Predição do índice de cetanas de diferentes tipos de biodiesel utilizando PCA e técnicas de *clustering*.

Diogo Viola De Nadai[1*] (PG), Paulo Cesar Muniz de Lacerda Miranda[1] (PQ), Carlos Eduardo Novo Gatts[2] (PQ). e-mail: *diogonadai@gmail.com*

[1]LCQUI, [2]LCFIS - Universidade Estadual do Norte Fluminense (UENF), Campos dos Goytacazes - RJ, CEP:28013602.
Palavras Chave: *biodiesel, índice de cetanas, RMN-[1]H, PCA, cluster hierárquico, cluster não hierárquico.*

Introdução

A Análise de Componentes Principais (PCA) é uma ferramenta clássica, largamente empregada em análise estatística. Neste método, as n-variáveis originais geram, através de suas combinações lineares, m-componentes principais (onde m < n).

Os métodos de "clustering" hierárquico são implementados usando um procedimento de maneira a aglomerar os objetos do conjunto de dados, dispondo os dados de maneira hierárquica, em um dendrograma. Já os de "clustering" não hierárquico são implementados de maneira a distribuir os n-dados em k-"clusters" (n > k).[1]

Neste trabalho buscamos fazer uma correlação entre um banco com dados de espectros teóricos de RMN-^{1}H e o índice de cetanas (IC) de diferentes tipos de biodiesel utilizando a PCA e métodos de cluster. O tratamento dos dados foi realizado com a utilização do ambiente estatístico R. [2]

Resultados e Discussão

Para compor a base de dados foram escolhidos ésteres e diferentes tipos de biodiesel com IC disponíveis na literatura, resultando num total de 44 elementos (35 ésteres e 9 tipos de biodiesel).

Utilizando o programa gNMR, foram simulados os espectros de RMN-^{1}H dos ésteres componentes. Os espectros foram fatiados em 7 intervalos, referentes aos diferentes tipos de hidrogênios presente nestas moléculas. O banco de dados foi formatado utilizando a integração dos sinais presentes em cada uma destas fatias. Sobre este banco foi feita uma PCA. O gráfico 1 mostra a existência de 5 componentes principais, sendo que o PC1 somado ao PC2 equivale a 83,65% da variância do sistema.

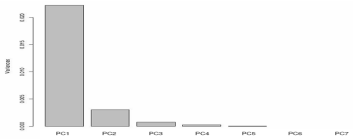

Gráfico 1. Resultado das componentes principais resultantes da PCA.

A figura 1 é uma representação dos resultados no plano PC1 x PC2. Avaliando a mesma, observa-se que a PC1 corresponde ao número de insaturações presentes nas cadeias carbônicas dos constituintes deste banco. Segundo Knothe[3], este é o principal fator que interfere no IC destes tipos de biocombustíveis.

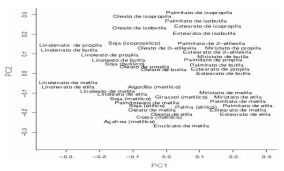

Figura 1. Resultados obtidos da PCA (PC1xPC2).

Depois foi feito um "*clustering*" hierárquico sobre o banco de dados, a figura 2 expõe estes resultados.

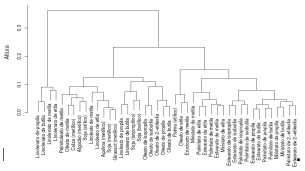

Figura 2. Dendrogr

Nesta figura observa-se que, inicialmente, este método separa os ésteres linolênicos dos demais elementos do banco de dados. Posteriormente há a separação entre os ésteres insaturados e saturados, ou seja, está havendo coerência no critério de separação dos dados em relação ao PC1. Construiu-se ainda um "*clustering*" não hierárquico sobre o banco de dados onde os resultados são mostrados na figura 3.

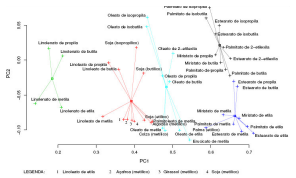

Figura 3. Resultados obtidos do "clustering" não hierárquico.

Nesta figura observa-se que o critério utilizado para a divisão dos ésteres em cada cluster também foi o mesmo aplicado nos tratamentos anteriores.

Conclusões

Em todos os tratamentos o principal critério para a discriminação dos dados foi o número de ligações duplas C=C presente nas cadeias carbônicas. Segundo Knothe[3] este é o principal fator que interfere no IC destes biocombustíveis.

[1] Xu J., Zhang Q., Shih C. *Molecular Diversity*. **2006**, 10, 463-478.

[2] http://www.r-project.org, acessado em 23/11/2007.

[3] Knothe, G., Matheaus, A. C., Ryan III, T. W. *Fuel*. **2003,** 82, 971-975.