

MATERIAL SUPLEMENTAR

Principal component analysis (PCA) para a avaliação de dados químicos e geração de heat maps: um tutorial

Dennis da Silva Ferreira^{a, , Leticia da Silva Rodrigues^b, Fabiola Manhas Verbi Pereira^b e Edenir Rodrigues Pereira-Filho^{a,*, }}

^aDepartamento de Química, Universidade Federal de São Carlos, 13565-905 São Carlos – SP, Brasil

^bInstituto de Química, Universidade Estadual Paulista, 14800-060 Araraquara – SP, Brasil

*e-mail: erpf@ufscar.br

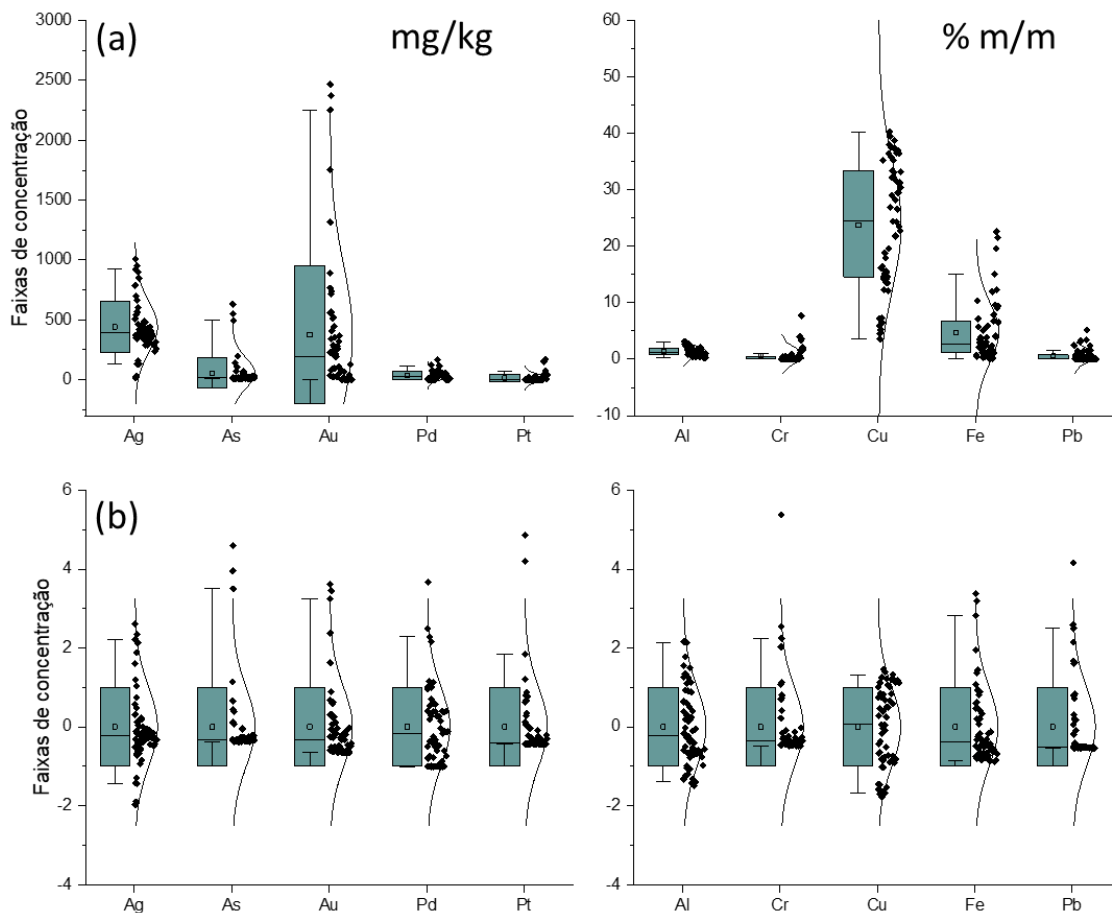


Figura 1S. Box-plot das concentrações dos 10 analitos determinados: (a) dados originais e (b) autoescalados

O elemento Cu, por exemplo, apresentou o menor valor de assimetria (-0,24). Com a Figura 1S(a) pode ser verificado, utilizando o *box-plot*, que a média dos dados para o Cu (quadrado no interior do *box-plot*) está próxima à linha horizontal no interior que representa a mediana.

O tamanho do box-plot da Figura 1S representa o desvio padrão dos dados e as barras de erros possuem 95% dos dados.

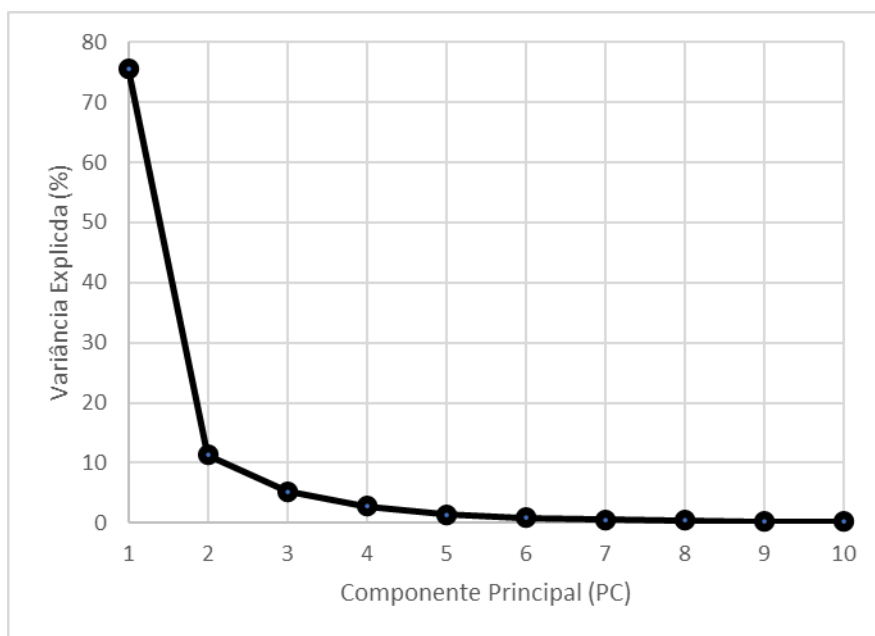


Figura 2S. Porcentagem de variância explicada em função do número de componentes principais para o pulso 5 (ver mais detalhes na Figura 1)

Descrição das rotinas computacionais utilizadas nos exemplos

LIBS_inverse:

```
function [MATRIZ] = LIBS_inverse(X,NP,NL,NC)
k=1;
s=0;
% criando matriz origem
M=ones(NL,NC*NP);
    for i=1:NL
        for j=1:NC*NP
            s=s+1;
            M(i,j)=s;
        end
    end
% mudança de ordem
for i=1:NL
    for j=1:NC*NP
        if mod(i,2)== 0
            Mc(i,j)=M(i,NC*NP+1-j);
        else
            Mc(i,j)= M(i,j);
        end
        Ml(k,1)=Mc(i,j);
        k=k+1;
    end
end
% substituindo na matriz inicial
MATRIZ= X(Ml,:);
```

data_pre:

```
function [Xauto,Xcm,xm,xstd] = data_pre(X);
    % [Xauto,Xcm,xm,xstd] = data_pre(X);
    % Esta função tem como objetivo centrar na média e
    autoescalar uma matriz X.
    % Inputs: Matriz X (m linhas e n colunas)
    % Outputs:
    % Xauto: Matriz autoescalada
    % Xcm: Matriz centrada na média
    % xm: vetor contendo a média das variáveis
    % xstd: vetor contendo o desvio padrão das variáveis
    % Efetivação dos cálculos

    [m,n]=size(X);      % Este comando permite a
    visualização do número de linhas e colunas da matriz X.
    xm=mean(X);          % Cálculo da média
    xstd=std(X);          % Cálculo do desvio padrão
    A=ones(m,1)*xm;      % Uso da função ones para
    regularizar o tamanho das matrizes.
    B=ones(m,1)*xstd;
    Xauto = (X-A)./B;    % Geração da matriz autoescalada
    Xcm=X-A;             % Geração da matriz centrada na
    médiam

    % Visualização dos dados antes e após o pré-
    processamento

    plot(X','-');
    xlabel ('Variáveis')
    ylabel ('Sinais')
    title ('Dados originais')

    figure
    plot(Xauto','-')
    xlabel ('Variáveis')
    ylabel ('Sinais autoescalados')
    title ('Dados autoescalados')

    figure
    plot(Xcm','-')
    xlabel ('Variáveis')
    ylabel ('Sinais centrados na média')
    title ('Dados centrados na média')
```

pca_dis:

```
function [scores, loads, var_exp] = pca_dis (X)
    % [scores, loads, var_exp] = pca_dis (X)
    % Esta função tem como objetivo efetuar os cálculos de
    uma Análise por Componentes Principais (PCA).
    % Neste caso está sendo utilizada a função svd já
    disponível no Octave.
    % Efetivação dos cálculos
    [m,n]=size(X);
    [U,S,V]=svd(X,'econ');

    % Cálculos dos scores, loadings e variância explicada
    scores = U*S;
    loads = V;
    var_exp=(diag(S).^2)*100/(sum(diag(S).^2));

    % Gráficos para visualização dos dados

    plot (var_exp,'-b^')
    xlabel ('Componente Principal (PC)')
    ylabel ('Variância Explicada (%)')
    title ('Variância Explicada')

    figure
    plot (scores(:,1),scores(:,2),'ro')
    xlabel ('PC1')
    ylabel ('PC2')
    title ('Scores PC1xPC2')
    grid
    % Nesta parte foi inserido um loop para identificar as
    amostras
    for i=1:m
        text(scores(i,1),scores(i,2),num2str(i))
    end

    figure
    plot (loads(:,1),loads(:,2),'gv')
    xlabel ('PC1')
    ylabel ('PC2')
    title ('Loadings PC1xPC2')
    grid
    % Nesta parte foi inserido um loop para identificar as
    variáveis
    for i=1:n
        text(loads(i,1),loads(i,2),num2str(i))
    end
end
```

libs:

```
function [MAPS,T,P,Var_exp] = libs(MATRIZ,NT,NPH,NPV,PCS)

% Parâmetros de entrada
% MATRIZ, matriz com todos os tiros e todos os pontos na
forma:
% NT, Número de tiros;
% NPH, Número de pontos na horizontal;
% NPV, Número de pontos na vertical;
% PCS, Número de mapas de escores;
% Parâmetros de saída
% MAPS, mapas de escores em tensor 4D na forma NT x PCS x NPV
x NPH
% T, escores na forma NT x NP x componentes principais
% P, loadings na forma NT x comprimento de onda x componentes
principais
% Var_exp, porcentagem de variância explicada por cada PC na
forma NT x
% Matriz X na forma: X(tiros, pontos, espectro)
% TPT, Total de pontos;

TPT = NPH * NPV * NT;
for j=1:NT; ctr=0;
for i=j:NT:TPT
ctr=ctr+1;
X(j,ctr,:)=MATRIZ(i,:);
end
end
% Fazendo o PCA para cada matriz:
for i = 1:NT
    [u,s,v]=svd(squeeze(X(i,:,:)),0);
    T(i,:,:)=u(:,1:PCS)*s(1:PCS,1:PCS);
    P(i,:,:)=v(:,1:PCS);
    Var_exp(i,:)=(diag(s).^2)./sum(diag(s).^2)*100;
end
% Mapa de scores
for i = 1:NT
    for j = 1:PCS; ctr = 0;
        for k = 1:NPV
            for l = 1:NPH
                ctr = ctr + 1;
                MAPS(i,j,k,l) = T(i,ctr,j);
            end
        end
    end
end
end
```

libs_plot:

```
function libs_plot(MAPS,TIRO,PCS)

% Parâmetros de entrada
% MAPS, mapas de escores em tensor 4D na forma NT x PCS x NPV
x NPH
% TIRO, número do tiro para o qual quer ver os mapas de
escores
% PCS, mapas até qual componente principal? Tem que ser menor
que o valor
% de PCS do MAPS calculado.

for i = 1:PCS
figure(i);contourf(squeeze(squeeze(MAPS(TIRO,i,:,:))));view([
0,270]);
xlabel('Axis x'); ylabel('Axis y');
title(['Score maps of PC',num2str(i),' from pulse
',num2str(TIRO),'']);
colorbar
end
```

Tabela 1S. Tabela de caracterização em ICP OES dos *e-waste* escolhidos para este tutorial

Amostra	ID	Tamanho (μm)	Ag (mg kg^{-1})	Au (mg kg^{-1})	Pt (mg kg^{-1})	Pd (mg kg^{-1})	As (mg kg^{-1})	Al % (m/m)	Cu % (m/m)	Fe % (m/m)	Cr % (m/m)	Pb % (m/m)
1	MB1	< 106	508,88	560,37	0,04	7,46	16,78	3,12	5,93	0,74	0,03	2,46
2	MB2	< 106	372,77	229,19	0,42	9,60	11,24	3,11	7,25	0,46	0,03	2,38
3	N1	< 106	544,56	33,57	0,04	0,01	12,25	2,63	4,47	10,39	0,44	0,05
4	N2	< 106	564,95	77,53	0,04	0,01	9,54	3,09	3,60	7,14	0,06	0,02
5	N3	< 106	465,43	563,77	0,04	0,01	14,47	2,40	5,22	5,74	0,25	0,66
6	N4	< 106	417,44	340,85	0,04	0,01	12,05	2,46	4,75	7,21	0,30	0,40
7	T1	< 106	449,83	765,77	7,31	18,41	139,00	2,26	16,19	2,12	0,26	0,02
8	T2	< 106	406,47	231,56	0,04	0,01	11,08	2,39	7,22	2,96	0,10	0,02
9	T3	< 106	352,82	213,82	0,04	0,48	11,73	2,58	6,52	3,43	0,09	0,02
10	S1	< 106	342,55	2468,06	17,96	126,00	70,63	1,69	16,50	3,65	0,04	1,49
11	S2	< 106	364,72	891,68	12,97	72,93	110,34	0,92	35,20	3,86	0,52	0,11
12	S3	< 106	388,18	717,60	0,04	22,73	105,05	2,81	14,63	1,85	0,14	0,05
13	S4	< 106	394,72	2254,93	1,71	77,88	551,75	1,55	14,20	2,17	0,24	0,04
14	S5	< 106	415,15	2372,41	5,08	71,03	494,96	1,27	12,34	2,01	0,19	0,04
15	S6	< 106	359,60	737,94	1,13	55,45	631,36	2,29	15,14	1,25	0,17	0,03
16	S7	< 106	389,35	1755,36	4,21	22,26	199,27	0,99	18,85	3,04	0,34	0,01
17	MB1	106-212	443,71	86,44	0,62	27,43	15,91	1,90	17,95	0,43	0,01	2,99

Tabela 1S. Tabela de caracterização em ICP OES dos *e-waste* escolhidos para este tutorial (cont.)

Amostra	ID	Tamanho (μm)	Ag (mg kg^{-1})	Au (mg kg^{-1})	Pt (mg kg^{-1})	Pd (mg kg^{-1})	As (mg kg^{-1})	Al % (m/m)	Cu % (m/m)	Fe % (m/m)	Cr % (m/m)	Pb % (m/m)
18	MB2	106-212	476,85	289,68	1,28	24,45	11,19	2,13	15,62	0,24	0,01	3,38
19	N1	106-212	490,43	21,53	0,04	0,01	14,69	1,86	13,59	5,11	0,19	0,05
20	N2	106-212	481,74	24,03	0,04	8,45	9,36	2,12	14,17	2,64	0,02	0,04
21	N3	106-212	400,17	275,89	0,04	20,23	11,24	1,37	15,37	2,09	0,04	0,94
22	N4	106-212	377,95	409,31	0,04	18,15	12,25	1,18	13,54	2,13	0,03	0,83
23	T1	106-212	429,79	353,89	16,34	72,49	8,92	0,34	14,57	0,51	0,11	0,15
24	T2	106-212	380,75	95,47	0,04	19,31	18,15	0,60	12,13	1,24	0,05	0,02
25	T3	106-212	424,17	91,60	0,04	14,75	11,91	1,63	19,63	2,65	0,06	0,01
26	S1	106-212	331,58	511,18	34,55	168,48	24,02	0,86	36,49	3,04	0,03	1,48
27	S3	106-212	346,07	191,27	3,60	58,28	9,75	1,21	40,28	0,58	0,09	0,02
28	S4	106-212	401,14	519,92	4,44	77,10	51,54	1,10	39,56	0,62	0,14	0,04
29	S5	106-212	287,00	1316,78	11,32	118,48	65,80	0,40	26,99	5,84	0,48	0,04
30	S6	106-212	377,64	253,44	6,27	114,12	48,69	1,11	37,61	1,26	0,28	0,03
31	S7	106-212	305,50	442,43	4,97	50,45	22,00	0,55	35,79	0,90	0,18	0,02
32	MB1	212-600	443,05	280,26	0,04	43,34	18,97	1,50	29,03	0,43	0,00	3,48
33	MB2	212-600	407,71	192,23	0,81	46,47	13,19	1,53	32,25	0,17	0,02	5,22
34	N1	212-600	922,29	23,17	0,04	50,06	21,74	1,31	33,41	1,62	0,03	0,03

Tabela 1S. Tabela de caracterização em ICP OES dos *e-waste* escolhidos para este tutorial (cont.)

Amostra	ID	Tamanho (μm)	Ag (mg kg^{-1})	Au (mg kg^{-1})	Pt (mg kg^{-1})	Pd (mg kg^{-1})	As (mg kg^{-1})	Al % (m/m)	Cu % (m/m)	Fe % (m/m)	Cr % (m/m)	Pb % (m/m)
35	N2	212-600	912,76	43,53	0,04	34,85	10,69	1,79	24,45	1,62	0,05	0,02
36	N3	212-600	1006,99	259,81	0,04	57,10	17,56	0,89	35,25	2,13	0,03	1,40
37	N4	212-600	950,83	318,91	0,04	50,33	17,42	1,07	32,16	1,38	0,02	0,79
38	T2	212-600	789,02	108,86	2,73	46,09	12,86	0,98	33,10	1,71	0,11	0,02
39	T3	212-600	335,97	98,96	0,04	30,32	11,44	0,83	37,42	2,41	0,18	0,02
40	S1	212-600	346,90	365,47	0,04	45,88	20,27	0,44	38,73	11,94	0,03	0,81
41	S2	212-600	293,18	216,67	23,47	61,76	23,58	0,77	37,96	3,82	0,94	0,03
42	MB1	> 600	13,41	93,48	0,04	41,93	16,40	1,58	28,24	0,04	0,00	2,41
43	MB2	> 600	30,00	63,20	0,04	32,36	12,39	2,09	21,84	0,29	0,00	1,53
44	N1	> 600	901,74	0,01	0,04	50,61	23,74	1,34	37,11	7,86	0,61	0,02
45	N2	> 600	700,32	0,70	0,04	49,77	13,31	0,90	36,53	6,83	0,19	0,02
46	N3	> 600	666,15	15,27	0,04	31,72	8,16	1,71	24,37	7,97	0,30	0,02
47	N4	> 600	850,31	67,02	7,52	50,41	16,34	0,85	31,44	1,12	0,18	0,50
48	T1	> 600	602,11	0,01	149,84	1,89	17,72	0,54	26,61	15,06	3,99	0,01
49	T2	> 600	378,15	0,01	39,97	32,11	11,80	0,91	29,56	9,64	1,57	0,03
50	T3	> 600	368,94	0,01	25,32	3,62	13,49	0,87	37,03	4,08	1,18	0,02
51	S1	> 600	131,70	0,01	53,10	0,01	16,41	0,25	29,44	19,64	3,32	0,07

Tabela 1S. Tabela de caracterização em ICP OES dos *e-waste* escolhidos para este tutorial (cont.)

Amostra	ID	Tamanho (μm)	Ag (mg kg^{-1})	Au (mg kg^{-1})	Pt (mg kg^{-1})	Pd (mg kg^{-1})	As (mg kg^{-1})	Al % (m/m)	Cu % (m/m)	Fe % (m/m)	Cr % (m/m)	Pb % (m/m)
52	S2	> 600	159,71	0,01	171,50	0,01	31,17	0,34	23,47	22,61	7,74	0,00
53	S3	> 600	292,78	0,01	73,80	0,01	18,30	0,78	36,47	12,31	3,61	0,01
54	S4	> 600	237,60	38,55	35,68	4,90	27,01	0,61	31,27	9,04	1,60	0,04
55	S5	> 600	317,44	129,75	4,70	0,01	20,66	0,21	22,77	21,58	2,06	0,01
56	S6	> 600	257,62	0,01	42,43	14,51	28,99	0,94	33,23	9,39	2,13	0,02
57	S7	> 600	128,58	0,01	39,02	9,96	13,24	0,28	30,43	6,54	1,71	0,01

