

Monitoramento resiliente de processos com reconciliação de dados e estimação simultânea de parâmetros sob falhas de sensores

Bruno C. Manhães^{a,*}, Gildeir L. Rabello^a, Marcelo F. de S. Alves^a, Thiago K. Anzai^b, Márcio Nele^a, José Carlos Pinto^a

^a Programa de Pós-Graduação em Engenharia de Processos Químicos e Bioquímicos, Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, CEP 21941-972, Brazil

^b Centro de Pesquisas Leopoldo Américo Miguez de Mello (CENPES), Petrobras – Petróleo Brasileiro S.A., Rio de Janeiro, RJ, CEP 21941-915, Brazil

* brunocmhs@peq.coppe.ufrj.br

RESUMO

A crescente complexidade dos processos industriais e a necessidade de operações seguras e eficientes têm elevado a importância dos sistemas de monitoramento. Sendo assim, a reconciliação de dados garante consistência nas medições, melhorando a confiabilidade das informações. Porém, seu desempenho pode ser comprometido por falhas nos sensores, o que dificulta a observabilidade do sistema. Este trabalho propõe uma metodologia inovadora de monitoramento resiliente, baseada em reconciliação de dados robusta, detecção de erros grosseiros e análise da observabilidade em tempo real, por meio da classificação de variáveis. Este último aspecto é o diferencial da metodologia, pois permite uma abordagem adaptativa à disponibilidade de medições, reavaliando a observabilidade e reconfigurando o monitoramento em uma rotina de janela móvel. Resultados obtidos no estudo de um CSTR simulado mostram que o método é capaz de manter a continuidade do monitoramento, reduzir a propagação de inconsistências e preservar condições mínimas de observabilidade, mesmo diante de falhas de sensores.

Palavras-chave: Reconciliação de dados; Monitoramento; Estimação de parâmetros; Tolerância a falhas .

1 Introdução

A crescente complexidade das plantas industriais tem aumentado a necessidade de sistemas avançados de monitoramento, nos quais a qualidade e a confiabilidade dos dados são fundamentais para a tomada de decisão (Miao et al., 2009). Nesse contexto, a Reconciliação de Dados (RD) desempenha um papel central ao impor consistência física às medições por meio de modelos e restrições, gerando estimativas mais confiáveis para controle, otimização e detecção de falhas (Crowe, 1989; Narasimhan e Jordache, 1999).

A RD é formulada como um problema de otimização com restrições, geralmente resolvido pelo método dos Mínimos Quadrados Ponderados, assumindo-se erros de medição normalmente distribuídos (Leibman et al., 1992). No entanto, a qualidade dos resultados depende não apenas da formulação matemática, mas também dos requisitos estruturais do sistema de medição, como a redundância e a observabilidade. Sem um conjunto suficiente de medições independentes, o problema torna-se indeterminado e a reconciliação perde significado físico (Crowe, 1989).

Na prática industrial, essas condições nem sempre são atendidas devido a falhas em sensores (Isermann, 2005). Esses problemas reduzem a observabilidade e comprometem a confiabilidade das estimativas, exigindo o uso de técnicas de detecção de erros grosseiros para evitar resultados inconsistentes (Bagajewicz e Jiang, 1997). Como consequência, é comum a adoção de estratégias operacionais conservadoras, o que reduz a eficiência do processo.

Apesar dos avanços na literatura, a maioria dos métodos adota uma estrutura de medições fixa, sem considerar a reconfiguração do sistema diante de falhas. Assim, a manutenção do monitoramento diante da perda de observabilidade permanece um desafio em aberto.

Para enfrentar esse problema, este trabalho propõe uma metodologia inovadora de monitoramento baseada na adaptação do problema de reconciliação em tempo real, utilizando análises de observabilidade para orientar decisões automáticas. A abordagem busca manter a continuidade do monitoramento, reduzir o conservadorismo

operacional e aumentar a confiabilidade das estimativas e é avaliada em um reator CSTR simulado.

2 Metodologia

2.1 Aquisição e caracterização de dados

Os dados foram obtidos da simulação de um CSTR por Melo, 2020, baseada em Feital et al., 2013, cujos detalhes de modelagem estão nos trabalhos citados. Para o tempo de testes, foram usados 960 pontos, em vez de 2880, com mudanças de regime a cada 320 amostras. Para tornar o cenário mais realista, os parâmetros E , k_0 e U foram considerados desconhecidos e estimados ao longo do monitoramento, enquanto os demais permaneceram fixos. Todas as variáveis foram inicialmente tratadas como medidas, exceto a constante cinética k , obtida pela relação de Arrhenius. Para avaliar o método sob perda de informação, foram introduzidas duas falhas, com substituição das medições por valores não numéricos (NaN) nos intervalos 150–400 e 600–850.

Para caracterizar as variáveis do estudo de caso, foi aplicado, fora do loop de monitoramento, um procedimento de filtragem com janelas móveis não sobrepostas, no qual valores não numéricos e pontos influentes (outliers) foram substituídos pelo último valor válido. Os outliers foram definidos como valores fora dos limites inferior e superior, calculados como a mediana da janela $\pm 3\sigma_i$, sendo σ_i o desvio padrão da própria janela. Essa abordagem impede que valores extremos distorçam a caracterização de Menezes et al., 2020.

Com os dados filtrados, foram então obtidos, para cada variável, os valores mínimo, máximo, mediano e o desvio padrão. Este último foi calculado pelo espectro de variância com janela de 3 pontos (Feital e Pinto, 2015).

2.2 Classificação de variáveis

A classificação de variáveis é um elemento central da metodologia proposta, pois permite avaliar a observabilidade e a redundância do sistema com base nas equações e medições disponíveis, requisitos para a reconciliação de dados. Neste trabalho, essa etapa é realizada pelo método da matriz de incidência, proposto por de Oliveira Junior, 2006, aplicável a sistemas lineares e não lineares e utilizado no presente estudo como base para as heurísticas de decisão adotadas.

Ressalta-se que apenas parte do procedimento de classificação foi empregada, pois o foco do estudo está na reconfiguração dinâmica do sistema de monitoramento e reconciliação, e não no desenvolvimento completo do algoritmo. Assim, a implementação considera dois pontos de dados nos modelos transientes, o que é suficiente para demonstrar a viabilidade da metodologia, mas pode resultar em resultados conservadores. Dessa forma, os desempenhos obtidos representam um limite inferior, indicando que melhorias na etapa de classificação podem aumentar sua eficiência.

A metodologia convencional de monitoramento baseada em reconciliação de dados é apresentada na Figura 1, na qual a classificação de variáveis atua como etapa de validação da reconciliação, integrada à detecção de erros grosseiros para o tratamento de medições inconsistentes.

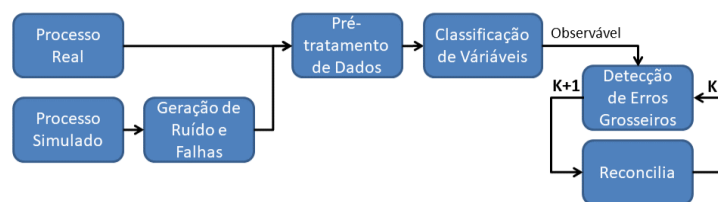


Figura 1: Metodologia convencional de Reconciliação de Dados.

Na metodologia proposta, a classificação de variáveis orienta a adaptação do algoritmo, o que impede instabilidades numéricas ou problemas mal condicionados quando a observabilidade ou a redundância mínima deixam de ser atendidas devido a falhas. Sempre que ocorre uma mudança no estado de um sensor (Online $\rightarrow$ Offline ou Offline $\rightarrow$ Online), a classificação é acionada para verificar se o sistema permanece observável. Simultaneamente, as falhas detectadas são impedidas de influenciar a otimização da reconciliação, conforme a Heurística 1. Essa interação entre o estado dos sensores, detecção de erros grosseiros, classificação de variáveis e reconciliação permite a reconfiguração adaptativa do sistema diante de perdas de medição. Caso seja detectada uma condição de não observabilidade, a Heurística 2 é ativada para restaurar a observabilidade do processo.

2.3 Heurísticas

A Heurística 1 verifica, a cada iteração da janela deslizante, o estado dos sensores e reduz o efeito de medições incorretas na reconciliação. As variáveis são tratadas em janelas de H pontos, e cada variável i é representada por um vetor de medições e um vetor de desvios-padrão com valores constantes, obtidos a partir do espectro de variância. Assim, todos os pontos da janela recebem o mesmo peso estatístico, assumindo que a variância é aproximadamente constante em curtos intervalos de tempo.

A detecção de erros grosseiros é realizada de forma simplificada, seguindo o mesmo critério adotado na filtragem da caracterização estatística para identificar valores não numéricos e anômalos. Em ambos os casos, o desvio padrão associado é substituído por 10^9 , o que reduz sua influência na reconciliação por meio de uma função objetivo robusta. No caso de valores não numéricos, a medição também é substituída pelo último valor válido da variável. Neste trabalho, adota-se a função de Welsch (Dennis Jr. e Welsch, 1978) como estimador robusto, cujas propriedades foram investigadas por de Menezes et al., 2021. Como o estudo utiliza janelas móveis e um modelo com equações diferenciais, a reconciliação dinâmica é formulada pela função objetivo baseada nesse estimador:

$$\min \sum_{t=t_{c-H}}^{t_c} \sum_{i=1}^{N_v} \rho_W(\xi_{t,i}, c_W) = \min \sum_{t=t_{c-H}}^{t_c} \sum_{i=1}^{N_v} c_W^2 \left[1 - \exp \left(- \left(\frac{\xi_{t,i}}{c_W} \right)^2 \right) \right], \quad (1)$$

Sujeito às restrições físicas do modelo h , às inequações g e às equações diferenciais f . O parâmetro H define o tamanho da janela, enquanto t_{c-H} e t_c indicam, respectivamente, o início e o fim da janela. A constante $c_W = 2,9846$, obtida por de Menezes et al., 2021, corresponde a uma eficiência relativa de 95% sob a hipótese de erros normalmente distribuídos. Além disso, o resíduo padronizado da variável i na janela t ($\xi_{t,i}$) é definido como a diferença entre o valor reconciliado e o valor medido, dividida pelo desvio padrão da variável i . Quando o desvio padrão de uma falha, σ_i , é elevado, o resíduo padronizado, $\xi_{t,i}$, tende a zero e o custo associado à variável passa a ser quase nulo. Dessa forma, o algoritmo reduz automaticamente a influência dessa medição, permitindo que seja estimada com base nas demais variáveis, sem alterar o problema de reconciliação. Esse mesmo mecanismo define o *status* dos sensores. Quando menos de 10% dos dados da variável i na janela t são confiáveis, isto é, não apresentam desvio de 10^9 , a variável é classificada como *Offline*; caso contrário, recebe o *status Online*.

Sempre que ocorre uma mudança de *status*, o algoritmo executa automaticamente a classificação de variáveis para verificar a observabilidade após a perda ou o ganho de uma medição. Se o sistema for observável, o algoritmo segue para a geração de limites e, em seguida, para a otimização. Se a perda de uma variável tornar o sistema não observável, a Heurística 2 é acionada para restaurar a observabilidade com base em duas ferramentas: *Hierarquia* e *Relações*.

A *Hierarquia* é uma análise de sensibilidade que mede o impacto de cada variável sobre a função objetivo do modelo. Para isso, seleciona-se um ponto de operação em regime estacionário, substituem-se os valores das variáveis pelos valores medianos e, em seguida, cada variável é perturbada individualmente com seus valores máximo e mínimo, mantendo as demais constantes. A influência de cada variável é medida pelo módulo da variação na função objetivo causada por essas perturbações. Quanto maior essa variação, maior a hierarquia da variável. Já as *Relações* descrevem as conexões estruturais entre variáveis, parâmetros e equações. Para cada variável, o algoritmo armazena o conjunto de elementos que aparecem nas equações do modelo, associado a ela. Esse mapeamento é obtido em Python a partir da leitura de um arquivo de texto com o modelo, do qual são extraídos, por meio de expressões simbólicas, as variáveis e os parâmetros de cada equação. O resultado é uma estrutura do tipo *varivel* $\rightarrow$ *equao* : *varivel* que explicita as dependências do sistema e permite identificar variáveis que compartilham equações.

O uso conjunto de *Hierarquia* e *Relações* formam o núcleo da Heurística 2. A *Hierarchy* ordena as variáveis pela influência no sistema, enquanto as *Relations* mostram suas conexões estruturais. Com isso, o algoritmo seleciona, de forma iterativa, variáveis ou parâmetros a serem fixados para recuperar graus de liberdade e restaurar a observabilidade. A escolha prioriza uma variável *Offline* associada ao máximo possível de outras variáveis *Offline*, que tenha a menor hierarquia possível, isto é, menos influente, pois fixar uma variável pouco influente tende a reduzir a propagação de erros.

A variável escolhida não é fixada de imediato; o algoritmo apenas a marca com o *status Fixada*. Após cada escolha, a classificação de variáveis é executada novamente para verificar se a observabilidade foi restaurada. Se ainda não for o caso, o processo se repete até que o sistema volte a ser observável. Dessa forma, a estrutura

de monitoramento é mantida mesmo diante de falhas de sensores, evitando a interrupção da reconciliação em aplicações em tempo real. Uma vez restaurada a observabilidade, o método segue para a geração dos limites de otimização.

A geração dos limites é integrada à Heurística 2 e utiliza os *status* atualizados a cada janela, de modo que, para variáveis *Online*, os limites são definidos em torno dos valores confiáveis, com $\pm 3\sigma_i$. Para variáveis com falha, os limites são definidos com base no último valor válido disponível ou, se houver, no limite válido anterior. Para as variáveis *Offline*, os limites também são dados por $\pm 3\sigma_i$, centrados no último valor válido salvo. Já as variáveis *Fixadas* recebem limites praticamente fixos, centrados no último valor válido, com variação de $\pm 10^{-12}$.

Neste trabalho, a otimização é feita com o método estocástico Dual Annealing (DA), implementado em `scipy.optimize` (Virtanen et al., 2020). Os hiperparâmetros `maxfun` e `maxiter` são iguais a 50.000 e controlam, respectivamente, o número de avaliações da função objetivo e o número de iterações. O chute inicial é dado pelos valores medidos na primeira iteração e, nas iterações seguintes, pelo ótimo da janela anterior. No estudo de caso, a otimização foi realizada em janelas de 80 pontos, totalizando 327 valores otimizados por vez: 320 variáveis de entrada, 4 condições iniciais das variáveis de estado e 3 parâmetros constantes em cada janela.

Após a descrição dos principais elementos do algoritmo, a rotina de monitoramento pode ser resumida na Figura 2. Em cada instante de medição K , correspondente a uma janela de dados, o algoritmo inicia a detecção de erros grosseiros com a Heurística 1. Em seguida, checka se houve mudança no *status* das variáveis monitoradas. Se não houver mudança, o algoritmo segue diretamente para a reconciliação ou para a estimação dos dados. Caso contrário, a classificação de variáveis é acionada para avaliar a observabilidade do sistema e, se necessário, definir as variáveis ou parâmetros a serem fixados pela Heurística 2.

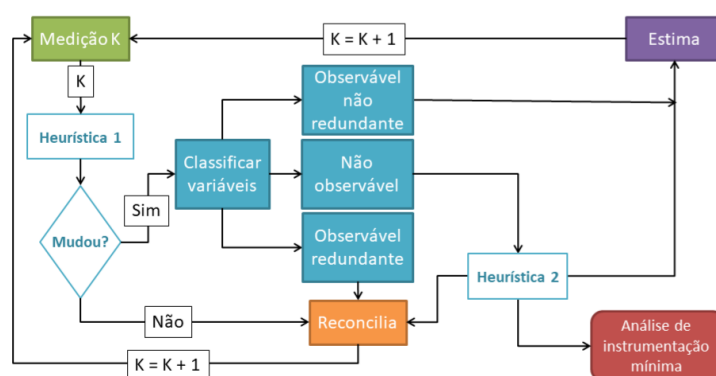


Figura 2: Esquema do monitoramento proposto

Vale ressaltar que o objetivo dessa estratégia não é substituir definitivamente as medições, mas evitar que perdas temporárias de informação tornem a operação excessivamente conservadora, mantendo uma base consistente para a tomada de decisão segura. Embora a ausência de algumas medições possa ser aceitável em determinadas situações, falhas persistentes ou agravadas reduzem progressivamente a confiabilidade dos resultados, aumentam a incerteza operacional e podem comprometer a validade das informações monitoradas, o que traz riscos à operação segura e eficiente do processo.

3 Resultados

Como discutido na Seção 2.2, a classificação adotada não abrange todos os dados da janela, o que pode tornar as decisões de fixação mais conservadoras, pois a análise estrutural não explora plenamente a redundância disponível em janelas mais longas. Para validar a metodologia, realizou-se a reconciliação de dados em um cenário com dois eventos distintos de falha ao longo da simulação, conforme descrito na Seção 2.1: (i) perda das medições do volume do reator V e da vazão de saída F ; e (ii) perda simultânea das variáveis de estado V , C_A e T .

Ambos os cenários foram analisados conjuntamente para avaliar a adaptabilidade da metodologia diante de diferentes condições estruturais, incluindo perdas sucessivas de medições e a posterior recuperação. A resposta das heurísticas a esses eventos é apresentada na Tabela 1. Observa-se que a primeira falha não comprometeu a

observabilidade, não sendo necessária a fixação de variáveis. Por outro lado, a segunda falha resultou na perda de observabilidade, exigindo a fixação dos parâmetros E e U para restabelecer a estrutura de reconciliação.

Tabela 1: Sistema de fixação para o CSTR com falhas em V e F , e em V , T e C_A .

Intervalo	Var On	Var Off	Var Fix	Observabilidade
1 → 160	$F, F_i, F_j, V, T, T_i, T_j, T_{ji}, C_A, C_{Ai}$	E, U, k_0, k	—	Observável
161 → 400	$F_i, F_j, T, T_i, T_j, T_{ji}, C_A, C_{Ai}$	E, U, k_0, k, F, V	—	Observável
401 → 640	$F, F_i, F_j, V, T, T_i, T_j, T_{ji}, C_A, C_{Ai}$	E, U, k_0, k	—	Observável
641 → 800	$F, F_i, F_j, T, T_i, T_j, T_{ji}, C_{Ai}$	E, U, k_0, V, T, k, C_A	E, U	Não observável
801 → 960	$F, F_i, F_j, V, T, T_i, T_j, T_{ji}, C_A, C_{Ai}$	E, U, k_0, k	—	Observável

A Figura 3 apresenta a reconciliação dinâmica das variáveis na presença de falhas de medição. Os pontos pretos representam os dados medidos, os círculos azuis indicam os pontos removidos para teste e os símbolos + em laranja correspondem aos valores reconciliados. Observa-se que o método mantém trajetórias contínuas e consistentes, mesmo na ausência de medições. No primeiro intervalo de falha (amostras 150–400), os valores reconciliados de V e F acompanham adequadamente a dinâmica do processo. No segundo intervalo, as trajetórias de V e T permanecem estáveis, enquanto C_A apresenta uma leve descontinuidade durante a transição para o terceiro estado estacionário, sem comprometer a identificação do novo regime. De forma geral, os dados reconstruídos mantêm coerência com a média original e apresentam um baixo nível de ruído, evidenciando a capacidade das heurísticas de recuperar informações relevantes de forma consistente.

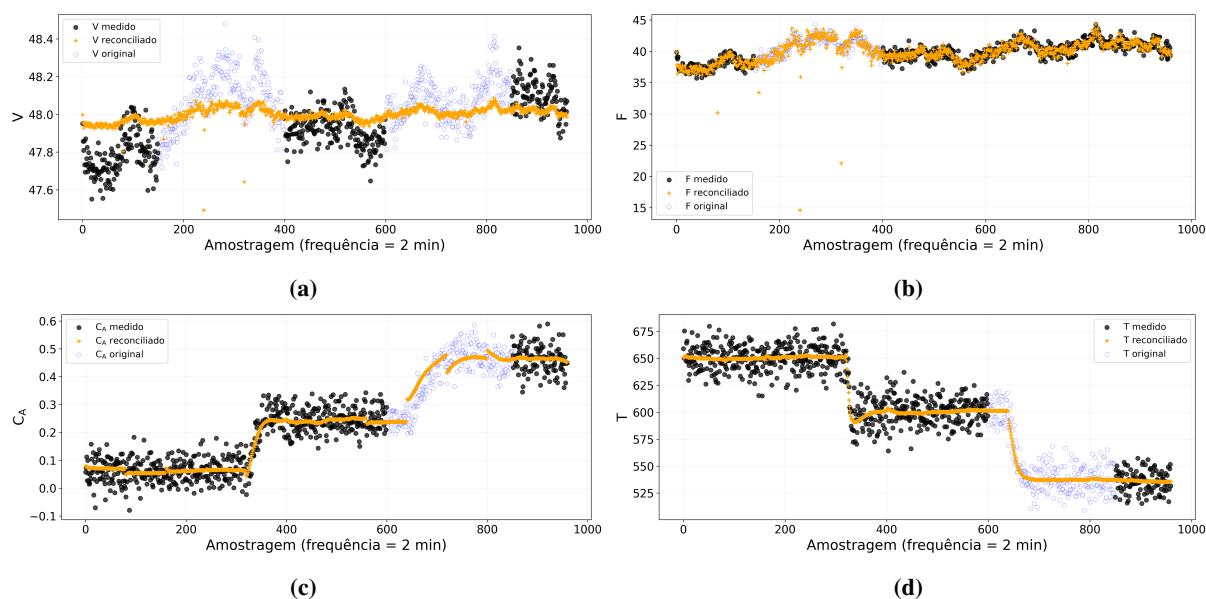


Figura 3: Sinais medidos e reconstruídos após a reconciliação sob condições de falha de sensores: (a) V medido e reconciliado; (b) F medido e reconciliado; (c) C_A medido e reconciliado; (d) T medido e reconciliado.

O tempo de otimização, avaliado em triplicata, apresentou média de 9,96 horas, inferior ao caso sem falhas (11,39 horas), para uma simulação equivalente a 32 horas de processo. Esse resultado indica que o método não aumenta o custo computacional, uma vez que as fixações reduzem o espaço de busca, além de evidenciar seu potencial para aplicações em tempo real.

4 Conclusões

O presente trabalho propôs e validou uma metodologia inovadora para aumentar a resiliência do monitoramento de processos industriais diante de falhas de sensores. A qual utiliza informações sobre o estado de

observabilidade do monitoramento para reduzir o impacto da perda de dados por meio de uma rotina em janela móvel que integra detecção de erros grosseiros, classificação de variáveis e reconciliação de dados com estimação de parâmetros. Ao combinar elementos *model-based* e *data-driven*, a metodologia configura uma abordagem híbrida capaz de adaptar a estrutura do problema diante de perdas de medição.

O estudo de caso com o CSTR demonstrou que o método mantém o monitoramento mesmo diante de falhas, evitando a propagação de erros e permitindo a reconfiguração de variáveis para preservar a observabilidade e a continuidade do sistema. Nesse contexto, a metodologia atua tanto na mitigação da degradação da qualidade dos dados, por meio da reconciliação baseada em modelo, quanto na manutenção da continuidade operacional. Como consequência, fornece estimativas fisicamente consistentes que reduzem conservadorismo desnecessário, mantendo uma base confiável para uma operação segura.

Entre as limitações e sugestões para trabalhos futuros, destaca-se que a etapa de classificação ainda não incorpora os ganhos de informação proporcionados pelas janelas móveis, o que pode levar a decisões conservadoras, como fixações desnecessárias, além da dependência da escolha de hiperparâmetros.

Agradecimentos: Os autores agradecem o financiamento oferecido pela Petrobras (projeto de P&D, SAP: 4600680239) e pela CAPES (financiamento #001).

Referências

- Bagajewicz, M. J., & Jiang, Q. (1997). Integral approach to plant linear dynamic reconciliation. *AIChE Journal*, 43(10), 2546–2558.
- Crowe, C. M. (1989). Observability and redundancy of process data for steady state reconciliation. *Chemical Engineering Science*, 44(12), 2909–2917.
- de Menezes, D. Q. F., Prata, D. M., Secchi, A. R., & Pinto, J. C. (2021). A review on robust M-estimators for regression analysis. *Computers & Chemical Engineering*, 147, 107254.
- de Menezes, D. Q. F., de Sá, M. C. C., Fontoura, T. B., Anzai, T. K., Diehl, F. C., Thompson, P. H., & Pinto, J. C. (2020). Modeling of Spiral Wound Membranes for Gas Separations—Part II: Data Reconciliation for Online Monitoring. *Processes*, 8(9).
- Dennis Jr., J. E., & Welsch, R. E. (1978). Techniques for nonlinear least squares and robust regression. *Communications in Statistics - Simulation and Computation*, 7(4), 345–359.
- de Oliveira Junior, A. M. (2006). *Estimação de parâmetros em modelos de processo usando dados industriais e técnica de reconciliação de dados* [tese de dout., Universidade Federal do Rio de Janeiro].
- Feital, T., Kruger, U., Dutra, J., Pinto, J. C., & Lima, E. L. (2013). Modeling and performance monitoring of multivariate multimodal processes. *AIChE Journal*, 59(5), 1557–1569.
- Feital, T., & Pinto, J. C. (2015). Use of variance spectra for in-line validation of process measurements in continuous processes. *The Canadian Journal of Chemical Engineering*, 93(8), 1426–1437.
- Isermann, R. (2005). *Fault-diagnosis systems: an introduction from fault detection to fault tolerance*. Springer Science & Business Media.
- Leibman, M., Edgar, T., & Lasdon, L. (1992). Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques [An International Journal of Computer Applications in Chemical Engineering]. *Computers & Chemical Engineering*, 16(10), 963–986.
- Melo, A. (2020). *EPV PEQ Aula 4 - Clusterização*. <https://www.kaggle.com/code/afnriomelo/epv-peq-aula-4-clusteriza-o>
- Miao, Y., Su, H., Gang, R., & Chu, J. (2009). Industrial Processes: Data Reconciliation and Gross Error Detection. *Measurement and Control*, 42(7), 209–215.
- Narasimhan, S., & Jordache, C. (1999). *Data reconciliation and gross error detection: An intelligent use of process data*. Elsevier.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17, 261–272.