

PREDIÇÃO DE COMPOSIÇÃO EM TEMPO REAL POR SENSOR VIRTUAL INDUSTRIAL BASEADO EM CAUSALIDADE FÍSICA

H. S. Vargens¹, A. F. Santos²

¹Universidade Federal do Paraná, Curitiba, Brasil (hermannvargens@gmail.com)

²Universidade Federal do Paraná, Curitiba, Brasil

A medição em tempo real da composição é vital para otimizar colunas de destilação, reduzindo energia e variabilidade. Este trabalho propõe um sensor virtual baseado em Regressão por Mínimos Quadrados Parciais (PLS), utilizando engenharia de atributos causal e alinhamento de atrasos. Em testes cegos, o modelo alcançou $R^2=0,9620$ e $RMSE=0,0396$. O desempenho é equivalente ao de redes neurais, porém com maior interpretabilidade, robustez e facilidade de implementação industrial.

Palavras-chave: Soft Sensor; PLS; Coluna Debutanizadora; Engenharia de *Features*; Validação Cruzada Temporal).

INTRODUÇÃO

A Indústria 4.0 busca integrar tecnologias inteligentes como a Inteligência Artificial e Internet das Coisas aos processos industriais, sendo a destilação um dos principais focos devido ao seu elevado consumo energético (SCHACK et al., 2020). Nesse contexto, a debutanização é o processo de destilação que visa fracionar a nafta não estabilizada, enriquecendo as correntes de fundo com compostos mais pesados (C5+) e removendo butanos e componentes mais leves pelo topo. A capacidade manter as especificações do produto de fundo está diretamente ligada ao gasto energético no refeedor, o qual precisa ser otimizado para não haver *overspecification* (FATIA, et al., 2022).

No entanto, a medição convencional, realizada através de cromatógrafos gasosos em linha, limita a qualidade do controle preditivo, principalmente por causa do tempo morto de análise e transporte, o qual pode chegar a 45 minutos (JIANG, et al., 2021; FORTUNA, et al., 2007). Assim, possíveis perturbações no processo não podem ser antecipadas de forma eficaz, tornando as malhas de controle reativas e não preditivas.

Diversos pesquisadores têm buscado superar essas limitações através de sensores virtuais (*soft sensors*). Soluções baseadas em redes neurais do tipo *Perceptron* Multicamadas (MLP), como os modelos NARX (*Nonlinear Autoregressive with Exogenous Inputs*) propostas por Fortuna et al. (2007), demonstram alta acurácia, alcançando coeficientes de correlação próximos a 0,982 (FORTUNA, et al., 2007). Contudo, tais arquiteturas resultam em elevado custo computacional e riscos de instabilidade

numérica quando integradas em malhas de controle industrial (PERERA, et al., 2023).

Nesse cenário, o presente trabalho propõe utilizar o algoritmo de Regressão por Mínimos Quadrados Parciais (PLS), procurando demonstrar que o adequado tratamento dos dados e engenharia de *features* fundamentada na causalidade física e no alinhamento dos atrasos de transporte é capaz de obter desempenhos preditivos comparáveis aos modelos neurais não-lineares, contornando a complexidade e a instabilidade inerentes às arquiteturas de caixa-preta.

MATERIAL E MÉTODOS

Base de Dados e Amostragem

O *dataset* utilizado neste trabalho pode ser encontrado como material suplementar do livro *Soft Sensors for Monitoring and Control of Industrial Processes (Advances in Industrial Control)* de Fortuna, et al., (2007). Tratam-se de históricos de operação de uma coluna debutanizadora, com amostragem contínua de variáveis de processo, como medições de temperatura, pressões e vazões, além da concentração de C4 no produto de fundo, fornecida por um analisador cromatográfico em linha as quais são coletadas a intervalos de 12 minutos.

A Tabela 1 apresenta a lista de variáveis constantes no *dataset*, bem como a descrição de cada uma, conforme a literatura de referência. A Tabela 2 apresenta as nomenclaturas utilizadas na construção dos nomes das variáveis utilizadas, e será importante para melhor entendimento no tópico **Resultados e Discussão**.

Tabela 1. Lista de variáveis e descrição

Variável	Descrição
Top_Temp	Temperatura do topo da coluna
Top_Pressure	Pressão no topo da coluna
Reflux_Flow	Vazão de refluxo
Flow_to_Next_Process	Fluxo para o próximo processo
Temp_Sixth_Tray	Temperatura da 6ª bandeja (zona de transição)
Bottom_Temp_1	Temperatura de fundo 1
Bottom_Temp_2	Temperatura de fundo 2
C4_Fundo	Teor de C4 no fluxo de fundo (variável alvo)

Tabela 2. Lista de sufixos

Sufixo/Termo	Significado Técnico
_lag_4	Atraso puro de 45 min
_sq	Termo polinomial de 2ª ordem
_roll_mean	Média móvel (14 pontos)
_roll_std	Desvio-padrão móvel
_deriv	Diferença finita de 1ª ordem

Processamento Causal de Sinais

Para reduzir ruídos de alta frequência presentes nas leituras sensores, aplicou-se a Média Móvel Exponencial (EMA), com fator de *span* igual a 18,

garantindo que o tratamento do sinal seja realizado estritamente com base no histórico e no valor presente, viabilizando a aplicação em tempo real (HUNTER, 1986).

Alinhamento de Lags e Engenharia de *Features*

A matriz de entrada foi estruturada para capturar as dinâmicas linear e não-linear da coluna, totalizando 36 atributos preditores. Inicialmente, foi realizado alinhamento temporal das *features* originais utilizando a Função de Correlação Cruzada. Em seguida, foram acrescentadas *features* com base nas derivadas temporais discretas e médias móveis locais (janela de 14 pontos), e a expansão polinomial de grau 2, (YUAN, et al., 2020). Com objetivo de acrescentar um termo autorregressivo, tendo o intervalo de amostragem de 12 minutos e o atraso físico de transporte e análise do cromatógrafo, 45 minutos, definiu-se o deslocamento temporal mínimo para evitar vazamento de dados (*shift*) em 4 passos para a variável resposta ($y(k-4)$), ($45/12 = 3,75$) (GRAZIANI, S.; XIBILIA, M. G., 2021), recurso também utilizado por Fortuna, et al.,(2007).

Validação Cruzada Temporal

Para evitar o sobreajuste, o conjunto de dados foi dividido cronologicamente: 80% para treinamento e 20% para teste cego final, onde utilizou-se a técnica de *TimeSeriesSplit* do pacote *Scikit-Learn* do Python com 4 subdivisões, permitindo uma validação cruzada que respeita a natureza sequencial dos dados (HYNDMAN, R. J.; ATHANASOPOULOS, G., 2021).

RESULTADOS E DISCUSSÃO

Convergência e Redução de Dimensionalidade

A Figura abaixo mostra a curva do erro RMSE (*Root Mean Squared Error*) em função do número de componentes latentes.

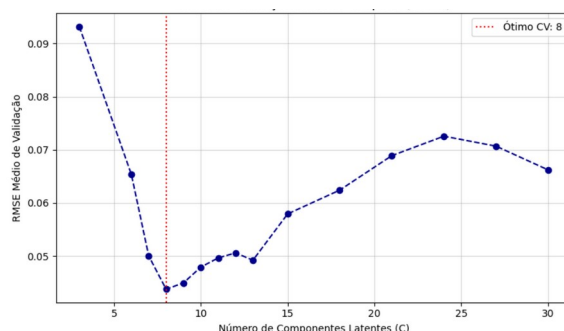


Figura 1. Curva de Validação Cruzada Temporal para o *dataset* de treino

É possível observar que o algoritmo PLS condensou eficientemente as 36 variáveis de entrada em apenas 8 componentes latentes, reduzindo a colinearidade e extraíndo as informações essenciais para a predição.

Desempenho no conjunto de teste

O modelo demonstrou alta capacidade preditiva na avaliação final, mantendo a consistência estatística mesmo em regime de operação não mapeado durante o treinamento. As métricas obtidas foram: MAE = 0,0306; RMSE = 0,0396; e $R^2 = 0,9620$.

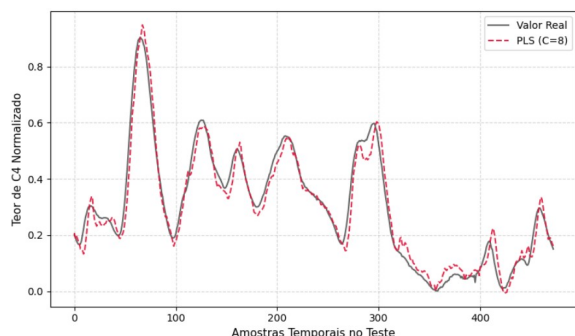


Figura 2. Comportamento dinâmico da predição do Soft Sensor PLS em comparação ao valor real de cromatógrafo no conjunto de teste.

Análise Comparativa com a Literatura

Ao comparar os resultados com os modelos MLP-NARX propostos por Fortuna *et al.* (2007), observa-se que o coeficiente de correlação do segundo modelo da literatura ($r = 0,982$, equivalente a $R^2 \sim 0,9643$) foi virtualmente igualado pelo modelo PLS aqui proposto ($R^2 = 0,9620$). Diferentemente das arquiteturas neurais, o PLS mitiga o efeito de acúmulo de erro dinâmico típico dos modelos NARX *multi-step*, garantindo estabilidade superior para a malha de controle.

Tabela 3. Comparativo de métricas de desempenho entre o modelo proposto e a literatura

Modelo	r	R^2
NARX (Fortuna et al., 2007)	0,982	0,9643
PLS	0,981	0,9620

Análise de Importância das Variáveis

A quantificação dos *VIP scores* para análise de importância das variáveis, revela a predominância de atributos que respeitam a física da coluna de destilação. A variável de defasagem pura (**C4_Fundo_lag_4**), que representa o atraso do analisador de processo, apresentou o maior impacto na predição ($VIP = 2,8$), validando o alinhamento temporal como o fator determinante para o sucesso do modelo. Além disso, a relevância das variáveis associadas à temperatura do 6º prato (**Temp_Sixth_Tray**) em suas diversas formas de processamento (valor original, quadrático e média móvel) demonstra que o modelo capturou com precisão a região de maior sensibilidade

termodinâmica da coluna. A Tabela 2 contém os sufixos de cada variável, necessários para compreender cada variável.

Tabela 4. Variáveis com *VIP Scores* >1

Variável	VIP Score
C4_Fundo_lag_4	2.827606
Temp_Sixth_Tray	1.751856
Temp_Sixth_Tray_sq	1.676645
Temp_Sixth_Tray_roll_mean	1.278630
Temp_Sixth_Tray_roll_std	1.142591
Top_Temp	1.135175
Top_Pressure	1.082342
Top_Temp_sq	1.031664
Top_Pressure_sq	1.024910
Reflux_Flow_sq	1.022110
Temp_Sixth_Tray_deriv	1.003802

Diagnóstico de Estabilidade e Resíduos

A análise de resíduos, fundamentada no teste de Shapiro-Wilk, resultou em uma estatística $W = 0,9908$, com *p-valor* de 0,0047. Embora o *p-valor* rejeite a hipótese nula — comportamento esperado devido ao elevado tamanho amostral do *dataset* —, a proximidade de W em relação a unidade valida matematicamente a eliminação de erros sistemáticos.

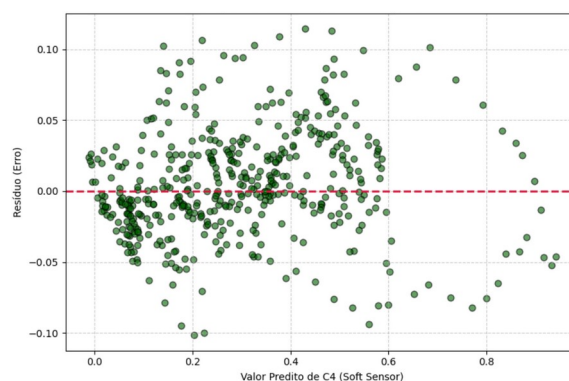


Figura 3. Dispersão de Resíduos

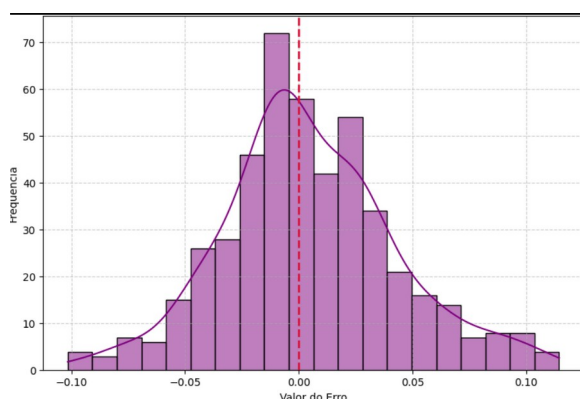


Figura 4. Distribuição Estatística dos Erros

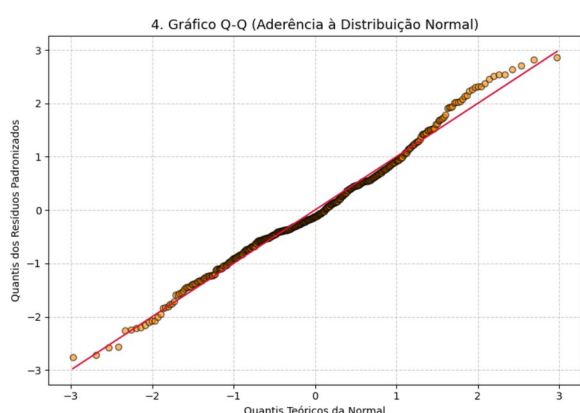


Figura 5. Gráfico Q-Q (Aderência à Distribuição Normal)

O ajuste apresenta robustez estatística, a qual confirma a aplicabilidade do modelo em condições de operação reais, assegurando que as estimativas em tempo real são capazes de otimizar os gastos energéticos da coluna debutanizadora.

CONCLUSÃO

O presente trabalho demonstrou que a implementação de sensores virtuais em colunas debutanizadoras não exige, necessariamente, a adoção de arquiteturas não-lineares complexas e de alto custo computacional. O modelo proposto, baseado em Regressão por Mínimos Quadrados Parciais (PLS), cumpriu integralmente os requisitos de confiabilidade operacional exigidos pela indústria de refino, apresentando um desempenho estatístico competitivo frente às soluções neurais consagradas na literatura.

A eficácia do *soft sensor* desenvolvido não reside na complexidade do algoritmo de aprendizado de máquina, mas na inteligência quimiométrica aplicada ao entendimento do fenômeno físico e dos atrasos temporais inerentes ao processo. A correta modelagem causal, integrada a um tratamento de dados que respeita as constantes de tempo da

instrumentação — notadamente o atraso de transporte do cromatógrafo de 45 minutos —, permitiu que algoritmos lineares bem estruturados competissem em pé de igualdade com redes neurais complexas.

Conclui-se, portanto, que a transformação digital pela qual a indústria vem passando, voltada à transição energética pode ser viabilizada por modelos interpretáveis, estáveis e de fácil manutenção. Esta abordagem reduz a variabilidade operacional e minimiza o consumo energético dos refeedores, provendo uma alternativa tecnicamente robusta e industrialmente escalável para o controle de qualidade em tempo real.

AGRADECIMENTOS

Agradeço à Universidade Federal do Paraná, ao Programa de Pós-Graduação em Engenharia Química (PPGEQ), ao Laboratório de Cinética Química Aplicada (LACTA) e ao Professor Doutor Alexandre Ferreira Santos.

REFERÊNCIAS

- FATIMA, S. A. et al. Intelligent Control of an Industrial Debutanizer Column. **Chem. Eng. Technol.**, v. 45, p. 667-677, 2022. DOI: 10.1002/ceat.202100039.
- FORTUNA, L. et al. **Soft Sensors for Monitoring and Control of Industrial Processes**. London: Springer-Verlag, 2007. [1, 2]
- GRAZIANI, S.; XIBILIA, M. G. On the use of correlation analysis in the estimation of finite-time delay in Soft Sensors design. In: IEEE INTERNATIONAL INSTRUMENTATION AND MEASUREMENT TECHNOLOGY CONFERENCE (I2MTC), 2021, Glasgow. **Anais [...]**. Piscataway: IEEE, 2021. p. 1-6. DOI: 10.1109/I2MTC50364.2021.9459807.
- HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3. ed. Melbourne: OTexts, 2021. Disponível em: <https://otexts.com/fpp3/>. Acesso em: 25 jun. 2025.
- HUNTER, J. S. The exponentially weighted moving average. **Journal of Quality Technology**, v. 18, p. 203-210, 1986.
- JIANG, Y. et al. A Review on Soft Sensors for Monitoring, Control, and Optimization of Industrial Processes. **IEEE Sensors Journal**, v. 21, n. 11, p. 12868-12881, 2021. DOI: 10.1109/JSEN.2020.3033153.
- PERERA, Y. S. et al. The role of artificial intelligence-driven soft sensors in advanced sustainable process industries: a critical review.

Engineering Applications of Artificial Intelligence,
v. 121, 105988, 2023. DOI:
10.1016/j.engappai.2023.105988.

SCHACK, D. et al. Energy-Efficient Distillation Processes by Additional Heat Transfer Derived From the FluxMax Approach. **Front. Energy Res.**, v. 8, 134, 2020. DOI: 10.3389/fenrg.2020.00134.

YUAN, X. et al. A Deep Supervised Learning Framework for Data-Driven Soft Sensor Modeling of Industrial Processes. **IEEE Trans Neural Netw Learn Syst.**, v. 31, n. 11, p. 4737-4746, 2020. DOI: 10.1109/TNNLS.2019.2957366.