

SELEÇÃO DE VARIÁVEIS SECUNDÁRIAS PARA MODELOS DE INFERÊNCIA EM UMA PLANTA INDUSTRIAL DE PRODUÇÃO DE ETILBENZENO

E. F. L. VIEIRA¹, P. R. A. MARIZ¹, BRANDÃO, W. Y. A¹, L. O. A. ROJAS¹, A. A. MORAIS JR¹

¹ Universidade Federal da Paraíba, Departamento de Engenharia Química
E-mail para contato: emanuella_vieira@hotmail.com

RESUMO – Muitas variáveis de processos químicos refletem o real estado de outras variáveis de saída, porém, quando se deseja construir modelos de identificação, a inserção de um número excessivo de variáveis converge para problemas recursivos de sobreajustes, bem como o oposto pode corromper a etapa de aprendizagem e boas estimativas poderão não ser alcançadas, como é o caso das estimativas por redes neurais artificiais (RNA). Muitas vezes, apenas o conhecimento fenomenológico do processo não é suficiente para selecionar variáveis secundárias importantes para o processo. Assim, desenvolveu-se e empregou-se em *software* Matlab® algoritmos de seleção de variáveis baseados em análise estatística multivariada, com objetivo de indicar os melhores modelos de inferência em um processo de produção de etilbenzeno (EB) para construção de sensores virtuais. Tal processo mostrou-se interessante para o presente trabalho, haja vista a ocorrência de efeitos transientes nas composições de alta pureza nas correntes de topo e de base da segunda coluna de destilação. Foram empregados os métodos da regressão por etapas (*stepwise regression*) e de todas as regressões possíveis (*all possible regressions*) para selecionar as variáveis de entrada dos modelos de inferência das composições de topo do EB e de base do dietilbenzeno (DEB). Com os resultados das técnicas utilizadas, concluiu-se que todas as variáveis analisadas exercem influência preponderante nas composições de saída dos compostos de interesse.

1. INTRODUÇÃO

A importância do monitoramento de variáveis críticas de processos utilizando dispositivos de medição adequados é evidente. Quando se deseja medir, em tempo real, variáveis químicas ou bioquímicas não facilmente mensuráveis, tais como composição dos produtos de colunas de destilação, surgem obstáculos de como efetuar-las (Fortuna *et al.*, 2007; Morais Jr, 2015).

O conhecimento a priori do processo (sentimento de engenharia) não pode e não deve ser descartado, mas por si só não é o suficiente para selecionar algumas variáveis secundárias importantes para o processo. Nesse caso, recorre-se comumente a técnicas de seleção de variáveis, a exemplo da análise estatística multivariada (busca exaustiva e sequencial) e métodos baseados em inteligência artificial (por exemplo, o algoritmo genético). Nesse

trabalho, foi desenvolvido um algoritmo de seleção de variáveis baseado na busca exaustiva em *software* Matlab®, através do método de todas as regressões possíveis (TRP), e empregado o uso de um método baseado em etapas sequenciais (regressão por etapas, do inglês *stepwise regression*). Tal seleção tem por objetivo indicar os melhores modelos de inferência para construção de sensores virtuais (*soft sensors*) de composição.

Na rota em estudo, o EB é produzido em dois reatores CSTR em série, a partir da reação do etileno (E) e do benzeno (B). Nesse processo, ocorre também a formação do composto dietilbenzeno (DEB), originado da reação secundária entre EB e B. A saída do segundo reator do processo é corrente de alimentação da primeira coluna de destilação multicomponente, sendo seu composto “chave leve” o benzeno (>99,9% molar), que é reciclado para o primeiro reator. A segunda coluna de destilação tem como alimentação a corrente de base da primeira coluna, fornecendo uma corrente de topo rica em EB (>99,9% molar) e uma corrente de base rica no DEB (>99,9% molar), sendo essa última uma corrente de reciclo para o segundo reator. Assim, o processo mostrou-se interessante para o presente trabalho, haja vista a ocorrência de efeitos transientes nas composições de alta pureza das correntes de topo e de base da segunda coluna de destilação.

2. SELEÇÃO DE VARIÁVEIS SECUNDÁRIAS

Dentre os métodos de seleção de variáveis destacam-se aqui os de busca exaustiva e os métodos sequenciais (*stepwise*).

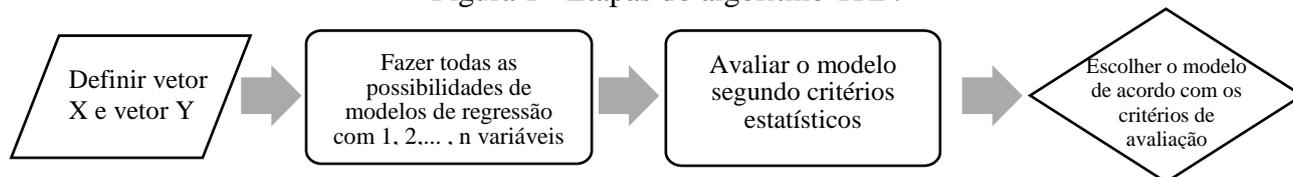
2.1. Todas as Regressões Possíveis (TRP)

Método de busca exaustiva que necessita do ajuste de todas as equações de regressão envolvendo 1, 2, ..., p variáveis candidatas. Considere um modelo de regressão linear múltipla e suas suposições, Equação 1:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (1)$$

A metodologia, cujas etapas estão apresentadas na Figura 1, avalia todos os submodelos compostos pelos possíveis subconjuntos das p variáveis e identifica os melhores desses subconjuntos através dos critérios de avaliação. Vale salientar que deverá existir um termo constante β_0 em qualquer subconjunto disponível para o modelo, e que, se existirem k regressores, serão gerados 2^k modelos de regressão. Assim, à medida que aumentam os candidatos, o número de equações aumenta rapidamente: dobra a cada adição de uma nova variável candidata (Montgomery e Runger, 2014).

Figura 1 - Etapas do algoritmo TRP.



2.2. Regressão em Etapas (*Stepwise Regression*)

É um procedimento iterativo que a cada etapa desenvolve uma sequência de modelos pela adição ou remoção de regressores candidatos, levando em consideração a influência da entrada ou saída de cada variável no modelo. O critério mais utilizado para avaliar a necessidade de adição ou remoção de regressores é o teste F parcial, conforme Equação 2:

$$F_j = \frac{SQR(\beta_2, \beta_0)}{MQE(x_1, x_2)} \quad (2)$$

Em que SQR é a soma quadrática dos resíduos e MQE é o erro médio quadrático.

Inicia-se o procedimento com o modelo que contém uma variável, aquela que possui a correlação mais alta com a variável de saída Y . Na próxima etapa, são analisados os regressores candidatos restantes, e a variável para qual a estatística parcial F é máxima é adicionada ao modelo se o valor de F calculado for maior que um valor f_{entra} estabelecido previamente. Analisa-se também se há a necessidade de remoção da variável adicionada na etapa anterior calculando-se a estatística F . Caso o resultado de F tenha sido menor que um valor pré-estabelecido f_{sai} , a variável adicionada anteriormente é removida e o modelo prossegue para a próxima etapa utilizando a mesma dinâmica.

2.3. Critérios Estatísticos de Avaliação

Geralmente, na seleção de modelos recorre-se a critérios de avaliação, tais como o coeficiente de determinação ajustado ou R^2_{aj} (Equação 3), erro médio quadrático ou MQE (Equação 4), além da estatística C_p , que é uma medida do erro quadrático médio total para a regressão (Equação 5).

$$R^2_{ajust} = 1 - \left(\frac{n-1}{n-p} \right) (1 - R^2) \quad (3)$$

$$MQE = \frac{SQE}{(n-p)} \quad (4)$$

$$C_p = \left(\frac{SQE(p)}{\sigma^2} \right) - n - 2p \quad (5)$$

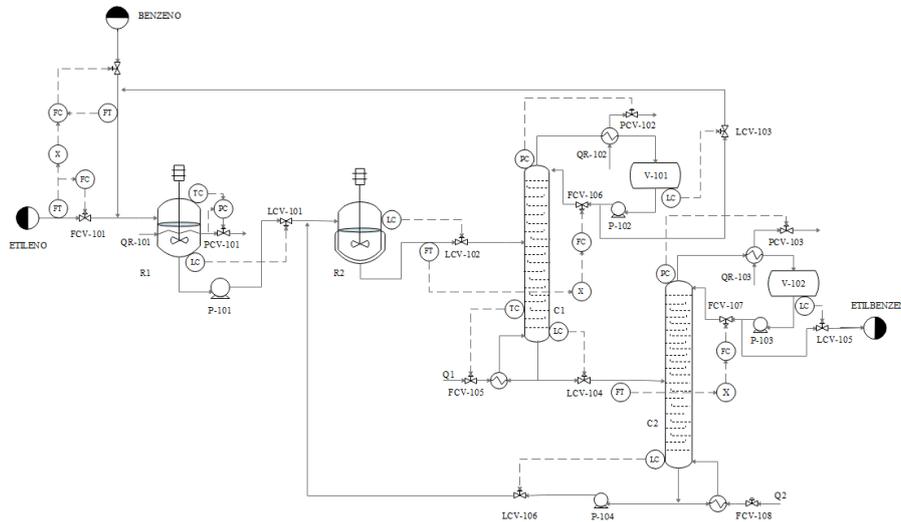
Sendo SQE a soma quadrática do erro, σ^2 a variância, R^2 o coeficiente de determinação, n o número de variáveis e p o número de amostras.

3. MODELAGEM E SIMULAÇÃO

O processo produtivo do EB, mostrado no fluxograma da Figura 2, foi simulado em regime estacionário no *software* Aspen PlusTM, com os dados de projeto e condições iniciais obtidos em Luyben (2011). Os reatores possuem volume de 200 m³ e operam a pressões de 19 e 20 atm, respectivamente. As colunas 1 e 2 (C1 e C2) possuem 21 e 25 estágios, com alimentações nos estágios 10 e 15, operando com pressões de 0,1 e 0,3 atm, respectivamente.

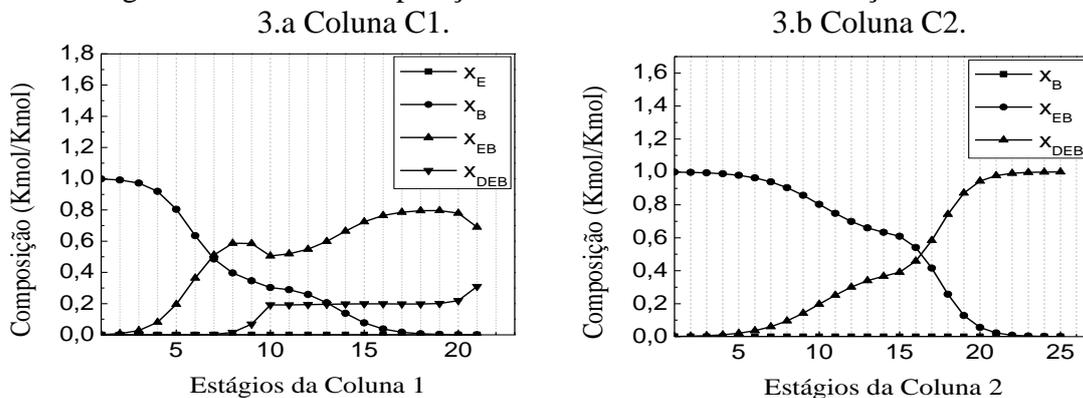
O modelo termodinâmico empregado para os cálculos das propriedades físicas foi de *Chao-Seader*. Os equipamentos do processo foram dimensionados para operar em regime transiente.

Figura 2: Fluxograma do processo produtivo do EB com controles típicos.



Os gráficos das Figuras 3.a e 3.b apresentam os perfis de composição ao longo dos estágios de C1 e C2. Verifica-se, de acordo com a figura 3.a, que em C1 o benzeno, x_B , sai em alta pureza no estágio 1 (topo). De acordo com a figura 3.b, o etilbenzeno e o dietilbenzeno, x_{EB} e x_{DEB} , estão em alta pureza nos estágios 1 e 25 (topo e base) de C2, respectivamente.

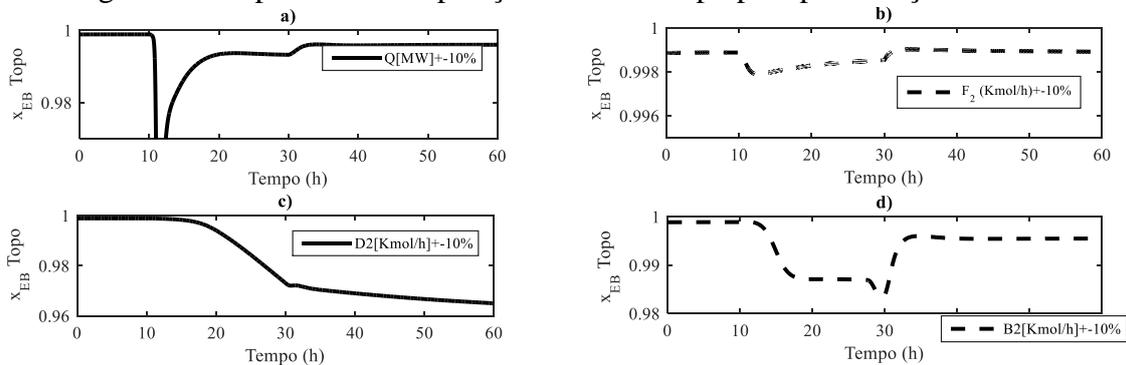
Figura 3. Perfis de composição molar nas colunas de destilação C1 e C2.



3.1. Procedimento para Coleta dos Dados

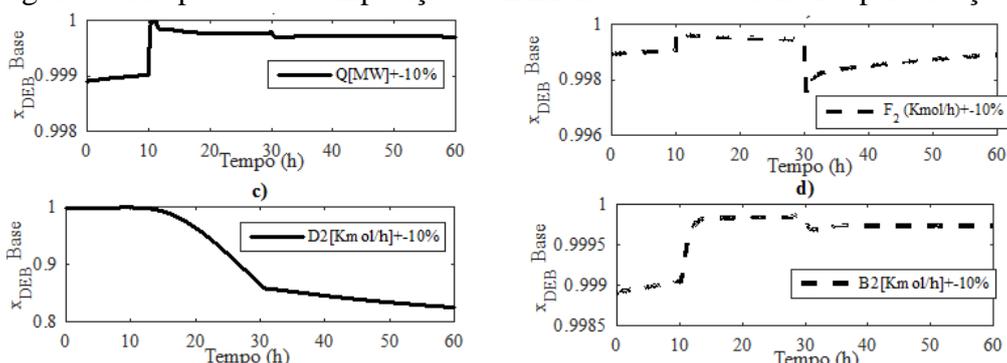
Após realizar e avaliar a simulação em regime estacionário, os equipamentos foram dimensionados e foi feita a conversão para o transiente, cuja simulação deu-se no *Aspen Dynamics™*, onde foram efetuadas perturbações nas seguintes variáveis de C2: carga térmica (Q_2), vazão de entrada (F_2), vazão de topo e de base (B_2) e (D_2), conforme gráficos das Figuras 4 e 5.

Figura 4 - Respostas da composição de EB no topo para perturbações em C2.



De acordo com os gráficos da Figura 4, verifica-se que a composição do etilbenzeno (x_{EB}) é sensível a mudanças nas variáveis Q_2 , F_2 , D_2 e B_2 , proporcionando para essa resposta de composição sobre-elevações percentuais consideráveis. Do mesmo modo, verifica-se que para os mesmos distúrbios efetuados, foram proporcionadas alterações na composição do dietilbenzeno na corrente de base da coluna 2 (x_{DEB}), Figura 5.

Figura 5 - Respostas da composição de DEB na base da coluna 2 a perturbações.



Para o procedimento de coleta dos dados, foram inseridos distúrbios com amplitude de $\pm 10\%$ em F_2 , D_2 e B_2 ; e sinais pseudo-randômico-aleatórios (PRBS) em Q_2 , com uma amplitude de 5%.

Após avaliar as respostas aos distúrbios, notou-se que todas as variáveis previamente selecionadas exerciam influência significativa sobre as composições. Formou-se com elas o conjunto de 8 regressores, como mostrado na Tabela 1, para a seleção de variáveis dos modelos de inferência das composições do EB e DEB no topo e na base da segunda coluna, respectivamente.

Tabela 1 – Lista de variáveis candidatas para seleção.

SAÍDAS	ENTRADAS (VARIÁVEIS CANDIDATAS)							
	u1	u2	u3	u4	u5	u6	u7	u8
y_{EB} (topo) e x_{DEB} (base) na coluna 2 (kmol/kmol)	Vazão de topo F_2 (kmol/h)	Carga térmica Q_2 (MW)	Vazão de destilado FD_2 (kmol/h)	Vazão de base FB_2 (kmol/h)	Vazão de refluxo R_2 (kmol/h)	Temperatura estágio 20 T_{20} (K)	Temperatura estágio 21 T_{21} (K)	Temperatura estágio 22 T_{22} (K)

4. RESULTADOS E DISCUSSÃO

Após a utilização das metodologias TRP e *stepwise regression* para seleção de variáveis secundárias em modelos de inferência, foram obtidos os cinco melhores modelos apresentados na Tabela 2. São destacados os dois melhores modelos para as duas variáveis de saída (y_{EB} - topo e x_{DEB} – base), juntamente com os critérios estatísticos de avaliação.

Tabela 2 – Melhores conjuntos de variáveis secundárias.

Modelos y_{EB} -topo	TRP			Regressão por Etapas	
	MQE	Cp	R ² aju	MQE	R ² aju
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₆ u ₇ u ₈]	0,000654	0,9970	0,8906	0,000654	0,8906
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₇ u ₈]	0,000673	0,9967	0,8875	0,000673	0,8875
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₆ u ₈]	0,000678	0,9967	0,8865	0,000678	0,8865
[u ₁ u ₂ u ₃ u ₄ u ₆ u ₇ u ₈]	0,000679	0,9967	0,8864	0,000679	0,8864
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₈]	0,000684	0,9963	0,8851	0,000684	0,8851
Modelos x_{DEB} -base	MQE	Cp	R ² aju	MQE	R ² aju
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₆ u ₇ u ₈]	0,00359	0,9969	0,9491	0,00359	0,9491
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₇ u ₈]	0,00375	0,9966	0,9469	0,00375	0,9469
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₆ u ₈]	0,00389	0,9966	0,9449	0,00389	0,9449
[u ₁ u ₂ u ₃ u ₄ u ₆ u ₇ u ₈]	0,00404	0,9966	0,9428	0,00404	0,9428
[u ₁ u ₂ u ₃ u ₄ u ₅ u ₆ u ₇]	0,00409	0,9966	0,9420	0,00409	0,9420

Os resultados da regressão por etapas confirmaram os de todas as regressões possíveis (TRP), tendo sido obtidos valores semelhantes para os critérios estatísticos de avaliação. Não foi necessária nenhuma etapa de retirada de variáveis para o método *stepwise*. Vale salientar que, quando comparados os dois métodos, recomenda-se o da TRP, haja vista que ele pode encontrar, de maneira global e entre todas as possibilidades, a melhor equação de regressão de acordo com os critérios apontados. Já a metodologia por etapas pode distorcer o procedimento por considerar as dependências entre os regressores. O modelo que contém as oito variáveis atingiu os melhores critérios: menor MQE e maiores Cp e R² ajustado. Além disso, foram selecionados os mesmos regressores para ambos os modelos de inferência de composição, a saber: vazões (alimentação, topo, base e refluxo), além da carga térmica e das temperaturas dos estágios 20, 21 e 22, todas referentes à coluna 2. Logo, verificou-se que essas variáveis exercem influência significativa nas composições de saída dos compostos de interesse.

5. REFERÊNCIAS

- LUYBEN, W. L. Design and Control of the Ethyl Benzene Process. *AIChE Journal*, v. 57, p. 655-670, 2011.
- MORAIS JR, A. A. M. Uso de Sensores Virtuais (Soft Sensors) para Estimativa e Controle de Impurezas em Colunas de Destilação de Alta Pureza. Tese de doutorado, 184 páginas. UFCG, 2015.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. 6th edition. New York: John Wiley & Sons, 2014.
- FORTUNA, L.; GRAZIANI, S.; RIZZO, A.; & XIBILIA, M.G. *Soft sensors for monitoring and control of industrial processes*. London, UK: Springer, 2007.