

# UM NOVO CRITÉRIO PARA SELEÇÃO DE VARIÁVEIS EMPREGANDO O ALGORITMO DAS PROJEÇÕES SUCESSIVAS

Sófacles Figueredo Carreiro Soares<sup>2\*</sup> (PG), Roberto Kawakami Harrop Galvão<sup>1</sup> (PQ), Mário Cesar Ugulino de Araújo<sup>2</sup> (PQ). [sofacles@gmail.com](mailto:sofacles@gmail.com).

<sup>1</sup> Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Eletrônica, São José dos Campos, SP

<sup>2</sup> Universidade Federal da Paraíba, CCEN, Departamento de Química – João Pessoa PB.

Palavras Chave: regressão linear múltipla, seleção de variáveis, algoritmo das projeções sucessivas, erro estatístico de previsão, espectrometria NIR.

## Introdução

Este trabalho propõe uma modificação no Algoritmo das Projeções Sucessivas (APS)<sup>1</sup>, com objetivo de aumentar a robustez dos modelos de Regressão Linear Múltipla (RLM) construídos. Na formulação original do APS, subconjuntos de variáveis são comparados entre si com base no erro quadrático médio obtido em um conjunto de validação. De acordo com o critério aqui proposto, a comparação é feita também com base no erro estatístico de previsão. Tal métrica leva em conta a *leverage* associada à amostra, bem como o resíduo da regressão.

Dois estudos de caso envolvendo análises de diesel e milho por espectrometria NIR são discutidos. Os resultados são avaliados em termos da raiz do erro quadrático médio em um conjunto de previsão independente (RMSEP).

Estudo de Caso 1: Óleo Diesel

- Espectros NIR de 170 amostras, na faixa de 880-1675 nm, obtidos em um espectrômetro FT-NIR Perkin Elmer GX com resolução de 2 cm<sup>-1</sup>. Os valores de referência para teor de enxofre, T10% e T90% foram determinados de acordo com as normas ASTM. Foram separadas 70 amostras para calibração, 50 para validação e 50 para previsão, usando o algoritmo SPXY<sup>2</sup>.

Estudo de Caso 2: Milho

- Espectros NIR de 80 amostras obtidos em dois aparelhos, denominados m5 e mp5, na faixa de 1100-2498 nm. Os dados incluem os valores de referência para umidade, óleo, proteína e amido. Foram separadas 40 amostras para calibração, 20 para validação e 20 para previsão nas medidas de ambos aparelhos, usando o algoritmo SPXY<sup>2</sup>.

## Resultados e Discussão

No primeiro estudo de caso foi adicionado ruído aos espectros das amostras de previsão e em seguida, foram previstos os parâmetros de enxofre, T10% e T90%.

O segundo estudo de caso foi realizado com base nas amostras medidas em dois aparelhos, o mp5 e o m5. Os modelos obtidos utilizando o aparelho mp5 foram usados para prever os parâmetros

32ª Reunião Anual da Sociedade Brasileira de Química

umidade, óleo, proteína e amido, usando os espectros obtidos a partir de medidas realizadas no aparelho m5. Os resultados obtidos com os critérios original e proposto para o APS em ambos estudos de caso encontram-se na **tabela 1**.

**Tabela 1.** RMSEP para os diversos parâmetros estudados.

Parâmetro	Critério original	Critério proposto
Enxofre (%)	18.4	10.7
T10% (°C)	8.9	6.9
T90% (°C)	7.8	6.7
Umidade (%)	1.684	0.884
Óleo (%)	0.388	0.332
Proteína (%)	0.615	0.498
Amido (%)	0.697	0.761

Ao observar-se os resultados apresentados na **tabela 1**, percebe-se que de um modo geral os modelos obtidos depois da modificação apresentaram um menor RMSEP. Tal tendência não foi seguida em apenas um caso (determinação de amido).

## Conclusões

Os resultados obtidos demonstraram que o novo critério proposto para uso com o APS foi capaz de melhorar de maneira substancial a robustez dos modelos RLM obtidos, no que diz respeito a ruído espectral e diferenças instrumentais.

A modificação proposta pode ser considerada como um melhoramento útil para a formulação básica do APS.

## Agradecimentos

CAPES, CNPq.

<sup>1</sup> Araújo, M. C. U.; Saldanha, T. C. B.; Galvão, R. K. H.; Yoneyama, T.; Chame, H. C. e Visani, V. *Chemom. Intell. Lab. Syst* **2001**, 57, 65.

<sup>2</sup> Galvão, R. K. H.; Araújo, M. C. U.; José, G. E.; Pontes, M. J. C.; Silva, E. C. e Saldanha, T. C. B. *Talanta*. **2005**, 67, 736.