

ESTUDO DE REGIÕES DE CONFIANÇA DE PARÂMETROS DE ISOTERMAS EM SISTEMA DE ADSORÇÃO DE CORANTE EM CARVÃO ATIVADO

I. S. V. SEGTOVICH¹, A. G. BARRETO JR.²

¹ Universidade Federal do Rio de Janeiro, Escola de Química, Tecnologia de Processos Químicos e Bioquímicos

² Universidade Federal do Rio de Janeiro, Escola de Química, Departamento de Engenharia Química
E-mail para contato: iurisegtovich@poli.ufrj.br

RESUMO – Técnicas de estimação de parâmetros são ferramentas para relacionar estruturas teóricas a observações, a partir da comparação entre previsões e medidas, segundo uma dada métrica. Neste trabalho é estudada a região de confiança de parâmetros de isotermas de adsorção de corante em carvão ativado. São comparadas regiões de confiança geradas para parâmetros estimados por meio de três metodologias: teste t-Student associado a aproximação quadrática da função objetivo em torno do ótimo; teste de Fisher baseado no valor final da função objetivo e associado a procedimento de otimização estocástico; e histograma de parâmetros ótimos obtidos a partir de amostragens de valores pseudo-experimentais, supondo conhecidos os valores reais e distribuição de erros das variáveis. Os resultados mostram que as imprecisões das medidas influenciam tanto o valor ótimo estimado para cada parâmetro, como a região de confiança desses. O efeito da forma da distribuição de erros experimentais também é analisado.

1. INTRODUÇÃO

O conhecimento científico é construído a partir de observações experimentais e da modelagem de sistemas e processos. Nesse contexto, técnicas de estimação de parâmetros constituem ferramentas básicas para o estabelecimento e interpretação da relação entre variáveis de um problema. Estimar parâmetros consiste em inferir valores a grandezas que não podem ser medidas diretamente, essa inferência é realizada a partir da comparação entre medidas experimentais e um modelo de referência utilizando determinada métrica. Toda medida experimental está sujeita a imprecisões e tais imprecisões são relevantes tanto no valor estimado para cada parâmetro quanto na região de confiança de predição do modelo. Para a estimação do valor dos parâmetros é necessária definição de uma métrica que correlacione valores dos parâmetros a graus de adequação do modelo aos dados experimentais. A principal métrica utilizada é a maximização de uma função de verossimilhança, que dadas as hipóteses de erros experimentais independentes e de distribuição normal, se resume a minimização de resíduos quadráticos ponderados.

Modelos lineares podem ser tratados analiticamente com facilidade, enquanto modelos não

lineares requerem a utilização de métodos numéricos variados e algumas aproximações. Para evitar o uso de procedimentos iterativos, alguns modelos podem ser manipulados matematicamente gerando um modelo linear em um novo conjunto de variáveis dependentes e independentes; porém tais transformações modificam as distribuições de erros das variáveis dependentes e independentes, invalidando hipóteses implícitas do procedimento de estimação, o que compromete a interpretação da significância estatística dos resultados. (Osmari et al., 2013, Cassol et al., 2014).

A abordagem mais comum para geração de intervalos de confiança é baseada em teste t-Student, na matriz de covariância dos parâmetros e em um grau de confiança arbitrário. Essa abordagem gera intervalos de confiança necessariamente simétricos para cada parâmetro. A matriz de covariância é calculada por uma aproximação quadrática da função objetivo em torno do ponto ótimo, dessa forma, a abordagem se torna imprecisa para modelos muito não-lineares. Procedimentos de otimização estocásticos possibilitam a obtenção de região de confiança por uma abordagem diferente. Avaliações guiadas por regras heurísticas em uma região de busca pré-definida conduzem ao conjunto de valores ótimos para os parâmetros estimados. A partir dessa, é determinada uma região de valores de parâmetros que resultam em valores da função objetivo inferiores a determinada tolerância, calculada a partir de um teste de Fisher sobre o valor final da função objetivo e um grau de confiança arbitrário. (Schwaab e Pinto, 2007).

2. METODOLOGIA

O presente trabalho pretende avaliar a distribuição “real” de parâmetros estimados utilizando variáveis de saída pseudo-experimentais, obtidas estocasticamente a partir de suas distribuições supostamente conhecidas. Esse tipo de procedimento, que consiste na repetição de uma operação utilizando alguma amostragem aleatória em cada iteração, visando avaliar o conjunto de resultados obtidos após um grande número de simulações, é conhecido como método de Monte Carlo. Na prática, a distribuição, ou seja, a função de densidade de probabilidade de obtenção de valores para dada medição, nunca será conhecida, mas poderá ser inferida em alguns casos a partir de um número elevado de experimentos, ou suposta como igual a distribuição inferida a partir de um grande número de “experimentos brancos” ou calibrações.

A metodologia apresentada será avaliada no contexto da adsorção, um fenômeno de superfície caracterizado pelo aumento da concentração de moléculas de um componente em uma interface, em relação a uma fase fluida, que pode ser estabilizada por interações eletrostáticas, efeitos entrópicos, ou ligações químicas efetivas. Indústrias exploram diversos processos de adsorção na remoção de umidade do ar, remoção de contaminantes de efluentes, separação de fármacos, e separação de misturas de gases. A modelagem de processos de adsorção consiste principalmente na correlação entre a concentração do adsorbato na fase fluida e a quantidade adsorvida em equilíbrio, por massa de adsorvente utilizada.

Foi discutida por Severo Jr. (2011), a discriminação entre variáveis dependentes e variáveis independentes no problema de avaliação de isotermas de adsorção por método de banho finito. São comumente tratadas como dependentes a concentração mássica adsorvida, (Q_{eq}), e como

independentes a concentração de equilíbrio (C_{eq}), porém o que é medido de fato são: como entrada, concentração e volume de solução de adsorbato ($C_0; V_{sol}$), e massa de adsorvente (m); e como saída, a concentração de equilíbrio.

As variáveis medidas estão relacionadas por um balanço de massa no experimento e pela relação de isoterma a ter seus parâmetros estimados. O modelo não linear resultante é dado por:

$$\frac{Q_{\max} k C_{eq}}{1 + k C_{eq}} = \frac{V_{sol} (C_0 - C_{eq})}{m} \quad (1)$$

Em que C_0 , V_{sol} e m são as variáveis de entrada; Q_{\max} e k são os parâmetros a serem estimados, e C_{eq} é a variável de saída.

Neste trabalho, é utilizada uma metodologia semelhante à descrita por Himmelblau (1970) e por Bard (1974). São realizadas simulações do modelo para obter valores “reais” de uma variável de saída a partir do modelo supostamente perfeito. A esses valores são acrescidas flutuações de medição gerando uma distribuição ou variância associada a obtenção experimental dessas variáveis. As variâncias base utilizada para as variáveis de saída foram as inferidas experimentalmente a partir de réplicas no trabalho de Severo Jr. (2011). Conhecidas as distribuições de erros das variáveis medidas, são gerados aleatoriamente um grande número de conjuntos experimentais, a função objetivo definida é minimizada em cada um desses conjuntos. Com isso é obtido um histograma de valores estimados para cada parâmetro, variáveis de saída calculadas, e respectivos valores finais da função objetivo. A função objetivo utilizada nos procedimentos de otimização foi a de mínimos quadrados ponderados pelas variâncias experimentais, mostrada a seguir:

$$f_{obj}(v^{\exp}; v^{\text{modelo}}; \sigma_v^2) = \sum_i^{N_{\exp}} \left\{ \frac{(v_i^{\exp} - v_i^{\text{modelo}})^2}{\sigma_i^2} \right\} \quad (2)$$

Foi utilizado procedimento de otimização de enxame de partículas (*Particle swarm optimization* - PSO), o qual proporcionou avaliações estocásticas para determinação de região de confiança pelo teste de Fisher posteriormente. O valor ótimo determinado para os parâmetros pelo algoritmo de enxame foi utilizado como estimativa inicial para o método de otimização de Gauss-Newton, a partir do qual pode ser obtida a matriz de covariância paramétrica para determinação de intervalos de confiança por teste t-Student.

3. RESULTADOS E DISCUSSÃO

Com base no planejamento experimental e valores estimados para os parâmetros de Severo Jr. (2011), foram gerados valores ideais para as variáveis de saída, considerando o modelo de Langmuir como modelo perfeito. Com esses valores ideais e as variâncias experimentais reportadas no trabalho, foram gerados 10.000 cenários experimentais, e o procedimento de estimação já descrito foi realizado em cada um desses cenários, sendo construído um histograma com 20 blocos englobando todos os

valores obtidos.

A Tabela 1 apresenta os valores de referência para as amostragens e estimações realizadas. As identificações plan(1), plan(2) e plan(3) representam o planejamento 1, resultante do trabalho sequencial de Severo Jr. (2011), e planejamentos teóricos alternativos 2 e 3, cobrindo menos eficientes a região relevante das variáveis experimentais. Todos os planejamentos constam de 5 condições experimentais diferentes. O planejamento 2 se concentra na região de comportamento linear do modelo, enquanto o planejamento 3 se concentra na região de saturação do modelo, apenas o planejamento 1 amostra toda a região relevante do modelo.

Tabela 1 – Valores de referência para amostragens e estimações realizadas.

Plan(*)	V_{sol} [L]	C_0 [g·L ⁻¹]	m [g]	C_{eq}^{modelo} [g·L ⁻¹]	σ_i^2 [g ² ·L ⁻²]
1; 2	0,05	0,5	0,2509	0,0351	$9,214 \cdot 10^{-8}$
1; 2	0,05	0,5	0,3017	0,0274	$4,683 \cdot 10^{-8}$
1; 2	0,05	0,5	0,3557	0,0221	$7,047 \cdot 10^{-7}$
1; 3	0,05	5,0	0,2517	3,7044	$1,785 \cdot 10^{-5}$
1; 3	0,05	5,0	0,7004	1,4842	$2,516 \cdot 10^{-6}$
2	0,05	1,0	0,2509	0,1270	$7,047 \cdot 10^{-7}$
2	0,05	1,0	0,3017	0,0877	$7,047 \cdot 10^{-7}$
3	0,05	2,0	0,2509	0,7853	$2,516 \cdot 10^{-6}$
3	0,05	5,0	0,2017	3,4490	$1,785 \cdot 10^{-5}$
3	0,05	5,0	0,3557	3,1742	$1,785 \cdot 10^{-5}$

As concentrações equilíbrio apresentadas são soluções exatas do modelo completo, obtidas a partir dos valores reais estipulados para os parâmetros: Q_{max} [g/g] = $2.618 \cdot 10^{-1}$; k [L/g] = $1.559 \cdot 10^1$.

Foram obtidas regiões de confiança segundo o teste de Fisher sobre o valor final da função objetivo utilizando os pontos gerados pelo enxame para quatro conjuntos constituídos por 5 e 15 (3 réplicas) pseudo experimentos.

A Figura 1 apresenta os histogramas de valores ótimos para os parâmetros estimados a partir de 10.000 conjuntos pseudo-experimentais seguindo o planejamento 1. Foi possível observar que dada a pequena variância experimental inserida na geração, e qualidade de amostragem do planejamento 1, os parâmetros estimados apresentaram distribuição próxima a normal, e apresentaram baixa correlação tanto para o cenário com 5 experimentos (-0.560) quanto para o cenário com 15 experimentos (-0.562). As identificações par(1) e par(2) representam os parâmetros da isoterma de Langmuir Q_{max} e k respectivamente.

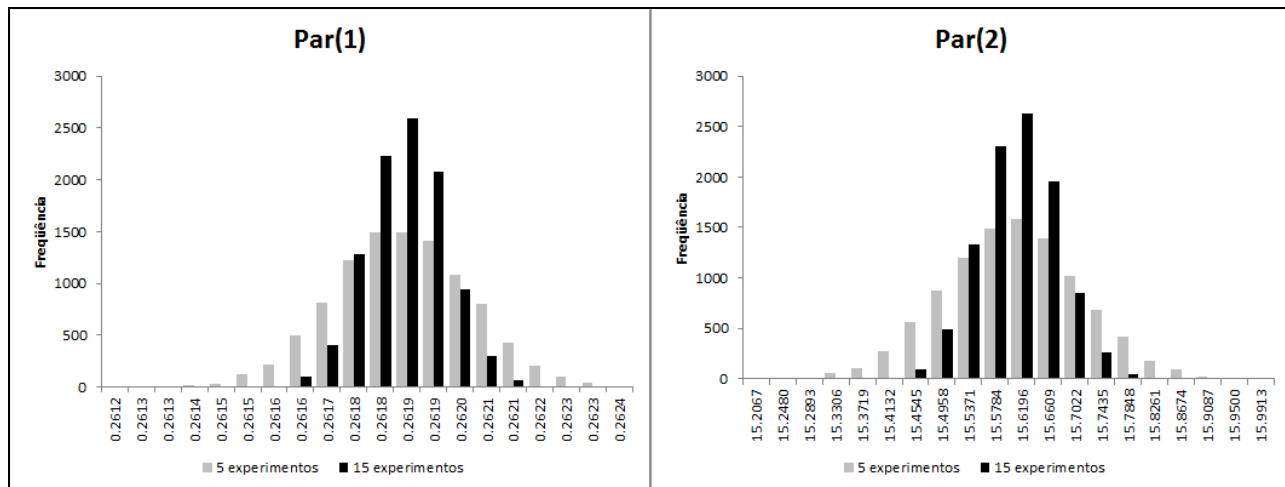


Figura 1 – Histograma de valores estimados para o planejamento 1

É possível observar que com o maior número de experimentos os parâmetros são estimados com maior precisão, o que se reflete na obtenção de um histograma mais estreito, a região de confiança dos parâmetros estimados foi reduzida mediante a disponibilidade de réplicas, devido a contribuição do maior número de experimentos ao procedimento de estimação.

A figura 2 apresenta regiões de confiança determinadas pelo teste de Fisher e avaliações estocásticas realizadas pelo algoritmo de enxame de partículas para 4 gerações aleatórias de conjuntos de pseudo-experimentos, considerando 1 experimento em cada uma das 5 condições experimentais, ou três réplicas em cada uma das 5 condições experimentais.

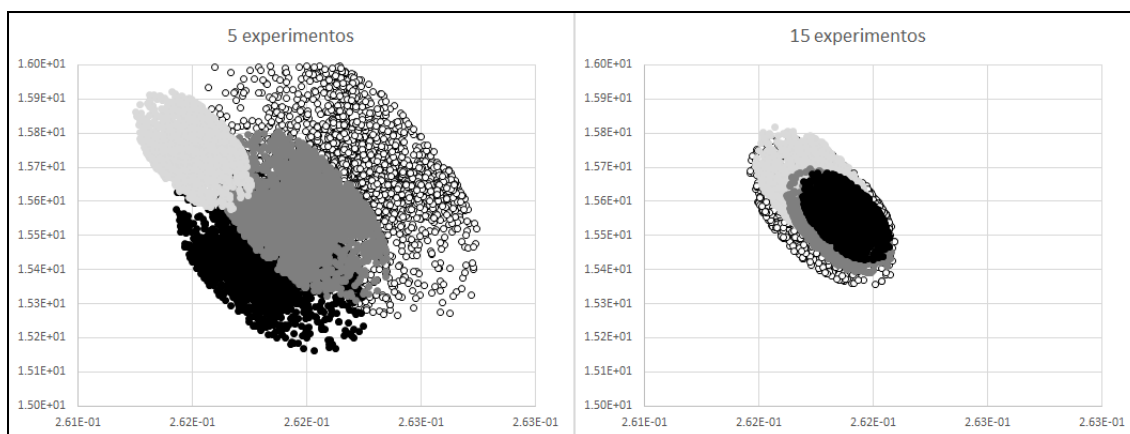


Figura 2 – Nuvem de confiança segundo teste de Fisher para o planejamento 1.

É possível observar que no cenário com 15 experimentos as regiões de confiança são menores e que regiões de confiança de diferentes conjuntos experimentais coincidem.

Com o intuito de observar resultados menos bem comportados, foram conduzidas novas rodadas

de geração e estimação seguindo os planejamentos 2 e 3 e utilizando variâncias 100 vezes maiores que as variâncias de referência da tabela. A Figura 3 apresenta histogramas gerados para cada parâmetro nos dois planejamentos alternativos.

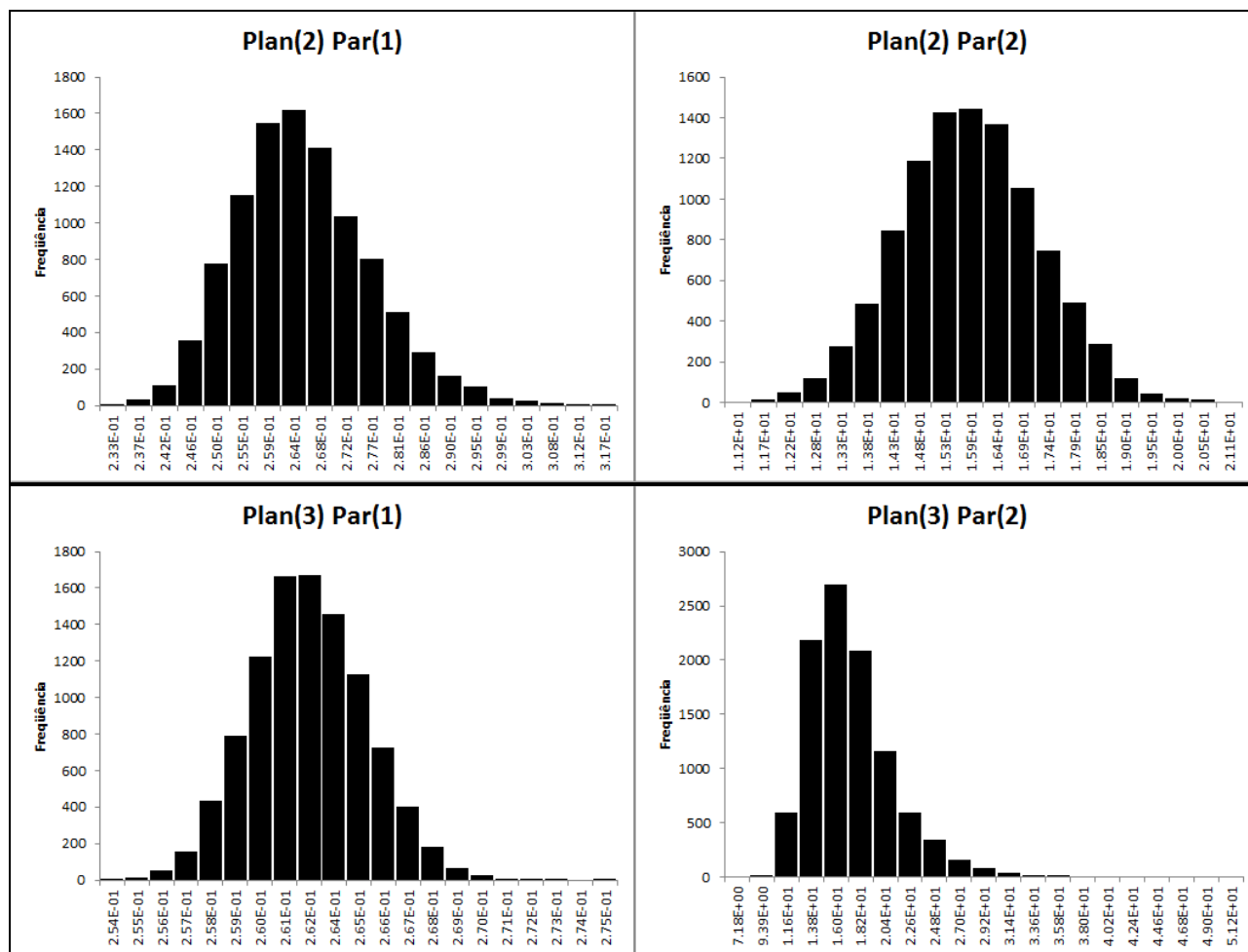


Figura 3 – Histograma de valores estimados para os planejamentos 2 e 3

Foram obtidos histogramas mais largos e apresentando distribuição assimétrica, principalmente para o parâmetro ao qual o modelo é menos sensível na região experimentada. No planejamento 2, que cobriu melhor a região linear, o parâmetro de saturação apresenta histograma mais largo, enquanto no planejamento 3, que cobriu melhor a região de saturação, o parâmetro de constante cinética apresentou histograma mais largo. Em ambos os casos foi observada alta correlação entre os parâmetros (-0.942 no plan(2) e -0.908 no plan(3)).

As Figuras 4 (a) e 4 (d) apresentam as regiões de confiança determinadas pelo teste de Fisher com avaliações estocásticas realizadas pelo algoritmo de enxame de partículas (identificada como Nuvem Fisher) para dois conjuntos de pseudo-experimentos gerados aleatoriamente, um conjunto para cada um dos planejamentos alternativos, em contraste com a nuvem de pares de parâmetros

ótimos calculados no procedimento de Monte Carlo (identificada como Nuvem Monte Carlo). A nuvem gerada pelo método de Monte Carlo no gráfico foi feita excluindo-se os 5% parâmetros estimados que geraram maiores desvios das variáveis previstas em dado conjunto de experimentos em relação aos valores médios dessas variáveis, ponderados pelas respectivas variâncias. Junto aos gráficos de nuvens são apresentadas as barras de erro para o intervalo de confiança pelo teste de t-Student. As Figuras 4 (b), 4 (c), apresentam os intervalos de confiança gerados para os parâmetros pelas três metodologias estudadas, utilizando o planejamento 2; e as Figuras 4 (e), 4 (f) utilizando o planejamento 3.

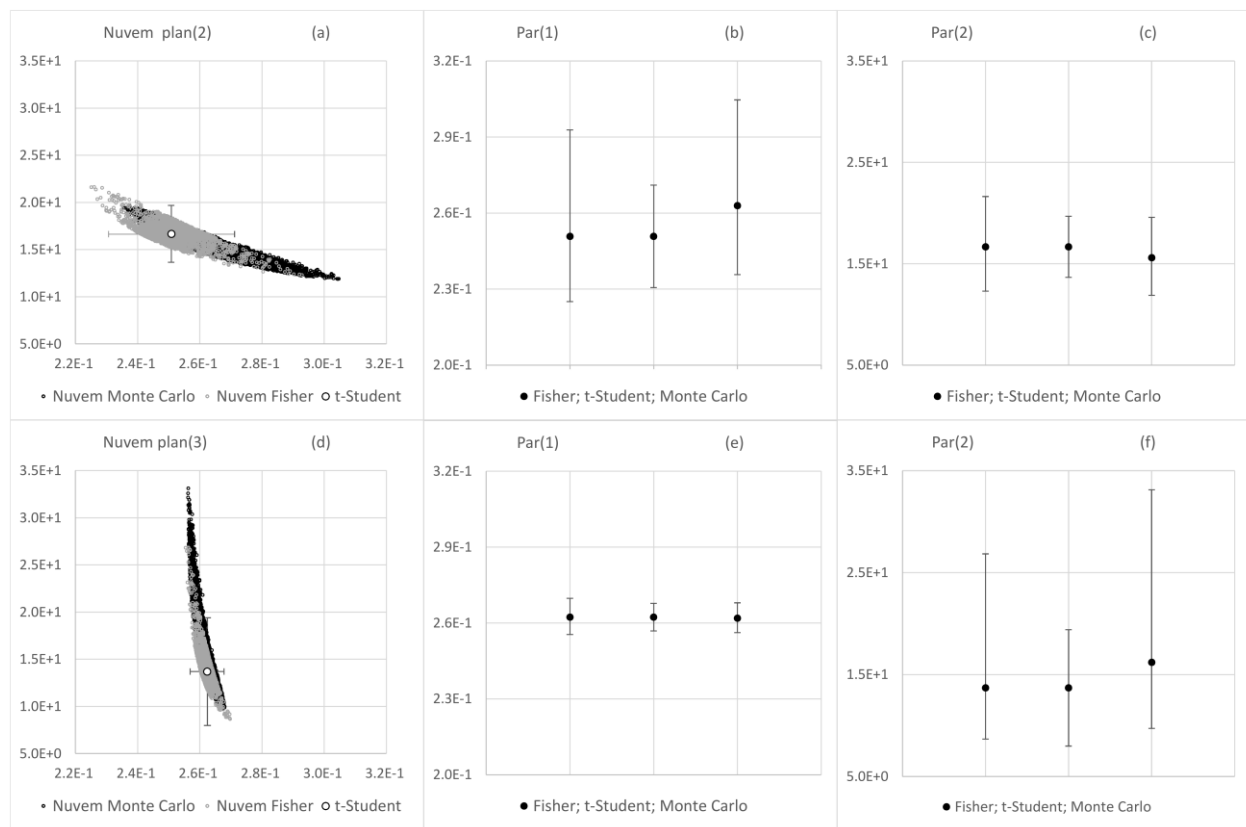


Figura 4 – Nuvens e intervalos de confiança determinados para os planejamentos 2 e 3

A nuvem determinada pelo método de Enxame/Fisher equivale a região de confiança determinada a partir de um conjunto de 15 experimentos (5 condições experimentais e 3 réplicas em cada condição), enquanto a nuvem de Monte Carlo representa os 95% pares de parâmetros ótimos gerando menores que melhor preveem as variáveis de saída, que poderiam ser encontrados por experimentadores dispondo de um conjunto de 15 experimentos obtidos a partir da distribuição suposta.

É possível observar sobreposição entre as nuvens determinadas pelos métodos de Enxame/Fisher e de Monte Carlo; porém a nuvem gerada pelo método de Monte Carlo está um pouco deslocada em relação a determinada a partir das amostras do enxame de partículas, pois essas foram

obtidas a partir de um único conjunto de 15 experimentos gerados aleatoriamente, logo é uma região em torno de um par de parâmetros ótimos para esse conjunto de experimentos, enquanto a nuvem do método de Monte Carlo está em torno do valor real dos parâmetros, utilizados na geração dos pseudo-experimentos. A boa sobreposição, então, indica adequação da metodologia de Enxame/Fisher para predição de região de confiança paramétrica em um modelo não linear.

Por outro lado, o intervalo de confiança pelo teste t-Student, subestima a região de confiança para os parâmetros, analisando globalmente, como mostram as Figuras 4 (b), 4 (c), 4 (e) e 4 (f), e ao mesmo tempo, analisando localmente, o intervalo de confiança para um parâmetro, considerando o outro fixo em seu valor ótimo, aparece superestimado. Isso ocorre pois o teste t-Student é feito considerando apenas as variâncias de cada parâmetro, e não a matriz de covariâncias paramétricas completa. É importante levar em consideração a correlação entre os parâmetros, evidenciada pelo formato das nuvens mostradas nas Figuras 4 (a) e 4 (d) e não tomar os intervalos de confiança de cada parâmetro mostrados nas Figuras 4 (b), 4 (c), 4 (e) e 4 (f) como independentes.

4. CONCLUSÃO

Foi possível comparar regiões de confiança obtidas a partir de testes de Fisher e t-Student utilizando histogramas de estimações de amostragens estocásticas. As regiões de confiança das três metodologias tendem a coincidir para cenários com variância experimental baixa e planejamento cobrindo as regiões de diferentes comportamentos do modelo, e conforme se aumenta o número de experimentos. Em cenários com planejamentos menos adequados e variâncias maiores são observado histogramas assimétricos, aumento da correlação, e nuvens de confiança não elípticas para os parâmetros.

5. REFERÊNCIAS

- BARD, Non-linear parameter estimation, Academic Press: New York and London, 1974.
- CASSOL, O. G., GALLON, R., SCHWAAB, M., BARBOSA-COUTINHO, E., SEVERO JR., J. B., PINTO, J. C., Statistical Evaluation of Nonlinear Parameter Estimation Procedures for Adsorption Equilibrium Models, Adsorption Science & Technology, aceito para publicação, 2014.
- HIMMELBLAU, Process analysis by statistical methods, New York: John Wiley & Sons, Inc., 1970.
- OSMARI, T. A., GALLON, R., SCHWAAB, M., BARBOSA-COUTINHO, E., SEVERO JR., J. B., PINTO, J. C., Statistical analysis of linear and non-linear regression for the estimation of adsorption isotherm parameters, Adsorption Science & Technology, v. 31, n. 5, p. 433-458, 2013.
- SCHWAAB, M., PINTO, J. C., Análise de Dados Experimentais I: Fundamentos de Estatística e Estimação de Parâmetros, Rio de Janeiro: e papers, 2007.
- SEVERO JR., J. B., Avaliação de técnicas de planejamento de experimentos no reconhecimento do equilíbrio de adsorção em sistemas cromatográficos, Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro, 2011.