ISSN: 2594-6153

Impacto dos Raios Cósmicos na Classificação dos Espectros Raman via LPA2v e PCA

Guilherme Mendes de Andrade[1], Rogério José Osti[1], Douglas de Jesus Passoni[1], Marcos Tadeu Tavares Pacheco[1,2], Dorotea Vilanova Garcia[1], Landulfo Silveira Jr.[1,2]

[1]Universidade Santa Cecília – UNISANTA, Santos-SP, Brasil
[2]Universidade Anhembi Morumbi – UAM, São José dos Campos, SP, Brasil

E-mail: guimean@gmail.com

Resumo: A espectroscopia Raman tem sido utilizada para analisar diversos tipos de substâncias, porém, a existência de ruídos pode interferir na análise dos picos Raman das amostras. Por meio das técnicas da lógica paraconsistente anotada de dois valores (LPA2v) e análise de componente principal e análise discriminante (PCA-DA) procurou-se avaliar se os raios cósmicos interferem na interpretação de sinais obtidos por Raman em amostras de gasolina sem e com presença de adulterantes. Os dados originais foram tratados, eliminando a linha de base de fluorescência. Foram adicionados manualmente picos equivalentes a raios cósmicos de diferentes intensidades e larguras de faixa. As técnicas de LPA2v e PCA-DA foram aplicadas nos dados sem e com raios cósmicos. Os erros de classificação das amostras de gasolina foram de 0% nas amostras sem raios cósmicos, 6% com raios fortes e 13% com raios largos para a técnica LPA2v. Para a técnica PCA, os erros foram de 6% nas amostras sem raios cósmicos, 13% com raios fortes e 6% com raios largos. Ambas as técnicas se mostraram promissoras para analisar a influência dos raios cósmicos nos picos Raman, porém ambas são afetadas pela existência destes ruídos.

Palavras-chave: espectroscopia Raman; gasolina; raios cósmicos; PCA; LPA2v.

Impact of Cosmic Rays on the Classification of Raman Spectra via LPA2v and PCA

Abstract: Raman spectroscopy has been used to analyze several types of substances, however, the existence of noise can interfere with the analysis of Raman peaks in the samples. Using the techniques of two-value annotated paraconsistent logic (LPA2v) and principal component analysis and discriminant analysis (PCA-DA), we sought to assess whether cosmic rays interfere with the interpretation of signals obtained by Raman in gasoline samples with and without presence of adulterants. The original data was treated, eliminating the fluorescence baseline. Peaks equivalent to cosmic rays of different intensities and band widths were manually added. The LPA2v and PCA-DA techniques were applied to data with and without cosmic rays. The classification errors of gasoline samples were 0% in samples without cosmic rays, 6% with strong rays and 13% with wide rays for the LPA2v technique. For the PCA technique, errors were 6% in samples without cosmic rays, 13% with strong rays and 6% with wide rays. Both techniques have shown promise for analyzing the influence of cosmic rays on Raman peaks, but both are affected by the existence of these noises.

Keywords: Raman spectroscopy; gasoline; cosmic rays; PCA; LPA2v.

Introdução

A tecnologia tem evoluído em diversos campos nas últimas décadas e com isto permitiu a melhoria na eficiência e na redução dos custos envolvidos dos componentes utilizados para medidas ópticas, o que possibilitou que as técnicas espectroscópicas fossem abundantemente usadas para a caracterização de materiais de diversos tipos [1].

Entre estas técnicas o espalhamento Raman se mostra promissora na análise comparativa destes materiais, pois é considerada uma técnica entre as mais modernas no qual não há necessidade do pré-tratamento ou preparação da amostra, tem um sistema mais rápido

de analise, não gera resíduos, possui custo baixo e pode produzir diagnóstico em tempo real [2].

A espectroscopia Raman é baseada no espalhamento inelástico da luz por determinada substância, onde a luz incidente polariza as moléculas da substância e a interação da vibração molecular com a luz incidente (durante a polarização) resulta em espalhamento com comprimento de onda diferente do incidente. Esta radiação espalhada, o sinal Raman, pode ser detectada por um espectrômetro Raman [3]. Ao medir a distribuição de intensidade e comprimento de onda da luz espalhada, é possível deduzir informações sobre a estrutura vibracional da substância irradiada [4].

A utilização do CCD (*charge-coupled device* ou dispositivo de carga acoplada) para realizar a medida do sinal difratado na grade de difração, leva a necessidade de considerarmos a presença de ruídos inerente a sistemas eletrônicos na captação do sinal luminoso por esse tipo de sensor CCD. Estes ruídos, ruído térmico, ruído eletrônico de leitura, ruído gerado pelo sinal óptico da amostra (ruído de fóton) e ruído gerado pelos raios cósmicos dificultam a interpretação dos espectros. [5].

Ao incidirem no CCD, raios cósmicos (partículas atômicas, núcleos de H e He de origem extraterrestre) que chegam à superfície da Terra interagem com o material do CCD e gera muitos sinais que se superpõe aos sinais gerados pelos fótons provenientes da amostra. Os raios cósmicos incidem aleatoriamente em pontos diferentes no CCD e interferem nos gráficos do espectro Raman com picos bem estreitos e de alta intensidade [6]. Para eliminar a interferência do ruído dos raios cósmicos, são utilizadas técnicas de processamento de dados, a maioria delas baseada na taxa de variação espectral, já que o raio cósmico caracteriza-se por apresentar alta intensidade em uma pequena largura de banda espectral.

As técnicas de lógica paraconsistente anotada com anotação de dois valores (LPA2v) e análise de componente principal (PCA) [7-9] são técnicas que podem ser utilizadas para classificar ou agrupar os espectros de um conjunto de amostras baseado nas diferenças espectrais relacionadas com a diferença na composição química das amostras. Os ruídos podem interferir na classificação, pois podem ser confundidos com sinal.

Objetivos

O objetivo deste trabalho foi analisar mediante a aplicação da LPA2v e da PCA se os ruídos provocados por raios cósmicos interferem na avaliação entre gasolina comum e gasolina adulterada na análise utilizando espectroscopia Raman no infravermelho próximo de 45 amostras de gasolina de 15 postos de abastecimentos.

Material e Métodos

O estudo utilizou espectros Raman de gasolina comum e adulteradas do estudo de Bezerra et al. (2018) [10]. Foram coletadas amostras de gasolina de 15 postos de combustíveis. As amostras foram submetidas à espectroscopia Raman (sistema Raman dispersivo modelo *Dimension* P-1, *Lambda Solutions*, MA, EUA, com excitação de 830 nm e laser com ajuste de potência de até 350 mW, resolução espectral de 2 cm^{-1}).

Cinco destes postos apresentaram gasolina sem adulteração, e 10 amostras apresentaram sinais de adulteração.

A base de dados foi estruturada em três matrizes: a primeira com espectros sem raios cósmicos, a segunda com adição de raios cósmicos de alta intensidade alta, e a terceira com adição de raios cósmicos de intensidade média, porém com bandas largas. Estes picos foram inseridos de forma aleatória, simulando o efeito que ocorre na coleta dos dados. A classificação das amostras em gasolina comum e gasolina adulterada foi feita utilizando as rotinas de LPA2v e PCA-DA. Em ambos os métodos foi utilizado o software MATLAB *Starter Application* 2018 versão 1.0.0.1 (*MathWorks*, MA, USA).

Resultados

A classificação dos espectros das amostras de gasolina pelas técnicas LPA2v e PCA-DA aplicada nos espectros sem ruído (Limpos) e com raios cósmicos fortes (Fortes) e largos (Largos) é apresentada na Tabela 1. A técnica LPA2v promoveu um erro de 0% em relação à premissa (classificação dos postos conforme trabalho de Bezerra et al., 2018 [10]) para os espectros Limpos. Já o grupo Fortes apresentou erro de 6%, enquanto o grupo Largos apresentou erro de 13%. A técnica PCA promoveu um erro de 6% em relação à premissa para os espectros Limpos. Já o grupo Fortes apresentou erro de 13%, enquanto que o grupo Largos apresentou erro de 6%.

Tabela 1. Classificação dos espectros dos postos de acordo com o tipo de gasolina (comum = C e adulterada = A) pelas técnicas LPA2v e PCA-DA.

Postos	Premissa	Limpos LPA2v	Fortes LPA2v	Largos LPA2v	Limpos PCA	Fortes PCA	Largos PCA
1	Adulterada	A	A	A	A	A	A
2	Adulterada	A	A	C	A	C	A
3	Comum	C	C	C	C	C	C
4	Comum	C	C	C	C	C	C
5	Comum	C	C	C	C	C	C
6	Adulterada	A	C	C	C	C	C
7	Adulterada	A	A	A	A	A	A
8	Adulterada	A	A	A	A	A	A
9	Adulterada	A	A	A	A	A	A
10	Adulterada	A	A	A	A	A	A
11	Adulterada	A	A	A	A	A	A
12	Adulterada	A	A	A	A	A	A
13	Comum	C	C	C	C	C	C
14	Adulterada	A	A	A	A	A	A
15	Comum	C	C	C	C	C	C
Erro em relação à premissa		0%	6%	13%	6%	13%	6%

A Tabela 2 mostra que a melhor classificação para a técnica PCA-DA foi obtida utilizando-se a distância quadrática com o emprego das variáveis da PCA (*Coefficients*) de 1 a 5 com os espectros do grupo Limpos, com 4,4% de erro. Devido aos raios cósmicos, os

autovetores (variáveis latentes) da PCA alteraram entre os grupos Fortes e Largos comparativamente ao grupo Limpos, com 99% da variância explicada no primeiro vetor (*Score* 1) para o grupo Limpos, 97% para o grupo Fortes e 91% para o grupo Largos.

Os resultados da LPA2v mostrados na Tabela 1 indicam que a gasolina do Posto 6 (gasolina adulterada) passou a ser classificado como gasolina comum nos grupos Fortes e Largos. Observa-se na Tabela 3 que o espectro da 17ª amostra, no qual a ocorrência de 286 e grau de frequência 0,75639, em comparação com o grupo Fortes, houve um aumento no número de ocorrências para 309 e uma redução no grau de frequência para 0,7368, assim como no grupo Largos, onde a ocorrência também aumentou para 309 e o grau de frequência diminuiu para 0,72487. Isso decorre da forma como o extrator de contradição trabalha: utilizando valores máximos e mínimos, e já que os raios cósmicos em determinado comprimento de onda vão sinalizar um valor máximo ou mínimo dependendo de onde ele incide no CCD, interfere no padrão do LPA2v e aumenta o número de ocorrências, diminuindo o grau de frequência [7, 11]. Portanto, é necessário que haja a extração dos raios cósmicos antes de utilizar a LPA2v.

Tabela 2. Classificação dos espectros via PCA-DA utilizando as distâncias linear, quadrática e de Mahalanobis para os grupos Limpos, Fortes e Largos.

	LINEAR				
	grupo quadract coeff 01	grupo quadract coeff 01 a 02	grupo quadract coeff 01 a 03	grupo quadract coeff 01 a 04	grupo quadract coeff 01 a 05
LIMPOS	28.9%	26.7%	22.2%	20.0%	26.7%
FORTES	31.1%	26.7%	28.9%	28.9%	28.9%
LARGOS	44.4%	42.2%	24.4%	24.4%	24.4%

	QUADRATIC				
	grupo quadract coeff 01	grupo quadract coeff 01 a 02	grupo quadract coeff 01 a 03	grupo quadract coeff 01 a 04	grupo quadract coeff 01 a 05
LIMPOS	22.2%	15.6%	13.3%	13.3%	4.4%
FORTES	22.2%	17.8%	11.1%	17.8%	11.1%
LARGOS	28.9%	35.6%	22.2%	13.3%	13.3%

	MAHAL				
	grupo quadract coeff 01	grupo quadract coeff 01 a 02	grupo quadract coeff 01 a 03	grupo quadract coeff 01 a 04	grupo quadract coeff 01 a 05
LIMPOS	24.4%	17.8%	26.7%	24.4%	24.4%
FORTES	20.0%	22.2%	20.0%	26.7%	24.4%
LARGOS	37.8%	31.1%	24.4%	20.0%	31.1%

Tabela 3. Classificação dos espectros via LPA2v: ocorrência e frequência para os espectros dos grupos Limpos, Fortes e Largos.

Posto 6 as amostras são GASOLINA ADULTERADA		Posto 6 as amostras são GASOLINA COMUM		Posto 6 as amostras são GASOLINA COMUM	
OCORRÊNCIA LIMPOS	FREQUÊNCIA LIMPOS	OCORRÊNCIA FORTES	FREQUÊNCIA FORTESS	OCORRÊNCIA LARGOS	FREQUÊNCIA LARGOS
16° amostra tem ocorrência = 233	GASOLINA ADULTERADA grau de frequência = 0.80153	16° amostra tem ocorrência = 253	GASOLINA ADULTERADA grau de frequência = 0.7845	16° amostra tem ocorrência = 253	GASOLINA ADULTERADA grau de frequência = 0.76491
17° amostra tem ocorrência = 286	GASOLINA ADULTERADA grau de frequência = 0.75639	17° amostra tem ocorrência = 309	GASOLINA COMUM grau de frequência = 0.7368	17° amostra tem ocorrência = 309	GASOLINA COMUM grau de frequência = 0.72487
18° amostra tem ocorrência = 314	GASOLINA COMUM grau de frequência = 0.73254	18° amostra tem ocorrência = 348	GASOLINA COMUM grau de frequência = 0.70358	18° amostra tem ocorrência = 348	GASOLINA COMUM grau de frequência = 0.69336

Discussão

A espectroscopia Raman tem sido utilizada para analisar diversos tipos de substâncias. A existência de ruídos de diversas origens pode interferir na análise dos picos das amostras analisadas [5, 6]. Por meio das técnicas LPA2v e PCA procurou-se avaliar se, particularmente, se os ruídos gerados por raios cósmicos interferem na interpretação de sinais obtidos pela técnica Raman nessas 45 amostras de gasolina (45 espectros em tréplica).

Os resultados da Tabela 3 mostram que é necessário que haja a limpeza dos raios cósmicos na LPA2v, pois mesmo que eles sejam estreitos ou largos a LPA2v não conseguiu analisar corretamente o espectro sem a retirada desse tipo de ruído.

Para a PCA, a Tabela 1 mostrou erro de 6% em relação à premissa, e este erro que ocorreu para o Posto 6 pode indicar na realidade a presença excessiva de etanol onde a metodologia estatística pode ser aprimorada para refletir exatamente os dados obtidos durante a geração dos espectros Raman [12].

Quando se analisa a influência dos raios cósmicos na classificação via PCA, o melhor resultado aparece utilizando *Coefficients* de 1 a 3 no grupo Fortes e *Coefficients* de 1 a 5 para o grupo Largos. Os raios cósmicos podem não interferir acentuadamente, pois no grupo quadratic e *Coefficients* de 1 a 3 há uma melhora na informação espectral de Limpos para Fortes, e no grupo quadratic e *Coefficients* de 1 a 4, os grupos Limpos e Largos tem praticamente a mesma informação espectral. No grupo Fortes, como nos Postos 2 e 6 a gasolina adulterada passou a ser classificada como gasolina comum, e no grupo Largos como o Posto 6 a gasolina adulterada passou a ser gasolina comum, sugere-se que os raios cósmicos interferem na interpretação dos espectros, desta forma deixando claro que a extração se faz necessária também para a PCA. Observou-se em todas as situação de erro na avaliação das amostras de um "falso negativo", isto é, amostras adulteradas produziram resultados de gasolina não adulterada.

Conclusões

A LPA2v se mostrou melhor na classificação dos espectros em relação a técnica PCA, uma vez que no grupo sem raios cósmicos (grupo Limpos) o erro foi de 0%, enquanto que para a PCA, o erro foi de 6% em relação à premissa. Ambas as técnicas se mostraram promissoras para analisar a influência dos raios cósmicos nos espectros Raman das 45 amostras de gasolina, porém, ambas são afetadas pela existência destes ruídos. Portanto recomenda-se que a análise seja feita após extração dos raios cósmicos.

Referências

1. Santos AR. et al. Aplicação da espectroscopia Raman na caracterização de minerais pertencentes a uma geocoleção. **Química Nova**. vol. 42, n° 5, p. 489-496, 2019. DOI: 10.21577/0100-4042.20170358.
2. Butler HJ. et al. Using Raman spectroscopy to characterize biological materials. **Nature Protocols**. vol. 11, n° 4, p. 664-687, 2016. DOI: 10.1038/nprot.2016.036.
3. Bettignies P. Optics/instrumentation Micro-Raman spectroscopy: theory and application. **Physical Sciences Reviews**. vol. 5, n° 1, p. 31, 2020. DOI: 10.1515/psr-2019-0027.
4. Tsukada S, Fujii Y. Multivariate curve resolution for angle-resolved polarized Raman spectroscopy of ferroelectric crystals. **Japanese Journal of Applied Physics**. vol. 59, n° 5, p. SKKA03, 2020. DOI: 10.35848/1347-4065/ab78e8
5. Zeña AC. et al. Fontes de ruídos e relação sinal-ruído em espectrômetro Raman dispersivo utilizando câmera CCD. **Anais do XXIV Congresso Brasileiro de Engenharia Biomédica – CBEB 2014**. vol. 1, p. 2770-2773, 2014. https://www.canal6.com.br/cbeb/2014/artigos/cbeb2014_submission_821.pdf.
6. Sabin GP. et al. Desenvolvimento de um algoritmo para identificação e correção de *spikes* em espectroscopia Raman de imagem. **Química Nova**. vol. 35, n° 3, p. 612-615, 2012. DOI: 10.1590/S0100-40422012000300030.
7. Silveira RS. et al. Simulação do reticulado da Lógica Paraconsistente Anotada com anotação de dois valores LPA2v aplicado em VB.net. **Anais do Encontro Nacional de Pós-Graduação – VI ENPG**. vol. 1, p. 46-50, 2017. https://ojs.unisanta.br/index.php/ENPG/article/download/1068/996.
8. Mendes TF, Souza DL. Application of principal component analysis method (PCA) for fault detection in chemical plants. **Research, Society and Development**. vol. 9, n° 8, e957986335, 2020. https://rsdjournal.org/index.php/rsd/article/view/6335/6062.
9. Hongyu K, Sandanielo VLM, Oliveira Junior GJ. Análise de componentes principais: resumo teórico, aplicação e interpretação. **Engineering and Science**. vol. 5, n° 1, p. 83-90, 2016. DOI: 10.18607/ES20165053.
10. Bezerra ACM. et al. Quantification of anhydrous ethanol and detection of adulterants in commercial Brazilian gasoline by Raman spectroscopy. **Instrumentation Science & Technology**. vol. 47, n° 1, p. 90-106, 2019. DOI: 10.1080/10739149.2018.1470535.
11. Abe JM. et al. Three decades of paracosistente annotated logics: a review paper on some applications. **Procedia Computer Science**. vol. 159, p. 1175-1181, 2019. DOI: 10.1016/j.procs.2019.09.286.
12. Martins JC. et al. Determinação da concentração de etanol na gasolina comum através da técnica de espectroscopia Raman. **Unisanta Science and Technology**. vol. 5, n° 3, p. 200-209, 2016. https://ojs.unisanta.br/index.php/sat/article/download/779/813.