

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

Recebido em 17/10/2022; aceito em 11/01/2023; publicado na web 08/03/2023

PRINCIPAL COMPONENT ANALYSIS (PCA) FOR CHEMICAL DATA EVALUATION AND HEAT MAPS PREPARATION: A TUTORIAL. This tutorial shows a step-by-step guide on handling big datasets using principal component analysis (PCA). A dataset of chemical elements' concentration, emission spectrum, and energy-dispersive X-ray fluorescence (EDXRF) of e-waste were used as examples. Five routines were proposed to apply data processing and PCA calculation focusing data from laser-induced breakdown spectroscopy (LIBS), EDXRF, and heat maps preparation. These routines can be used in various softwares such as MatLab, Octave, R, and Python. PCA was applied in three examples; the first was for concentrations, and the other two were for spectra. An example of heat maps assembling a hyperspectral image of a printed circuit was also described. In addition, a playlist was created on YouTube using the available examples. Therefore, with this tutorial, it may be possible to learn how to deal with a large volume of data by applying PCA. The authors hope to contribute to those researching in the area.

Keywords: exploratory analysis; data mining; data visualization; direct solid sample analysis; laser; X-ray fluorescence.

INTRODUÇÃO

A tecnologia na geração de dados está presente em diversos segmentos da sociedade contemporânea e impacta significativamente o cotidiano das pessoas.¹ De forma geral, uma única pessoa pode gerar e/ou transmitir cerca de 3 Mbytes de dados por segundo. A geração de dados atingiu cifras elevadas, onde cerca de 64 Zettabytes (10^{21}) foram gerados no ano de 2020 e estima-se que essa marca possa atingir mais de 180 até 2025.² A área de química não fica em desvantagem neste aspecto, pois a instrumentação analítica moderna aliada ao rápido desenvolvimento tecnológico tem permitido a obtenção de sinais de diversas regiões do espectro eletromagnético e de outras fontes, tais como fluorescência de raios-X (*X-ray fluorescence*, XRF), ultravioleta (*ultraviolet*, UV), visível (*Vis*), infravermelho (*infrared*, IR), de massas, entre outros. Todos esses dados podem ser obtidos em um curto intervalo de tempo e com uma facilidade jamais vista.^{3,4}

A correta organização, transformação e interpretação dos dados em informações é uma tarefa crucial para qualquer tipo de análise. A Quimiometria⁵ e a análise exploratória de dados⁶ têm papel fundamental nesse aspecto com a apresentação de uma gama diversificada de ferramentas, como exemplo, a *principal component analysis* (PCA)^{7,8} que possibilita a projeção dos dados multidimensionais em um número reduzido de variáveis denominadas *principal component* (PC), como mostra a equação 1:^{8,9}

$$X = TL^T \quad (1)$$

onde a matriz X contém os dados de interesse, que são decompostos em duas matrizes: a de *scores* T e a matriz ortonormal de pesos ou *loadings* L. *Scores* e *loadings* trazem informações das amostras e das variáveis, respectivamente. Com a análise exploratória e mineração de dados (*data mining*) é possível também ter uma melhor compreensão das informações, identificando a correlação entre as

variáveis e informações subjacentes. Assim, é possível a identificação das principais características dos dados, possibilitando a tomada de decisões. A PCA¹⁰ ocupa um papel importante neste tipo de tratamento de dados sendo a ferramenta mais comumente empregada na química.

Assim, neste terceiro tutorial de uma série que teve início em 2018,¹¹ e uma continuidade em 2021,¹² o objetivo é mostrar a aplicação da PCA em três exemplos: (i) dados de concentrações de elementos químicos; (ii) espectros de emissão obtidos por *laser-induced breakdown spectroscopy* (LIBS) e de ED-XRF (*energy-dispersive XRF*) e (iii) obtenção de *heat maps* por meio de imagens hiperspectrais.¹³ Os tópicos que serão abordados visam também, complementar as publicações sobre a PCA e anteriormente apresentadas pelos grupos dos professores Poppi,¹⁴ Ferreira,¹⁵ Fragozo¹⁶ e Valderrama.¹⁷

PARTE EXPERIMENTAL**Variáveis provenientes de concentrações de elementos químicos**

Nesta parte do tutorial foram utilizados resíduos eletrônicos (*e-waste*) compostos por computadores (*desktops* e *laptops*), celulares e *tablets*. O resíduo obtido foi desmontado e segregado em diferentes partes: polímeros, cabos e circuitos eletrônicos (*printed circuit board*, PCB). O objetivo foi a determinação de analitos, tais como elementos nobres (Ag, Au, Pt e Pd), convencionais ou metais base (Al, Cu e Fe) e potencialmente tóxicos (As, Cr e Pb) nos PCB's. Ao todo foram obtidos dois *desktops*, quatro *laptops*, sete celulares e três *tablets*. As PCB's foram moídas em moinho de facas (Tecnal TE-650, Piracicaba, Brasil), peneiradas e segregadas em quatro diferentes tamanhos de partícula: (i) < 106 µm; (ii) 106-212 µm; (iii) 212-600 µm e (iv) > 600 µm. Ao final do processo de peneiramento foram obtidas 57 amostras.

O material foi mineralizado em bloco digestor (Marconi MA4025, Piracicaba, Brasil) com o emprego de água régia diluída 1:1 com água desionizada Milli-Q e frascos fechados de tetrafluoroethylene-perfluoro (alkoxy vinyl ether) (PFA). A temperatura de aquecimento

*e-mail: erpf@ufscar.br

dos tubos foi de 95 °C e o tempo de preparo foi da ordem de 2 horas. O extrato ácido resultante foi diluído para 50 mL com água desionizada e as determinações foram executadas em ICP OES (Thermo Fisher, iCap 7000 series, EUA). A Tabela 1 mostra as condições instrumentais empregadas nas determinações por ICP OES. A acurácia dos dados foi verificada por meio da utilização de uma amostra de referência de *e-waste* preparada por Andrade *et al.*^{18,19} Ao final do processo foi possível obter uma matriz X com 57 amostras (n) e 10 variáveis (m) que representam as concentrações dos 10 analitos determinados. Assim, a matriz X terá a dimensão n:m (57:10).

Tabela 1. Condições instrumentais do ICP OES empregado nas determinações dos elementos de interesse e os LoD para cada analito

Parâmetros do instrumento	Condições operacionais	
Tempo de integração (s)	15 a 5	
Potência aplicada (kW)	1,15	
Vazão gás refrigerante (L min ⁻¹)	12,0	
Vazão do gás auxiliar (L min ⁻¹)	0,5	
Vazão do gás no nebulizador (L min ⁻¹)	0,7	
Vazão da amostra (L min ⁻¹)	2,10	
Nebulizador	Mira Mist	
Câmara de estabilização	Ciclônico	
Analito e comprimento de onda	Ag I 328.068, ^a Al I 308.215, ^a	
	As I 189.042, ^b Au I 267.595, ^a	
	Cr I 283.563, ^b Cu I 327.396, ^a	
	Fe II 259.940, ^a Pb I 220.353, ^b	
	Pd I 340.458, ^a Pt I 265.945 ^a	
LoD (mg kg ⁻¹)	Ag - 0,001	Cu - 0,006
	Al - 0,02	Fe - 0,004
	As - 0,01	Pb - 0,01
	Au - 0,004	Pd - 0,002
	Cr - 0,001	Pt - 0,01

^aModo de visualização axial; ^bmodo de visualização radial.

Variáveis provenientes das técnicas LIBS e EDXRF

Os espectros estudados neste tutorial foram compostos por dados espectrais das técnicas analíticas LIBS²⁰ e EDXRF. Para as medidas LIBS das 57 amostras de *e-waste* descritas na seção anterior, foi necessário obter pastilhas por meio de uma prensa (Shimadzu, SSP-10). O diâmetro e a espessura das pastilhas obtidas tiveram cerca de 10 e 2 mm, respectivamente. O instrumento LIBS utilizado foi o modelo J200 (Applied Spectra, Fremont, EUA). As condições operacionais foram: laser de Nd:YAG com comprimento de onda de 1064 nm; *spot size* de 100 µm; *delay time* de 1 µs e energia de 80 mJ. Com os parâmetros descritos foi possível obter uma fluência teórica de cerca de 1000 J cm⁻². Ao todo foram adquiridos 130 espectros por amostra utilizando o modo varredura de análise (*raster*). Para este instrumento, os espectros de emissão abrangem a faixa entre 186 e 1042 nm. O sistema de detecção é constituído por 6 espectrômetros: (i) 186 até 311 nm (resolução de 0,059 nm); (ii) 311 até 465 nm (resolução de 0,073 nm); (iii) 465 até 591 nm (resolução de 0,062 nm); (iv) 591 até 693 nm (resolução de 0,050 nm); (v) 693 até 884 nm (resolução de 0,094 nm) e (vi) 884 até 1042 nm (resolução de 0,079 nm). A quantidade de variáveis registradas é de 12288.

Os resultados empregados nos cálculos subsequentes foram as médias aritméticas de cada conjunto de espectros obtidos por amostra. Assim, foi estudada uma matriz X de 57 amostras por 12288 variáveis (matriz 57:12288).

No caso da EDXRF foi utilizado um espectrômetro NEX QC⁺, (Rigaku, Austin, Texas, EUA). As 57 amostras foram analisadas

em ar e atmosfera de gás hélio (He) em 3 condições instrumentais distintas: (i) número atômico alto (faixa do Ru até o Pr e K até o Br), tensão de 50 kV, corrente 10 µA e filtro tipo B; (ii) número atômico intermediário (faixa do K até o Mo e Sn até o U), tensão 30 kV, corrente 10 µA e filtro tipo A e (iii) número atômico baixo (faixa do Sn até o U e Na até o Cl), tensão 6,5 kV, corrente 50 µA e sem filtro. Os espectros gerados cobriram a faixa de 0 até 49,94 keV (0,024 keV de resolução) e foi possível obter 2048 canais de energia (keV) para cada uma das condições instrumentais descritas anteriormente. Ao final foram organizadas seis matrizes de dados com 57 amostras e 2048 variáveis (matrizes 57:2048). Nos cálculos exemplificados nas próximas seções foram utilizados os dados obtidos em atmosfera de He, pois apresentaram maior sensibilidade.

Imagens hiperespectrais

Imagens hiperespectrais podem gerar informações tanto espaciais quanto espectrais. No caso específico do exemplo deste tutorial cada pixel da imagem foi composto por um espectro de emissão obtido pela LIBS. Desta forma, as imagens hiperespectrais²¹ são bastante úteis para investigar a composição química de uma amostra e a distribuição de determinados compostos ou elementos ao longo da sua superfície.^{22,23} Neste tutorial foi utilizada uma amostra de PCB de placa de som de um dos *laptops* desmontados e, juntamente com a LIBS, a proposta é mostrar e ilustrar o tratamento de dados deste tipo de problema. As condições do laser foram as mesmas descritas anteriormente, mas a aquisição dos espectros foi realizada ponto a ponto, cobrindo uma área de 100 mm². O arranjo de dados obtidos foi 10 por 10 e 12288 variáveis. A Figura 1 mostra a superfície da amostra antes (Figura 1(a)) e após (Figura 1(b)) a obtenção dos espectros. Em cada ponto de incidência do laser (ver marcações na Figura 1(b)) foram dados 10 pulsos e a quantidade final de espectros obtidos foi 10 × 10 × 10 = 1000. Ao incidir diversos pulsos de laser no mesmo ponto da amostra é possível fazer inferências sobre a composição da amostra em diferentes profundidades. Esse tipo de informação é importante para amostras de WEEE que são formadas por várias camadas com propriedades distintas: polímeros, metais, materiais isolantes, entre outros. A matriz X de dados foi de 1000:12288.

Rotinas computacionais utilizadas

Para o desenvolvimento desse terceiro tutorial e cálculos dos dados gerados, foram preparadas cinco rotinas computacionais para os programas Matlab (R2019b, The Mathworks, EUA) e Octave (programa livre). Além disso, foi utilizado o Microsoft Excel[®] para a organização dos dados. Além do Octave, podem ser empregados também os programas livres R e Python. Todas as rotinas estão descritas no Material Suplementar e podem ser implementadas, livremente, nos programas computacionais descritos.

A primeira rotina foi denominada LIBS_{inverse} e tem como objetivo organizar de forma adequada o arranjo de dados das imagens hiperespectrais. Durante a aquisição dos dados para a formação do arranjo, o instrumento LIBS emprega como padrão uma configuração em zigue-zague que não é adequada para o tratamento matemático das rotinas subsequentes. Assim, a LIBS_{inverse} regulariza as linhas de obtenção dos dados fazendo com que cada uma inicie da esquerda para a direita. A linha de comando da LIBS_{inverse} é:

$$[\text{MATRIZ}] = \text{LIBS}_{\text{inverse}}(\text{X}, \text{NP}, \text{NL}, \text{NC})$$

As informações entre parênteses representam os dados de entrada (*input*), onde o parâmetro X representa a matriz com os dados e,

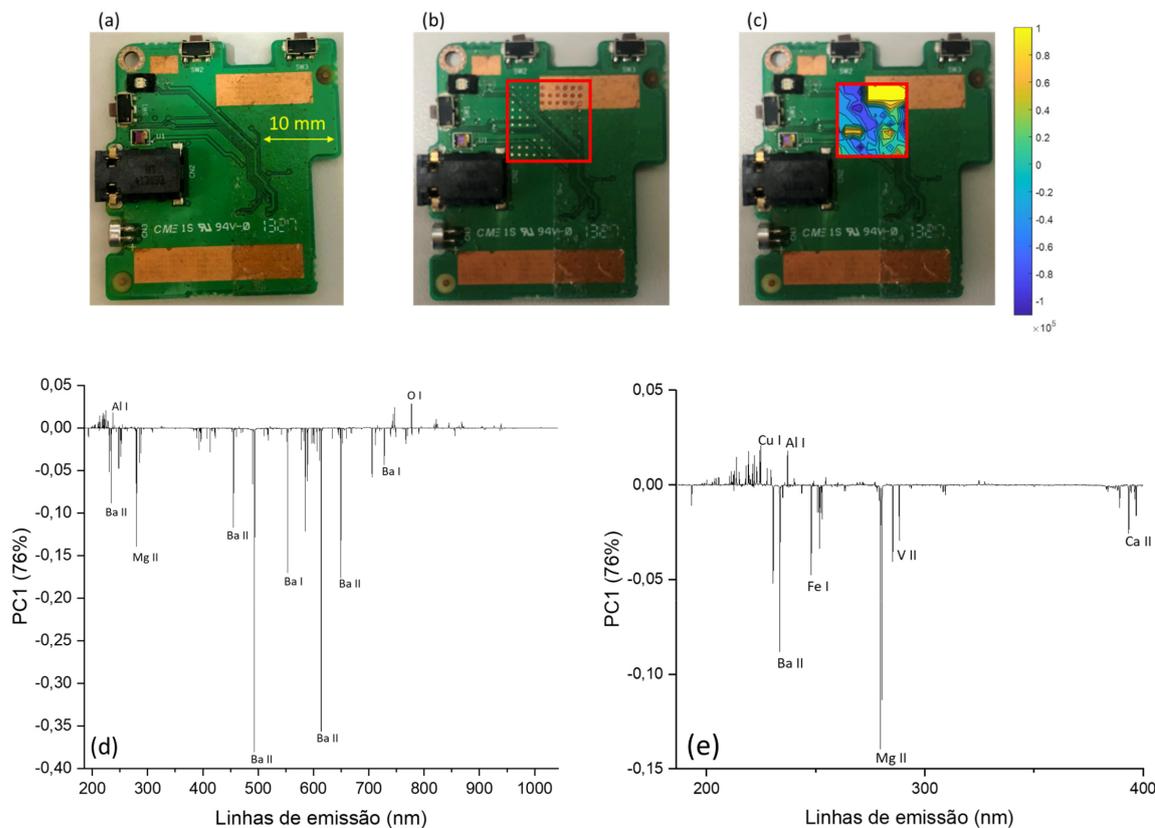


Figura 1. PCB antes do mapeamento elemental no LIBS (a), após o mapeamento (b), o *heat map* dos scores (c) da PC1 e para o pulso 5, os loadings da PC1 (d) e pulso 5 e uma ampliação dos valores de loadings para as linhas emissão de 200 a 400 nm (e)

neste exemplo em específico, as dimensões são 1000:12288. Já os parâmetros NP, NL e NC representam o número de pulsos por ponto, número de linhas e o número de colunas, respectivamente. No exemplo da Figura 1(b) temos que NP, NL e NC são 10, 10, e 10, respectivamente. Como dado de saída (*output*) temos o parâmetro MATRIZ que representa a matriz X inicial com a direção correta da obtenção dos dados.

A segunda rotina é denominada `data_pre` e tem como objetivo calcular a média aritmética, desvio padrão e realizar o pré-processamento dos dados: centrar na média e autoescalar. A linha de comando da `data_pre` é:

```
[Xauto,Xcm,xm,xstd] = data_pre(X)
```

De forma análoga à rotina `LIBS_inverse`, a matriz de dados X (entrada dos dados) será pré-processada. Como resultado teremos duas matrizes: `Xauto` e `Xcm` que representam a matriz X autoescalada e centrada na média, respectivamente. Além disso, são apresentados dois objetos (`xm` e `xstd`) que mostram a média e o desvio padrão das variáveis da matriz X. A escolha da matriz `Xauto` ou `Xcm` para os cálculos subsequentes dependerá da natureza dos dados que estão sendo estudados.

Os dados gerados são então utilizados na terceira rotina, denominada `pca_dis` que efetua os cálculos da PCA, gerando gráficos de variância explicada, *scores* e *loadings*. No caso específico dos *scores* e *loadings*, são mostrados os gráficos para as duas primeiras PC's. As informações das demais PC's podem ser acessadas nas matrizes de saída da função (*scores* e *loads*). A linha de comando da `pca_dis` é:

```
[scores,loads,var_exp] = pca_dis (X)
```

A matriz X representa os dados de entrada e as matrizes *scores* e *loads* mostram os valores de *scores* (n:n) e *loadings* (m:n) calculados. O vetor `var_exp` mostra a variância explicada para as n PC's (o vetor é n:1). Na análise dos resultados obtidos pela PCA devem ser utilizadas as informações das PC's que apresentam mais informações. Na decisão sobre quantas PC's serão avaliadas, é importante verificar, a partir de qual, a porcentagem de variância explicada apresenta uma queda monotônica.

A quarta rotina denominada `libs` efetua cálculos que geram mapas de *scores* ou *heat maps* em tensor 4D na forma: número de pulsos (NP) versus PC's versus número de linhas (NL) versus número de colunas (NC). A linha de comando da `libs` é:

```
[MAPS,T,P,Var_exp] = libs (MATRIZ,NP,NC,NL,PCS)
```

Na rotina `libs` é calculada uma PCA para cada matriz de pulso obtida. No exemplo específico da Figura 1(b) foram calculadas 10 PCAs (uma para cada pulso de laser). Assim, é possível verificar também as diferenças na composição química da amostra, ao passo que são obtidas informações sobre a superfície (pulso 1) e o seu interior (pulso 10). Os dados necessitam ser pré-processados antes da utilização da rotina `libs`.

A quinta rotina denominada `libs_plot` é responsável por gerar gráficos *heat maps* provenientes da rotina anterior. `MAPS` é o tensor gerado na rotina anterior, `PULSO` e `PCS` são referentes ao pulso e a quantidade de PC's que serão visualizadas. A linha de comando da `libs_plot` é:

```
libs_plot(MAPS,PULSO,PCS)
```

Para verificar o mapa de *scores* do pulso 2 e das três primeiras PCs, por exemplo deve ser utilizado: `libs_plot(MAPS,2,3)`.

RESULTADOS E DISCUSSÃO

Variáveis provenientes de concentrações de elementos químicos

A Tabela 1 mostra as condições instrumentais do instrumento ICP OES e os limites de detecção (*limits of detection*, LoD) para os 10 analitos determinados. A Tabela 1S, no Material Suplementar, mostra todos os valores de concentrações obtidos. Assim, o leitor poderá reproduzir os cálculos. Além disso, foram preparados vídeos (ver próximas seções) para auxiliar na utilização dos programas computacionais e das rotinas preparadas.

As faixas de concentração variaram de 0,01 mg kg⁻¹ para o Au e Pd e até 40% (m/m) para o Cu. As concentrações possuíam magnitudes em faixas diferentes e, assim, os dados foram autoescalados, como mostra a Equação 2:

$$x_{auto} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2)$$

onde $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$ e $s_j = \sqrt{\frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}{n-1}}$ são a média e o desvio padrão da coluna dos dados (variáveis), respectivamente.

A Figura 1S, no Material Suplementar, mostra os dados originais (Figura 1S(a)) e autoescalados (Figura 1S(b)) para os 10 analitos determinados. Os elementos Ag, As, Au, Pd e Pt apresentaram as menores concentrações e estão representados na unidade de concentração de mg kg⁻¹. Já o Al, Cr, Cu, Fe e Pb estão em faixas superiores e, assim, foram representados em % m/m.

Na Figura 1S(a) a distribuição estatística das amostras possui características assimétricas²⁴ para quase todos analitos, com exceção

da Ag, Al e Cu, este parâmetro informa quão semelhantes são os valores de média, mediana (valor localizado no centro do conjunto de dados) e moda (valor que mais se repete) de cada uma das variáveis. Outro parâmetro importante que pode ser discutido é a curtose.²⁴ A curtose descreve se os dados apresentam uma distribuição achatada (valores baixos de curtose) ou em cume (valores altos de curtose). Para os dados apresentados na Figura 1S temos que as distribuições mais em cume são as do Cr e Pt e a mais achatada é a do Cu. A Figura 1S(b) mostra uma representação dos dados autoescalados e nota-se que a média (quadrado no interior do *box-plot*) e o desvio padrão são iguais a 0 e 1, respectivamente. Assim, todas as variáveis possuem a mesma escala e podem contribuir de maneira semelhante para os cálculos dos *scores* e *loadings*. Dados de concentração como os apresentados nesta seção devem ser, preferencialmente, autoescalados.⁹ Outros exemplos de dados autoescalados são propriedades físico-químicas.²⁵⁻²⁷

Com os dados autoescalados obtidos foi calculada uma PCA e a Figura 2 mostra os resultados obtidos. As Figuras 2(a) e 2(b) mostram os valores de *scores* onde as classes investigadas foram o tipo de *e-waste* e tamanho de partícula, respectivamente. Na Figura 2(c) temos os *loadings*. Com os *scores* e *loadings* é possível fazer inferências sobre o comportamento das amostras. No caso da Figura 2(a), por exemplo, temos um certo agrupamento das amostras de telefones celulares (triângulos azuis) e *tablets* (triângulos invertidos verdes). Para a Figura 2(b), temos que as amostras com tamanho de partícula > 600 μm (triângulos invertidos verdes) apresentam um comportamento distinto das demais. Com o gráfico de *loadings* (Figura 2(c)) podemos verificar as correlações entre as variáveis. As variáveis Cr, Fe e Pt, por exemplo, apresentam valores de coeficiente de correlação (*r*)²⁴ variando de 0,6599 (Fe *versus* Pt) até 0,9283 (Cr *versus* Pt), ou seja, os valores de *r* são positivos

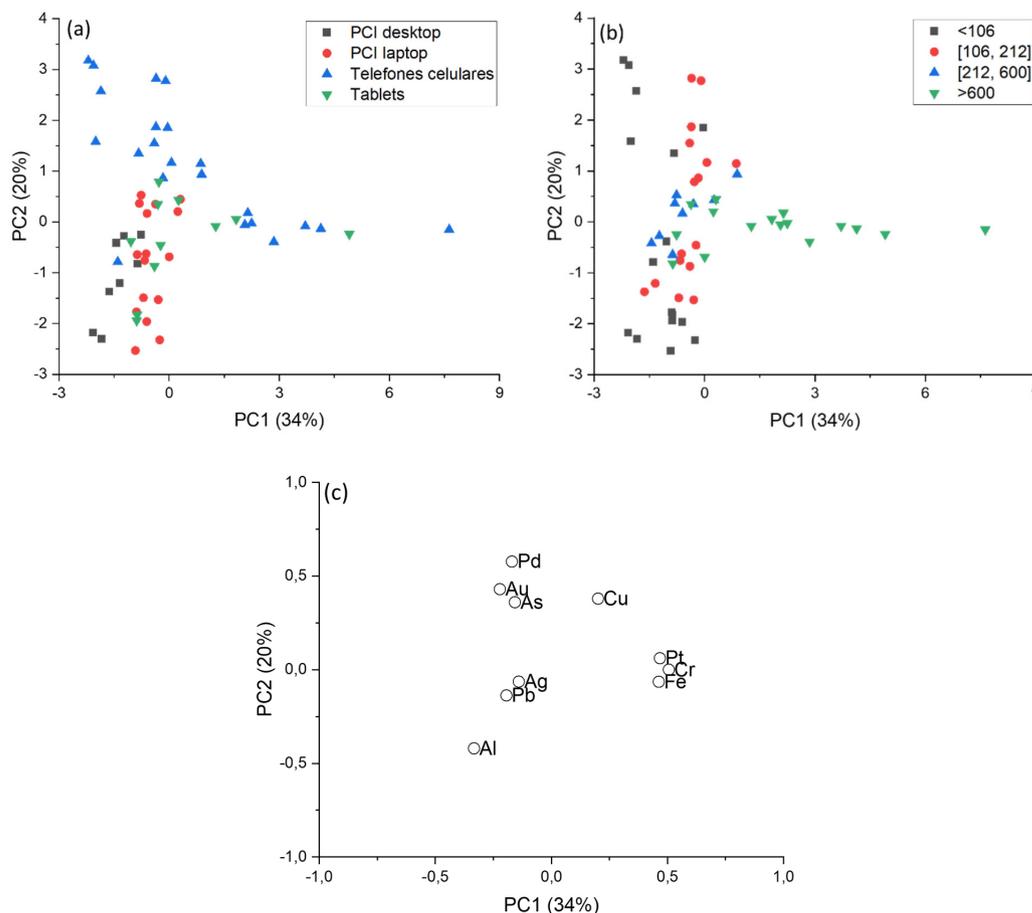


Figura 2. Scores das concentrações determinadas em função do tipo de amostra (a), tamanho de partícula (b) e seus respectivos valores de loadings (c)

e isso pode denotar fontes semelhantes dos 3 elementos. Já os elementos Al e Cu apresentaram um r negativo, igual a $-0,7025$, correspondendo ao fato de amostras que apresentaram concentrações altas de Al, tiveram concentrações baixas de Cu e vice-versa. Além de correlações entre os elementos, é possível também observar quais amostras possuem concentrações mais elevadas de alguns elementos. As amostras situadas no quadrante negativo da PC1 e PC2 da Figura 2(a), por exemplo, apresentam concentrações mais altas de Al (também presente no quadrante negativo da PC1 e PC2 para o gráfico de *loadings*, Figura 2(b)). De fato, tais amostras apresentam concentrações de Al três vezes maiores que as demais (ver Tabela 1S do Material Suplementar).

Variáveis provenientes das técnicas LIBS e EDXRF

Os dados de espectros obtidos por LIBS e EDXRF foram centrados na média. As Figuras 3 e 4 mostram espectros típicos da LIBS e de EDXRF, respectivamente, para uma amostra de PCB de telefone celular.

Na Figura 3 é verificado que os espectros de emissão da LIBS apresentam maiores intensidade para os elementos Ba, Ca, Cu, Mg e Na. A Figura 4 apresenta os três espectros de EDXRF: (i) alto número atômico (Figura 4(a)), apresentou maiores intensidades para os elementos Br, Cu e Sn; (ii) número atômico intermediário (Figura 4(b)), apresentou maiores intensidades para os elementos Br, Cu e Pb e (iii) baixo número atômico (Figura 4(c)), apresentou maiores intensidades para os elementos Ag, Al, Ca e Si. Os espectros mostrados na Figura 4 foram obtidos em atmosfera de He.

Os elementos Ag e Cu fazem parte dos circuitos eletrônicos das amostras de PCB.²⁸ Já os elementos Pb e Sn são encontrados nas soldas.²⁶ O Al e o Ba estão presentes em dispositivos da PCB e em capacitores, respectivamente. Já os elementos Ca, Mg e Si fazem parte do isolamento das PCBs. Compostos de Br são constituintes de retardantes de chama.^{26,28-30}

Os espectros LIBS foram organizados, e uma PCA foi calculada, como mostra os *scores* na Figura 5(a) e os *loadings* na Figura 5(b). O instrumento LIBS possui uma maior sensibilidade para diferenciar os tamanhos de partícula das amostras quando comparado ao EDXRF. Isto fica evidente na Figura 5(a), com o resultado obtido na PCA é possível identificar diferenças nos padrões de espectro de emissão para os diferentes tamanhos de partículas. As amostras com tamanho de partícula $> 600 \mu\text{m}$ se agruparam a esquerda (triângulos verdes invertidos), as com tamanho $< 106 \mu\text{m}$ (quadrados pretos) e $[106, 212] \mu\text{m}$ (círculos vermelhos) possuem padrões espectrais semelhantes e se agruparam mais à direita, enquanto as com tamanho $[212, 600] \mu\text{m}$ (triângulos azuis) se agruparam mais ao centro da Figura 5(a). Nos *loadings* (Figura 5(b)) foi observado que os elementos presentes com maior sinal foram o Ba e o Mg e que ambos estão positivamente correlacionados.

Os espectros de EDXRF foram organizados da mesma forma e foi escolhida a condição (i) alto número atômico e atmosfera de He, como exemplo, para calcular a PCA. O resultado obtido é apresentado na Figura 6, com os *scores* (Figura 6(a)) e os *loadings* (Figura 6(b)). Nos *scores* (Figura 6(a)) houve um leve agrupamento por diferença de tamanho de partícula e foi notado uma anormalidade quanto aos espectros das PCBs de celulares, que se agruparam independente do seu tamanho de partícula. Esse comportamento é devido aos sinais dessas amostras serem muito semelhantes e se destacarem das demais. Nos *loadings* (Figura 6(b)) os elementos com maiores contribuições foram o Br, Cu, Fe e Sn. O Cu apresentou correlação negativa com os demais elementos, principalmente o Br, como mostra a PC1. Portanto as amostras que estão do lado positivo da PC1 na Figura 6(a) possuem maiores concentração Cu e menores de Br.

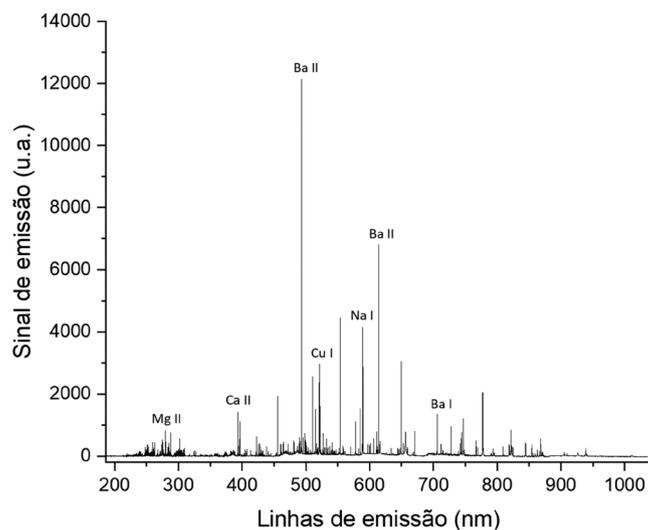


Figura 3. Espectro de emissão LIBS de uma amostra de celular moída e prensada

Imagens hiperespectrais

A Figura 1(c) mostra o mapa de *scores* da PC1 para o pulso 5. A PC1 para o pulso 5 apresentou uma variância explicada de 76% e foi a que melhor representou espacialmente a amostra real (Figuras 1(a) e 1(b)). As demais PC's mostraram valores de variância explicada muito baixos (da ordem de 23%). A Figura 2S mostra os valores de variância explicada (%) para as 10 primeiras componentes principais e nota-se que a partir da PC2 há uma queda monotônica na porcentagem de variância.

Na Figura 1(c) temos uma escala de cores que representa a magnitude dos valores de *scores*. Os dados foram centrados na média e os valores positivos e negativos de *scores* são representados pelas cores amarela e azul, respectivamente. Assim, valores positivos de *loadings* (Figura 1(d)) representam elementos químicos identificados na parte em amarelo da Figura 1(c). Já os valores negativos de *loadings*, representam elementos identificados na parte azul da amostra. A Figura 1(e) mostra uma ampliação da região de 200 a 400 nm sendo possível notar que os elementos Al e Cu estão presentes na parte alaranjada da PCB (ver Figuras 1(a) e 1(b)). Já os demais elementos, Ba, Ca, Fe, Mg e V, estão presentes na parte polimérica da PCB. Tais informações são úteis para efetuar uma análise qualitativa da composição química da amostra e, no caso específico do resíduo eletrônico, propor rotas diferentes de segregação e reciclagem, viabilizando a recuperação de elementos, podendo ser aplicadas em escala industrial. A estratégia ilustrada permite a análise e reconhecimento de padrões no processamento de resíduos eletrônicos, facilitando a avaliação e o controle industrial.

O uso de imagens hiperespectrais e LIBS associadas com a geração de mapas de *scores* constituem ferramenta importante para a obtenção de informações sobre a composição da amostra, sendo utilizada em diversas aplicações, tais como a classificação de alimentos³¹ e análise de minerais.³²

CONCLUSÕES

Com este tutorial é possível aprender de maneira didática como utilizar a ferramenta PCA para análises de variáveis provenientes de concentrações e de espectros (de emissão e de fluorescência de raios-X). Além disso, foram também elaboradas as rotinas para outras linguagens e programas computacionais, tais como Python e R (mais detalhes são apresentados no Material Suplementar). Os

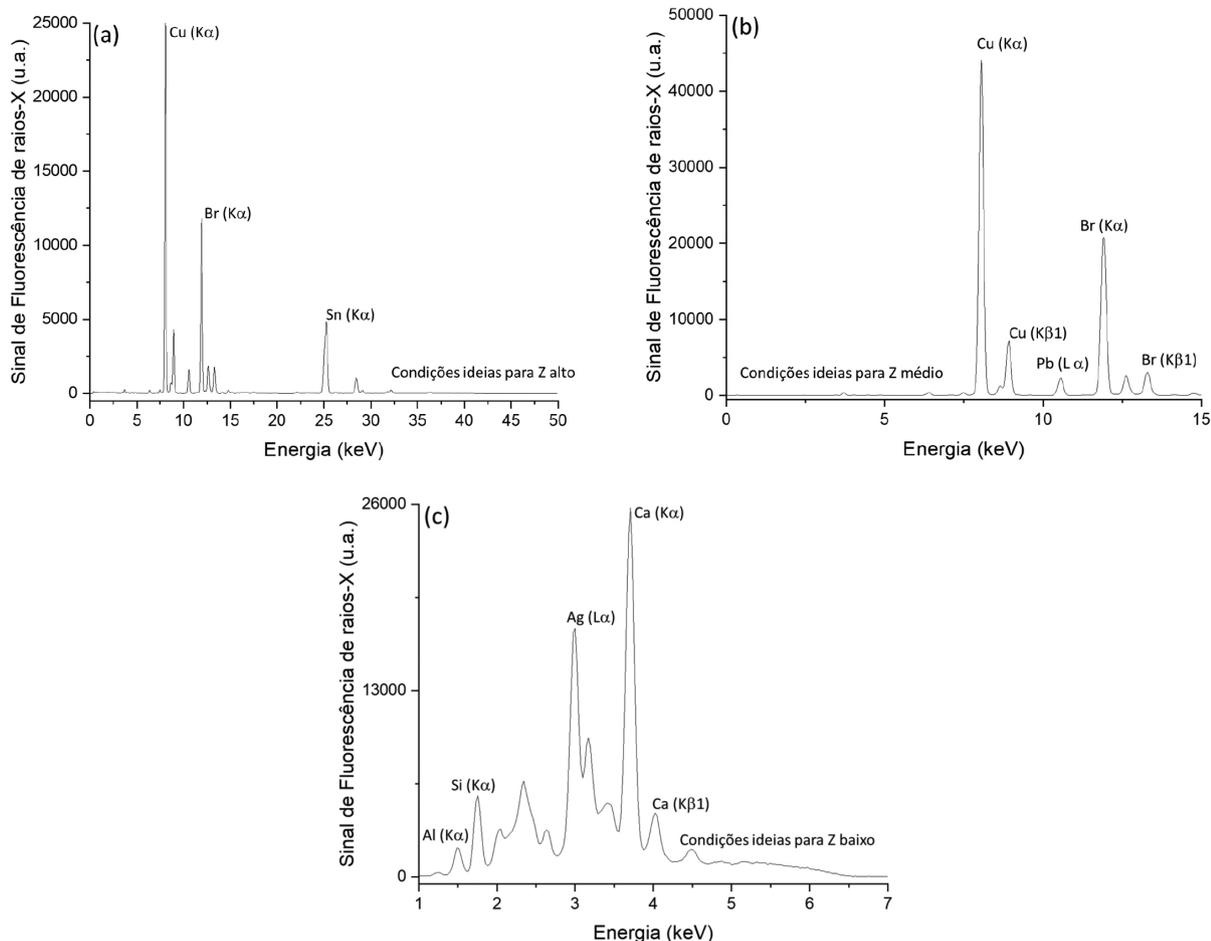


Figura 4. Espectros de fluorescências de raios-X de uma amostra de celular em atmosfera de He, nas condições (i) número atômico alto (a), (ii) número atômico intermediário (b) e (iii) número atômico baixo (c)

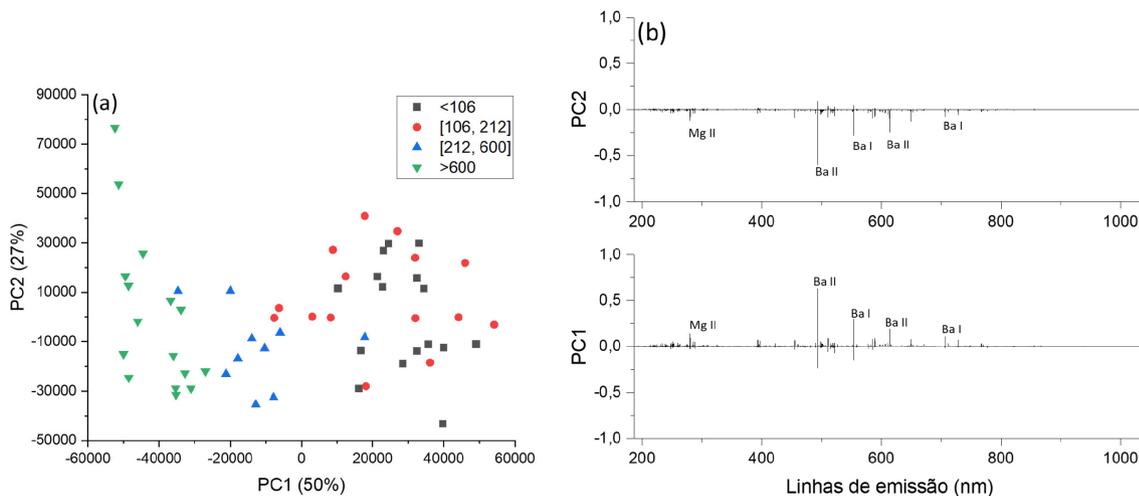


Figura 5. Scores (a) e loadings (b) dos espectros de emissão LIBS das 57 amostras em função do tamanho de partícula

exemplos mostrados neste terceiro tutorial podem ser visualizados de forma detalhada na playlist 28 do canal no YouTube do Prof. Edénir R. Pereira Filho (<https://www.youtube.com/c/EdenirPereiraFilho>). Nesta playlist são disponibilizadas também rotinas computacionais em R e Python e o leitor pode reproduzir os cálculos. Além disso, é mostrada também a utilização do Octave online, que não requer a instalação do programa no computador do analista. A playlist 28 visa complementar as informações já divulgadas nas playlists de 22 até 25 e a lista a seguir mostra uma descrição das mesmas:

Playlist 22 apresenta o programa livre Octave e os detalhes de como utilizá-lo:

https://youtube.com/playlist?list=PL4CuftF41_fAzIX7O373iO-Sf_BHR1_Zd.

Playlists 23 e 24 mostram o programa R e a linguagem Python, respectivamente, e as suas utilizações em planejamento fatorial:

https://youtube.com/playlist?list=PL4CuftF41_fCL_NicWd5Ax5xWIEUn7Zd3 (R) e

https://youtube.com/playlist?list=PL4CuftF41_

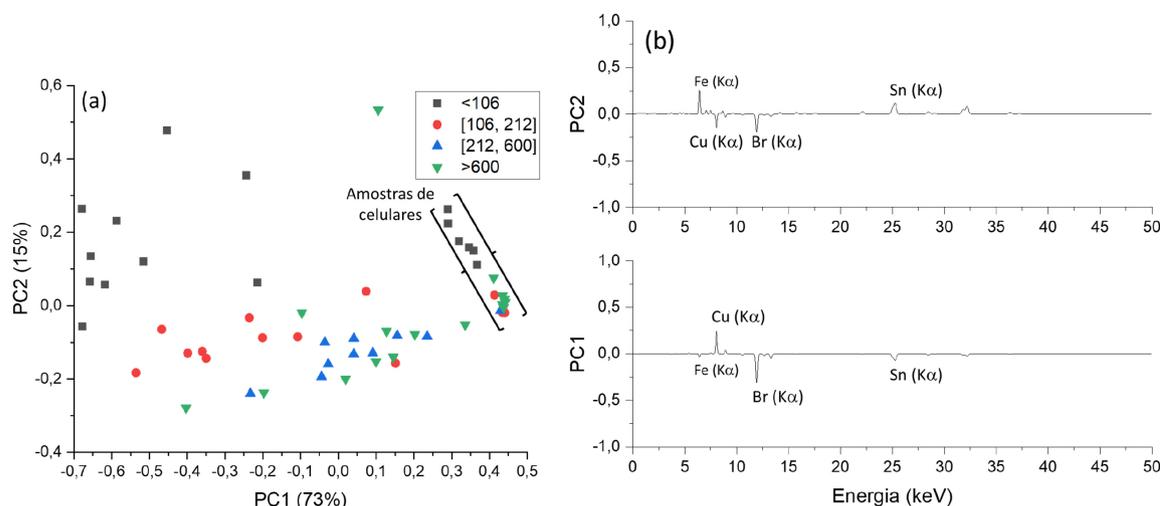


Figura 6. Scores (a) e loadings (b) dos espectros de fluorescência de raios-X das 57 amostras em função do tamanho de partícula

fCsdCBY4eaLcTDFhhMaH6VN (Python).

Playlist 25 apresenta diversos detalhes sobre os cálculos efetuados com a PCA empregando o Octave:

https://youtube.com/playlist?list=PL4CuftF41_fAoYAae3a6sucVbfS2f0GiS.

Playlist 28 apresenta os exemplos deste tutorial:

https://youtube.com/playlist?list=PL4CuftF41_fAsDkW2SEIgtQ9QqV5fPmB.

MATERIAL SUPLEMENTAR

O Material Suplementar apresenta os seguintes itens:

- Figura 1S. *Box-plot* das concentrações dos 10 analitos determinados: dados originais (a) e autoescalados (b).
- Figura 2S. Porcentagem de variância explicada em função do número de componentes principais para o pulso 5.
- Rotinas computacionais: (i) LIBS_inverse; (ii) data_pre; (iii) pca_dis; (iv) libs e (v) libs_plot.
- Tabela 1S. Tabela das concentrações das amostras de *e-waste* determinados em ICP OES.

Todo conteúdo está disponível em <http://quimicanova.sbj.org.br>, na forma de arquivo PDF, com acesso livre.

AGRADECIMENTOS

Os autores são gratos à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processos 2016/17221-8, 2019/24223-5, 2021/10882-7 e 2022/02232-5) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processos 140867/2021-0, 302719/2020-2 e 307328/2019-8). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- Berisha, B.; Mëzhu, E.; Shabani, I.; *Journal of Cloud Computing* **2022**, *11*, 24. [Crossref]
- Hastie, T.; Tibshirani, R.; Friedman, J.; *The Elements of Statistical Learning Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, 2009.
- Artavia, G.; Cortés-Herrera, C.; Granados-Chinchilla, F.; *Foods* **2021**, *10*, 1081. [Crossref]
- Tavares, T. R.; Molin, J. P.; Nunes, L. C.; Alves, E. E. N.; Krug, F. J.; de Carvalho, H. W. P.; *Data Brief* **2022**, *41*, 108004. [Crossref]
- Neto, B. D. B.; Scarminio, I. S.; Bruns, R. E.; *Quim. Nova* **2006**, *29*, 1401. [Crossref]
- Brandt, S.; *Data Analysis: Statistical and Computational Methods for Scientists and Engineers*, 3rd ed.; Springer: New York, 1998.
- Beebe, K. R.; Pell, R. J.; Seasholtz, M. B.; *Chemometrics: A Practical Guide*, 1st ed.; John Wiley & Sons: New Jersey, 1998.
- Wold, S.; Esbensen, K.; Geladi, P.; *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37. [Crossref]
- Ferreira, M. M. C.; *Quimiometria: Conceitos, Métodos e Aplicações*, 1^a ed.; Editora da Unicamp: Campinas, 2015.
- Bro, R.; Smilde, A. K.; *Analytical Methods* **2014**, *6*, 2812. [Crossref]
- Manhas, F. M. V.; Pereira-Filho, E. R.; *Quim. Nova* **2018**, *41*, 1061. [Crossref]
- Hilário, F. F.; Castro, J. P.; Barros, T. E.; Pereira-Filho, E. R.; *Quim. Nova* **2021**, *44*, 874. [Crossref]
- Amigo, J. M.; Babamoradi, H.; Elcoroaristizabal, S.; *Anal. Chim. Acta* **2015**, *896*, 34. [Crossref]
- de Souza, A. M.; Poppi, R. J.; *Quim. Nova* **2012**, *35*, 223. [Crossref]
- Ferreira, M. C. F.; *Quim. Nova* **2022**, *45*, 1251. [Crossref]
- Lyra, W. S.; Silva, E. C.; Araújo, M. C. U.; Frago, W. D.; Veras, G.; *Quim. Nova* **2010**, *33*, 1594. [Link] acessado em fevereiro 2023
- Valderrama, L.; Paiva, V. B.; Março, P. H.; Valderrama, P.; *Quim. Nova* **2016**, *39*, 245. [Crossref]
- Andrade, D. F.; Machado, R. C.; Pereira-Filho, E. R.; *J. Anal. At. Spectrom.* **2019**, *34*, 2402. [Crossref]
- Andrade, D. F.; Machado, R. C.; Bacchi, M. A.; Pereira-Filho, E. R.; *J. Anal. At. Spectrom.* **2019**, *34*, 2394. [Crossref]
- Pereira-Filho, E. R. Em *Laser-Induced Breakdown Spectroscopy (LIBS): Applications and Calibration Strategies*; Cruz, J. A. S., ed.; Editora Ibero-Americana de Educação: Bauru, 2021.
- Carneiro, R. L.; Poppi, R. J.; *Spectrochim. Acta, Part A* **2014**, *118*, 215. [Crossref]
- Chang, C. I.; Song, M.; Zhang, J.; Wu, C. C.; *Remote Sens.* **2019**, *11*, 2012. [Crossref]
- Tang, S.; Chen, Z.; Zhang, M.; *11th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Amsterdam, 2021. [Crossref]
- Hash-Vaughn, D. L.; Lomax, R. G.; *An Introduction to Statistical Concepts*, 4th ed.; Routledge: United Kingdom, 2020.
- Andrade, D. F.; de Almeida, E.; de Carvalho, H. W. P.; Pereira-Filho, E. R.; Amarasiriwardena, D.; *Talanta* **2021**, *225*, 122025. [Crossref]
- Castro, J. P.; Pereira-Filho, E. R.; *Detritus* **2020**, *13*, 131. [Crossref]

27. Ciulu, M.; Serra, R.; Caredda, M.; Salis, S.; Floris, I.; Pilo, M. I.; Spano, N.; Panzanelli, A.; Sanna, G.; *Talanta* **2018**, *190*, 382. [Crossref]
28. Carvalho, R. R. V.; Coelho, J. A. O.; Santos, J. M.; Aquino, F. W. B.; Carneiro, R. L.; Pereira-Filho, E. R.; *Talanta* **2015**, *134*, 278. [Crossref]
29. Costa, V. C.; Castro, J. P.; Andrade, D. F.; Babos, D. V.; Garcia, J. A.; Sperança, M. A.; Catelani, T. A.; Pereira-Filho, E. R.; *TrAC, Trends Anal. Chem.* **2018**, *108*, 65. [Crossref]
30. Andrade, D. F.; Castro, J. P.; Garcia, J. A.; Machado, R. C.; Pereira-Filho, E. R.; Amarasiriwardena, D.; *Chemosphere* **2022**, *286*, 131739. [Crossref]
31. Ning, J.; Sun, J.; Li, S.; Sheng, M.; Zhang, Z.; *Int. J. Food Prop.* **2017**, *20*, 1515. [Crossref]
32. Moncayo, S.; Duponchel, L.; Mousavipak, N.; Panczer, G.; Trichard, F.; Bousquet, B.; Pelascini, F.; Motto-Ros, V.; *J. Anal. At. Spectrom.* **2018**, *33*, 210. [Crossref]