

Relatório Pré-Final – F 590A – Iniciação Científica I

Aluno: Raphael Alves Silva **RA:** 103898



Email: alvesraphael x(arroba)x gmail.com

Professor/Coordenador: Prof. Dr. Jose J. Lunazzi

Orientador: Dr. Carlos E. Driemeier

1. Resumo

O relatório aqui apresentado descreve as atividades relativas ao projeto de pesquisa intitulado “Componentes desordenados de lignocelulose: caracterização por espectroscopia de infravermelho associada à dinâmica de troca isotópica”, realizado durante o estágio do aluno Raphael Alves Silva, e desenvolvido ao longo do primeiro semestre de 2014. O projeto vigorou sob a orientação do Dr. Carlos Eduardo Driemeier, e é vinculado ao Laboratório Nacional de Ciência e Tecnologia do Bioetanol (CTBE) do Centro Nacional de Pesquisa em Energia e Materiais (CNPEM).

Foram analisados espectros de radiação infravermelha para amostras de bagaço de cana-de-açúcar, avicel e línter de algodão. Os resultados foram interpretados utilizando a teoria de PCA (Principal Component Analysis), buscando-se a melhor visualização da dinâmica de trocas isotópicas ao longo do tempo. Observar e compreender essa dinâmica química revela-se um importante instrumento para melhorar o aproveitamento da energia que pode ser gerada pelo bagaço residual da cana-de-açúcar, bem como demais resíduos biovegetais.

2. Introdução

2.1 Equipamentos

- **Espectrofotômetro FT-IR:** O **espectrofotômetro** é um instrumento de análise, amplamente utilizado em laboratórios de pesquisa, capaz de medir e comparar a quantidade de luz (radiação eletromagnética) absorvida, transmitida ou refletida por uma determinada amostra, seja ela solução, sólido transparente ou sólido opaco. Existem duas classes de espectrofotômetros: o de feixe simples e o de duplo feixe e eles são constituídos,

essencialmente, por cinco componentes principais: as fontes de radiação, o monocromador, o porta-amostras, os detectores e o indicador de sinal [1].

No nosso caso, utilizamos um espectrofotômetro IR (*infrared*), equipamento que trabalha com radiação infravermelha. Especificamente, trata-se de radiação de infravermelho média, de número de onda k na faixa de 4000 cm^{-1} a 400 cm^{-1} . A figura 1 mostra o equipamento utilizado em laboratório. Trata-se de um modelo Frontier da Perkin Elmer, associado ao módulo ATR (*Attenuated Total Reflectance*) universal com polarizador e ao software Spectrum TimeBase, no qual são adquiridos e manipulados os dados das séries temporais dos espectros.

A técnica utilizada para a obtenção dos dados espectrais é conhecida como espectroscopia FTIR (*Fourier transform infrared spectroscopy*), na qual uma transformada de Fourier é requerida para converter os dados brutos e gerar o espectro de radiação propriamente dito.



Figura 1 – Espectrofotômetro Frontier da Perkin Elmer, utilizado em laboratório.

Gerador de vapor de H₂O e D₂O: trata-se de um dispositivo utilizado para produzir vapor aplicando energia térmica à água. Neste projeto, os geradores foram responsáveis por produzir atmosferas de umidade, de forma controlada, de maneira a propiciar trocas isotópicas entre o ambiente e as amostras analisadas (em outras palavras, as amostras absorvem a água do ambiente controlado). Essas interações químicas ocorreram utilizando água e água deuterada (D₂O).

Os geradores de vapores utilizados são o modelo GenRH-A [2], produzidos pela empresa Surface Measurement Systems. Ver figura 2.



Figura 2 – Gerador de vapor GenRH-A.

2.2 Montagem experimental

A figura 3 ilustra o aparato instrumental montado em laboratório. Uma espécie de saia (linha azul tracejada na figura) no interior da qual se injeta o fluxo de vapor é suficiente para criar o ambiente que expõe a amostra à umidade controlada. Espectrômetro e geradores de vapor trabalharão à temperatura ambiente.

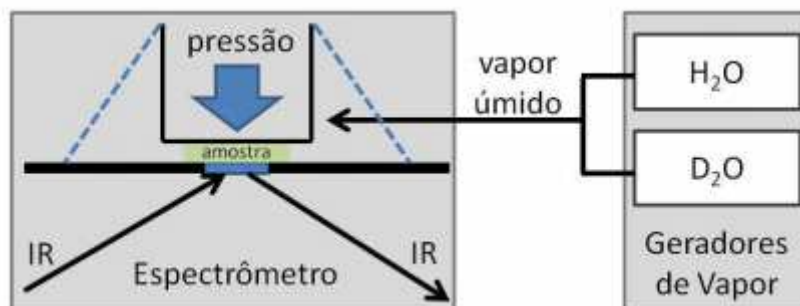


Figura 3 – Esquema do instrumental. O espectrômetro ATR-FTIR tem uma saia (linha azul tracejada) que cria uma atmosfera controlada no entorno da amostra analisada (verde). Os geradores de vapor controlam a umidade relativa dessa atmosfera, com umidade gerada a partir de H₂O ou D₂O.

3. Teoria

Espectroscopia FTIR: Espectroscopia de infravermelho é o estudo da interação da radiação infravermelha com a matéria. A luz é composta de ondas elétricas e magnéticas. Estas duas ondas encontram-se em planos perpendiculares entre si, e a onda de radiação se move através do espaço num plano perpendicular aos planos contendo as ondas elétrica e magnética. É a parte elétrica da luz, denominada *vetor elétrico*, responsável por interagir com moléculas de matéria. [3]

A radiação infravermelha é outra forma de se referir ao calor. Todos os objetos do universo que se encontram a uma temperatura acima do zero absoluto emitem radiação infravermelha. Quando a radiação de infravermelho interage com a matéria, pode ser absorvida, fazendo com que as ligações químicas no material vibrem. A presença dessas ligações no material é essencial para que ocorra absorção de infravermelho. Existe uma correlação entre os números de onda nos quais uma molécula absorve radiação, e a sua respectiva estrutura. Tal correlação permite que a estrutura de moléculas desconhecidas seja identificada através do espectro infravermelho da mesma. É neste fato que reside a grande importância da espectroscopia de infravermelho para a análise química.

Um gráfico de intensidade de radiação infravermelha medida em função do número de onda é chamado *espectro de infravermelho*.

Além de estruturas químicas, o espectro de infravermelho é capaz de fornecer informação quantitativa da mesma forma, tais como a concentração de determinada molécula em uma amostra.

Espectroscopia FTIR é uma forma menos intuitiva de obter essas informações. Ao invés de se incidir um feixe monocromático de radiação sobre a amostra, esta técnica consiste em incidir um feixe contendo várias frequências de luz de uma só vez, e então medir o quanto deste feixe é absorvido pela amostra. Em seguida, o feixe é modificado de tal forma que contenha diferentes combinações de frequências, resultando num segundo conjunto de dados.

O feixe mencionado é gerado através de uma fonte de radiação de banda larga, a qual contém todo o espectro de comprimentos de onda a serem medidos. A radiação incide sobre um interferômetro de Michelson, dispositivo físico que conta com uma configuração específica de espelhos, os quais combinados são capazes de bloquear ou transmitir periodicamente cada comprimento de onda emitido, por meio de interferências de ondas. Diferentes comprimentos de onda são modulados sob diferentes taxas. A cada momento, o feixe que sai do interferômetro possui um espectro diferente [4].

O processamento computacional é requerido a fim de converter os dados brutos (a luz absorvida pelos espelhos) no resultado desejado (a luz absorvida para cada comprimento de onda). O processo necessário trata-se de um algoritmo comum denominado *Transformada de Fourier*. Os dados brutos são muitas vezes chamados “interferogramas” [5].

Análise de componentes principais (PCA): *Principal component analysis* (PCA) é um procedimento estatístico que se utiliza de transformações ortogonais para converter um conjunto de observações de variáveis possivelmente correlacionadas em um conjunto de valores de variáveis linearmente não correlacionadas, chamadas **componentes principais**. O número de componentes principais é menor ou igual ao número de variáveis originais. Esta transformação é definida de forma que o primeiro componente principal tem a maior variância possível (ou seja, é responsável pelo máximo de variabilidade nos dados), e cada componente seguinte, por sua vez, tem a máxima variância sob a restrição de ser ortogonal a (i.e., não correlacionado com) os componentes anteriores [6].

O PCA é matematicamente definido como uma transformação linear ortogonal que transforma os dados para um novo sistema de coordenadas de forma que a maior variância por qualquer projeção dos dados fica ao longo da primeira coordenada (o chamado *primeiro componente*), a segunda maior variância fica ao longo da segunda coordenada, e assim por diante [5]. Cada componente principal é uma combinação linear das variáveis originais. Todas as componentes principais são ortogonais entre si, de forma a não haver redundância de informação. Como um todo, as componentes principais formam uma base ortogonal para o espaço vetorial dos dados

4. Metodologia Experimental:

Amostras de bagaço de cana-de-açúcar, avicel e línter de algodão foram preparadas em laboratório, para posterior medição de seu espectro de radiação infravermelha. Nos casos do bagaço e da avicel, foram cortadas seções muito finas – *feixes vasculares* – da amostra, utilizando-se um bisturi, de forma a penetrar melhor as camadas do material. Posteriormente, estando cada uma das amostras corretamente identificadas, foram medidas no espectrofotômetro, segundo a configuração mostrada na figura 3. Os espectros foram obtidos através dos softwares Spectrum TimeBase (o software fornece diretamente o espectro de infravermelho obtido da transformada de Fourier aplicada sobre o interferograma gerado pelo equipamento).

A princípio, o foco inicial do trabalho estava em observar a dependência angular do espectro do bagaço da cana-de-açúcar. A amostra de bagaço era colocada sobre o porta-amostra do espectrofotômetro, e através de um goniômetro simples, medida em diferentes inclinações em relação ao eixo longitudinal do aparelho, em intervalos de $\Delta\theta = 10^\circ$, $0 \leq \theta \leq 90^\circ$. De fato, encontramos resultados indicando uma presença de variação dos espectros de radiação de infravermelho em função do ângulo da amostra, em regiões onde aparecem grupos funcionais próprios do bagaço. A figura 4 apresenta uma dessas medidas, para intervalo de 650 cm^{-1} a 4000 cm^{-1} . Note a diferença de intensidade de radiação transmitida para os picos nos intervalos $(2500-3000) \text{ cm}^{-1}$ e $(3000-3500) \text{ cm}^{-1}$, para diferentes ângulos configurados.

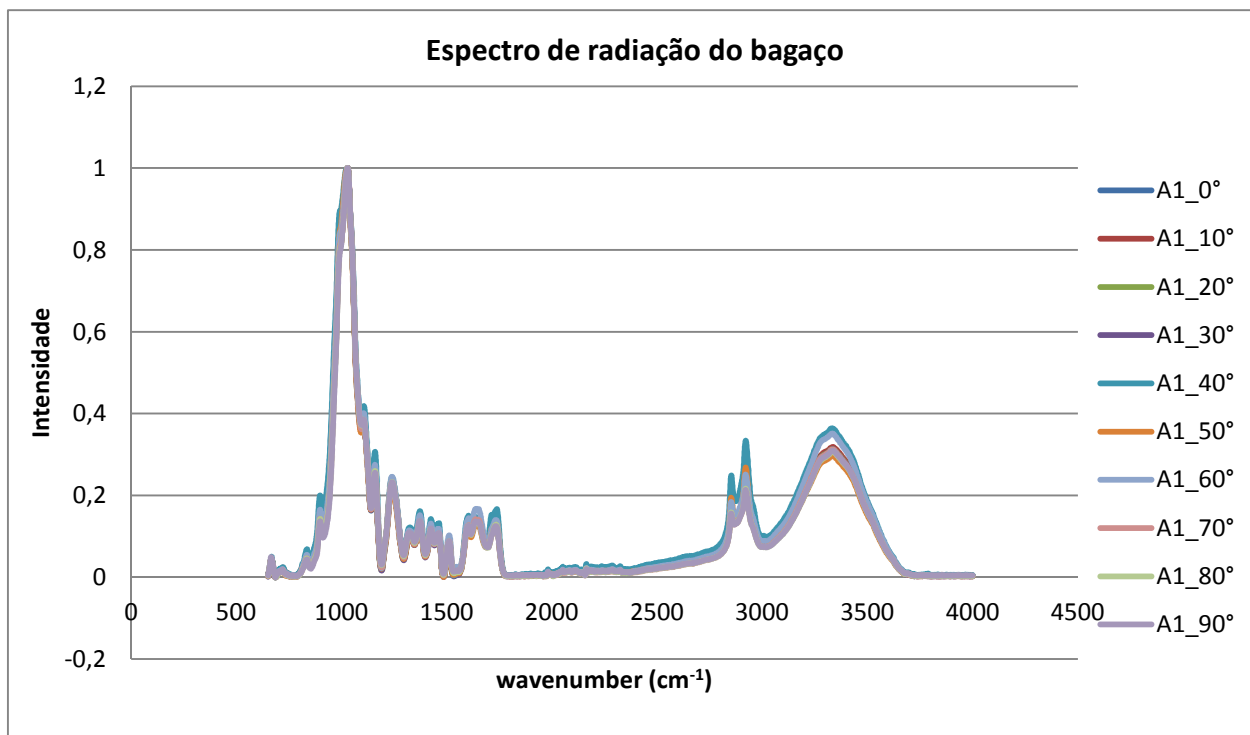


Figura 4 – Espectro de radiação para o bagoço, medido em diferentes ângulos de inclinação da amostra.

Na figura acima, os dados experimentais foram normalizados e a linha de base foi subtraída pelo próprio software do TimeBase. No entanto, constatou-se não ser esta a normalização adequada para análise qualitativa dos dados. Efetivamente, não obtivemos até o presente momento êxito na determinação da melhor normalização para o espectro bruto gerado. A normalização é necessária devido às interferências do instrumento sobre a medição, mas principalmente porque cada configuração angular em que a amostra é posta requer que a mesma seja manipulada pelo operador do experimento, i.e., a amostra é tocada por quem opera o equipamento, e muitas vezes a região medida para um determinado ângulo será diferente ou estará sob uma pressão distinta de outra região da mesma amostra. Tais fatores geram fortes discrepâncias no resultado final da medida.

Concentramos nosso esforço na questão das trocas isotópicas para amostras de bagoço e avicel. Nesse caso, a amostra é mantida intacta dentro da atmosfera produzida pelos geradores de vapor, e os dados obtidos podem ser mais bem normalizados. Foram obtidas evoluções temporais do espectro. Ao longo do tempo, queremos enxergar as interações químicas entre os grupos funcionais presentes no bagoço/avicel, o que equivale avaliar o comportamento de determinados comprimentos de onda do espectro.

Os dados obtidos foram tratados em Excel e MATLAB. No Excel, obtivemos as médias espectrais para cada tempo medido. Aplicamos a normalização $x_i^n - \langle x^n \rangle$, onde x_i^n é o valor do dado bruto do i -ésimo comprimento de onda e n -ésimo tempo, e $\langle x^n \rangle$ é o respectivo valor médio espectral para o n -ésimo tempo. O resultado obtido foi aplicado em um algoritmo construído no MATLAB. O código-fonte é exposto abaixo:

```
clear all
```

```

close all
clc

[nndata, variable, alldata] = xlsread('Arquivo de dados.csv');

[lin,col] = size(alldata);
variable = variable(1,2:col); % vector of variables
sample = alldata(2:lin,1); % vector of materials

data_0 = cell2mat(alldata(2:lin,2:col)); % the data matrix without
weight
[l,c] = size(data_0); % l: the number of materials, c: the number of
variables

% Compute the factor Score, the cumulative variance
[coefs, scores,variances,t] = princomp(data_0);

```

Atente-se para a última linha, onde a função 'princomp' é requerida. Esta é a função responsável por gerar os componentes principais da matriz sobre a qual o algoritmo é aplicado. A função princomp gera, no MATLAB, quatro matrizes de resultado, a saber: 'coefs', 'scores', 'variances' e 't', conforme é apresentado no código. Para os nossos objetivos, as três primeiras matrizes são as mais importantes. Estas definem, respectivamente:

- Component coefficients: contêm os coeficientes da combinação linear das variáveis que geraram os componentes principais.
- Component scores: contêm as coordenadas dos dados originais no novo sistema de coordenadas definido pelas componentes principais.
- Component variances: é um vetor contendo a variância da componente principal correspondente.

5. Resultados e análise de dados:

A figura 5 apresenta um dos resultados obtidos para uma amostra de Avicel. Foi aplicado PCA sobre a matriz de dados $x - \langle x \rangle$ desta amostra, no intervalo de (700-1800) cm^{-1} . O espectro de absorvância de radiação infravermelha da avicel foi medido de forma a acompanhar a evolução temporal para 206 tempos diferentes (Abs 1 a Abs206). A figura apresenta sete intervalos de tempo de medida diferentes. Cada tempo medido apresenta uma variação ao longo de 2min. Facilmente, podemos acompanhar a dependência temporal do espectro, através das diferenças de intensidade para picos localizados nas mesmas faixas de número de onda.

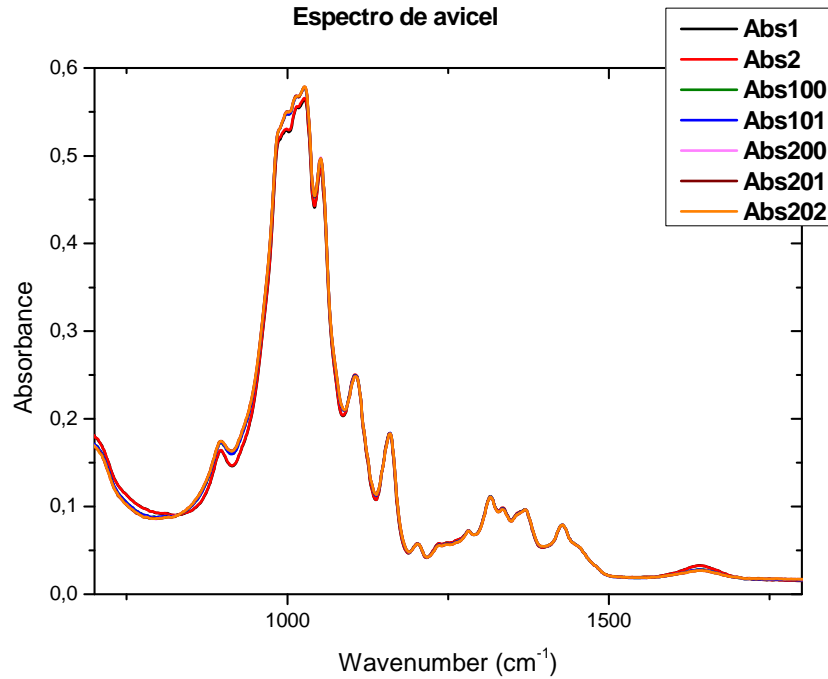


Figura 5 – Espectro de avicel, para sete tempos diferentes.

O gráfico da figura 6 apresenta a distribuição dos coeficientes (*component coefficients*) da combinação linear das variáveis originais ao longo do espectro, que geraram os componentes principais (vide Teoria), para as três primeiras componentes principais, fornecidas pelo algoritmo do MATLAB.

Repare como terceira componente principal (PC3) apresenta uma distribuição cheia de ruídos, com a curva pouco suave. Isto é um forte indicativo de que podemos descartar as componentes principais superiores. Com isto, reduziu-se drasticamente a dimensão do problema, de 206 variáveis (206 espectros diferentes), para somente duas variáveis, ou seja, duas dimensões, representadas por PC1 e PC2, os quais teoricamente carregam as principais informações do sistema.

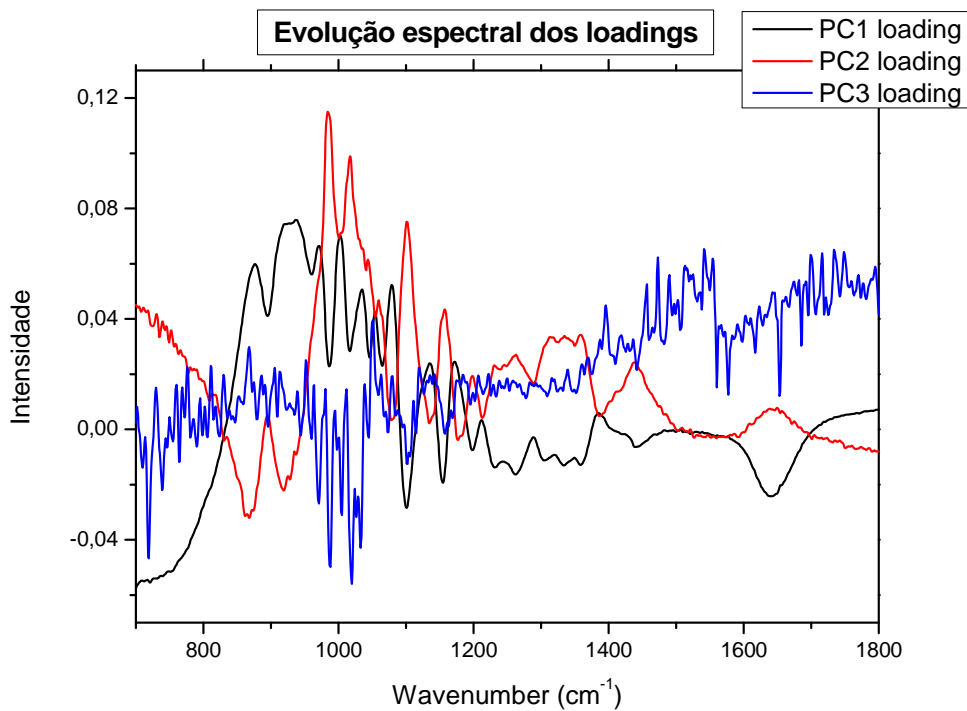


Figura 6 – Evolução espectral das três primeiras componentes principais para uma amostra de Avicel. Espectro entre 700 cm^{-1} e 1800 cm^{-1} .

Abaixo, são apresentados em gráfico os *scores* gerados pelo PCA, i.e., as coordenadas dos dados originais expressas no novo sistema de coordenadas definido pelos componentes principais.

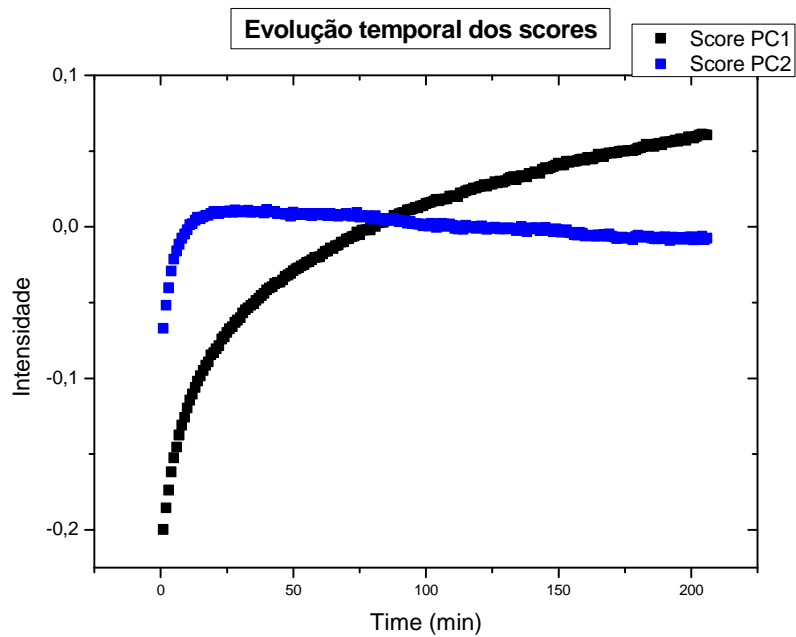


Figura 7 – Scores gerados para PC1 e PC2.

No gráfico, é possível acompanhar o crescimento assintótico da curva em PC1, o que é bastante razoável, uma vez que podemos supor, a princípio, que a curva denota as trocas isotópicas ocorrendo ao longo do processo, através do aumento de alguma componente química na amostra. Por outro lado, a curva de PC2 apresenta um comportamento de aparente oscilação, o que não é esperado para este experimento. Nossa primeira hipótese é a de que PC2 está relacionada com a existência de uma linha de base, sendo então uma informação que esta linha de base carrega ao longo do tempo.

Tomou-se a derivada numérica espectral dos dados brutos de avicel. Aplicando PCA sobre os respectivos valores, esperava-se a redução do número de componentes principais obtidas. De fato, foi observada essa redução, e nossa primeira hipótese foi a de que a componente principal subtraída está relacionada a linha de base do sistema de dados. Por derivada espectral, entenda-se a derivada em relação à evolução do espectro, não a temporal. Foi calculada segundo a expressão $\frac{\partial y}{\partial x} = (f_{v+1} - f_{v-1}) / 2$. A figura 8 apresenta o gráfico dos *scores* para as derivadas. Note que o sinal da curva mudou em relação àquela mostrada na figura 7. Podemos notar também o predomínio da evolução assintótica.

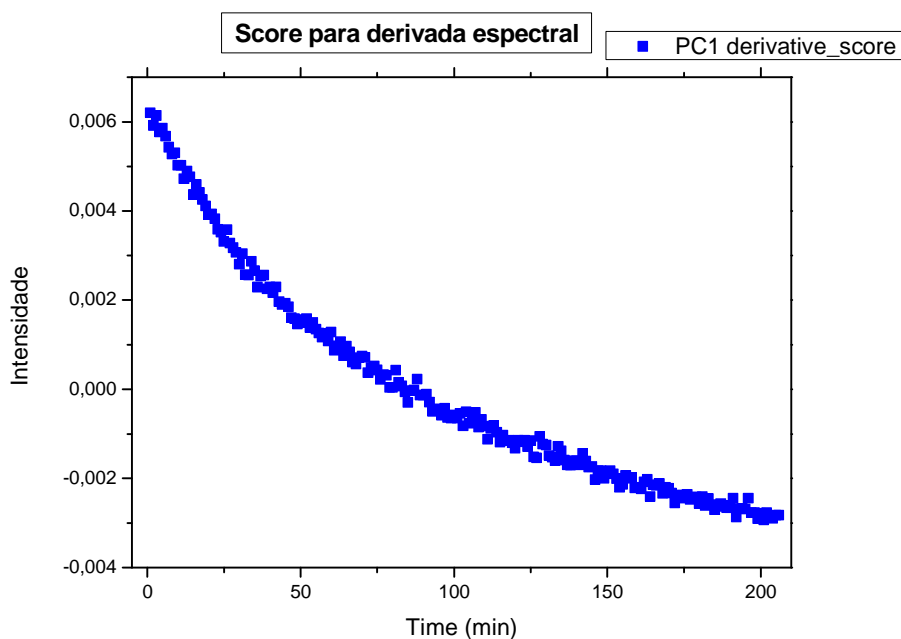


Figura 8 – Scores gerado por PC1 da curva derivada.

Em seguida, foram tomados intervalos específicos do espectro de IR da avicel, e realizou-se a análise por PCA para cada um deles. Os *scores* obtidos são apresentados na figura abaixo, para cada um dos intervalos (vide legenda).

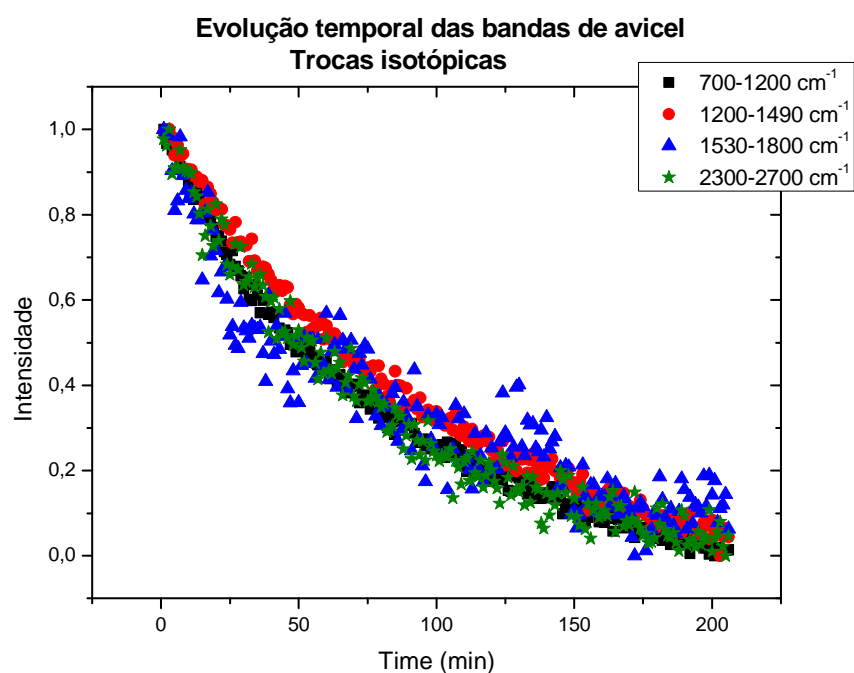


Figura 9 – Scores de PC1 para diferentes bandas da avicel. O PCA foi tomado sobre a derivada espectral dos respectivos intervalos.

A queda assintótica das curvas, para cada uma das bandas, é um grande indicativo da ocorrência das trocas isotópicas entre o meio e o material analisado.

6. Conclusão

Os fatos aqui expostos descrevem os passos realizados até o presente momento ao longo do trabalho de pesquisa realizado pelo aluno. Verificou-se que a aplicação do método de PCA sobre os dados gerados através de espectroscopia de infravermelho aparentam coerência, revelando-se um caminho muito viável e capaz de conduzir à melhor interpretação do fenômeno químico de trocas isotópicas. Pretende-se, nesse sentido, dar continuidade ao método de análise aplicado, desenvolvendo melhores formas de refinamento dos dados gerados.

O orientador concorda com o expressado neste relatório parcial e deu a seguinte opinião:

“O relatório apresenta de forma muito resumida os passos tomados até aqui ao longo do trabalho de pesquisa.”

Abaixo, segue o comentário feito pelo coordenador da matéria acerca do relatório parcial.

“Aprovado, mas lembre que a disciplina é F 590, não F 530. Envie mensagem com nova indicação da data para painel.”

7. Referências

[1] <http://pt.wikipedia.org/wiki/Espectrofot%C3%B4metro>

[2] <http://genrh.jimdo.com/humidity-generator-products/gen-rh-ambient/>

[3] SMITH, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*. 5th ed. CRC Press LLC; c 1996. pp 1-5.

[4] http://en.wikipedia.org/wiki/Fourier_transform_infrared_spectroscopy

[5] http://en.wikipedia.org/wiki/Principal_component_analysis

[6] Jolliffe I.T. *Principal Component Analysis*, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus. ISBN 978-0-387-95442-4