

Classificação de biodiesel por espectrometria NIR utilizando SIMCA e KNN com seleção de variáveis

Adenilton Camilo da Silva¹ (IC), Germano Vêras¹ (PQ)*, Gean Bezerra da Costa¹ (IC), Lorena Cristina Nóbrega Felix¹ (IC)

1- DQ - CCT- LQAQ - Universidade Estadual da Paraíba Av Juvêncio Arruda s/n, Campina Grande-PB

*germano@uepb.edu.br

Palavras Chave: Biodiesel, espectrometria NIR, APS, SIMCA, KNN.

Introdução

A produção de biodiesel tornou-se uma forma para diversificar a matriz energética mundial, sendo uma alternativa aos combustíveis minerais oriundos do petróleo. O biodiesel é um combustível que pode ser definido como sendo um mono-álcool éster de ácidos graxos derivados de fontes renováveis, como óleos vegetais e gorduras animais.

Atualmente, o biodiesel consumido no Brasil tem como principais matérias primas: óleo de soja, sebo bovino e óleo de algodão¹. Porém, outras culturas, como girassol, mamona, amendoim, canola, entre outros, surgem como proposta ao PNPB (Programa Nacional de Produção e Uso de Biodiesel).

Dentro deste contexto, este trabalho tem como objetivo apresentar um método para a classificação de biodiesel utilizando espectrometria (NIR), em conjunto com as técnicas de reconhecimento de padrões: SIMCA e KNN, avaliando a capacidade classificatória das mesmas, ao se aplicar a técnica de seleção de variáveis APS.

Resultados e Discussão

Para o referido estudo, foram analisadas 9 amostras de biodiesel de algodão, 10 de girassol, 9 de soja e 9 de canola, sendo estas obtidas por óleos de diferentes marcas e lotes.

Os espectros foram registrados em triplicata na região de 1100-2200 nm, com resolução de 0,5nm, utilizando o instrumento Vis/NIR FOSS XDS MasterLab™. Os dados foram pré-processados aplicando o algoritmo de Savitzky-Golay usando uma janela de 21 pontos, primeira derivada e ajuste com polinômio de segundo grau, através do software The Unscrambler 9.8®.

Utilizando o algoritmo de Kennard Stones (KS), no MatLab®, dividiu-se os espectros em dois grupos: calibração (56 espectros) e validação (55 espectros).

Os modelos SIMCA, foram construídos fazendo PCA para cada tipo de biodiesel do grupo de calibração no software The Unscrambler 9.8®. O número de PCs para cada modelo garantiu uma variância explicada maior ou igual a 98%.

A classificação por knn foi realizada no software MatLab®, onde o grupo de calibração foi utilizado para treinamento e o grupo de validação foi utilizado para teste.

34ª Reunião Anual da Sociedade Brasileira de Química

Para a seleção de variáveis utilizou-se o Algoritmo das Projeções Sucessivas (APS) desenvolvido no software MatLab®. Com base no gráfico de custo de validação x número de variáveis 12 variáveis foram selecionadas.

Para verificar o poder discriminatório das variáveis selecionadas pelo algoritmo APS, foram observados os erros de classificação SIMCA em três níveis de significância (5%, 10% e 25%), e KNN em três níveis k (1, 2 e 3) antes (tabela-1) e depois da seleção de variáveis (Tabela-2).

Tabela 1 – Erros encontrados para classificação utilizando o espectro completo.

Erros	Técnica de Classificação					
	SIMCA			KNN		
	5%	10%	25%	k=1	k=2	k=3
tipo I	2	2	11	2	8	12
tipo II	4	1	0	2	1	8
total	20			33		

Tabela 2 – Erros encontrados para classificação utilizando variáveis selecionadas.

Erros	Técnica de Classificação					
	SIMCA			KNN		
	5%	10%	25%	k=1	k=2	k=3
tipo I	0	0	4	0	0	0
tipo II	0	0	0	0	0	0
total	4			0		

De acordo com as tabelas 1 e 2, ao aplicar-se a seleção de variáveis, verificou-se erros apenas na classificação em SIMCA a 25% de significância, deixando evidente o poder discriminatório dos comprimentos de onda selecionados.

Conclusões

Conclui-se que o uso de seleção de variáveis associada a SIMCA e KNN, são eficientes para classificação de biodiesel.

Agradecimentos

Agradecimento ao CNPq pelas bolsas de Iniciação Tecnológica e Industrial e pelo financiamento do projeto n°: 576416/2008-8.

¹ Agência Nacional de Petróleo – ANP. Boletim Mensal de Biodiesel-SRP de 12/2010. Disponível em <<http://www.anp.gov.br/?dw=40428>>. Acesso em 27/01/2011