PIPELINE QUIMIOMÉTRICO PARA AVALIAÇÃO DE PRÉ-PROCESSAMENTO EM ESPECTROS ATR-FTIR DE PETRÓLEOS

Lucimara Ribeiro Venial, Flaviane Mendonça Ambrozim, Ester Ramos Pancieri, Poliany de Oliveira Barbosa, Bárbara Lima Sartorio, Maristela de Araújo Vicente, Maria de Fátima Pereira dos Santos.

Universidade Federal do Espírito Santo (UFES), Programa de Pós-Graduação em Energia, CEUNES, São Mateus, ES, Brasil. E-mail do autor remetente: venialluci@gmail.com.

Resumo
A espectroscopia no infravermelho com transformada de Fourier por refletância total atenuada (ATR-FTIR) associada à quimiometria é relevante para caracterizar matrizes complexas, mas variações de linha de base, escala e ruído tornam o pré-processamento uma etapa crítica. Neste trabalho, com o objetivo de reduzir a subjetividade na escolha do tratamento espectral, foi estruturado um pipeline quimiométrico para avaliar estratégias de tratamento de espectros ATR-FTIR de petróleos, integrando correção de linha de base, suavização, derivação, normalização e PCA. A comparação considerou critérios complementares de variância, estabilidade da projeção, preservação espectral e interpretabilidade química. As rotas avaliadas produziram respostas multivariadas distintas, mostrando que o pré-processamento interfere na organização dos dados. A abordagem permitiu selecionar o tratamento de forma sistemática e reprodutível, com potencial aplicação em rotinas de controle de qualidade de petróleos brutos.
Palavras-chave: ATR-FTIR. Quimiometria. Pré-processamento espectral. PCA. Petróleo.

Área do Conhecimento: Ciências Exatas e da Terra - Química.

Introdução

Petróleos brutos constituem matrizes químicas de elevada complexidade, compostas por hidrocarbonetos saturados e aromáticos, resinas, asfaltenos. A espectroscopia ATR-FTIR permite registrar informações relacionadas a grupos funcionais e ao perfil molecular dessas amostras de maneira rápida e sem preparo prévio, gerando conjuntos de dados adequados à exploração quimiométrica (Mohammadi et al., 2020). A análise de componentes principais, PCA, é particularmente útil nesse contexto por reduzir a dimensionalidade do conjunto espectral e evidenciar fontes dominantes de variação entre amostras (Wold; Esbensen; Geladi, 1987).

A qualidade da interpretação multivariada, entretanto, depende da etapa de pré-processamento. Deslocamentos de linha de base, diferenças de escala e ruído instrumental podem introduzir variabilidade não associada à composição química. Técnicas como correção de linha de base, suavização de Savitzky-Golay, derivação e transformação Standard Normal Variate, SNV, são amplamente empregadas para minimizar esses efeitos, mas sua aplicação deve ser avaliada de forma criteriosa, pois tratamentos excessivos também podem atenuar ou distorcer informações espectrais relevantes (Savitzky; Golay, 1964; Barnes; Dhanoa; Lister, 1989; Rinnan; Van den Berg; Engelsen, 2009; Butler et al., 2018).

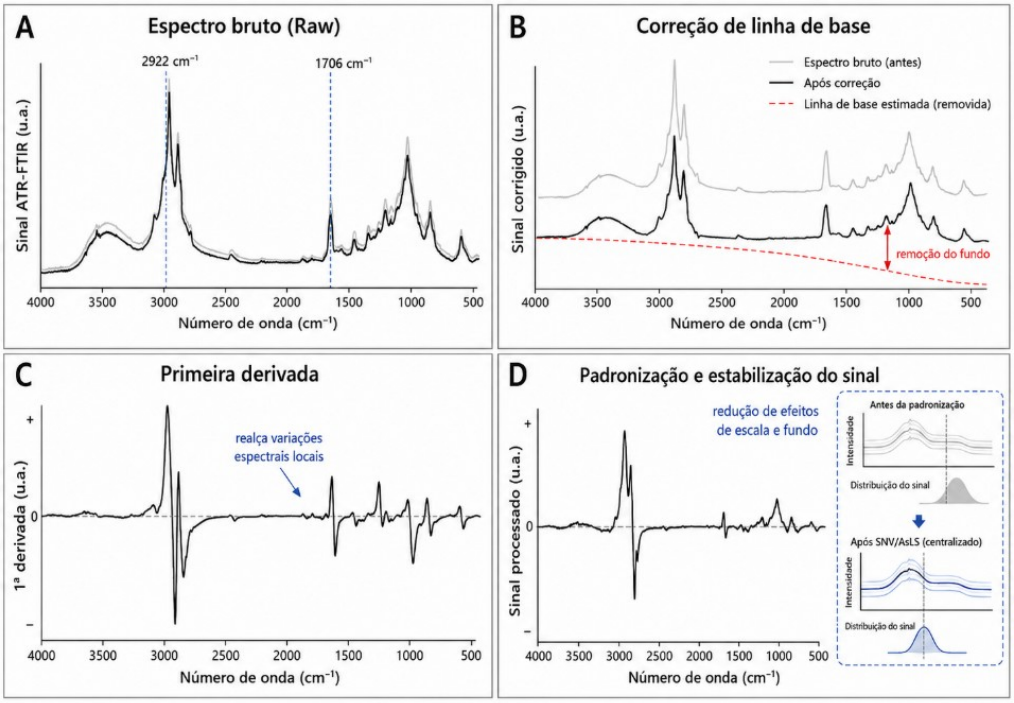
Nesse cenário, a adoção de um pipeline computacional capaz de comparar sistematicamente diferentes estratégias de pré-processamento amplia a rastreabilidade da análise e reduz a subjetividade na escolha do tratamento espectral. Assim, o objetivo deste trabalho foi desenvolver um pipeline quimiométrico para avaliação de estratégias de pré-processamento em espectros ATR-FTIR de petróleos.

Metodologia

O conjunto de dados foi constituído por espectros ATR-FTIR de amostras de petróleo bruto obtidos em triplicata. A aquisição espectral foi realizada no intervalo de 4000 a 650 cm^{-1}, com resolução de 4 cm^{-1} e 64 varreduras por espectro. Os dados foram organizados em grade espectral regular e submetidos a rotinas computacionais. O pipeline foi estruturado em módulos sucessivos de leitura, padronização da matriz, pré-

processamento, modelagem por PCA, avaliação de estabilidade e integração dos critérios de decisão (Figura 1).

Figura 1 - Representação didática do efeito de etapas de pré-processamento em espectros ATR-FTIR. O painel A mostra o espectro bruto. O painel B exemplifica a correção de linha de base. O painel C ilustra a primeira derivada. O painel D sintetiza a padronização e a estabilização do sinal, destacando a redução de efeitos de escala e fundo antes da modelagem multivariada.



Fonte: Elaborado pelo autor (2026).

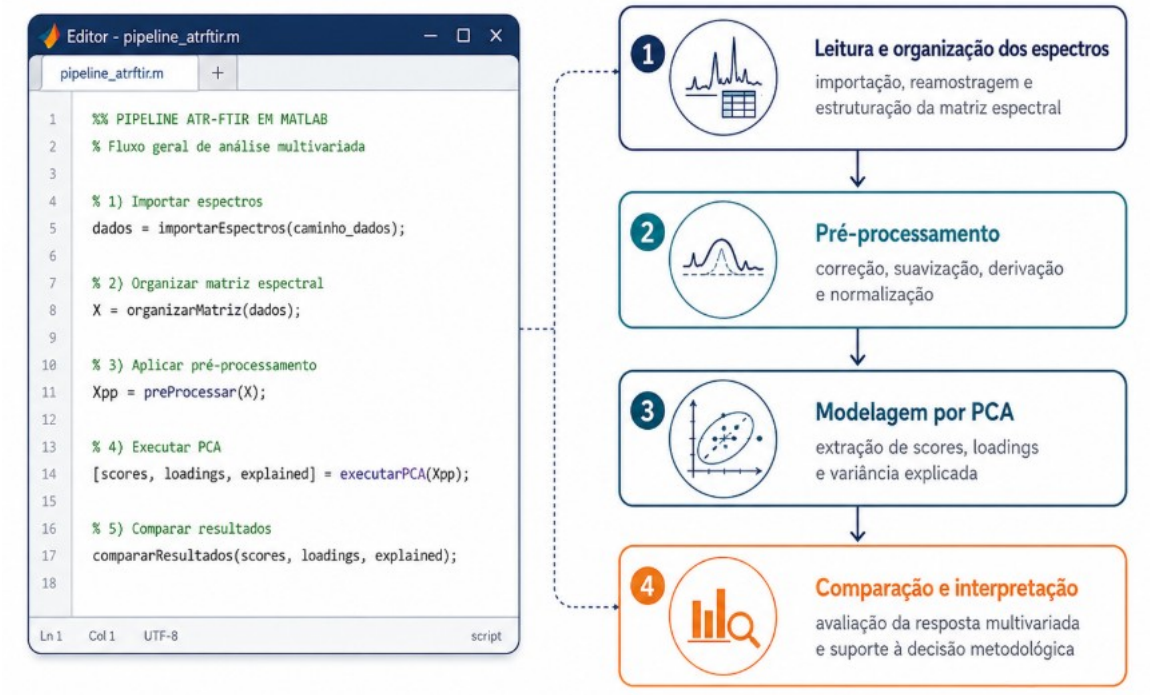
Foram comparadas rotas envolvendo dados de referência, correção polinomial de linha de base, suavização de Savitzky-Golay, primeira derivada, transformação Standard Normal Variate, SNV, e correção por mínimos quadrados assimétricos, AsLS (Eilers; Boelens, 2005). Para as rotinas baseadas em Savitzky-Golay foi empregada janela de 11 pontos e polinômio de segunda ordem. Na correção AsLS foram utilizados parâmetros previamente definidos no pipeline. Para fins exclusivamente gráficos, o sinal de referência foi deslocado verticalmente sem normalização adicional, evitando mascarar as diferenças produzidas pelas etapas de pré-processamento.

Para cada matriz pré-processada foi calculada uma PCA após centralização dos dados na média. Foram considerados, de forma integrada, a distribuição relativa da variância nas primeiras componentes, a estabilidade da estrutura de scores mediante validação cruzada leave-one-out, LOO-CV, a coerência química dos loadings, a preservação do perfil espectral e a ocorrência de ruído ou artefatos. Essa estrutura permitiu comparar tratamentos de complexidade distinta sem assumir previamente que maior intensidade de pré-processamento corresponde a melhor desempenho analítico.

Resultados

A execução do pipeline mostrou que as etapas de pré-processamento modificaram de maneira perceptível os perfis espectrais e a distribuição da informação no espaço multivariado. Como mostrado na Figura 2, correção de linha de base, derivação e padronização atuaram sobre aspectos distintos do sinal, tornando mais explícita a lógica de preparação dos dados para a PCA.

Figura 2 - Representação generalista da estrutura computacional do pipeline, incluindo leitura e organização da matriz espectral, pré-processamento, modelagem por PCA e comparação interpretativa dos resultados.



Fonte: Elaborado pelo autor (2026).

A comparação das PCAs confirmou que o tratamento escolhido altera a concentração relativa da informação nas primeiras componentes e a interpretação química dos loadings. Algumas rotas apresentaram comportamento mais favorável em determinados critérios, mas a decisão foi baseada no equilíbrio entre estabilidade, preservação espectral e coerência química, e não em uma métrica isolada. Essa leitura qualitativa evidencia, de forma consistente, que a etapa de pré-processamento é determinante para a organização multivariada dos dados espectrais de petróleo.

O fluxo computacional automatizou a aplicação dos tratamentos e a geração dos diagnósticos, favorecendo repetibilidade e rastreabilidade. Como sintetizado na Figura 2, a estrutura modular permitiu incorporar novas estratégias de correção e métricas sem alterar a lógica central de decisão.

Discussão

O pré-processamento não deve ser tratado como etapa meramente operacional, pois cada transformação altera a relação entre sinal químico e variabilidade indesejada (Wang et al., 2022). Como evidenciado na Figura 1, a correção de linha de base atua sobre o fundo, a derivação realça variações locais e a padronização reduz efeitos de escala, de modo que a escolha da rota interfere diretamente na interpretabilidade do espaço multivariado.

A leitura integrada de critérios constitui o principal diferencial do pipeline. A seleção metodológica deve considerar conjuntamente estabilidade, preservação da forma espectral, distribuição relativa da informação e coerência química, evitando decisões baseadas em uma única métrica (Khalighi et al., 2024). Essa abordagem torna a escolha do pré-processamento mais rastreável e criteriosa, fortalecendo a robustez metodológica dessa etapa.

Sob a perspectiva computacional, a implementação modular ilustrada na Figura 2 favorece reprodutibilidade e expansão para novos conjuntos de dados. O mesmo fluxo pode incorporar outras estratégias de correção, regiões espectrais, métricas de diagnóstico e técnicas multivariadas, funcionando como etapa de controle metodológico anterior à classificação, regressão ou integração com propriedades físico-químicas (Mohammadi et al., 2022).

Conclusão

O pipeline desenvolvido permitiu avaliar de forma estruturada e reprodutível o impacto de diferentes estratégias de pré-processamento sobre espectros ATR-FTIR de petróleos. As análises demonstraram que as transformações alteram a organização multivariada e a interpretabilidade dos dados, reforçando a necessidade de seleção criteriosa antes de aplicações quimiométricas subsequentes. Nesse sentido, o pipeline proposto contribui como ferramenta de apoio à caracterização multivariada de petróleos brutos, com potencial aplicação em rotinas de controle de qualidade e triagem analítica em contextos industriais e de pesquisa.

Referências

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. *Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra.* ***Applied Spectroscopy***, v. 43, p. 772-777, 1989.

BUTLER, H. J. et al. *Optimised spectral pre-processing for discrimination of biofluids via ATR-FTIR spectroscopy.* ***The Analyst***, 2018. DOI: 10.1039/c8an01384e.

EILERS, P. H. C.; BOELENS, H. F. M. *Baseline correction with asymmetric least squares smoothing.* ***Leiden University Medical Centre Report***, 2005.

KHALIGHI, S. et al. *Evaluating the impact of data pre-processing methods on classification of ATR-FTIR spectra of bituminous binders.* ***Fuel***, 2024. DOI: 10.1016/j.fuel.2024.132701.

MOHAMMADI, M. et al. *Rapid determination and classification of crude oils by ATR-FTIR spectroscopy and chemometric methods.* ***Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy***, 2020. DOI: 10.1016/j.saa.2020.118157.

MOHAMMADI, M. et al. *Classification and determination of sulfur content in crude oil samples by infrared spectrometry.* ***Infrared Physics & Technology***, 2022. DOI: 10.1016/j.infrared.2022.104382.

RINNAN, Å.; VAN DEN BERG, F.; ENGELSEN, S. B. *Review of the most common pre-processing techniques for near-infrared spectra.* ***TrAC Trends in Analytical Chemistry***, v. 28, p. 1201-1222, 2009.

SAVITZKY, A.; GOLAY, M. J. E. *Smoothing and differentiation of data by simplified least squares procedures.* ***Analytical Chemistry***, v. 36, p. 1627-1639, 1964.

WANG, H.-P. et al. *Recent advances of chemometric calibration methods in modern spectroscopy: algorithms, strategy, and related issues.* ***TrAC Trends in Analytical Chemistry***, 2022. DOI: 10.1016/j.trac.2022.116648.

WOLD, S.; ESBENSEN, K.; GELADI, P. *Principal component analysis.* ***Chemometrics and Intelligent Laboratory Systems***, v. 2, p. 37-52, 1987.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES). Os autores também agradecem ao Laboratório de Prepara de Amostras e Análises Químicas (LaPAQui), ao Programa de Pós-Graduação em Energia (PPGEN), ao Centro Universitário Norte do Espírito Santo (CEUNES) e à Universidade Federal do Espírito Santo (UFES) pelo suporte institucional e infraestrutura concedidos para a realização desta pesquisa.