



PREDIÇÃO DO TEOR DE PROTEÍNAS EM BEBIDAS ENERGÉTICAS POR ESPECTROMETRIA DE MASSAS DE ALTA RESOLUÇÃO E PLSR

Giovana S. Pinheiro^{1,2}; José M. S. Netto¹; Camila Assis¹.

¹British American Tobacco (BAT), Cachoeirinha, RS, Brasil, 94930-000

²Programa de Pós-Graduação em Química, Instituto de Química, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brasil, 91540-000
giovana_pinheiro@bat.com

Palavras-Chave: quimiometria, bebidas energéticas, espectrometria de massas.

Introdução

A caracterização da composição de alimentos e bebidas requer métodos analíticos capazes de fornecer informações confiáveis sobre diferentes parâmetros de interesse. Nesse contexto, ferramentas quimiométricas têm sido amplamente empregadas para explorar dados instrumentais complexos, permitindo estabelecer relações entre respostas analíticas e propriedades químicas de interesse. A regressão por mínimos quadrados parciais (PLS) destaca-se entre as técnicas de calibração multivariada por possibilitar a construção de modelos quantitativos a partir de conjuntos de dados de elevada dimensionalidade e colinearidade [1,2].

A espectrometria de massas de alta resolução (HRMS) apresenta elevado potencial para a caracterização de matrizes complexas devido à sua capacidade de fornecer informações composicionais abrangentes. Quando associada a ferramentas quimiométricas, essa técnica permite explorar fingerprints químicos e extrair informações relacionadas a propriedades de interesse que podem não ser diretamente acessíveis por abordagens univariadas [3,4].

A determinação do teor de proteínas por métodos baseados em Kjeldahl constitui uma abordagem de referência para a quantificação de proteína total em diferentes matrizes alimentares [5]. Entretanto, a utilização dos dados multivariados provenientes de técnicas instrumentais de alta resolução pode ampliar as possibilidades de avaliação da composição de matrizes complexas. Dessa forma, este trabalho teve como objetivo avaliar o potencial da espectrometria de massas de alta resolução (HRMS), combinada à quimiometria, para a quantificação do teor de proteínas em bebidas energéticas, utilizando os dados espectrais obtidos por HRMS como preditores e os resultados determinados pelo método de Kjeldahl como referência.



Material e Métodos

Foram avaliadas 18 amostras únicas de bebidas energéticas, adquiridas em triplicata, totalizando 54 espectros. Inicialmente, o teor de proteínas foi determinado por um método de referência baseado em Kjeldahl, sendo utilizado o valor médio das determinações para representar cada amostra. Os resultados obtidos foram comparados aos valores declarados nas respectivas tabelas nutricionais. Para fins de modelagem, os resultados inferiores ao limite de quantificação do método de referência foram atribuídos ao valor zero, uma vez que não apresentavam confiabilidade metrológica suficiente para quantificação.

Após a determinação de referência, as amostras foram preparadas e analisadas em espectrômetro de massas de alta resolução, sendo os espectros obtidos utilizados como matriz de preditores (X). Os valores de teor de proteínas determinados pelo método de referência constituíram a variável resposta (Y). Para construção do modelo PLS, os dados de X foram submetidos a escalonamento Pareto e centralização na média, enquanto Y foi submetido à centralização na média.

A separação dos conjuntos de calibração e validação externa foi realizada ao nível de amostra, mantendo-se todas as réplicas espectrais de uma mesma amostra integralmente no mesmo conjunto. Dessa forma, evitou-se o compartilhamento de informações entre calibração e validação, assegurando a independência da avaliação preditiva. A seleção das amostras para cada conjunto foi realizada pelo algoritmo Kennard-Stone considerando simultaneamente as matrizes X e Y (KSXY), na proporção de 80:20, com o objetivo de assegurar que ambos os conjuntos apresentassem distribuição representativa da variabilidade espectral e da faixa de valores de referência. A validação cruzada do conjunto de calibração foi conduzida pelo método Venetian Blinds. O desempenho do modelo foi avaliado por meio dos parâmetros RMSECV, RMSEP, coeficientes de determinação (R^2) e Bias. A seleção de variáveis foi realizada com base nos VIP Scores (Variable Importance in Projection), considerando-se as variáveis com $VIP > 1$. A significância estatística do modelo foi avaliada por meio de teste de permutação, adotando-se nível de confiança de 95%.

Resultados e Discussão

A comparação entre os resultados experimentais e os valores declarados nas tabelas nutricionais evidenciou diferenças entre a rotulagem e os teores determinados pelo método de referência. Nas 18 amostras, os valores experimentais variaram de <LQ a 6,31 g/100 mL, enquanto os declarados variaram de 0 a 6,45 g/100 mL. Algumas amostras sem declaração de proteínas apresentaram valores quantificáveis (0,44 a 1,04 g/100 mL), enquanto quatro apresentaram resultados <LQ, reforçando a importância da caracterização analítica da matriz.

O modelo PLS baseado nos dados de HRMS apresentou 5 variáveis latentes (LVs). Na validação cruzada pelo método Venetian Blinds, foram obtidos RMSECV de 0,6483, R^2 CV de 0,9143 e Bias CV de 0,0701. Na validação externa, o modelo apresentou RMSEP de



1,1981, R^2 Pred de 0,9430 e BiasPred de -0,3615. As cinco variáveis latentes explicaram, cumulativamente, 76,57% da variância em X e 98,63% da variância em Y. A análise dos VIP Scores permitiu identificar as variáveis mais relevantes para a construção do modelo.

O teste de permutação indicou significância estatística do modelo, com *p-values* inferiores a 0,05 nos testes de Wilcoxon, Sign Test e Rand t-test, tanto para a auto-predição quanto para a validação cruzada. Esses resultados indicam que a relação entre os dados de HRMS e o teor de proteínas é estatisticamente significativa e demonstra o potencial dos dados multivariados para a predição desse parâmetro.

Conclusões

Os resultados demonstraram o potencial da associação entre espectrometria de massas de alta resolução (HRMS) e regressão PLS para a estimativa do teor de proteínas em bebidas energéticas. A abordagem apresentou desempenho preditivo satisfatório nas etapas de validação cruzada e validação externa, além de significância estatística do modelo, indicando que os dados de HRMS apresentam informação relevante para a modelagem desse parâmetro.

Considerando o número limitado de amostras avaliadas, os resultados apresentam caráter exploratório e indicam a necessidade de aumentar a diversidade amostral para ampliar a robustez e a capacidade de generalização do modelo. Como futuras etapas, propõe-se a inclusão de amostras com maior variabilidade composicional, bem como a adição de proteínas de diferentes origens por meio de fortificação. Essas estratégias permitirão expandir a faixa de variabilidade do conjunto de calibração e contribuir para uma avaliação mais abrangente da aplicabilidade do modelo.

Referências

1. ARANCIBIA, J. A.; DAMIANI, P. C.; IBÁÑEZ, G. A.; OLIVIERI, A. C. Recent applications of first- and second-order multivariate calibration to analytical chemistry. **Journal of AOAC International**, 97(1), 39–49, 2014.
2. VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model: a case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, 55(21), 8331–8338, 2007.
3. DÍAZ-GALIANO, F. J.; MURCIA-MORALES, M.; RODRÍGUEZ FERNÁNDEZ-ALBA, A. From sound check to encore: A journey through high-resolution mass spectrometry-based food analyses and metabolomics. **Comprehensive Reviews in Food Science and Food Safety**, 23(2), e13325, 2024.
4. KAUFMANN, A. The current role of high-resolution mass spectrometry in food analysis. **Analytical and Bioanalytical Chemistry**, 403(5), 1233–1249, 2012.
5. TOMÉ, Daniel; CORDELLA, Christophe; DIB, Omar; PÉRON, Christine. Nitrogen and protein content measurement and nitrogen to protein conversion factors for dairy and soy protein-based foods: a systematic review and modelling analysis. Geneva: World Health Organization; **Food and Agriculture Organization of the United Nations**, 2019. Disponível em: <https://iris.who.int/bitstream/handle/10665/331206/9789241516983-eng.pdf>. Acesso em: 20 ago. 2026.