

Enhancing EDXRF Quantification Through Automated Matrix Determination

Pedro Manguinhas,^a Pedro Catalão Moura,^a João Silva,^a António Alberto Dias,^a Sofia Barbosa,^b and Sofia Pessanha^{a,*}

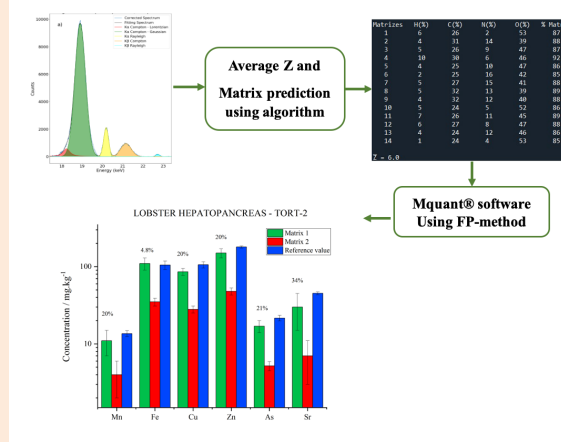
^a Laboratory for Instrumentation, Biomedical Engineering and Radiation Physics (LIBPhys-UNL), Department of Physics, NOVA School of Science and Technology, NOVA University of Lisbon, Campus FCT-UNL, 2829-516, Caparica, Portugal

^b GeoBioTec (UNL) – GeoBiosciences, GeoTechnologies and GeoEngineering, Department of Earth Sciences, NOVA School of Science and Technology, Campus FCT-UNL, 2829-516, Caparica, Portugal

Received: February 02, 2025; Revised: March 12, 2025; Accepted: March 15, 2025; Available online: March 15, 2025.

DOI: 10.46770/AS.2025.008

ABSTRACT: Energy dispersive X-ray fluorescence spectrometry has been widely used for the analysis of trace and minor elements in applications that extend from biomedical to environmental assessment, due to its non-destructive nature, rapid analysis and suitable detection limits. However, EDXRF quantification is frequently hampered by matrix effects, introducing significant inaccuracies in the obtained results. In this work, we present an automated methodology for sample matrix determination using EDXRF spectra, leveraging on the analysis of the Compton and Rayleigh peaks of the characteristic lines of the X-ray tube anode. First, a model was created to fit these scattering peaks, allowing the plotting of a calibration curve that correlated the Compton-to-Rayleigh ratio with the average atomic number (Z) of the sample. Matrix composition was quantified using a developed algorithm combining support vector regression (SVR) and bootstrapping to optimize the determination of the best matrix composition. SVR with a Radial Basis Function kernel was applied to handle non-linear data, and Bootstrapping was utilized to train the algorithm, enhancing model generalization. The study demonstrates that the developed algorithm and matrix-based approach effectively quantified elemental compositions across various certified reference materials (CRMs). The chosen Matrix provided more accurate results, especially for heavier elements like Fe, Cu, and Zn, while deviations of around 20% were observed for lighter elements in biological matrices. In geological samples like phosphate rock and clay, heavier elements aligned well with reference values, but trace elements like Cu and Ni showed larger deviations due to low concentrations. Despite discrepancies for some elements like Pb in wood, the methodology proved effective, particularly for elements like Cr with minimal deviation, highlighting its versatility across diverse matrices. The methodology successfully integrated computational tools and open-source libraries to establish a reliable, efficient workflow for average atomic number determination and spectral analysis, achieving strong agreement with reference materials.



INTRODUCTION

Several contemporary fields are in constant pursuit for newer, faster, more accessible and reliable methodologies that allow for accurate elemental quantification of varied matrices. X-ray

Fluorescence spectrometry technique has been often used in the determination of the chemical composition of many samples due to characteristics like analytical simplicity, rapidness of analysis,

non-invasive and non-destructive nature, and good detection limits for a large range of elements in many fields of application. These fields comprise environmental research¹ and precision farming,² biomedical research and the identification of disease biomarkers,³ geological research, such as the classification of samples⁴ or the identification of rare earth elements,⁵ or even industrial applications, like the classification of wood⁶ samples or other materials of manufacturing relevance.⁷

While the qualitative analysis of XRF spectra is quite straightforward, the quantification procedures present many challenges.^{8,9} Absorption and fluorescence by the matrix often affects the measured fluorescence of the analyte, and for this reason, XRF is often subject to significant matrix effects. Strategies for accounting for matrix effects include sample dilution, standard addition or the use of matrix matched certified reference materials (CRM) for calibration.¹⁰ However, these methods require either sample destruction/alteration or the existence of a significant and suitable number of CRMs for the development of a calibration curve. This requisite often has a preponderant impact on the accuracy of the results obtained using calibration curves for quantification.^{10,11} Some mathematical procedures have been exploited as a way of encompassing the matrix effects, namely, matrix correction methods – the Fundamental Parameters method.^{11–13} This method based on the Sherman's equation¹⁴ in which the mathematical model calculates the expected intensity for an element in a given matrix and compares it with the values obtained in the real sample, due to the non-linear nature of the equations, the elemental concentrations are calculated by iteration. An estimate of the starting set of elemental concentrations, usually from concentrations of similar samples, is required to make the iteration converge.¹³ However, the major shortcoming of the Fundamental Parameters approach is that prior knowledge of the sample areal density (g/cm²) is required.^{15,16} Many approaches to simultaneously determine the areal density of a sample and its elemental composition have been proposed, including adding an internal standard¹⁷ or combining XRF with other techniques.¹⁸ Some softwares require the establishment of an approximate initial composition for the matrix.^{19,20} Rindby²¹ developed the XRFAES (X-ray Fluorescence Automatic Evaluation System) package, a FP-based software for the evaluation of XRF spectra. In this software, the intensity of the coherent (Rayleigh) and incoherent (Compton) scattering peaks is used for the estimation of the light-element fraction of the sample. Other commercial systems have inbuilt calibrated curves for certain generic matrices, namely, soil,²² mining,²³ or Environmental samples.²⁴ However, these methods are often used as black-boxes and quantifications cannot be taken as accurate for matrices that deviate from the calibrated one or need correction with other techniques²⁵ or with Machine Learning algorithms.²⁶

This work aims at developing an automated, rapid and effective methodology capable of determining the matrix of unknown

samples by analyzing and exploring a specific region of the EDXRF spectrum. To do so, the developed methodology aims at automatically determine the matrix of unknown samples considering the scattering peaks of the characteristic lines of the X-ray tube anode, and the average atomic number (Z) to plot calibration curves correlating these values with the Compton-to-Rayleigh ratio¹⁸. The method was applied to different matrices, of biological, geological and industrial interest.

EXPERIMENTAL

Samples. In order to establish a calibration curve for the atomic number as a function of the Compton-to-Rayleigh scattering peaks, a set of 21 model samples consisting on different proportions of reference materials of HAp [$\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$] (Sigma-Aldrich, lot #BCBS8492V), and Boric Acid [H_3BO_3] (for conservation-restoration purposes) was created. The different proportions ranged from 100% Boric Acid (Average $Z=7.1$) adding 5% HAp up to 100% HAp (Average $Z=14.2$). Powders were ground with a mortar and pestle and 3 replicas of each model sample were obtained in order to reduce uncertainty due to inhomogeneity. The average atomic number, $\bar{Z}$, of each material has been calculated according to the equation (1):

$$\bar{Z} = \sum_i w_i Z_i \quad (1)$$

Where w_i the mass fraction and Z_i the atomic number of the element, i . The list of model samples with corresponding $\bar{Z}$ is presented in Pessanha *et al.*²⁷

Furthermore, a total of 13 biological (soft tissues), geological (soils, sediments and phosphate rocks), and industrial (wood) Certified Reference Materials (CRMs) were considered during this work (Table S1).

All CRMs were prepared as pellets by compressing ~2g of powder with a 10-ton manual hydraulic press (Specac, UK), into pellets with a diameter of 20 mm and a thickness of ~3 mm. The obtained thickness was a compromise between the needed thickness to consider the pellets infinitely thick considering the system's geometry, average energy of incoming radiation, the different matrices and the characteristic lines of the elements, and the use of enough material to have a stable pellet. Furthermore, for the geological CRMs 1 drop of mowiol agglutinate was added to the powder inside the press, similarly to the procedure of preparing unknown geological samples.

Energy Dispersive X-Ray Fluorescence Instrumentation. A benchtop micro-XRF spectrometer, the M4 TORNADO by Bruker (Berlin, Germany) was used. The system makes use of a low-power X-ray tube with a Rh anode, operated at 50 kV and 300

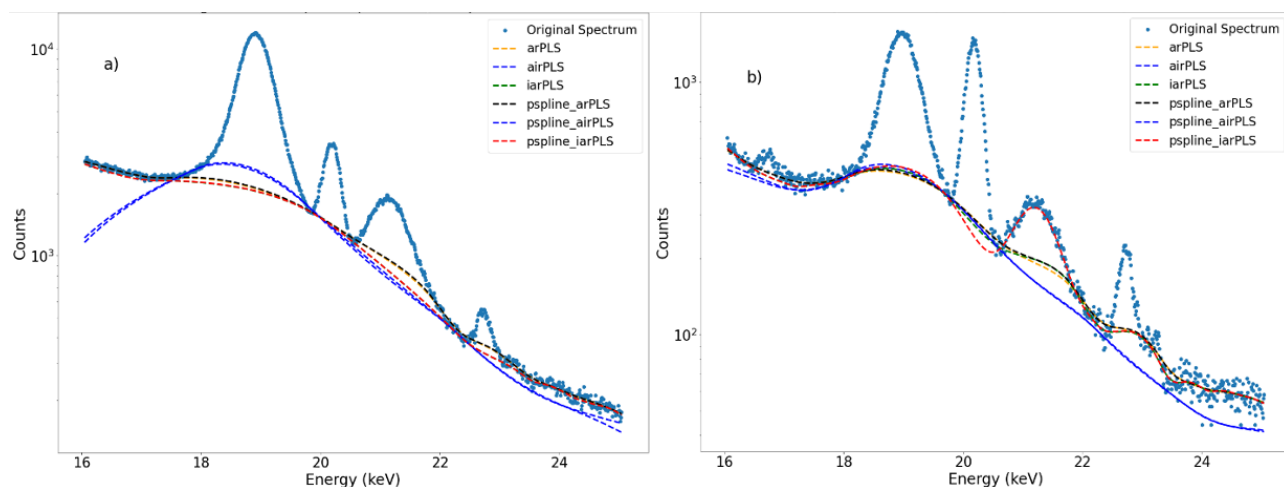


Fig. 1 Comparison of all the baseline correction procedures applied to two samples. a) 10%hydroxyapatite - 90%boric acid and b) phosphate rock.

μA and a poly-capillary lens focused the beam on a $40\ \mu\text{m}$ spot for Mo-K α emission lines. As a way of minimizing eventual contaminations due to the presence of atmospheric elements like argon (Ar) and improve the detection limits, the analysis chamber was kept at a vacuum pressure of 20 mbar. Additionally, a $12.5\ \mu\text{m}$ Al filter was used between the X-ray tube and the samples to better improve the signal-to-noise ratio of the lower region of the spectrum.²⁸

The analyses were performed using an area of interest of $1.4 \times 1.4\ \text{mm}^2$ and point-by-point measurements were performed on the samples with a step size of $20\ \mu\text{m}$ and 30 ms acquisition time per pixel, and the cumulative spectrum was retrieved. Two areas were mapped by pellet and a total of four and six replicates were measured for model and CRMs samples, respectively, in an overall analysis time of 60 h. Considering the different matrices under study, Table S2 presents the Limits of Detection calculated according to equation (2):

$$LoD = 3 * C_i * \frac{\sqrt{N_b}}{N_p} \quad (2)$$

Where C_i is the concentration of element, i , in the CRM, N_b are the net counts of the background and N_p are the net counts of the peak, for the given element, i .

Spectra pre-processing, fitting and determination of the average atomic number. All the developed code was implemented in the interface *Spyder*® using *Python* programming language and applying two major libraries; *NumPy* (useful to treat arrays and perform more complex mathematical calculations) and *Matplotlib* (to plot and visualize graphs).

Baseline correction. Among the several options for baseline correction, one of the most used procedures was developed by

Whittaker²⁹ adapted to the penalized least squares (PLS) method by Eilers.³⁰ This approach allows to correct the baseline by penalizing or minimizing the spectrum roughness and as some important advantages that contribute to its suitability and simplicity, namely, the fact of being capable of smoothing the spectrum by considering a single varying parameters, its capacity of dealing with missing values by considering iterative weights and its computational rapidness.

Besides the original smoothing procedure, several variations have been proposed, namely, the asymmetrically reweighted penalized least squares (arPLS), the adaptive iteratively reweighted penalized least squares (airPLS), and the improved asymmetrically reweighted penalized least squares (iarPLS).³¹ All these variations were compared to define the most suitable one for the developed algorithm. Additionally, the spline version of all the aforementioned methods were also considered and compared. The spline function of a graph is defined by connecting different points of the spectrum with a smooth curve.

Several variations of the work proposed by Whittaker and improved by Eilers were compared as a way of assessing the most suitable variation for the developed methodology. To do so, an open-source library (*pybaselines*) capable of comparing all the versions was used.³² For the implementation of the procedure, a second order matrix was considered and the smoothing parameter of the equations developed by the authors was varied in accordance with the spectrum under processing.^{29, 30}

Figure 1 compares the results for the application of all the variations of baseline corrections to an EDXRF spectrum for two samples. In a) a model sample composed of 10% hydroxyapatite and 90% boric acid (low Z) was considered, and in b) a CRM of phosphate rock (high Z) was analyzed. By comparing both spectra, one can infer that the airPLS method and the spline-penalized air

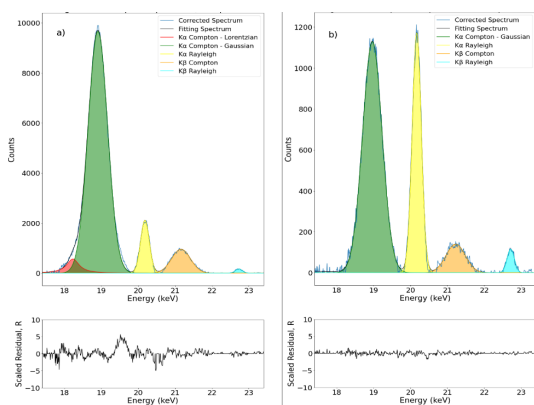


Fig. 2 Peaks fitting and the respective scaled residue for a) 10%hydroxyapatite - 90%boric acid and b) phosphate rock samples.

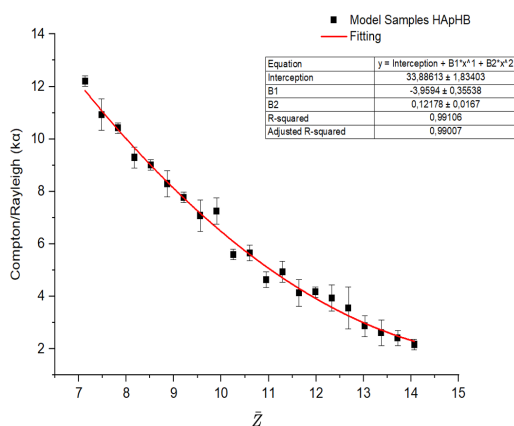


Fig. 3 Calibration curve plotted for the model samples considering the Compton-to-Rayleigh ratios and their respective average atomic values.

PLS provided the worst corrections for both samples. Despite the fact that the spline-penalized airPLS exhibits one of the best fittings for the sample with lower average Z that is not the case to the second sample, in which an overfitting of the $K\beta$ Compton peaks is visible. From all the functions, the spline-penalized version of the arPLS (with $\lambda=100$)^{29, 30} corresponds to the option in which the peaks intensity is better preserved, hence, it was the selected for baseline correction purposes.

Peak fitting. The characteristic peaks and the Rayleigh scattering peaks can often be fitted with a Gaussian function. nonetheless, the Compton scattering generates more complex and asymmetric peaks in the final spectrum whose fitting is often challenging.³³ The asymmetry of the Compton peak can be caused by two phenomena, the multiple dispersion of Compton, and an incomplete charge collection in the detector. In the first phenomenon, if the photons involved in the first dispersion have enough energy available, they can be involved in a second dispersion (second order scattering), resulting in photons in lower energy - more common in samples with lower average Z .³⁴ Regarding incomplete charge collection, a portion of the photons is not considered by the detector, resulting in lower number of

counts in the lower energy side of the peak.³⁵

To mitigate these effects, Lorentzian functions were used. This type of function is known for its long tails, characteristic useful to properly fit the asymmetry of the Rh $K\alpha$ dispersion peak caused by the second order dispersion and by the incomplete charge collection. In this way, a Lorentzian function was added to a Gaussian for purposes of fitting the Rh $K\alpha$ Compton peak. The Rh $K\beta$ peak, due to its less significative asymmetry, was properly fitted with just a Gaussian function³⁶. The Gaussian and Lorentzian functions were defined as follows (A – maximum amplitude of the curve, μ – centroid, FWHM – full width at half maximum):

$$G(x, A, \mu, FWHM) = A \exp \left[-4 \ln 2 \frac{(x-\mu)^2}{FWHM^2} \right] \quad (3)$$

$$L(x, A, \mu, FWHM) = \frac{A}{1 + 4 \frac{(x-\mu)^2}{FWHM^2}} \quad (4)$$

For the calculation of the ratio itself, the intensity of both Compton and Rayleigh peaks were assessed with a specific library (*trapz*) of an open-source programming module (*scipy.integrate*), which uses the composite trapezoidal rule to obtain the peaks area.³⁷

Figure 2 exhibits the fitted curves for two model samples, namely, a 10%hydroxyapatite + 90%boric acid sample and b) phosphate rock sample and the respective scaled residuals. Table S3 summarizes the amplitude (counts), centroid (keV) and FWHM (keV) for both samples. In order to assess the quality of the fitting, two other statistical tools were used to assess the quality of the fitting, the root mean square error (RMSE) and the reduced chi-squared χ^2_{red} also included in Table S4. The RMSE values for both $K\alpha$ and $K\beta$ peaks are considerably lower than the respective amplitudes. This fact allows to infer that the developed model apply a proper fitting to the EDXRF spectra considered during the study. Regarding the values of χ^2_{red} , since most of the values are around 1, the fitting of the peaks was achieved successfully.

Assessment of matrix composition and quantification. In order to perform quantitative analysis and implement the Fundamental Parameter method, the methodology used by Ensina *et al.*³⁸ and by Braga³⁹ was followed: determination of the combination and composition of the elements in the matrix that would exhibit a given average Z .

The average Z values were determined considering the calibration curve plotted using the Rh $K\alpha$ Compton-to-Rayleigh ratios determined for all the model samples and their respective average Z values.

Figure 3 represents the curve plotted with the software

developed in this work, in which each point corresponds to the average value determined from a total of four measurements and the error bars correspond to the average deviation to the mean. The fitted curve results are presented as inset in the figure.

Preliminary quantification. From literature, preliminary 20-30 matrix compositions were tested for each type of CRM. Once the matrix was defined, it was introduced ad-hoc in the MQuant software for purposes of quantifying the trace and minor elements by the Fundamental Parameters method. The data treatment aimed at finding the most suitable matrices which exhibited the best results when compared with the CRMs expected values. To assess that suitability of the different matrices, the weighted percentual deviations ($|\Delta|_{\text{weighted}}\%$) between the concentrations obtained for each matrix and the reference values of the CRMs were estimated as follows:

$$|\Delta|_{\text{weighted}} = \sum_{i=1}^N (|\Delta| * f_i) * 100 = \sum_{i=1}^N \left(\frac{|C_{i\text{obs}} - C_{i\text{ref}}|}{C_{i\text{ref}}} * 100 \right) \quad (5)$$

Here, $|\Delta|$ represents the relative deviation, N is the total number of trace and minor elements, f_i corresponds to the mass fraction of the i element, $C_{i\text{obs}}$ is the concentration of the element i obtained during the quantification and $C_{i\text{ref}}$ is the reference concentration of the element i . The matrices with lower values of $|\Delta|_{\text{weighted}}\%$ and consequently, the best quantifications, were considered for the development of the optimization script.

Algorithm. In order to determine the best combination of elements for the matrix, an algorithm was developed based on two specific tools: support vector regression (SVR) and bootstrapping and the *scikit-learn* open-source library was used during the development of the entire script.⁴⁰

The SVR tool allows to compensate for errors by employing a loss function which combines the errors of the training observations (ϵ), responsible for defining a range in which the errors are disregarded, and a regularization term (C) that controls the complexity of the model. As a result, the final model is considerably smoother and less sensible to small data variations, *i.e.*, it is robust to treat both linear and even nonlinear data.

To treat nonlinear data, as it is the EDXRF data, this machine learning tool uses kernel functions that transform the original dimensional space into a new one with larger dimension where a linear relationship between the data can be found. The nonlinear function can be translated as follows (α_i^* and α_i are Lagrange multipliers, $k(x_i, x)$ is the kernel function, and b is a bias term which allows to adjust the model to the data):

$$f(x) = \sum_{i=1}^N (\alpha_i^* - \alpha_i) k(x_i, x) + b \quad (6)$$

The loss function, L_ϵ , in its turn, can be defined as (y corresponds to the real value):

$$L_\epsilon = f(x) = \begin{cases} 0, & \text{if } |f(x) - y| < \epsilon \\ |f(x) - y| - \epsilon, & \text{other cases} \end{cases} \quad (7)$$

The SVR was implemented by providing the parameters C and ϵ to the *sklearn.svm* module. A specific kernel function, the radial basis function (RBF), was used in the implementation of the software. This variation measures the similarity among two observations considering their Euclidian distance⁴¹. Higher values of the kernel allow to consider a wider variety of arrays and, on the other hand, lower kernel values imply that only close arrays can be considered as similar. The RBF can be represented as follows (x and x' correspond to the characteristics arrays of the input data. and σ is the range of the function that defines the kernel):

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (8)$$

Since the model under development exhibits multiple outputs (all the concentration levels for the several elements existent in the matrix) for a single average Z input, it was necessary to use the *MultiOutputRegressor* class of the *sklearn.multioutput* module, in order to process all the outputs. Then, to define the best values for the parameters C and ϵ , the *RandomizedSearchCV* class of the *sklearn.multioutput* module was used. This class generates random values and evaluates the suitability of each pair until the best parameters are obtained. In this work, $C = 500$ and $\epsilon = 1.3$. Finally, bootstraps datasets of 1000 observations were considered to teach the algorithm. The bootstrapping generates multiple bootstrap samples from a single original sample by randomly selecting, with replacement, n data points, where n is the size of the original dataset.⁴² These bootstrap samples, were generated using the *resample* function of the module *sklearn.utils* and then used to train the SVR algorithm. Fig. 4 describes the implemented workflow.

This approach effectively implements an ensemble learning technique where bootstrapping is used to enhance the generalization of the SVR. All the repeated matrices were disregarded.

RESULTS AND DISCUSSION

The obtained results will be discussed in different sections considering the type of matrix.

Soft tissue and blood samples have a dark matrix composed of elements like hydrogen, carbon, nitrogen, and oxygen, which explains the low levels of average Z for these CRMs. To obtain the

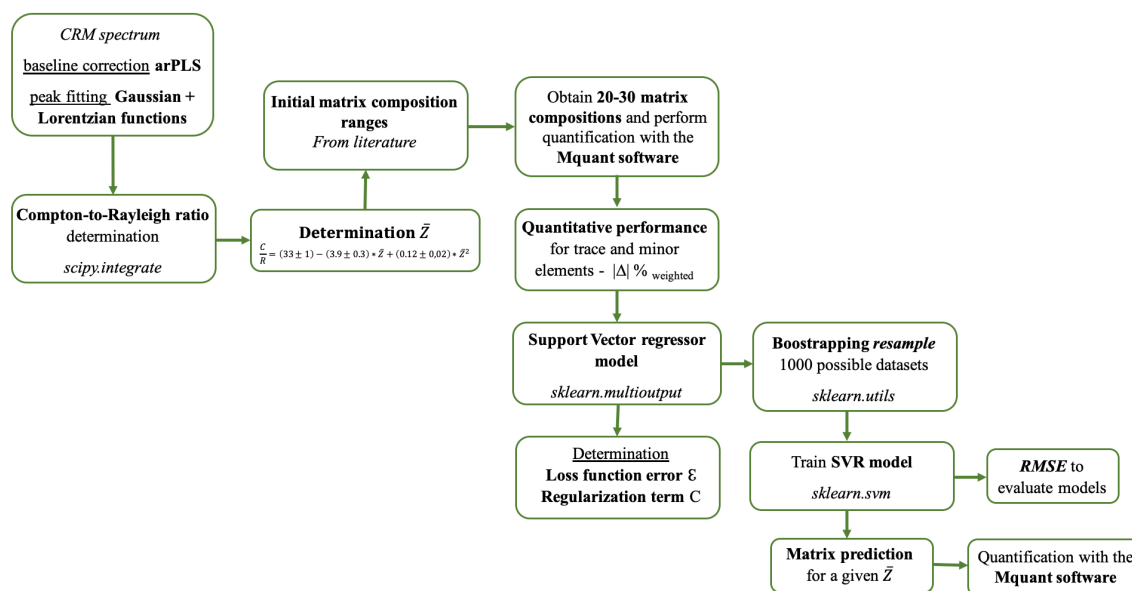


Fig. 4 Schematic representation of the workflow.

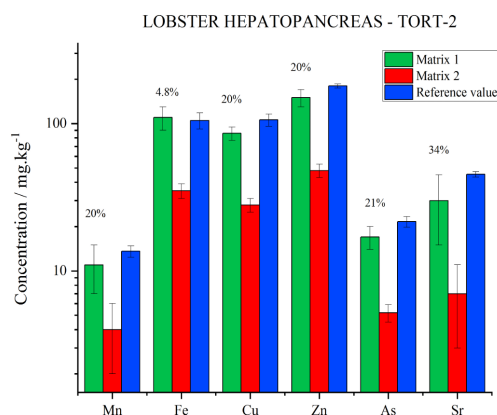


Fig. 5 Elemental concentrations (mg.kg⁻¹) obtained with both 1 and 2 matrices for the CRM lobster hepatopancreas (TORT-2), compared to the reference values. Error bars correspond to the Standard Deviation of the results for the 8 measured spectra. There is also presented the weighted deviation between the best matrix and the certified value.

reference concentrations of these elements in CRMs, a literature search was conducted for the marine life CRMs (IAEA MA-A-2, DORM-4, TORT-2, NIST-1566b).^{34–36} Since the CRMs animal blood (IAEA-A-13), bovine liver (NIST-1577a) and pig kidney (ERM-BB186) are of mammal origin, it was expected that they exhibited concentration levels similar to the ones found in the human organism, similar to the matrixes tested in Ensina *et al.*³⁸ The considered concentration ranges were 1 – 10% (H), 15 – 50% (C), 1 – 20% (N), and 20 – 70% (O). These concentration levels may vary depending on the type of tissue, in this way, the values

available in the literature may not reflect their exact concentration, nonetheless, the differences are not significative.

Then, the algorithm developed for purposes of creating the different combinations of elements in regard to the matrix composition of each CRM was applied considering the concentration ranges and the average atomic numbers of reference.

After applying the algorithm, 10 different combinations of elements were tested for each CRM. Additionally, normalized and non-normalized versions of the FP-method were compared, being that the non-normalized values exhibited better results, meaning that some of the elements of the matrix can be present in higher concentrations than expected or there could be a missing element.

Considering all the CRMs, the matrices with better (matrix 1) and worst (matrix 2) quantifications are presented Table S3. The concentrations of the main elements are shown, for the CRM lobster hepatopancreas (TORT-2), in Fig. 5 (the remaining results are present in Table S5).

By analyzing the results, one can conclude that matrix 1 allows more accurate quantifications when compared with matrix 2, nonetheless, for some elements, the obtained and certified values do not converge even considering the uncertainties. Considering all the CRMs, more accurate quantifications were achieved for the elements K, Ca, Fe, Cu, Zn, and Mn - if present at higher concentrations. Additionally, lighter elements like P, S and Cl and also As and Mn if present at lower concentrations, exhibit quantifications less accurate (*i.e.* with weighted deviations of

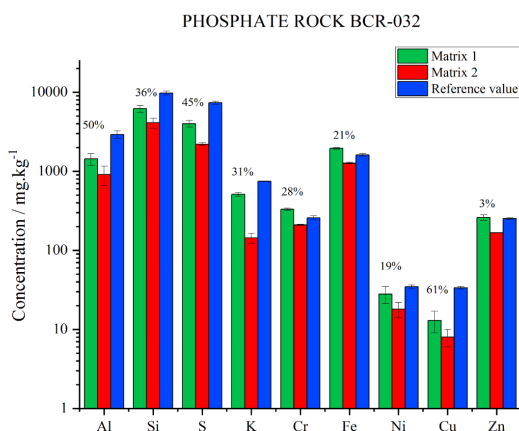


Fig. 6 Elemental concentrations (mg.kg^{-1}) obtained with both 1 and 2 matrices for the CRM phosphate rock (BCR-032) compared to the reference values. Error bars correspond to the Standard Deviation of the results for the 8 measured spectra. There is also presented the weighted deviation between the best matrix and the certified value.

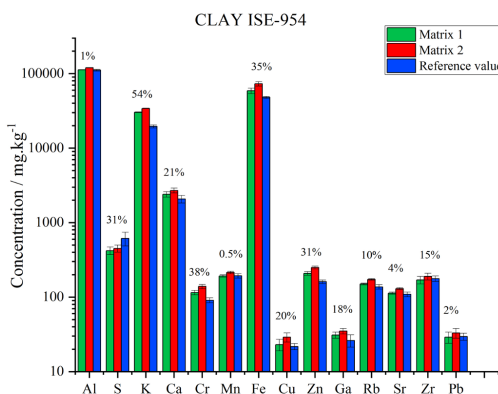


Fig. 7 Elemental concentrations (mg.kg^{-1}) obtained with both 1 and 2 matrices for the CRM clay (ISE-954) compared to the reference values. Error bars correspond to the Standard Deviation of the results for the 8 measured spectra. There is also presented the weighted deviation between the best matrix and the certified value.

around 20%). Regarding Br and Sr, lower accuracy can also be due to the thickness of the pellets, not fully complying with the infinitely thick regime.^{43,8, 16}

Geological samples have elemental compositions that are considerably different from soft tissues since they are often composed of heavier elements and, consequently, have a higher average atomic number. Phosphate rock, for example, is mostly composed of S, O, P, and Ca, nonetheless, the contribution of sulfur was disregarded since it exhibited a very low quantity in the sample. As for the previous cases, the concentration ranges were defined based on the literature; they are 30 – 60% (O), 10 – 20% (P), and 30 – 50% (Ca).⁴⁴

Differently from the previous case studies, it was not possible to

consider a matrix solely containing invisible elements like oxygen and performing an automatic quantification of P and Ca later in the processing. This way, it was necessary to add these two elements and find a matrix that complied with the already mentioned criteria.

The matrices with better (matrix 1) and worst (matrix 2) quantifications are presented in Table S6, while the quantification results achieved are presented in the bar charts of Fig. 6 and compared with the reference values.

Despite still existing discrepancies between the best matrix and the reference values for lighter elements, the selected matrix was the one that allow to minimize them. The remaining elements exhibit quantification values similar to the reference values proving the suitability of the methodology. The exception would be Cu, but this element is present in quantities very close to the limit of detection (LOD).

Regarding composition, the studied samples are composed by a variety of elements among which one can highlight Al, Ca, K, and Fe. The high quantities of these elements in the samples as well as their dispersion between distinct samples impose several challenges in the definition of a proper matrix for purposes of quantification that complies with all the criteria. Nonetheless, three main elements were verified to be always present in this kind of samples: C, O and Si and by consulting the literature it was possible to define concentration ranges for these elements, from 1 – 45% (C), 10 – 50% (O), and 10 – 35% (Si).⁴⁵

Similarly, to the case study involving phosphate rock, the removal of Si from the matrix worsened the results. This way, this element was considered during the selection of the most suitable matrix. Considering all the CRMs, the matrices with better (matrix 1) and worst (matrix 2) quantifications are presented Table S7. The concentrations of the main elements are shown, for the CRM Clay (ISE-954), in Fig. 7 (the remaining results are present in Table S8). Ni in the clay (ISE-954) and the Cu in the soil (IAEA-SOIL-7) CRMs exhibit the worst results. This can be justified by their considerably lower concentrations when compared with the remaining elements. For the San Joaquin soil (SRM-2709a) CRM, the elements K, Ca, and Fe were overestimated, while the element K was underestimated for the river sediment (NIST-1645) CRM.

Mostly composed of H, C, N, and O a reference material of wood was also included in this study to assess the robustness of the developed protocol. As for previous cases, the theoretical concentration ranges for the elements were gathered from the literature and varied from 1 – 10% (H), 40 – 55% (C), 30 – 45% (O) and 0 – 4% (N).⁴⁶

Similarly to the previous cases, several elemental compositions were obtained for the potential composition of these CRMs matrices, considering the Z of each element and the reference

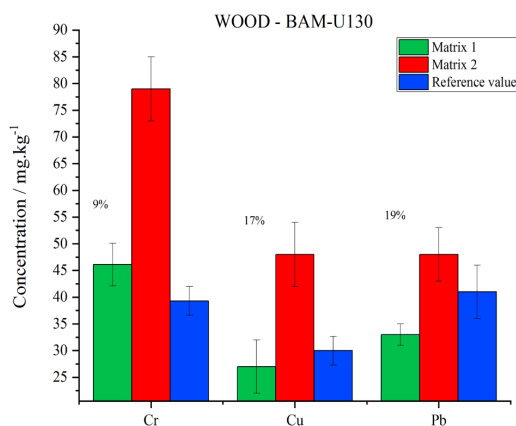


Fig. 8 Elemental concentrations (mg.kg^{-1}) obtained with both 1 and 2 matrices for the CRM wood (BAMU130) compared to the reference values. Error bars correspond to the Standard Deviation of the results for the 8 measured spectra. There is also presented the with the weighted deviation between the best matrix and the certified value, as well as, the respective reference values.

average Z of the sample. The selected matrices were the ones that exhibited lower percentual deviations and whose average atomic number was closer to the reference value (Table S9).

The quantification results achieved for the selected matrix during the statistical processing of the CRM of wood, as well as the reference values and the $|\Delta|_{\text{weighted}}\%$ were plotted in the graph of the Fig. 8.

Interestingly, the quantification of Pb exhibited the worst results for the wood while Cr was the element with better results since it showed lower percentual deviation when compared to the other two elements, so it is suitable to say that the achieved results are accurate, proving the potential of the developed methodology to also quantify elements in wood-based samples.

CONCLUSION

This automated and non-destructive approach integrates open-source computational tools and algorithms into a reliable workflow for elemental analysis. The combination of spectral analysis and matrix composition quantification has proven its potential to streamline workflows and enhance accuracy across diverse applications, including environmental, biomedical, and industrial contexts. The developed methodology for matrix determination using EDXRF spectra has proven to be an effective and versatile solution to address quantification, especially when suitable calibration curves cannot be developed. By leveraging the analysis of Compton and Rayleigh scattering peaks, combined with computational tools like support vector regression (SVR) and bootstrapping, the approach enabled the determination of the

matrix composition with enhanced reliability and precision. The results demonstrated strong agreement with reference values, particularly for heavier elements such as Fe, Cu, and Zn, which showed minimal deviations across biological, geological, and wood-based samples. Similarly, in geological samples such as phosphate rock and clay, heavier elements aligned well with reference values, but trace elements like Cu and Ni showed larger deviations, likely due to their low concentrations relative to the sample matrix. In wood samples, discrepancies were observed for certain elements, such as Pb, but the methodology performed well for Cr, highlighting its adaptability to different matrix types.

While the proposed methodology demonstrates strong potential, certain areas highlight room for improvement, offering opportunities to refine its accuracy and versatility. One notable challenge lies in the quantification of lighter elements, such as Mn, Sr, P, S, and Cl, particularly in biological matrices. These elements often exhibit deviations around 20%, likely due to their low atomic numbers and fluorescence yields. Addressing this requires enhanced spectral deconvolution techniques (the MQuant software was used and shows clear limitations) and the use of high-resolution detectors, namely microcalorimeter detectors, that could improve the sensitivity and separation of these elements, minimizing deviations.

Expanding the training dataset to include a broader range of certified reference materials (CRMs) would enhance its robustness. Combining the current SVR model with other machine learning techniques, such as neural networks, may also improve adaptability.

Despite some discrepancies for some trace elements, the developed methodology is a significant step forward in the field of EDXRF analysis, offering an efficient and robust solution for quantifying complex matrices.

ASSOCIATED CONTENT

The supporting information (Tables S1–S9) is available at www.at-spectrosc.com/as/home.

AUTHOR INFORMATION



Sofia Pessanha obtained her PhD degree in Physics at the Faculty of Sciences, University of Lisbon in December 2013 and is an Assistant Researcher at the Physics Department of NOVA School of Science and Technology since 2020. She is an expert in Raman and X Ray Fluorescence spectroscopy and is a member of the Advisory Board of X Ray Spectrometry

journal and Editor for X Ray Spectrometry in Encyclopedia of Analytical Chemistry. Her research focus lies in refining methodologies for determining biomarkers in human diseases and assessing the impact of environmental exposures on both biomarkers and human health. She has co-authored over 100 scientific publications and presented over 35 communications in international conferences: she was invited speaker at the main conferences X ray Spectroscopy: European X-Ray Spectrometry conference (Belgium, 2022), Denver X-ray Conference (USA, 2018) and Latin American Conference of Analysis by X-ray Techniques (Chile, 2018 & Brazil, 2024).

Corresponding Author

* S. Pessanha

Email address: sofia.pessanha@fct.unl.pt

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been financially supported by Fundação para a Ciência e a Tecnologia through the LIBPhys funding UID/FIS/04559/2020 and GeoBioTec funding UIDP/GEO/04035/2020. CEECIND/00278/2018/CP1564/CT0007, <https://doi.org/10.54499/CEECIND/00278/2018/CP1564/CT0007> and the project ERAMIN3/0008/2021, FCT/0008/PG2CRM: Phosphogypsum Processing to Critical Raw Materials.

REFERENCES

1. V. M. Chubarov, A. S. Maltsev, A. A. Amosova, E. V. Chuparina, S. N. Prosekin, and J. V. Sokolnikova, *Xray Spectrom.*, 2024, **53**, 374–381. <https://doi.org/10.1002/xrs.3384>
2. V. K. Singh, N. Sharma, and V. K. Singh, *Xray Spectrom.*, 2022, **51**, 304–327. <https://doi.org/10.1002/xrs.3260>
3. D. S. Almeida, M. M. Brígido, M. J. Anjos, S. T. Ferreira, A. S. Souza, and R. T. Lopes, *Xray Spectrom.*, 2019, **48**, 452–464. <https://doi.org/10.1002/xrs.3044>
4. A. K. Maurya and P. Deb Barman, *Xray Spectrom.*, 2024, **54**, 38–46. <https://doi.org/10.1002/xrs.3422>
5. C. A. M. Dingle, J. F. M. Jecong, F. C. Hila, M. E. S. V. Ramo, N. R. D. Guillermo, M. R. Vasquez and V. A. I. Samson, *Xray Spectrom.*, 2019, **48**, 513–521. <https://doi.org/10.1002/xrs.3029>
6. T. Zielenkiewicz, J. Zawadzki, and A. Radomski, *Xray Spectrom.*, 2012, **41**, 371–373. <https://doi.org/10.1002/xrs.2416>
7. J. Romero-Pastor, A. García-Porras, R. Van Grieken, S. Potgieter-Vermaak, J. Coll-Conesa, and C. Cardell, *Xray Spectrom.*, 2015, **44**, 426–435. <https://doi.org/10.1002/xrs.2613>
8. Rafał Sitko and Beata Zawiska. Quantification in X-Ray Fluorescence Spectrometry, X-Ray Spectroscopy, Dr. Shatendra K Sharma (Ed.) 2012
9. A. G. Revenko and G. V. Pashkova, X-Ray Fluorescence Spectrometry: Current Status and Prospects of Development, Pleiades Publishing, 2023
10. C. Bowers, *J. Chem. Educ.*, 2019, **96**, 2597–2599. <https://doi.org/10.1021/acs.jchemed.9b00630>
11. S. Pessanha, C. Fonseca, J. P. Santos, M. L. Carvalho, and A. A. Dias, *Xray Spectrometry*, 2018, **47**, 108–115. <https://doi.org/10.1002/xrs.2819>
12. I. Otel, K. Dias, R. Pereira, M. Fonseca, A. P. Jesus, A. Mata, V. Vassilenko, J. M. Silveira and S. Pessanha, *J. Trace Elem. Med. Biol.*, 2022, **71**, 126938. <https://doi.org/10.1016/j.jtemb.2022.126938>
13. P. Kump, *Xray Spectrometry*, 2024, **53**, 16–26. <https://doi.org/10.1002/xrs.3352>
14. J. Sherman, *Spectrochim. Acta*, 1955, **7**, 283–306. [https://doi.org/10.1016/0371-1951\(55\)80041-0](https://doi.org/10.1016/0371-1951(55)80041-0)
15. P. M. Wróbel, S. Bała, M. Czyżycki, M. Golasik, T. Librowski, B. Ostachowicz, W. Piekoszewski, A. Surówka, and M. Lankosz, *Talanta*, 2017, **162**, 654–659. <https://doi.org/10.1016/j.talanta.2016.10.043>
16. M. Szczerbowska-Boruchowska, *Xray Spectrom.*, 2012, **41**, 328–337. <https://doi.org/10.1002/xrs.2407>
17. I. Purwadi, L. W. Casey, C. G. Ryan, P. D. Erskine, A. van der Ent, *Plant Methods*, 2022, **18**, 139–156. <https://doi.org/10.1186/s13007-022-00958-z>
18. S. Lagomarsino, S. Iotti, G. Farruggia, A. Cedola, V. Trapani, M. Fratini, I. Bukreeva, A. Notargiacomo, L. Mastrotoaro, C. Marraccini, A. Sorrentino, I. McNulty, S. Vogt, D. Legnini, S. Kim, A. Gianoncelli, J. A. M. Maier, and F. Wolf Minotti, *Spectrochim. Acta B*, 2011, **66**, 834–840. <https://doi.org/10.1016/j.sab.2011.11.00>
19. S. Pessanha, A. Guilherme, and M. L. Carvalho, *Appl. Phys. A-Mater.*, 2009, **97**, 497–505. <https://doi.org/10.1007/s00339-009-5251-x>
20. M. Wang, Y. Gu, H. Lu, L. Ge, Q. Zhang, and G. Zeng, *Spectrochim. Acta B*, 2022, **193**, 106438. <https://doi.org/10.1016/j.sab.2022.106438>
21. A. Rindby, *Xray Spectrometry*, 1989, **18**, 113–118. <https://doi.org/10.1002/xrs.1300180308>
22. R. Williams, G. Taylor, and C. Orr, *Archaeometry*, 2020, **62**, 1145–1163. <https://doi.org/10.1111/arc.12583>
23. S. Wolfs and R. Dekkers, *Xray Spectrom.*, 2023, **52**, 52–61. <https://doi.org/10.1002/xrs.3315>
24. A. Aldrian, A. Ledersteger, and R. Pomberger, *Waste Manage.*, 2015, **36**, 297–304. <https://doi.org/10.1016/j.wasman.2014.10.025>
25. M. Qu, J. Chen, W. Li, C. Zhang, M. Wan, B. Huang, and Y. Zhao, *Env. Pollution*, 2019, **254**, 112993. <https://doi.org/10.1016/j.envpol.2019.112993>
26. Y. Sapkota, B. L. Drake, L. M. McDonald, T. C. Griggs, and T. J. Basden, *J. Environ. Qual.*, 2020, **49**, 472–482. <https://doi.org/10.1002/jeq2.20013>
27. S. Pessanha, S. Silva, L. S. Martins, J. P. Santos, and J. M. Silveira, *J. Anal. At. Spectrom.*, 2019, **34**, 854–859. <https://doi.org/10.1039/C8JA00370J>
28. S. Pessanha, A. Samouco, R. Adão, M. L. Carvalho, J. P. Santos, and P. Amaro, *Xray Spectrom.*, 2017, **46**, 102–106. <https://doi.org/10.1002/xrs.2737>
29. E. T. Whittaker, *P. Edinburgh Math. Soc.*, 1922, **41**, 63–75. <https://doi.org/10.1017/S0013091500077853>
30. P. H. C. Eilers, 2003, *Anal. Chem.*, 2003, **75**, 3631–3636. <https://doi.org/10.1021/ac034173t>
31. Ł. Górski and M. Jakubowska, *Chemometr. Intell. Lab. Syst.*, 2023, **238**, 104848. <https://doi.org/10.1016/j.chemolab.2023.104848>

32. E. Donald, pybaselines: A Python library of algorithms for the baseline correction of experimental data. <https://zenodo.org/records/10676584>
33. M. Van Gysel, P. Lemberge, and P. Van Espen, *Xray Spectrom.*, 2003, **32**, 139–147. <https://doi.org/10.1002/xrs.628>
34. B. P. E. Tee, B. Ganly, J. D. McIlquham, P. Giang, and Y. Van Haarlem, *Xray Spectrom.*, 2025, **54**, 133–141. <https://doi.org/10.1002/xrs.3441>
35. B. Ganly, Y. Van Haarlem, and J. Tickner, *Xray Spectrom.*, 2017, **46**, 336–340. <https://doi.org/10.1002/xrs.2756>
36. G. H. Major, V. Fernandez, N. Fairley, and M. R. Linford, *Surf. Interface Anal.*, 2022, **54**, 262–269. <https://doi.org/10.1002/sia.7050>
37. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G. L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko and Y. Vázquez-Baeza, *Nat. Methods*, 2020, **17**, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
38. A. Ensina, P. M. Carvalho, J. Machado, M. L. Carvalho, D. Casal, D. Pais, J. P. Santos, A. A. Dias, and S. Pessanha, *J. Trace Elem. Med. Bio.*, 2021, **68**, 126837. <https://doi.org/10.1016/j.jtemb.2021.126837>
39. S. Pessanha, D. Braga, A. Ensina, J. Silva, J. Vilchez, C. Montenegro, S. Barbosa, M. L. Carvalho and A. Dias, *Talanta*, 2023, **260**, 124605. <https://doi.org/10.1016/j.talanta.2023.124605>
40. F. Pedregosa, V. Michel, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, J. Vanderplas, D. Cournapeau, F. Pedregosa, G. Varoquaux, A. Gramfort, B. Thirion, O. Grisel, V. Dubourg, A. Passos, M. Brucher, M. Perrot, É. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830. <https://doi.org/10.48550/arXiv.1201.0490>
41. X. Ding, J. Liu, F. Yang, and J. Cao, *J. Franklin I.*, 2021, **358**, 10121–10140. <https://doi.org/10.1016/j.jfranklin.2021.10.005>
42. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning*, Springer Texts in Statistics, 2013, New York, USA. <https://link.springer.com/book/10.1007/978-1-0716-1418-1>
43. S. Pessanha, A. Veiga, D. Douteil, F. Silva, J. Silva, P. M. Carvalho, S. Barbosa, J. P. Santos, A. Félix, J. Machado, *Spectrochim. Acta B*, 2023, **205**, 106704. <https://doi.org/10.1016/j.sab.2023.106704>
44. U. Ryszko, P. Rusek, and D. Kołodyńska, *Materials*, 2023, **16**, 793–807. <https://doi.org/10.3390/ma16020793>
45. D. Dent, B. P. Boincean, and I. A. Krupenikov, in *The Black Earth*, Springer, New York, 2011. <https://www.nhbs.com/the-black-earth-book>
46. R. C. Pettersen, in *The chemistry of solid wood*, ACS publications, Washington, 1984. <https://doi.org/10.1021/ba-1984-0207.ch002>