

Análise de Componentes Principais Aplicada à Espectroscopia no Infravermelho de Misturas de Diesel e Biodiesel: Estudos de Casos

Ana Luiza Pessanha Blanco (Química – FTESM)

Alexandre Nelson Martiniano Carauta (Química – FTESM)

RESUMO

Um estudo da aplicação de métodos multivariados com destaque para a análise de componentes principais em misturas complexas, a partir de estudos de casos, foi realizado. Destaque foi dado à associação desses métodos às técnicas de espectroscopia no infravermelho aplicadas principalmente ao biodiesel devido ao seu interesse tecnológico e ambiental como fonte de energia alternativa e renovável. A utilização da análise de componentes principais como técnica exploratória e de reconhecimento de padrões foi mostrada em diversos casos diferentes como na análise de blends de biodiesel e diesel, na avaliação de adulteração de biodiesel e na tentativa de desenvolvimento de modelos de calibração para a determinação de parâmetros de controle de qualidade de biodiesel. Por último, foi demonstrada didaticamente a aplicação desta técnica com o uso de um programa comercial.

Palavras-chave: Análise de Componentes Principais, Infravermelho, Biodiesel.

1. INTRODUÇÃO

A análise de dados de sistemas cada vez mais complexos em que muitas variáveis podem ser obtidas por instrumentos cada vez mais sofisticados com muita rapidez tem levado à necessidade da utilização de técnicas de tratamentos de dados por métodos estatísticos mais complexos e robustos. Uma única amostra pode fornecer muitas variáveis a serem medidas por tais

instrumentos e somente técnicas multivariadas são apropriadas para tratar adequadamente esses dados (FERREIRA et al., 1999).

A quimiometria surgiu exatamente dessa necessidade crescente de análise de dados multivariados e pode ser definida como o “ramo da ciência cujo objetivo é o de utilizar técnicas matemáticas (principalmente estatísticas) no tratamento e interpretação de dados químicos” (PASQUINI, 2003), ou ainda, segundo Kowalski, uma disciplina da química que usa métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação na análise de dados de natureza multivariada (VANDEGINSTE, 1987 apud FERREIRA, 1999), não sendo uma disciplina matemática e sim da química, pois se destina a resolver problemas de origem química. A matemática, a estatística e as ciências da computação fornecem as ferramentas.

Dentre as diversas técnicas existentes utilizadas em tratamento de dados multivariados, duas se destacam por sua relativa simplicidade e que têm sido empregadas com bastante êxito: a análise de componentes principais (PCA, do inglês Principal Component Analysis) que é uma técnica exploratória e que permite entre outras coisas relevar a existência ou não de amostras anômalas, de relações entre as variáveis medidas e de relações ou agrupamentos de amostras; e a regressão por mínimos quadrados parciais (PLS, do inglês Partial Least Square) que pode ser usada para calibração e que é derivada do PCA. (LYRA et al., 2010).

A utilização dessas técnicas associadas às análises espectrofotométricas quantitativas, principalmente à espectroscopia na região do infravermelho, tem tido grande êxito. Pode ser destacada a calibração multivariada de métodos de espectroscopia no infravermelho próximo (Near Infra-red – NIR) que por ser uma técnica analítica que apresenta como vantagens, rapidez, robustez, necessitando de quantidades mínimas de amostra, além de ser não destrutiva e não invasiva, tem sido muito utilizada para análise de misturas de múltiplos componentes, como no caso de misturas complexas de hidrocarbonetos, o que caracteriza o petróleo e seus derivados. De grande interesse também tem sido a utilização dessas técnicas na caracterização e no controle de qualidade do biodiesel que surge como uma fonte de energia renovável para substituir ou ser adicionada ao óleo diesel proveniente do petróleo (FERREIRA et al., 1999).

Este trabalho tem por objetivo realizar uma análise da aplicação de métodos multivariados associados principalmente à espectroscopia no

infravermelho na análise de misturas complexas com destaque para o biodiesel através de estudos de casos.

2. QUIMIOMETRIA

A Quimiometria é uma ferramenta matemática e estatística utilizada para o tratamento de dados químicos visando extrair o máximo de informações possíveis, possibilitando a compressão dos dados, permitindo a visualização gráfica de todo o conjunto de elementos e a presença ou ausência de agrupamentos naturais entre as amostras. É de suma relevância no monitoramento online de processos industriais quando associada à Tecnologia Analítica de Processos (PAT) (GUIMARÃES, 2014; CORREA, 2012; HENRIQUES, 2011).

2.1 História da Quimiometria no Brasil e sua Importância na Indústria

A Quimiometria surgiu durante a década de 70 com o advento dos computadores nos laboratórios, pois não existiam meios tecnológicos para calcular a pesada carga de cálculos que as análises forneciam, principalmente com o avanço dos equipamentos. Ainda assim, a primeira dissertação defendida sobre o assunto só ocorreu em fevereiro de 1981, por Ieda S. Scarminio, e, meses depois, Elías Zagatto apresentou a primeira tese utilizando técnicas quimiométricas no país.

Em 1985, o primeiro curso regular semestral foi oferecido no programa de pós-graduação do IQ-UNICAMP, com apenas dois alunos inscritos. Apesar desta adversidade, as indústrias nacionais como a Souza Cruz S. A., a Docegeo e a 3M do Brasil começavam a se interessar pelo uso dos métodos quimiométricos e tal era este interesse, que a 3M do Brasil solicitou um curso sobre otimização de misturas sendo que nenhuma faculdade tinha explorado este tema na área de química ainda. Então, eles forneceram o material bibliográfico para que o curso fosse ministrado em 1988 na própria empresa.

Após esse evento, o interesse pela técnica cresceu drasticamente, não ficando restrito apenas à área acadêmica, mas indústrias como a Oxiteno, Pirelli, Braskem, Clariant, Nitroquímica, Unilever, Petroflex e Masterfoods do Brasil as têm empregado, e continuam empregando. Um exemplo prático

do uso é no caso da Petrobrás, que já possui nas refinarias um analisador em linha para monitoramento das correntes de gasolina utilizando a espectroscopia no infravermelho próximo e métodos multivariados para prever parâmetros de qualidade do derivado de petróleo como destilação, octanagem e concentrações de aromáticos, olefinas e saturados. Esse trabalho foi resultado da tese de Aerenton Ferreira Bueno (2011) intitulada “Desenvolvimento de um analisador de processo por espectroscopia no infravermelho próximo (NIR) para previsão de propriedades de derivados de petróleo”.

Hoje é possível dizer que o Brasil tem grupos de pesquisa e desenvolvimento consolidados em quimiometria, não só academicamente como em empresas de grande porte. Isso se deve aos esforços de publicação de vários pesquisadores e as três escolas de verão, organizadas pela PUC-RJ, sob coordenação dos professores Pércio Farias (1999), José Godoy (2000) e Ricardo Aurélio (2004). Mas um fato marcante, que provavelmente é a prova do engajamento e interesse por essa linha de estudo, foi a realização do encontro internacional Chemometrics in Analytical Chemistry Conference (CAC) que ocorreu no período de 10-14 de setembro de 2006 na UNICAMP, em São Paulo, coordenada pela professora Márcia Ferreira (BARROS NETO, SCARMÍNIO & BRUNS, 2006).

2.2 Métodos Quimiométricos

O desenvolvimento da quimiometria irrompeu a necessidade de se desenvolver novas metodologias que pudessem ser aplicadas a objetivos distintos. Para isso, ela foi dividida de acordo com as diferentes linhas de pesquisa em processamento de sinais analíticos, planejamento e otimização de experimentos, reconhecimento de padrões e classificação de dados, calibração multivariada e métodos de inteligência artificial.

Dentro dessa ramificação há muitos métodos que são rotineiramente utilizados, como HCA, ANOVA, redes neurais, PCR, entre outros, mas os principais são PCA e PLS.

3. ANÁLISE DE COMPONENTES PRINCIPAIS

Um método para o cálculo de duas ou três variáveis foi desenvolvido em 1901 por Karl Pearson. Denominado de análise de componentes

principais (PCA), era acreditado ser a solução de alguns dos problemas da época. Em 1933, Hotteling aperfeiçoou o método, mas ainda assim os cálculos eram muito complexos para mais do que algumas variáveis, pois tinham que ser feitos a mão. Somente com o advento dos computadores que a técnica começou a ser amplamente utilizada.

Com o objetivo de visualizar a estrutura de dados, encontrar similaridades entre amostras, detectar amostras anômalas (outlier) e reduzir a dimensionalidade do conjunto de dados, a PCA se tornou uma das técnicas primordiais da Quimiometria, sendo a base para os mais diversos métodos de reconhecimento de padrões, classificação e calibração multivariada (TEÓFILO, 2013).

A análise de componentes principais consiste em reduzir a dimensionalidade através de componentes principais (PCs), permitindo a compressão dos dados através da combinação linear das variáveis originais com o intuito de explicitar as informações mais importantes, facilitando sua interpretação (CORREIA & FERREIRA, 2007).

Matematicamente, esta técnica decompõe a matriz de dados X em outras três matrizes de modo que $X = TPt + E$, onde T é a matriz de escores, P é a matriz de pesos e E é a matriz de resíduos. O t sobrescrito (t) é o operador de transposição da matriz (LYRA et al., 2010).

“Os escores representam as novas coordenadas dos objetos nas novas variáveis que são as PCs” (LYRA et al., 2010). Os coeficientes de combinação, denominados pesos, são os cossenos dos ângulos entre as componentes principais e as variáveis originais, retratando quanto cada variável original contribui para uma determinada PC (TEÓFILO, 2013).

O estudo dos pesos mostra quais variáveis mais auxiliam os agrupamentos na observação do gráfico de escores. Por meio desta análise conjunta entre gráfico de escores e pesos, há possibilidade de verificar quais são os autores das diferenças mostradas entre as amostras. O número de componentes principais a ser utilizado no modelo PCA será demonstrado pela porcentagem de variância explicada. Por isso, escolhe-se um número de componentes de tal maneira que a maior porcentagem da variação presente no conjunto de dados originais seja capturada (SANTANA, 2013).

A primeira componente principal (PC1) é traçada no sentido da maior variação no conjunto de dados, sendo descrita como $PC1 = t1p1t$; a segunda (PC2) é traçada ortogonalmente à primeira, sendo descrita como $PC2 = t2p2t$ que corresponde à direção de maior variância no espaço multi-

variado não modelada por PC1, obtida através do resíduo de X descontada a primeira componente, $E1 = X - t1p1t$. A terceira PC é obtida através do resíduo deixado pelas duas primeiras componentes, $E2 = E1 - t2p2t$, de forma a modelar sempre a direção de maior variância não determinada pelas PCs anteriores e sempre ortogonais à elas (TEÓFILO, 2013; LYRA et al., 2010).

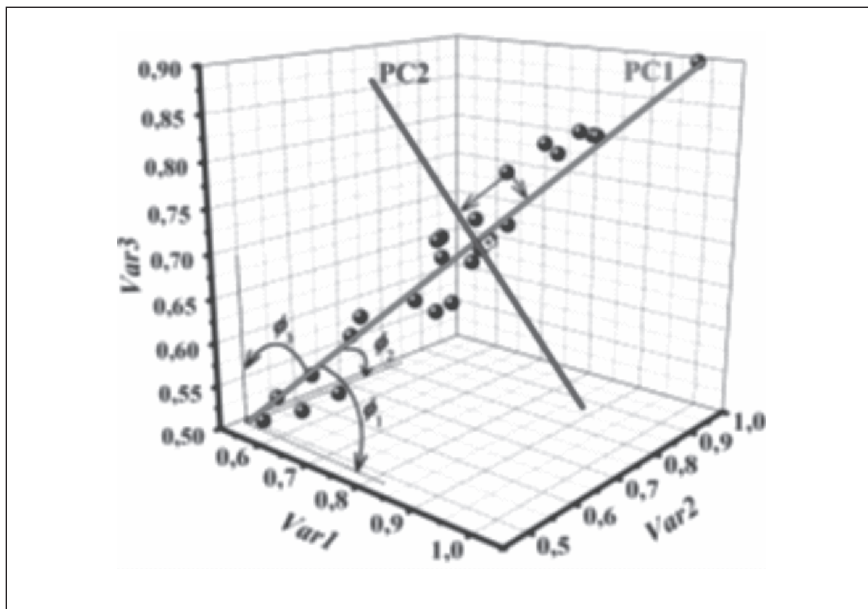


Figura 1: Representação geométrica de duas componentes principais em um espaço com três variáveis (3D) (TEÓFILO, 2013).

4. ESPECTROSCOPIA NO INFRAVERMELHO

A espectroscopia no infravermelho é uma técnica versátil que pode ser usada para diversas determinações qualitativas e quantitativas de compostos bioquímicos, orgânicos e inorgânicos (CORREA, 2012). Seu objetivo é o estudo dos níveis de energia de átomos ou moléculas e consiste em estudar a interação da radiação eletromagnética com a matéria (LUZ, 2003).

Situada entre a região do visível e das micro-ondas no espectro eletromagnético (Figura 2), é usualmente dividida em infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (FIR).

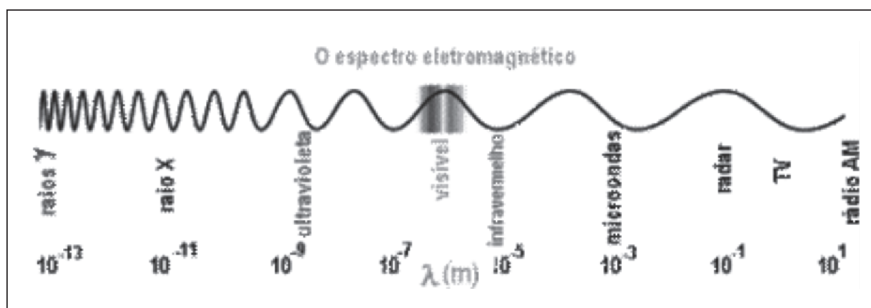


Figura 2: Faixa do Espectro Eletromagnético (Fraquelli, 2015).

As vibrações moleculares ocorrem quando uma radiação de comprimento de onda na faixa do espectro do infravermelho incide sobre uma molécula com a ocorrência de transferência de energia. Isso provoca uma variação da amplitude da vibração molecular desde que resulte na variação do seu momento de dipolo elétrico (o momento de dipolo é determinado pela magnitude da diferença de carga e a distância entre dois centros de carga), ou seja, deve haver assimetria suficiente no modo vibracional para que haja um momento de dipolo resultante diferente de zero e provoque uma oscilação do campo elétrico, o que resulta na absorção da radiação infravermelha. Somente assim, um modo vibracional será ativo e um sinal será gerado no espectro de infravermelho. Por isso, moléculas totalmente simétricas como o tetracloreto de carbono (CCl_4), iodo (I_2) e cloro (Cl_2) não apresentam espectros de infravermelho no estado gasoso, e no estado líquido, pequenos sinais podem ser vistos, mas por pequenas quebras dessas simetrias devido às forças intermoleculares, pois nenhuma variação efetiva no momento de dipolo ocorre durante a vibração ou rotação dessas moléculas (SALA, 1995).

As vibrações moleculares podem ser classificadas em estiramento, deformação angular e torções, e podem ser simétricas ou assimétricas. As deformações angulares podem ainda ser classificadas como no plano ou fora do plano (CARAUTA, 1996).

4.1 Espectroscopia no Infravermelho Próximo (NIR – Near Infrared)

As bandas de absorção que aparecem na região do espectro vibracional acima de 4000 cm^{-1} são consideradas bandas da região do infravermelho próximo. Praticamente, essa região do espectro não apresenta

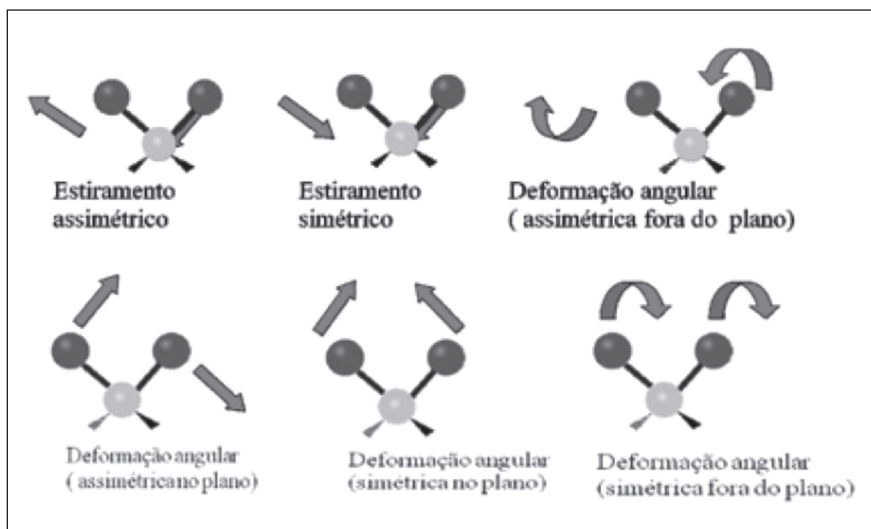


Figura 3: Modos vibracionais de uma molécula de geometria angular (OLIVEIRA, 2008).

bandas de absorção de vibrações moleculares do estado fundamental e sim bandas de combinação e sobretons.

A absorção de radiação no NIR só se verifica se uma molécula receber uma radiação numa frequência com a exata energia para a transição entre dois níveis energéticos ou para a ocorrência de sobretons ou combinações de duas ou mais vibrações. Se uma radiação não possuir essa quantidade exata de energia, não haverá absorção. A absorção da energia radiante irá aumentar a amplitude da vibração molecular, sem, entretanto, alterar sua frequência.

A intensidade de absorção da radiação está relacionada com a magnitude da alteração do momento dipolar e ao grau de anarmonicidade da ligação. Modos de estiramento das ligações CH, CO, CS e CN apresentam bandas de absorção na região do infravermelho médio entre 3000 e 4000 cm^{-1} e produzem bandas de combinação e sobretons na região do NIR, pois suas ligações apresentam grande anarmonicidade com grande variação no momento de dipolo.

As intensidades das bandas de absorção dos sobretons são bem inferiores às bandas fundamentais, conforme apresentado na Figura 4 (SIESLER et al., 2002).

Transição (v_0 a v_n)	Transição	Absorbância Relativa
1	Fundamental	100
2	Primeiro sobretom	9
3	Segundo sobretom	0,3
4	Terceiro sobretom	0,01

Figura 4: Intensidade das bandas dos sobretoms em relação às bandas fundamentais (BUENO, 2004).

A utilização da espectroscopia na região do infravermelho próximo tem sido cada vez mais frequente em vários campos dos mais variados setores tais como, indústria de petróleo e combustível (analisadores em linha para controle de qualidade de derivados), polímeros (espessura de filmes poliméricos), agricultura (teor energético, proteína e açúcar em feno e pastagem), indústria têxtil (controle de qualidade da lã), indústria farmacêutica (monitoramento de biorreações), saúde e medicina (estudos de tecido, identificação de bactérias) entre outros.

A quantidade de informações obtidas por uma análise espectroscópica no infravermelho é tão grande que torna a interpretação dos dados trabalhosa e propensa a erros. À vista disso, o uso de ferramentas quimiométricas permite utilizar todo o conjunto de dados de forma a obter informações úteis e facilmente interpretáveis (SKROBOT, MARIANO & FORTES, 2005).

5. BIODIESEL

A variação do preço do petróleo e a poluição relacionada com sua queima têm dificultado o suprimento energético mundial. A procura por novas fontes mais limpas vem aumentando devido à preocupação ambiental entre os países, reforçando a necessidade de novas formas de produção de energia. Perante isso, retomou-se a ideia de utilizar óleos vegetais como combustíveis, mas os convertendo a biodiesel de forma a eliminar os problemas causados pelos óleos, tais como contaminação do óleo lubrificante, combustão incompleta e formação de depósitos de carbono nos sistemas de injeção.

Segundo a Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), o biodiesel é definido como um composto de alquil ésteres

de ácidos carboxílicos de cadeia longa, derivados de gorduras de origem vegetal ou animal. É produzido pela reação de transesterificação entre triglicerídeos e um álcool de cadeia curta na presença de um catalisador ácido ou básico, sendo mais comumente utilizado hidróxido de sódio (NaOH) ou hidróxido de potássio (KOH) (CORREA, 2012).

Dos alcoóis utilizados em sua fabricação, metanol (CH₄) e etanol (C₂H₈) são os mais comuns. O metanol, por apresentar maior reatividade que o etanol, vem sendo mais aplicado, pois implica menor tempo e temperatura de reação. Entretanto, o etanol é menos tóxico, produz um biodiesel com maior índice de cetano e com lubrificidade mais semelhante a do diesel de petróleo. Além disso, possui uma produção consolidada no Brasil, podendo ser obtido de fontes renováveis (SOUZA, 2014; GUIMARÃES, 2014).

Estequiometricamente, a reação global (Figura 5) necessita de três mols de álcool e um mol de triglicerídeo, mas devido seu caráter reversível excesso de álcool é utilizado, aumentando o rendimento do éster e permitindo a separação da glicerina formada (CORREA, 2012).

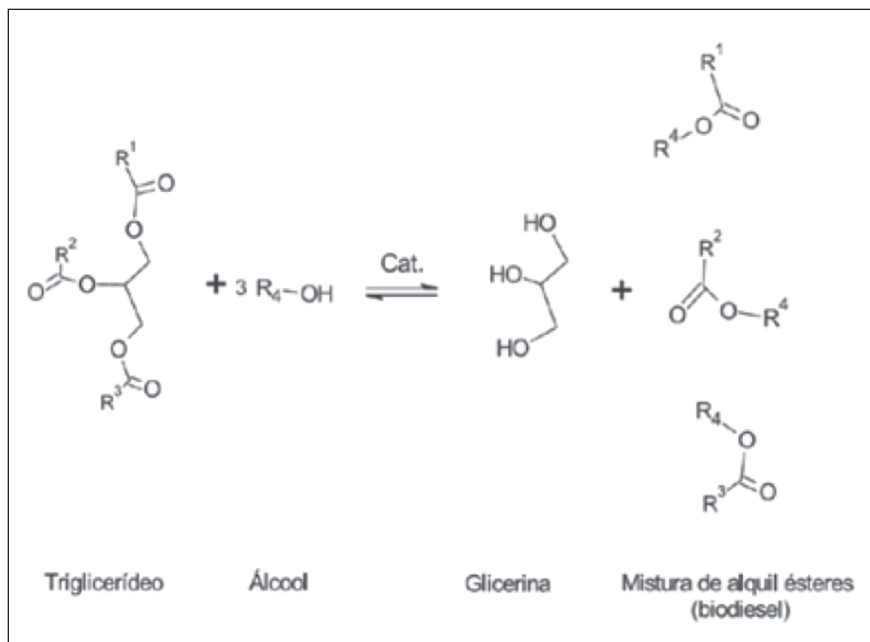


Figura 5: Reação global de transesterificação para produção de biodiesel (GUIMARÃES, 2014).

Matérias-Primas

Diversas matérias-primas podem ser utilizadas na produção do biodiesel, como é possível observar na figura 6, mas é preciso levar em conta diversos fatores antes de escolhê-la para definir a viabilidade do processo. Custo/benefício, composição química e razão óleo/área plantada estão entre eles.

O Brasil é o segundo maior exportador de soja, sendo cultivada em larga escala nas regiões Centro-Oeste, Sul e Sudeste. Por consequência, é a oleaginosa mais utilizada no processo, apresentando maior potencial no desenvolvimento da produção de biodiesel no país (OLIVEIRA, 2008).

Aspectos Econômicos

O uso do biodiesel traz importantes benefícios à sociedade, contribuindo para o desenvolvimento ambiental, econômico e social do Brasil através da redução da emissão de gases durante o processo produtivo e resíduos sólidos na combustão, e no aumento da geração de empregos principalmente na agricultura familiar e pequenos produtores (GUIMARÃES, 2014).

Além disso, para que seja viável a sua fabricação, pleiteando acessibilidade técnica, ele precisa estar de acordo com as condições de proteção ambiental. Sua produção é de suma importância para o Brasil, pois acarreta

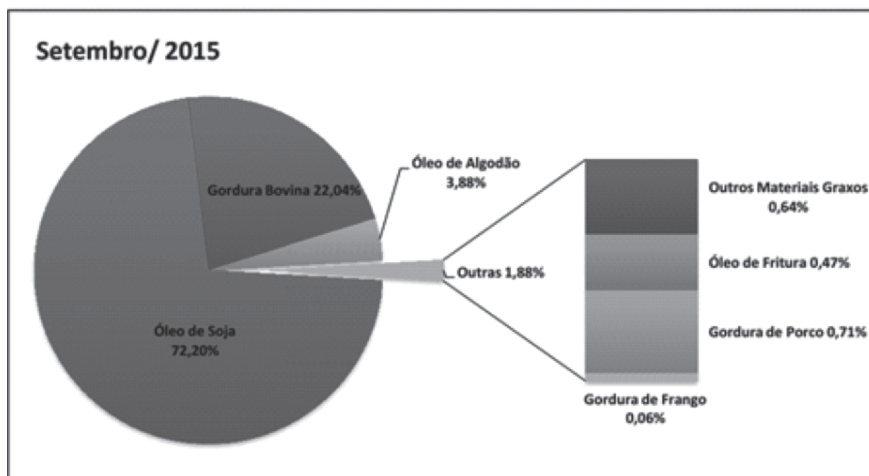


Figura 6: Matérias-primas utilizadas na fabricação de biodiesel no mês de setembro de 2015 (ANP, 2015).

numa diminuição na importação do petróleo e de seus derivados, já que esta ainda é a opção mais utilizada para consumo no transporte rodoviário do país (GUARIEIRO et al., 2008).

De acordo com a Lei nº 13.033 de 24 de Setembro de 2014 da ANP, o Programa Nacional de Produção e Uso do Biodiesel fixou o uso de 7% de biodiesel misturado ao diesel, trazendo uma economia de mais de US\$ 152 milhões e a geração de mais de 153.000 novos empregos (GUIMARÃES, 2014).

Proveniente disso é possível observar na figura abaixo a evolução da produção de biodiesel do ano de 2005 até 2014 de acordo com a ANP, constatando-se uma tendência de continuação do aumento da produção.

Aprimoramento do Processo Produtivo

Mesmo sendo conhecidos os mecanismos de transesterificação, otimizações das variáveis da produção do biodiesel são necessárias para obtenção do melhor rendimento possível. As metodologias utilizadas nessas avaliações requerem grande esforço e tempo, como no caso da cromato-

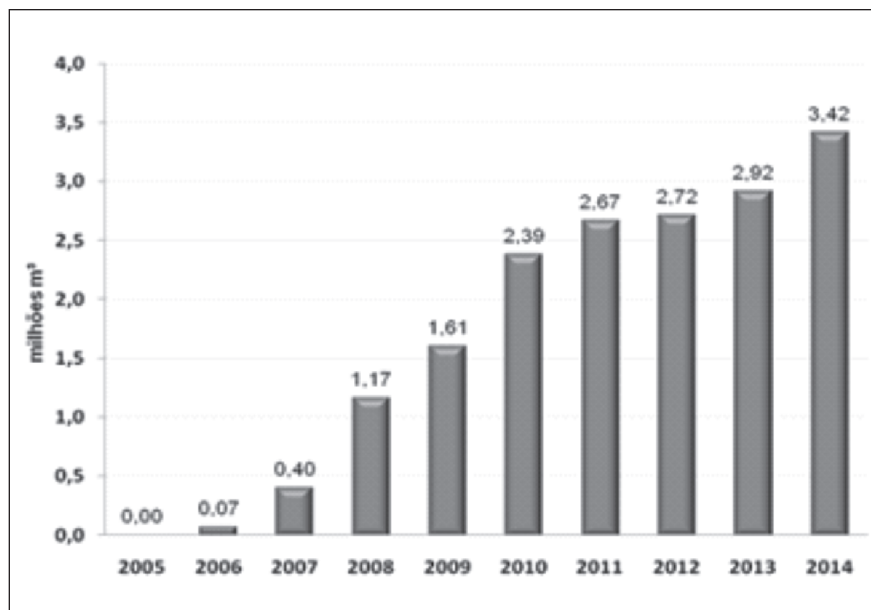


Figura 7: Evolução da produção de biodiesel de 2005 – 2014 (ANP, 2015).

grafia ou da comparação das massas do produto final e da matéria-prima (SKROBOT, MARIANO & FORTES, 2005).

Diante disso, é possível encontrar relatos na literatura científica sobre o uso da espectroscopia no infravermelho durante o processo de alcoólise para determinação da taxa de conversão da reação (GUARIEIRO et al., 2008). As vantagens dessa técnica são tantas, praticidade, caráter não destrutivo, não necessidade de preparo da amostra, que faz com que não seja utilizada apenas durante o processo produtivo do biodiesel, mas também na agricultura, na indústria de alimentos, farmacêutica, e nos mais diversos laboratórios de controle de qualidade (CARNEIRO & BORIN, 2012).

6. ESTUDOS DE CASOS

Avaliação de adulteração de misturas biodiesel-diesel empregando espectroscopia no infravermelho e análise por componentes principais (VIEIRA et al., 2010).

Nesse trabalho, os autores utilizam a espectroscopia no infravermelho médio e a análise de componentes principais a fim de estabelecer uma metodologia para avaliar possíveis adulterações de misturas diesel/biodiesel. A importância desse trabalho se justifica devido à produção de biodiesel contar com isenções fiscais para o incentivo da produção, o que pode motivar à adulteração nas misturas Diesel/Biodiesel.

Os autores utilizaram 81 amostras binárias em diferentes proporções de biodiesel, óleo de girassol bruto, óleo de soja degomado ou óleo residual de fritura em diesel, sendo empregadas uma amostra de interior e outra metropolitana empregando a espectroscopia no infravermelho com transformada de Fourier e reflexão total atenuada (FTIR-ATR).

O espectro IV foi tomado entre as regiões de 650 a 4000 cm^{-1} e foi dividido em 32 regiões distintas, como mostrado pela Figura 8. A avaliação da faixa espectral foi realizada através do algoritmo IPCA (análise de componentes principais por intervalos) e o intervalo 5 do gráfico, que corresponde à faixa de 1074-1178 cm^{-1} , foi escolhido por possuir a maior variância percentual, ou seja, por possuir a melhor descrição da variabilidade das informações.

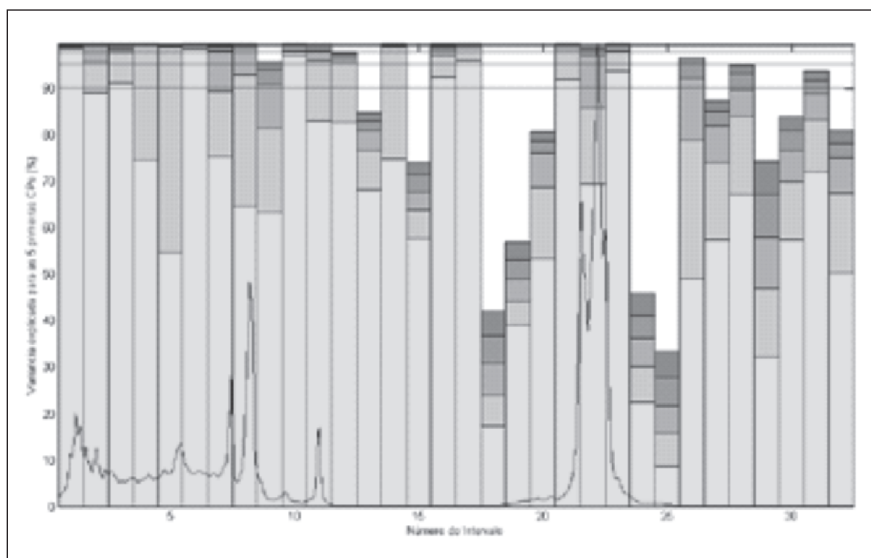


Figura 8: Variância percentual para os espectros FTIR/ATR dividido em 32 intervalos (VIEIRA et al., 2010).

A Figura 1 do referido trabalho mostra o gráfico de escores entre a primeira componente principal (PC1) e a terceira componente principal (PC3).

Como os autores afirmam,

Essa região espectral é capaz de distinguir os distintos componentes das misturas (biodiesel, óleo de girassol bruto, óleo de soja degomado e óleo residual de fritura). Observa-se que a componente principal 1 (com 54,5% da variância) foi responsável pela separação mais pronunciada entre o biodiesel e os demais óleos, tornando esse intervalo apto para a identificação de possíveis blends adulteradas. Já a componente principal 3 (com 0,26% da variância) discriminou o tipo de adulterante empregado (VIEIRA et al., 2010).

A Figura 10 mostra PC1 X PC2 e permite observar que a concentração de biodiesel, óleo de girassol bruto, óleo de soja degomado e óleo

O uso combinado da espectroscopia (FTIR/ATR) e da técnica de análise de componentes principais (PCA) nesse trabalho conseguiu separar grupos formados pelos óleos vegetais (adulterantes) das amostras contendo biodiesel.

Análise exploratória aplicada a espectros de reflexão total atenuada no infravermelho com transformada de Fourier (ATR-FTIR) de blendas de biodiesel/diesel (RUSCHEL et al., 2014).

Nesse trabalho os autores utilizaram ferramentas de análise de agrupamento hierárquico, do inglês, Hierarchical Cluster Analysis (HCA) e de análise de componentes principais, do inglês, Principal Component Analysis (PCA) com o objetivo de classificar blendas preparadas a partir de óleo diesel metropolitano (S500) e de vários tipos de biodiesel metílico e etílico, a partir do tratamento de dados obtidos por espectroscopia por refletância total atenuada horizontal no infravermelho com transformada de Fourier (HATR-FTIR).

A Figura 11 aponta o conjunto de espectros IV das diferentes misturas. A região entre 650 e 1800 cm^{-1} , apresenta as principais diferenças espectrais entre as misturas e devido a isso, essa faixa espectral é a que foi tratada pelos métodos multivariados.

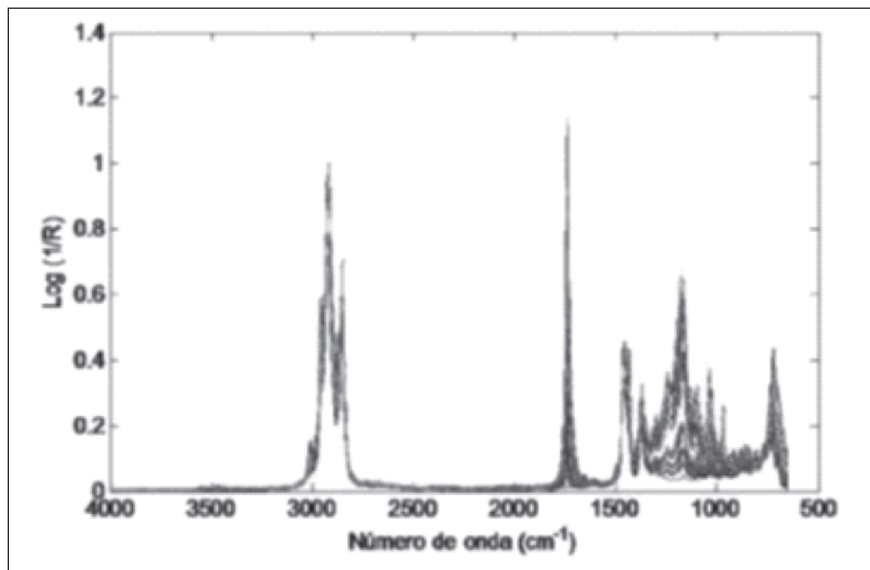


Figura 11: Conjunto de espectros de infravermelho de todas as amostras (RUSCHEL et al., 2014).

O gráfico da Figura 12 retrata os escores de PC1 em relação à PC2. A distribuição das amostras se apresenta em função do aumento gradativo do percentual de biodiesel nas misturas, ou seja, é possível identificar esse aumento percentual ao longo da PC1 que responde por 89,07% da variância do sistema. Vale frisar, que os autores identificaram as misturas como sendo B0, aquela que possui diesel puro, e B100 a que só tem biodiesel presente.

Outro resultado importante a partir da análise da componente principal 2 (PC2), que responde por 8,59% da variância, é que pode ser diferenciado o diesel metílico do etílico, pois o metílico apresenta predominantemente, valores positivos para essa PC enquanto o etílico têm valores negativos. Além disso, as amostras constituídas com 50% de cada tipo de biodiesel se localizam próximo ao valor zero da PC, quer dizer, exatamente na região intermediária entre os dois grupos.

A Figura 12 (Figura 5b do trabalho) aponta o resultado obtido a partir da análise da terceira componente principal (PC3) que responde por 2,14% da variância do sistema. A partir do eixo da PC3 é possível perceber a separação das amostras das blendas produzidas com biodiesel a partir da gordura vegetal hidrogenada daquelas em que o biodiesel foi oriundo de fontes de outros óleos vegetais.

Conforme os autores afirmam:

As blendas de biodiesel de gordura possuem valores positivos para a PC3 e as blendas de biodiesel provenientes dos óleos apresentaram valores negativos para a PC3. De acordo com o gráfico de escores, Figura 5b, os pesos positivos para a PC3 referem-se às blendas provenientes de biodiesel produzido a partir de gordura vegetal hidrogenada, que é saturada, e os pesos negativos para a PC3 são referentes às blendas de biodiesel dos óleos de soja e de fritura, que são mais insaturados (RUSCHEL et al., 2014).

Esse trabalho faz uma análise muito interessante relacionando, diretamente, as três primeiras componentes principais com o espectro na região de interesse. Isso é mostrado na Figura 13 (Figura 6a do trabalho). É possível observar que para a primeira componente (PC1) as bandas referentes ao diesel têm valores positivos enquanto àquelas relativas ao biodiesel têm valores negativos. Isto pode ser identificado na análise das bandas relativas

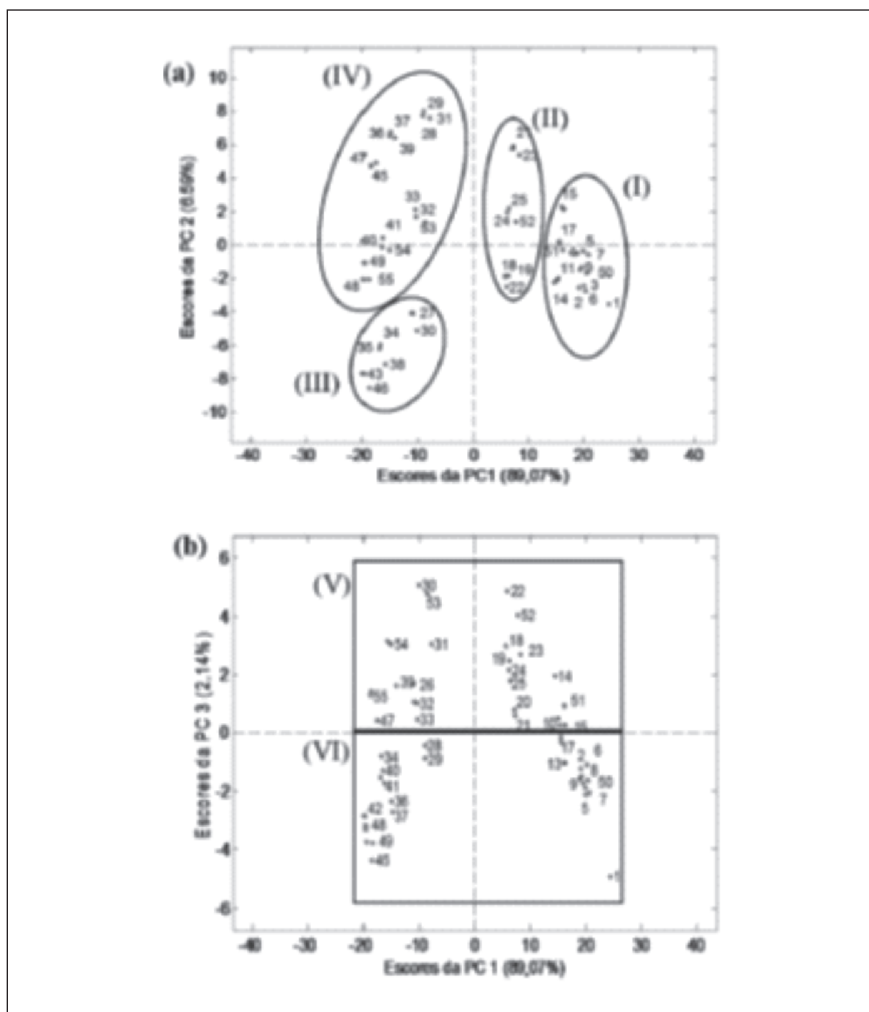


Figura 12: Gráficos de escores da PCA (RUSCHEL et al., 2014).

às deformações angulares simétricas e antissimétricas dos grupos metila (-CH₃) e metileno (-CH₂) na região entre 1300 e 1500 cm⁻¹ para o diesel enquanto que para o biodiesel as bandas do estiramento (C=O) de éster alifático em 1741 cm⁻¹ e as bandas do estiramento (C-O), na normalmente complicada região para atribuição de bandas entre 1000 e 1300 cm⁻¹, são as mais importantes para o biodiesel.

O gráfico de pesos para PC2 (Figura 13 - 6b do trabalho) a partir da análise das bandas do estiramento (C=O) consegue diferenciar o biodiesel metílico do etílico, pois no caso do metílico este grupo está adjacente a um grupo metila que possui valores positivos para a PC2 enquanto para o etílico, o grupo adjacente é uma etila que possui valores negativos em PC2.

O último gráfico (Figura 13 - 6c do trabalho) explica o já exposto que é a diferenciação da origem do biodiesel em relação ao seu óleo antes da reação de transesterificação. A banda em 700 cm⁻¹ é relativa à deformação angular no plano tipo balanço “rocking” do grupo (-CH₂) presente nos óleos, que possui peso negativo para a PC3 e como os autores afirmam:

Além disso, observam-se também no espectro de infravermelho do biodiesel metílico de gordura vegetal hidrogenada que estão presentes as mesmas bandas com valores positivos para a PC3, conforme o gráfico de pesos da Figura 6c, indicando que assim como os pesos, os escores positivos da PC3 correspondem à matéria-prima empregada na produção do biodiesel (RUSCHEL et al., 2014).

Para finalizar a análise desse estudo de caso pode-se concluir que a utilização da espectroscopia no infravermelho associada com uma análise multivariada de dados através das técnicas HCA e PCA foi capaz de classificar blendas de óleo diesel e biodiesel, onde possibilitaram diferenciar essas blendas a partir dos seus espectros de infravermelho, mesmo essas misturas sendo compostas por vários tipos de biodiesel provenientes de três matérias-primas distintas: óleo de soja, óleo residual de fritura e gordura vegetal hidrogenada, além de duas rotas alcoólicas: a metílica e a etílica.

Desenvolvimento de modelos de calibração multivariada para determinação de índice de acidez e viscosidade em biodiesel empregando espectroscopia no infravermelho médio (CARNEIRO & BORIN, 2012).

O índice de acidez e a viscosidade são parâmetros físico-químicos importantes para o controle de qualidade do biodiesel. O índice de acidez pode indicar a presença de subprodutos da reação de transesterificação durante o processo de produção do biodiesel entre outras coisas que podem prejudicar o bom desempenho dos motores. A viscosidade é um parâmetro

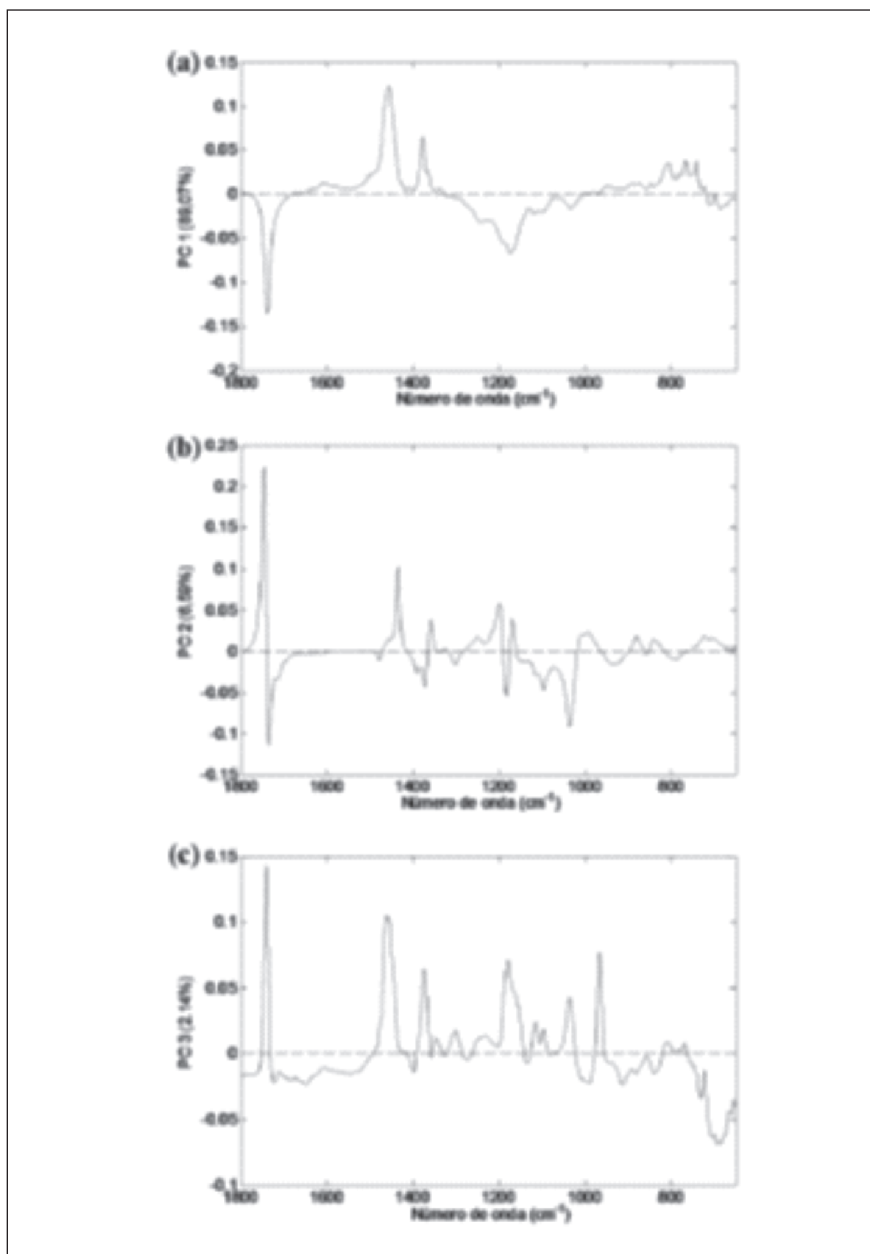


Figura 13: Gráfico de pesos da PCA (RUSCHEL et al., 2014).

importantíssimo para o controle do petróleo e derivados e também o é para o biodiesel. Viscosidade fora das especificações pode levar a formação de depósitos no motor e entupimento dos bicos injetores, porém, ambas as medidas consomem bastante tempo, consumo de reagentes, preparação, padronização de soluções, entre outros.

Os autores então empregaram a espectroscopia no infravermelho com refletância total atenuada associada à análise de métodos multivariados (PCA e PLS) a fim de desenvolver um modelo de calibração multivariada para a determinação desses dois parâmetros. Os espectros no infravermelho médio de 12 amostras de óleos vegetais (canola, girassol e soja) e 24 amostras de biodiesel produzidos a partir destes óleos vegetais foram tomados e seus dados tratados pelos métodos multivariados citados acima.

A Figura 14 mostra os espectros no IV médio das 36 amostras na região entre 600 a 2000 cm^{-1} .

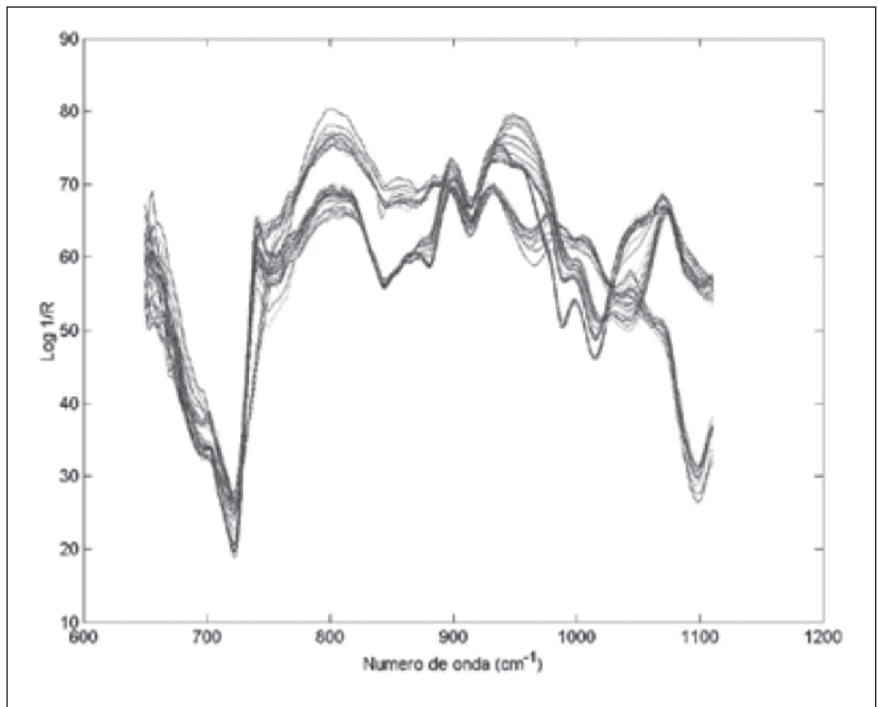


Figura 14: Espectros no infravermelho médio de óleos vegetais e biodiesel (CARNEIRO & BORIN, 2012).

A Figura 15 aponta o resultado do tratamento dos dados pela PCA. Nesse gráfico, são apresentadas as duas primeiras componentes principais (PC1 versus PC2) que respondem por mais de 96% da variância do sistema. Os autores afirmam então que “é possível observar uma nítida separação das amostras de biodiesel e óleos vegetais confirmando a produção do biodiesel e também a diferença espectral entre amostras dos biodiesel oriundas de diferentes óleos”.

Os vários grupos formados representam: G1 – óleo de algodão; G2 – óleo de canola; G3 – óleo de girassol; G4 – óleo de soja; G5 – biodiesel de algodão; G6 – biodiesel de canola; G7- biodiesel de girassol; G8 – biodiesel de soja.

Um modelo PLS para quantificação da viscosidade do conjunto de amostras de biodiesel foi também desenvolvido com cinco variáveis latentes (98.62%) para a viscosidade. Para o desenvolvimento desse modelo 16 amostras foram escolhidas aleatoriamente para comporem o conjunto de calibração e 8 amostras o conjunto de previsão.

O modelo se mostrou adequado, apresentando erros relativos para o conjunto de previsão de até 14%. Este valor de erro é aceitável por ser um método multivariado.

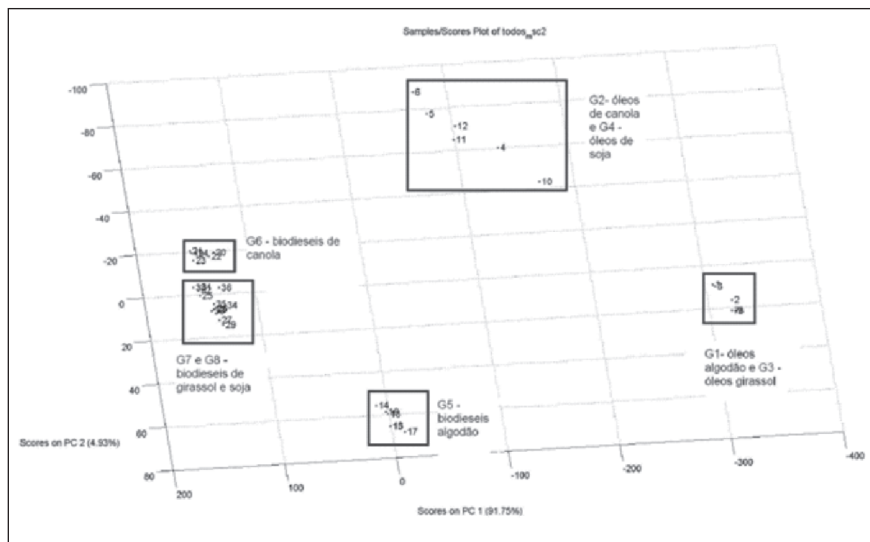


Figura 15: Análise das componentes principais das 36 amostras de óleos vegetais e biodiesel (CARNEIRO & BORIN, 2012).

A Figura 16 representa a tabela com os valores teóricos e previstos pelo modelo.

Para o desenvolvimento do modelo PLS para índice de acidez foram utilizadas um conjunto de 21 amostras sendo 14 para o conjunto de calibração e 7 para previsão. No entanto, o modelo PLS para a calibração de um método para a determinação do índice de acidez não foi adequado apresentando erros relativos de até 50%.

Por fim, os autores afirmam que foi possível observar uma nítida separação das amostras de óleos vegetais e do biodiesel com a utilização do PCA e consideram adequado um modelo de calibração multivariada, baseado no PLS, para a quantificação da viscosidade, mas não para a quantificação do índice de acidez baseados nos erros relativos obtidos.

Classificação periódica: Um exemplo didático para ensinar análise de componentes principais (LYRA et al., 2010).

Tipo de óleo vegetal	VISCOSIDADE TEORICA [mPa.s]	VISCOSIDADE PREVISTA [mPa.s]	ERRO RELATIVO [%]
ALG	3,87	3,81	1,55
ALG	3,86	3,87	-0,26
CAN	3,42	2,93	14,03
CAN	2,94	3,06	-4,08
GIR	3,26	3,10	4,90
GIR	3,23	3,25	-0,62
SOJ	3,22	3,42	-5,90
SOJ	3,49	3,28	6,01

Figura 16: Valores da viscosidade teórica e prevista pelo modelo PLS com seus respectivos erros relativos (CARNEIRO & BORIN, 2012).

Este último estudo de caso tem por objetivo explicar passo a passo através de um exemplo simples, porém didático, a aplicação da análise de componentes principais na análise dos elementos tradicionalmente conhecidos como semi-metais, na Tabela Periódica, como metais ou ametais, de acordo com a recomendação da IUPAC.

Apesar de diferente dos demais estudos de caso, onde a ênfase foi a aplicação dos métodos multivariados na análise de biodiesel a partir de dados de espectroscopia no infravermelho, a análise desse trabalho é uma oportunidade de mostrar o “como fazer” uma análise de componentes principais a partir de dados que podem ser obtidos de Handbooks, por exemplo.

A partir de valores tabelados das seguintes propriedades: primeira energia de ionização (EI), raio atômico (RA), afinidade eletrônica (AE), eletronegatividade de Pauling (EN), densidade (D), calor específico (CE), entropia padrão (So) e condutividade térmica (CT) para 88 elementos naturais da Tabela Periódica foi realizada uma análise de componentes principais (PCA) com o objetivo de verificar se essa análise exploratória é capaz de agrupar os elementos, antes conhecidos como semi-metais, na atual classificação proposta pela IUPAC em que germânio, antimônio e polônio são metais, enquanto boro, silício, arsênio e telúrio são ametais.

PASSO 1 – Elaboração da matriz de variáveis.

A matriz de variáveis abaixo foi montada a partir dos dados tabelados, onde os elementos da tabela (objetos) são dispostos nas linhas e as propriedades (variáveis) nas colunas.

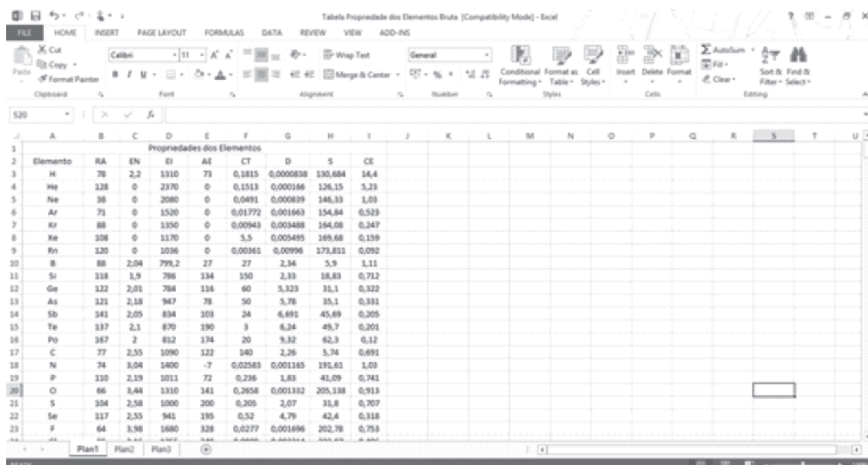


Figura 17: Matriz do Excel com os dados brutos.

PASSO 2 – Pré-processamento.

Devido à diversidade dos valores das propriedades, os dados foram pré-processados por auto-escalonamento que consiste em centrar a matriz na média dos valores, subtraindo-se o valor de cada elemento da matriz pela média de cada variável, e dividindo pelo desvio-padrão (normalização).

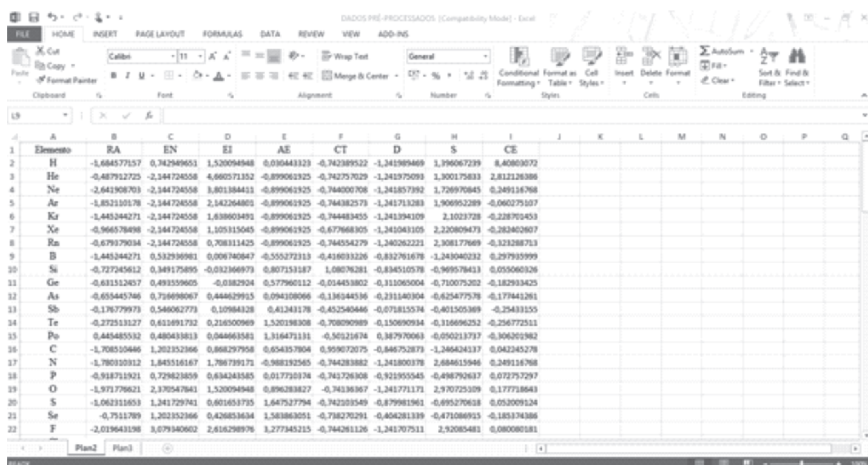


Figura 18: Matriz do Excel com os dados pré-processados.

PASSO 3 – Carregar a matriz de dados pré-processados para o programa Statistica®.

O programa permite carregar arquivos diretamente do Excel como mostrado abaixo:

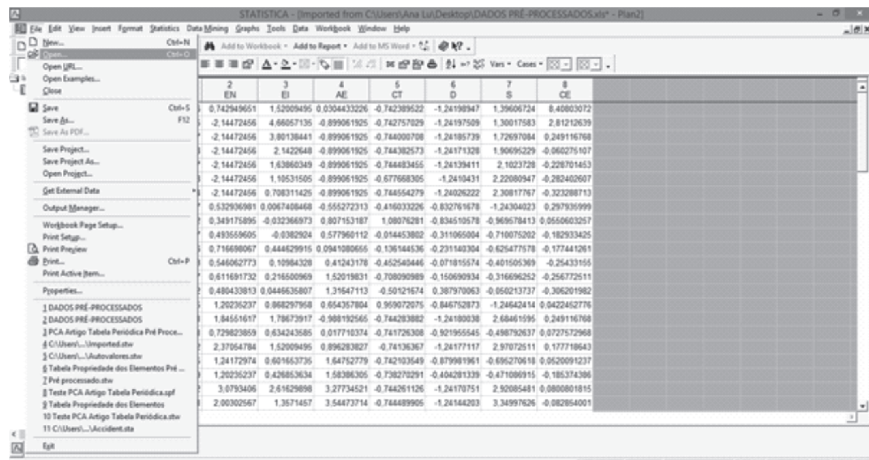


Figura 19: Abertura da matriz de dados pré-processados do Excel.

E assim, a matriz de dados pré-processados pode ser carregada.

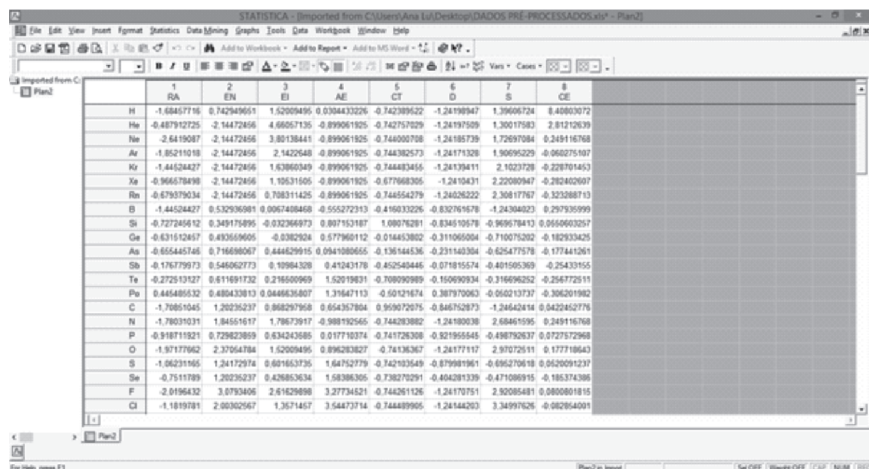


Figura 20: Matriz pré-processada carregada.

PASSO 4 – Obter as componentes principais a partir do software.

Com a matriz das variáveis carregada, o software permite através de um procedimento simples iniciar o tratamento dos dados por PCA.

Selecionando: *Statistics – Multivariate Exploratory Techniques – Principal Components & Classification Analysis*, conforme a figura abaixo:

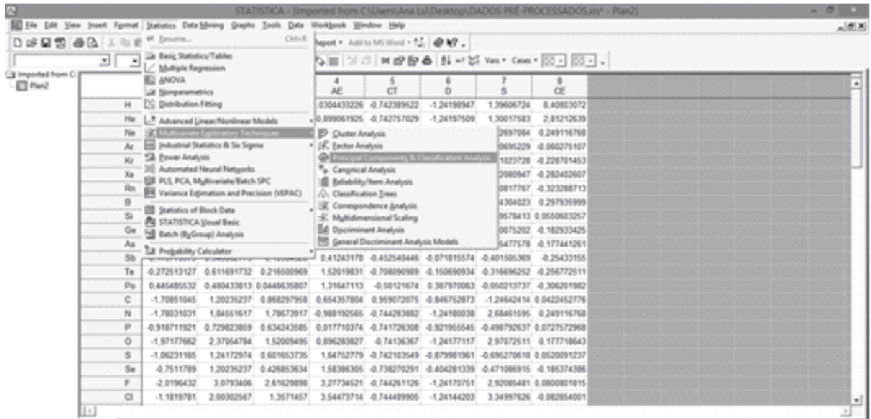


Figura 21: Carregando o programa para a Análise de Componentes Principais.

PASSO 5 – Seleção das variáveis.

No caso, temos 8 variáveis selecionadas que são as propriedades periódicas em estudo.

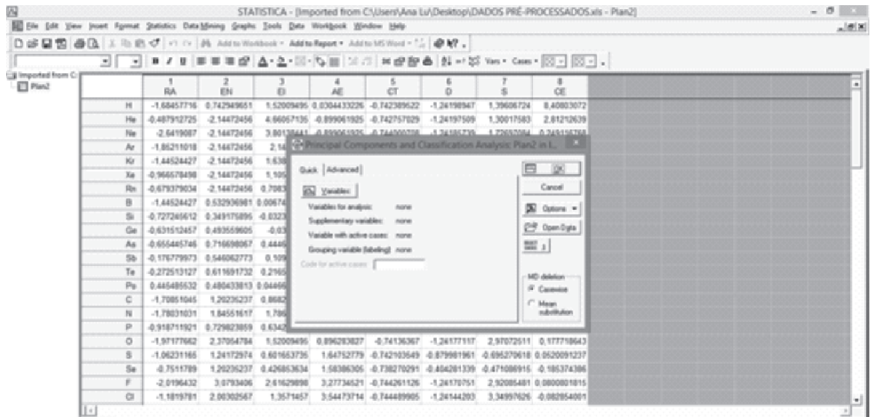


Figura 22: Caixa de seleção de variáveis.

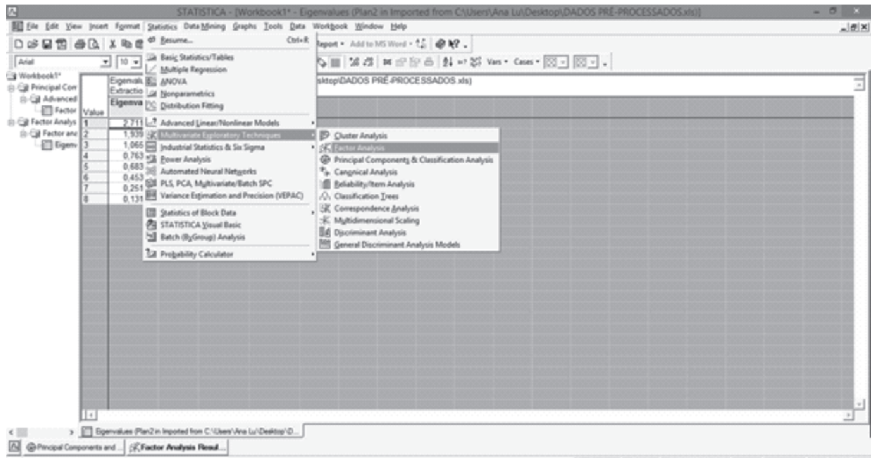


Figura 23: Variáveis selecionadas.

PASSO 6 – Análise da variância explicada.

Para determinar a variância explicada de cada componente, necessita-se calcular os autovalores através do procedimento abaixo:

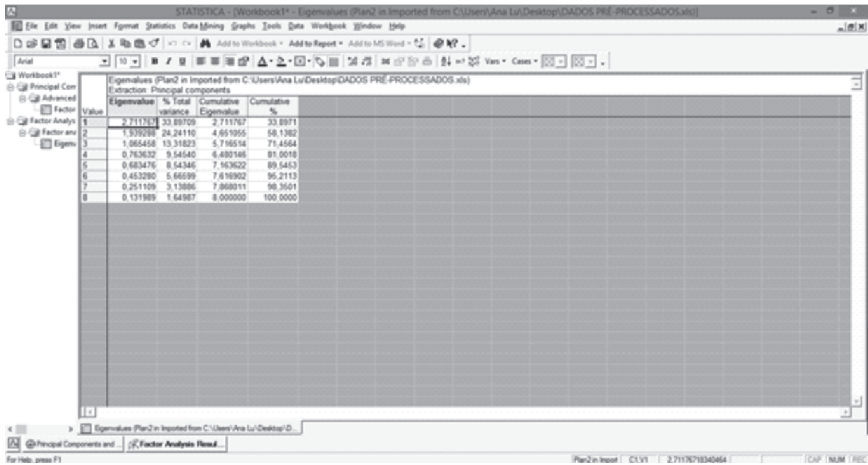


Figura 24: Carregando o programa para o cálculo dos autovalores.

Por esse procedimento as equações matriciais do PCA são resolvidas e os autovalores para cada componente são calculados. Os resultados podem ser observados na Figura 25 abaixo:

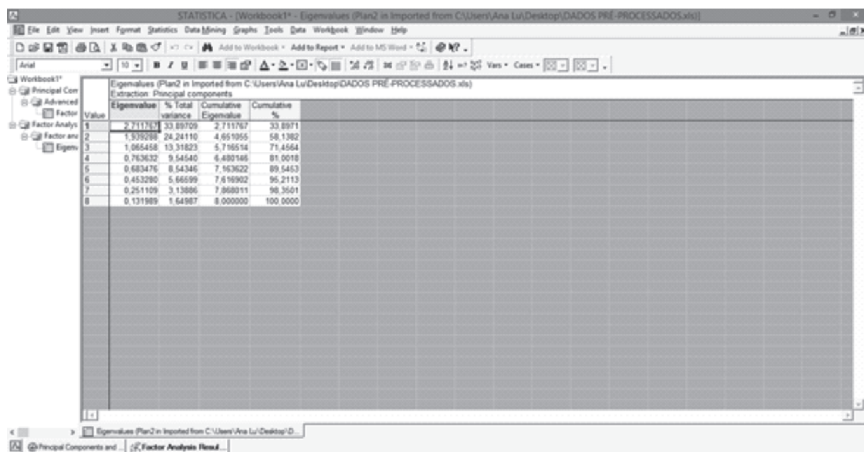


Figura 25: Autovalores e percentual da variância explicada para cada componente.

Pode ser observado que a componente 1 (PC1) responde por 33,89% da variância do sistema enquanto a PC2 responde por aproximadamente 24%. Podem ser observadas também, pela variância acumulada, que as três primeiras componentes respondem por 71% da variância do sistema e pela análise dos autovalores são as três que possuem autovalores acima de 1. Então essas três componentes são suficientes para uma boa descrição dos dados.

PASSO 7 – Encontrando as componentes principais.

Após o passo 5, a janela da Figura 26 abaixo aparecerá.

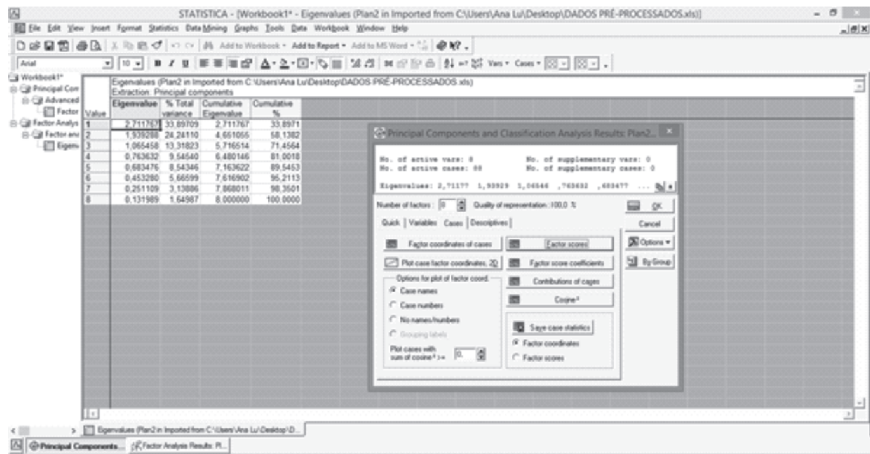


Figura 26: Encontrando as componentes principais.

Selecionando Variables/Factor & variable correlations, as componentes principais são calculadas como na Figura 27 abaixo:

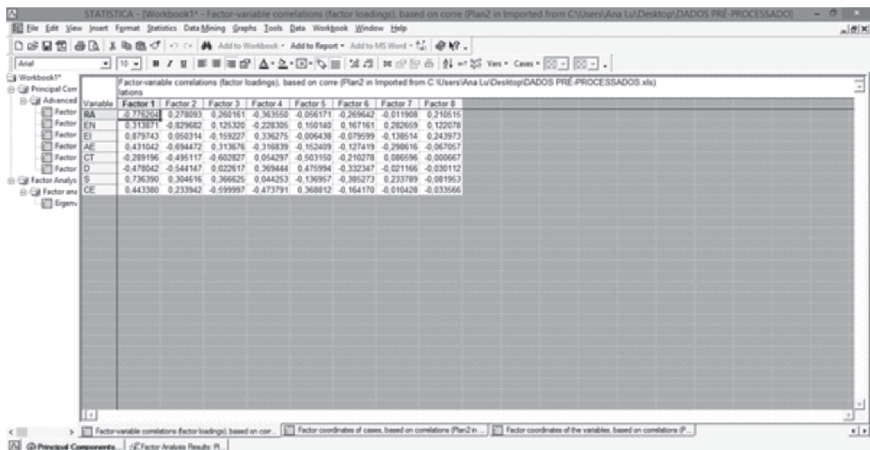


Figura 27: Componentes Principais.

Cada componente (fator) dá o peso correspondente de cada variável original.

PASSO 8 – Plotando e analisando os resultados nos gráficos de escores entre as PCs.

A partir da janela mostrada na Figura 28, pode ser selecionado Plot case factor coordinates 2D, que mostrará a relação entre as diversas componentes em um gráfico de escores.

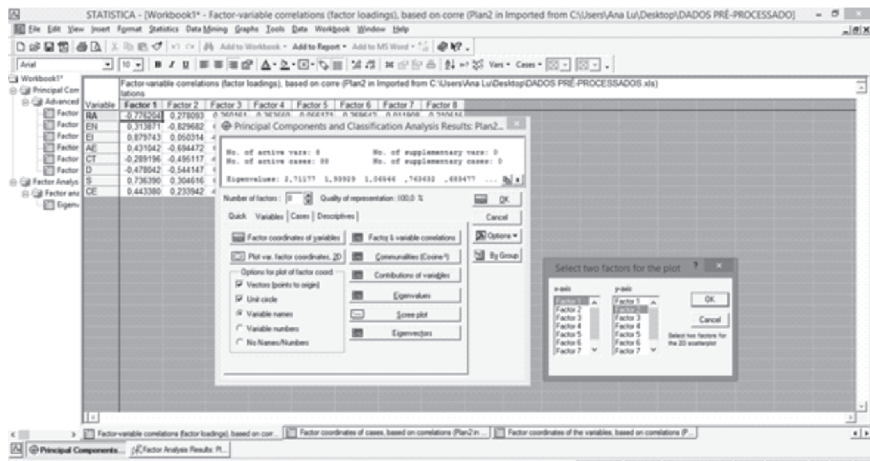


Figura 28: Plotando o gráfico de escores entre as componentes.

A figura 29 mostra a distribuição dos 88 elementos no plano das duas coordenadas principais que respondem por aproximadamente 58% da variância do sistema.

A distribuição dos elementos químicos no plano das duas primeiras componentes, PC1 e PC2, é retratada nas figuras 29 e 30. Pode ser observado certo padrão de distribuição dos elementos de acordo com a sua classificação na Tabela Periódica. Nota-se que os elementos gasosos estão dispostos mais acima e a direita, em valores positivos da PC1, enquanto que os metais estão concentrados à esquerda para valores mais negativos de PC1. Os ametais aparecem na região mais a direita e abaixo do gráfico.

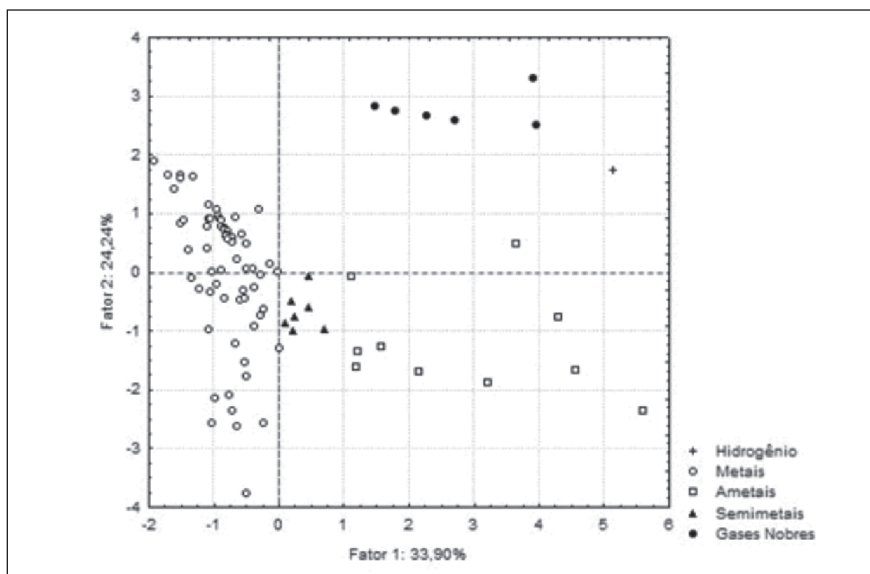


Figura 29: Gráfico dos escores de PC1 e PC2 para os elementos químicos.

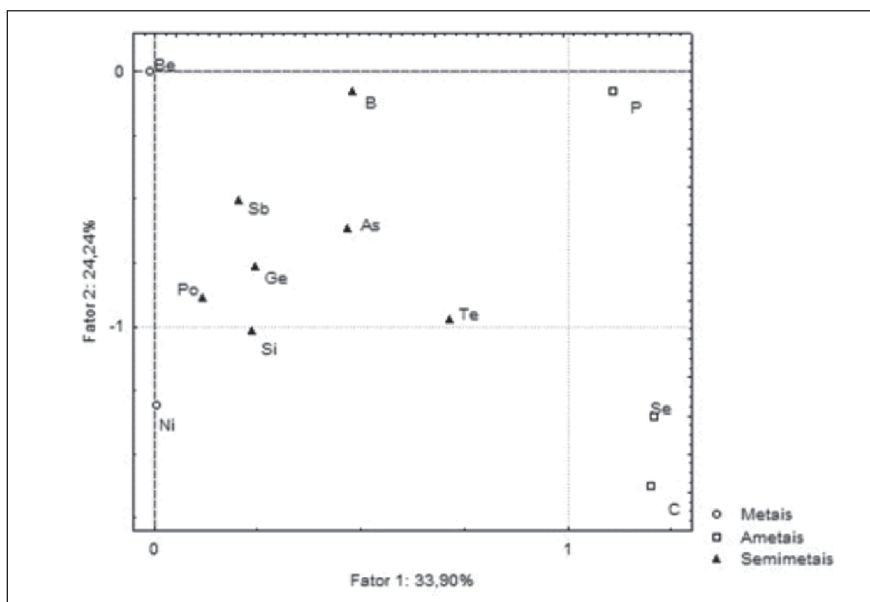


Figura 30: Gráfico dos escores de PC1 e PC2 para os elementos químicos. Escala modificada.

No mesmo gráfico com a escala ampliada é possível notar a posição dos elementos que eram considerados semimetais. O Boro aparece próximo aos elementos classificados como ametais, assim como o Arsênio e o Telúrio, o que condiz com a classificação atual. O Polônio, mais a esquerda do gráfico, encontra-se na região dos metais podendo ser classificado como um por estar próximo destes.

Apesar de sabermos que o Germânio é um metal e o Silício é um ametal, as propriedades utilizadas nesta análise não foram suficientes para obtermos um padrão que concorde com as atribuições da IUPAC. Ainda assim, o resultado obtido foi satisfatório e condizente com o presente artigo, havendo apenas os sinais das PC's invertidos, o que não inviabilizou a análise nem sua interpretação.

7. CONSIDERAÇÕES FINAIS

A partir da análise de uma seleção de trabalhos publicados em revistas ou apresentados em congressos em que métodos multivariados, em especial a Análise de Componentes Principais (PCA), são aplicados para tratar dados de espectros de infravermelho a fim de analisar amostras de diesel, biodiesel e seus óleos constituintes, foi possível compreender melhor a importância e o poder que esses tipos de técnicas têm na análise de amostras de múltiplos componentes em que muitas propriedades necessitam ser consideradas para uma boa compreensão do seu comportamento físico-químico. Em todos os casos, foi possível observar, o quanto uma análise exploratória como a PCA pode auxiliar na identificação de outliers, e na separação de amostras em agrupamentos que pode, entre outras coisas, identificar se, bem calibrada, algumas das misturas estão fora do padrão, o que pode servir como um método de identificação de amostras adulteradas sendo de suma importância para o controle de qualidade de combustíveis e de outros derivados de petróleo. Por fim, foi possível aplicar a técnica com o uso de um software comercial bastante popular nos meios acadêmicos e industriais, explicando passo a passo, de uma forma bem didática, como fazer essa análise e mostrar, a partir de um exemplo simples, mas que pode ser utilizado até mesmo no ensino de quimiometria, a versatilidade e eficiência que esta técnica de tratamento de dados multivariados tem na exploração de sistemas complexos.

REFERÊNCIAS

ANP, 2015. Disponível em: <<http://www.anp.gov.br>>. Acesso em: 04 dez. 2015.

BARROS NETO, B. de; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. *Química Nova*, vol. 29, n. 6, p. 1401-1406, 2006.

BUENO, A. F. Caracterização de petróleo por espectroscopia no infravermelho próximo. 2004. 122 f. Dissertação (Mestrado em Química) – Universidade Estadual de Campinas, Campinas, 2004.

CARAUTA, A. N. M. Espectro vibracional do nitrato de diamin uranila. 1996. 72 f. Dissertação (Mestrado em Química Inorgânica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1996.

CARNEIRO, G. P.; BORIN, A. Desenvolvimento de modelos de calibração multivariada para determinação de índice de acidez e viscosidade em biodiesel empregando espectroscopia no infravermelho médio. In: Encontro de Iniciação em Desenvolvimento Tecnológico e Inovação, 2., 2012, Campinas, SP. Anais... Campinas, MG: PUC – Campinas, 2012.

CORREA, C. Avaliação do uso de espectroscopia no infravermelho para controle e garantia da qualidade da blenda biodiesel/diesel usando cartas de controle multivariadas. 2012. 47 f. Trabalho de Conclusão de Curso (Graduação em Química) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: Explorando procedimentos quimiométricos para tratamento de dados analíticos. *Química Nova*, vol. 30, n. 2, p. 481-487, 2007.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: Calibração multivariada, um tutorial. *Química Nova*, vol. 22, n. 5, p. 724-731, 1999.

FRAQUELLI, H. A. O experimento de Herschel. 2015. Disponível em: <http://www.if.ufrgs.br/~riffel/notas_aula/ensino_astro/roteiros/Roteiro_Experimento_Herschel.htm>. Acesso em: 06 dez. 2015.

GUARIEIRO, L. L. N.; PINTO, A. C.; AGUIAR, P. F. de; RIBEIRO, N. M. Metodologia analítica para quantificar o teor de biodiesel na mistura biodiesel:diesel utilizando espectroscopia na região do infravermelho. Química Nova, vol. 31, n. 2, p. 421-426, 2008.

GUIMARÃES, E. Aplicação de espectroscopia no infravermelho médio e calibração multivariada para quantificação de adulterantes em biodieseis de óleo de soja e residual e suas misturas com diesel. 2014. 86 f. Dissertação (Mestrado em Química) – Universidade Federal de Uberlândia, Uberlândia, 2014.

HENRIQUES, C. B. Caracterização prévia de petróleo com vistas à otimização de processos. 2011. 284 f. Tese (Doutorado em Engenharia Química) – Universidade Estadual de Campinas, Campinas, 2011.

LUZ, E. R. da. Predição de propriedades de gasolinas usando espectroscopia FTIR e regressão por mínimos quadrados parciais. 2003. 109f. Dissertação (Mestrado em Química) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2003.

LYRA, W. da S.; SILVA, E. C. da; ARAÚJO, M. C. U. de; FRAGOSO, W. D. Classificação periódica: Um exemplo didático para ensinar análise de componentes principais. Química Nova, vol. 33, n. 7, p. 1594-1597, 2010.

OLIVEIRA, I. K. de. Aplicação de quimiometria e espectroscopia no infravermelho no controle de qualidade de biodiesel e mistura biodiesel/diesel. 2008. 110 f. Dissertação (Mestrado em Química) – Universidade Estadual de Campinas, Campinas, 2008.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. J. Braz. Chem. Soc., vol. 14, n. 2, p. 198-219, 2003.

RUSCHEL, C. F. C.; HUANG, C. T.; SAMIOS, D.; FERRÃO, M. F. Análise exploratória aplicada a espectros de reflexão total atenuada no infravermelho com transformada de Fourier (ATR-FTIR) de blendas de biodiesel/diesel. Química Nova, vol.37, n. 5, p. 810-815, 2014.

SALA, O. Fundamentos da espectroscopia Raman e no infravermelho. 1 ed. São Paulo: UNESP, 1995. 223 p.

SANTANA, G. P. Quimiometria com R. [S.I.], 2013. 99 p.

SIESLER, H. W.; OZAKI Y.; KAWATA, S.; HEISE, H. M. Near-infrared spectroscopy: Principles, instruments, applications. *Journal of Chemometrics*, vol. 16, n. 12, p. 636-638, 2002.

SKROBOT, L. V.; MARIANO, L. I.; FORTES, I. C. P. Estudo da utilização de espectrometria no infravermelho e quimiometria na otimização da obtenção de biodiesel de óleo residual. In: Congresso Brasileiro de Plantas Oleaginosas, Óleos, Gorduras e Biodiesel, 2., 2005, Varginha, MG. Anais... Varginha, MG: UFLA, 2008. p. 783-787.

SOUZA, L. M. de. Uso de espectroscopia no infravermelho médio, calibração multivariada e seleção de variáveis na quantificação de adulterantes em biodieseis e suas misturas com diesel. 2014. 104 f. Dissertação (Mestrado em Química) – Universidade Federal de Uberlândia, Uberlândia, 2014.

TEÓFILO, R. F. Métodos quimiométricos: Uma visão geral – Conceitos básicos de quimiometria. Universidade Federal de Viçosa, Viçosa, vol. 1, 2013. 112 p.

VANDEGINSTE, B. G. M., *Top. Curr. Chem.* 1987, 141, 1.

VIEIRA, M. de S.; FRANCESQUETT, J. C.; FACHINI, D.; GERBASE, A. E.; FERRÃO, M. F. Avaliação de adulteração de misturas biodiesel diesel empregando espectroscopia no infravermelho e análise por componentes principais. In: Encontro Nacional de Engenharia de Produção, 30., 2010, São Carlos, SP. Anais... São Carlos, SP: ABEPRO, 2010. p. 1-12.