

Estimativa do intervalo de confiança em calibração multivariada utilizando PLS por *boosting ensemble*

Paulo R. Filgueiras*¹ (PG), Luciana A. Terra¹ (PG), Eustáquio V. R. de Castro² (PQ), Lize M. S. L. de Oliveira³ (PQ), Júlio C. M. Dias³ (PQ), Ronei J. Poppi¹ (PQ).

¹Instituto de Química, Universidade Estadual de Campinas, Caixa Postal 6154, Campinas/SP, CEP 13083-970;

²Departamento de Química, Universidade Federal do Espírito Santo, ³CENPES/PETROBRAS. pauloiuna@hotmail.com

Palavras Chave: *Boosting, Ensemble, PLS, Intervalo de confiança.*

Introdução

O intervalo de confiança para o método de calibração multivariada por mínimos quadrados parciais (PLS) depende diretamente do erro padrão de calibração (SEC) e da raiz da distância da amostra ao centro do conjunto de calibração (*leverage*).¹ Portanto, o intervalo de confiança é fortemente afetado por amostras de calibração com resultados discrepantes. Uma alternativa a este problema é utilizar um *ensemble*: procedimento estatístico que combina múltiplos modelos individuais, construídos a partir de reamostragens, para gerar uma única resposta. Este procedimento aliado à metodologia *boosting*, pondera as amostras de calibração durante a reamostragem para construção do modelo final.²

Este trabalho estima o intervalo de confiança para modelos de calibração utilizando PLS por *boosting ensemble* aplicado na determinação das temperaturas de massa recuperada a 10% (T10), 50% (T50) e 90% (T90) de petróleos por espectroscopia ¹H RMN.

Resultados e Discussão

Foram medidas as temperaturas de massa recuperada a 10% (T10), 50% (T50) e 90% (T90) por destilação simulada³ e tomou-se espectros de ¹H RMN (Instrumento Varian MR-400 com frequência de 399,8 MHz e padrão de CDCl₃), para 35 amostras de petróleo bruto com gravidade API variando de 17,0 a 54,0. As amostras foram divididas em 24 para calibração e 11 para validação pelo algoritmo Kennard-Stone.⁴

O modelo ensemble PLS (ePLS) foi construído utilizando 100 reamostragens do conjunto de calibração para construção dos regressores individuais. A estimativa de previsão e do intervalo de confiança da temperatura de perda de massa foi determinada respectivamente pelo valor médio dos resultados de previsão dos regressores e sua distribuição com 95% de confiança.

Os resultados mostram uma boa concordância entre os valores estimados pelos modelos PLS e ePLS com os valores de referência (Figura 1). Uma amostra de calibração (propriedade T10) apresentou resultado discrepante, porém foi mantida para modelagem, resultando num aumento do intervalo

de confiança para o modelo PLS (Figura 1a). Os modelos PLS e ePLS apresentam erros de previsão respectivamente de: 15,6 °C e 15,1 °C para T10; 24,2 °C e 23,4 °C para T50 e 39,0 °C e 39,9 °C para T90, não apresentando diferença estatisticamente significativa pelo teste F de Fisher-Snedecor.

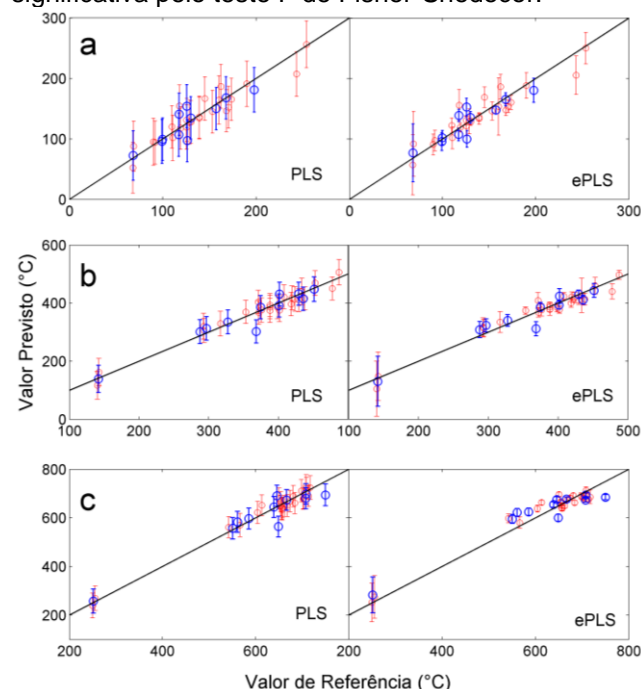


Figura 1. Gráficos da previsão das temperaturas de perda de massa, para as propriedades T10 (a), T50 (b) e T90 (c) em petróleos, em função dos valores de referência. As barras verticais representam os intervalos de confiança de 95%. Em vermelho as amostras de calibração e azul as amostras de validação.

Conclusões

Os intervalos de confiança obtidos pela metodologia *boosting ensemble* apresentam bons resultados e são menos influenciados por amostras anômalas presentes no conjunto de calibração.

Agradecimentos

LABPETRO-UFES, PETROBRAS e CNPq.

¹ ASTM E 1655. West Conshohocken, PA: ASTM International; 2005.

² Cao, D. S.; Xu, Q. S.; Liang, Y. Z.; Zhang, L. X.; Li, H. D. *Chemom. Intell. Lab. Syst.* **2010**, *100*, 1.

³ ASTM D 7169. West Conshohocken, PA: ASTM International; 2005.

⁴ Kennard, R. W.; Stone, L. A. *Tecnometrics*. **1969**, *11*, 137.