

EXPERIMENTO DIDÁTICO DE QUIMIOMETRIA PARA ANÁLISE EXPLORATÓRIA DE ÓLEOS VEGETAIS COMESTÍVEIS POR ESPECTROSCOPIA NO INFRAVERMELHO MÉDIO E ANÁLISE DE COMPONENTES PRINCIPAIS: UM TUTORIAL, PARTE I

André Marcelo de Souza e Ronei Jesus Poppi*

Departamento de Química Analítica, Instituto de Química, Universidade Estadual de Campinas, CP 6154, 13084-971 Campinas – SP, Brasil

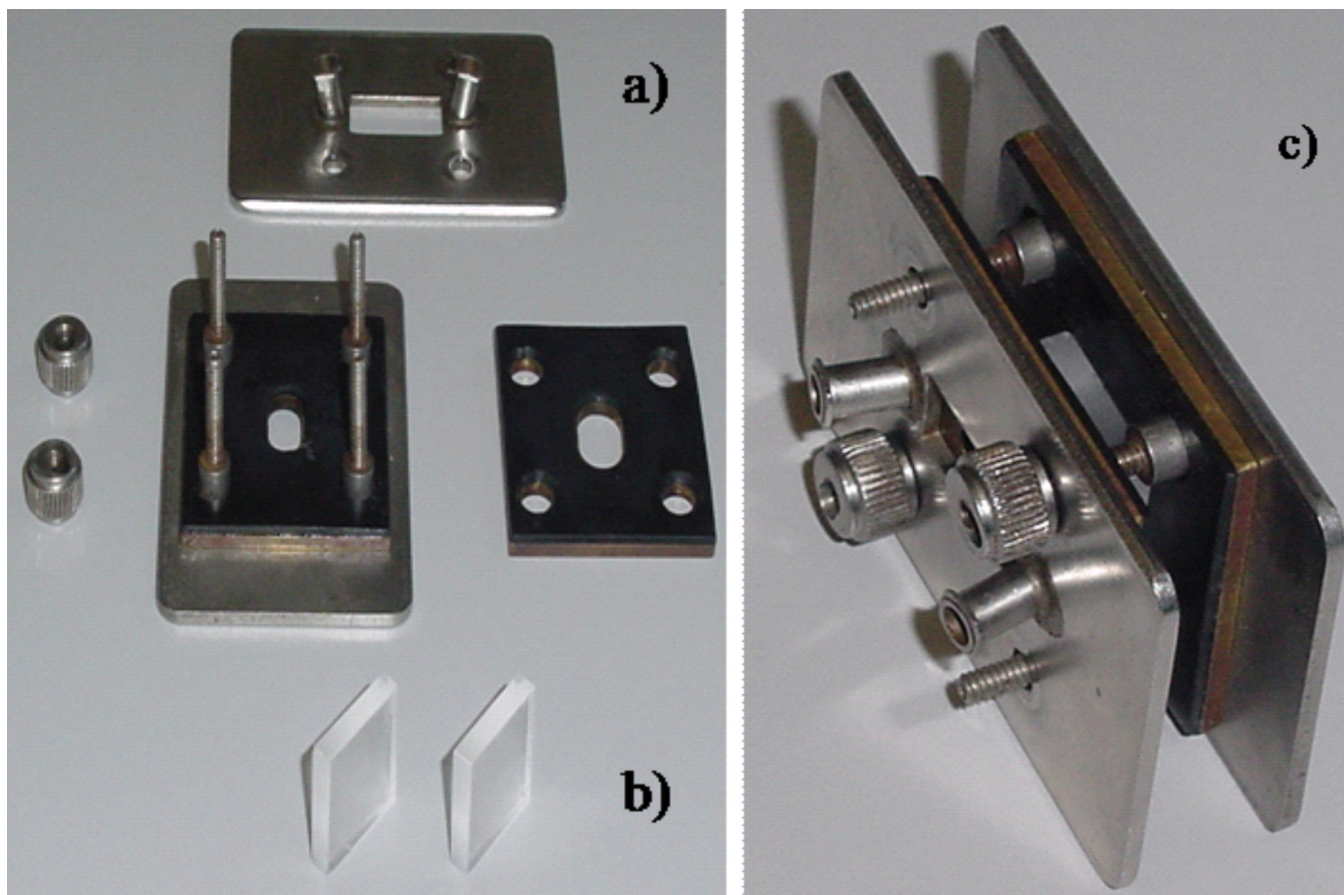


Figura 1S. Cella desmontável para amostras líquidas: a) componentes da cela; b) janela de KBr; c) janelas de KBr

*e-mail: ronei@iqm.unicamp.br

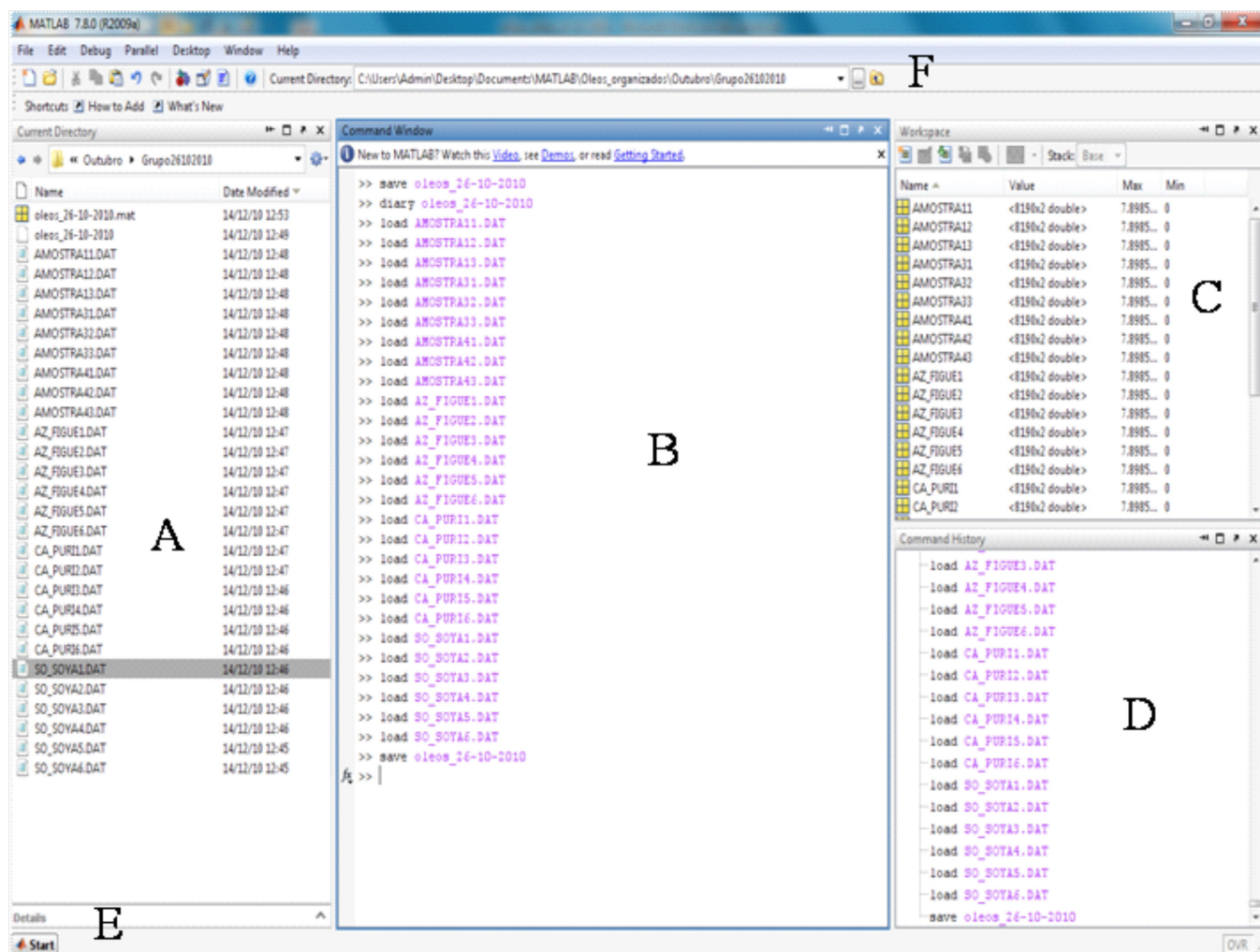


Figura 2S. A) [Current directory]: área na qual é exibida a lista dos arquivos contidos no diretório corrente; B) [Command window]: área de trabalho na qual as instruções são digitadas; C) [Workspace]: área na qual são exibidas as variáveis definidas na área de trabalho do Matlab; D) [Command history]: área na qual ficam armazenadas todas as instruções executadas no Matlab; E) Menu pop-up: comandos para configuração do ambiente Matlab e, F) menu de ícone: barra de ferramentas padrão do Matlab

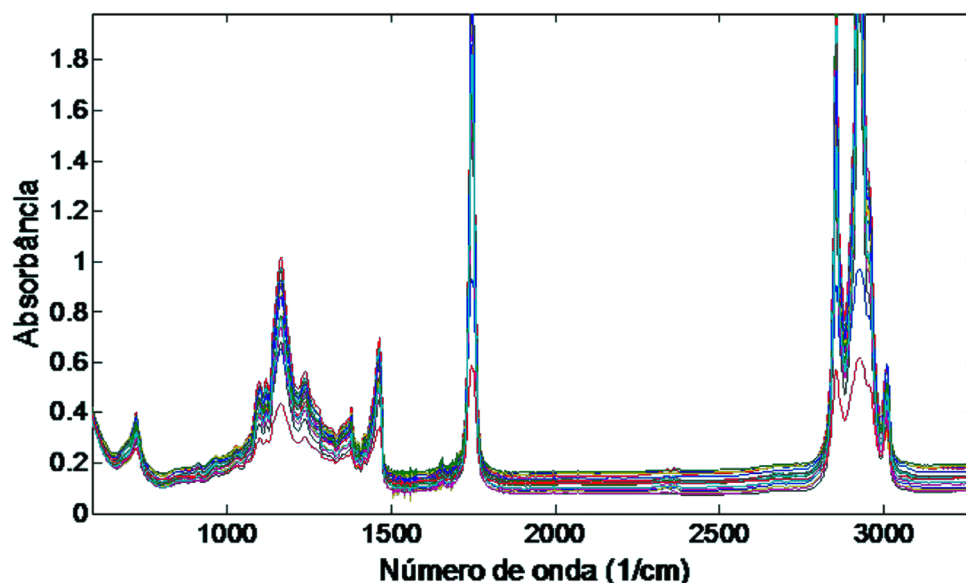


Figura 3S. Gráfico da matriz de espectros das 18 amostras de óleo

Quadro 1S. Sequência de comandos executados para carregar e pré-processar, respectivamente, os 18 espectros das amostras de óleo

```
%% Carregando os espectros das 18 amostras de óleo%% Carregando
os dados
load a1.dat;load a2.dat;load a3.dat;load a4.dat;load a5.dat;load a6.dat;load
c1.dat;load c2.dat;load c3.dat;load c4.dat;load c5.dat;load c6.dat;load
s1.dat;load s2.dat;load s3.dat;load s4.dat;load s5.dat;load s6.dat;
%% Montando a matriz X
x=[a1(:,2);a2(:,2);a3(:,2);a4(:,2);a5(:,2);a6(:,2);c1(:,2);c2(:,2);c3(:,2)
';c4(:,2);c5(:,2);c6(:,2);s1(:,2);s2(:,2);s3(:,2);s4(:,2);s5(:,2);s6(:,2)'];
%% Criando eixo x em número de ondas utilizando a primeira coluna:
num=a1(:,1);
%% Selecionando a faixa espectral de interesse
x1=x(:,600:3400);
num1=num(600:3400);
figure
plot(num1,x1) %
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
title('Espectros das 18 amostras de óleo')
%% Selecionando a faixa de impressão digital
x2=x1(:,400:950);
num2=num1(400:950);
figure
plot(num2,x2) %
title('Faixa de impressão digital')
xlabel('Número de onda (1/cm)');ylabel('Absorbância')
%% Subplots
figure
subplot(3,1,1), plot(num,x);title('a');grid;subplot(3,1,2),
plot(num1,x1);title('b');grid;
subplot(3,1,3), plot(num2,x2);title('c');grid;
ylabel('Absorbância');
xlabel('Número de onda (1/cm)');
%% Préprocessamentos
% Centrando na média
x2m=mean(x2);
for i=1:18
x2mc(i,:)=x2(i,:)-x2m;
end
figure
plot(num2,x2mc)
title('Centrado na média')
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
% 1ª derivada
%[dx] = deriv(x,der.window,order)
[dx1x2nmc] = deriv(x2mc,2,15,2);
figure
plot(num2,dx1x2nmc) %
title('Centrado na média + 1a derivada')
xlabel('Número de onda (1/cm)')
ylabel('1a derivada')
```

Quadro 2S. Sequência de comandos para a execução da PCA a partir dos dados centrados na média e derivados pela primeira derivada (matriz de dados dx1x2nmc)

```
%% PCA
[t,p,percent] = kpca(dx1x2nmc);
% Gráfico dos scores
% PC1xPC2
figure
plot(t(1:6,1),t(1:6,2),'xg',t(7:12,1),t(7:12,2),'+r',t(13:18,1),t(13:18,2),'*b')
grid
for i=1:18;
text(t(i,1),t(i,2),num2str(i))
end
xlabel('PC1 (79.8789%)');
ylabel('PC2(17.7318%)');
title('Centrado na média + 1a derivada + EVD')
legend('Azeite','Canola','Soja');
% Gráfico dos loadings
% PC2
figure
plot(num2,p(:,2))
xlabel('1/cm');
ylabel('PC2(17.7318%)');
title('Centrado na média + 1a derivada + EVD')
% Gráfico dos scores
% PC1xPC3
figure
plot(t(1:6,1),t(1:6,3),'xg',t(7:12,1),t(7:12,3),'+r',t(13:18,1),t(13:18,3),'*b')
grid
for i=1:18;
text(t(i,1),t(i,3),num2str(i))
end
xlabel('PC1 (79.8789%)');
ylabel('PC3 (1.1825%)');
title('Centrado na média + 1a derivada + EVD')
legend('Azeite','Canola','Soja');
% Gráfico dos loadings
% PC3
figure
plot(num2,p(:,3))
xlabel('1/cm');
ylabel('PC3 (1.1825%)');
title('Centrado na média + 1a derivada + EVD')
```

Quadro 3S. Sequência de comandos para a execução da PCA a partir dos dados centrados na média e derivados através da segunda derivada

```
% 2ª derivada
%[dx] = deriv(x,der>window,order)
[dx2x2nmc] = deriv(x2mc,2,15,2); % x2mc é a matriz centrada na média
figure
plot(num2,dx2x2nmc)
title('2a derivada')
xlabel('Número de onda (1/cm)')
ylabel('2a derivada')
%% PCA
[t,p,percent] = kpca(dx2x2nmc);
% Gráfico dos scores PC1xPC2
figure
plot(t(1:6,1),t(1:6,2),'xg',t(7:12,1),t(7:12,2),'+r',t(13:18,1),t(13:18,2),'*b')
grid
for i=1:18;
text(t(i,1),t(i,2),num2str(i))
end
xlabel('PC1 (79.8789%)');
ylabel('PC2(17.7318%)');
title('Centrado na média + 2a derivada + EVD')
legend('Azeite','Canola','Soja');
% Gráfico dos loadings na PC2
figure
plot(num2,p(:,2))
xlabel('1/cm');
ylabel('PC2(17.7318%)');
title('Centrado na média + 2a derivada + EVD')
```

Quadro 4S. Sequência de comandos para a execução da PCA, sendo os espectros pré-processados com o MSC e centrados na média, respectivamente

```
% Pré-processamento da matriz X2
% MSC
[xmsc]=msc(x2);
figure
plot(num2,xmsc)
title('MSC')
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
% centrando na média
x2m=mean(xmsc);
for i=1:18
xmscmc(i,:)=xmsc(i,:)-x2m;
end
figure
plot(num2,xmscmc)
title('MSC + centrado na média')
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
%% PCA
[t,p,percent] = kpca(xmscmc);
% Gráfico dos scores PC1xPC2
figure
plot(t(1:6,1),t(1:6,2),'xg',t(7:12,1),t(7:12,2),'+r',t(13:18,1),t(13:18,2),'*b')
grid
for i=1:18;
text(t(i,1),t(i,2),num2str(i))
end
xlabel('PC1 (50.6736%)');
ylabel('PC2(39.4630%)');
title('MSC + centrado na média + EVD')
legend('Azeite','Canola','Soja');
% Gráfico dos loadings na PC2
figure
plot(num2,p(:,2))
xlabel('1/cm');
ylabel('PC2(39.4630%)');
title('MSC + centrado na média + EVD')
```

Quadro 5S. Sequência de comandos executados para carregar os 27 espectros das amostras de óleo e selecionar as faixas espectrais de interesse

```
%% Carregando os vinte e sete espectros do conjunto de dados
load a1.dat;load a2.dat;load a3.dat;load a4.dat;load a5.dat;load a6.dat;load
c1.dat;
load c2.dat;load c3.dat;load c4.dat;load c5.dat;load c6.dat;load s1.dat;load
s2.dat;
load s3.dat;load s4.dat;load s5.dat;load s6.dat;load am11.dat;load am12.dat;
load am13.dat;load am31.dat;load am32.dat;load am33.dat;load am41.dat;
load am42.dat;load am43.dat;
%% Montando a matriz X
xc=[a1(:,2);a2(:,2);a3(:,2);a4(:,2);a5(:,2);a6(:,2);c1(:,2);c2(:,2);
c3(:,2);c4(:,2);c5(:,2);c6(:,2);s1(:,2);s2(:,2);s3(:,2);
s4(:,2);s5(:,2);s6(:,2);am11(:,2);am12(:,2);am13(:,2);am31(:,2);
am32(:,2);am33(:,2);am41(:,2);am42(:,2);am43(:,2)];
%% Criando eixo x em número de ondas utilizando a primeira coluna:
num=a1(:,1);
%% Selecionando a faixa espectral de interesse
xc1=xc(:,600:3400);
num1=num(600:3400);
figure
plot(num1,xc1)
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
title('Espectros das 27 amostras de óleo')
%% Selecionando a faixa de impressão digital
xc2=xc1(:,400:950);
num2=num1(400:950);
figure
plot(num2,xc2)
title('Faixa de impressão digital')
xlabel('Número de onda (1/cm)')
ylabel('Absorbância')
%% subplots
figure
subplot(3,1,1), plot(num,xc);title('a');grid;subplot(3,1,2), plot(num1,xc1)
;title('b');grid;subplot(3,1,3), plot(num2,xc2);title('c');
grid
ylabel('1a derivada');
xlabel('Número de onda (1/cm)');
```


Quadro 6S. Sequência de comandos para executar a PCA com os 27 espectros a partir dos dados centrados na média e derivados através da primeira derivada

```
%% Preprocessamento da matriz XC2
% Centrando na média
xc2m=mean(xc2);
for i=1:27
xc2mc(i,:)=xc2(i,:)-xc2m;
end
figure
plot(num2,xc2mc)
title('Centrado na média');xlabel('Número de onda (1/cm)');ylabel('Absorbância')
%% 1ª derivada
%[dx] = deriv(x,der.window,order)
[dx1xc2nmc] = deriv(xc2mc,2,15,2);
figure
plot(num2,dx1xc2nmc)
title('Centrado na média + 1a derivada');xlabel('Número de onda (1/cm)')
ylabel('1a derivada')
%% PCA - conhecidas
[t,p,percent] = kpca(dx1xc2nmc);
% Gráfico dos scores
% PC1xPC2 (Figura 7)
figure
plot(t(1:6,1),t(1:6,2),'xg',t(7:12,1),t(7:12,2),'+r',t(13:18,1),t(13:18,2),
'*b',t(19:21,1),t(19:21,2),'ok',t(19:21,1),t(19:21,2),'xk',t(22:24,1),
t(22:24,2),'om',t(22:24,1),t(22:24,2),'+m',t(25:27,1),t(25:27,2),'oc',t(25:27,1),t(25:27,2),'*c')
grid
for i=1:27;
text(t(i,1),t(i,2),num2str(i))
end
xlabel('PC1 (74.5796%)');ylabel('PC2(21.7213%)');title('Centrado na média + 1a derivada');legend('Azeite','Canola','Soja');
% PC1xPC3
figure
plot(t(1:6,1),t(1:6,3),'xg',t(7:12,1),t(7:12,3),'+r',t(13:18,1),t(13:18,3),
'*b',t(19:21,1),t(19:21,3),'ok',t(19:21,1),t(19:21,3),'xk',t(22:24,1),
t(22:24,3),'om',t(22:24,1),t(22:24,3),'+m',t(25:27,1),t(25:27,3),'oc',t(25:27,1),t(25:27,3),'*c')
grid
for i=1:27;
text(t(i,1),t(i,3),num2str(i))
end
xlabel('PC1 (74.5796%)');ylabel('PC3 (1.6158%)');title('Centrado na média + 1a derivada');legend('Azeite','Canola','Soja');
% Gráfico dos loadings
% PC2
figure
plot(num2,p(:,2))
xlabel('1/cm');ylabel('PC2(21.7213%)');title('Centrado na média + 1a derivada')
```

Tabela 1S. Tipos de óleos vegetais e quantidade de medidas realizadas por cada subgrupo

| Tipo de óleo | Código das medidas | Posição na matriz X |
|--------------|--------------------|---------------------|
| Azeite | a1;a2;a3;a4;a5;a6 | 1 a 6 |
| Canola | c1;c2;c3;a4;a5;a6 | 7 a 12 |
| Soja | s1;s2;s3;s4;s5;s6 | 13 a 18 |
| am1 | am11;am12;am13 | 19 a 21 |
| am2 | am21;a,22;am23 | 22 a 24 |
| am3 | am31;am32;am33 | 25 a 27 |