aumentam), embora também identificássemos que as maiores incertezas (menores  $QI$ s) podem aumentar (diminuir); contudo a incerteza (qualidade) global sempre melhora, ou seja, sempre diminui (aumenta).

O resultado da reconciliação deve ser discutido com os engenheiros e operadores mais graduados da planta industrial para avaliar se as  $QI$ s antes da reconciliação são adequadas; caso seja determinado que novas  $QI$ s iniciais devam ser estabelecidas, a reconciliação de dados deve ser novamente executada. Esse processo de diálogo técnico é um momento de forte interação homem técnica é uma grande oportunidade para conhecer e melhorar o processo. Faz parte da metodologia RECLIM está sempre revendo os pontos de medições ou estimativas utilizados no cálculo e, se possível,

melhorar aqueles que suscitarem dúvidas. Essa etapa permite identificar erros no sistema de medição, pontos onde o consumo está acima do recomendado e pontos do sistema que necessitem de um investimento para melhorar a  $QI$ .

A metodologia apresentada nesse artigo foi e continua sendo aplicada com êxito em diversas empresas, mas pode ser melhorada em alguns aspectos:

- Método para estabelecer a  $QI$ . Nesse item vem sendo desenvolvido uma metodologia que usa lógica *fuzzy* para estimar mais consistentemente a  $QI$ .
- Desenvolvimento da metodologia RECLIM e propagação da incerteza para o caso com restrições não lineares.
- Uma análise da melhor métrica para quantificar a qualidade de informação global ou da incerteza global antes e depois da reconciliação.
- Aplicação da inferência bayesiana na reconciliação de dados e na determinação do  $QI$ .

### AGRADECIMENTOS

Os autores gostariam de agradecer à Fundação de Amparo à Pesquisa do Estado da Bahia pelo apoio financeiro.

### REFERÊNCIAS

- [1] S. Narasimhan, C. Jordarche, *Data Reconciliation and Gross Error Detection*. 1ª ed. Houston: Gulf Publishing Company, 2000.
- [2] BIPM, IEC, IFCC, ILAC, ISO, IUPAC, IUPAP e OIML. *Evaluation of Measurement Data - Guide to the Expression of Uncertainty in Measurement* (GUM 1995 with minor corrections). Joint Committee for Guides in Metrology, JCGM 2008:100.
- [3] Leonardo Silva Souza, *Propagação da incerteza na reconciliação de dados com restrições lineares*. Monografia (Especialização em Controle e Automação de Processos Industriais) – Escola Politécnica da Universidade Federal da Bahia, Salvador, 2011.
- [4] I. Lira, *Evaluating the measurement uncertainty: fundamentals and practical guidance*. Bristol: Institute of Physics Publishing, 2002.
- [5] M. A. F. Martins, C. A. Amaro, L. S. Souza, R. Kalid, A. Kiperstok, “New objective function for data reconciliation in water balance from industrial processes”, *Journal of Cleaner Production*, vol. 18, pp. 1184-1189, 2010.
- [6] B. Weidema, M. Wesnas, “Data quality management for life cycle inventories - An example of using data quality indicators”, *Journal of Cleaner Production*, vol.4, pp.167-174, 1996.
- [7] B. P. Weidema, “Multi-user test of the data quality matrix for product life cycle inventory data”, *The International Journal of Life Cycle Assessment*, vol. 3, pp. 259-265, 1998.
- [8] K. Weise, W. Woger, “A Bayesian theory of measurements uncertainty”, *Metrologia*, vol. 3, pp.1-11, 1993.